EFFECT OF DATA CACHING ON SYSTEMWIDE ANONYMITY WITH USERS SENDING AND RECEIVING MULTIPLE MESSAGES

A Thesis by

Ahsan Ahmad Khan

B.E., NED University of Engineering & Technology, Pakistan, 2007

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

May 2012

© Copyright 2012 by Ahsan Ahmad Khan

All Rights Reserved

EFFECT OF DATA CACHING ON SYSTEMWIDE ANONYMITY WITH USERS SENDING AND RECEIVING MULTIPLE MESSAGES

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Electrical Engineering.

_____
Rajiv Bagai, Committee Chair

_____
Murtuza Jadliwala, Committee Member

_____
Esra Buyuktahtakin, Committee Member

ACKNOWLEDGEMENTS

I would like to thank many people without the support of whom this thesis would not have been possible. First, I thank my advisor, Dr. Rajiv Bagai, for his concern, support, help, and advice with every part of my thesis. I also thank Dr. Bin Tang for his support and guidance. I express gratitude to all my friends, especially Abdus Samad, for his kind help and support throughout the course of my thesis work.

ABSTRACT

Data caching is a well-acknowledged method for reducing access times to commonly requested data. Data caching yields considerable results when employed in communication systems that exhibit bidirectional communication, such as web browsing. This technique has been used in anonymous networks to increase the overall anonymity provided by the system. With the increasing number of anonymous networks in place today, the need arises for a metric in order to differentiate the degree of anonymity provided by these systems. This thesis presents a method to compute the degree of anonymity provided by such systems. The model focuses on an anonymous system employing data caching and builds on existing related work in order to allow senders to send multiple messages and receivers to receive multiple messages. A systemwide metric is proposed for measuring the anonymity provided by such systems and is then tested under special conditions. The thesis concludes with an analysis of a pool mix system employing data caching.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Anonymous Networks

Privacy and security were not among the main concerns for engineers who designed the Internet and its accompanying communication protocols, for the simple reason that security was not a critical factor at that time. Today, that has all changed with the advent of applications such as anonymous web browsing and emailing, e-voting, online banking, and various others for which users demand privacy and secrecy. To make such applications possible, engineers have had to fabricate anonymous systems and networks over the prevalent infrastructure. Anonymous networks enable users to perform various functions with a level of assurance that their identity is unknown.

Chaum's Mix-Net [7] design was one of the first modern anonymity networks. This system employed a chain of proxy servers called "mixes." The sender-receiver relationship was hidden by encrypting messages in layers of public-key cryptography. Each mix along the path decrypts, delays, and re-orders messages before forwarding them to achieve a semblance of anonymous communication.

Pfitzmann et al. [25] constructed a system to anonymize ISDN telephone conversations. This work was later generalized to provide an outline for real-time, mixed communications [19]. Later, concepts from both ISDN and real-time mixes were reformed for anonymous web browsing. These networks were known as Web MIXes [5]. Web MIXes always use cascades of mixes, which ensure that each message is handled by all mixed in the same order. The benefit of this mechanism is that it eliminates the need for routing information to be passed along with the messages.

Further examples of anonymous networks include Babel [17] and Mixmaster [23], which were designed in the mid-nineties. Both of these networks follow a message-based approach, which means sending single messages, usually email, through a mix network. Babel employs the concept of "forward path" to achieve sender anonymity and "return path" to achieve receiver anonymity. The forward part is assembled by the sender by wrapping a message in layers of encryption. The message can also include a return address to be used for replies. Replies use the return part in order to protect the identity of the receiver. Mixmaster has been evolving since its inception and is the most widely used remailer system. It can only provide sender anonymity. Messages are made bitwise unlinkable by hybrid RSA and EDE 3DES encryption, whereas the message length is kept constant by adding noise at the end of the message.

Crowds [26] was developed by Reiter and Rubin at AT&T Laboratories. Its purpose is to provide an anonymous way of accessing the web. A client wishing to surf the web anonymously contacts a central server and receives the list of participants, the "crowd." The client's web request is relayed to a randomly selected node in the crowd. Upon receipt of a web request, each node decides whether the request is to be further relayed or sent to the final recipient. Eventually, the reply is sent back to the originating client via the path previously established through the crowd.

Tor [12], a variant of Chaum's Mix-Net design, is another anonymous network in wide use today. Tor utilizes the concept of building a circuit through the network in which each node (onion router) is only aware of its predecessor and successor node. Encrypted traffic travels down this circuit with symmetric-key decryption at each node (like layers on an onion). This is an example of onion routing. Further examples of anonymous networks that employ onion routing are Tarzan [14], MorphMix [27], and several others [8, 18, 24]. Tarzan is a peer-to-peer

network in which every node is an onion router. MorphMix has a similar architecture as that of Tarzan with a major difference being that the route through the network is not specified by the source but chosen by intermediate nodes.

1.2 Anonymity Metrics

Anonymity metrics are used to differentiate between the levels of anonymity provided by different anonymous networks. Various approaches have been developed to determine the level of anonymity provided by an anonymous system.

Some metrics proposed for measuring anonymity look at anonymity from the viewpoint of a single message or user of the system. Serjantov and Danezis [28], and Diaz et al. [11] proposed anonymity metrics from the standpoint of a single message or user. In contrast, Edman et al. [13] proposed a systemwide metric for measuring anonymity. With the latter case, the purpose of the attacker is to correlate all messages entering the anonymous system to all messages exiting the same system. They used a complete bipartite graph to represent an anonymous system where vertices represent the input and output messages, and edges represent a possible relationship between the input and output messages. After an attack renders a number of these relationships to be infeasible, anonymity is measured by comparing the number of feasible perfect matchings to the total number of all possible perfect matchings.

Gierlichs et al. [16] took the work of Edman et al. [13] to the next level by measuring anonymity as the relationship between senders and receivers, not just messages. They argued that the eventual goal of an attacker is to reveal the sender-to-receiver relationship. Gierlichs et al. [16] illustrated that when users send or receive multiple messages, the metric of Edman et al. [13] overestimates the anonymity of the system because it does not take into consideration the equivalence relationship ~ induced on the set of all feasible perfect matchings. Gierlichs et al.

[16] calculated the size of these equivalence classes that lead to the attacker's probability distribution over such communication systems. From there, the actual anonymity of the system is determined using the Shannon entropy technique of Diaz et al. [11].

Bagai et al. [2] reviewed the work of Edman et al. [13] and Gierlichs et al. [16], and took a new approach toward the multiple senders and receivers problem by considering the equivalence relationship $\bowtie$ over all possible perfect matchings instead of just the feasible perfect matchings. The normalized weight of each equivalence class is then calculated, resulting in the probability distribution of all sender-receiver associations. The widely accepted Shannon entropy technique is then used to calculate the anonymity metric. This generic approach toward calculating the anonymity metric made this metric applicable to all classes of biadjacency matrices. Bagai et al. [2] also revealed that the approach of Gierlichs et al. [16] only covered a specific class of biadjacency matrices known as leveled biadjacency matrices.

Berthold et al. [6] used the expression $A = log_2(N)$ to define the degree of anonymity offered by a system, where N is the number of users of the system. This is a very primitive metric and obviously does not represent the anonymity properties of various systems. This metric is also known as the anonymity set size.

A Crowd- based metric was put forth by Reiter and Rubin [26], which has also been used in other contexts. They define the degree of anonymity, A, as a value between 0 (probably exposed) and 1 (absolute privacy), where $A = 1 - p_i$. Possible innocence—$\acute{p}_i$ that u_i is not the sender—is non-negligible; therefore, $\acute{p}_i > 0 + \sigma$, where the threshold $\sigma > 0$. Hence, $A = 1 - p_i = \acute{p}_i > 0 + \sigma$.

The source-hiding property, Θ, is another anonymity metric defined as the greatest probability that can be assigned to any user u_i of being the sender of a message. Therefore, it can

be said that $\Theta = \max(\mathcal{P})$. The term $\mathcal{P}$ is defined as the set of probabilities assigned to each output message as being a certain input message. This assignment is based on the history of the system as defined by Toth and Hornak [29]. Understandably, Θ lies between $1/N$ and 1, where $\Theta = 1/N$ denotes maximum anonymity. The number of users in the system is defined by N.

