Entropy Bounds for Traffic Confirmation

Luke O'Connor*

October 21, 2008

Abstract

Consider an open MIX-based anonymity system with N participants and a batch size of b. Assume a global passive adversary who targets a given participant Alice with a set $\mathcal{R}_A$ of m communicating partners. Let $H(\mathcal{R}_A \mid \mathcal{B}_t)$ denote the entropy of $\mathcal{R}_A$ as calculated by the adversary given t message sets (MIX batches) where Alice is a sender in each message set. Our main result is to express the rate at which the anonymity of Alice (as measured by $\mathcal{R}_A$) degrades over time as a function of the main parameters N, b and m. Assuming $m, b < \sqrt{N}$, we prove that there is a threshold $t^* = O(m \ln N)$ such that when $t = t^* + 5cm$, for any integer $c > 0$, then

$$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0) \quad > \quad 1 - e^{-c} \tag{1}$$

where the probabilities are computed over the random communication model. Thus once the attacker has collected a threshold t^* of messages sets, each additional $5m$ message sets collected decreases $H(\mathcal{R}_A \mid \mathcal{B}_t)$ towards zero geometrically. We provide formulas for computing the implied constant in $t^* = O(m \ln N)$ and also to improve the constant 5 in $t = t^* + 5cm$.

When $N > O(b\, m \ln m)$, or in general for large N, the threshold can be improved to $t^* = O(m(\ln b + \ln m))$, which is independent of N. Further, if $b = O(m)$ then the threshold simplifies to $t^* = O(m \ln m)$ and this bound is asymptotically optimal since the coupon collector problem indicates that $t^* = \Omega(m \ln m)$.

*The author can be contacted via email at `lukejamesoconnor@gmail.com`.

1 Introduction

Traffic analysis is a collection of techniques for inferring communication relationships without having access to the content that is being communicated. *Traffic confirmation* is a special subcase that attempts to infer communication relationships amongst users whose communications are mediated through an anonymity system. We assume that that the underlying anonymity system is a threshold MIX, with batch size b, supporting N participants, each of whom may act as both a sender or receiver. We consider the class of traffic confirmation attacks where an attacker targets a specific user Alice and attempts to infer her communication partners from observing her communication patterns through the MIX. The attacker is assumed to be a global passive adversary, and is therefore able to determine the set of senders and receivers for each message set (batch) submitted to the MIX. Let $\mathcal{R}_A$ be the set of Alice's communicating partners where $|\mathcal{R}_A| = m$, and let $H(\mathcal{R}_A \mid \mathcal{B}_t)$ be the entropy of $\mathcal{R}_A$ after observing t rounds of the MIX. Intuition suggests that $H(\mathcal{R}_A \mid \mathcal{B}_t)$ will tend to zero as t increases since the attacker can ignore message sets where Alice is not a sender.

Our main results are to express the rate at which the anonymity of Alice (as measured by $\mathcal{R}_A$) degrades over time as a function of the main parameters N, b and m. Assuming $m, b < \sqrt{N}$, we prove that there is a threshold $t^* = O(m \ln N)$ such that when $t = t^* + 5cm$, for any integer $c > 0$, then

$$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0) \quad > \quad 1 - e^{-c}, \tag{2}$$

where the probabilities are computed over the random communication model (RCM) [1]. This result has the following interpretation. Once the attacker has collected a threshold t^* of messages sets, each additional $5m$ message sets collected decreases $H(\mathcal{R}_A \mid \mathcal{B}_t)$ towards zero at a geometric rate governed by $1/e$. The constant 5 is an upper bound and can be improved in practice. Using the benchmarking parameters $N = 20,000$, $b = 50$ and $m = 20$, our results yield that when $t = 292 + 22c$ the bound of (2) applies. When $t = 292 + 22 \cdot 10 = 512$, for example, then $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_{512}) = 0) > 0.9999$.

With a threshold of $t^* = O(m \ln N)$, the number of messages required to identify the communicating partners of Alice increases as the number of participants increases, albeit only logarithmically. This is somewhat counterintuitive since larger values of N should actually decrease the likelihood of having hitting sets different from $\mathcal{R}_A$. We show that if $N > O(b\, m \ln m)$ then the threshold can be improved to $t^* = O(m(\ln b + \ln m))$, which is independent of N. Further, if $b = O(m)$ then the threshold simplifies to $t^* = O(m \ln m)$ and this bound is asymptotically optimal since the coupon collector problem indicates that $t^* = \Omega(m \ln m)$.

