An Accurate System-Wide Anonymity Metric for Probabilistic Attacks

Rajiv Bagai, Huabo Lu, Rong Li, and Bin Tang

Department of Electrical Engineering and Computer Science
Wichita State University, Wichita, KS 67260-0083, USA
{rajiv.bagai,hxlu,rxli1,bin.tang}@wichita.edu

Abstract. We give a critical analysis of the system-wide anonymity metric of Edman et al. [3], which is based on the permanent value of a doubly-stochastic matrix. By providing an intuitive understanding of the permanent of such a matrix, we show that a metric that looks no further than this composite value is at best a rough indicator of anonymity. We identify situations where its inaccuracy is acute, and reveal a better anonymity indicator. Also, by constructing an information-preserving embedding of a smaller class of attacks into the wider class for which this metric was proposed, we show that this metric fails to possess desirable generalization properties. Finally, we present a new anonymity metric that does not exhibit these shortcomings. Our new metric is accurate as well as general.

Keywords: System-wide anonymity metric, Probabilistic attacks, Combinatorial matrix theory.

1 Introduction

Measuring the amount of anonymity that remains in an anonymity system in the aftermath of an attack has been a concern ever since a need for web anonymity systems was first recognized. Much of the work on anonymity metrics, such as that of Serjantov and Danezis [1] or of Diaz, Seys, Claessens and Preneel [2], has focused on measuring anonymity from the point of view of a single message or user. In contrast, Edman, Sivrikaya and Yener [3] proposed a *system-wide* metric for measuring an attacker's uncertainty in linking each input message of a system with the corresponding output message it exited the system as. They employ the framework of a complete bipartite graph between the system's input and output messages. Any perfect matching between nodes of this graph is a possible message communication pattern of the system. Anonymity in this framework is measured as the extent to which the single perfect matching reflecting the system's true communication pattern is hidden, after an attack, among all perfect matchings in the graph.

Edman et al. [3] gave metrics for measuring anonymity after two kinds of attacks, which we name as *infeasibility* and *probabilistic* attacks. Infeasibility attacks determine infeasibility of some edges in the system's complete bipartite

S. Fischer-Hübner and N. Hopper (Eds.): PETS 2011, LNCS 6794, pp. 117–133, 2011.

graph and arrive at a reduced graph by removing such edges. Probabilistic attacks, on the other hand, arrive at probabilities for each edge in the complete bipartite graph of being the actual communication pattern. Both metrics of [3] are based upon *permanent* values of certain underlying matrices.

Contributions of our paper are two-fold. We first demonstrate that while the metric given in [3] for infeasibility attacks is sound, the one for probabilistic attacks has two major shortcomings. We then propose a new, unified anonymity metric for both classes of attacks that overcomes these shortcomings.

By presenting an intuitive understanding of the permanent of a matrix for probabilistic attacks, we show that the first shortcoming of the metric in [3] for such attacks is that the permanent, which is a composite value, is at best a rough indicator of the system's anonymity level. We highlight situations in which the permanent is especially inadequate, and show that a better anonymity indicator is the breakdown of the permanent as a probability distribution on the graph's perfect matchings.

The second shortcoming shown of the metric in [3] for probabilistic attacks is that it is not a generalization of their metric for infeasibility attacks. We present an information-preserving embedding of infeasibility attacks into the wider class of probabilistic attacks to show that the former are just special cases of the latter, a relationship ideally reflected in the metrics of [3], but is not.

The rest of this paper is organized as follows. Section 2 contains an overview of the two metrics proposed by Edman et al. [3], namely for infeasibility and probabilistic attacks. Section 3 analyzes the metric of [3] for probabilistic attacks and exposes two shortcomings of it. The inadequacy of permanent as an indicator of anonymity is explained in Section 3.1, and its failure to correctly generalize infeasibility attacks in Section 3.2. These sections also develop much of the mathematical framework that is used to construct our new, unified metric, which is then presented in Section 4. Finally, Section 5 concludes our work and mentions some directions for future work.

2 Overview of a System-Wide Metric

In this section we give an overview of the anonymity metrics proposed by Edman, Sivrikaya, and Yener [3]. Their metrics give 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.

Let S be the set of n input messages observed by an attacker having entered an anonymity system, and T be the set of output messages observed by the attacker having exited from that system. It is assumed that every input message eventually appears at the output, i.e. $|S| = |T| = n$. The anonymity system attempts to hide from the attacker which input message in S exited the system as which output message in T. It may employ a number of techniques to this end, such as outputting messages in an order other than the one in which they arrived to prevent sequence number association, or modifying message encoding by encryption/decryption to prevent message bit-pattern comparison, etc. The

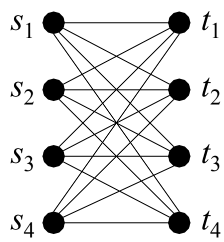


Fig. 1. Complete anonymity, when all edges in the complete bipartite graph between the system's input and output messages are equally likely

maximum anonymity this system can strive to achieve is when for any particular input message in S, each of the output messages in T is equally likely to be the one that input message in S exited the system as. This situation is depicted by the complete bipartite graph $K_{n,n}$ between S and T, as shown in Fig. 1 for $n = 4$. Any edge $\langle s_i, t_j \rangle$ in this graph indicates that the incoming message s_i could possibly have been the outgoing message t_j. All edges in the graph are considered equally likely.

