Effectiveness of Probabilistic Attacks on Anonymity of Users Communicating via Multiple Messages

Rajiv Bagai, Bin Tang, *Member, IEEE,* and Euna Kim

***Abstract*—A major objective of any system-wide attack on an anonymity system is to uncover the extent to which each user of the system communicated with each other user. A probabilistic attack attempts to achieve this objective by arriving at some probability values for each of the system's possible input–output message pairs of reflecting actual communication. We show that these values lead to a probability distribution on the set of all possible system-wide communication patterns between users, and develop a combinatorial technique to determine this distribution. We give a method to measure from this distribution the effectiveness of any such attack or, alternatively, the level of anonymity remaining in the system in the aftermath of the attack. We also compare our metric with three earlier attempts in the literature to solve a similar problem, and demonstrate that the scope of our metric is far wider than those of all earlier ones.**

***Index Terms*—Combinatorial matrix theory, probabilistic attacks, system-wide anonymity metric.**

I. Introduction

THE IMPORTANCE of accurate techniques for measuring the effectiveness of attacks on anonymity systems has been well recognized for a long time. Such techniques help determine the amount of anonymity that still remains in an anonymity system in the aftermath of an attack.

Much of the well-known work in this area, such as that of Serjantov and Danezis [1] or of Diaz *et al.* [2], has focused on measuring effectiveness of attacks that aim to uncover the sender (or receiver) of a single message passing through the system. In contrast, Bagai *et al.* [3] and Edman *et al.* [4] studied system-wide attacks that attempt to link each input message of an anonymity system with the corresponding output message it exited the system as. Specifically, they studied the following two classes of attacks.

1) **A-Inf:** Attacks whose analysis determines infeasibility of some of the system's input–output message pairings of being actual linkages. Attacks in this class are called infeasibility attacks.
2) **A-Prob:** Attacks whose analysis arrives at probability values for each system's input–output message pairing of being an actual linkage. Attacks in this class are called probabilistic attacks.

Measuring the effectiveness of any attack in the above classes results in a system-wide measure of the level of anonymity provided by the anonymity system. By presenting an information-preserving embedding of the class **A-Inf** into the class **A-Prob**, Bagai *et al.* [3] showed that the latter class is significantly wider and more general than the former. They also constructed a unified metric for measuring the anonymity left in the system upon conclusion of any attack in these classes. For attacks in **A-Inf**, this unified metric was shown to coincide with an earlier metric proposed by Edman *et al.* [4], while for attacks in **A-Prob** the unified metric of Bagai *et al.* [3] was shown to be more accurate than one that appeared in [4].

Gierlichs *et al.* [5] demonstrated the need for considering linkages between the senders and receivers of the anonymity system, instead of just its input and output messages. They argued that since users may send/receive multiple messages and the system attempts to hide the number of messages any particular sender sent to any particular receiver, the anonymity of linkages between senders and receivers is, in general, lower than that of exact linkages between messages. A modification of the basic metrics, namely the metric of Edman *et al.* [4] for attacks in **A-Inf**, and the unified metric of Bagai *et al.* [3] for attacks in **A-Inf** and/or **A-Prob**, thus became necessary in order to take multiple messages sent/received into account.

To date, three works have addressed the issue of modifying the basic metrics for the multiple messages sent/received scenario. Grégoire and Hamel [6] only reiterated the need for a modified metric for attacks in **A-Inf**, but did not propose any metric that takes an attack into account. Gierlichs *et al.* [5] presented a modified metric for attacks in **A-Inf**, but their metric was later shown to be applicable not to this entire class, rather just a portion of it, whose size becomes smaller as the numbers of messages sent/received by users grow. Bagai *et al.* [7] succeeded in developing a modified metric for the entire class **A-Inf**. None of these earlier works, however, considered probabilistic attacks, i.e., attacks in the wider class **A-Prob**. Later in our paper we explain in more detail the exact limitations of each of these related works.

The main contribution of our paper is a technique to measure the system-wide anonymity in the wake of any probabilistic attack, when users may send/receive multiple messages. Our technique is built upon the basic unified metric of Bagai *et al.* [3] that employs Shannon entropy [8] to

Manuscript received September 30, 2011; revised March 17, 2012; accepted June 3, 2012. This work was supported in part by the U.S. National Science Foundation under Grant CNS-1116849 and by the U.S. Navy Engineering Logistics Office under Contract N41756-08-C-3077.

R. Bagai and E. Kim are with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260 USA (e-mail: rajiv.bagai@wichita.edu; exkim@wichita.edu).

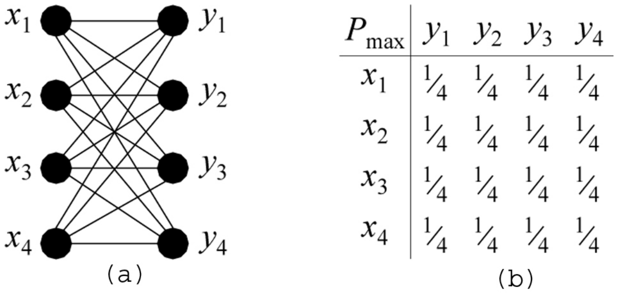
B. Tang is with the Department of Computer Science, Azusa Pacific University, Azusa, CA 91702 USA (e-mail: btang@apu.edu).

Digital Object Identifier 10.1109/JSYST.2012.2221855

measure such an attacker's uncertainty of which matching between the system's input and output messages is the actual communication pattern. In the presence of multiple messages sent/received by users, an equivalence relation is induced by message multiplicities on the set of all such matchings. Equivalence classes of this relation correspond to the possible levels of communication between senders and receivers. While only one of these equivalence classes reflects the actual level of communication, we develop a method to compute the probability distribution arrived at by the attacker on the set of all equivalence classes, of being the actual one. The attacker's uncertainty inherent in this distribution correctly measures the anonymity of the actual communication level between the system's senders and receivers.

Our technique for measuring anonymity is significant both practically as well as theoretically. Its practical significance is on two accounts. First, as already pointed out by Gierlichs *et al.* [5], its underlying system model that takes into consideration multiple messages sent/received by users accurately reflects reality. In contrast, the basic metric of Bagai *et al.* [3] ignores that important practical aspect. Second, it is well-known and reiterated in [9] that most attacks on real anonymous systems are probabilistic, i.e., members of the class **A-Prob**. Our technique is tailored for such attacks, while all earlier attempts at measuring anonymity in the multiple messages sent/received scenario, namely attempts by Gierlichs *et al.* [5], Grégoire and Hamel [6], and Bagai *et al.* [7], consider attacks only in the subclass **A-Inf**. The theoretical significance of our technique lies mainly in its scope, and stems from the fact that the class **A-Inf** is just a finite subclass of the uncountably infinite class **A-Prob**. As shown later in this paper, even for the significantly modest class **A-Inf**, only one of the three earlier attempts at arriving at a metric for measuring anonymity for the multiple messages scenario succeeded completely. The other attempts were at best partial. On the other hand, the technique developed in this paper, based on certain regions of probability matrices and extracts within those regions described in Section V, results in accurate anonymity measurement for any attack in **A-Prob**.

The rest of this paper is organized as follows. Section II gives an overview of the model of the anonymity system considered by Bagai *et al.* [3], an example of a probabilistic attack on such a system, and their basic anonymity metric. Section III extends this model by adding users that may send and/or receive multiple messages. This section then introduces the central concept of an equivalence relation induced by message multiplicities on the set of all matchings between messages, discusses the number of equivalence classes of this relation, and gives a method to compute the size of any class. Section IV presents our new anonymity metric, which is based on a probability distribution, arrived at by the attacker, on these classes. A technique for determining this distribution is then developed in Section V, and an application of our method to measure anonymity provided by a pool mix is contained in Section VI. Section VII compares our method with the three existing attempts in the literature for solving a similar problem. It first establishes that each of the earlier works only attempted to measure effectiveness of attacks in the finite class **A-Inf**, which is essentially a subclass of the uncountably infinite class **A-Prob**, to which our new method is applicable. It then characterizes the exact scope, within **A-Inf**, of the earlier methods. Finally, Section VIII concludes our work and mentions some directions for future work.



$$\begin{array}{c|cccc} P_{\max} & y_1 & y_2 & y_3 & y_4 \\ \hline x_1 & 1/4 & 1/4 & 1/4 & 1/4 \\ x_2 & 1/4 & 1/4 & 1/4 & 1/4 \\ x_3 & 1/4 & 1/4 & 1/4 & 1/4 \\ x_4 & 1/4 & 1/4 & 1/4 & 1/4 \end{array}$$

Fig. 1. Example for $t = 4$. (a) System's complete bipartite graph $K_{t,t}$ between X and Y. (b) $t \times t$ probability matrix $P_{\max}$ for the maximum anonymity case.

