A system-wide anonymity metric with message multiplicities

Rajiv Bagai*

Department of Electrical Engineering and Computer Science,
Wichita State University,
Wichita, KS 67260-0083, USA
Fax: +1-316-978-3984
E-mail: rajiv.bagai@wichita.edu
*Corresponding author

Bin Tang

Computer Science Department,
California State University,
Dominguez Hills,
Carson, CA 90747, USA
E-mail: btang@csudh.edu

Ahsan A. Khan and Abdus Samad

Juniper Networks,
Sunnyvale, CA 94089, USA
E-mail: ahsankhan@juniper.net
E-mail: asamad@juniper.net

Abstract: An important goal of an anonymity system is to hide the *level* of communication between two users, i.e., the number of messages one user sent to another. We construct an anonymity metric that measures the extent to which the system-wide communication pattern of all users of an anonymity system is hidden, among other possible patterns, in the aftermath of an attack. Our model allows users to send or receive multiple messages via the anonymity system, and our metric handles attacks that are capable of determining infeasibility of some of the system's input-output message combinations. We also analyse two earlier attempts in the literature at arriving at such a metric, and show that one of those attempts is limited to only a small class of such attacks, while the other fails to take any attack into account. In comparison, our metric is comprehensive, as it is applicable to all attacks mentioned above.

Keywords: anonymity; privacy; measuring anonymity; anonymity metric; infeasibility attack; graph theory; combinatorial matrix theory; Shannon entropy.

Reference to this paper should be made as follows: Bagai, R., Tang, B., Khan, A.A. and Samad, A. (2015) 'A system-wide anonymity metric with message multiplicities', *Int. J. Security and Networks*, Vol. 10, No. 1, pp.20–31.

Biographical notes: Rajiv Bagai received his BS in Computer Science from the Birla Institute of Technology and Science (BITS), Pilani, India, and his MS and PhD in Computer Science from the University of Victoria, Canada. He is presently an Associate Professor in the Department of Electrical Engineering and Computer Science at the Wichita State University, USA. His current research area is web anonymity, but in the past he has worked in logic programming and paraconsistent databases.

Bin Tang received a BS in Physics from Peking University, China, in 1997, an MS in Materials Science, and MS and PhD in Computer Science from Stony Brook University in 2000, 2002 and 2007, respectively. He is currently an Assistant Professor in the Computer Science Department at the California State University, Dominguez Hills, USA. His research interests include algorithmic aspects of data intensive sensor networks.

Ahsan A. Khan received his BE in Computer and Information Systems Engineering from the NED University of Engineering and Technology, Karachi, Pakistan, and his MS in Electrical Engineering

with major in Computer Networks from the Wichita State University, USA. His current research area is web anonymity.

Abdus Samad received his BS in Computer Science and Information Technology from the NED University of Engineering and Technology, Karachi, Pakistan, and his MS in Computer Networking from the Wichita State University, USA. His research interests include information security, privacy and anonymity.

1 Introduction

Accurate measurement of the degree of anonymity remaining in an anonymity system upon conclusion of an attack has been a fundamental problem in the area of computer communication privacy. Most anonymity metrics that have been proposed for this task, such as the ones by Serjantov and Danezis (2002), and Diaz et al. (2002), measure anonymity from the point of view of a single message of the system. Edman et al. (2007), on the other hand, gave a *system-wide* metric for measuring an attacker's uncertainty in linking all incoming messages of a system with its outgoing ones. They modelled possible associations between these messages as edges of a complete bipartite graph, whose vertices represent the system's input and output messages. An attack results in labelling some edges of this graph as infeasible, thus also some perfect matchings between input and output messages. Anonymity in this framework is measured by comparing the number of matchings that remain feasible, after the attack, to that of all possible ones.

This metric of Edman et al. (2007) was reviewed by Gierlichs et al. (2008), who argued for the need to go a step further and measure the attacker's uncertainty in linking the system's users, i.e., *senders* and *receivers*, as against just *messages*. Indeed, as the attacker's ultimate goal is to uncover the communication pattern between users, this extra step is essential. Gierlichs et al. (2008) showed that when users send/receive multiple messages, an equivalence relation $\sim$ is induced on the set of all feasible perfect matchings, any equivalence class of which essentially represents a feasible communication pattern between the system's users. The class sizes lead to the attacker's probability distribution over such communication patterns, from which the anonymity of the actual pattern is measured by the well-known Shannon-entropy technique of Diaz et al. (2002). A method for computing the all important class sizes is also presented in Gierlichs et al. (2008).

Contributions of our paper are two-fold. Our minor contribution is to show that the method of Gierlichs et al. (2008) for computing sizes of equivalence classes of $\sim$ is limited to a small family of attacks. We accurately characterise this attack family, and show that it shrinks in size as the number of messages sent/received by users grows. Our major contribution is an improved technique for arriving at the attacker's probability distribution over feasible communication patterns. The improvement achieved by our technique is in its generality, as it works for all possible attacks.

We depart from the approach of Gierlichs et al. (2008) by considering the equivalence relation $\bowtie$ on the set of *all* perfect matchings of the complete bipartite graph of Edman et al. (2007). Equivalence classes of $\bowtie$ represent all possible communication patterns and, as we show, are aptly denoted by nonnegative integer matrices, whose individual rows add up to numbers of messages sent by the senders, and columns to numbers of messages received by the receivers. When feasibility information of perfect matchings, resulting from any attack, is also taken into account, it becomes necessary to determine the number of *feasible* perfect matchings in classes associated with such matrices. We present a recursive method to compute these values, and a system-wide anonymity metric based on them to measure the anonymity of the communication pattern between the system's users.

The rest of this paper is organised as follows. Section 2 gives an overview of the model of the anonymity system considered by Edman et al. (2007), some example attacks on such a system, and the basic anonymity metric proposed by them. Section 3 extends this model by adding users that may send and/or receive multiple messages, and introduces the nonnegative integer matrices that denote equivalence classes of $\bowtie$. It then presents our anonymity metric, which is based on the number of feasible perfect matchings in these classes. A combinatorial technique for computing this number is also developed. This technique works for any attack. This section ends with a complexity analysis of our metric. Section 4 studies two alternative approaches proposed in the literature for solving this problem. It shows that the method of Gierlichs et al. (2008) is limited to a small family of attacks, and characterises that family. Another method, proposed later by Grégoire and Hamel (2009), does not consider any attack, and is thus shown to just determine the *number* of equivalence classes of $\bowtie$. Finally, Section 5 concludes our work and gives some directions for future work.

2 An existing system-wide anonymity metric

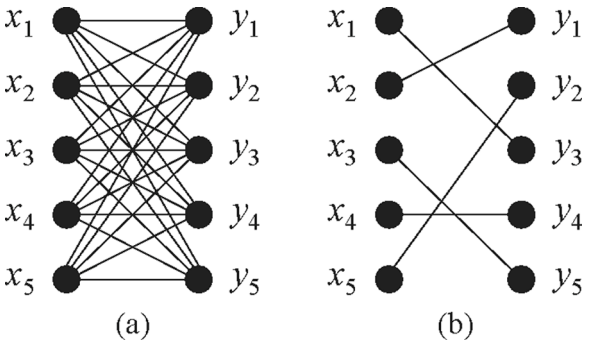
In this section we give an overview of the anonymity metric proposed by Edman et al. (2007). Their metric gives a *system-wide* measure of the anonymity provided to the messages sent via an anonymity system, rather than to any *single* message going through it.