1.3 Data Caching

Data-caching techniques are widely employed in today's communication systems to reduce service times. With data caching, users requiring certain data can be served by an intermediate node maintaining a cache repository. Therefore, users do not have to traverse the entire path to reach the required resource. In another work, Bagai et al. [1] showed that data caching can be used by anonymous systems for achieving greater anonymity. Data caching employed in anonymous networks enables certain input messages to be served within the anonymous system. Therefore, not all input messages exit the anonymous system, thereby improving anonymity. The specifics of data caching are beyond the scope of this thesis.

1.4 Thesis Contributions

Data caching coupled with anonymous systems is a fairly new concept. Bagai et al. [1] have described how data caching can be used to leverage the level of anonymity provided by a system for senders transmitting a single message and receivers receiving a single message. This thesis takes the former work one step further by considering a system in which senders can send multiple messages and receivers can receive multiple messages. This thesis puts forth an approach on how to model such a system and also formulate a method for determining the level of anonymity provided by such systems.

1.5 Thesis Organization

The remainder of the thesis begins with Chapter 2, which covers related work in the area of anonymity metrics. An overview of the anonymity system model proposed by Edman et al. [13] is provided. This is followed by the work of Bagai et al. [2], which basically takes the work of Edman et al. [13] a step further by proposing an anonymity metric for a system in which users send and receive multiple messages. Chapter 3 proposes a new systemwide metric for anonymous networks employing data caching, which is the main contribution of this thesis. Chapter 4 provides a conclusion for this thesis and presents directions for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Edman's Systemwide Anonymity Metric

A considerable amount of work in the area of anonymity metrics has been done from the viewpoint of a single user or message of a system. Edman et al. [13] proposed a systemwide metric in order to form an expression for the overall anonymity of a system. Their metric was based on the permanent of a matrix, which basically measures the extent of information required by an attacker to determine the relationships between incoming and outgoing messages.

Edman et al. [13] considered an anonymous system as a combination of mixes. An attacker was able to see the messages entering the system and the messages leaving the system. It was assumed that the incoming and outgoing messages had a one-to-one relationship, meaning that every incoming message to the system corresponded to one of the messages exiting the system. A system exhibiting complete anonymity would mean that any input message is equally likely to be any output message.

The goal of an attacker is to break the anonymity of the system by performing an attack that would render some of the input-to-output mappings as infeasible. The attack may be an analysis of message latencies across the system where the attacker labels certain input-to-output pairings as infeasible based on the upper and lower latency bounds. The attack may also comprise analyzing the system's route-length restrictions or comparison of input and output message sizes if the anonymous system does not pad all exiting messages to be of equal size.

Consider an anonymous system with S as the set of t input messages and R as the set of t output messages. It is assumed that each input message corresponds to one of the output messages; therefore, the sizes of set S and R are equal, i.e., $|S| = |R| = t$. Given a set of possible

input-to-output mappings, a bipartite graph is constructed to represent the system. Each edge in the graph indicates a possible association between an input and output message. The bipartite graph can be represented by its adjacency matrix, which is a (0, 1) matrix of size $t \times t$. The adjacency matrix has a row for each input message and a column for each output message. If an input-to-output mapping exists between sender i and receiver j where $i \in S$ and $j \in R$, then the entry $A(i,j) = 1$; otherwise, it equals 0, where A represents the adjacency matrix.

The correct input-to-output relationship corresponds to a perfect matching on the bipartite graph. An anonymous system providing complete anonymity can be represented by a complete bipartite graph, $B = K_{t,t}$, where t is the number of inputs (and outputs). An example of this is shown in Figure 2.1(a). Such a system would have $t!$ perfect matchings because there are $t!$ possible ways to map the set of t input messages to t output messages. On the other hand, a bipartite graph with only one edge between senders and receivers would represent an anonymous system with no anonymity. An example of this is shown in Figure 2.1(b). The bipartite graph of such a system would have only one perfect matching.

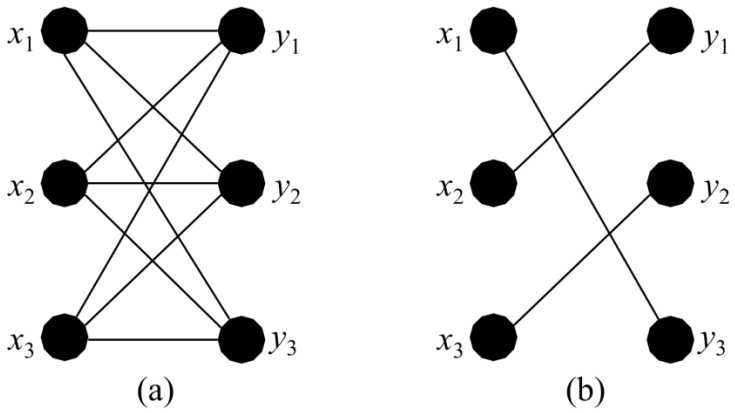

Figure 2.1. (a) Complete anonymity, (b) no anonymity

Therefore, it can be understood that given a bipartite graph, the number of perfect matchings indicate the anonymity provided by the anonymous system. Also, counting the

number of perfect matchings of the bipartite graph is equivalent to the permanent of its corresponding adjacency matrix. The permanent of a $t \times t$ matrix of real numbers is defined as

$$per(A) = \sum_{\pi} \prod_{i=1}^{t} A\big(i, \pi(i)\big), \tag{2.1}$$

where the summation is over all permutations of $\{1,2,\ldots,t\}$.

Based on the matrix permanent, Edman et al. [13] defined the anonymity metric as

$$d(A) = \begin{cases} 0 & \text{if } t = 1, \\ \dfrac{\log(per(A))}{\log(t!)} & \text{otherwise.} \end{cases} \tag{2.2}$$

As can be seen, the above anonymity metric is a comparison of the feasible number of perfect matchings, per(A), with the maximum number of possible perfect matchings, $t!$. A value of 0 means no anonymity is offered by the system, whereas a value of 1 means the system offers complete anonymity.

An attack on an anonymous system will result in certain input-output mappings as infeasible. This attack will produce a bipartite graph that lies somewhere between Figure 2.1(a) and Figure 2.1(b). Let this graph of probable input-output pairings after an attack be called the candidacy graph of the attack. The amount of information obtained after the attack will determine which edges can be removed from the complete bipartite graph. Certain attacks make use of the message length to correlate between input and output messages. To guard against such an attack, many anonymous systems pad their messages to become equal in size before being transmitted. However, such systems exhibit maximum route-length constraints. This is because each message contains the addresses of all nodes through which it will traverse upon leaving the sender. Serjantov and Danezis [28] describe an attack that takes advantage of knowledge of the maximum route-length. The example of Figure 2.2 shows this attack in action.

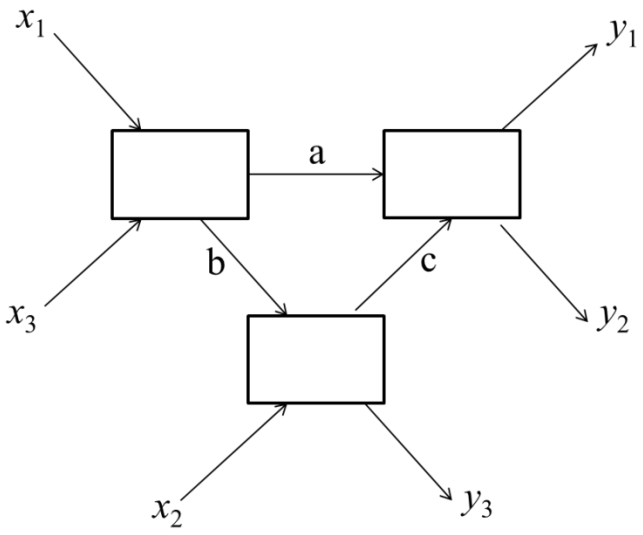

Figure 2.2. Route length attack

Figure 2.2 shows an anonymous system employing three mix nodes, incoming messages $X = \{x_1, x_2, x_3\}$, and outgoing messages $Y = \{y_1, y_2, y_3\}$. Messages *a, b,* and *c* are internal to the system. Suppose that the attacker knows the maximum route length for this system, which is equal to 2. This means that a message can only traverse two nodes. The attacker can observe messages entering and leaving each node. Knowing that the maximum route length is 2, the attacker can safely conclude that message *c* must be x_2. If message *c* were either x_1 or x_3, this would mean that the messages would have violated the maximum route length constraint by traversing three nodes. Therefore, y_3 cannot be x_2. Figure 2.3(a) shows the resulting candidacy graph with the edge missing between x_2 and y_3. Fig 2.3(b) shows the resulting adjacency matrix. For the system in Figure 2.3(a), the anonymity offered by the system can now be calculated as $\log(4)/\log(6) \approx 0.77$.