II. A System-Wide Anonymity Metric

In this section we give an overview of the technique of Bagai *et al.* [3] for measuring the amount of anonymity remaining in an anonymity system, after a probabilistic attack has been carried out. Their method determines the system-wide level of anonymity provided to messages sent via the system, rather than to any particular message going through it.

A. The Underlying Model

Let $X = \{x_1, x_2, \ldots, x_t\}$ be the set of t input messages observed by an attacker having entered an anonymity system, and $Y = \{y_1, y_2, \ldots, y_t\}$ be the set of output messages observed by the attacker having exited from that system. We assume that every input message eventually appears at the output, and that no message originates from within the system, thus $|X| = |Y| = t$.

Definition 1 (Matchings): A matching is any one-to-one correspondence between X and Y. The set of all matchings is denoted by $\mathfrak{M}$. Of all the $t!$ matchings in $\mathfrak{M}$, the unique one reflecting the system's actual communication pattern is called genuine; all others are called fake.

The foremost goal of the anonymity system is to hide from the attacker which input message in X exited the system as which output message in Y. In other words, the system attempts to hide the genuine matching by making fake matchings seem probable. 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 associations, or modifying message encoding by encryption/decryption to prevent message bit-pattern comparisons, etc.

While any given message $x_i \in X$ exits the system as some unique message $y_j \in Y$, all matchings are possible, and can therefore be modeled by the complete bipartite graph $K_{t,t}$ between X and Y, as shown in Fig. 1(a) for an example value of $t = 4$. In this graph, any vertex x_i is connected to all members of Y, indicating that x_i may have exited the system as any of its output messages. However, for any given x_i or

y_j, exactly one of the edges connected to that vertex is part of the genuine matching.

Maximum anonymity is achieved when for any message $x_i \in X$, each output message in Y appears to the attacker to be equally likely for being the one that x_i exited the system as. If each edge of the complete bipartite graph is assigned by the attacker a probability of its being a member of the genuine matching, then for this maximum anonymity situation, all edges uniformly get the value $1/t$ as their probability. We use real values from the closed interval $[0, 1]$ for probabilities, and employ a $t \times t$ probability matrix for storing attacker's probabilities for all edges in this graph. Fig. 1(b) shows the probability matrix $P_{\max}$ corresponding to maximum anonymity, for the example complete graph of Fig. 1(a).

The attacker, however, attempts to use some information gained about the system and/or messages to arrive at nonuniform probabilities for the graph's edges, i.e., some probability matrix that is preferably different than $P_{\max}$. Given that the genuine matching has to be a one-to-one correspondence between X and Y, any probability matrix resulting from an attack must be doubly stochastic, i.e., the sum of all values in any of its rows or columns is 1.

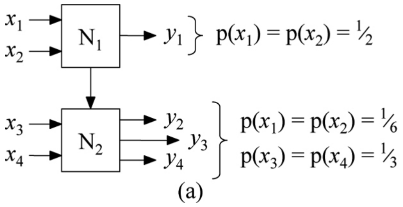
Ideally, the attacker would like to gain enough information for arriving at a matrix in which each row and column contains a single occurrence of 1 and $(t-1)$ occurrences of 0, as such a matrix pinpoints the genuine matching, and the system would then provide no anonymity. In general, however, the attacker has some partial information resulting in a probability matrix corresponding to an anonymity level that lies somewhere between the two extremes of maximum anonymity and no anonymity. Section II-B contains an example attack and its probability matrix, and Section II-C outlines the method of Bagai *et al.* [3] for measuring the anonymity level corresponding to any given probability matrix resulting from an attack.

B. Attack Example

As an example of a probabilistic attack, consider the simple anonymity system shown in Fig. 2(a), with two proxy nodes, N_1 and N_2, and four input as well as output messages. The message from node N_1 to N_2 is internal to the network. As discussed in [1], suppose each proxy node randomly shuffles all its input messages before sending them out, i.e., a message entering any proxy node is equally likely to appear as any of that node's output messages. If this characteristic of proxy 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. 2(a). The probability matrix P containing all these probabilities is shown in Fig. 2(b). Any entry P_{ij} in this matrix contains the probability that the system's input message x_i appeared as its output message y_j. Note that P is doubly stochastic.

C. Anonymity Metric

We briefly review the metric of Bagai *et al.* [3] for measuring the anonymity remaining in the system upon conclusion of a probabilistic attack that results in some $t \times t$ doubly stochastic probability matrix P. In a nutshell, their approach first recognizes the fact that P induces a certain probability value for any given matching of being the genuine one, and then employs the well-known technique of Shannon entropy [8] to measure the attacker's uncertainty contained in this probability distribution of which matching is indeed genuine.


x_1
x_2
N_1
y_1
$\mathrm{p}(x_1) = \mathrm{p}(x_2) = \frac{1}{2}$
x_3
x_4
N_2
y_2
y_3
y_4
$\mathrm{p}(x_1) = \mathrm{p}(x_2) = \frac{1}{6}$
$\mathrm{p}(x_3) = \mathrm{p}(x_4) = \frac{1}{3}$
(a)

P	y_1	y_2	y_3	y_4
x_1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
x_2	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
x_3	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
x_4	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

(b)

Fig. 2. (a) Example message flow via an anonymity system, observed by attacker to arrive at probabilities of input–output message pairings. (b) Probability matrix resulting from this attack.

Definition 2 (Lines): A line in P is any subset of its cells that contains exactly one cell from each row of P. The set of all lines in P is denoted by $\mathcal{L}(P)$.

Each line therefore has exactly t cells. As any cell in the matrix P corresponds to an edge of the system's complete bipartite graph between X and Y, a line corresponds to a subgraph of that graph obtained by removing all but one edge connected to each member of X, i.e., a function from X to Y.

Definition 3 (Line Weights): The weight of any line l in P, denoted by $\mathcal{W}(l)$, is the product of values in all cells of l.

We first make the following observation about the sum of weights of all lines.

Proposition 1: For any probability matrix P

$$\sum_{l \in \mathcal{L}(P)} \mathcal{W}(l) = 1.$$

Proof: From the definition of weights, and by algebraic rearrangement, we have

$$\sum_{l \in \mathcal{L}(P)} \mathcal{W}(l) = \sum_{j_1=1}^{t} \sum_{j_2=1}^{t} \cdots \sum_{j_t=1}^{t} P_{1j_1} P_{2j_2} \cdots P_{tj_t}$$

$$= \prod_{i=1}^{t} (P_{i1} + P_{i2} + \cdots + P_{it}) \;=\; 1.$$

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

Definition 4 (Diagonals): A line in P is called a diagonal if no two of its cells lie in the same column of P. The set of all diagonals in P is denoted by $\mathcal{D}(P)$.

Just as a line corresponds to a function from X to Y, a diagonal can be seen to correspond to a matching, i.e., a bijection between X and Y. Clearly, P has t^t lines, of which $t!$ are diagonals. The sum of weights of all diagonals in P is of special significance.

Definition 5 (Permanent): The permanent of P is given by

$$\mathrm{per}(P) = \sum_{d \in \mathcal{D}(P)} \mathcal{W}(d).$$

Permanents of square matrices have been the subject of much mathematical study (see [10] and [11]). As P is doubly stochastic, it is well known that $0 < t!/t^t \leq \text{per}(P) \leq 1$.

We now formulate the probabilities induced by P on each diagonal (i.e., matching) of being genuine. Consider the set Y^X of all t^t functions $f : X \rightarrow Y$. Suppose a function f is randomly chosen from Y^X by choosing independently, for each i

$$f(x_i) = y_j \text{ with probability } P_{ij}.$$

Since the sum of each row in P is 1, i.e., each input message in X must appear as some output message in Y, and the above choice is performed independently for each x_i, the probability of the event that our chosen function $f = f_0$, for any given function $f_0 \in Y^X$, is $\prod\{P_{ij} \mid f_0(x_i) = y_j\}$, i.e., the weight of the line in P corresponding to f_0. By Proposition 1, all such weights add up to 1, and we have a probability distribution on the entire set Y^X. It follows that, per(P) is the probability of the event that our randomly chosen function f is a bijection between X and Y, i.e., a matching. Furthermore, the weight of the diagonal corresponding to any given bijection $b_0 \in Y^X$ is the probability of the event that f chosen as above was b_0.

Of particular interest to us, however, is the probability of the event $f = b_0$, given the event f is a bijection. This is due to the additional constraint that the sum of each column of P is also 1, i.e., each output message in Y must have been some input message in X. This value is the probability of b_0 being the genuine matching, given that exactly one matching is genuine, and is now seen to be the weight of the diagonal corresponding to b_0, normalized by the sum of weights of all diagonals, per(P).