2.1 Model of the anonymity system

Let X be the set of t messages observed by an attacker having entered an anonymity system, and Y be the set of messages observed by the attacker having exited from that system. Under the assumption that the system neither destroys nor generates any messages of its own, we have that $|X| = |Y| = t$. The anonymity system attempts to hide from the

attacker which input message in X exited the system as which output message in Y. It may employ a number of techniques to conceal this association, such as modifying message encoding by encryption/decryption to prevent straightforward message bit-pattern comparison, or sending messages out in an order other than the one in which they arrived to prevent sequence number association, etc. The maximum anonymity this system can strive to achieve is when for any particular input message $x \in X$, each of the output messages in Y appears to be a feasible candidate for being the one that x exited the system as. This situation is depicted by the complete bipartite graph $K_{t,t}$ between X and Y, as shown in Figure 1(a) for $t = 5$. Any edge $\langle x_i, y_j \rangle$ in this graph indicates that the incoming message x_i could possibly have been the outgoing message y_j.

Figure 1 (a) Complete anonymity and (b) an instance of no anonymity



An attacker, on the other hand, attempts to eliminate as many edges as possible from the complete bipartite graph of Figure 1(a), due to their infeasibility concluded from the attack. After a completely successful attack, for each outgoing message $y \in Y$, the attacker would have identified exactly one possible incoming message that could have exited the system as y. In other words, the attacker would have obtained a *perfect matching* between the input and output messages of the system. In this case, the system is thus considered to provide no anonymity. There are $t!$ possible perfect matchings between X and Y, an arbitrary one of which is shown in Figure 1(b).

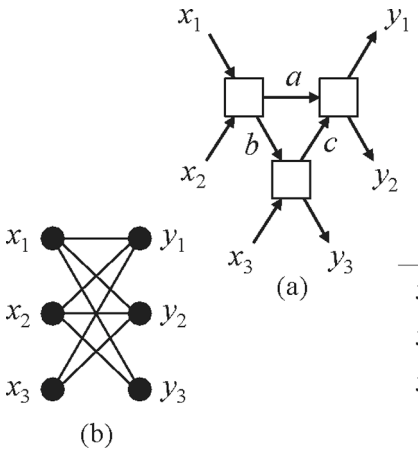
Figures 1(a) and (b) correspond to the two extreme situations, namely complete anonymity and no anonymity at all. In general, after having detected some input-output pairings as infeasible, an attack would result in a bipartite graph that lies somewhere in between these two extreme ends. Exactly which edges of the complete bipartite graph are missing from the graph resulting from the attack will depend upon how much information is available to that attack.

2.2 Some attack examples

We now look at two examples of attacks on such anonymity systems.

Example 1: If messages passing through a system are not padded to become of the same size, then message sizes can clearly be used to rule out input-output message pairings that are of incompatible sizes. Anonymity systems are therefore recommended to pad their messages to become of equal size. However, systems in which all messages are of the same size are then constrained to have some maximum route length for messages, due to the fact that each message (upon leaving the sender) contains in it the addresses of all proxy servers it will go via. Serjantov and Danezis (2002) present an attack that exploits knowledge of this maximum route length to render certain input-output pairings infeasible. Figure 2(a) shows an anonymity system with three mix nodes, $X = \{x_1, x_2, x_3\}$, and $Y = \{y_1, y_2, y_3\}$. Messages a, b, and c are within the system.

Figure 2 (a) Route length attack; (b) graph resulting from this attack and (c) biadjacency matrix of this graph



	y_1	y_2	y_3
x_1	1	1	1
x_2	1	1	1
x_3	1	1	0

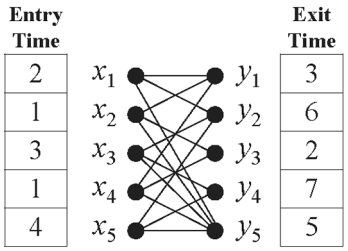
(c)

Suppose the maximum route length for messages in this system is 2, i.e., any x_i can pass through at most 2 mix nodes. If an attacker knows the maximum route length of this system, and can observe messages entering and leaving *each* mix node, then he can infer that message c must be x_3, because if it were either x_1 or x_2 the route length condition is violated as then that message would have passed through 3 mix nodes. Therefore, y_3 cannot be x_3. Figure 2(b) shows the graph resulting from this analysis, from which the edge $\langle x_3, y_3 \rangle$ is thus absent. The *biadjacency matrix* of this graph, a 0-1 matrix with a row for each input message and a column for each output message, is given in Figure 2(c). □

Example 2: As another example of an attack on such systems, consider an attacker that notes the times at which messages enter and exit the system, and uses its knowledge of the minimum and maximum latency of messages in the system.

In this example, suppose each message entering the system always comes out after a delay of between 1 and 4 time units, and this characteristic of the system is known to the attacker. If 5 messages enter and exit this system at times shown in Figure 3(a), then x_1 must be either y_1, y_2, or y_5, because the other outgoing messages, namely y_3 and y_4, are outside the possible latency window of x_1. Similar reasoning can be performed on all other messages to arrive at the reduced graph produced by this attack, shown in Figure 3(a). Note that in this graph x_1 is connected to only y_1, y_2, and y_5, and not to y_3 or y_4, since the edges $\langle x_1, y_3 \rangle$ and $\langle x_1, y_4 \rangle$ were determined by the attack to be infeasible. Figure 3(b) shows the biadjacency matrix of this graph. □

Figure 3 (a) Graph resulting from a timing analysis attack, for a system with latency between 1 and 4 time units and (b) biadjacency matrix of this graph



(a)

	y_1	y_2	y_3	y_4	y_5
x_1	1	1	0	0	1
x_2	1	0	1	0	1
x_3	0	1	0	1	1
x_4	1	0	1	0	1
x_5	0	1	0	1	1

(b)

2.3 The existing metric

The number of perfect matchings between the system's input and output messages allowed by the bipartite graph resulting from an attack indicates the level of anonymity left in the system after the attack. It is well known (see, for example, Asratian et al. (1998)) that this number is the same as the permanent of the biadjacency matrix of that graph. The *permanent* of any $k \times k$ matrix M of real numbers is defined as:

$$\text{per}(M) = \sum_{\phi \in \Phi_k} M_{1\phi(1)} M_{2\phi(2)} \cdots M_{k\phi(k)},$$

where Φ_k is the set of all bijections $\phi : \{1, 2, \ldots, k\} \to \{1, 2, \ldots, k\}$, i.e., permutations of the first k positive integers. It can be seen that the graph of Figure 3(a) allows 12 perfect matchings, and that is also the permanent of its biadjacency matrix in Figure 3(b).

Given any $t \times t$ bipartite graph G resulting from an attack, it is assumed that G contains at least one perfect matching between the input and output messages, the one that corresponds to the true communication pattern. The minimum value of the permanent of its biadjacency matrix A is thus 1, when A contains exactly one 1 in each of its rows and columns. In this case, the system is considered to provide no anonymity as the attacker has identified the actual perfect matching, by ruling out all others. The largest number of perfect matchings in G is $t!$, when G is the complete bipartite graph $K_{t,t}$. Therefore, the maximum value of $\text{per}(A)$ is $t!$, when all entries in A are 1. In this case, the system is considered to provide maximum anonymity as the attacker has been unable to rule out any perfect matching as being the actual one.