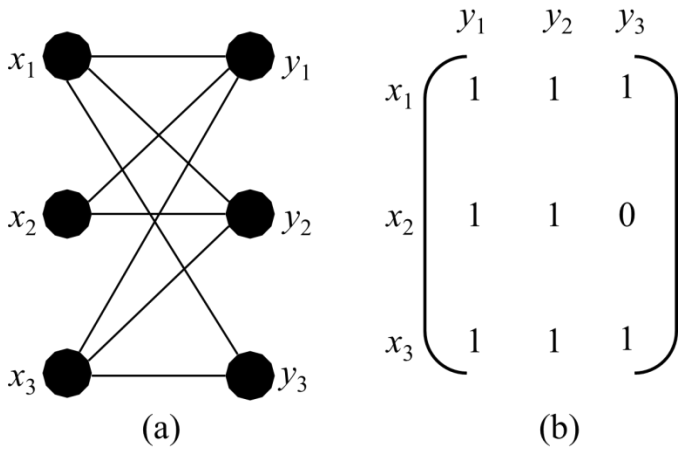

Figure 2.3. (a) Bipartite graph B, (b) resulting adjacency matrix A

2.2 Data Caching for Improved Anonymity

A growing number of anonymous systems are used for applications where the communication between the sender and receiver is bidirectional, such as anonymous web-surfing. In such applications, clients send anonymous web requests to servers and receive responses accordingly. If such an anonymous system employs an internal cache to store the most frequently requested content so that subsequent users requesting the same content can be served by the system's cache, the sender's request does not have to travel all the way to the web server for a response. This results in the number of incoming messages to such an anonymous system being greater than or equal to the number of outgoing messages, with the difference in messages said to be served by the system's internal cache.

Bagai et al. [1] considered such an anonymous system employing data caching and worked toward an expression for the anonymity provided by this system. Let S be the set of input messages entering a system, and T be the set of output messages exiting the system. An anonymous system employing data caching exhibits $|S| > |T|$. Out of the $|S| = m$ input messages to the system, only $|T| = n$ messages appear as output messages, while the remaining $m - n$

messages are assumed to be served by the system's cache, where $m \geq n$. The candidacy graph of an attack will now be a bipartite graph between S and T whose set of edges will be a subset of the set of edges in the complete $K_{m,n}$ bipartite graph. As an example, consider the system of Figure 2.3(a). Suppose that the incoming message x_3 has been served by the system's cache and therefore does not appear as an output message. The resulting candidacy graph is shown in Figure 2.4(a).

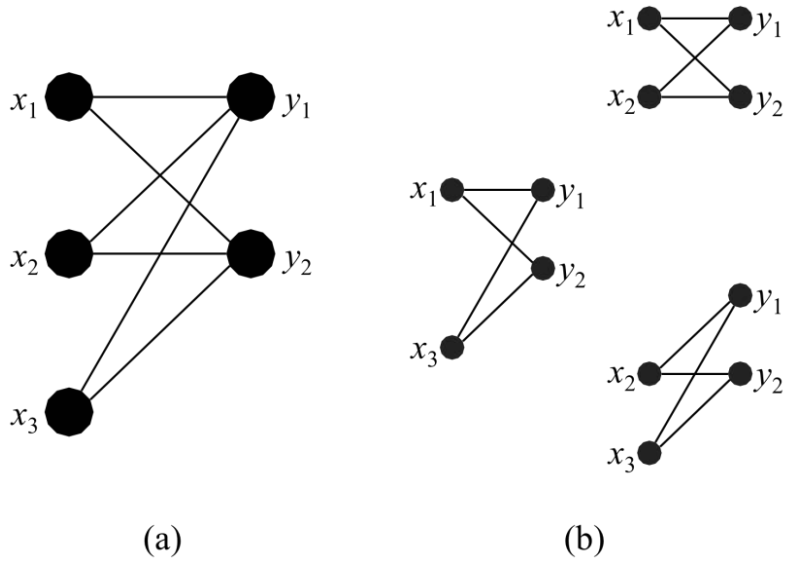

Figure 2.4. (a) Bipartite graph G, (b) left-projections of G

For a system employing data caching, Bagai et al. [1] noted that since the $m - n$ messages do not appear as output, they cannot be mapped to any of the outgoing messages. A perfect matching between S and T is not possible because of their different sizes by virtue of data caching. However, the n outgoing messages must be some n out of m incoming messages. The total number of possible matchings between T and any subset of S of size n is a good indication of the anonymity provided by the system after an attack.

An anonymous system employing data caching can be represented by a bipartite graph $G = (U, V, E)$, where $U = S$, $V = T$, and E is the set of edges representing all possible sender-to-

receiver mappings. For any set U, and $n \geq 0$, let $S_n(U)$ be the set of all subsets of U of size n. A *left-projection* of a bipartite graph $G = (U, V, E)$ is defined as a bipartite graph $P = (W, V, F)$ such that

$$W \in S_{|V|}(U), \qquad and$$

$$F = \{\langle u, v\rangle \in E \mid u \in W\}$$

In simple words, P is a subgraph obtained from G by removing any $|U| - |V|$ vertices for U along with the edges connected to those vertices. Let $\mathcal{L}(G)$ represent the set of all left-projections of G. Bagai et al. [1] defined the anonymity provided by a system employing data caching after an attack that results in a candidacy graph $G = (S, T, E)$, where $|S| = m \geq 1$, $|T| = n \geq 0$, and $m \geq n$ as

$$d'(G) = \begin{cases} 0 & \text{if } m = n = 1, \\ 1 & \text{if } n = 1, \\ \frac{1}{m}\left[(m-n) + \frac{n \log(\sum_{P \epsilon \mathcal{L}(G)} \hat{P})}{\log(\binom{m}{n} n!)}\right] & \text{otherwise.} \end{cases} \tag{2.3}$$

The value of the above metric lies between 0 (for no anonymity) and 1 (for full anonymity). As can be seen, $m - n$ of the m input messages are served by the system's internal cache and hence the anonymity provided to those $m - n$ messages is maximal, i.e., 1. The total number of perfect matchings within G for the remaining n messages is the sum of the number of perfect matchings within all left-projections of G, i.e., $\sum_{P \epsilon \mathcal{L}(G)} \hat{P}$. There are also $\binom{m}{n}$ left-projections, each of which contains $n!$ perfect matchings.

2.3 Anonymity Metric for Multiple Senders and Receivers

Gierlichs et al. [16] were the first to revisit the approach of Edman et al. [13]. They argued that instead of determining the exact input and output message pairings, a more realistic goal for an attacker would be to determine the sender-to-receiver relationship. They showed that

the metric of Edman et al. [13] overestimated the system anonymity when nodes send or receive multiple messages. Gierlichs et al. [16] discovered that an equivalence relation, represented by ~, is induced on the set of all feasible perfect matchings. This equivalence relationship was used to formulate a revised expression for the level of anonymity in a system involving nodes sending or receiving multiple messages.