The motivation for our work was to investigate if there are relatively simple relationships between the main parameters and the level of anonymity they provided as the number of observations by the attacker increases. Previous work on traffic confirmation

[5, 1, 2, 3, 6, 7] has not convincingly undercovered these fundamental relationships in a manner that is instructive for both system users and designers. This is somewhat surprising given the simplicity and symmetry of the RCM. Some of the previous work might be best described as "proof by plotting", often because the bounds on degrading anonymity are derived from system simulation or the evaluation of complex formula that have no known closed form.

2 Preliminaries

Let the total set of communicating users be $\mathcal{U}$, where $|\mathcal{U}| = N$ and N is large, say at least several tens of thousands. The set of potential senders will be denoted as $\mathcal{S}$ and similarly define $\mathcal{R}$ for the recipients. We assume that $|\mathcal{S}| = |\mathcal{R}| = N$, meaning that each user may send sender or receive messages. For each sender S, let $\mathcal{R}_S \subset \mathcal{R}$ denote the communicating partners of S. Communication between the users is mediated by a MIX that operates over a series of rounds, indexed by a parameter $t \geq 1$ which can be thought of as symbolizing time. At each round the MIX collects a message set (batch) B_t consisting of b messages sent by a collection of senders S_t. After processing the MIX delivers the messages of B_t to a collection of recipients R_t. It is assumed that the MIX perfectly hides the communication patterns between the members of S_t and R_t. The value of $b = |B_t|$ is referred to as the threshold of the MIX. When an example is helpful to clarify a point we will use $N = 20,000$, $b = 50$ and $m = 20$, which will be referred to as the *standard parameters*. These parameters have been used for previously for benchmarking search algorithms [5, 1, 6].

The attacker is assumed to be a global passive adversary, and can therefore determine S_t and R_t for each message set B_t collected by the MIX. A collection of t message sets observed by the attacker will be denoted as $\mathcal{B}_t = \{ B_1, B_2, \ldots, B_t \}$, where each B_i corresponds to a batch. To analyze traffic confirmation attacks we also require a model of the traffic patterns amongst the users of the MIX. We will adopt the *random communication model* (RCM) as proposed in [1], which has been used to study the disclosure attack for example. The defining properties of the model are:

- Each message batch is B_t formed by b distinct senders, which determines a receiver set R_t for the batch.
- When a sender, other than Alice, sends a message then the recipient is chosen uniformly amongst all N users.
- The user targetted by the attacker, Alice, has m regular communicating peers, denoted by $\mathcal{R}_A = \{ A_1, A_2, \ldots, A_m \} \subset \mathcal{R}$. The recipient of each message sent by Alice is chosen uniformly from $\mathcal{R}_A$.

- The attacker is able to observe the set of senders for each message batch B_t, and the corresponding set of recipients for R_t.

The last property of the RCM enables an attacker to discard all message batches where Alice is not a sender, and to focus on recipient sets that are guaranteed to contain at least one communicating peer of Alice. That is, the attacker can restrict their attention to message batches B_t for which $R_t \cap \mathcal{R}_A \neq \emptyset$. Our analysis will rely extensively on probabilistic arguments, and the relevant probability space will always be defined over the random choices for recipients as defined by the RCM. If Z is some property of a collection of message sets $\mathcal{B}_t$, then we will use the notation $\Pr(Z \mid \mathcal{B}_t)$ to denote the probability that $\mathcal{B}_t$ has property Z according to the recipient choices of the RCM.

3 Entropy and Hitting Sets

We are interested in deriving a bound on the number message sets that must be observed by an attacker in the RCM to determine $\mathcal{R}_A$ with high probability. Let $H(\mathcal{R}_A \mid \mathcal{B}_t)$ be the conditional entropy of $\mathcal{R}_A$ given a collection of $t \geq 1$ message sets $\mathcal{B}_t$ generated according to the RCM. A fundamental question is to determine how $H(\mathcal{R}_A \mid \mathcal{B}_t)$ evolves as a function of increasing t. By definition an attacker has sufficient information from $\mathcal{B}_t$ to compromise $\mathcal{R}_A$ if $H(\mathcal{R}_A \mid \mathcal{B}_t) = 0$. However, for any value of t there exist message sets such that $H(\mathcal{R}_A \mid \mathcal{B}_t) > 0$.

Lemma 3.1 For all $t, \Pr(|\mathcal{H}_m(\mathcal{B}_t)| = 1) > 1$.