Definition 1 (Edman et al., 2007)**:** Define a system's *degree of anonymity* after an attack that results in a $t \times t$ biadjacency matrix A as:

$$d(A) = \begin{cases} 0 & \text{if } t = 1, \\ \dfrac{\log(\text{per}(A))}{\log(t!)} & \text{otherwise.} \end{cases} \qquad \square$$

The above anonymity metric compares the number of perfect matchings deemed feasible by the attack with their maximum number. Note that $0 \leq d(A) \leq 1$. Also, $d(A) = 0$ iff A has just one perfect matching, i.e., the system provides no anonymity, and $d(A) = 1$ iff $t > 1$ and A has $t!$ perfect matchings, i.e., full anonymity.

Example 3: The matrices of Figures 2(c) and 3(b) contain 4 and 12 perfect matchings, out of the 3! and 5! maximum possible matchings, respectively. By the above metric, the degrees of anonymity of these systems after their respective attacks are log(4) / log(3!) $\approx$ 0.774, and log(12) / log(5!) $\approx$ 0.519. $\square$

3 A new model with message multiplicities

The metric of Definition 2.3 measures the extent to which the anonymity system, upon conclusion of the attack, is still hiding the actual communication pattern, represented by a *unique* perfect matching between the system's input and output messages, among all such perfect matchings that still seem feasible to the attacker. Gierlichs et al. (2008) pointed out that the attacker usually has a more modest goal, especially in the commonly-occurring scenarios where system users may send and/or receive multiple messages. The following is a simple example.

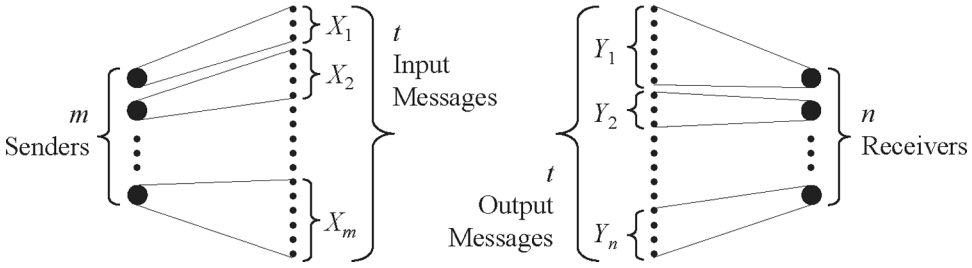
Example 4: Suppose the attacker knows that messages x_3 and x_5 of Figure 3 were both sent by the same sender, say *Alice*, and that message y_4 was received by the receiver *Bob*. Consider now the anonymity of the sender-receiver pair $\langle \textit{Alice}, \textit{Bob} \rangle$. Half the perfect matchings allowed by the graph of Figure 3(a) contain the edge $\langle x_3, y_4 \rangle$, and the other half contain $\langle x_5, y_4 \rangle$. Thus, although the attacker is unsure of whether y_4 was message x_3 or x_5 (each event has a 50% probability), he has determined with 100% certainty the *sender*, namely *Alice*, of a message received by *Bob*. From the point of view of the anonymity of the $\langle \textit{Alice}, \textit{Bob} \rangle$ pair, the attacker does not care to determine *which* of the two messages, x_3 or x_5, was y_4, but is content to know that *Alice* sent a message to *Bob*. $\square$

To measure the extent to which the system is still hiding the *level* of communication between its users, who may be sending/receiving multiple messages, a modification of the metric of Definition 2.3 is therefore needed, as this metric considers relationships only between the system's input and output messages. While an attack, such as those of Examples 1 and 2, still determines infeasibility of *input-output message relationships*, leading to an arbitrary $t \times t$ biadjacency matrix A, the modified metric needs to measure anonymity of *sender-receiver relationships*. We now make this more precise, and show that sender-receiver relationship anonymity is usually lower than that of input-output message relationship.

3.1 Associations between senders and receivers

Let m be the number of senders of a system, As shown in Figure 4, and for any $i \in \{1, 2, \ldots, m\}$, let X_i be the set of messages sent by sender i. Similarly, let n be the number of receivers and, for any $j \in \{1, 2, \ldots, n\}$, let Y_j be the set

Figure 4 Users sending and/or receiving multiple messages



of messages received by receiver j. Note that, $\sum_{i=1}^{m} |X_i| = \sum_{j=1}^{n} |Y_j| = t$. Clearly,

$$\{X_i \times Y_j : 1 \leq i \leq m, 1 \leq j \leq n\}$$

is a partition of $X \times Y$. Any member $X_i \times Y_j$ of this partition is the set of all edges in $K_{t,t}$ from sender i to receiver j.

For any set $E \subseteq X \times Y$ of edges in $K_{t,t}$, we define the *(sender-receiver) association matrix* of E, denoted $\mathbf{Z}(E)$, as the $m \times n$ matrix of nonnegative integers given by:

$$\mathbf{Z}(E)_{ij} = |E \cap (X_i \times Y_j)|.$$

Any entry $\mathbf{Z}(E)_{ij}$ of this matrix is simply the number of edges (i.e., input-output message associations) in E from sender i to receiver j. It follows that the sum of all entries in $\mathbf{Z}(E)$ is $|E|$.

In all, there are $2^{(t^2)}$ subsets of $X \times Y$, of which as we already know, $t!$ are perfect matchings between X and Y. If E is a perfect matching, then its association matrix $\mathbf{Z}(E)$ has an additional property. Its row- and column-sums are then the same as the sender and receiver multiplicities in the system, respectively, i.e.,

$$\sum_{k=1}^{n} \mathbf{Z}(E)_{ik} = |X_i|, \quad 1 \leq i \leq m, \text{ and}$$
$$\sum_{k=1}^{m} \mathbf{Z}(E)_{kj} = |Y_j|, \quad 1 \leq j \leq n.$$

Let $\mathfrak{P}$ be the set of all $t!$ perfect matchings (between X and Y) of the system. We now define an equivalence relation $\bowtie$ on $\mathfrak{P}$.

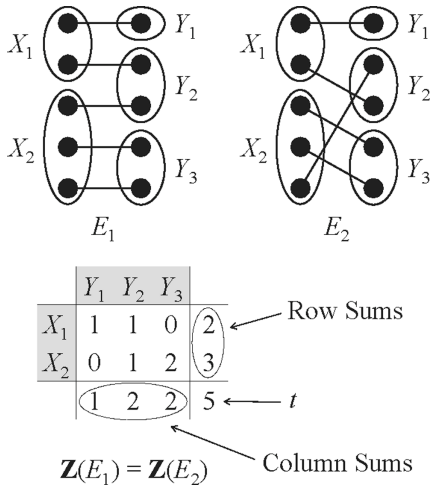
Definition 2: Let $E_1, E_2 \in \mathfrak{P}$. Then E_1 and E_2 are *equivalent*, denoted $E_1 \bowtie E_2$, if they have the same association matrix, i.e., $\mathbf{Z}(E_1) = \mathbf{Z}(E_2)$. □

Perfect matchings are thus considered equivalent if they have the same *number* of messages going from each sender to each receiver. The exact input-output message relationship in equivalent perfect matchings may, however, be different.

Example 5: In the system of Figure 3, suppose an attacker observed that $X_1 = \{x_1, x_2\}, X_2 = \{x_3, x_4, x_5\}, Y_1 = \{y_1\}$, $Y_2 = \{y_2, y_3\}, Y_3 = \{y_4, y_5\}$. In other words, the system has $m = 2$ senders, $n = 3$ receivers, and its $t = 5$ messages are broken down into sender multiplicities of $|X_1| = 2$ and $|X_2| = 3$, and on the other side, receiver multiplicities of $|Y_1| = 1$, $|Y_2| = 2$ and $|Y_3| = 2$.