Definition 6 (Profiles): Let $\langle d_1, d_2, \ldots , d_{t!}\rangle$ be an arbitrary sequence of all diagonals of P. Then, a diagonal weight profile (or just profile) of P is the normalized sequence of weights of diagonals in the above sequence

$$\frac{1}{\text{per}(P)}\langle \mathcal{W}(d_1), \mathcal{W}(d_2), \ldots , \mathcal{W}(d_{t!})\rangle.$$

As all doubly stochastic matrices have nonzero permanents, the above sequence is well defined. A profile of P is the probability distribution arrived at by the attacker on the diagonals of P, i.e., matchings, of being the genuine matching. From the point of view of a system-wide anonymity metric, this is the most vital piece of information contained in P.

Strictly speaking, P may have multiple profiles, but all its profiles are simply permutations of each other and, as we will see shortly, lead to the same measure of anonymity. We therefore ignore the order of values in profiles and consider P to have a unique profile.

For the example matrix P of Fig. 2(b), $t = 4$ and there are $t! = 24$ matchings. Exactly 12 of these matchings can be seen to have weight 1/108 each, and the remaining 12 have zero weight each. The sum of weights of all matchings is 12/108 = 1/9 = per(P). Thus, the profile of P is a sequence with exactly 12 occurrences of the value (1/108) / (1/9) = 1/12, and 12 occurrences of the value 0. The 24 values in this sequence are the attacker's probabilities for each of the matchings of being genuine.

We can now state the anonymity metric of Bagai *et al.* [3] for a system's degree of anonymity after an attack that results in a probability matrix P. Ever since the works of Serjantov and Danezis [1] and Diaz *et al.* [2], Shannon entropy [8] of a probability distribution is a well-accepted measure of anonymity. Bagai *et al.* [3] employ the same technique over the profile of the matrix as a measure of the attacker's uncertainty of which matching is indeed genuine.

Definition 7 (Basic Metric): Let P be a $t \times t$ probability matrix resulting from an attack, with profile $\langle w_1, w_2, \ldots , w_{t!}\rangle$. The system's degree of anonymity after this attack is

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

In the above summation, a subexpression $0 \cdot \log(0)$ is interpreted as 0. This metric measures the extent to which the genuine matching is still hidden, after the attack, among all fake ones. It is easily verified that $\Delta(P) = 1$ iff $P = P_{\max}$, as in Fig. 1(b), and $\Delta(P) = 0$ iff P contains a single occurrence of the value 1 in each of its rows and columns. For the matrix P of Fig. 2(b), $\Delta(P) = \log(12)/\log(24) \approx 0.78$.

III. Sending/Receiving Multiple Messages

The basic metric of Definition 7 measures a system's effectiveness in thwarting an attacker's attempt at determining the unique genuine matching. Stated alternatively, it measures the extent to which the attacker falls short of achieving the goal of pinpointing the genuine matching. Gierlichs *et al.* [5] pointed out that the attacker, in fact, usually has a more modest goal, especially in the commonly occurring scenarios where system users send and/or receive multiple messages. In such scenarios, the attacker is content with figuring out just how many messages each sender sent to each receiver, and not necessarily which messages.

As a simple example, suppose an attacker knows that messages x_1 and x_2 of Fig. 2 were both sent by the same sender, say Alice, and that message y_1 was the only message received by some receiver, say Bob. Consider now the anonymity of the Alice-Bob pair. Although the attacker is still unsure of whether y_1 was the message x_1 or x_2 (each event has a 50% probability), he has determined with 100% certainty that Alice sent one message to Bob. From the point of view of the anonymity of relationships between senders and receivers, the attacker does not care to determine exactly which message sent by Alice went to Bob, first or second, but is content to determine the number of such messages.

To measure the anonymity of relationships between a system's senders and receivers, who may be sending/receiving multiple messages, a modification of the metric of Definition 7 is therefore needed, as this basic metric considers relationships only between the system's input and output messages. It is worth noting that, while an attack, such as the example in Section II-B, still arrives at probabilities of message relationships, leading to a $t \times t$ probability matrix P, the modified metric needs to measure sender–receiver relationship anonymity.

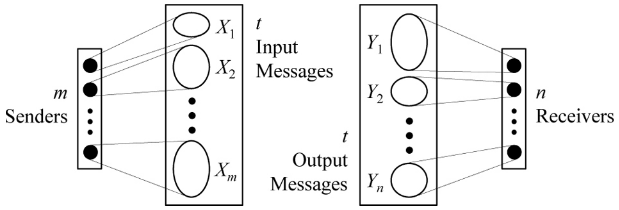
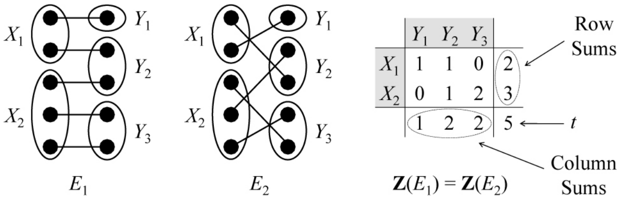


Fig. 3. Users sending and receiving multiple messages.



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

Fig. 4. Example of two equivalent matchings and their common association matrix.

A. Sender–Receiver Associations

As shown in Fig. 3, let m be the number of senders in the system 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 of messages received by receiver j. It is easy to see that, $\sum_{i=1}^{m} |X_i| = t = \sum_{j=1}^{n} |Y_j|$, and that

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

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.

Definition 8 (Association Matrices): For any subset $E \subseteq X \times Y$ of edges in $K_{t,t}$, the (sender–receiver) association matrix of E, denoted $\mathbf{Z}(E)$, is the $m \times n$ matrix of nonnegative integers given by

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

In other words, any entry $\mathbf{Z}(E)_{ij}$ of this matrix is 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 $|E|$.

There are a total of $2^{(t^2)}$ subsets of $X \times Y$. We already know that exactly $t!$ of these subsets are matchings between X and Y. If E is any of the $t!$ matchings, 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, that is

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

Recall from Definition 1 that $\mathfrak{M}$ is the set of all $t!$ matchings between X and Y. We now define a binary relation on $\mathfrak{M}$.

Definition 9 (Equivalent Matchings, $\bowtie$): Let $E_1, E_2 \in \mathfrak{M}$. Then E_1 and E_2 are equivalent, denoted by $E_1 \bowtie E_2$, if they have the same association matrix, i.e., $\mathbf{Z}(E_1) = \mathbf{Z}(E_2)$.

Clearly, $\bowtie$ is an equivalence relation and, for any matching $E \in \mathfrak{M}$, the equivalence class of E is the set

$$\{E' \in \mathfrak{M} \mid E \bowtie E'\}$$

of all matchings equivalent to E. In other words, matchings are considered equivalent if they have the same number of messages going from any given sender to any given receiver. The exact input–output message relationship in equivalent matchings may, however, be different. Fig. 4 shows an example of two matchings in a system with $m = 2$ senders, $n = 3$ receivers, and $t = 5$ messages. These matchings are equivalent because they have the same association matrix.

The five messages in this system 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 $|Y_3| = 2$. These values are also shown in the figure as the row- and column-sums of the association matrix of these matchings. It is instructive at this point to consider the following two questions.

1) How many equivalence classes does $\bowtie$ partition the set $\mathfrak{M}$ into?
2) Given a particular equivalence class, how many matchings does that class contain?

The rest of Section III addresses the above questions. An understanding of these issues assists in understanding our modified metric, given in Sections IV and V.

B. Number of Equivalence Classes

Let S and R denote, respectively, the sender and receiver multiplicity vectors of the system, that is

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

Associated with every equivalence class of $\bowtie$ over the set $\mathfrak{M}$ is then a unique $m \times n$ association matrix of nonnegative integers, with S as its row-sums vector and R as its column-sums vector. The number of equivalence classes is thus the number of such matrices.

Let us first consider the complexity of determining this number. Valiant [12] introduced the complexity class #P, of counting problems associated with the decision problems in the class NP. For example, the subset-sum decision problem in NP is to determine if there exists a subset of a given list of integers, such that members of that subset add up to some given value. The corresponding counting problem in #P is to count the exact number of such subsets. Clearly, any problem in #P is at least as hard as its counterpart in NP because its decision variant only seeks to know whether or not the count is greater than zero. Valiant [12] then showed that computing the permanent of any matrix, even if all its entries are just 0 or 1, is #P-complete. Consequently, as in Jerrum *et al.* [13], attention has been given to efficient approximation of permanents to an acceptable degree of accuracy.

Dyer *et al.* [14] established the well-known result that exactly counting the number of nonnegative integer matrices with given row- and column-sums is #P-complete. Although several exact counting algorithms are known, due to this result, one does not expect to find any polynomial-time ones, as that would imply P = NP. Gail and Mantel [15] gave a straightforward technique based on a recurrence. Greselin [16] has another recursive method for counting and generating such matrices. Macdonald's [17] method employed complete symmetric functions to obtain this count. Grégoire and Hamel [6] essentially restate that method in our context of anonymity systems.