Proof. If $|\mathcal{R}_A(\mathcal{B}_t)| = k < m$ then a proper subset of $\mathcal{R}_A$ is a hitting set for $\mathcal{B}_t$, and this set can be extended to $\binom{N-k}{m-k} > 1$ hitting sets of size m for $\mathcal{B}_t$. The probability that Alice fails to send to all of her recipients in t messages is greater than $(1 - 1/m)^t > 0$. □

Thus an attacker can never be certain that collecting a particular number of messages sets $\mathcal{B}_t$ will in fact compromise $\mathcal{R}_A$. On the other hand, if the attacker can determine a value of t such that $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0)$ is sufficiently small, an attack can be undertaken on the (grounded) assumption that most collections of t message sets will compromise $\mathcal{R}_A$. The probabilities are defined over the (random) choices of the senders for their recipients according to the RCM.

Let $M_1, M_2, \ldots, M_n$ be the $n = \binom{N}{m}$ distinct m-sets of N recipients. Then $H(\mathcal{R}_A \mid \mathcal{B}_t)$ is defined as

$$H(\mathcal{R}_A \mid \mathcal{B}_t) \;=\; -\sum_{M_i} \Pr(M_i = \mathcal{R}_A \mid \mathcal{B}_t) \cdot \log(\Pr(M_i = \mathcal{R}_A \mid \mathcal{B}_t)). \tag{3}$$

Evaluating the probabilities $\Pr(M_i = \mathcal{R}_A \mid \mathcal{B}_t)$ exactly may be difficult, and will be unnecessary for our purposes. In fact it will be sufficient to use the hitting set test, as

first suggested in [6], to distinguish between the zero and non-zero terms of (3). The m-set M_i is a hitting set for $\mathcal{B}_t$ if $M_i \cap B_j \neq \emptyset$ for each $B_j \in \mathcal{B}_t$, and let $\mathcal{H}_m(\mathcal{B}_t)$ denote all hitting sets of size m for $\mathcal{B}_t$. If $M_i \in \mathcal{H}_m(\mathcal{B}_t)$ then $\Pr(M_i = \mathcal{R}_A \mid \mathcal{B}_t) > 0$, as there is no evidence to exclude M_i as a candidate for $\mathcal{R}_A$. On the other hand, if $M_i \notin \mathcal{H}_m(\mathcal{B}_t)$ then clearly $\Pr(M_i = \mathcal{R}_A \mid \mathcal{B}_t) = 0$. We can therefore bound $H(\mathcal{R}_A \mid \mathcal{B}_t)$ as a function of $|\mathcal{H}_m(\mathcal{B}_t)|$.

Lemma 3.2 $H(\mathcal{R}_A \mid \mathcal{B}_t) \leq \log(|\mathcal{H}_m(\mathcal{B}_t)|)$.

Proof. If $|\mathcal{H}_m(\mathcal{B}_t)| = s$ then there are s non-zero points in the probability space of (3). The maximal entropy on any discrete space with s non-zero probabilities is less than $\log(s)$. □

Since Alice is a sender in all observed message sets, then $\mathcal{R}_A$ is always a hitting set for $\mathcal{B}_t$, and it follows that $|\mathcal{H}_m(\mathcal{B}_t)| \geq 1$ for all $t \geq 1$. If $|\mathcal{H}_m(\mathcal{B}_t)| = 1$ then we will say that there is a *unique* hitting set for $\mathcal{B}_t$, which by Lemma 3.2 implies $H(\mathcal{R}_A \mid \mathcal{B}_t) = 0$. We now show that the probability of a unique hitting manifesting for $\mathcal{B}_t$ tends to one as t increases.

Lemma 3.3 $\Pr(|\mathcal{H}_m(\mathcal{B}_t)| = 1) \to 1$ as $t \to \infty$.

Proof. Each message set consists of one message from Alice and $b-1$ messages from her peer senders. The probability that a given peer sender selects a recipient not included in M_i is $(1 - m/N) < 1$. Since $M_i \neq \mathcal{R}_A$ then Alice selects a recipient not in M_i with probability at least $1/m$. Then set p to be

$$p \;=\; 1 - \frac{1}{m}\left(1 - \frac{m}{N}\right)^{b-1}. \tag{4}$$

Clearly $p < 1$ and observe that $\Pr(M_i \in \mathcal{H}(\mathcal{B}_t)) \leq p^t$ for all $M_i \neq \mathcal{R}_A$. Then as $t \to \infty$,

$$\Pr(|\mathcal{H}_m(\mathcal{B}_t)| > 1) \;\leq \sum_{M_i \neq \mathcal{R}_A} \Pr(M_i \in \mathcal{H}_m(\mathcal{B}_t)) \;<\; p^t \cdot \binom{N}{m} \;\to\; 0. \tag{5}$$