The two perfect matchings $E_1, E_2 \in \mathfrak{P}$, shown in Figure 5, are then equivalent because they have the same association matrix, also shown in the figure. The row- and column-sums of this association matrix can be seen to be, respectively, the sender and receiver message multiplicities. □

Figure 5 An example of two equivalent perfect matchings and their common association matrix



	Y_1	Y_2	Y_3	
X_1	1	1	0	2
X_2	0	1	2	3
	1	2	2	5 ← t

3.2 A new metric for message multiplicities

We now develop a metric for measuring the extent to which the system's sender-receiver communication pattern is hidden among other such patterns, upon conclusion of a two-pronged attack that has performed the following:

- Determined infeasibility of some input-output message pairings. This portion of the attack is carried out at the level of the system's input and output messages, as in Examples 1 and 2, thereby resulting in an arbitrary $t \times t$ biadjacency matrix A.
- Associated the correct sender with each input message of the system, and receiver with each output message, thereby resulting in the sender and receiver multiplicity vectors S and R.

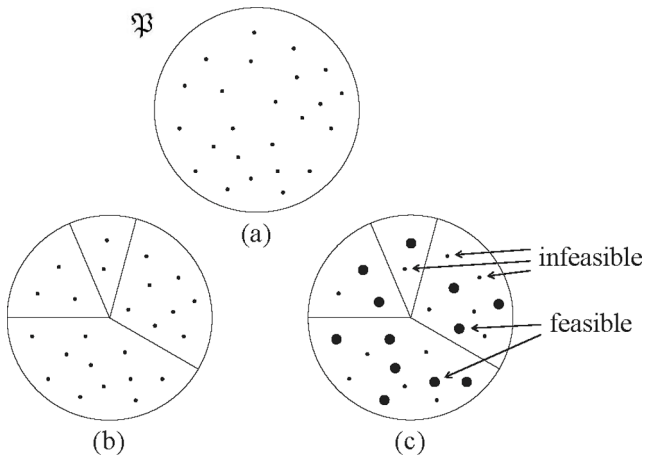
It is important to note that the above segments of the attack are orthogonal to each other and each of them, independently, has

an adverse effect on the system's anonymity level. Our metric measures their combined effect.

Gierlichs et al. (2008) were the first to realise a need for measuring their combined effect, and even proposed a metric to that end. However, their metric works for only a small class of biadjacency matrices. Grégoire and Hamel (2009) considered just the effect of the multiplicity vectors, and did not take any biadjacency matrices into account. In Section 4 we analyse these approaches in detail, and show the limitation in their scopes.

We already understand the independent effects of each of these attack segments. Recall that $\mathfrak{P}$ is the set of all $t!$ perfect matchings, as shown in Figure 6(a).

Figure 6 (a) The set $\mathfrak{P}$ of all $t!$ perfect matchings; (b) equivalence classes on $\mathfrak{P}$ induced by the multiplicity vectors S and R and (c) feasible/infeasible labelling of perfect matchings in $\mathfrak{P}$ caused by the biadjacency matrix A



The sender and receiver multiplicity vectors S and R induce the equivalence relation $\bowtie$ on $\mathfrak{P}$, whose classes are as depicted in Figure 6(b). Each class represents a sender-receiver association scenario, captured by a unique association matrix with S as its row-sums vector and R as its column-sums vector. In the absence of the first attack component, namely the one based on the biadjacency matrix, just the sizes of these classes lead directly to a probability distribution over all possible association scenarios. Anonymity can be measured from this distribution by standard techniques, such as by Shannon entropy, as mentioned a little later.

The biadjacency matrix A resulting from the first attack component, however, alters this probability distribution in a subtle, yet significant, way. By rendering some input-output message pairings of the system infeasible, A in fact ends up labelling some perfect matchings in $\mathfrak{P}$ infeasible (if at least one edge in that matching is infeasible). This effect is depicted in Figure 6(c). For our task of measuring anonymity, what is important now is not the raw *size* of each class, but the number of *feasible* perfect matchings in each class.

Example 6: The two perfect matchings shown in Figure 5, namely E_1 and E_2, are equivalent, as both belong to the same equivalence class. The matching E_1 is infeasible, since some of its edges (e.g., $\langle x_2, y_2\rangle$, $\langle x_3, y_3\rangle$, etc.) were determined by the attacker to be infeasible and are thus not contained in the graph of Figure 3. On the other hand, the matching E_2 is feasible, as all its edges are also contained in the graph of Figure 3. □

Given any multiplicity vectors S and R, let $\mathcal{Z}_{S,R}(\mathfrak{P})$ be the set of all association matrices of perfect matchings in $\mathfrak{P}$, i.e., the set of all $m \times n$ nonnegative integer matrices with S as their row-sums vector and R as their column-sums vector. Additionally, given a $t \times t$ biadjacency matrix A, let the *weight* assigned by A to any matrix $Z \in \mathcal{Z}_{S,R}(\mathfrak{P})$, denoted $\mathcal{W}_A(Z)$, be the number of perfect matchings in the equivalence class of $\bowtie$ associated with Z that are declared feasible by A. We already know that:

$$\sum_{Z \in \mathcal{Z}_{S,R}(\mathfrak{P})} \mathcal{W}_A(Z) = \text{per}(A),$$

as in all there are $\text{per}(A)$ perfect matchings that appear feasible to the attacker. In Section 3.3, we give a method to compute $\mathcal{W}_A(Z)$. For now, we let

$$\omega_A(Z) = \mathcal{W}_A(Z)/\text{per}(A)$$

be the *normalised weight* of Z, i.e., the fraction of feasible perfect matchings present in the class denoted by Z. It is easy to see that the values $\omega_A(Z)$ add up to 1, over all Z, and we have a probability distribution on the set $\mathcal{Z}_{S,R}(\mathfrak{P})$ of all sender-receiver association scenarios. For any Z, the value $\omega_A(Z)$ is the likelihood assigned by the attacker to the scenario represented by Z, of being the actual communication pattern.

Ever since the works of Serjantov and Danezis (2002) and Diaz et al. (2002), Shannon entropy of a probability distribution is a well accepted measure of the system's degree of anonymity. We employ the same technique over the probability distribution given by $\omega_A(Z)$ values, for all $Z \in \mathcal{Z}_{S,R}(\mathfrak{P})$, as a measure of the attacker's uncertainty of which of the system's sender-receiver association scenarios is the actual one.

Definition 3: Let a $t \times t$ biadjacency matrix A, and multiplicity vectors S and R, be the result of an attack. We define the underlying system's *degree of anonymity* after this attack as:

$$\delta_{S,R}(A) = \begin{cases} 0 & \text{if } t = 1, \\ \dfrac{-\sum_{Z \in \mathcal{Z}_{S,R}(\mathfrak{P})} \omega_A(Z) \cdot \log(\omega_A(Z))}{\log(t!)} & \text{otherwise.} \end{cases}$$

It is easily verified that $\delta_{S,R}(A)$ is always between 0 and 1. We first establish that, if all message multiplicity values in S and R are 1, the above metric coincides with that of Edman et al. (2007), given by Definition 2.3, and that higher multiplicity values reduce anonymity. The following proposition follows from the definitions:

Proposition 1: *For all S, R, and A, $\delta_{S,R}(A) \leq d(A)$, with equality iff all multiplicity values in S and R are 1.*

In the next subsection we develop a method for computing the weight, $\mathcal{W}_A(Z)$, of an association matrix Z, i.e., the number of matchings deemed feasible by A in the equivalence class associated with Z.

