Probabilistic Treatment of MIXes to Hamper Traffic Analysis

Dakshi Agrawal
IBM Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532, USA
agrawal@us.ibm.com

Dogan Kesdogan and Stefan Penz
Aachen University of Technology
Computer Science Department Informatik IV
Ahornstr. 55, D-52074 Aachen, Germany
{kesdogan, penz}@informatik.rwth-aachen.de

Abstract

The goal of anonymity providing techniques is to preserve the privacy of users, who has communicated with whom, for how long, and from which location, by hiding traffic information. This is accomplished by organizing additional traffic to conceal particular communication relationships and by embedding the sender and receiver of a message in their respective anonymity sets. *If the number of overall participants is greater than the size of the anonymity set and if the anonymity set changes with time due to unsynchronized participants, then the anonymity technique becomes prone to* traffic analysis *attacks. In this paper, we are interested in the statistical properties of the* disclosure attack, *a newly suggested traffic analysis attack on the MIXes. Our goal is to provide analytical estimates of the number of observations required by the disclosure attack and to identify fundamental (but avoidable) 'weak operational modes' of the MIXes and thus to protect users against a traffic analysis by the disclosure attack.*

1. Introduction

Anonymity techniques to prevent network traffic analysis date back to the 1970's and 1980's when David Chaum and others suggested several revolutionary techniques including broadcast and implicit addresses, MIXes, DC-Networks, and PIR [5, 4, 6, 7, 14, 18]. Among these techniques, the MIX concept can be considered as the most popular and deployment friendly [1, 2, 8]. It has been proposed for networks like GSM, ISDN and the Internet [3, 9, 10, 11, 13, 16, 17, 20, 21]. However, in such *open environments* the network traffic is usually unconstrained and can be vary in each time epoch. The variation in traffic can be exploited by an attacker to gain information about a targeted user, e.g., the attacker can establish statistics about the on-line and off-line behavior of users.

Kesdogan, Agrawal, and Penz [15] have suggested a new type of traffic analysis attack, the so called *disclosure attack*, that re-identifies all hidden peer partners of the targeted user. Thus the attack determines a protection limit of anonymity techniques, that is, it provides an upper bound on the number of observations required for an attacker to 'break' a given anonymity technique. In [15], the authors applied disclosure attack on the popular MIX technique and presented simulation results to show protection limits of the MIXes for a typical set of system parameters.

The goal of this paper is to present a comprehensive analysis of the disclosure attack and compute system parameters for which the anonymity of systems like the MIXes can be compromised by a few acts of communication by the targeted user. To that end, we will follow [15] as required, and then present our new results on the disclosure attack. The paper is organized as follows. In the next section, we will provide basic terminology and overview the formal model of the MIXes and the disclosure attack as described in [15]. In Section 3, contributions of this paper are highlighted. In Section 4, we present a proof that the underlying problem of the disclosure attack is NP-hard. After this, an analysis of the disclosure attack is given in Section 5. Next, in Section 7, we define 'weak operational region' of an anonymity system and provide an analytical formula for the boundary of the weak operational region induced by the disclosure attack. In Section 8, we discuss the impact of assumptions made during our analysis, and finally, in Section 9, we conclude with a summary of this paper and ideas for the future research work.

2. Background

The basic mechanism to provide anonymity is to organize additional traffic and conceal specific communication relationships amidst the additional traffic. In particular, the sender and/or receiver of a message can be embedded in a so-called *anonymity set* [16]. The size of the anonymity set can be used as a measure of anonymity provided by a technique.

Definition 1 *Assume an attacker model E and a finite set of all users Ψ. Let R be a role for the user (sender or recipient) with respect to a message M. If, for an attacker according to model E, the a-posteriori probability p that a user $\mathfrak{u} \in \Psi$ has the role R with respect to M is non-zero ($p > 0$), then $\mathfrak{u}$ is an element of the anonymity set $\mathfrak{A} \subseteq \Psi$. A technique (method) provides an anonymity set of size n if the cardinality of $\mathfrak{A}$ is n ($n \in \mathbb{N}$).*

Thus a sender or a receiver is anonymous only within their anonymity sets. In open environments, the anonymity set of a sender or a receiver would change with time. Since the intersection of two different anonymity sets is likely to be smaller than either of the anonymity sets, different intersections of anonymity sets could be used to gain information about the targeted user. Effectively, this leads to an anonymity set whose size shrinks as the attacker observes additional acts of communication by the targeted user[1]. The disclosure attack proposed by Kesdogan, Agrawal and Penz can re-identify all hidden peer partners of a targeted user deterministically after observing sufficient acts of communication by the user[15].

The disclosure attack is powerful in that it assumes a formal model of the MIXes and is independent of specific implementation weakness often exploited by other attacks. In the following, we overview the formal model of the MIXes presented in [15] and the corresponding disclosure attack.

2.1. The Formal Model of the MIXes

A MIX is a special intermediary network station which provides untraceability between the sender and recipient of a massage. A MIX collects b messages of equal length from b distinct senders, discards repeats, changes their appearance(i.e., the bit pattern), and forwards the messages to the recipients in a different order [7]. This measure hides the relationship between the sender and recipient of a message from everybody but the MIX and the sender of the message. By using more than one MIXes to forward a message from the sender to the recipient, the relation is hidden from all attackers in the network who do not control all MIXes through which the message passed, or who do not have the cooperation of all the other sender [7].

Clearly, MIXes should be carefully designed to implement the procedures for discarding repeat messages, changing message appearance, and reordering messages in a batch. The disclosure attack assumes a formal model for the MIXes which is inherently secure except for the unsynchronized users. Specifically, the random communication model proposed in [15] makes the following assumptions:

- The untraceability providing system, i.e. a MIX, provides perfect untraceability between incoming and outgoing packets.
- There are N users in the system. The batch size of the system is b, where $1 < b \ll N$, and a batch may contain a receiver more than once. Thus, the size n of the anonymity set fulfills the condition $n \leq b$ (see also Definition 1).
- The b packets in a batch are created by b different senders.
- Alice is one of the senders and she uses the system to hide her m, $1 \leqslant m \ll N$, communication partners.
- Alice chooses her communication partner in each communication uniformly among her m partners, while the other senders choose their communication partners uniformly among all N recipients.
- The attacker E takes notice of each untraceable communication act of Alice. This triggers the attacker to write down all recipients who are involved in this untraceable communication process, that is, the attacker simply records only those *recipient sets* which include a communication partner of Alice.

  For the sake of simplicity, we will enumerate the time t with increasing integer numbers whenever Alice sends a message. Thus, when Alice communicates for the first time, $t = 1$, when she communicates for the second time, $t = 2$, and so on. We will denote the recipient set at time t by $R_t = \{R_t^1, \ldots, R_t^n\}$.