The proof now follows since $\Pr(|\mathcal{H}_m(\mathcal{B}_t)| = 1) = 1 - \Pr(|\mathcal{H}_m(\mathcal{B}_t)| > 1)$. □

To prove the lemma it would have been sufficient to merely observe that p must be less than one. However since we have given the exact value of p in (4), we can provide a coarse upper bound on the number of observations required to compromise the recipients of Alice with an arbitrarily high probability.

Lemma 3.4 Let $d = (1 - m/N)^{b-1}$ and let $t^* = m^2 \ln N/d$. If $t = t^* + cm/d$, for some constant $c > 0$, then

$$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0) \;>\; 1 - e^{-c}. \tag{6}$$

Proof. Recalling that $(1-x) \leq e^{-x}$, p may be bound $p \leq e^{-d/m}$. Then

$$\Pr(|\mathcal{H}_m(\mathcal{B}_t)| > 1) \;<\; p^t \cdot N^m \;\leq\; \exp\left(m \ln N - \frac{dt}{m}\right). \tag{7}$$

Define $t^* = m^2 \ln N/d$ and observe that $p^t \cdot N^m < 1$ when $t = t^*$. Letting $t = t^* + cm/d$, for some constant $c > 0$, it than follows that

$$\Pr(|\mathcal{H}_m(\mathcal{B}_t)| > 1) \;<\; e^{-c}. \tag{8}$$

The lemma follows since $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0) = 1 - \Pr(|\mathcal{H}_m(\mathcal{B}_t)| > 1)$. □

We can state the result of the lemma informally as follows. After $O(m^2 \ln N)$ messages have been sent by Alice, the anonymity of her recipient set decreases exponentially towards zero for each additional $O(m)$ messages that she sends. For the standard parameters $N = 20,000$, $b = 50$ and $m = 20$, the lemma yields that $\lceil t^* \rceil = 4161$ and $\lceil m/d \rceil = 22$. Thus when $t = 4161 + 22c$, $c > 0$, the probability that $H(\mathcal{R}_A)$ is zero is greater than $1 - e^{-c}$.

4 A General Bound

In this section we improve show how to improve the previous threshold by a factor of m from $t^* = O(m^2 \ln N)$ to $t^* = O(m \ln N)$. This is achieved by replacing the single probability p with m probabilities tailored to the size of the intersection each m-set has with $\mathcal{R}_A$.

Theorem 4.1 Let $H_m = |\mathcal{H}_m(\mathcal{B}_t) - \mathcal{R}_A|$ be the number of hitting sets of size m for $\mathcal{B}_t$ that are distinct from $\mathcal{R}_A$. Then for $t \geq 1$,

$$\mathbf{E}[H_m] \;=\; \sum_{k=0}^{m-1} \binom{m}{k} \binom{N-m}{m-k} \left(1 - \left(1 - \frac{k}{m}\right)\left(\frac{N-m}{N}\right)^{b-1}\right)^t. \tag{9}$$

Proof. Let $M_1, M_2, \ldots, M_n$ be the $n = \binom{N}{m}$ m-sets of the N recipients. These m-sets can be partitioned according to the number of recipients k, $0 \leq k \leq m$, that each M_i has in common with $\mathcal{R}_A$. Consider a specific m-set M_i such that $|M_i \cap \mathcal{R}_A| = k < m$. If $M_i \cap B_j = \emptyset$ then both Alice and the peer senders selected recipients distinct from M_i, and these independent events jointly occur with probability

$$q_k \;=\; \left(1 - \frac{k}{m}\right)\left(\frac{N-m}{N}\right)^{b-1}. \tag{10}$$

Thus $\Pr(M_i \cap B_j \neq \emptyset) = 1 - q_k$ and $\Pr(M_i \in \mathcal{H}(\mathcal{B}_t)) = (1-q_k)^t$ since message sets are generated identically and independently. Since there are $\binom{m}{k}\binom{N-m}{m-k}$ m-sets M_i such that $|M_i \cap \mathcal{R}_A| = k$, then

$$\mathbf{E}[H_m] \;=\; \sum_{M_i \neq \mathcal{R}_A} \Pr(M_i \in \mathcal{H}(\mathcal{B}_t)) \;=\; \sum_{k=0}^{m-1} \binom{m}{k}\binom{N-m}{m-k}(1-q_k)^t. \tag{11}$$