3.3 Weight of an association matrix

For a given $t \times t$ biadjacency matrix A, and an $m \times n$ association matrix Z that corresponds to some equivalence class of $\bowtie$, we now develop an expression for $\mathcal{W}_A(Z)$.

The row- and column-sums vectors of such a matrix Z are, respectively, the system's sender and receiver multiplicity vectors,

$$S = \langle |X_1|, |X_2|, \ldots, |X_m| \rangle, \text{ and}$$
$$R = \langle |Y_1|, |Y_2|, \ldots, |Y_n| \rangle.$$

Without loss of generality, we assume that the first $|X_1|$ rows of A correspond to messages sent by the first sender, the next $|X_2|$ rows to messages sent by the second sender, and so on. Similarly, the first $|Y_1|$ columns of A correspond to messages received by the first receiver, etc. If necessary, the rows and/or columns of A can be permuted to achieve this without affecting the result of the weight computation method developed in this section.

There is now a natural partition of A into mn mutually disjoint *regions*, denoted $Reg_{(A;\, i \to j)}$, for $1 \le i \le m$, and $1 \le j \le n$, as shown in Figure 7(a). Any particular region, $Reg_{(A;\, i \to j)}$, is the submatrix of A made up of $|X_i|$ contiguous rows starting from row number $1 + \sum_{k=1}^{i-1} |X_k|$, and $|Y_j|$ contiguous columns starting from column number $1 + \sum_{k=1}^{j-1} |Y_k|$. This region contains feasibility information determined by the attacker of the subset $X_i \times Y_j$ of edges in the system's complete graph $K_{t,t}$.

We also need square submatrices of A that are not necessarily contiguous. Let I and J be any subsets of size s of the set $\{1, 2, \ldots, t\}$, i.e., the row and column index set of A. We use $A[\![I; J]\!]$ to denote the *s-extract* obtained by extracting the $s \times s$ submatrix made up of elements of A at row numbers in I and column numbers in J. The s-extract $A[\![I; J]\!]$ represents the situation of the s input messages given by indices in I corresponding, in any order, to the s output messages given by indices in J. Observe that, according to the attacker, there are exactly

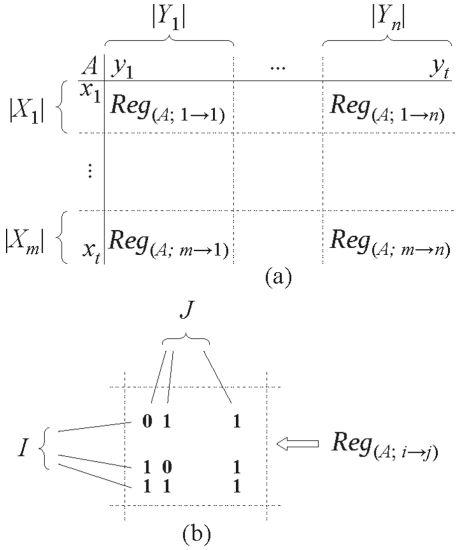
$$\text{per}(A[\![I; J]\!])$$

feasible ways of achieving this correspondence, where $0 \le \text{per}(A[\![I; J]\!]) \le s!$.

In general, an s-extract may intersect with multiple regions of A. However, for the method we are developing in this section to evaluate $\mathcal{W}_A(Z)$, we will be interested in only a small class of them. For any value Z_{ij} of the association matrix Z, the Z_{ij}-extracts that are completely contained within their corresponding region $Reg_{(A;\, i \to j)}$ are of concern to us. As depicted in Figure 7(b) for an example value of $Z_{ij} = 3$, any Z_{ij}-extract $A[\![I; J]\!]$ can be easily seen to be completely contained within the region $Reg_{(A;\, i \to j)}$ if and only if

$$I \subseteq \left\{ c + \sum_{k=1}^{i-1} |X_k| : 1 \le c \le |X_i| \right\}, \text{ and}$$
$$J \subseteq \left\{ c + \sum_{k=1}^{j-1} |Y_k| : 1 \le c \le |Y_j| \right\}.$$

Figure 7 (a) Regions of A, $Reg_{(A;\, i \to j)}$, for $1 \le i \le m$, and $1 \le j \le n$ and (b) an example 3-extract completely contained within a region



As Z_{ij} number of rows and columns may be chosen from $Reg_{(A;\, i \to j)}$ in $\binom{|X_i|}{Z_{ij}} \binom{|Y_j|}{Z_{ij}}$ ways, that is also the number of such extracts. Let $\mathcal{E}_{(A;\, i \to j)}$ denote the set of all these Z_{ij}-extracts that are completely contained within the region $Reg_{(A;\, i \to j)}$. From the above explanation, the following proposition now becomes straightforward:

Proposition 2: *According to A, the total number of feasible ways of sending Z_{ij} messages from sender i to receiver j is:*

$$\sum_{A[\![I;J]\!] \in \mathcal{E}_{(A;\, i \to j)}} \text{per}(A[\![I; J]\!]).$$

To determine $\mathcal{W}_A(Z)$, i.e., the number of all perfect matchings declared feasible by A in the equivalence class associated with Z, a product of the above count for all sender-receiver combinations can now be employed. A little caution needs to be exercised though, as choosing a Z_{ij}-extract $A[\![I; J]\!] \in \mathcal{E}_{(A;\, i \to j)}$ makes all input messages in I unavailable for receivers other than j. Similarly, all output messages in J become unavailable for senders other than i. We accomplish this by zeroing-out all rows in I and columns in J of the biadjacency matrix A for all subsequent counting.

For any extract $A[\![I; J]\!]$, we let $\widehat{A}[\![I; J]\!]$ denote the matrix identical to A, except it contains zeroes for all elements of rows in I and columns in J, i.e.,

$$\widehat{A}[\![I; J]\!]_{pq} = \begin{cases} 0 & \text{if } p \in I \text{ or } q \in J, \\ A_{pq} & \text{otherwise.} \end{cases}$$

The permanent of any extract from $\widehat{A}[\![I;J]\!]$ that has at least one row from I and/or at least one column from J will clearly be 0, thus avoiding duplicate counting resulting from associating any input message with multiple output messages, or vice versa. We also let $\widehat{Z}[\![i;j]\!]$ denote the matrix identical to Z, except it contains a 0 for the element at row i and column j, i.e.,

$$\widehat{Z}[\![i;j]\!]_{pq} = \begin{cases} 0 & \text{if } p = i \text{ and } q = j, \\ Z_{pq} & \text{otherwise.} \end{cases}$$

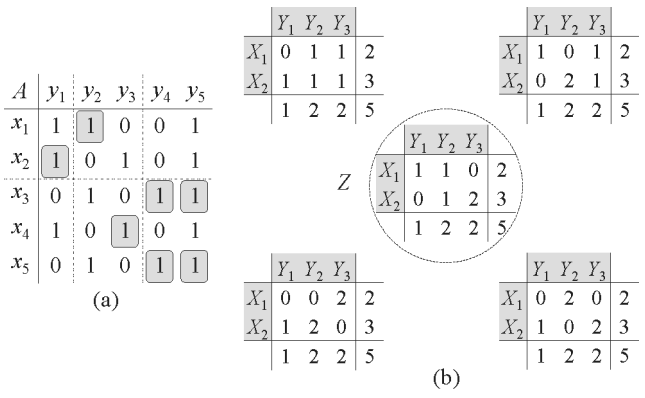
The number of feasible perfect matchings in the class associated with Z can now be expressed as the following recursive formula, which effectively multiplies the number of feasible 'pieces' corresponding to each sender-receiver combination:

$$\mathcal{W}_A(Z) = \begin{cases} \displaystyle\sum_{A[\![I;J]\!] \in \mathcal{E}_{(A;\, i\to j)}} \Big[\mathrm{per}(A[\![I;J]\!]) \cdot \mathcal{W}_{\widehat{A}[\![I;J]\!]}(\widehat{Z}[\![i;j]\!])\Big] \\ \qquad\qquad \text{if } Z_{ij} \neq 0, \text{ for some } i \text{ and } j, \\ 1 \qquad\qquad \text{otherwise.} \end{cases}$$