The formal model contains three essential parameters, namely user population N, batch size b, and the number m of communication partners of the intended target. These parameters can be easily identified in other anonymity providing techniques. For a detailed discussion of the assumptions involved in the formal model, we refer the reader to [15].

2.2. The Disclosure Attack

In the disclosure attack on the MIXes, a potential attacker proceeded in two stages: *the learning phase* and *the excluding phase*. In the learning phase the attacker waits until he observes m mutually disjoint recipient sets $(R_1, \ldots, R_m)$, i.e., for all $i \neq j$, $R_i \cap R_j = \varnothing$. After the learning phase, the attacker can be sure that in each set R_i, there is only one peer communication partner of Alice. In the excluding phase of the attack, the recipient sets $(R_1, \ldots, R_m)$ are refined using further observations. This can be done by using a new recipient set R which intersects with only one prior recipient set, that is, if $R \cap R_i \neq \varnothing$ and $R \cap R_j = \varnothing$ for all $j \neq i$. In that case, R_i can be

[1] In other anonymity evaluations [19, 22, 23, 12], a weak attacker model is assumed, i.e., the attacker is not global. The disclosure attack investigates the pure hiding functionality of MIXes: the attacker is global and only one MIX is not corrupt.

refined to $R_i \cap R$. The refinement process is continued until each of the sets $R_1, \dots, R_m$ contains only one user. It is clear that the remaining m users in $R_1, \dots, R_m$ are the communication partners of Alice.

Since the senders are not coordinated, the probability that one of the phases does not find the needed batches (i.e. m disjoint batches or batches that overlap with only one of these m disjoint batches) converges to zero as the number of observations grows. Hence, we can deduce that with probability one the attack succeeds after a finite number of observations. Let T_l and T_e respectively be the average number of observations needed by the disclosure attack to complete the learning and excluding phases. The total average number of observations $(T_l + T_e)$ required by the attacker can be considered as the protection limit of the MIXes.

3. Our Contribution

In this paper, we present an extensive analysis of the disclosure attack which includes:

- a proof that disclosure attack solves an NP-complete problem and consequently, simulations of the disclosure attack require significant computing resources.

- analytical estimates of the number of observations that an adversary needs in the learning and excluding phases of the disclosure attack.

For the learning phase of the attack our estimate is quite tight, effectively eliminating the need for time and resource consuming simulations. For the excluding phase, our estimate provides an upper bound on the required number of observations.

We develop analytical formulas which explain the shape of performance curves[2] obtained by simulations in [15]. These formulas are used to relate the sharp rise in the number of observations in the performance curves to the non-peer recipients in the system.

We define the concept of weak operational regions of an anonymity system. In weak operational regions, the number of observations required by an attack becomes independent of the system parameters. We present analytical formula which provides the boundary of the weak operation region induced by disclosure attack. Finally, we extend simulation results presented in [15] by providing performance curves for two typical parameter settings of the MIXes. In each case, we show that our analysis effectively predicts the boundary of the weak operational region.

4. NP-Completeness of the Disclosure Attack

Fortunately, the disclosure attack is an NP-complete problem. Specifically, finding m mutually disjoint sets among t given sets with b elements each is an NP-complete problem. The proof consists of two parts. First, we show that this problem is in NP, i.e., in the class of all problems whose solutions can be verified in polynomial time. Second, we reduce another NP-complete problem, the CLIQUE problem, to our problem.

The first part of the proof is easy to see. A solution consists of m sets that have to be checked pairwise for disjointness. There are $O(m^2)$ pairs of sets, and for each pair $O(b^2)$ elements must be checked for equality in order to find common elements. Altogether, the verification of a solution is possible in polynomial time and therefore the problem is in NP.

To complete the proof, we have to show that another NP-complete problem can be reduced to our problem in polynomial time. We chose the CLIQUE problem as its structure is very similar to our problem. A CLIQUE problem consists of an undirected Graph $G = (V, E)$ and a natural number k. The goal is to find a k-clique inside the graph, that is a set of k vertices that are fully connected by the edges E of the graph.

We assume $V = \{v_1, v_2, \dots, v_t\}$ for any natural number t. Let $\{i, j\}$ denote the edge joining v_i and v_j. For each vertex $v_i \in V$, construct a set A_i which consists of all edge representations $\{i, j\}$ that are not included in E:

$$\forall 1 \le i < j \le t: \quad A_i = \{\{i, j\} | i \neq j \wedge \{v_i, v_j\} \notin E\}$$

Obviously, A_i can easily be constructed by checking for every other vertex v_j, if $\{v_i, v_j\} \in E$. Only if this check fails, $\{i, j\}$ is included into A_i. Clearly, this procedure can be performed in polynomial time.

According to this procedure, two sets A_i and A_j can only have one common element, that is $\{i, j\}$. This element is included in both sets if and only if $\{v_i, v_j\} \notin E$. Hence, A_i and A_j are disjoint if and only if v_i and v_j are connected. If we find k mutually disjoint sets, the corresponding vertices form a k-clique. Thus, the problems are equivalent and the construction is a valid problem reduction.

Figure 1 shows an example reduction. We see a graph with six vertices $v_1, \dots, v_6$, which are transformed to the corresponding sets $A_1, \dots, A_6$ on the right. The graph contains a 4-clique $\{v_2, v_3, v_5, v_6\}$. Hence, the sets A_2, A_3, A_5 and A_6 are mutually disjoint.

Altogether, the problem of finding m mutually disjoint sets turns out to be NP-complete. Therefore, simulating the disclosure attack becomes computationally infeasible when the parameters m and b are large. The goal of this paper is to derive analytical results which permit an analysis of the disclosure attack without simulations.

[2] performance curves plot the number of observation required to break anonymity versus a system parameter.

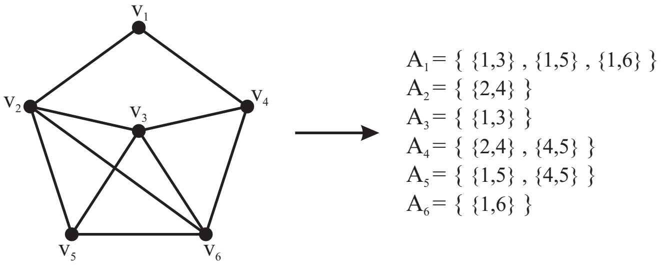

Figure 1. Example of the reduction procedure

5. Analytical Estimates

In the following subsections, we will develop estimates for T_l and T_e, the average number of observations needed respectively for the learning and excluding phases. In developing these estimates, we often encounter a situation where analytical calculations can be made easier by making some simplifying assumptions or by making an approximation. In order to maintain the flow of main ideas, we will postpone the discussion of these assumptions and approximations to the Section 8.