Substituting (10) into (11) completes the proof. □

By definition $H_m = 0$ implies $H(\mathcal{R}_A \mid \mathcal{B}_t) = 0$, and if $\mathbf{E}[H_m] \to 0$ as a function of t, then $\Pr(H_m = 0) \to 1$ since

$$\Pr(H_m = 0) \;=\; 1 - \Pr(H_m \geq 1) \;\geq\; 1 - \mathbf{E}[H_m]. \tag{12}$$

However it is clear that $\mathbf{E}[H_m] \to 0$ for increasing t since the binomial terms of (9) are independent of t, and each term $(1-q_k)$ is less than 1. Evaluating $\mathbf{E}[H_m]$ directly for the standard parameters $N = 20,000$, $b = 50$ and $m = 20$, yields that the smallest value of t for which $\mathbf{E}[H_m] < 1$ is $t^* = 268$ (a significant reduction from the corresponding bound $t^* = 4161$ given in previous section from Lemma 7).

We now use $\mathbf{E}[H_m]$ to bound the convergence of $H(\mathcal{R}_A \mid \mathcal{B}_t)$ to zero.

Theorem 4.2 Let $d = (1-m/N)^{b-1}$ and let $t^* = m(\ln N + \ln m + 1)/d$. If $t = t^* + cm/d$, for some constant $c > 0$, then

$$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0) \;>\; 1 - e^{-c}. \tag{13}$$

Proof. We begin by bounding the binomial and $(1-q_k)^t$ terms in (9). Combining the bounds $\binom{n}{k} \leq (ne/k)^k$ and $\binom{n}{k} \leq n^k$, and recalling that $\binom{n}{k} = \binom{n}{n-k}$, it follows that

$$\binom{m}{k}\binom{N-m}{m-k} \;\leq\; (Nm)^{m-k} \cdot \min\left\{1, \left(\frac{e}{m-k}\right)^{2(m-k)}\right\}. \tag{14}$$

Also since $(1-q_k)^t$ can be bound as

$$(1-q_k)^t \;=\; \left(1 - \frac{d(m-k)}{m}\right)^t \;<\; \exp\left(-\frac{td(m-k)}{m}\right) \tag{15}$$

it follows that $(1-q_k)^t < (Nm)^{k-m}$ when $t = m(\ln N + \ln m)/d$. Combining this result with (14) yields that

$$\mathbf{E}[H_m] \;=\; \sum_{k=0}^{m-1} \binom{m}{k}\binom{N-m}{m-k}(1-q_k)^{m(\ln N + \ln m)/d}$$

$$
\begin{aligned}
&< \sum_{k=m-2}^{m-1} 1 + \sum_{k=0}^{m-3} \left(\frac{e}{m-k}\right)^{2(m-k)} \\
&< 2 + \left(\frac{e}{3}\right)^6 + \sum_{k=0}^{m-4} \left(\frac{e}{4}\right)^{2(m-k)} \\
&= 2.5534 + \sum_{k=0}^{m-4} \left(\frac{e^2}{16}\right)^{m-k} \\
&< 2.5534 + \sum_{k\geq 0} \left(\frac{e^2}{16}\right)^{k} - \sum_{k=0}^{3} \left(\frac{e^2}{16}\right)^{k} \\
&= 2.6379 \\
&< e.
\end{aligned}
$$

Thus when $t > m(\ln N + \ln m)/d$, it follows that $\mathbf{E}[H_m]$ can be bounded as

$$\mathbf{E}[H_m] \;<\; e \cdot \Big[\max_k (1-q_k)\Big]^{t-m(\ln N+\ln m)/d}. \tag{16}$$

But from (15) we see that

$$\max_k (1-q_k) \;=\; (1-q_{m-1}) \;<\; e^{-d/m}. \tag{17}$$

Finally, let $t^* = m(\ln N + \ln m)/d + m/d$ and set $t = t^* + cm/d$ for some constant $c > 0$. Then substituting (17) into (16) yields that

$$\mathbf{E}[H_m] \;<\; e \cdot (1-q_{m-1})^{m/d+cm/d} \;<\; (e^{-d/m})^{cm/d} \;=\; e^{-c}. \tag{18}$$

The proof is completed by recalling that $\Pr(|\mathcal{H}_m(\mathcal{B}_t)| = 1) = \Pr(H_m = 0)$ and then applying (12). □

What Theorem 4.2 actually shows that the rate at which $\mathbf{E}[H_m] \to 0$ is essentially governed by the rate at which $m(N-m)(1-q_{m-1})^t \to 0$. This is to be expected since the m-sets corresponding to $(1-q_{m-1})^t$ are those m-sets that differ from $\mathcal{R}_A$ by only a single recipient, and intuitively these m-sets require the largest number of observations to eliminate as false positives. We can simplify the results of Theorem 4.2 by replacing d with a constant.