Edman et al. in [3] consider two different classes of attacks. The first class is of attacks that label some of the edges (i.e. input-output pairings) in the above complete bipartite graph as infeasible. Removal by the attacker of these infeasible edges from the graph results in decreased anonymity. The latency-based attack of [3] and the route length attack of Serjantov and Danezis [1] are examples of such attacks. The second class considered in [3] is of attacks that arrive at probabilities for the edges in the graph of Fig. 1 of being the actual communication pattern. This also reduces the anonymity provided by the system, and an example of such a probabilistic attack is given in [3] as well.

For both of these classes of attacks, Edman et al. [3] propose anonymity metrics to reflect the level of anonymity remaining in the system in the aftermath of an attack. While our work in this paper is an improvement of just the second metric of [3], namely for probabilistic attacks, here we give an overview of both metrics of [3] as they are related.

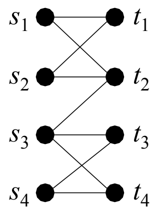
2.1 A Metric for Infeasibility Attacks

An infeasibility attack removes from the system's complete bipartite graph, like the one shown in Fig. 1, edges that are determined by the attack to be infeasible due to some attacker's observation.

Edman et al. [3] give an example of such an attack that notes the times at which messages enter and exit the system, and uses its knowledge of the minimum and/or 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 4 messages enter and exit this system at times shown in Fig. 2(a), then s_1 must be either t_1 or t_2, because the other outgoing messages, namely t_3

Entry times	Exit times
$s_1 = 1$	$t_1 = 4$
$s_2 = 2$	$t_2 = 5$
$s_3 = 4$	$t_3 = 7$
$s_4 = 5$	$t_4 = 8$

(a)



(b)

	t_1	t_2	t_3	t_4
s_1	1	1	0	0
s_2	1	1	0	0
s_3	0	1	1	1
s_4	0	0	1	1

(c)

Fig. 2. (a) Message entry and exit times observed by attacker. (b) Graph resulting from the attack, which removed edges it determined to be infeasible from system's complete bipartite graph. (c) Biadjacency matrix of this graph.

and t_4, are outside the possible latency window of s_1. Similar reasoning can be performed on all other messages to arrive at the reduced graph produced by this attack, shown in Fig. 2(b). Note that in this graph s_1 is connected to only t_1 and t_2, and not to t_3 or t_4, since the edges $\langle s_1, t_3 \rangle$ and $\langle s_1, t_4 \rangle$ were determined by the attack to be infeasible. 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 Fig. 2(c).

The number of perfect matchings between the system's input and output messages allowed by the bipartite graph resulting from such an attack is a good indication of the level of anonymity left in the system after the attack. It is well known (see, for example, Asratian et al. [4]) that this number is the same as the permanent of the biadjacency matrix of that graph. The *permanent* of any $n \times n$ matrix $M = [m_{ij}]$ of real numbers is defined as:

$$\text{per}(M) = \sum_{\pi \in S_n} m_{1\pi(1)} m_{2\pi(2)} \cdots m_{n\pi(n)},$$

where S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$. It can be seen that the graph of Fig. 2(b) allows 4 perfect matchings, and that is also the permanent of its biadjacency matrix in Fig. 2(c).

Given any n by n 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 $n!$, when G is the complete bipartite graph $K_{n,n}$. Therefore, the maximum value of per(A) is $n!$, 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 (Infeasibility Attacks Metric). *Edman et al. [3] define a system's* degree of anonymity *after an infeasibility attack that results in an $n \times n$ biadjacency matrix A as:*

$$d(A) = \begin{cases} 0 & \textit{if } n = 1, \\ \dfrac{\textit{log}(\textit{per}(A))}{\textit{log}(n!)} & \textit{otherwise.} \end{cases}$$

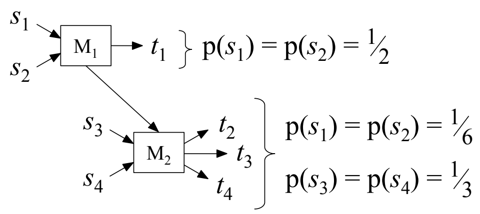
The above anonymity metric is reasonable as it compares the number of perfect matchings deemed feasible by the attack with their maximum number. Note that $0 \leq d(A) \leq 1$. Also, $d(A) = 0$ iff A has just one perfect matching, i.e. the system provides no anonymity, and $d(A) = 1$ iff $n > 1$ and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

The matrix of Fig. 2(c) contains 4 perfect matchings out of the 24 maximum possible. By the above metric, the system's degree of anonymity after that attack is $\log(4) / \log(24) \approx 0.436$.

2.2 A Metric for Probabilistic Attacks

Unlike infeasibility attacks, that simply label edges of the system's complete bipartite graph as being feasible or infeasible, probabilistic attacks assign to each edge of the graph a real value between 0 and 1 as that edge's probability of being a part of the actual communication pattern.

As an example of this attack, consider the simple mix network shown in Fig. 3(a), with two mix nodes, M_1 and M_2, and four input as well as output messages. The message from mix M_1 to M_2 is internal to the network. As discussed in Serjantov and Danezis [1], suppose each mix node randomly shuffles all its input messages before sending them out, i.e. a message entering any mix node is equally likely to appear as any of that node's output messages. If this characteristic of mix nodes is known to the attacker, and the entire message flow pattern of the network (including internal messages) is visible to the attacker, the attacker can arrive at probabilities for each input-output message pairing of the system, as shown next to the output messages in Fig. 3(a). These probabilities are essentially labels produced by the attack on edges of the system's



(a)

P	t_1	t_2	t_3	t_4
s_1	1/2	1/6	1/6	1/6
s_2	1/2	1/6	1/6	1/6
s_3	0	1/3	1/3	1/3
s_4	0	1/3	1/3	1/3

(b)

Fig. 3. (a) Message flow via a mix network, observed by attacker to arrive at probabilities of input-output message pairings. (b) Probability matrix of this network.

complete bipartite graph, and can be arranged as a *probability matrix* $P = [p_{ij}]$, as shown in Fig. 3(b). Any entry p_{ij} in this matrix contains the probability that the system's input message s_i appeared as its output message t_j. Real values from the closed interval $[0, 1]$ are used for probabilities.