5.1. An Estimate for the Learning Phase

In the learning phase, an attacker wants to collect m mutually exclusive recipient sets in which Alice participates. An attacker with unlimited resources could follow the following brute-force strategy: after observing a new recipient set in which Alice participates, construct all possible collections of mutually exclusive recipient sets. The attacker would complete the learning phase when one of these collections contains m recipient sets.

More formally, let $\mathcal{G}^k$ be the set of all possible collections of mutually exclusive recipient sets after observing k batches in which Alice and only one peer communication partner of Alice participate. Let $G^k(i)$, $i = 1, 2, \ldots, m$ and $k = 1, 2, 3, \ldots$, be the number of collections in $\mathcal{G}^k$ with i recipient sets. After observing the $(k+1)$-th batch, the attacker would update $\mathcal{G}^k$ to $\mathcal{G}^{(k+1)}$ as follows: the attacker would check if the recipient set of the $(k+1)$-th batch is compatible with any of the collections in $\mathcal{G}^k$, that is, the attacker will check if the recipient set is mutually exclusive with all recipient sets in a collection C, $C \in \mathcal{G}^k$. If the $(k+1)$-th recipient set is compatible with the collection C, then the recipient set is included in the collection C and the cardinality of the collection C goes up by one. The original as well as the updated collection C is then copied over to $\mathcal{G}^{(k+1)}$. If the collection C is not compatible, then it is not updated and is copied over to $\mathcal{G}^{(k+1)}$. Note that $\mathcal{G}^k \subset \mathcal{G}^{(k+1)}$.

Clearly, we are interested in the average value of the smallest k for which the number of collections with m recipient sets is at least one, that is, $G^k(m) \geqslant 1$. Let $\pi(i)$ is the probability of finding a new recipient set (in which Alice and only one peer communication partner of Alice participate) compatible with a collection of $i-1$ mutually exclusive recipient sets. Since each recipient set in the collection contains exactly one peer communication partner of Alice, there are only $(m-i+1)$ choices for the peer partner of Alice for the i-th mutually exclusive set. For the other $(b-1)$ recipients, neither a partner of Alice nor the other non-peer recipients in already collected $(i-1)$ sets can be used. This results in *at least* $N-m-(i-1)(b-1)$ choices for the other $b-1$ recipients. There could be more choices for the non-peer $(b-1)$ recipients since a recipient set in the collection may contain a non-peer recipient more than once. However, in order to derive our estimate we will assume that we have *exactly* $N-m-(i-1)(b-1)$ choices for the other $(b-1)$ recipients. Later in Section 8, we will discuss the effect of this assumption on our calculations.

By simple combinatorics, there are total $(N-m-(i-1)(b-1))^{(b-1)}(m-i+1)$ compatible and equally likely choices for the i-th mutually exclusive recipient set. Since there are a total of $(N-m)^{(b-1)}m$ equally likely recipient sets with only one peer partner of Alice, $\pi(i)$, the probability that a new recipient set would be compatible with a collection of $i-1$ mutually exclusive recipient sets is given by

$$\pi(i) = \frac{(N-m-(i-1)(b-1))^{(b-1)}(m-i+1)}{(N-m)^{(b-1)}m} \quad (1)$$

A collection with i recipient sets in $\mathcal{G}^{(k+1)}$ either belongs to $\mathcal{G}^k$ or is a result of updating a collection with $(i-1)$ recipient sets in $\mathcal{G}^k$. Consequently, the expected number of collections with i recipient sets in $\mathcal{G}^{(k+1)}$ is given by:

$$E[G^{(k+1)}(i)|G^k(i-1), G^k(i)] = G^k(i-1)\pi(i) + G^k(i) \quad \text{for } i = 2, \ldots, m,$$
$$G^{(k+1)}(1) = G^k(1) + 1 \quad (2)$$

where $E[\cdot]$ denotes the expected value. Note that (2) implies that the probability of $(k+1)$-th recipient set being identically equal to one of the earlier recipient set is zero. In Section 8, we will show that this probability is indeed negligible and can be assumed to be zero without much loss in the accuracy of our results.

The above equations suggest that a heuristic estimate of the number of observations required for the learning phase can be obtained by replacing random variables in these equations by their expected values:

$$\begin{aligned}
&E\Big[G^{(k+1)}(i)|E[G^k(i-1)],E[G^k(i)]\Big]\\
&\quad = E\Big[G^k(i-1)|E[G^{k-1}(i-2)],E[G^{k-1}(i-1)]\Big]\pi(i)\\
&\qquad + E\Big[G^k(i)|E[G^{k-1}(i-1)],E[G^{k-1}(i)]\Big]\\
&\qquad\qquad\qquad\qquad\text{for } i=2,\ldots,m,\\
&E[G^{(k+1)}(1)] = E[G^k(1)]+1
\end{aligned} \tag{3}$$

Recall that the learning phase stops when there is a collection with m mutually exclusive sets. Using (3), we can recursively calculate the value of k for which

$$\operatorname*{argmin}_k E\Big[G^k(m)|E[G^{(k-1)}(m-1)],E[G^{(k-1)}(m)]\Big] \geqslant 1$$

Let $\mathcal{K}$ denote this value of k. We treat $\mathcal{K}$ as an estimate of the average value of smallest k for which $G^k(m) \geqslant 1$.

The above estimate only considers the batches in which Alice and only one of the peer partners of Alice participate. Since there is a small probability that a sender other than Alice would send a message to a peer communication partner of Alice, the total number of recipient sets required for the learning phase is given by:

$$\frac{\mathcal{K}}{\text{Prob(One peer partner participates)}} \tag{4}$$

where

$$\text{Prob(One peer partner participates)} = \left(\frac{N-m}{N}\right)^{b-1} \tag{5}$$

The estimate derived above can be improved slightly for certain values of N, b, and m by realizing that the number of required observations cannot be less than the number of observations required to see all m peer communication partners of Alice. This correction results in the following improved estimate:

$$T_l \approx \max\left[\frac{\mathcal{K}}{\left(\frac{N-m}{N}\right)^{b-1}}, \sum_{i=1}^{m}\frac{m}{m-i+1}\right] \tag{6}$$

5.2. An Upper Bound on the Learning Phase

In this section, we will develop an upper bound on the number of observations needed at the learning stage. Imagine a genie that knows peer communication partners of Alice. The genie observes batches in which Alice participates and keeps the recipient set of a observation if it contains only one partner of Alice and if it is mutually disjoint from all the other recipient sets the genie has kept so far. The genie discards the recipient set if it either contains more than one partner of Alice or intersects with already kept recipient sets.