Corollary 4.1 When $m, b < \sqrt{N}$ then $1/d < 5$ and $t^* = O(m \ln N)$.

Proof. Observe that if $d \geq z$ then $1/d \leq 1/z$. Recalling that $(1-x) \geq e^{-x-x^2/2}$ for all $0 \leq x \leq 1/2$ then $(1-m/N) \geq \exp(-m/N - m^2/(2N^2))$ since $m < \sqrt{N}$. Then since

$mb < N$ it follows that

$$1/d \;\leq \exp\left(\frac{bm}{N} + \frac{bm^2}{2N^2}\right) \;\leq\; \exp\left(1 + 1/2\right) \;<\; 5. \tag{19}$$

Then $t^* = O(m \ln N)$ since $t^* < 5m(\ln N + \ln m + 1)$. □

For the standard parameters the relevant values are $1/d = 1.05024$, $t^* = 292$ and $\lceil m/d \rceil = 22$. Thus when $t = 292 + 22c$, $c > 0$, the probability that $H(\mathcal{R}_A)$ is zero is greater than $1 - e^{-c}$. Table 1 shows the smallest values of c for which $(1 - e^{-c}) \geq 1 - 10^k$, $1 \leq k \leq 6$. For example, consider the row $c = 10$ corresponding to $t = 292 + 22 \cdot 10 = 512$ observations. In this case $e^{-c} = 0.45399 \times 10^{-4}$ which shows that there is a 99.99% chance $H(\mathcal{R}_A)$ is zero since $1 - e^{-c} > 1 - 10^{-4}$. Theorem 4.2 bounds $\mathbf{E}[H_m]$ by e^{-c} when the number of observation exceeds t^*, and Table 1 also includes a comparison between these two values.

c	$t^* + cm$	e^{-c}	$\mathbf{E}[H_m]$	$e^{-c}/\mathbf{E}[H_m]$	$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0)$
3	358	0.49787×10^{-1}	0.10425×10^{-1}	4.7752	> 0.9
5	402	0.67379×10^{-2}	0.12179×10^{-2}	5.5323	> 0.99
7	446	0.91188×10^{-3}	0.14238×10^{-3}	6.4042	> 0.999
10	512	0.45399×10^{-4}	0.56925×10^{-5}	7.9752	> 0.9999
12	556	0.61442×10^{-5}	0.66558×10^{-6}	9.2312	> 0.99999
14	600	0.83152×10^{-6}	0.77821×10^{-7}	10.685	> 0.999999

Table 1: Lower bounds on $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t) = 0)$ using the standard parameters.

5 A Bound independent of N

With a threshold of $t^* = O(m \ln N)$, the number of messages required to identify the communicating partners of Alice increases as the number of participants increases, albeit only logarithmically. This is somewhat counterintuitive since larger values of N should actually decrease the likelihood of finding hitting sets different from $\mathcal{R}_A$. Consider an attacker who has observed t message sets $\mathcal{B}_t$ that determines a set of receivers $\mathcal{R}_t$. Our main observation is that if t is sufficiently large so that $\mathcal{R}_A \subseteq \mathcal{R}_t$ with high probability, then it may be more efficient to apply the hitting set test to m-sets from $\mathcal{R}_t$ rather than to m-sets from all possible N receivers. In this section we show that if $N > O(b\, m \ln m)$ then considering the receivers of $\mathcal{R}_t$ is a better strategy, and N can be eliminated from the threshold to yield $t^* = O(m(\ln b + \ln m))$.