The depth of recursion in the above is exactly the number of nonzero entries in Z. The order in which these entries are considered is not important. At any stage, one such entry Z_{ij} is chosen nondeterministically and, for each Z_{ij}-extract $A[\![I;J]\!] \in \mathcal{E}_{(A;\, i\to j)}$, the number of perfect matchings that are extensions of some partial matching from $A[\![I;J]\!]$ is determined. This value is added over all such extracts for Z_{ij}.

Example 7: For the system of Figure 3(a) in Example 2, let m, n, X_1, X_2, Y_1, Y_2, and Y_3 be as given in Example 5. Thus, $S = \langle 2, 3\rangle$ and $R = \langle 1, 2, 2\rangle$ are the observed message multiplicity vectors for this system. These values induce six regions on the biadjacency matrix A, as shown in Figure 8(a), and partition the set of all $5! = 120$ matchings of this system into five equivalence classes, whose association matrices are shown in Figure 8(b).

Figure 8 (a) Regions induced on the biadjacency matrix by multiplicity vectors, and an example extract collection for the circled association matrix Z and (b) association matrices of all equivalence classes



These are all the $m \times n$ nonnegative integer matrices that have S as their row-sums vector and R as their column-sums vector. Incidentally, the circled association matrix Z of Figure 8(b) is identical to the one already encountered in Figure 5. We first step through the computation of $\mathcal{W}_A(Z)$, by our method.

There are four nonzero entries in Z, namely $Z_{11} = Z_{12} = Z_{21} = 1$, and $Z_{23} = 2$. Thus any perfect matching in the equivalence class associated with Z is made up of four extracts completely contained within regions of A with pairwise disjoint row-sets and column-sets: a 1-extract from each of the regions

$$\mathit{Reg}_{(A;\, 1\to 1)},\ \mathit{Reg}_{(A;\, 1\to 2)},\ \text{and } \mathit{Reg}_{(A;\, 2\to 1)},$$

and a 2-extract from the region

$$\mathit{Reg}_{(A;\, 2\to 3)}.$$

An example collection of such extracts within the regions induced on A is shown in Figure 8(a). The product of permanents of extracts in this collection is:

$$1 \cdot 1 \cdot 1 \cdot (1 \cdot 1 + 1 \cdot 1) = 2.$$

Our recursive method given in this section for computing the weight of Z essentially adds this value for *all* such collections of extracts. It can be seen from a quick enumeration of all such collections that $\mathcal{W}_A(Z) = 4$.

The weights $\mathcal{W}_A$ of the other four association matrices shown in Figure 8(b) can be similarly evaluated by our method to be 2, 4, 0, and 2. The sum of weights of all five association matrices is $\mathrm{per}(A) = 12$, and dividing their individual weights by this value leads to the normalised weights ω_A of these matrices: $1/3, 1/6, 1/3, 0$, and $1/6$. From our metric of Definition 3 for the system's degree of anonymity, we get $\delta_{S,R}(A) \approx 0.278$. □

It is instructive to compare the value of the metric computed in the above example with that of Edman et al. (2007) based on just the matrix A, computed in Example 3, as $d(A) \approx 0.519$. The reduction in anonymity captured by our metric is due to message multiplicities. In general, the higher the message multiplicity values in vectors S and R, the lower the anonymity. Figure 9 depicts this phenomenon over all possible message multiplicity vectors for this particular biadjacency matrix A. In the figure,

$$\alpha = \frac{\sum_{i=1}^{m} |X_i| + \sum_{j=1}^{n} |Y_j|}{m+n}$$

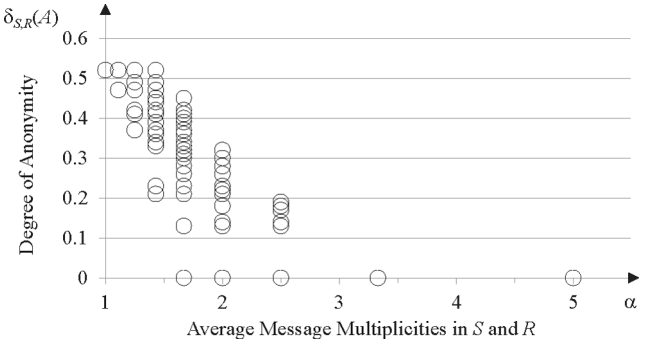
is the average multiplicity count in S and R. As shown, for some values of α the anonymity is even reduced to zero.

3.4 Complexity of computing the new metric

Given a $t \times t$ biadjacency matrix A, and multiplicity vectors S and R, we now study the complexity of computing $\delta_{S,R}(A)$, which mainly depends upon computing the permanents of

several s-extracts of A. For any s, $1 \leq s \leq t$, there are at most $\binom{t}{s}^2$ s-extracts of A. Although most of these s-extracts straddle over multiple regions of A, and are therefore not relevant for the computation of the metric, for the worst case analysis, we go by this number of $s \times s$ submatrices of A whose permanents need to be computed. The maximum number of possible extracts is thus $\sum_{s=1}^{t} \binom{t}{s}^2$.

Figure 9 Degree of anonymity of biadjacency matrix A of Figure 3(b), for all possible multiplicity vectors S and R



Valiant (1979) showed that computing the permanent of a matrix, even if all values of the matrix are just 0 or 1, is #P-complete. A polynomial-time solution for computing the permanent is thus unlikely, as that would imply $P = NP$. The fastest known exact algorithms for computing the permanent of a $s \times s$ real matrix have time complexity $\Theta(s2^s)$. An example is the method of Ryser that appeared on Page 122 of Minc (1978). Much attention has consequently been given to arriving at approximations to permanents more efficiently, as in Jerrum and Vazirani (1996) and Chien et al. (2003). However, even the state-of-the-art polynomial-time approximation method of Jerrum et al. (2004) runs in $\mathcal{O}(s^{22})$, resulting in the worst-case complexity of $\mathcal{O}(\sum_{s=1}^{t} \binom{t}{s}^2 s^{22})$ of approximately computing $\delta_{S,R}(A)$. This strongly suggests the need for arriving at a more efficient heuristic for our metric, whose development is beyond the scope of this paper and left as future work.

4 Comparison with existing approaches

There have been two other attempts at measuring the system-wide anonymity after such attacks when users send and/or receive multiple messages. Gierlichs et al. (2008) were the first to observe a need for modifying the basic metric of Edman et al. (2007) for such a multiple message scenario. However, the metric they proposed is applicable, as shown in this section, to only a small class of biadjacency matrices resulting from attacks. Later, Grégoire and Hamel (2009) revisited this problem, but stopped at just giving a method for counting the number of equivalence classes. Their counting method actually appeared earlier in Macdonald (1979), and]vfill was also summarised in Diaconis and Gangolli (1995). In this section, we briefly explain these approaches and demonstrate the limitation of their scopes.