Given the complexity of obtaining the exact count, much work has also been done on getting approximate answers efficiently and on asymptotic expressions. Some recent techniques are in [18]–[20].

Informative surveys on this topic are contained in [21], and more recently in [22].

C. Cardinality of an Equivalence Class

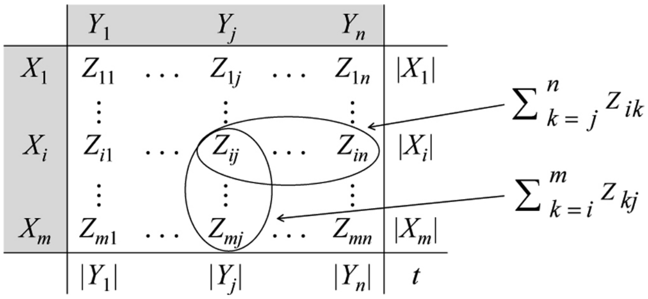
We now turn our attention to counting the number of matchings contained in the equivalence class identified by a given $m \times n$ association matrix Z. In other words, the number of matchings E, for which $\mathbf{Z}(E) = Z$. As before, we let $S = \langle |X_1|, |X_2|, \ldots, |X_m| \rangle$ and $R = \langle |Y_1|, |Y_2|, \ldots, |Y_n| \rangle$ be the sender and receiver multiplicity vectors. We start with a special case, and proceed to the most general one.

If all sender and receiver multiplicities are 1, i.e., $m = t = n$ and $|X_i| = 1 = |Y_j|$, for all i, j, then Z can only be a 0–1 matrix, with exactly one 1 in each of its rows and columns. There are $t!$ such matrices, thus equivalence classes, which is also the total number of matchings. Any class in this case, including the one identified by Z, therefore contains exactly one matching.

Now suppose $m < t = n$, i.e., some senders sent multiple messages but each receiver received only one message. As all column-sums of Z are 1, in this case also Z can only be a 0–1 matrix, with exactly one 1 in each of its columns. Some rows of Z, however, have multiple occurrences of 1. Consider, arbitrarily, the ith row, with $k = |X_i|$ occurrences of 1, in columns $j_1, j_2, \ldots, j_k$. The k messages of this sender could have been sent in any of $k!$ ways to these receivers, while still maintaining the association matrix Z. The messages sent by any other sender can also be shuffled among their receivers in the same way, while still maintaining the matrix Z. The number of matchings whose association matrix is Z is thus $\prod_{i=1}^{m} |X_i|!$. Note that this value is independent of Z, i.e., all equivalence classes have the same size, which can be determined just from S.

The case $m = t > n$ is similar, i.e., when each sender sent exactly one message but some receivers received multiple messages. By a symmetric reasoning, we have that the cardinality of each class in this case is $\prod_{j=1}^{n} |Y_j|!$.

In the most general case, $m < t > n$. Without loss of generality, we start at the top-left corner of Z and proceed to its bottom-right corner, but any other order of considering all its entries leads to the same result. Consider Z_{11}, the number of messages from the first sender to the first receiver. Of all the $|X_1|$ messages sent by the first sender, these Z_{11} messages can be chosen in $\binom{|X_1|}{Z_{11}}$ ways, i.e., $\binom{\sum_{k=1}^{n} Z_{1k}}{Z_{11}}$ ways. Similarly, of all the $|Y_1|$ messages received by the first receiver, these Z_{11} messages can be chosen in $\binom{|Y_1|}{Z_{11}}$ ways, i.e., $\binom{\sum_{k=1}^{m} Z_{k1}}{Z_{11}}$ ways. Once the exact Z_{11} source and destination messages between these two users are identified, there are $Z_{11}!$ matchings between them. Thus



$$\begin{array}{c|ccccc|c} & Y_1 & & Y_j & & Y_n & \\ \hline X_1 & Z_{11} & \ldots & Z_{1j} & \ldots & Z_{1n} & |X_1| \\ & \vdots & & \vdots & & \vdots & \\ X_i & Z_{i1} & \ldots & Z_{ij} & \ldots & Z_{in} & |X_i| \\ & \vdots & & \vdots & & \vdots & \\ X_m & Z_{m1} & \ldots & Z_{mj} & \ldots & Z_{mn} & |X_m| \\ \hline & |Y_1| & & |Y_j| & & |Y_n| & t \end{array}$$

Fig. 5. Typical association matrix Z.

$$\binom{\sum_{k=1}^{n} Z_{1k}}{Z_{11}} \binom{\sum_{k=1}^{m} Z_{k1}}{Z_{11}} Z_{11}!$$

is the total number of ways of sending Z_{11} messages from the first sender to the first receiver. Now consider Z_{12}. On the sender side, we now only have $|X_1| - Z_{11}$ messages to choose from, which is $\sum_{k=2}^{n} Z_{1k}$. This choice can be performed in $\binom{\sum_{k=2}^{n} Z_{1k}}{Z_{12}}$ ways, leading to

$$\binom{\sum_{k=2}^{n} Z_{1k}}{Z_{12}} \binom{\sum_{k=1}^{m} Z_{k2}}{Z_{12}} Z_{12}!$$

as the total number of ways of sending Z_{12} messages from the first sender to the second receiver. Proceeding in this fashion to the bottom-right of the association matrix Z, we can get an expression of the number of ways in which Z_{ij} messages may be sent by any sender i to receiver j. Fig. 5 displays the remaining choices for Z_{ij}, given that choices to its left and top have already been made.

The total number of ways all messages may be forwarded by the system, while adhering to the associations given by the matrix Z, is thus

$$\prod_{i=1}^{m} \prod_{j=1}^{n} \binom{\sum_{k=j}^{n} Z_{ik}}{Z_{ij}} \binom{\sum_{k=i}^{m} Z_{kj}}{Z_{ij}} Z_{ij}!$$

and that is clearly the cardinality of the equivalence class identified by Z, of the equivalence relation $\bowtie$ over the set $\mathfrak{M}$.

The cardinality of the equivalence class corresponding to the example association matrix of Fig. 4 can be computed by the above expression to be 24. Two of these 24 equivalent matchings are displayed in that figure.

IV. New Metric

In this section we develop a metric for measuring the anonymity of the system's sender–receiver communication

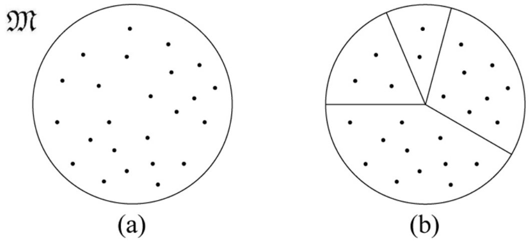


Fig. 6. (a) Set $\mathfrak{M}$ of all $t!$ matchings. (b) Equivalence classes on $\mathfrak{M}$ induced by the multiplicity vectors S and R.

pattern, in the aftermath of a two-pronged attack that has performed the following.

1) Arrived at probabilities of all 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 the example of Section II-B. It thus results in a $t \times t$ probability matrix P.
2) Associated the correct sender with each input message of the system, and receiver with each output message. This results in the sender and receiver multiplicity vectors S and R.

The above segments of the attack are orthogonal to each other and, independently, each has an adverse effect on the system's anonymity level. The metric we now develop measures their combined effect.

We already understand the effects of each of these attack segments separately. Recall that $\mathfrak{M}$ is the set of all $t!$ matchings, as shown in Fig. 6(a). The sender and receiver multiplicity vectors S and R induce the equivalence relation $\bowtie$ on $\mathfrak{M}$, whose classes are as in Fig. 6(b). Each class represents a sender–receiver association scenario, and is identified by a unique association matrix with S as its row-sums vector and R as its column-sums vector. In Section III-C we gave an expression for the size of a class in terms of its association matrix. In the absence of the other attack component, namely the one based on the probability matrix, these class sizes lead directly to a probability distribution over all possible association scenarios. The probability of any particular scenario, in this case, is simply proportional to the number of matchings in its corresponding equivalence class. From this distribution, anonymity can be measured by standard techniques, such as by Shannon entropy, as mentioned a little later.

The probability matrix P resulting from the other part of the attack, however, alters this probability distribution. By inducing a probability on each matching, as explained in Section II-C, P in fact ends up assigning a probability value to each equivalence class of $\bowtie$. The value assigned to any class is the probability of the genuine matching being contained in that class, which is now the sum of probabilities of all matchings in that class. From an anonymity point of view, what is important now is thus not the raw size of each class, but the sum of the probabilities of all matchings in each class.

Definition 10 (Weights of Association Matrices): For any given multiplicity vectors S and R, let $\mathcal{Z}_{S,R}(\mathfrak{M})$ denote the set of all association matrices of matchings in $\mathfrak{M}$, 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. Now, for any given $t \times t$ probability matrix P, the weight assigned by P to any matrix $Z \in \mathcal{Z}_{S,R}(\mathfrak{M})$, denoted $\mathcal{W}_P(Z)$, is the sum of weights of all matchings in the equivalence class of $\bowtie$ associated with Z.