The equivalence relation ~ of Gierlichs et al. [16] is induced by the sender and receiver multiplicity vectors and a given biadjacency matrix. Thus, the equivalence relation ~ is defined on the set of all perfect matchings reckoned feasible by the given biadjacency matrix. Bagai et al. [2] showed that a major limitation of this approach is that the resultant equivalence class sizes from ~ work correctly for a small number of biadjacency matrices called leveled biadjacency matrices. This behavior prompted Bagai et al. [2] to consider a new approach for the problem of nodes sending or receiving multiple messages. They considered an equivalence relation $\bowtie$ over the set of all possible perfect matchings of the complete bipartite graph of Edman et al. [13]. To illustrate this effect on the level of anonymity in such systems, consider that messages x_1 and x_2 of Figure 2.1(a) belong to the same sender A. Looking at the perfect matchings resulting from Figure 2.1(a), half of them have the edge $\langle x_1, y_1 \rangle$, while the other half have the edge $\langle x_2, y_1 \rangle$. It can therefore be concluded that y_1 was sent by A because x_1 and x_2 were both sent by A. The attacker is not concerned whether y_1 is x_1 or x_2 because the attacker is interested in the sender-to-receiver relationship, not the individual input-to-output message relationship. This phenomenon effectively brings down the anonymity of the system, which is not reflected by the metric of Edman et al. [13]. Bagai et al. [2] arrived at a new expression to measure the level of anonymity in such communication systems that basically measures the sender-to-receiver relationship anonymity instead of input-to-output message relationship anonymity.

Consider a system having m senders and n receivers. Let X_i be the set of messages sent by sender i such that $i \in \{1,2,\ldots,m\}$. Similarly, let Y_j be the set of messages received by receiver j such that $j \in \{1,2,\ldots,n\}$. Also, the total number of messages sent, t, equals the total number of messages received (also t). Let β be the set of all possible $t!$ perfect matchings between X and Y. Bagai et al. [2] defined an equivalence relation $\bowtie$ over β. E_1 is said to be equivalent to E_2 if they have the same association matrix, where E_1 and E_2 are subsets of $X \times Y$. An association matrix of E, written as $\mathbf{Z}(E)$, is a $m \times n$ matrix of nonnegative integers, where the entry $\mathbf{Z}(E)_{i,j}$ represents the number of messages from sender i to receiver j. In simple words, an association matrix is a possible instance of messages sent by m senders to n receivers where rows represent the number of messages sent by senders and columns represent the number of messages received by receivers. Perfect matchings are therefore considered equivalent if they contain the same number of messages travelling from each sender to each receiver. Figure 2.5 shows an example of two equivalent perfect matchings and their resulting common association matrix.

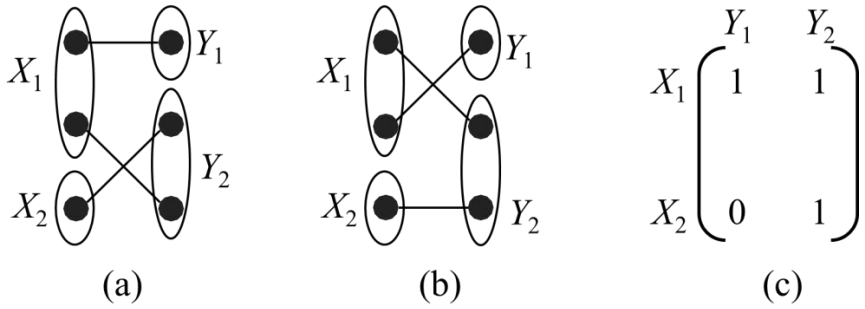

Figure 2.5. (a) E_1, (b) E_2, (c) common association matrix $\mathbf{Z}(E_1) = \mathbf{Z}(E_2)$

Bagai et al. [2] has provided several references for work done in the direction of calculating the number of equivalence classes into which the equivalence relation $\bowtie$ partitions the set of all possible perfect matchings β. Valiant [30] and Jerrum et al. [20] work toward this problem in the context of counting problems. Gail and Mantel [15] put forth a method centered

on recurrence. MacDonald [22] came up with a technique that engaged symmetric functions. Further methods are discussed in Kijima and Matsui [21], Barvinok et al. [4], and Barvinok and Hartigan [3].

Cardinality of an equivalence class is defined as the number of perfect matchings present in an equivalence class identified by its $m \times n$ association matrix. Bagai et al. [2] developed an iterative technique for accurately calculating the cardinality for the most generic case where $m \times t > n$. Based on the number of senders m, the number of receivers n, and the association matrix Z for the equivalence class in question, cardinality is

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

After an attack has been carried out on a system, the attacker has determined the infeasibility of certain input-to-output message pairings and the relationship of the senders to the input messages and the relationship of the receiver to output messages of the system. The former results in a $t \times t$ biadjacency matrix A, whereas the latter brings forth the sender and receiver multiplicity vectors S and R. The cardinality of an equivalence class is dependent only on the multiplicity vectors S and R and the corresponding association matrix. Let $Z'_{S,R}(\beta)$ denote the set of all $m \times n$ association matrices, with S as their row-sum vector and R as their column-sum vector. After an attack has resulted in a biadjacency matrix A, rendering some input-to-output message pairings as infeasible, it is now interesting to focus on the feasible number of perfect matchings within an equivalence class. Recall that the cardinality of an equivalence class represents the number of all possible perfect matchings. Bagai et al. [2] defined weight $W_A(Z)$ to represent the number of feasible perfect matchings in an equivalence class represented by its association matrix Z.

The following expression is a recursive formula whose depth of recursion is exactly the number of nonzero entries in Z:

$$W_A(Z) = \begin{cases} \sum_{A[\![I;J]\!] \in \mathcal{E}_{(A;\, i \to j)}} \left[per(A[\![I;J]\!]) \,.\, W_{\hat{A}[\![I;J]\!]}\left(\hat{Z}[\![i;j]\!]\right)\right] & \text{if } Z_{ij} \neq 0 \text{ for some } i \text{ and } j, \\ 1 & \text{otherwise.} \end{cases} \tag{2.4}$$

Since the number of feasible perfect matchings equals the permanent of the biadjacency matrix A, it can be said that

$$\sum_{Z \epsilon Z'_{S,R}(\beta)} W_A(Z) = per(A) \tag{2.5}$$

Let $w_A(Z) = W_A(Z)/per(A)$ represent the normalized weight of Z. These normalized weights provide the probability distribution on the set of $Z'_{S,R}(\beta)$ of all feasible sender-to-receiver associations. Bagai et al. [2] used the well-accepted Shannon entropy technique over the probability distribution given by $w_A(Z)$ to measure the level of the attacker's improbability of determining the correct sender-to-receiver relationship. Given a $t \times t$ biadjacency matrix A and the multiplicity vectors S and R resulting from an attack, Bagai et al. [2] defined the system's degree to anonymity as

$$\delta_{S,R}(A) = \begin{cases} 0 & \text{if } t = 1, \\ \dfrac{-\sum\{w_A(Z) \,.\, \log\left(w_A(Z)\right) \,:\, Z \in Z'_{S,R}(\beta)\}}{\log(t!)} & \text{otherwise.} \end{cases} \tag{2.6}$$

The value of $\delta_{S,R}(A)$ will always fall between 0 and 1, where 0 represents no anonymity and 1 represents complete anonymity. Bagai et al. [2] later went on to show that the anonymity metric of Gierlichs et al. [16] is limited to a limited class of biadjacency matrices referred to as leveled matrices.

2.4 Computation of Anonymity Metric

Consider a system with four input and four output messages. Figure 2.6 (a) illustrates the bipartite graph G of this system after an attack has been executed. Figure 2.6 (b) displays the resulting biadjacency matrix A.

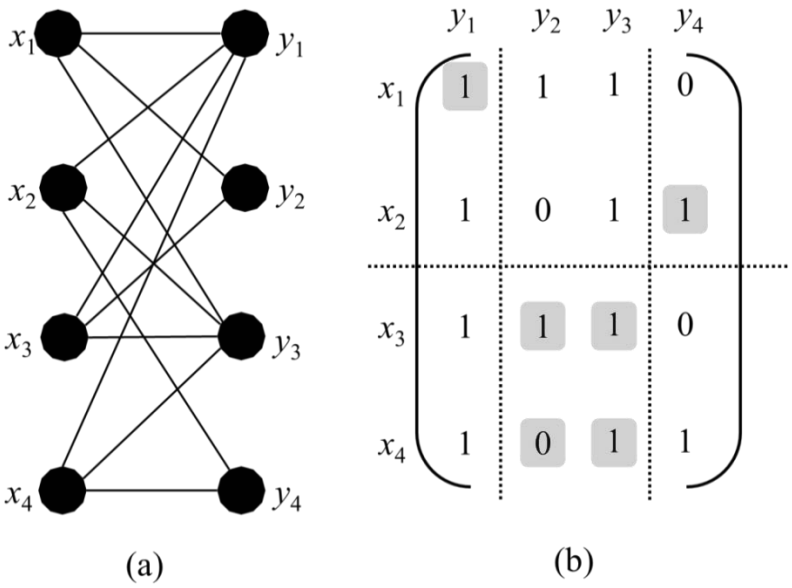

Figure 2.6. (a) Bipartite graph G; (b) biadjacency matrix A, its regions induced by multiplicity vectors S and R, and an example extract collection shaded in grey.