4.1 The metric of Gierlichs et al. (2008)

We first note that our equivalence relation $\bowtie$ is induced solely by the multiplicity vectors S and R, and is defined over the set $\mathfrak{P}$ of all $t!$ perfect matchings between X and Y. The *size* of any equivalence class of $\bowtie$ thus depends only on S and R. Once a biadjacency matrix A is introduced into the picture, the *weight* of any class, as declared by A, is computed to measure anonymity. Gierlichs et al. (2008), on the other hand, defined an equivalence relation $\sim$ that is induced by S, R, *and* a given biadjacency matrix A. While the underlying definitions of $\bowtie$ and $\sim$ are the same, their relation $\sim$ is defined on the set $\mathfrak{F}$ of all per(A) perfect matchings deemed *feasible* by A. Their metric is similar to ours given by Definition 3, with the exception that theirs is based upon the sizes of classes of $\sim$, which are essentially the weights of the corresponding classes of $\bowtie$. For the purpose of measuring anonymity, this difference between the two approaches is therefore inconsequential.

However, a major limitation of the method given in Gierlichs et al. (2008) to compute the equivalence class sizes of $\sim$ is that it works correctly for only a small class of biadjacency matrices.

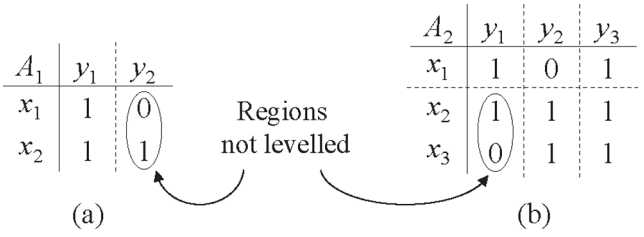
Recall that m is the total numbers of senders and n is the total number of receivers. For any sender i and receiver j, let the region $Reg_{(A;\, i \to j)}$ of a $t \times t$ biadjacency matrix A be called *leveled* if all its elements are identical, i.e., either all values in it are 0 or all values are 1. Also, the matrix A is called *leveled* if each of its mn regions are leveled. Clearly, of all the $2^{(t^2)}$ possible biadjacency matrices, only 2^{mn} are leveled, and this number shrinks as multiplicities in the S and R vectors grow. We now show that the class size computation method of Gierlichs et al. (2008) is limited to leveled biadjacency matrices.

Gierlichs et al. (2008) first consider the special case of all receiver multiplicities being 1, i.e., $|Y_j| = 1$, for all j, $1 \leq j \leq n$. For this particular case, they assert the following:

Assertion 1: If $|Y_j| = 1$, for all j, $1 \leq j \leq n$, then all equivalence classes of $\sim$ are of equal size, given by $\prod_{i=1}^{m} |X_i|!$.

Two counterexamples to the above assertion are shown in Figure 10.

Figure 10 Two counterexamples to Assertion 1, of Gierlichs et al. (2008)



For the example of Figure 10(a), $t = 2$, $m = 1$, $S = \langle 2 \rangle$, per$(A_1) = 1$, and the only feasible perfect matching is in its

own class, i.e., there is just one equivalence class of $\sim$, and its size is 1. However, $\prod_{i=1}^{1} |X_i|! = 2! = 2$, which is not the class size. For the example of Figure 10(b), $t = 3, m = 2, S = \langle 1, 2\rangle$, per$(A_2) = 3$, and the three feasible perfect matchings are partitioned into two classes of *unequal* sizes, namely size 1 and 2, respectively. Again, $\prod_{i=1}^{2} |X_i|! = 1! \cdot 2! = 2$ is not the size of *each* class.

The reason why Assertion 1 fails to hold in the above examples is that each of these biadjacency matrices contains at least one region that is not leveled. This leads us to the following proposition:

Proposition 3: *If $|Y_j| = 1$, for all j, $1 \leq j \leq n$, then all equivalence classes of $\sim$ are of equal size, given by $\prod_{i=1}^{m} |X_i|!$, iff A is leveled.*

Proof: Since $|Y_j| = 1$, for all j, all regions of A are just one-column wide. For the only-if part, suppose A is not leveled. Then, for some sender s and receiver r, $Reg_{(A;\, s \to r)}$ is not leveled, i.e., this region contains a 0 as well as 1 value. Consider any perfect matching deemed feasible by A that contains the edge corresponding to this 1 value, and let C be the equivalence class in $\sim$ of that matching. As this region contains $k < |X_s|$ number of 1 values, $|C|$ is at most $(\prod_{i=1}^{s-1} |X_i|!) \cdot k! \cdot (\prod_{i=s+1}^{m} |X_i|!)$, which is strictly smaller than $\prod_{i=1}^{m} |X_i|!$. We leave out the proof of the if part as it is not relevant to our discussion here. □

The next special case considered in Gierlichs et al. (2008) is when all sender multiplicities are 1, i.e., $|X_i| = 1$, for all $i, 1 \leq i \leq m$. For this case, they asserted that all equivalence classes of $\sim$ are of equal size, given by $\prod_{j=1}^{n} |Y_j|!$. By a reasoning symmetric to the above proposition, it can be shown that this is also true only when A is leveled.

For the general case of any arbitrary sender and receiver multiplicities, Gierlichs et al. (2008) contains a recursive algorithm for computing all class sizes. One of the base cases of that algorithm (Line 9) corresponds to the above particular situations and gives a correct answer only when A is leveled. The other base case (Line 5) corresponds to a solitary sender or receiver and can also be shown to work only when A is leveled. Their entire method is thus limited to leveled biadjacency matrices.

On the other hand, our method of computing weights of equivalence classes of $\bowtie$, given in Section 3.3, works for all biadjacency matrices.

To better understand the scope enhancement of our method over that of Gierlichs et al. (2008), we define the following five sets of biadjacency matrices:

Ψ = the set of all $t \times t$ biadjacency matrices. Then, $|\Psi| = 2^{(t^2)}$. We already know that the permanents of matrices in Ψ are integer values between 0 and $t!$.

$\Psi_{>0}$ = the set of all matrices $A \in \Psi$, such that per$(A) > 0$. These are matrices whose permanent is at least 1. The basic metric of Edman et al. (2007) is for such matrices as they label at least one perfect matching feasible, the system's true communication pattern. Ideally, any extension to that metric, such as the current one for sending/receiving multiple messages, should cover this set. To obtain an estimate of the number of such matrices, we define the next two sets.

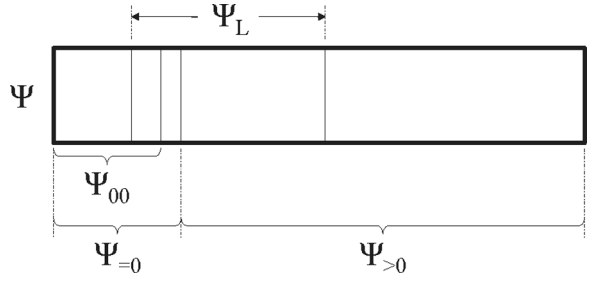
$\Psi_{=0}$ = the set of all matrices $A \in \Psi$, such that per$(A) = 0$. In other words, $\{\Psi_{=0}, \Psi_{>0}\}$ is a partition of Ψ.