We already know that

$$\sum_{Z\in\mathcal{Z}_{S,R}(\mathfrak{M})} \mathcal{W}_P(Z) = \sum_{d\in\mathcal{D}(P)} \mathcal{W}(d) = \operatorname{per}(P).$$

In Section V, we give a method to compute $\mathcal{W}_P(Z)$. For now, let $\omega_P(Z) = \mathcal{W}_P(Z)/\operatorname{per}(P)$ be the normalized weight of Z. Clearly, the values $\omega_P(Z)$ add up to 1, over all Z, and we have a probability distribution on the set $\mathcal{Z}_{S,R}(\mathfrak{M})$ of all sender–receiver association scenarios. For any Z, the value $\omega_P(Z)$ is the likelihood assigned by the attacker to the scenario represented by Z.

As done in [3] for the probability distribution over all matchings, we employ Shannon entropy [8] of the probability distribution given by $\omega_P(Z)$, for all Z $\in \mathcal{Z}_{S,R}(\mathfrak{M})$, as a measure of the attacker's uncertainty of which of the system's sender–receiver association scenarios is the actual one.

Definition 11 (New Metric): Let a $t \times t$ probability matrix P, 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

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

It is easily seen that $\delta_{S,R}(P)$ is always between 0 and 1. We first establish that, if all multiplicities are 1, the above metric coincides with the basic metric of Bagai *et al.* [3], given by Definition 7, and that higher multiplicity values reduce anonymity. The following proposition is verified easily from the definitions.

Proposition 2: For all S, R, and P, $\delta_{S,R}(P) \leq \Delta(P)$, with equality iff all multiplicity values in S and R are 1.

It remains to compute the weight, $\mathcal{W}_P(Z)$, of an association matrix Z, i.e., the sum of weights of all matchings in the class associated with Z. The next section gives a method for this.

V. Weight of an Equivalence Class

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

The row- and column-sums vectors of Z are, respectively, the system's sender and receiver multiplicity vectors, $S = \langle|X_1|, |X_2|, \ldots, |X_m|\rangle$ and $R = \langle|Y_1|, |Y_2|, \ldots, |Y_n|\rangle$. Without loss of generality, we assume that the first $|X_1|$ rows of P 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 P correspond to messages received by the first receiver, etc. If necessary, the rows and/or columns of P 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 P into regions.

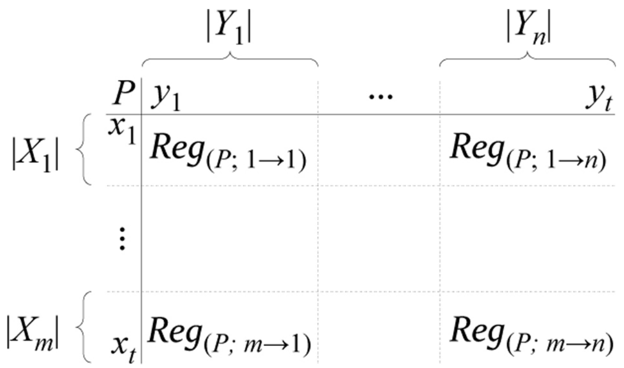


Fig. 7. Regions of P, $Reg_{(P;\ i\to j)}$, for $1 \le i \le m$, and $1 \le j \le n$.

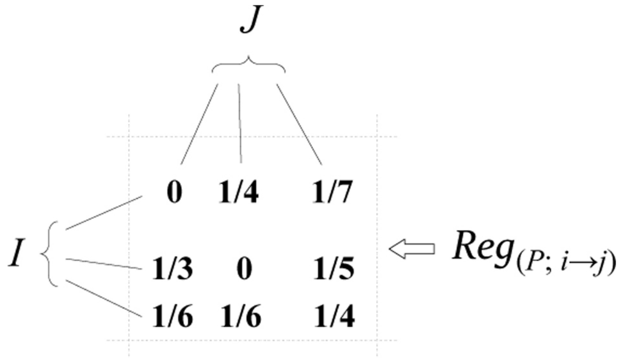


Fig. 8. Example extract $P[\![I;J]\!]$ of region $Reg_{(P;\ i\to j)}$, if $Z_{ij} = 3$.

Definition 12 (Regions): For any sender i and receiver j, $1 \le i \le m$, and $1 \le j \le n$, a region of P, denoted $Reg_{(P;\ i\to j)}$, is the submatrix of P 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 $1 + \sum_{k=1}^{j-1} |Y_k|$.

There are mn regions of P, all pairwise mutually disjoint, as shown in Fig. 7. Any particular region, $Reg_{(P;\ i\to j)}$ contains probability information arrived at by the attacker of the subset $X_i \times Y_j$ of edges in the system's complete graph $K_{t,t}$.

Regions are contiguous submatrices of P and are not necessarily square. They are induced solely by the multiplicity vectors S and R. Given, additionally, an association matrix Z, our method will also employ square submatrices of P that are not necessarily contiguous. These square submatrices are extracts contained within its regions and are of sizes stated by corresponding values in Z.

Definition 13 (Extracts): Let I be any subset of size Z_{ij} of the set of all row indices of some region $Reg_{(P;\ i\to j)}$. Also, let J be any subset of the same size of the set of all column indices of that region. An extract of that region, denoted $P[\![I;J]\!]$, is the $Z_{ij} \times Z_{ij}$ submatrix obtained by extracting all elements of P at row numbers in I and column numbers in J. We also let $\mathcal{E}_{(P;\ i\to j)}$ denote the set of all extracts of the region $Reg_{(P;\ i\to j)}$.

Fig. 8 shows an example extract $P[\![I;J]\!]$ of some region $Reg_{(P;\ i\to j)}$, assuming $Z_{ij} = 3$. This extract represents the situation of the Z_{ij} input messages given by indices in I corresponding, in any order, to the Z_{ij} output messages given by indices in J. Each of the $Z_{ij}!$ diagonals of $P[\![I;J]\!]$, i.e., matchings between I and J, is essentially some partial matching between X and Y. Observe that

$$\mathrm{per}(P[\![I;J]\!])$$

is the sum of weights of all these $Z_{ij}!$ partial matchings between X and Y. In general, $P[\![I;J]\!]$ is not doubly stochastic, but all its row- and column-sums are always nonnegative and at most 1. Thus, $0 \le \mathrm{per}(P[\![I;J]\!]) \le 1$.

As Z_{ij} rows and columns may be chosen from $Reg_{(P;\ i\to j)}$ in $\binom{|X_i|}{Z_{ij}}\binom{|Y_j|}{Z_{ij}}$ ways, that is also the number of extracts of this region. The following proposition now becomes straightforward:

Proposition 3: The sum of weights given by P to partial matchings of Z_{ij} messages from sender i to receiver j is

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

To determine $\mathcal{W}_P(Z)$, i.e., the sum of weights given by P to all matchings in the equivalence class associated with Z, due to distributive law, 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 an extract $P[\![I;J]\!] \in \mathcal{E}_{(P;i\to j)}$ 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 probability matrix P for all subsequent weight evaluation.

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

$$\widehat{P}[\![I;J]\!]_{uv} = \begin{cases} 0, & \text{if } u \in I \text{ or } v \in J \\ P_{uv}, & \text{otherwise.} \end{cases}$$

The permanent of any extract from $\widehat{P}[\![I;J]\!]$ that has at least one row from I and/or at least one column from J will clearly be 0, thus avoiding duplicate weight evaluation 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, that is

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

The sum of weights, $\mathcal{W}_P(Z)$, of all matchings in the class associated with Z can now be expressed as a recursive formula, which effectively multiplies the sum of weights of all partial matchings corresponding to each sender–receiver combination. As the base case, if all values in the Z matrix are 0, we let $\mathcal{W}_P(Z) = 1$, the identity of multiplication. In general, if for some i and j, $Z_{ij} \neq 0$, then $\mathcal{W}_P(Z)$ is the following value:

$$\sum_{P[\![I;J]\!] \in \mathcal{E}_{(P;i\to j)}} \left[\mathrm{per}(P[\![I;J]\!]) \cdot \mathcal{W}_{\widehat{P}[\![I;J]\!]}(\widehat{Z}[\![i;j]\!])\right].$$

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 extract $P[\![I;J]\!] \in \mathcal{E}_{(P;i\to j)}$, the product of the following is obtained.

1) $\mathrm{per}(P[\![I;J]\!])$, i.e., the sum of weights of all partial matchings within $P[\![I;J]\!]$.

2) $\mathcal{W}_{\widehat{P}[\![I;J]\!]}(\widehat{Z}[\![i;j]\!])$, i.e., the sum of weights of all partial matchings that can be extensions of partial matchings within $P[\![I;J]\!]$.