A probability matrix produced by an attack is *doubly-stochastic*, i.e. the sum of all values in any of its rows or columns is 1. This follows from the assumption that each input message must appear as some output message, and each output message must have been one of the input messages. The maximum value of the permanent of an $n \times n$ probability matrix P is 1 (see Propositions 1 and 2 in Section 3.1), when P 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 determined all input-output message pairings with full certainty. The minimum value of per(P) is well known to be $n!/n^n$, when all entries in P are $1/n$ (see, for example, Egorychev [5]). This corresponds to the system providing full anonymity.

Definition 2 (Probabilistic Attacks Metric). *For any probabilistic attack resulting in an $n \times n$ probability matrix P, Edman et al. [3] define the system's* degree of anonymity *after that attack as:*

$$D(P) = \begin{cases} 0 & \textit{if } n = 1, \\ \dfrac{\textit{log}(\textit{per}(P))}{\textit{log}(n!/n^n)} & \textit{otherwise.} \end{cases}$$

The permanent of the matrix of Fig. 3(b) works out to $1/9 \approx 0.11111$, while the minimum value of the permanent of a 4×4 probability matrix is $4!/4^4 = 0.09375$. By the above metric, the system's degree of anonymity after this attack is $\log(1/9)\ /\ \log(4!/4^4) \approx 0.9282$.

A Note on Our Naming Convention and Figures. As the rest of this paper deals with two different types of matrices, namely *biadjacency* matrices that have 0 and 1 entries and *probability* matrices with real values in the closed interval $[0, 1]$ as their entries, we adopt a consistent naming convention while discussing them. The name A is always used for discussing any biadjacency matrix, and P for any probability matrix. When the type of a matrix under consideration is not important, we use the name M.

In figures, biadjacency matrices are displayed in the plain format, as in Fig. 2(c), and probability matrices with shaded row and column titles, as in Fig. 3(b).

Finally, the infeasibility attacks metric d of Edman et al. [3], given in Definition 1, is defined for biadjacency matrices, while their probabilistic attacks metric D, given in Definition 2, is for probability matrices.

3 Shortcomings of Metric for Probabilistic Attacks

It is instructive to recapitulate the ranges of the permanent of matrices considered so far. These ranges are shown in Fig. 4. There are some similarities

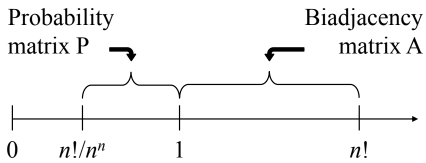


Fig. 4. Ranges of permanent: For an $n \times n$ biadjacency matrix A, per(A) is an integer from the set $\{1, 2, \ldots, n!\}$, and for an $n \times n$ probability matrix P, per(P) is a real value in the range $[n!/n^n, 1]$

between the metric expressions proposed by Edman et al. [3] for infeasibility attacks given by Definition 1 and probabilistic attacks given by Definition 2. First, in both cases, the argument of the logarithm in the denominator is the permanent of the matrix that corresponds to full anonymity. Second, the farther away from 1 the permanent of the underlying matrix (A for an infeasibility attack, and P for a probabilistic attack), the larger the system's degree of anonymity.

Despite these similarities, while the metric for infeasibility attacks in Definition 1 is sound, we show that the metric for probabilistic attacks in Definition 2 is not a good one. In this section, we demonstrate some shortcomings of this metric and, in the next section, we propose a better metric for probabilistic attacks.

3.1 Inadequacy of Matrix Permanent

The first shortcoming of the metric in Definition 2 for probabilistic attacks is that it is a function of just the permanent of the probability matrix. While the value of the permanent is necessary to take into account, we will show that it is not sufficient.

An Intuitive Understanding of Permanent. We begin by gaining a better understanding of the permanent of a matrix. Recall that S and T are the sets of n input and output messages of the system. Given any $n \times n$ biadjacency or probability matrix M, we define a *thread* of M to be any subset of its cells that contains exactly one cell from each row of M. Each thread therefore has exactly n cells. Additionally, a thread of M is a *diagonal* if no two of its cells lie in the same column of M. Let $\mathcal{T}(M)$ and $\mathcal{X}(M)$ denote, respectively, the sets of all threads and diagonals of M. Note that, a cell in the matrix M corresponds to an edge of the system's complete bipartite graph between S and T, a thread corresponds to a subgraph of that graph obtained by removing all but one edge connected to each $s \in S$ (i.e. a function from S to T), and a diagonal corresponds to a perfect matching between S and T. Clearly, M has n^n threads, of which $n!$ are diagonals.

Let the *weight* of any thread t of M, denoted $\mathcal{W}(t)$, be the product of values in all cells of t. The following proposition follows immediately from the definitions so far.