This attack also reveals the sender-to-input message and receiver-to-output message associations. Input messages x_1 and x_2 belong to sender X_1, whereas input messages x_3 and x_4 belong to sender X_2. Output message y_1 belongs to receiver Y_1, y_2 and y_3 belong to receiver Y_2, and y_4 belongs to receiver Y_3. This results in the sender multiplicity vector $S = \langle 2, 2 \rangle$ and receiver multiplicity vector $R = \langle 1, 2, 1 \rangle$. Figure 2.7 displays the association matrices of the four equivalence classes resulting from the sender and receiver multiplicity vectors. These equivalence classes are constructed with S as their row-sums vector and R as their column-sums vector.

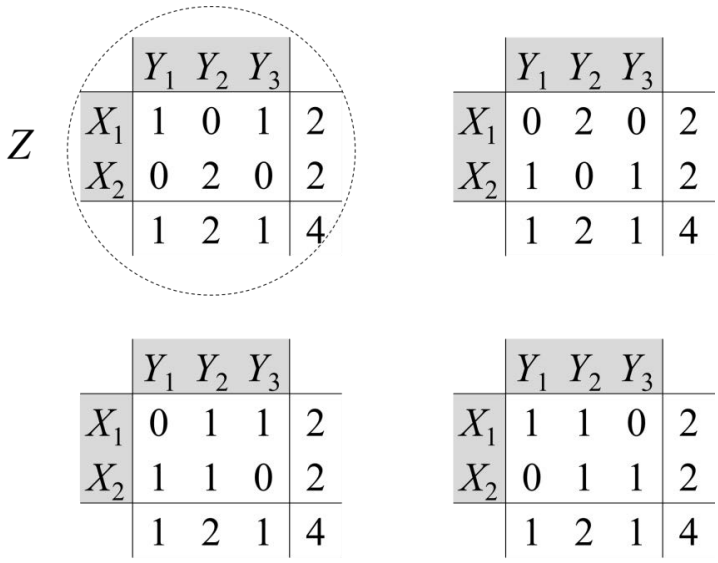

Figure 2.7. Association matrices of all equivalence classes

The sender S and receiver R multiplicities induce six regions on the biadjacency matrix A, as shown in Figure 2.6(b). Now it is important to calculate the weight $W_A(Z)$ of equivalence class Z, shown in equation (2.4) and equation (2.5) by Bagai et al. [2]. Taking the equivalence class Z circled in Figure 2.7, there are three nonzero entries in this class, namely $z_{11} = z_{13} = 1$, and $z_{22} = 2$. Therefore, any matching in this equivalence class constitutes three extracts of regions of A with pairwise disjoint row-sets and column-sets: a 1×1 extract from regions $Reg_{(A;\, 1\to1)}$, $Reg_{(A;\, 1\to3)}$, and a 2×2 extract from region $Reg_{(A;\, 2\to2)}$. An example collection of such extracts is shown shaded in grey in Figure 2.6(b). The product of permanents of extracts in this collection is

$$1 \bullet 1 \bullet (1 \bullet 1 + 1 \bullet 0) = 1.$$

The recursive method of Bagai et al. [2] of computing the weight of Z adds the above value for all such collections of extracts. It can be easily seen that for all such collections, $W_A(Z) = 1$. The weights $W_A(Z)$ for the other three association matrices can be evaluated in a similar manner. The resulting weights are 1, 3, and 3. The sum of the weights of all four association

matrices is per(A) = 8. Dividing the individual weights by per(A) leads to the normalized weights w_{A} of these matrices: 1/8, 1/8, 3/8, and 3/8. The metric defined in Section 2.3 yields $\delta_{S,R}(A) = 0.393$.

CHAPTER 3

DATA CACHING WITH USERS SENDING/RECEIVING MULTIPLE MESSAGES

3.1 Introduction

Bagai et al. [1] showed that anonymous systems employing data caching can be a means of increasing the degree of anonymity provided by the system. The gain in anonymity is proportional to the level of data caching performed by the system. Such a system is especially effective for bidirectional communications such as web browsing. Consider a scenario where clients access web servers for anonymous web surfing. An anonymous system employing data caching has the ability to store frequently demanded content in its cache. This cached content will then be used to serve succeeding demands. Thus, demands that can be served by the cached content do not need to be sent to the end server. The phenomenon of data caching results in a system where the number of incoming messages m are equal to or less than the number of outgoing messages n with $m - n$ messages being served by the system cache.

3.2 System Model

Consider an anonymous system employing data caching. For the purposes of this thesis, we are not concerned with the specifics of how and what data is cached. It is presumed that the system maintains one or more internal caches of the most frequent content requested by users. Every incoming message is evaluated as to whether it can be served by the system's internal cache. If it can, then the incoming message does not appear as an outgoing message for the system.

Since the number of incoming messages is greater than the number of outgoing messages, a perfect matching between the incoming and outgoing messages is not possible here. To work around this, the bipartite graph G is divided into its corresponding left-projections, as described

in Section 2.2. A left-projection of G is its subgraph and is obtained by removing any $m - n$ vertices from G (and edges connected to those vertices). The bipartite graph G can be represented by an $m \times n$ non-square biadjacency matrix, whereas the resulting left-projections can be represented by respective $n \times n$ square biadjacency matrices.

Consider the system in Figure 3.1 employing data caching. This system takes m messages as input and delivers n messages as output, with $m - n$ messages said to be served by the system's internal cache. It is also assumed that senders can send multiple messages and receivers can receive multiple messages. Let $X = \{X_1, X_2, .., X_\sigma\}$ represent the set of senders with sender multiplicities $S = \langle S_1, S_2, ..., S_\sigma \rangle$ and $Y = \{Y_1, Y_2, .., Y_\rho\}$ represent the set of receivers with receiver multiplicities $R = \langle R_1, R_2, .. R_\rho \rangle$.

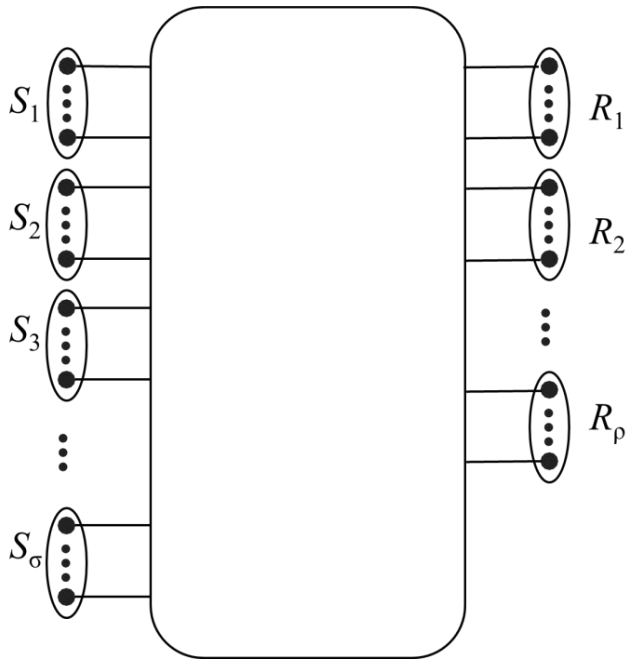

Figure 3.1. Anonymous system employing data caching

The following relations can be drawn from Figure 3.1:

$$S_1 + S_2 + \ldots + S_\sigma = m \tag{3.1}$$

$$R_1 + R_2 + \ldots + R_\rho = n \tag{3.2}$$

$$m \geq n \tag{3.3}$$

3.3 Anonymity Metric

Now, an expression for the degree of anonymity provided by the system in Figure 3.1 is derived. After an attack has been carried out, rendering certain input-to-output message pairings as infeasible, we obtain a non-square $m \times n$ biadjacency matrix A. The resulting left-projections can be represented by $n \times n$ square biadjacency matrices A_t, where $1 \leq t \leq \binom{m}{n}$. The attack also results in multiplicity vectors S and R, which impose an equivalence relation $\bowtie$ over the set of all possible perfect matchings. We use weight $W_{A_t}(Z_u^t)$, as defined by Bagai et al. [2], to determine the number of feasible perfect matchings for a given $n \times n$ biadjacency matrix A_t and a $\sigma \times \rho$ association matrix Z_u^t that corresponds to some equivalence class of $\bowtie$, where $1 \leq t \leq \binom{m}{n}$ and $1 \leq u \leq \theta_t$. Let θ_t be the number of equivalence classes within a particular left-projection. For the $\binom{m}{n}$ number of left-projections, there are $\binom{m}{n}$ number of resulting new sender multiplicities defined as

$$L^1 = \langle L_1{}^1, L_2{}^1, \ldots, L_\sigma{}^1 \rangle \ (1^{\text{st}} \text{ left-projection}) \tag{3.4}$$

$$L^2 = \langle L_1{}^2, L_2{}^2, \ldots, L_\sigma{}^2 \rangle \ (2^{\text{nd}} \text{ left-projection}) \tag{3.5}$$

and so on:

$$L^{\binom{m}{n}} = \langle L_1^{\binom{m}{n}}, L_2^{\binom{m}{n}}, \ldots, L_\sigma^{\binom{m}{n}} \rangle \ (\binom{m}{n}^{\text{th}} \text{ left-projection}) \tag{3.6}$$

Each left-projection is divided into its resulting equivalence classes. Taking the first left-projection with Senders $X = \{X_1, X_2, \ldots, X_\sigma\}$ and Receivers $Y = \{Y_1, Y_2, \ldots, Y_\rho\}$ with sender

multiplicities $L^1 = \langle L_1^1, L_2^1, \ldots, L_\sigma^1 \rangle$ and receiver multiplicities $R = \langle R_1, R_2, \ldots R_\rho \rangle$, a particular equivalence class within the first left-projection can be represented by the $\sigma \times \rho$ association matrix Z shown in Figure 3.2.

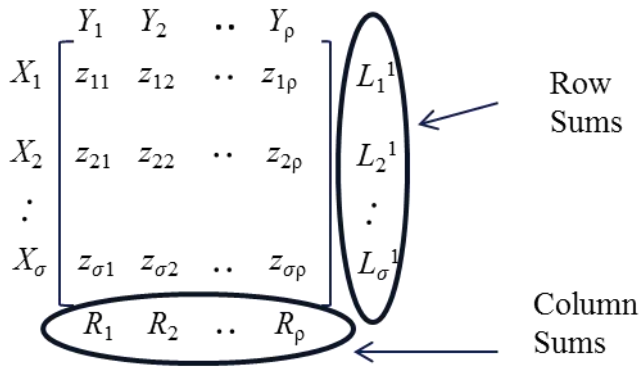

Figure 3.2. $\sigma \times \rho$ association matrix Z for a particular equivalence class

The number of equivalence classes, θ_t, within a left-projection is defined as the number of possible non-negative $\sigma \times \rho$ association matrices that follow the margins L (row sums) and R (column sums):

$$(z_{i1} + z_{i2} + \ldots + z_{i\rho} = L_i) \text{ and } (z_{1j} + z_{2j} + \ldots + z_{\sigma j} = R_j) \tag{3.7}$$

where $(1 \leq i \leq \sigma)$ and $(1 \leq j \leq \rho)$.

The weights resulting from the θ_1 equivalence classes within the first left-projection are defined as

$$\Psi_{A_1} = W_{A_1}(Z_1^1) + W_{A_1}(Z_2^1) + \ldots + W_{A_1}(Z_{\theta_1}^1) = per(A_1) \tag{3.8}$$

Similarly, for the second left-projection, the number of respective equivalence classes will be θ_2 with the following weights:

$$\Psi_{A_2} = W_{A_2}(Z_1^2) + W_{A_2}(Z_2^2) + \ldots + W_{A_2}(Z_{\theta_2}^2) = per(A_2) \tag{3.9}$$

And so on. For the $\binom{m}{n}^{\text{th}}$ left-projection, the number of respective equivalence classes will be $\theta_{\binom{m}{n}}$ with the following weights:

$$\Psi_{A_{\binom{m}{n}}} = W_{A_{\binom{m}{n}}}(Z_1^{\binom{m}{n}}) + W_{A_{\binom{m}{n}}}(Z_2^{\binom{m}{n}}) + \ldots + W_{A_{\binom{m}{n}}}(Z_{\theta_{\binom{m}{n}}}^{\binom{m}{n}}) = per(A_{\binom{m}{n}}) \tag{3.10}$$

The weight profile of the $m \times n$ system employing data caching after an attack A can be written as

$$\begin{aligned} profile(A) = {} & W_{A_1}\left(Z_1^1\right), W_{A_1}\left(Z_2^1\right), \ldots, W_{A_1}\left(Z_{\theta_1}^1\right), W_{A_2}\left(Z_1^2\right), W_{A_2}\left(Z_2^2\right), \ldots, W_{A_2}\left(Z_{\theta_2}^2\right), \\ & \ldots, W_{A_{\binom{m}{n}}}(Z_1^{\binom{m}{n}}), W_{A_{\binom{m}{n}}}(Z_2^{\binom{m}{n}}), \ldots, W_{A_{\binom{m}{n}}}(Z_{\theta_{\binom{m}{n}}}^{\binom{m}{n}}) \end{aligned} \tag{3.11}$$

The above is a sequence of the weights. The normalized weights are

$$\psi = \frac{profile(A)}{per\ A_1 + per\ A_2 + \cdots + per\ A_{\binom{m}{n}}} \tag{3.12}$$

$$\psi = (\psi_1, \psi_2, \psi_3, \ldots, \psi_\Xi)\ ; \text{where } \Xi = \theta_1 + \theta_2 + \cdots + \theta_{\binom{m}{n}} \tag{3.13}$$

The values of ψ_Ξ all add up to 1 over all left-projections. Thus a probability distribution on the set $\mathbb{Z}$ is defined as

$$\mathbb{Z} = \bigcup_{t=1}^{\binom{m}{n}} [\, \acute{Z}_t : \ \acute{Z}_t = \{Z_1^t, Z_2^t, \ldots, Z_{\theta_t}^t\}] \tag{3.14}$$

Each left-projection has a set of equivalence classes of perfect matchings. The set $\mathbb{Z}$ is the union of the sets of equivalence classes. The degree of anonymity provided by the system in Figure 3.1 is

$$d'_{(S,R)}(A) = \begin{cases} 0 & if\ m = n = 1, \\ 1 & if\ n = 0, \\ \frac{1}{m}\left[(m-n) + \frac{n\left(-\sum_{i=1}^{\Xi} \psi_i \,.\log(\psi_i)\right)}{\log\left(\binom{m}{n} n!\right)} \right] & ;otherwise \end{cases} \tag{3.15}$$

The value of the above anonymity metric lies between 0 (for no anonymity) and 1 (for full anonymity). As can be seen in Figure 3.1, $m - n$ of all the m incoming messages are served by the system's internal cache. The anonymity provided for those $m - n$ messages is therefore 1. The probability distribution given by the normalized weights ψ is used to determine the anonymity provided by the system to the n outgoing messages. The well-accepted Shannon entropy of a probability distribution is applied over the probability distribution given by ψ to the n outgoing messages as a measure of the attacker's uncertainty of which of the sender-receiver associations is the actual one. The quantity $[\binom{m}{n} n!\,]$ represents the total number of possible perfect matchings within the entire $m \times n$ anonymous system of Figure 3.1.