This value is added over all such extracts for the chosen Z_{ij} giving, due to distributive law, the desired result.

While the above recursive formula gives the exact value of $\mathcal{W}_P(Z)$, it is based on permanent values of extracts of regions of P. Given the inherent complexity mentioned in Section III-B of computing the permanent of a matrix, this formula shows that exactly computing the weight of an equivalence class cannot, in general, be performed efficiently. However, as successfully accomplished by Lakshmanan *et al.* [23] for a different problem involving matrix permanents, it may be possible to arrive at efficient heuristic methods that compute approximate weights of equivalence classes within specified degrees of accuracy, or even be possible to identify some special classes of P and/or Z, for which exact values can be computed efficiently. We consider development of such techniques beyond the scope of this paper, and leave them for future work.

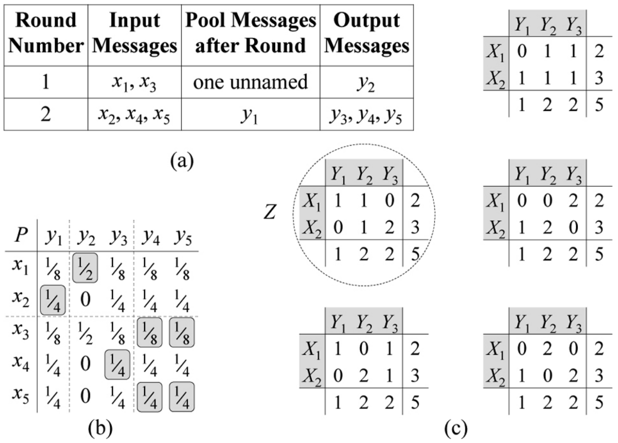
VI. Application to Pool Mixes

Pool mixes were originally introduced by Cottrell [24] as a high-latency strategy to counter attacks aimed at correlating a system's input and output messages. Several generalizations of this strategy have been analyzed by Diaz and Serjantov [25], all of which operate in iterative rounds. In each round, any pool mix first collects a certain number of incoming messages and places them in its internal message pool, then flushes out a randomly selected fraction of all messages contained in this pool. Messages left in its pool after any round are candidates for flushing in future rounds.

In this section we employ our method to determine the anonymity provided by an example pool mix after its first two rounds of operation. As shown in Fig. 9(a), suppose input messages x_1 and x_3 enter the mix in Round 1, and only one of them, called y_2, is output by the mix in that round. The other input message is retained in its internal pool. We do not need to assign an explicit name to this message in the pool after Round 1, because we are interested in measuring anonymity across two completed rounds. Also, suppose input messages x_2, x_4, and x_5 arrive in Round 2, and messages y_3, y_4, and y_5 exit in that round. We let y_1 denote the message left in the pool after Round 2. Fig. 9(b) shows the 5 × 5 probability matrix P resulting from these two rounds of operation. As an example, as there is a 50% chance of message x_3 being the same as y_2, we have $P_{32} = 1/2$. All other entries of P are arrived at similarly.

Now, suppose the attacker observed that messages x_1 and x_2 were both sent by the same user, and x_3, x_4, and x_5 by another user, i.e., $X_1 = \{x_1, x_2\}$, and $X_2 = \{x_3, x_4, x_5\}$. Also, treating the mix pool after Round 2 as a receiver, suppose the attacker observed that $Y_1 = \{y_1\}$, $Y_2 = \{y_2, y_3\}$, and $Y_3 = \{y_4, y_5\}$. This observation of the attacker results in the following values:

$$m = 2, \qquad S = \langle 2, 3\rangle,$$
$$n = 3, \qquad R = \langle 1, 2, 2\rangle.$$



Round Number	Input Messages	Pool Messages after Round	Output Messages
1	x_1, x_3	one unnamed	y_2
2	x_2, x_4, x_5	y_1	y_3, y_4, y_5

(a)

$$\begin{array}{c|c|cc|cc} P & y_1 & y_2 & y_3 & y_4 & y_5 \\ \hline x_1 & 1/8 & \boxed{1/2} & 1/8 & 1/8 & 1/8 \\ x_2 & \boxed{1/4} & 0 & 1/4 & 1/4 & 1/4 \\ \hline x_3 & 1/8 & 1/2 & 1/8 & \boxed{1/8} & \boxed{1/8} \\ x_4 & 1/4 & 0 & \boxed{1/4} & 1/4 & 1/4 \\ x_5 & 1/4 & 0 & 1/4 & \boxed{1/4} & \boxed{1/4} \end{array}$$

(b)

$$\begin{array}{c|ccc|c} & Y_1 & Y_2 & Y_3 & \\ \hline X_1 & 0 & 1 & 1 & 2 \\ X_2 & 1 & 1 & 1 & 3 \\ \hline & 1 & 2 & 2 & 5 \end{array}$$

$$Z \quad \begin{array}{c|ccc|c} & Y_1 & Y_2 & Y_3 & \\ \hline X_1 & 1 & 1 & 0 & 2 \\ X_2 & 0 & 1 & 2 & 3 \\ \hline & 1 & 2 & 2 & 5 \end{array} \qquad \begin{array}{c|ccc|c} & Y_1 & Y_2 & Y_3 & \\ \hline X_1 & 0 & 0 & 2 & 2 \\ X_2 & 1 & 2 & 0 & 3 \\ \hline & 1 & 2 & 2 & 5 \end{array}$$

$$\begin{array}{c|ccc|c} & Y_1 & Y_2 & Y_3 & \\ \hline X_1 & 1 & 0 & 1 & 2 \\ X_2 & 0 & 2 & 1 & 3 \\ \hline & 1 & 2 & 2 & 5 \end{array} \qquad \begin{array}{c|ccc|c} & Y_1 & Y_2 & Y_3 & \\ \hline X_1 & 0 & 2 & 0 & 2 \\ X_2 & 1 & 0 & 2 & 3 \\ \hline & 1 & 2 & 2 & 5 \end{array}$$

(c)

Fig. 9. (a) Message history of the first two rounds of an example pool mix. (b) Probability matrix after these rounds, its regions induced by multiplicity vectors, and an example extract collection for the circled association matrix Z. (c) Association matrices of all equivalence classes.

These values induce six regions on P, as shown in Fig. 9(b), and partition the set of all 5! = 120 matchings of this system into five equivalence classes, whose association matrices are shown in Fig. 9(c). 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 Fig. 9(c) is identical to the one already encountered in Fig. 4. We first step through the computation of $\mathcal{W}_P(Z)$, by our method of Section V.

There are four nonzero entries in Z, namely $Z_{11} = Z_{12} = Z_{21} = 1$, and $Z_{23} = 2$. Thus any matching in the equivalence class associated with Z is made up of four extracts of regions of P with pairwise disjoint row-sets and column-sets: a 1 × 1 extract from each of the regions $Reg_{(P;\ 1\to 1)}$, $Reg_{(P;\ 1\to 2)}$, and $Reg_{(P;\ 2\to 1)}$, and a 2 × 2 extract from the region $Reg_{(P;\ 2\to 3)}$. An example collection of such extracts of the regions induced on P is shown in Fig. 9(b). The product of permanents of extracts in this collection is

$$\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{4}\cdot\left(\frac{1}{8}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{4}\right)=\frac{1}{512}.$$

Our recursive method given in Section V of 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 $\mathcal{W}_P(Z) = 5/512$.

The weights $\mathcal{W}_P$ of the other four association matrices shown in Fig. 9(c) can be similarly evaluated by our method to be $1/128, 5/256, 1/256$, and $3/512$. The sum of weights of all five association matrices is $\mathrm{per}(P) = 3/64$, and dividing their individual weights by this value leads to the normalized weights ω_P of these matrices: $5/24$, $1/6$, $5/12$, $1/12$, and $1/8$. From our metric of Definition 11, we get $\delta_{S,R}(P) \approx 0.3$.

To appreciate the reduction in anonymity caused by message multiplicities, we also compute the anonymity level by the basic metric of Bagai *et al.* [3]. It is easily seen that, of the 120 diagonals of P, 48 diagonals have weight 1/1024 each, and the remaining 72 have 0 weight. The sum of weights of all its diagonals is $48/1024 = 3/64 = \mathrm{per}(P)$. Thus, the profile of P is a sequence with exactly 48 occurrences of the value $(1/1024)/(3/64) = 1/48$, and 72 occurrences of the value

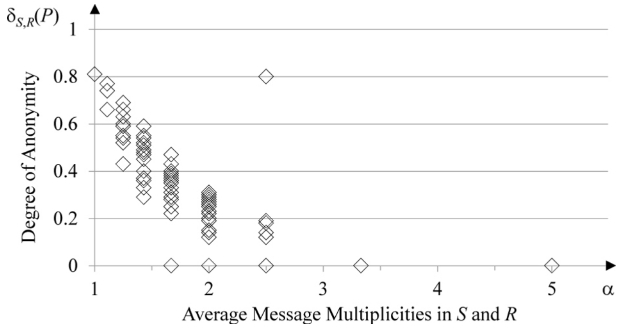


Fig. 10. Degree of anonymity of example probability matrix P for all possible multiplicity vectors S and R.