Proposition 1. *For any biadjacency or probability matrix M,*

$$\sum_{x \in \mathcal{X}(M)} \mathcal{W}(x) = per(M).$$

In other words, per(M) is the composite sum of weights of all diagonals of M. We first make the following important observation:

> *The values in M induce not just its permanent, but also a weight distribution on all its threads, including diagonals.*

Next, we improve our intuitive understanding of the permanent of a probability matrix by taking a closer look at the *information content* in it. The following proposition is also straightforward.

Proposition 2. *For any probability matrix P,*

$$\sum_{t \in \mathcal{T}(P)} \mathcal{W}(t) = 1.$$

Proof. Let P be $n \times n$. By definitions and algebraic rearrangement we have,

$$\sum_{t \in \mathcal{T}(P)} \mathcal{W}(t) = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} p_{1j_1} p_{2j_2} \cdots p_{nj_n} = \prod_{i=1}^{n} (p_{i1} + p_{i2} + \cdots + p_{in}) = 1.$$

The last equality follows from the fact that the sum of each row of P is 1. $\square$

Consider the set T^S of all n^n functions $f : S \to T$. By assigning a probability to each edge in the set $S \times T$, the matrix P ends up inducing a probability on each function in T^S. The probability that P associates with any function $f \in T^S$ is $\prod\{p_{ij} \mid f(s_i) = t_j\}$, i.e. the weight of the thread in P corresponding to f. By Proposition 2, these weights add up to 1, i.e. we have a probability distribution on the entire set T^S. If a function f is now picked randomly from the set T^S according to the probability distribution defined by P, then by Proposition 1, per(P) is the probability that f is a bijection, i.e. a perfect matching between S and T. The weights of the individual diagonals of P are the probabilities associated by P to their corresponding perfect matchings of being the true communication pattern of the system.[1]

A Better Indicator of Anonymity. Since the system's goal is to blend the true message communication pattern among others, the system's degree of anonymity should not be determined by simply answering the question:

> *What is the composite permanent of P?*

[1] As all column sums of P are also 1, P induces a similar probability distribution on the set S^T of all n^n functions $f : T \to S$. However, the bijections in S^T correspond to the bijections in T^S, and get identical probabilities in both distributions. This distribution therefore casts no further light on the meaning of per(P).

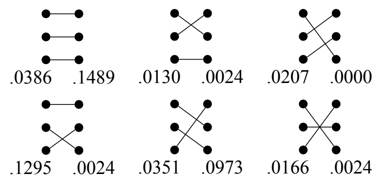
The quintessential question is, rather:

How evenly is the permanent of P distributed as its diagonal weights?

By Proposition 1, it is possible for two matrices, say P_1 and P_2, to have identical permanents, but a significantly different diagonal weight distribution. If the weights of all diagonals of P_1 are closer to each other in comparison with those of P_2, then the system underlying P_1 should be considered as providing better anonymity, because the attack has better succeeded in exposing some of the perfect matchings of P_2 as being the likely ones.

The example in Fig. 5 illustrates this phenomenon on 3×3 matrices. The

$\boldsymbol{P_1}$	t_1	t_2	t_3
s_1	.53	.25	.22
s_2	.20	.28	.52
s_3	.27	.47	.26



$\boldsymbol{P_2}$	t_1	t_2	t_3
s_1	.53	.46	.01
s_2	.01	.53	.46
s_3	.46	.01	.53

Fig. 5. Two probability matrices with nearly identical permanent, 0.2535, but significantly different diagonal weight distributions (for each perfect matching, weights according to P_1 and P_2 shown of its corresponding diagonal)

diagonal weight distributions of these two matrices, in non-decreasing order, are:

$$P_1\colon \langle 0.0130, 0.0166, 0.0207, 0.0351, 0.0386, 0.1295\rangle,$$
$$P_2\colon \langle 0.0000, 0.0024, 0.0024, 0.0024, 0.0973, 0.1489\rangle.$$

Clearly, the weights of the diagonals of P_1 are more evenly distributed than those of P_2. Yet, $D(P_1) \approx D(P_2)$, because $\mathrm{per}(P_1) \approx \mathrm{per}(P_2)$. Later, in Section 4, we propose another metric that, by taking the diagonal weight distribution into account, ends up assigning almost twice as high degree of anonymity to the system underlying P_1 than to that of P_2.

Region of Acute Inadequacy of Permanent. Let the *diameter* of an $n \times n$ probability matrix P be the largest difference between weights of any two of its diagonals, i.e.

$$\max\{\mathcal{W}(x_1) - \mathcal{W}(x_2) \mid x_1, x_2 \in \mathcal{X}(P)\}.$$

Just as the permanent of P, its diameter is another rough indicator of the degree of anonymity of the underlying system. In general, the smaller the diameter, the higher the anonymity.

For any possible permanent value $p \in [n!/n^n, 1]$, let $\mathfrak{M}(p)$ be the set of all $n \times n$ probability matrices with permanent p. As illustrated in Fig. 6 for $n = 3$, for any value of p that is close to 1 or extremely close to $n!/n^n$, the diameters of all matrices in $\mathfrak{M}(p)$ are roughly the same. Using just p to determine the system's anonymity level for such matrices, although inaccurate, is somewhat acceptable. However, for any other value of p, i.e. in the middle range, matrices in $\mathfrak{M}(p)$ vary

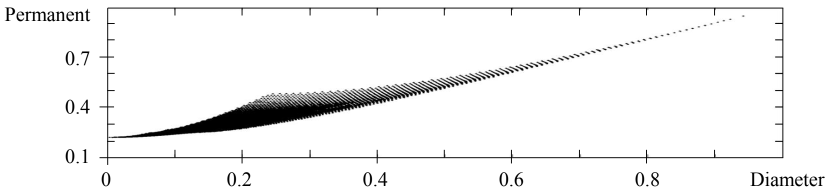


Fig. 6. Diameter spread of possible permanent values of 3×3 probability matrices

significantly in their diameters. It is in this region, where it is critical to consider the entire diagonal weight distribution of a probability matrix to determine the system's anonymity level, rather than just its permanent.