3.4 Results

As mentioned earlier, the value of $d'_{(S,R)}(A)$ lies between 0 and 1. Now, how this anonymity metric performs under special conditions is examined. Consider a system offering no data caching, i.e., the number of incoming messages m equals the number of outgoing messages n. For such a system, the expression for the degree of anonymity provided by the metric developed here reduces down to

$$d'_{(S,R)}(A) = \frac{-\sum_{i=1}^{\Xi} \psi_i \,.\log(\psi_i)}{\log(n!)} \tag{3.16}$$

The above expression is equivalent to the anonymity metric for users sending and receiving multiple messages, as given by Bagai et al. [2].

The next special case considers a system offering data caching with sender and receiver multiplicities equal to 1. For this case, the expression for the degree of anonymity becomes

$$d'_{(S,R)}(A) = \frac{1}{m}\left[(m-n) + \frac{n\left(-\sum_{i=1}^{\Xi}\psi_i \,.\log(\psi_i)\right)}{\log\left(\binom{m}{n} n!\right)}\right] \tag{3.17}$$

It is important to note here that since the sender and receiver multiplicities are 1, there are no equivalence classes. Although the above equation is similar to equation (3.15), both equations differ because the term Ξ in equation (3.15) represents the number of equivalence classes over all left-projections whereas the term Ξ in equation (3.17) represents the number of left-projections. The above expression is equivalent to equation (2.3) provided by Bagai et al. [1] for a system employing data caching. Equation (2.3) uses the number of perfect matchings over the set of all left-projections to determine the anonymity level provided to the *n* number of messages exiting the system. On the other hand, this thesis uses the concept of entropy as proposed by Diaz et al. [11] to determine the effective anonymity over the set of the set of all left-projections. Both techniques serve the common purpose to determine the amount of information required for an attacker to determine the true sender-to-receiver relationship.

3.5 Analysis

This section presents an analysis of the developed metric when applied to pool mixes. Cottrell [9] introduced pool mixes as a high-latency strategy to counter attacks aimed at determining the associations between a system's input and output messages. Diaz and Serjantov [10] presented various generalizations of this strategy which function in iterative rounds. In each round, a certain number of messages is collected by the pool mix as incoming messages. These messages are then stored in the internal message pool. The pool mix then sends out a fraction of the messages contained in the message pool as outgoing messages, and the remainder of the

messages left in the pool after any round are contenders for being sent out in future rounds. It is assumed that the pool mix employs an internal data cache, which can be used to serve input messages during rounds of operation.

Now, the developed method is implemented to determine the anonymity provided by an example pool mix after its first two rounds of operation. As shown in Figure 3.3(a), suppose input messages x_1 and x_3 enter the pool mix in Round 1. Only one message, y_2, is output by the pool mix in that round, leaving the other input message retained by the internal message pool for the next round. In Round 2, suppose input messages x_2 and x_4 enter the pool and message y_3 exits the pool. One of the input messages in this round has been served by the pool's internal data cache and hence will not appear as an output message or a message retained by the internal message pool. After Round 2, one message remains in the internal message pool, namely y_1.

Round Number	Input Messages	Number of Messages served by Cache	Pool Messages after Round	Output Messages
1	x_1, x_3	0	one unnamed	y_2
2	x_2, x_4	1	y_1	y_3

(a)

A	y_1	y_2	y_3
x_1	1	1	1
x_2	1	0	1
x_3	1	1	1
x_4	1	0	1

(b)

A_1	y_1	y_2	y_3
x_1	1	1	1
x_2	1	0	1
x_3	1	1	1

A_2	y_1	y_2	y_3
x_1	1	1	1
x_2	1	0	1
x_4	1	0	1

A_3	y_1	y_2	y_3
x_2	1	0	1
x_3	1	1	1
x_4	1	0	1

A_4	y_1	y_2	y_3
x_1	1	1	1
x_3	1	1	1
x_4	1	0	1

(c)

Figure 3.3. (a) Message history of first two rounds of example pool mix employing data caching, (b) resulting biadjacency matrix A, (c) left-projections of A, its regions induced by multiplicity vectors L^t, and R

Figure 3.3(b) shows the resulting 4 × 3 biadjacency matrix A after two rounds of operation of the pool mix. As an example, since message y_2 exited the pool mix before messages x_2 and x_4 arrived as input, therefore it is understood that y_2 cannot be x_2 or x_4. Hence, this observation results in entries of 0 in cells A_{22} and A_{42} of the biadjacency matrix A. Now, suppose the attacker observed that messages x_1 and x_2 were both sent by the same sender, and x_3 and x_4 were sent by another sender, i.e., $X_1 = \{x_1, x_2\}$ and $X_2 = \{x_3, x_4\}$. Also, suppose that the attacker observed that messages y_2 and y_3 were received by the same receiver, i.e., $Y_1 = \{y_1\}$ and $Y_2 = \{y_2, y_3\}$. This results in sender multiplicities $S = \langle 2, 2 \rangle$ and $R = \langle 1, 2 \rangle$. It can also be noted that $m = 4$ and $n = 3$.

To further proceed with determining the anonymity of this system, the biadjacency matrix A is broken down into its left-projections, as described in Section 3.2. Each left-projection results in a new sender multiplicity L^t vector, where $1 \leq t \leq \binom{m}{n}$. The receiver multiplicity vector R remains the same. The new sender multiplicity vectors are as follows:

$L^1 = \langle 2, 1 \rangle$ (for left-projection A_1)

$L^2 = \langle 2, 1 \rangle$ (for left-projection A_2)

$L^3 = \langle 1, 2 \rangle$ (for left-projection A_3)

$L^4 = \langle 1, 2 \rangle$ (for left-projection A_4)

Figure 3.3(c) shows all four possible left-projections along with the regions induced upon them by multiplicity vectors L^t and R. Next, the left-projections are divided into their respective equivalence classes. Taking the example of left-projection A_1, Figure 3.4 shows its resulting equivalence classes based on vectors L^1 and R.

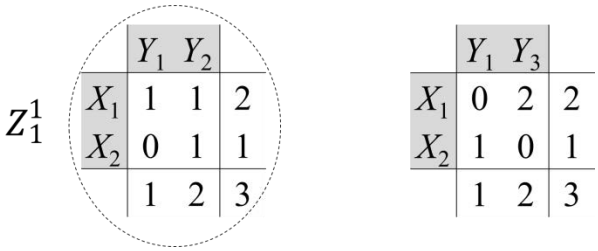

Figure 3.4. Equivalence classes of left-projection A_1.

Now, the weight $W_{A_1}(Z_1^1)$ of equivalence class Z_1^1 , circled in Figure 3.4, is calculated. There are three nonzero entries in this class, namely $z_{11} = z_{12} = z_{22} = 1$. Therefore, any matching in this equivalence class constitutes three extracts of regions of A_1 with pairwise disjoint row-sets and column-sets: a 1 × 1 extract from regions $Reg_{(A;\,1\to1)}$, $Reg_{(A;\,1\to2)}$, and $Reg_{(A;\,2\to2)}$. An example collection of such extracts is shown shaded in grey in Figure 3.3(c). The product of permanents of extracts in this collection is

$$1 \bullet 1 \bullet 1 = 1$$

The recursive method of Bagai et al. [2] for computing the weight of equivalence class adds the above value for all such collections of extracts. It can be easily seen that for all such collections, $W_{A_1}(Z_1^1) = 3$. The weight for the second association matrix can be evaluated in a similar manner, and the resulting weight is 1. The sum of the weights of all two association matrices of left-projection A_1 is $\Psi_{A_1} = W_{A_1}(Z_1^1) + W_{A_1}(Z_2^1) = 3 + 1 = 4 = per(A_1)$. Following the same approach for the other three left-projections yields the following results:

$$\Psi_{A_2} = W_{A_2}(Z_1^2) + W_{A_2}(Z_2^2) = 1 + 1 = 2 = per(A_2)$$

$$\Psi_{A_3} = W_{A_3}(Z_1^3) + W_{A_1}(Z_2^3) = 1 + 1 = 2 = per(A_3)$$

$$\Psi_{A_4} = W_{A_4}(Z_1^4) + W_{A_4}(Z_2^4) = 3 + 1 = 4 = per(A_4)$$

The weight profile as described in Section 3.3 of the pool mix system employing data caching after an attack A can be written as

$$profile(A) = 3, 1, 1, 1, 1, 1, 3, 1$$

Dividing the weight profile by the sum of permanents of all left-projections results in the following normalized weights:

$$\psi = (\frac{1}{4}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{4})$$

According to the metric presented in Section 3.3, the anonymity provided by the pool mix system of Figure 3.3(a) is $d'_{(S,R)}(A) \approx 0.707$. To appreciate the reduction in anonymity caused by message multiplicities, the anonymity level is also computed using the basic metric of Bagai et al. [1], also described in Section 2.2. It can easily be seen that the permanents of the four left-projections are 4, 2, 2, and 4. Hence, the anonymity level achieved by the expression of Bagai et al. [1] is

$$d'(A) = \frac{1}{4}\left[(4-3) + \frac{3\log(4+2+2+4)}{\log(\binom{4}{3}3!)}\right] \approx 0.836$$

The above value is considerably higher than the anonymity level achieved by the metric presented in Section 3.3, i.e., $d'_{(S,R)}(A) \approx 0.707$.