It is clear that the number of observations required by the genie to collect m mutually disjoint recipient sets is an upper bound on the number of observations required by an attacker. An unbounded attacker would form all possible combinations of m recipient sets from all the recipient sets in which Alice participates, and thereby find the mutually exclusive set found by the genie. Since an unbounded attacker does not discard recipient sets, it is likely that the attacker would find the required m mutually exclusive sets earlier than the genie.

Assume that the genie has already collected $(i-1)$ mutually exclusive recipient sets and it is looking for the i-th mutually exclusive recipient set. The average number of observations required by the genie to find the i–th mutually exclusive set is given by

$$\begin{aligned}
T_l^u(i) &= \sum_{j=1}^{\infty} j\Big(1-\pi(i)\Big)^{j-1}\pi(i)\\
&= \frac{1}{\pi(i)},
\end{aligned} \tag{7}$$

where $\pi(i)$ is the probability of finding the i-th mutually exclusive set given by (1). It follows that an upper bound on the average number of observations in the learning phase is given by

$$\begin{aligned}
T_l^u &= \sum_{i=1}^{m}\frac{1}{\pi(i)}\\
&= \sum_{i=1}^{m}\frac{(N-m)^{b-1}}{\Big(N-m-(i-1)(b-1)\Big)^{b-1}}\cdot\frac{m}{m-i+1}
\end{aligned} \tag{8}$$

We will show later that the bound given by (8) can be used for asymptotic analysis to explain the characteristics of the performance curves shown in [15]. In particular, it can explain the sharp rise in the number of observation required for the learning phase once the system parameters cross certain thresholds.

5.3. An Estimate for The Excluding Phase

Recall that in the excluding phase, the attacker observes new batches and checks if its recipient set R_t intersects with only one of the m recipient sets $O_l, 1 \leqslant l \leqslant m$, collected during the learning phase. If the recipient set R_t intersects with more than one set then it is kept in a pool $\mathcal{P}$ of the recipient sets that may be useful in the future. On the other hand, if the set R_t intersects with only one recipient set O_l, then O_l is replaced by the intersection of R_t and O_l and the set R_t is discarded. Each time after a set $O_l, 1 \leqslant l \leqslant m$ is replaced with a smaller set, all recipient sets from the pool $\mathcal{P}$ are checked again to see if any of them intersects with only one set from $O_1, \ldots, O_m$. This process is repeated until $O_1, \ldots, O_m$ each has only one recipient.

In order to obtain an upper bound on the number of observations required for the excluding phase, we will consider a storage limited attacker who does not maintain the pool $\mathcal{P}$ of recipient sets. Such an attacker would necessarily require more observations than the attacker who maintains the pool $\mathcal{P}$, and thereby provide an upper bound on the attacker with unlimited resources.

We can model the excluding phase of the limited attacker as a discrete stochastic process with state $X = (O_1, O_2, \ldots, O_m)$. Let X_t denote the state of the process after observing the t-th batch. Since the state X_t only depends on the previous state X_{t-1} and the t-th batch, X_t is a Markov process.

If we assume that the non-peer recipients in a new batch occur uniformly over all N users, then the process Y given by $Y = (|O_1|, |O_2|, \ldots, |O_m|)$ also becomes a Markov process. This is because the recipient sets $O_1, O_2, \ldots, O_m$ are mutually exclusive at the end of the learning phase and they remain mutually exclusive during the excluding phase. The mutual exclusivity of $O_1, O_2, \ldots, O_m$ and uniformly occurring non-peer recipients in a new batch render the individual identities of the recipient unimportant for determining the sizes of $O_1, O_2, \ldots, O_m$. In other words, the mutual exclusivity of $O_1, O_2, \ldots, O_m$ and uniformly occurring non-peer recipients in a batch imply that the number of recipients in the set O_l, $1 \leqslant l \leqslant m$, after observing the t-th batch depends only on the number of recipients in these sets before observing the t-th batch. Since we are only interested in the average number of observation required to reach the state $Y = (1, 1, \ldots, 1)$, it suffices to analyze the Markov process Y.

In theory, the recognition of Y as a Markov process enables us to obtain an upper bound on the average number of observations required for the excluding phase. However, this statistics is hard to compute due to the large size, given by b^m, of the state-space of Y. The state space of Y is large even for moderate values of b and m, for example, $b = 25$ and $m = 10$.

To efficiently compute the required upper bound, we just look at the number of recipients in the first recipient set, $Y_1 = |O_1|$, and stipulate that the total number of recipients in other recipient sets is given by γ. Under this stipulation, the transition probability of Y_1 can be computed easily. In particular, $P(Y_{1t} = r | Y_{1(t-1)} = r)$ is the sum of the probabilities of the following three disjoint events:

1. R_t does not contain the peer communication partner of Alice present in O_1,
2. R_t contains the peer communication partner of Alice present in O_1, but it intersects with γ non-peer recipients in $O_2, \ldots, O_m$, and
3. R_t contains peer communication partner of Alice present in O_1, it does not intersect with γ non-peer recipients in $O_2, \ldots, O_m$, but it contains all $r-1$ non-peer partners in O_1.

In each of these cases, the size of O_1 will remain the same. The transition probability, $P(Y_{1t} = c | Y_{1(t-1)} = r), c < r$ is given by the probability of the following event: R_t contains the peer partner in O_1, it does not intersect with $O_2, \ldots, O_m$, and it intersects with O_1 in $c-1$ non-peer recipients. It is clear that the state transition probability of Y is given by

$$P(Y_{1t} = c | Y_{1(t-1)} = r) = \begin{cases} \frac{(m-1)}{m} + \frac{1}{m}\Psi(\gamma) + \\ \qquad \frac{1}{m}\Big(1 - \Psi(\gamma)\Big)\Omega(c-1, r-1) & c = r \\ \frac{1}{m}\Big(1 - \Psi(\gamma)\Big)\Omega(c-1, r-1) & c < r \\ 0 & c > r \end{cases} \tag{9}$$

where $\Psi(i)$ is the probability that a recipient set contains any of the specified i recipients as its non-peer recipients, and $\Omega(\beta, \alpha)$ is the probability that a recipient set intersects a set of α non-peer recipients in exactly β recipients, given that it does not intersect with γ other given non-peer recipients. It is easy to check that that $\Psi(i)$ is given by