According to the RCM, Alice selects her communicating partners uniformly and at random for each message. Therefore the number of messages that she needs to send before each of the recipients receives at least one message is an example of the coupon collector problem [4]. The coupon collector problem considers the number X_n of uniform samples (with replacement) required to observe each of n items at least once. It is well-known [8] that $\mathbf{E}[X_n] = n \ln n + O(n)$ and that for any real constant $c > 0$,

$$\lim_{n\to\infty} \Pr(|X_n - n \ln n| \leq cn) \;=\; e^{-e^c} - e^{-e^{-c}} \tag{20}$$

When required, the exact distribution of X_n can be computed via

$$\Pr(X_n \leq t) \;=\; \sum_{k=0}^{n} \binom{n}{k} \left(1 - \frac{k}{n}\right)^t (-1)^k \;=\; 1 - \sum_{k=1}^{n} \binom{n}{k} \left(1 - \frac{k}{n}\right)^t (-1)^k. \tag{21}$$

For the standard parameters $N = 20,000$, $b = 50$ and $m = 20$, we can then determine that if Alice sends 150 messages then each of her 20 recipients receives at least one message with probability greater than 0.99. Thus the attacker need only consider m-sets from amongst at most $50 \cdot 150 = 7500$ recipients, rather than the full 20,000 potentials recipients, to succeed with high probability. We will also use the following simple non-asymptotic bound

$$\Pr(X_n > t) \;<\; m\,(1 - 1/m)^t \;<\; \exp\left(\ln m - \frac{t}{m}\right). \tag{22}$$

Defining t_α as $t = \alpha m \ln m$, it then follows that $\Pr(\mathcal{R}_A \not\subseteq \mathcal{R}_t) \geq 1 - m^{-(\alpha-1)}$. We can now recast the threshold of Theorem 4.2 to be independent of N.

Theorem 5.1 Let α be a constant $\alpha > 1$ and let $d = (1 - m/N)^{b-1}$. Let $t^* = t_\alpha + m(\ln(b \cdot t_\alpha) + \ln m + 1)/d$. Assuming that $m, b < \sqrt{N}$, then when $t = t^* + t_{\alpha-1}/d$

$$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t = 0) \;>\; 1 - \frac{2}{m^{\alpha-1}}. \tag{23}$$

Proof. After observing the first t_α message sets, the attacker then collects additional message sets and constructs candidate m-sets only from $\mathcal{R}_{t_\alpha}$. Let $\mathbf{E}[H_m]$ be the expected number of such hittings sets generated by this process. Then

$$\mathbf{E}[H_m] \;=\; \sum_{k=0}^{m-1} \binom{m}{k} \binom{b \cdot t_\alpha - m}{m - k} \left(1 - \left(1 - \frac{k}{m}\right)\left(1 - \frac{m}{N}\right)^{b-1}\right)^{t - t_\alpha}. \tag{24}$$

Now if $\mathcal{R}_A \in \mathcal{R}_{t_\alpha}$, then the rate at which $\mathbf{E}[H_m] \to 0$ follows the same asymptotics as in Theorem 4.2 except that N is replaced by $b \cdot t_\alpha$ in a binomial term of24) . To complete the proof we need only bound $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t = 0)$, which can be done as follows

$$\begin{aligned} \Pr(H(\mathcal{R}_A \mid \mathcal{B}_t = 0) &= \Pr(\mathcal{R}_A \subseteq \mathcal{R}_{t_\alpha}) \cdot \Pr(H(\mathcal{R}_A \mid \mathcal{B}_{t-t_\alpha} = 0) \\ &> \left(1 - \frac{1}{m^{\alpha-1}}\right)\left(1 - \frac{1}{m^{\alpha-1}}\right) \\ &> 1 - \frac{2}{m^{\alpha-1}}. \end{aligned} \tag{25}$$

The bound on $\Pr(H(\mathcal{R}_A \mid \mathcal{B}_{t-t_\alpha} = 0)$ follows by substituting $c = (\alpha - 1)\ln m$ into (13). □

Corollary 5.1 Let α be a constant $\alpha > 1$ and let $mb < N$. If $b = O(m)$ then $t^* = O(m \ln m)$ and this bound is aymptotically optimal.

Proof. Corollary 4.1 applies when $mb < N$ and therefore $1/d$ is a bound by a constant. Since $t_\alpha = O(m \ln m)$ and $\ln(b \cdot t_\alpha) = O(\ln b + \ln m)$ then $t^* = O(m(\ln b + \ln m))$. Finally, if $b = O(m)$ the threshold simplifies to $t^* = O(m \ln m)$ and this bound is asymptotically optimal since the coupon collector problem indicates that $t^* = \Omega(m \ln m)$. □

For the last time, consider the standard parameters $N = 20,000$, $b = 50$ and $m = 20$,

$\lceil m \ln m \rceil$	α	t_α	$\Pr(\mathcal{R}_A \subseteq \mathcal{R}_{t_\alpha})$	t^*	$t_{\alpha-1}/d$	t	$\Pr(H(\mathcal{R}_A \mid \mathcal{B}_t = 0)$
60	2	120	0.958	185	127	432	0.900
60	3	180	0.998	194	190	564	0.995
60	4	240	0.9999	200	253	693	0.999