CHAPTER 4

DISCUSSION

4.1 Limitations

For the purposes of this thesis, how the system performs data caching is not of concern. Different data caching algorithms will provide varying results, which in turn will affect the anonymity of the overall system. It is also not detailed how the boundaries of an anonymous system are determined. The types of attacks that can be launched upon an anonymous system and that result in the biadjacency matrix, the sender-to-input message associations, and receiver-to-output message associations are also not described.

4.2 Future Work

This thesis utilizes the work of Bagai et al. [2] to calculate the weight of an equivalence class. This is a recursive formula, the depth of which is the number on nonzero entries in the association matrix Z of that respective equivalence class. Hence, this computation becomes complex with large association matrices. A future line of work could be to improve on the calculation of weight of an equivalence class.

4.3 Conclusion

Depending upon the nature of communication, an anonymous system can employ data caching to improve the overall anonymity provided by the system. For example, in web browsing, an anonymous system utilizing data caching can use its internal cache to serve web requests for which it has cached content. This behavior results in certain input messages of the system not appearing as output messages. The input messages not appearing as output messages are said to have been served by the system's internal cache. This was studied by Bagai et al. [1], and an expression was proposed to express the anonymity provided by such a system.

This thesis takes the work of Bagai et al. [1] one step further by considering a system in which senders can send multiple messages and receivers can receive multiple messages. After breaking down the system into its respective left-projections, this thesis works toward determining the probability distribution over all left-projections for all possible sender-receiver association scenarios. Having obtained the probability distribution, the Shannon entropy technique of Diaz et al. [11] is used to formulate an expression for the anonymity provided by this system. Testing this expression under special cases provides expected results, which prove the validity of our anonymity metric.

REFERENCES

LIST OF REFERENCES

[1] R. Bagai and B. Tang, Data caching for enhancing anonymity. In *Proc. 25th IEEE Int. Conf. on Advanced Information Networking and Applications (AINA)*, pp. 135–142, 2011.

[2] R. Bagai, B. Tang, A. Khan, and A. Samad, A system-wide anonymity metric for users sending and receiving multiple messages. *IEEE Trans. Dependable and Secure Computing*, 2011, submitted.

[3] A. Barvinok and J. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Transactions of the American Mathematical Society*, 2011, to appear.

[4] A. Barvinok, Z. Luria, A. Samorodnitsky, and A. Yong, An approximation algorithm for counting contingency tables. *Random Structures & Algorithms*, vol. 37, pp. 25-66, 2010.

[5] O. Berthold, H. Federrath, and Stefan Kopsell, Web MIXes: A system for anonymous and unobservable Internet access. *Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability*, pp. 115-129, July 2000.

[6] O. Berthold, A. Pfitzmann, and R. Standtke, The disadvantages of free MIX routes and how to overcome them. *Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability*, pp. 30-45, July 2000.

[7] D. Chaum, Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, vol. 4, no. 2, February 1981.

[8] G. Ciaccio, Improving sender anonymity in a structured overlay with imprecise routing. In *Privacy Enhancing Technologies: 6th International Workshop, PET 2006*, pp. 190-207, 2006.

[9] L. Cottrell, Mixmaster and remailer attacks, *Obscura Information Security, Tech. Rep*., 1994.

[10] C. Diaz and A. Serjantov, Generalising mixes. In *Proceedings of 3rd Privacy Enhancing Technologies Workshop, ser. Lecture Notes in Computer Science*, vol. 2760, pp. 18-31, 2003.

[11] C. Diaz, S. Seys, J. Claessens, and B. Preneel, Towards measuring anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop*, pp. 54-58, 2002.

[12] R. Dingledine, N. Mathewson, and P. Syverson, Tor: the second-generation onion router. In *Proceedings of the 13th SENIX Security Symposium*, August 2004.

LIST OF REFERENCES (continued)

[13] M. Edman, F. Sivrikaya, and B. Yenner, A combinatorial approach to measuring anonymity. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, pp. 356-363, 2007.

[14] M. Freedman and R. Morris. Tarzan, A peer-to-peer anonymizing network layer. In *2002 Proceedings of the 9th ACM Conference on Computer and Communications Security, CCS*, pp. 193-206, 2002.

[15] M. Gail and N. Mantel, Counting the number of *r* x *c* contingency tables with fixed margins. *Journal of the American Statistical Association*, vol. 72, pp. 859-862, 1977.

[16] B. Gierlichs, C. Troncoso, C. Diaz, B. Preneel, and I. Verbauwhede, Revisiting a combinatorial approach towards measuring anonymity. In *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society*, Alexandria, VA, 2008, pp. 111-116.

[17] C. Gulcu and G. Tsudik, Mixing E-mail with Babel. In *Proceedings of the Symposium on Network and Distributed Security Symposium – NDSS '96*, pp. 2-16, February1996.

[18] N. Hopper, E. Vasserman, and E. Chan-Tin, How much anonymity does network latency leak? In *CCS'07: Proceedings of the 14th ACM Conference on Computer and Communications Security*, pp. 82–91, 2007.

[19] A. Jerichow, J. Muller, A. Pfitzmann, B. Pfitzmann, and M. Waidner, Real-time MIXes: A bandwidth-efficient anonymity protocol. *IEEE Journal on Selected Areas in Communications*, pp. 495-509, May 1998.

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

[21] S. Kijima and T. Matsui, Approximate counting scheme for *m* x *n* contingency tables. *IEICE Transactions on Information and Systems*, vol. E87-D, pp. 308-314, 2004.

[22] J. Macdonald, *Symmetric Functions and Hall Polynomial*. Clarendon Press, 1979.

[23] U. Moller, L. Cottrell, P. Palfrader, and L. Sassaman, Mixmaster protocol - version 3. IETF Internet Draft, 2003.

[24] A. Nambiar and M. Wright, Salsa: A structured approach to large-scale anonymity. In *CCS'06: Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 17–26, 2006.

LIST OF REFERENCES (continued)

[25] A. Pfitzmann, B. Pfitzmann, and M. Waidner, ISDNMIXes: Untraceable communication with very small bandwidth overhead. In *Kommunikation in Verteilten Systemen, Grundlagen, Anwendungen, Be-trieb, GI/ITG-Fachtagung*, vol. 267, pp.451-463, 1991.

[26] M. Reiter and A. Rubin, Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security (TISSEC)*, pp. 66–92, 1998.

[27] M. Rennhard and B. Plattner, Introducing MorphMix: Peer-to-peer based anonymous internet usage with collusion detection. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society, WPES 2002*, pp. 91–102, 2002.

[28] A. Serjantov and G. Danezis, Towards an information theoretic metric for anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop*, pp. 41-53, 2002.

[29] G. Toth and Z. Hornak, Measuring anonymity in a non-adaptive, real-time system. In *Proceedings of the Workshop on Privacy Enhancing Technologies (PET 2004),* pp. 26-28, May 2004.

[30] L. Valiant, The complexity of computing the permanent. *Theoretical Computer Science*, vol. 8, no. 2, pp. 189-201, 1979.