$$\Psi(i) = 1 - \Big(\frac{N-i}{N}\Big)^{(b-1)} \tag{10}$$

Computing $\Omega(\beta, \alpha)$ turns out to be tricky. There are $\binom{\alpha}{\beta}$ different sets of β non-peer recipients in a set with α non-peer recipients. Let G denote a set of β non-peer recipients. A recipient set which contains G can have members of G in the places of s recipients, $s = \beta, \ldots, b-1$, which can be chosen in $\binom{b-1}{s}$ ways. The number of ways members of G can appear for s given recipients is given by $\beta! \mathcal{S}_s^{(\beta)}$, where $\mathcal{S}_s^{(\beta)}$ is the Sterling number of second kind and equals to the number of ways of partitioning a set of s elements into β non-empty subsets. The factor of $\beta!$ takes into account the fact that these β subsets are further assigned β different identities of non-peer communication partners. For each of the other $b-s-1$ members of the recipient set, there are $N-\alpha-\gamma$ distinct choices. Putting this all together, we get the following formula for $\Omega(\beta, \alpha)$:

$$\Omega(\beta, \alpha) = \frac{\beta! \binom{\alpha}{\beta} \sum_{s=\beta}^{b-1} \binom{b-1}{s} \mathcal{S}_s^{(\beta)} (N-\alpha-\gamma)^{b-s-1}}{(N-\gamma)^{b-1}} \tag{11}$$

Using (9), (10), and (11), we can compute the transition probability matrix $M_{b \times b}$ of the Markov process Y_1, where $M_{ij} = P(Y_{1t} = j | Y_{1(t-1)} = i)$, for $1 \leqslant i, j \leqslant b$.

By the properties of Markov processes, the probability of state $|O_1| = 1$ after observing the k-th batch is given by $e_1(M^T)^k \Pi$, where $e_1 = [1, 0, 0, \ldots, 0]$, $\Pi = [\Pi_1, \ldots, \Pi_b]$ and Π_i is the probability of $|O_1| = i$ at the end of the learning phase. We would discuss how to calculate the vector of initial probabilities Π in the Section 8.

It follows that with probability $e_1[(M^T)^k - (M^T)^{k-1}]\Pi$, the excluding phase ends after observing k batches. Therefore an upper bound on the average number of batches required by the attacker for the excluding phase is given by:

$$T_e^u = \sum_{k=1}^{\infty} k \cdot e_1[(M^T)^k - (M^T)^{k-1}]\Pi \qquad (12)$$

It is tempting to hope that the above expression can be quickly computed by using geometric series summation formula to compute the sums $S_1 = \sum_{k=1}^{\infty} k \cdot e_1[(M^T)^k]\Pi$ and $S_2 = \sum_{k=1}^{\infty} k \cdot e_1[(M^T)^{k-1}]\Pi$, and then by putting $T_e^u = S_1 - S_2$. A careful investigation shows that the state transition matrix M always has an eigenvalue 1, and therefore neither of above geometric series converge, and $S_1 = S_2 = \infty$. Thus T_e^u should be calculated by computing partial sums $\sum_{k=1}^{K} k \cdot e_1[(M^T)^k - (M^T)^{k-1}]\Pi$ for a sufficiently large value of K which meets the desired accuracy goal.

6. Simulation Results

In this section, we will compare our analytical estimates of the number of observations required for the learning and excluding phases with the simulation results. We will consider three typical cases: (a) N = 20000, b = 50, m=20, (b) N = 400, b = 10, m = 10, (c) N = 200,000, b = 100, m = 40. Simulation result for the case (a) were presented in [15], while the simulation results for two other cases are new. Case (a) represents a typical case, while cases (b) and (c) may represent extremes of an anonymity providing system working in an open environment.

6.1. The Learning Phase

For case (a), Figure 2 shows the number of observations required at the learning stage as a function of m, b, and N as the two other parameters are kept fixed. For all three parameters, the analytical estimate fits closely with the actual number of observations required for the learning stage. Note that the analytical estimate is slightly less than the actual value obtained by simulation, with the difference becoming the most prominent around the knee of the curves. It turns out that the knee of these curve coincides with the change of term in (6). For the flat part of the curve, it is the second term which provides the number of observations, while for the rising part the first term provides the number of observations. As the number of observations shifts from the first term to the second, our estimate becomes less accurate.

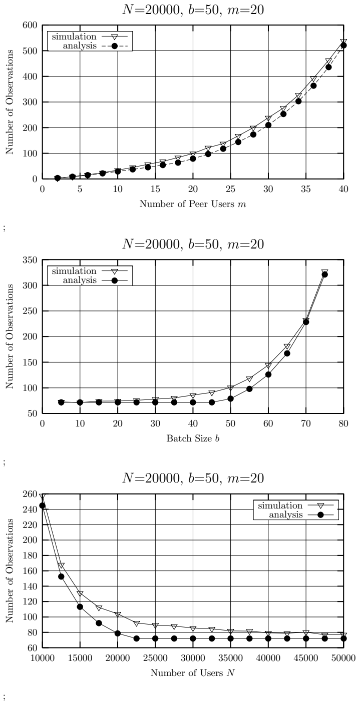

Figure 2. Number of observation required for the learning phase for Case (a)

Figure 3 shows the number of observations required for the learning stage for case (b). Once again, for all three parameters, analytical estimates for learning phase is quite close to the simulation results. Finally, Figure 4 shows the number of observations required for the case (c). In this case too, our estimates are quite close to the simulation results.

The three cases considered above span a wide range of anonymity system parameters. The close match of analytical results to the simulation results for the learning phase in these cases gives us confidence that one can rely on the analytical estimate presented here and avoid the computationally intensive simulations.

6.2. The Excluding Phase

For the excluding phase, Figures 5–7 compare the number of observations obtained by simulations to that of obtained by the analysis given in Section 5. These figures

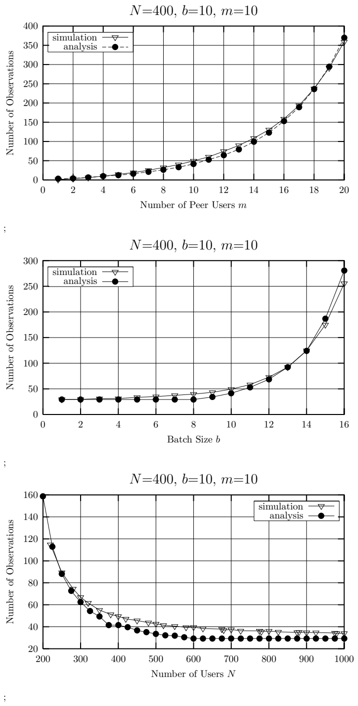

Figure 3. Number of observation required for the learning phase for Case (b)