Ψ_{00} = the set of all matrices $A \in \Psi$ that contain at least one row or column, all of whose elements are 0. Clearly, $\Psi_{00} \subseteq \Psi_{=0}$, but not vice versa, as there are matrices with 0 permanent, yet each row and column of which contains at least one 1. However, Erdős and Rényi (1964) showed that when t is large, all sufficiently non-sparse matrices in $\Psi_{=0}$ are also in Ψ_{00}. It is thus reasonable to assume that $\Psi_{=0} \approx \Psi_{00}$. This provides an estimate of $|\Psi_{>0}|$ as the number of matrices that do not contain a row or column of all 0 elements.

Ψ_{L} = the set of all leveled matrices $A \in \Psi$. This set depends upon the multiplicity vectors S and R. It is easily seen that $|\Psi_{\mathrm{L}}| = 2^{mn}$. If all multiplicity values in S and R are 1, we have that $m = n = t$, and $\Psi_{\mathrm{L}} = \Psi$. However, m, n, and Ψ_{L}, all shrink as these multiplicities grow.

Figure 11 shows all above sets. The method of Gierlichs et al. (2008) works for all matrices in Ψ_{L}. As multiplicity values in S and R become larger, Ψ_{L} becomes smaller and fails to cover $\Psi_{>0}$, which is the set of all biadjacency matrices that may result from an attack. Our method of Section 3.3, on the other hand, is applicable to all matrices in Ψ.

Figure 11 Scope of our method, Ψ, compared with that of Gierlichs et al. (2008), Ψ_{L}



4.2 The metric of Grégoire and Hamel (2009)

Grégoire and Hamel (2009) proposed

$$\log(\text{COUNT})/\log(t!)$$

as a 'metric', where COUNT is essentially the number of equivalence classes of our $\bowtie$ relation. This is not a metric in the true sense of the term as it does not take into account any biadjacency matrix resulting from an attack or, alternately, assumes that the biadjacency matrix contains all 1 values. Their formula thus captures just the *maximum* degree of anonymity one could expect, given some sender and receiver multiplicities.

They describe a method based on symmetric functions for determining COUNT. This method was originally given by Macdonald (1979) for counting nonnegative integer matrices with prescribed row and column sums. It is also well summarised in the survey of Diaconis and Gangolli (1995).

Moreover, as can be seen from the discussion leading to our metric in Definition 3, the value of COUNT is only of academic interest and not needed to compute the system's degree of anonymity.

5 Conclusions and future work

Edman et al. (2007) developed a system-wide anonymity metric based upon the extent to which the association between an anonymity system's input and output messages corresponding to the actual communication pattern is hidden, after an attack, among all associations that still seem feasible after the attack. Gierlichs et al. (2008) made the case for modifying this metric for the situation where system users send and/or receive multiple messages. Clearly, as an attacker's goal is to uncover the communication pattern between users, and not simply messages, this modification of the basic metric of Edman et al. (2007) is necessary.

In this paper, we developed a metric for measuring anonymity of associations between senders and receivers of a system after an attack that is still carried out at the level of messages. When multiple messages are sent and/or received by users, this induces an equivalence relation on the set of all perfect matchings between the system's input and output messages. Our main offering here is a method to compute the number of perfect matchings, in any equivalence class, that still seem feasible after the attack. These numbers lead to a probability distribution on the set of all possible communication patterns between senders and receivers, from which the system-wide anonymity level is determined by the Shannon-entropy method of Diaz et al. (2002).

Although a metric for this was originally proposed in Gierlichs et al. (2008), we showed that their metric works only for a small family of attacks. We characterised that family and showed that its size in fact shrinks as the number of messages sent/received by users increases. We also showed that another related work by Grégoire and Hamel (2009) simply counts equivalence classes and, since it does not take any attack into account, is not an anonymity metric.

Recently, Bagai et al. (2011) developed a generalisation of the metric of Edman et al. (2007) for attacks that end up assigning probabilities to perfect matchings between the system's input and output messages, as against declaring some infeasible. One direction for future work is to generalise our metric of this paper for such attacks. In another work, Bagai and Tang (2011) showed that adding data caching abilities to the system model of Edman et al. (2007) results in increased anonymity. Another direction for future work is to extend the metric of this paper for this additional system ability. We are currently working in both of these directions.

Acknowledgements

This work was partially supported by the US Navy Engineering Logistics Office contract no. N41756-08-C-3077.

References

Asratian, A., Denley, T. and Häggkvist, R. (1998) *Bipartite Graphs and their Applications*, Cambridge University Press, Cambridge, UK.

Bagai, R., Lu, H., Li, R. and Tang, B. (2011) 'An accurate system-wide anonymity metric for probabilistic attacks', in Fischer-Hübner, S. and Hopper, N. (Eds): *Proceedings of the 11th International Privacy Enhancing Technologies Symposium (PETS)*, Waterloo, Canada, Vol. 6794 of *Lecture Notes in Computer Science*, Springer-Verlag, pp.117–133.

Bagai, R. and Tang, B. (2011) 'Data caching for enhancing anonymity', *Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, Singapore, pp.135–142.

Chien, S., Rasmussen, L. and Sinclair, A. (2003) 'Clifford algebras and approximating the permanent', *Journal of Computer and System Sciences*, Vol. 67, No. 2, pp.263–290.

Diaconis, P. and Gangolli, A. (1995) 'Rectangular arrays with fixed margins', in Aldous, D., Diaconis, P., Spencer, J. and Steele, J. (Eds): *Discrete Probability and Algorithms*, Vol. 72 of *The IMA Volumes in Mathematics and its Applications*, Springer-Verlag, pp.15–41.

Diaz, C., Seys, S., Claessens, J. and Preneel, B. (2002) 'Towards measuring anonymity', in Dingledine, R. and Syverson, P., (Eds): *Proceedings of the 2nd Privacy Enhancing Technologies Workshop*, San Francisco, USA, Lecture Notes in Computer Science, Springer-Verlag, Vol. 2482, pp.54–68.

Edman, M., Sivrikaya, F. and Yener, B. (2007) 'A combinatorial approach to measuring anonymity', *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, New Brunswick, USA, pp.356–363.

Erdős, P. and Rényi, A. (1964) 'On random matrices', *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, Vol. 8, pp.455–461.

Gierlichs, B., Troncoso, C., Diaz, C., Preneel, B. and Verbauwhede, I. (2008) 'Revisiting a combinatorial approach toward measuring anonymity', *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society*, Alexandria, VA, USA, pp.111–116.

Grégoire, J.-C. and Hamel. A. (2009) *A Combinatorial Enumeration Approach for Measuring Anonymity*, http://arxiv.org/abs/0902.1663

Jerrum, M. and Vazirani, U. (1996) 'A mildly exponential approximation algorithm for the permanent', *Algorithmica*, Vol. 16, pp.392–401.

Jerrum, M., Sinclair, A. and Vigoda, E. (2004) 'A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries', *Journal of the ACM*, Vol. 51, No. 4, pp.671–697.

Macdonald, I. (1979) *Symmetric Functions and Hall Polynomials*, Clarendon Press, Oxford, NY, USA.

Minc, H. (1978) *Permanents*, Vol. 6 of *Encyclopedia of Mathematics and its Applications*, Addison-Wesley, Reading, MA, USA.

Serjantov, A. and Danezis, G. (2002) 'Towards an information theoretic metric for anonymity', in Dingledine, R. and Syverson, P. (Eds.): *Proceedings of the 2nd Privacy Enhancing Technologies Workshop*, San Francisco, USA, Lecture Notes in Computer Science, Springer-Verlag, Vol. 2482, pp.41–53.

Valiant, L. (1979) 'The complexity of computing the permanent', *Theoretical Computer Science*, Vol. 8, No. 2, pp.189–201.