Table 2: Parameters from Theorem (5.1) using the standard parameters.

and the relevant parameters from Theorem (5.1) are shown in Table 2 for a few values value of α. If we compare these results to Table 1, we actually see that less messages are required using the previous $t^* = O(m \ln N)$ bound than the bound from Theorem (5.1). The reason for this is that the bound Theorem (5.1) only uses the first t_α message sets to determine a recipient set, and does not apply the hitting set test to filter candidate m-sets. We are working on a model that can use all observed message sets for hitting set filtering.

6 Conclusion

Traffic confirmation attacks remain a common approach to assess how the level of anonymity in a given anonymity system changes over time. Since most practical systems leak in-

formation over time, anonymity degrades with the number of observations made on the system. The inherent structure of the rate at which leakage occurs follows a threshold and tail structure from the coupon collector problem - a certain minimum number of messages are required to observe most of $\mathcal{R}_A$ and then some additional observations (the tail) are required to gather the remaining recipients. The basic coupon collector threshold is $t^* = O(m \ln m)$. The coupon collector problem corresponds to a traffic confirmation problem where $b = 1$. For larger (and more practical values) of b the threshold increases as the number of participants N increases but then reaches a limit since participants not contained in $\mathcal{R}_A$ are less likely to be observed as recipients. We have shown that $t^* = O(m \ln N)$ is a general threshold applicable when $mb < N$ and $t^* = O(m(\ln b + \ln m))$ is appropriate threshold for N larger than $O(b\, m \ln m)$. In both cases the rate at which $H(\mathcal{R}_A \mid \mathcal{B}_t)$ converges to zero beyond the respective thresholds is the same, and is governed by

$$p \;=\; 1 - \frac{1}{m}\left(1 - \frac{m}{N}\right)^{b-1} \;=\; 1 - \frac{d}{m}$$

where d tends to a constant with increasing N.

Our results rely on the presence of an global passive adversary, the symmetry and simplicity of the RCM as well as the structure provided by threshold mixes with respect to hitting set attacks. All, or any, of these assumptions can be criticized as being impractical or unrealistic, and we willingly accept such criticism. Our results are designed to show relatively simple relationships between the main system parameters N, b, m and the anonymity provided by threshold mixes against targetted traffic conformation attacks. Previously work on this topic has not convincingly undercovered these fundamental relationships in a manner that is instructive for both system users and designers. With a rigorous analysis of a simple system completed we can now consider more complicated systems that use more complex traffic models, weaker adversaries or more sophisticated anonymity mechanisms such as pool mixes.

Our approach has focussed on the degradation of $H(\mathcal{R}_A \mid \mathcal{B}_t)$ as a function of increasing t. Since we have used an information-theoretic approach then our bounds do not give an indication of the amount of work that must be undertaken to recover $\mathcal{R}_A$ when its entropy is practically zero. We are currently working on a new analysis of the statistical disclosure attack, based on Chernoff bounds, which we hope will show that $t^* = O(m \ln N)$ is also a threshold for distinguishing $\mathcal{R}_A$ from all recipients based on simple counting arguments.

References

[1] Dakshi Agrawal, Dogan Kesdogan, and Stefan Penz. Probabilistic Treatment of MIXes to Hamper Traffic Analysis. In *Proceedings of the 2003 IEEE Symposium*

on Security and Privacy, May 2003.

[2] George Danezis. Statistical disclosure attacks. In Samarati Katsikas Gritzalis, Vimercati, editor, *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–426, Athens, May 2003. IFIP TC11, Kluwer.

[3] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In *Proceedings of 6th Information Hiding Workshop (IH 2004)*, LNCS, Toronto, May 2004.

[4] W. Feller. *An Introduction to Probability Theory and its Applications*. New York: Wiley, 3rd edition, Volume 1, 1968.

[5] Dogan Kesdogan, Dakshi Agrawal, and Stefan Penz. Limits of anonymity in open environments. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, October 2002.

[6] Dogan Kesdogan and Lexi Pimenidis. The hitting set attack on anonymity protocols. In *Proceedings of 6th Information Hiding Workshop (IH 2004)*, LNCS, Toronto, May 2004.

[7] Nick Mathewson and Roger Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, volume 3424 of *LNCS*, May 2004.

[8] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge Univerity Press, 1995.