0. According to Definition 7, $\Delta(P) = \log(48)/\log(120) \approx 0.81$, which is significantly higher than $\delta_{S,R}(P) \approx 0.3$, the anonymity level computed previously by our method.

In general, the higher the message multiplicity values in vectors S and R, the lower the anonymity.

Fig. 10 depicts this phenomenon over all possible message multiplicity vectors for this probability matrix P. 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.

VII. Comparison With Existing Approaches

There have been three other attempts at measuring the system-wide anonymity when users send and/or receive multiple messages. Gierlichs *et al.* [5] were the first to observe a need for arriving at a metric for such a multiple message scenario. This problem was then revisited, first by Grégoire and Hamel [6] and, more recently, by Bagai *et al.* [7]. We now compare our work with each of these.

A. Enhanced Scope

The main difference between our result and these previous attempts lies in the scope of each work. While the metric developed in this paper measures anonymity after any attack in the class **A-Prob**, as defined in Section I, all existing metrics have only addressed the attack class **A-Inf**. Bagai *et al.* [3] showed that the class **A-Inf** is much smaller and properly contained in **A-Prob**.

An estimate of the sizes of these attack classes is obtained easily. An attack from **A-Inf** simply rules out edges in the system's complete bipartite graph as infeasible, thus resulting in a $t \times t$, 0–1 matrix. There are $2^{(t^2)}$ such matrices, i.e., different attacks, for a system with t input and output messages. As attacks in **A-Prob** result in $t \times t$ doubly stochastic matrices, there are an uncountably infinite number of them.

To see that **A-Inf** is a subclass of **A-Prob**, we employ an elegant infinite procedure of Sinkhorn and Knopp [26]. Let f, g, and h be functions from and to $t \times t$ real matrices, defined as follows:

$$\begin{aligned} f(M)_{ij} &= M_{ij} \;/\; \textstyle\sum_{k=1}^{t} M_{ik} \quad \text{(row normalization)} \\ g(M)_{ij} &= M_{ij} \;/\; \textstyle\sum_{k=1}^{t} M_{kj} \quad \text{(column normalization)} \\ h(M) &= g(f(M)). \end{aligned}$$

Sinkhorn and Knopp [26] showed that, for any square matrix A with nonnegative values, the matrix $\lim_{k\to\infty} h^k(A)$ is doubly stochastic and has the same diagonal weight profile as A. Thus, a procedure that alternately normalizes all rows followed by all columns of any $t \times t$, 0–1 matrix A resulting from an attack in **A-Inf**, ad infinitum, converges to a matrix, which:

1) by being doubly stochastic, is a matrix resulting from some attack in **A-Prob**;
2) by having the same profile as A, i.e., by assigning the same probability value to each matching as A does, has the same information content as A.

The infeasibility attack resulting in A and the probabilistic attack resulting in $\lim_{k\to\infty} h^k(A)$ are thus equivalent. Fig. 11 shows a 3×3 example of a 0–1 matrix from an attack in **A-Inf**, and the doubly stochastic matrix from its equivalent attack in **A-Prob**. Reference [3] contains an alternative characterization, based on matrix scalings, of the unique equivalent attack in **A-Prob** for any given attack in **A-Inf**.

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

$\lim_{k\to\infty} h^k(A)$	y_1	y_2	y_3
x_1	0	$(\sqrt{5}-1)/2$	$(3-\sqrt{5})/2$
x_2	$(\sqrt{5}-1)/2$	0	$(3-\sqrt{5})/2$
x_3	$(3-\sqrt{5})/2$	$(3-\sqrt{5})/2$	$\sqrt{5}-2$

Fig. 11. Example 0–1 matrix resulting from an attack in **A-Inf**, and the doubly stochastic matrix resulting from its equivalent attack in **A-Prob**.

B. Metric of Grégoire and Hamel [6]

The scope of the work of Grégoire and Hamel [6] is, in fact, the smallest of the existing works. While they discussed the attack class **A-Inf**, they proposed the formula $\log(\text{COUNT})/\log(t!)$ as an anonymity 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 any attack from **A-Inf** into account or, alternatively, assumes that the attack's 0–1 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 [17] for counting nonnegative integer matrices with prescribed row and column sums. It is also well summarized in the survey of Diaconis and Gangolli [21].

Moreover, as shown in [7], the value of COUNT is only of academic interest and not needed to compute the system's degree of anonymity after an attack in **A-Inf**.

C. Metric of Gierlichs et al. [5]

The most significant contribution of Gierlichs *et al.* [5] was their demonstration of the need for a metric for the multiple

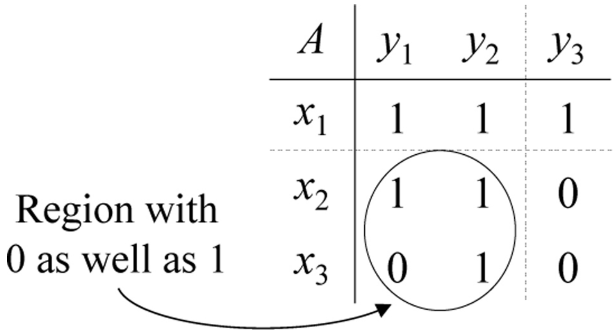


Fig. 12. Example 0–1 matrix resulting from an attack in **A-Inf**, for which the method of Gierlichs *et al.* [5] fails.

message scenario. For this scenario, they even proposed a method for measuring the system's anonymity after any attack in the class **A-Inf**. However, their method was shown by Bagai *et al.* [7] to be insufficient, as it works for only those attacks in **A-Inf** that result in a 0–1 matrix, each of whose regions contains either all 0s or all 1s. Fig. 12 contains an example 0–1 matrix for which the method of Gierlichs *et al.* [5] fails to work correctly.

This is due to the highlighted region that contains at least one occurrence of both 0 as well as 1. In this example, $t = 3$, $m = n = 2$, $S = \langle 1, 2\rangle$, and $R = \langle 2, 1\rangle$.

In all, there are $2^{(t^2)}$, $t \times t$, 0–1 matrices, of which $2^{(mn)}$ are such that each of their regions contains either all 0s or all 1s. The method of Gierlichs *et al.* [5] works for these $2^{(mn)}$ matrices. However, as sender and receiver multiplicity values grow (i.e., values in S and R), m and n become smaller (i.e., lengths of S and R), thereby shrinking the scope of their method as message multiplicity values grow.

D. Metric of Bagai et al. [7]

To date, the work of Bagai *et al.* [7] has been the most comprehensive, as their method is capable of measuring the effectiveness of any attack in the class **A-Inf**. However, this class is finite (with $2^{(t^2)}$ matrices) and, as shown earlier, is a subclass of the uncountably infinite class **A-Prob**, for which our method is designed.

Bagai *et al.* [7] showed that an attack in **A-Inf** has the effect of labeling certain matchings as infeasible, namely those that contain at least one edge (of the system's complete bipartite graph) that was determined by the attack to be infeasible. Some matchings contained in a given equivalence class of $\bowtie$ over $\mathfrak{M}$ are thus infeasible, while others are feasible, as depicted in Fig. 13.

They gave a method for counting the number of feasible matchings in a given equivalence class of $\bowtie$, which formed the basis of their anonymity metric.

Interestingly, the odds of the genuine matching belonging to any equivalence class are proportional to the number of feasible matchings in that class. Thus, our method of determining effectiveness of attacks in **A-Prob**, coincides with the method of [7] for attacks in **A-Inf**, for those $2^{(t^2)}$ special $t \times t$ doubly stochastic matrices that are equivalent to 0–1 matrices, as described earlier in Section VII-A. Stated precisely, let A be any $t \times t$, 0–1 matrix. Then, the effectiveness of A according to Bagai *et al.* [7] is the same as the effectiveness of $\lim_{k\to\infty} h^k(A)$ determined by our method. Our method is thus an accurate generalization of theirs, from **A-Inf** to **A-Prob**.

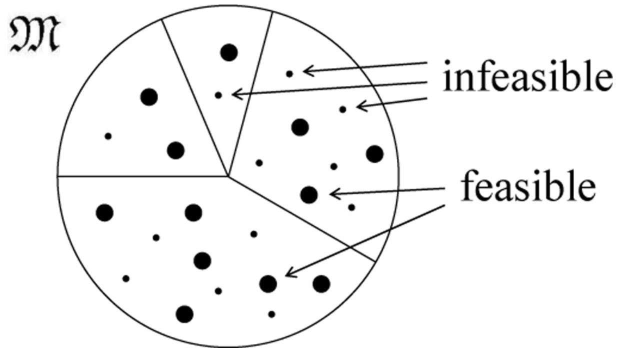


Fig. 13. Feasible and infeasible labeling of matchings in equivalence classes.