We end this discussion with the observation that the permanent of matrices in the example of Fig. 5 is approximately 0.2535. From Fig. 6 we can tell that the diameters of these two matrices are in fact not as far apart from each other as can be for some other two matrices with permanent, say around 0.4. Thus, even more convincing examples can be constructed to demonstrate the inadequacy of permanents as sole indicators of the anonymity level.

3.2 Incorrect Generalization of Infeasibility Attacks Metric

Another shortcoming of the metric in Definition 2 for probabilistic attacks is that it is not a generalization of the metric in Definition 1 for infeasibility attacks, despite the fact that probabilistic attacks are, in a sense, a generalization of infeasibility ones. We state this more precisely by giving an information-preserving embedding of infeasibility attacks into the wider class of probabilistic ones.

Diagonal Weight Profile. Let $\langle X_1, X_2, \ldots, X_{n!} \rangle$ be the sequence of diagonals of any $n \times n$ matrix M, ordered by the lexicographic ordering on their underlying index sets. In other words, if $\{(1, i_1), (2, i_2), \ldots, (n, i_n)\}$ is the set of indices of cells in a diagonal X_i, and $\{(1, j_1), (2, j_2), \ldots, (n, j_n)\}$ is the set of indices of cells in a diagonal X_j, then $i < j$ iff for some c, $i_c < j_c$ and for all $k < c$, $i_k = j_k$.

We define the *diagonal weight profile* (or just *profile*) of M to be the normalized sequence of weights of diagonals in the above sequence, given by:

$$\text{profile}(M) = \frac{1}{\text{per}(M)} \langle \mathcal{W}(X_1), \mathcal{W}(X_2), \ldots, \mathcal{W}(X_{n!}) \rangle.$$

As this paper only deals with matrices that have strictly positive permanents, the above sequence is well defined. A fixed ordering of diagonal weights in profiles, such as the lexicographic one given above, together with normalization, enable us to compare weights of corresponding diagonals across matrices.

From Proposition 1, it is seen that profile(M) is a probability distribution on the diagonals of M, i.e. perfect matchings of its underlying bipartite graph. From the point of view of a *system-wide* anonymity metric, this is the most vital piece of information contained in M.

A Profile-Preserving Embedding. Let A be an $n \times n$ biadjacency matrix resulting from an infeasibility attack. Exactly per(A) values in profile(A) are 1/per(A), and the remaining values are 0. The metric $d(A)$ of Definition 1 is based on the premise that each of the per(A) feasible perfect matchings corresponding to the nonzero values in profile(A) are equally likely, and the remaining are not possible. We now proceed to construct a unique probability matrix $\mathcal{C}_A$ with the same profile as A. We will then show that while it is desirable and expected that $D(\mathcal{C}_A) = d(A)$, in general it is not so.

We begin by observing that the reduced bipartite graph underlying A may contain edges that do not appear in any perfect matching as, for example, the edge $\langle s_3, t_2 \rangle$ in Fig. 2(b) and (c). Such nonzero entries in A are harmless since, by not being on any diagonal with nonzero weight, their presence affects neither per(A) nor profile(A), thus also not $d(A)$. Let $\hat{A} = [\hat{a}_{ij}]$ be the matrix identical to A, except that $\hat{A}$ contains a 0 entry for all such edges.

Now, let $\mathfrak{P}(A)$ be the set of all possible (doubly-stochastic) probability matrices conforming to the graph underlying A, i.e.

$$\mathfrak{P}(A) = \{n \times n \text{ probability matrix } P = [p_{ij}] \mid \\ p_{ij} = 0 \text{ if } \hat{a}_{ij} = 0, \text{ for all } i, j\}.$$

In other words, $\mathfrak{P}(A)$ contains all possible probability distributions on the edges declared feasible by A. It is well known that $\mathfrak{P}(A)$ is nonempty iff per$(A) > 0$ (see, for example, Theorem 2.2.3 in Bapat and Raghavan [6]). Observe that any $P \in \mathfrak{P}(A)$ has no less information than A as it contains some probability distribution *in addition to* the feasibility information in A, i.e. an attack resulting in P is at least as strong as one resulting in A. It is therefore expected and desirable that $D(P) \leq d(A)$, but that does not always hold as the example matrix in Fig. 7 illustrates. This matrix, P, is chosen arbitrarily from $\mathfrak{P}(A)$, for the biadjacency matrix A in Fig. 2(c). While $d(A) \approx 0.436$, as computed at the end of Section 2.1, we have that $D(P) \approx 0.491$, a larger value. This phenomenon does not conform to the intuition behind anonymity metrics.