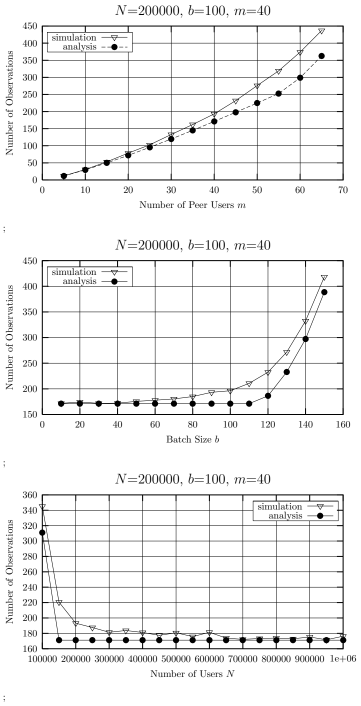

Figure 4. Number of observation required for the learning phase for Case (c)

show that once the number of observations start their sharp rise, our estimate is an overestimate. This is expected since in estimating the number of observations needed for the excluding phase, we did not maintain the pool of unused observations for the future use. However, our estimates are still within one order of magnitude of the numbers obtained by the simulations, and can be used to compute a protection limit for the anonymity systems.

7. Operational Regions of Anonymity Systems

Figures 2–7 show that for both phases of the disclosure attack, required number of observations rises sharply as the parameters m and b increase above a threshold and the parameter N decreases below a threshold. Before this sharp rise, the number of observations required at both stages is too low—often less than 50—to provide adequate anonymity. Furthermore, before the sharp rise, the number of required observations remain almost constant as a function of N and b. Clearly, the insensitivity of the required number of observations on the system parameters is undesirable as it takes away control dimensions from the operators of anonymity systems by which they can control the anonymity of the system. These facts motivate us to define the weak operational region of an anonymity system as follows:

Definition 2 *Weak operational region of an anonymity system is a region of the parameter space where the number of observations required by an attack is independent of one or more system parameters N, m, and b.*

The independence from a system parameter in a certain region of the parameter space implies that as the parameter is varied, the number of observations change at most by $O(1)$. We caution that unless the attack is optimal, parameter tuples outside the weak operational region do not guarantee a high level of anonymity against an attack which may be more efficient than the attack used to derive the weak operational region. Therefore, avoiding weak operational region is necessary, but not sufficient for providing adequate anonymity.

In the rest of this section, we will analyze characteristics of the operational region rendered weak by the disclosure attack. The first step toward this analysis is to show that in a certain region of the parameter space, the number of observations at the learning stage become independent of the parameters N and b.

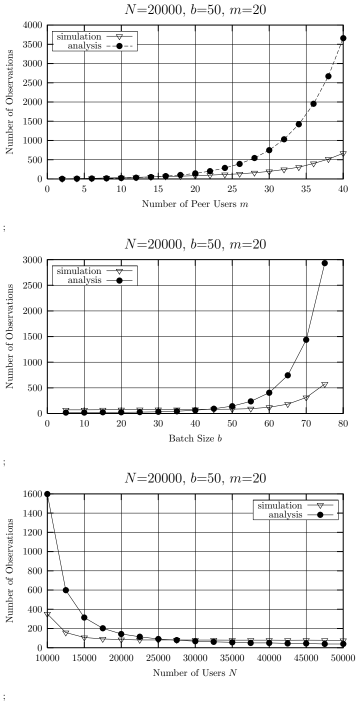

Figure 5. Number of observation required for the excluding phase for Case (a)

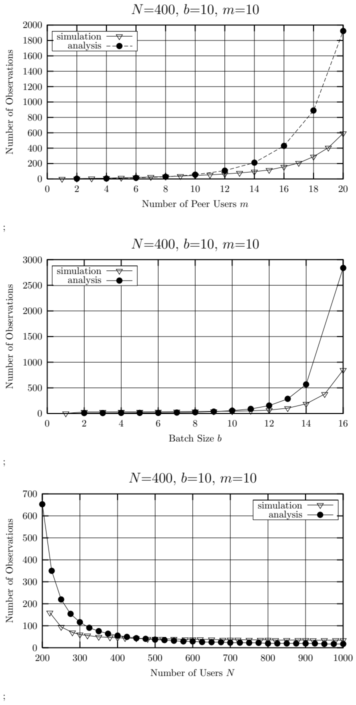

Figure 6. Number of observation required for the excluding phase for Case (b)

7.1. Asymptotic Analysis of the Learning Phase

Recall that a new recipient set observed by an attacker may not be compatible with a given collection of k mutually exclusive recipient sets if any of the following occurs: it contains a non-peer recipient that is already present in one of the k given recipient sets, the new recipient set has more than one peer communication partner of Alice, or it contains a peer communication partner of Alice that is already present in one of the k given recipient sets.

If the bottleneck in the learning phase is due to already observed non-peer recipients or due to more than one peer communication partners of Alice, then the dominant term in (8) becomes

$$\sum_{i=1}^{m} \frac{(N-m)^{(b-1)}}{(N-m-(i-1)(b-1))^{(b-1)}}. \tag{13}$$

where the already observed non-peer recipients decrease the denominator of the summand by $(i-1)(b-1)$ while the peer partners of Alice decrease it by m. Clearly, for $b > 10$, the largest contribution of non-peer recipients ($(m-1)(b-1)$) is far more than the contribution of peer partners (m) in reducing the denominator. Since for most anonymity systems $b > 10$, henceforth we will refer to this bottleneck as *Type N* bottleneck indicating that the non-peer recipients are the main cause of the bottleneck.

On the other hand, if the bottleneck is due to already observed peer communication partners of Alice, then the dominant term in (8) becomes

$$\sum_{i=1}^{m} \frac{m}{m-i+1}. \tag{14}$$

Henceforth, this bottleneck will be referred to as the *Type P* bottleneck indicating that the peer partners of Alice are the cause of the bottleneck.