VIII. Conclusion and Future Work

Bagai *et al.* [3] presented a metric for measuring the system-wide anonymity provided by an anonymity system in the aftermath of a probabilistic attack. Their metric measured the extent to which the genuine matching between all system's input and output messages is still hidden, after the attack, among fake ones. Gierlichs *et al.* [5] made the case for modifying such a metric for the scenario where system users send and/or receive multiple messages. As an attacker's goal is to uncover the communication pattern between users, and not simply messages, this modification of the basic metric of Bagai *et al.* [3] was necessary.

In this paper, we developed a metric for measuring the extent to which the communication pattern between senders and receivers of a system was hidden, after a probabilistic attack that is still carried out at the level of messages. When multiple messages were sent and/or received by users, these message multiplicities induce an equivalence relation on the set of all matchings between the system's input and output messages. We showed that a probabilistic attack ends up assigning a probability value to each equivalence class of this relation, which is essentially the likelihood of the genuine matching being contained in that class. Our main offering in the paper was a method for computing this probability value for any given class. We also demonstrated that these equivalence classes in fact correspond to possible communication patterns between senders and receivers, and then gave a method based on Shannon entropy [8] to determine the system-wide anonymity level after the attack.

Three approaches already exist in the literature for a similar problem, and we compared our results with all three. We established that the scope of our method is by far the widest, as it measures the level of anonymity remaining after any of the uncountably infinite possible probabilistic attacks is carried out. In contrast, each of the existing methods only achieved this, to varying degrees, for any of the finite possible infeasibility attacks. Not only is the class of infeasibility attacks finite, it is also a subclass of probabilistic ones.

Our metric gives an exact measure of the system's anonymity level. Computation of this exact value is not efficient, mainly due to the inherent complexity of computing permanent values of matrices. We left development of efficient heuristics that provide an approximate measure of anonymity within specified error bounds as future work. Another direction for future work is to identify special subclasses of attacks for which anonymity levels can be determined efficiently.

Recently, Bagai and Tang [27] showed that adding data caching abilities to the system model of Bagai *et al.* [3] results in increased anonymity. It is possible to extend the metric of this paper for this additional system ability, and we are currently working in this direction.

References

[1] A. Serjantov and G. Danezis, "Toward an information theoretic metric for anonymity," in *Proc. 2nd Privacy Enhancing Technologies Workshop*, LNCS 2482. 2002, pp. 41–53.

[2] C. Diaz, S. Seys, J. Claessens, and B. Preneel, "Toward measuring anonymity," in *Proc. 2nd Privacy Enhancing Technologies Workshop*, LNCS 2482. 2002, pp. 54–68.

[3] R. Bagai, H. Lu, R. Li, and B. Tang, "An accurate system-wide anonymity metric for probabilistic attacks," in *Proc. 11th Int. PETS*, LNCS 6794. 2011, pp. 117–133.

[4] M. Edman, F. Sivrikaya, and B. Yener, "A combinatorial approach to measuring anonymity," in *Proc. IEEE Int. Conf. Intell. Security Inform.*, May 2007, pp. 356–363.

[5] B. Gierlichs, C. Troncoso, C. Diaz, B. Preneel, and I. Verbauwhede, "Revisiting a combinatorial approach toward measuring anonymity," in *Proc. 7th ACM Workshop Privacy Electron. Soc.*, 2008, pp. 111–116.

[6] J.-C. Grégoire and A. M. Hamel. (2009). *A Combinatorial Enumeration Approach for Measuring Anonymity* [Online]. Available: http://arxiv.org/abs/0902.1663.

[7] R. Bagai, B. Tang, A. Khan, and A. Samad, "A system-wide anonymity metric for users sending and receiving multiple messages," *Int. J. Information Security*, 2012, under review.

[8] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[9] G. Danezis, C. Diaz, and P. Syverson, "Systems for anonymous communication," in *CRC Handbook of Financial Cryptography and Security* (Series CRC Cryptography and Network Security Series), B. Rosenberg and D. Stinson, Eds. New York: Chapman and Hall, 2010, pp. 341–390.

[10] H. Minc, *Permanents* (Series Encyclopedia of Mathematics and Its Applications, vol. 6). Reading, MA: Addison-Wesley, 1978.

[11] A. Asratian, T. Denley, and R. Häggkvist, *Bipartite Graphs and Their Applications*. Cambridge: Cambridge Univ. Press, 1998.

[12] L. Valiant, "The complexity of computing the permanent," *Theor. Comput. Sci.*, vol. 8, no. 2, pp. 189–201, 1979.

[13] M. Jerrum, A. Sinclair, and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries," *J. ACM*, vol. 51, no. 4, pp. 671–697, 2004.

[14] M. Dyer, R. Kannan, and J. Mount, "Sampling contingency tables," *Random Struc. Algor.*, vol. 10, no. 4, pp. 487–506, 1997.

[15] M. Gail and N. Mantel, "Counting the number of $r \times c$ contingency tables with fixed margins," *J. Am. Statist. Assoc.*, vol. 72, no. 360a, pp. 859–862, 1977.

[16] F. Greselin, "Counting and enumerating frequency tables with given margins," *Statist. Appl.*, vol. 1, no. 2, pp. 87–104, 2003.

[17] J. Macdonald, *Symmetric Functions and Hall Polynomial*. Oxford, U.K.: Clarendon Press, 1979.

[18] S. Kijima and T. Matsui, "Approximate counting scheme for $m \times n$ contingency tables," *IEICE Trans. Inform. Syst.*, vol. E87-D, pp. 308–314, Feb. 2004.

[19] A. Barvinok, Z. Luria, A. Samorodnitsky, and A. Yong, "An approximation algorithm for counting contingency tables," *Random Structures Algor.*, vol. 37, no. 1, pp. 25–66, 2010.

[20] A. Barvinok and J. Hartigan, "An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums," *Trans. Am. Math. Soc.*, 2012, vol. 364, pp. 4323–4368.

[21] P. Diaconis and A. Gangolli, "Rectangular arrays with fixed margins," in *Discrete Probability and Algorithms* (The IMA Volumes in Mathematics and Its Applications, vol. 72), D. Aldous, P. Diaconis, J. Spencer, and J. Steele, Eds. Berlin: Springer-Verlag, 1995, pp. 15–41.

[22] A. Barvinok, "Matrices with prescribed row and column sums," *Linear Algebra Appl.*, vol. 436, no. 4, pp. 820–844, 2012.

[23] L. Lakshmanan, R. Ng, and G. Ramesh, "To do or not to do: The dilemma of disclosing anonymized data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2005, pp. 61–72.

[24] L. Cottrell, "Mixmaster and remailer attacks," Obscura Information Security, La Mesa, CA, Tech. Rep., 1994.

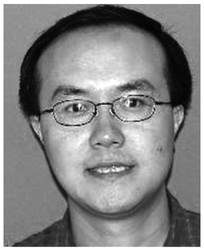
[25] C. Diaz and A. Serjantov, "Generalising mixes," in *Proc. 3rd Privacy Enhancing Technol. Workshop*, LNCS 2760. 2003, pp. 18–31.

[26] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pacific J. Math.*, vol. 21, no. 2, pp. 343–348, 1967.

[27] R. Bagai and B. Tang, "Data caching for enhancing anonymity," in *Proc. 25th IEEE Int. Conf. AINA*, Mar. 2011, pp. 135–142.

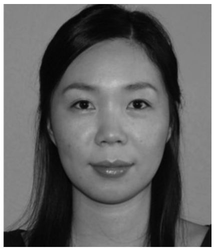
Rajiv Bagai received the B.S. degree in computer science from the Birla Institute of Technology and Science (BITS), Pilani, India, and the M.S. and Ph.D. degrees in computer science from the University of Victoria, Victoria, BC, Canada.

Currently, he is an Associate Professor with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS. His current research interests include web anonymity, but in the past he has worked in logic programming and paraconsistent databases.



Bin Tang (M'11) received the B.S. degree in physics from Peking University, Beijing, China, in 1997, and the M.S. degree in materials science and the M.S. and Ph.D. degrees in computer science from Stony Brook University, Stony Brook, NY, in 2000, 2002, and 2007, respectively.

Currently, he is an Assistant Professor with the Department of Computer Science, Azusa Pacific University, Azusa, CA. His current research interests include algorithmic aspects of data intensive sensor networks.



Euna Kim received the Bachelor's degree in computer graphic design from Bellevue University, Council Bluffs, IA. She is currently pursuing the M.S. degree in computer science with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS.

Her current research interests include web anonymity and privacy.