Let an $n \times n$ matrix $S = [s_{ij}]$ be called a *scaling* of an $n \times n$ matrix $M = [m_{ij}]$ if for some multiplier vectors $R = \langle r_1, r_2, \ldots, r_n \rangle$ and $C = \langle c_1, c_2, \ldots, c_n \rangle$ with strictly positive values, $s_{ij} = r_i m_{ij} c_j$, for all i, j. It is easily verified that the weight of any diagonal of S is the weight of the corresponding diagonal of M,

P	t_1	t_2	t_3	t_4
s_1	½	½	0	0
s_2	½	½	0	0
s_3	0	0	¼	¾
s_4	0	0	¾	¼

Fig. 7. An example $P \in \mathfrak{P}(A)$, for biadjacency matrix A of Fig. 2(c)

multiplied by the scaling factor $\lambda = \prod_{i=1}^{n} r_i c_i$. Thus, $\mathrm{per}(S) = \lambda \cdot \mathrm{per}(M)$ as well. This leads to the following proposition.

Proposition 3. *If S is a scaling of M, then profile(S) = profile(M).*

We let $\mathfrak{S}(M)$ denote the set of all scalings of M.

Theorem 1. *For any $n \times n$ biadjacency matrix A resulting from an infeasibility attack, $\mathfrak{P}(A) \cap \mathfrak{S}(\hat{A})$ is a singleton set.*

Proof. When $\mathrm{per}(A) > 0$, that the intersection is nonempty was established by Brualdi, Parter and Schneider [7]. Uniqueness, when nonempty, follows from the fact that distinct doubly-stochastic matrices cannot have identical profiles, given as Corollary 2.6.6 in Bapat and Raghavan [6]. □

The sole member of $\mathfrak{P}(A) \cap \mathfrak{S}(\hat{A})$ is the unique *canonical* probability matrix for A, denoted $\mathcal{C}_A$. It is the only doubly-stochastic matrix whose profile is identical to that of A. Fig. 8 shows an example matrix A, along with its $\mathcal{C}_A$. The matrix $\mathcal{C}_A$

A	t_1	t_2	t_3
s_1	0	1	1
s_2	1	0	1
s_3	1	1	1

$\mathcal{C}_A$	t_1	t_2	t_3
s_1	0	$(\sqrt{5}-1)/2$	$(3-\sqrt{5})/2$
s_2	$(\sqrt{5}-1)/2$	0	$(3-\sqrt{5})/2$
s_3	$(3-\sqrt{5})/2$	$(3-\sqrt{5})/2$	$\sqrt{5}-2$

Fig. 8. A biadjacency matrix A and its canonical probability matrix $\mathcal{C}_A$

can be viewed as the result of a probabilistic attack that has arrived at the same conclusion as the infeasibility attack resulting in A, in that the sets of perfect matchings called feasible by these attacks coincide and all those feasible perfect matchings are deemed equally likely by both attacks. As these two attacks are equally strong (in fact, identical), it is desirable that $D(\mathcal{C}_A) = d(A)$.

For the matrices shown in Fig. 8, $\mathrm{per}(A) = 3$ and $\mathrm{per}(\mathcal{C}_A) = 3(5\sqrt{5} - 11)/2$. However, $\mathrm{profile}(A) = \mathrm{profile}(\mathcal{C}_A) = \langle 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0 \rangle$. And while $d(A) \approx 0.6131$, we have that $D(\mathcal{C}_A) \approx 0.8693 \neq d(A)$. Again, an undesirable behavior of the D metric. In Section 4, we present a new metric Δ that has the property $\Delta(\mathcal{C}_A) = d(A)$, for all biadjacency matrices A.

Construction of Canonical Probability Matrix. As for the construction of $\mathcal{C}_A$ from a given A, recall that $\mathcal{C}_A$ is a scaling of $\hat{A}$ with some row-multiplier vector R and column-multiplier vector C. For the example of Fig. 8, $A = \hat{A}$, and let $R = \langle r_1, r_2, r_3 \rangle$ and $C = \langle c_1, c_2, c_3 \rangle$. As the sums of the rows and columns of $\mathcal{C}_A$ should be 1, we get the following 6 equations:

$$
\begin{array}{lll}
r_1(c_2 + c_3) = 1 & r_2(c_1 + c_3) = 1 & r_3(c_1 + c_2 + c_3) = 1 \\
c_1(r_2 + r_3) = 1 & c_2(r_1 + r_3) = 1 & c_3(r_1 + r_2 + r_3) = 1
\end{array}
$$

We seek solutions to the above system of equations in which all r_i's and c_i's are positive. One solution for this particular scaling is:

$$R = \left\langle \frac{3-\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}, \sqrt{5}-2 \right\rangle, \quad C = \left\langle \frac{1+\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}, 1 \right\rangle .$$

Although there are multiple such solutions, Sinkhorn [8] showed that all solutions are unique up to a scalar factor, i.e. if (R_1, C_1) and (R_2, C_2) are solutions to the above, then for some $\alpha > 0$, $R_2 = R_1\alpha$ and $C_2 = C_1/\alpha$. However, due to the uniqueness of $\mathcal{C}_A$ all solutions lead to the same resulting matrix.

Sinkhorn and Knopp [9] gave another interesting characterization of $\mathcal{C}_A$ as the limit of an infinite sequence of matrices. Let f, g and h be functions from and to $n \times n$ real matrices, defined as follows:

$$\begin{aligned} f(M)_{ij} &= M_{ij} \;/\; \textstyle\sum_{k=1}^{n} M_{ik} && (f \text{ normalizes each row of } M) \\ g(M)_{ij} &= M_{ij} \;/\; \textstyle\sum_{k=1}^{n} M_{kj} && (g \text{ normalizes each column of } M) \\ h(M) &= g(f(M)) \end{aligned}$$

Then, $\mathcal{C}_A = \lim_{k\to\infty} h^k(A)$. In other words, a procedure that alternately normalizes all rows followed by all columns of A, ad infinitum, would converge to $\mathcal{C}_A$. The accumulated row and column multipliers along the way also converge to the correct R and C values. However, as A contains just 0-1 values, multipliers accumulated after any finite number of iterations are only rational. As the example in Fig. 8 shows, the final solution can be irrational, the limit of an infinite sequence of rational approximations. So in general, this procedure requires an infinite number of iterations. A number of efficient algorithms have therefore been considered, as in Kalantari and Khachiyan [10] and Linial, Samorodnitsky and Wigderson [11], for producing in a finite number of steps, approximate solutions that are within acceptable error bounds.