Depending on the values of $N, b,$ and m, the dominant bottleneck in the learning phase could either be that of Type N or that of Type P. These two types of bottlenecks have significantly different impact on the required number of observations. Specifically,

$$\sum_{i=1}^{m} \frac{(N-m)^{(b-1)}}{(N-m-(i-1)(b-1))^{(b-1)}} = \mathrm{Poly}\Big(\frac{1}{c_m-m}, \frac{1}{N-c_N}\Big), \mathrm{Exp}(b) \tag{15}$$

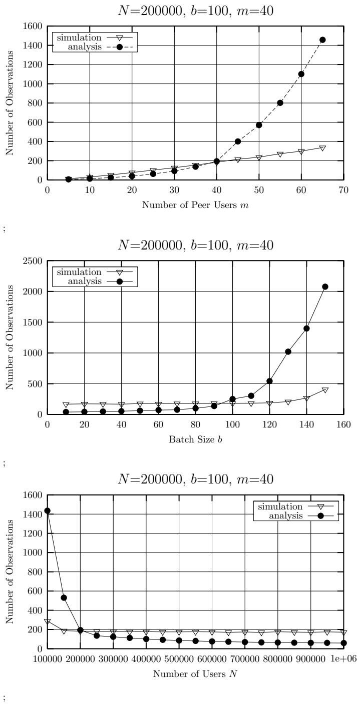

Figure 7. Number of observation required for the excluding phase for Case (c)

and

$$\sum_{i=1}^{m} \frac{m}{m-i+1} \sim O(m \log m) \tag{16}$$

where c_m, and c_N are constants which depend on two fixed parameters. Thus, while a Type N bottleneck causes the number of observations to increase at least polynomially in functions of b, m, and N, the Type P bottleneck causes the number of observations to increase only as $m \log m$. Furthermore, the number of observations under Type P bottleneck is largely independent of N or b.

We caution that the asymptotic behavior indicated by (15) and (16) is valid for the *upper bound* on the number of observations, not on the number of observations itself. However, these results help us intuitively understand the nature of performance curves in Figures 2–4. In particular, these figures show that before the sharp rise caused by non-peer recipients, the number of required observations remain almost constant as a function of N and b as suggested by (16). Once the bottleneck changes over to the Type N, the number of observations rise sharply as suggested by (15).

These figures also show that the transition from one type of bottleneck to another occurs in the same parameter range for both the learning and excluding phases. Since it is important that the anonymity provided by the system be sensitive to the parameters N and b, in the following we will present a formula for the transition region where the dominance of the bottlenecks changes for the learning phase.

parameters	N	m	b
N = 20000, m = 20, b = 50	15718	28	56
N = 400, m = 10, b = 10	369	11	10
N = 200000, m = 40, b = 100	105602	91	138

Table 1. Threshold values of N, m, and b for three typical cases

7.2. Threshold of Transition Region

A threshold for transition from Type N bottleneck to Type P bottleneck can be obtained by comparing the terms responsible for these bottlenecks. Specifically, we use the difference

$$D(i) = \frac{(N-m)^{(b-1)}}{(N-m-(i-1)(b-1))^{(b-1)}} - \frac{m}{m-i+1}$$

to compare the two terms. Since for both bottlenecks the most number of observations are required for finding the m-th compatible recipient set, the boundary of the transition region can be obtained by setting $D(m) = 0$,

$$\frac{(N-m)^{(b-1)}}{(N-m-(m-1)(b-1))^{(b-1)}} = m \tag{17}$$

By fixing two parameters in (17), we can compute the threshold value of the third parameter when the learning phase goes from one bottleneck to another. The Table 1 shows the threshold values for three typical cases considered in this paper. A comparison of these thresholds with the performance curves in Figures 2–4 shows that (17) precisely estimates the boundary of transition region for all nine performance curves. Furthermore, a comparison of these thresholds with the knee of performance curves in Figures 5–7 shows that they are also an excellent predictor of the transition region in the excluding phase.

8. Issues in Computing Analytical Estimates

In this section, we will address certain issues that arise in computing analytical estimates. We start by addressing the approximation of distinct non-peer recipients in a recipient set made for deriving the analytical estimate for the learning phase.

Recall that we assumed that all non-peer recipients in a recipient set were distinct or in other words, we assumed that the number of distinct non-peer communication partners in a recipient set is $b-1$. This approximation allows us to set the number of non-peer recipients in $i-1$ mutually exclusive recipient sets to be $(i-1)(b-1)$. In practice, there is a small probability that a non-peer partner would be the intended recipient for more than one message, making the number of distinct non-peer recipients in a recipient set to be l, where $l = 1, \ldots, b-1$. These l non-peer recipients can be chosen from the pool of $N-m$ non-peer partners in $\binom{N-m}{l}$ ways. Furthermore, a particular group of l non-peer recipients can be put in $b-1$ places in $l!\mathcal{S}_{b-1}^{(l)}$ ways, where $\mathcal{S}_{b-1}^{(l)}$ is the Sterling number of second kind and equals to the number of ways of partitioning a set of $(b-1)$ elements into l non-empty subsets. The factor of $l!$ takes into account the fact that these l subsets are further assigned l different identities of non-peer communication partners. It now follows that the probability of obtaining l distinct non-peer communication partners in a recipient set is given by

$$\frac{l!\binom{N-m}{l}\mathcal{S}_{b-1}^{(l)}}{(N-m)^{b-1}}$$

In an open environment, we expect that the chances of a non-peer communication partner occurring more than once in a batch would be very small, and a quick calculation for typical sets of parameters by using the above formula shows that indeed this is true, and assuming $b-1$ non-peer recipients in a batch results in very little loss of accuracy. If desired, the analytical estimates can be made more precise by substituting $E[\text{distinct non-peer partners}]$, the expected number of distinct non-peer partners, in place of $b-1$:

$$E[\text{distinct non-peer partners}] = \sum_{l=1}^{b-1} l \frac{l!\binom{N-m}{l}\mathcal{S}_{b-1}^{(l)}}{(N-m)^{b-1}}$$

Another assumption used in deriving the analytical estimate for the learning phase was that the probability of seeing the same recipient set more than once is negligible. Given a recipient set, the probability of a new recipient set being the same as the given set is

$$\frac{1}{m}(b-1)!\frac{1}{N^{(b-1)}}$$

For $b \ll N$ this probability is negligible: it evaluates to 5.4×10^{-150}, 1.38×10^{-19}, and 3.7×10^{-371} for the cases (a), (b), and (c), respectively.

Recall that in deriving the analytical estimate for the excluding phase we made two critical assumption: First, we assumed that the attacker does not maintain the pool of unused recipient sets for the future use. This attacker would necessarily require more observation than the unbounded attacker, and therefore our estimate produces an upper bound on the required number of observations. Second, in order to make our calculations tractable, we decoupled the state vector Y by assuming that the number of non-peer recipients left in $O_2, \ldots, O_m$ is given by γ, and proceeded to calculate the number of recipients left in O_1. In our calculations, we assumed that on the average other recipients sets have the same number of recipient as O_1 and put $\gamma = (m-1)(|O_1|-1)$.

It turns out that our upper bound for the excluding phase is quite loose. For the simulated cases discussed in this paper, our upper bound is up to six times higher than the values obtained by simulations. The main source of high estimate is our assumption that the attacker does not keep the pool of unused recipient sets for the future use. We are currently working on a solution which removes this restriction and promises a much tighter upper bound.