4 A More Accurate Metric for Probabilistic Attacks

We now present a new metric for probabilistic attacks that overcomes the shortcomings mentioned in the previous section of the metric D of Edman et al. [3]. By being sensitive to the distribution of the permanent of a given probability matrix over its diagonals, the new metric results in a more accurate measurement of the underlying system's degree of anonymity. Furthermore, this metric has the welcome trait of correctly treating probabilistic attacks as generalizations of infeasibility attacks. This feature is exploited to make just this one metric suffice for both kinds of attacks.

The fundamental premise upon which our metric is constructed is that the permanent of a matrix can be broken down into a probability distribution over its diagonals, i.e. the perfect matchings of the system's complete bipartite graph. The profile of the matrix is essentially that distribution.

Ever since the works of Serjantov and Danezis [1] and Diaz et al. [2], 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 profile of the matrix as a measure of the attacker's uncertainty of which perfect matching is the system's true communication pattern.

Definition 3 (Unified Metric). *Let M be a given $n \times n$ biadjacency or probability matrix resulting from an attack, with profile$(M) = \langle w_1, w_2, \ldots, w_{n!} \rangle$. We define the underlying system's* degree of anonymity *after this attack as:*

$$\Delta(M) = \begin{cases} 0 & \text{if } n = 1, \\ \dfrac{-\sum_{i=1}^{n!} w_i \cdot log(w_i)}{log(n!)} & \text{otherwise.} \end{cases}$$

In the above summation, a subexpression $0 \cdot \log(0)$ is interpreted as 0.

Observe that the above metric Δ is for biadjacency *as well as* probability matrices, whereas the metrics of Edman et al. [3] for these two kinds of matrices were separate. Their metric d, given in Definition 1, was for biadjacency matrices, while their metric D, given in Definition 2, was for probability matrices. We first establish that for biadjacency matrices, our Δ coincides with d.

Theorem 2. *For any biadjacency matrix A, $d(A) = \Delta(A) = \Delta(\mathcal{C}_A)$.*

Proof. The second equality follows from the fact that A and $\mathcal{C}_A$ have identical profiles. To show the first equality, we recall from Section 3.2 that exactly per(A) values in profile(A) are $1/\text{per}(A)$, and the remaining values are 0. The numerator of the expression in Definition 3 thus becomes:

$$-\text{per}(A) \left[\frac{1}{\text{per}(A)} \cdot \log \left(\frac{1}{\text{per}(A)} \right) \right] = \log(\text{per}(A)),$$

which is the numerator of the expression in Definition 1 of Section 2.1. □

To understand the properties of our new metric better, we revisit some of our earlier examples. For the probability matrices P_1 and P_2 of Fig. 5 with equal permanent value of about 0.2535, we had that $D(P_1) \approx D(P_2) \approx 0.9124$. However, $\Delta(P_1) \approx 0.8030$, about twice as high as $\Delta(P_2) \approx 0.4544$. Our new metric Δ recognizes that the profile of P_2 is significantly more uneven than that of P_1, thus assigning the system underlying P_2 a far lower degree of anonymity.

For the biadjacency matrix of Fig. 2(c), we have $\Delta(A) = d(A) \approx 0.436$. The probability matrix P of Fig. 7 was arbitrarily chosen from the set $\mathfrak{P}(A)$. Of the 24 values in profile(P), $\langle \frac{1}{20}, \frac{9}{20}, \frac{1}{20}, \frac{9}{20} \rangle$ is the subsequence of nonzero values. While we saw that $D(P) \approx 0.491 > d(A)$, we have that $\Delta(P) \approx 0.3204 < d(A)$. This behavior conforms with our intuition that P has more information than A. The following theorem shows that this phenomenon is guaranteed by Δ.

Theorem 3. *For any biadjacency matrix A and $P \in \mathfrak{P}(A)$, such that $P \neq \mathcal{C}_A$, $\Delta(P) < \Delta(A)$.*

Proof. Let per$(A) = t$. Then, profile(A) has t nonzero values, and each of those values is $1/t$. Let $p_1, p_2, \ldots, p_t$ be the corresponding values in profile(P). As these are the only diagonals of P that may have nonzero weights, their sum is 1. We need to show that:

$$-\sum_{i=1}^{t} p_i \cdot \log(p_i) < -\sum_{i=1}^{t} (1/t) \cdot \log(1/t).$$

Although this property of Shannon entropy is well known in information theory (see, for example, Kapur [12] for a proof based on Jensen's inequality), here we give a short proof.

It is easily seen that, for all $\beta > 0$, we have $2\beta \leq 2^{\beta}$, with equality iff $\beta = 1$. Taking logarithms to the base 2 gives $1+\log(\beta) \leq \beta$. As we interpret $0 \cdot \log(0) = 0$, we can substitute $\beta = (1/t)/p_i$, and simplify, to get that for all i, $p_i - p_i \cdot \log(p_i) \leq (1/t) - p_i \cdot \log(1/t)$, with equality iff $p_i = 1/t$. Summation over all i gives:

$$-\sum_{i=1}^{t} p_i \cdot \log(p_i) \leq \log(t) = -\sum_{i=1}^{t} (1/t) \cdot \log(1/t).$$

As $P \neq \mathcal{C}_A$ and distinct doubly-stochastic matrices cannot have the same profile, we have that for some i, $p_i \neq (1/t)$, leading to a strict inequality. □

We end this section with an example that demonstrates how different our new metric Δ can be from the old metric D of Edman et al. [3]. Fig. 9 shows two matrices, P_1 and P_2 for which, according to the D metric, P_1 seems to result in less anonymity than P_2, as $D(P_1) \approx 0.5658 < 0.7564 \approx D(P_2)$. However,

P_1	t_1	t_2	t_3
s_1	.04	.04	.92
s_2	.48	.49	.03
s_3	.48	.47	.05

P_2	t_1	t_2	t_3
s_1	.65	.01	.34
s_2	.01	.34	.65
s_3	.34	.65	.01

Fig. 9. Two probability matrices for which $D(P_1) < D(P_2)$, but $\Delta(P_1) > \Delta(P_2)$

$\Delta(P_1) \approx 0.4132 > 0.2750 \approx \Delta(P_2)$, i.e. according to our new metric, P_1 results in higher anonymity than P_2.

5 Conclusions and Future Work

Edman, Sivrikaya and Yener [3] introduced a method for arriving at a system-wide measure of the level of anonymity provided by a system. Their approach is based upon a complete bipartite graph that models all possible input and output message associations of the system. By rendering infeasible some edges of this graph, an *infeasibility* attack results in a reduced graph, thereby lowering anonymity. They proposed adopting the permanent of the biadjacency matrix of this reduced graph to determine the amount of anonymity remaining in the system in the aftermath of the attack.

Edman et al. [3] then suggest adopting a similar technique for a wider class of *probabilistic* attacks that, instead of removing infeasible edges from the system's complete bipartite graph, assign probabilities to all edges.

In this paper, we argue that while the metric given in [3] for the narrower class of infeasibility attacks is sound, their metric for probabilistic attacks has shortcomings. We show why using just the permanent of the underlying matrix for probabilistic attacks is inaccurate, as it at best gives only a rough measure of the system's anonymity level. We also show that this technique fails to correctly treat probabilistic attacks as generalizations of infeasibility ones.

We then present a new metric that overcomes these shortcomings. By recognizing that the permanent of a matrix can be broken down into a probability distribution on the perfect matchings of the underlying bipartite graph, our new metric provides an accurate measure of anonymity. It also has the desirable property of being a unified metric for both classes of attacks.

The basic metric of [3] for infeasibility attacks has since been extended for modified scenarios. Gierlichs et al. [13] enhanced it for situations where system users send or receive multiple messages. The equivalence relation on perfect matchings, induced by such multiplicity, causes a reduction in anonymity. Bagai and Tang [14] analyzed the effect of employing data caching within the mix network. Their modified metric captures an increase in anonymity due to such caching. We leave such extensions to the new metric proposed in this paper as future work.

Acknowledgements. We would like to thank Andrew Bradley of Stanford University for helpful discussions on matrix scaling that led to the example of Fig. 8. The research described in this paper has been partially supported by the United States Navy Engineering Logistics Office contract no. N41756-08-C-3077.

References

1. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In: Proceedings of the 2nd Privacy Enhancing Technologies Workshop, San Francisco, USA, pp. 41–53 (2002)
2. Diaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring Anonymity. In: Proceedings of the 2nd Privacy Enhancing Technologies Workshop, San Francisco, USA, pp. 54–68 (2002)
3. Edman, M., Sivrikaya, F., Yener, B.: A combinatorial approach to measuring anonymity. In: IEEE International Conference on Intelligence and Security Informatics, New Brunswick, USA, pp. 356–363 (2007)
4. Asratian, A., Denley, T., Häggkvist, R.: Bipartite graphs and their applications. Cambridge University Press, Cambridge (1998)
5. Egorychev, G.: The solution of van der Waerden's problem for permanents. Advances in Mathematics 42(3), 299–305 (1981)
6. Bapat, R., Raghavan, T.: Nonnegative matrices and applications. Cambridge University Press, Cambridge (1997)

7. Brualdi, R., Parter, S., Schneider, H.: The diagonal equivalence of a nonnegative matrix to a stochastic matrix. J. Mathematical Analysis and Applications 16(1), 31–50 (1966)
8. Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. Annals of Mathematical Statistics 35, 876–879 (1964)
9. Sinkhorn, R., Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices. Pacific Journal of Mathematics 21(2), 343–348 (1967)
10. Kalantari, B., Khachiyan, L.: On the complexity of nonnegative matrix scaling. Linear Algebra and its Applications 240, 87–103 (1996)
11. Linial, N., Samorodnitsky, A., Wigderson, A.: A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. Combinatorica 20(4), 545–568 (2000)
12. Kapur, J.: Maximum entropy models in science and engineering, 2nd edn. New Age International Publishers (2009)
13. Gierlichs, B., Troncoso, C., Diaz, C., Preneel, B., Verbauwhede, I.: Revisiting a combinatorial approach toward measuring anonymity. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society, Alexandria, USA, pp. 111–116 (2008)
14. Bagai, R., Tang, B.: Data caching for enhancing anonymity. In: Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications, Singapore, pp. 135–142 (2011)