At present, we have not calculated the impact assuming $\gamma = (m-1)(|O_1|-1)$ since the impact is likely to be much smaller than the impact of not maintain the pool of unused recipient sets for the future use.

9. Conclusions

In this paper, we presented a comprehensive analysis of the disclosure attack on anonymity providing systems. We showed that the disclosure attack is NP-complete which causes the simulations of the attack to be significantly expensive in computing resources. As an alternative, we derived analytical estimates for both the learning and excluding phases of the disclosure attack. These estimates were compared against the simulation results for three different sets of system parameters. For the learning phase, our estimate is tight and can serve as a substitute for simulation results. For the excluding phase, our estimates provide an upper bound which can be used as a protection limit of anonymity systems.

We defined *weak operational region* of anonymity providing systems as the parameter space where anonymity provided by the system becomes independent of one of the system parameters. Simulation results show that disclosure attack renders a large parameter space as weak operational region. We presented an asymptotic characterization of the performance curves and showed that indeed the number of observations required at the learning stage become independent of the number of users and the batch size for certain values of system parameters.

An analytical formula was developed to estimate the boundary of weak operational regions. The values estimated by this formula were in a precise agreement with the simulation results, eliminating the need for expensive simulations to avoid the weak operational region.

The future work may proceed in several directions. To completely eliminate dependencies on the simulations, better estimates are needed for the number of observations at the excluding phase. Although the disclosure attack is quite powerful in its scope, it is still suboptimal, and a better deterministic attack may be found. Finally, a more refined notion of privacy needs to be developed in order to assess the erosion of privacy as an attacker observes progressively more acts of communication by the targeted user.

References

[1] M. Abe, "Universally verifiable mix-net with verification work independent of the number of mix-servers," *In Advances in Cryptology, Eurocrypt '98*, volume 1403 of Lecture Notes in Computer Science, pages 437–447, Helsinki, Finland, 31 May, 4. June 1998. Springer-Verlag.

[2] M. Abe, "Mix-network on permutation networks," *In Advances in cryptology ASIACRYPT'99*, volume 1716, pages 258-273. Springer-Verlag, 1999.

[3] O. Berthold, H. Federrath, and S. Köpsell, "Web MIXes: A System for Anonymous and Unobservable Internet Access," *International Workshop on Design Issues in Anonymity and Unobservability*, Berkley, 2009 LNCS, Springer-Verlag, 2001.

[4] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the A.C.M.,* 24(2):84-88, February 1981

[5] D. Chaum, "The dining cryptographers problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, 1:65-75, 1988.

[6] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan:" Private information retrieval," *In 36th IEEE Conference on the Foundations of Computer Science*, pages 41-50. IEEE Computer Society Press, 1995.

[7] D. A. Cooper and K. P. Birman: "Preserving privacy in a network of mobile computers," *In 1995 IEEE Symposium on Research in Security and Privacy,* pages 26-38. IEEE Computer Society Press, 1995.

[8] L. Cottrell: "Mixmaster," http://www.obscura.com/ loki/.

[9] H. Federrath, A. Jerichow, A. Pfitzmann: "MIXes in Mobile Communication Systems: Location Management with Privacy," *Information Hiding, LNCS 1174*, Springer-Verlag, Berlin 1996, 121-135.

[10] I. Goldberg and A. Shostack: "Freedom network whitepapers".

[11] C. Gulcu and G. Tsudik: "Mixing E-mail with BABEL," *In Symposium on Network and Distributed Systems Security (NDSS '96)*, San Diego, California, February 1996.

[12] Y. Guan, X. Fu, R. Bettati, and W. Zhao: "An Optimal Strategy for Anonymous Communication Protocols," *Proceedings of the 22th International Conference on Distributed Computing Systems,* Vienna, Austria, July 2002.

[13] A. Jerichow, J. Müller, A. Pfitzmann, B. Pfitzmann, M. Waidner: "Real-Time Mixes: A Bandwidth-Efficient Anonymity Protocol," *IEEE Journal on Selected Areas in Communications,* 1998.

[14] P. A. Karger: "Non-Discretionary Access Control for Decentralized Computing Systems", Master Thesis, Massachusetts Institute of Technology, Laboratory for Computer Science, 545 Technology Square, Cambridge, Massachusetts 02139, Mai 1977, Report MIT/LCS/TR-179.

[15] D. Kesdogan, D. Agrawal and S. Penz: "Limits of Anonymity in Open Environments," *IH 2002, 5th international workshop on information hiding,* Noordwijkerhout, The Netherlands, 7–9 October 2002. Lecture Notes in Computer Science (to be published), Springer-Verlag, 2002

[16] D. Kesdogan, J. Egner, and R. Büschkes, "Stop-and-go mixes providing probabilistic security in an open system," In David Aucsmith, editor, *Information Hiding: Second International Workshop*, volume 1525 of Lecture Notes in Computer Science, pages 83-98. Springer-Verlag, Berlin, Germany, 1998.

[17] A. Pfitzmann, B. Pfitzmann, and M. Waidner: ISDN-mixes: "Untraceable communication with very small bandwidth overhead," *In GI/ITG Conference: Communication in Distributed Systems,* pages 451-463. Springer-Verlag, Heidelberg 1991, February 1991.

[18] A. Pfitzmann and M. Waidner, "Networks without user observability, design options. In Advances in Cryptology," *Eurocrypt '85, volume 219 of Lecture Notes in Computer Science*, Spinger-Verlag, 1985.

[19] M. K. Reiter and A. D. Rubin: "Crowds: Anonymity for Web Transactions," *ACM Transactions on Information and System Security,* volume 1, pages 66-92, 1998.

[20] M. G. Reed, P F Syverson, and D M Goldschlag: "Anonymous connections and onion routing," *IEEE Journal on Special Areas in Communications,* 16(4):482–494, May 1998.

[21] M. G. Reed, P. F. Syverson, and D. M. Goldschlag: "Protocols using Anonymous Connections: Mobile Applications, Security Protocols," *5th International Workshop Proceedings,* B. Christianson, B. Crispo, M. Lomas, and M. Roe (editors), Springer-Verlag LNCS 1361, 13–23, 1998.

[22] C. Shields and B. N. Levine: "A Protocol for Anonymous Communication Over the Internet," *Proceedings of the 7th ACM Conference on Computer and Communication Security,* Athens, Greece, 1-4, November 2000.

[23] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr: "Towards an Analysis of Onion Routing Security," *Workshop on Design Issues in Anonymity and Unobservability,* Berkeley, CA, July 2000.