On the Anonymity of Chaum Mixes

Parvathinathan Venkitasubramaniam
Electrical and Computer Engineering
Cornell University
Ithaca, NY 14850
Email: pv45@cornell.edu

Venkat Anantharam
Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, CA 94720
Email: ananth@eecs.berkeley.edu

***Abstract*— The information-theoretic analysis of Chaum mixing under latency constraints is considered. Mixes are relay nodes that collect packets from multiple users and modify packet timings to prevent an eavesdropper from identifying the sources of outgoing packets. In this work, an entropy-based metric of *anonymity* is proposed to quantify the performance of a mixing strategy under strict delay constraints. Inner and outer bounds on the maximum achievable anonymity are characterized as functions of traffic load and the delay constraint. The bounds are shown to have identical first derivatives at low traffic loads.**

***Index Terms*— mixing, traffic analysis, anonymity, timing channels.**

I. Introduction

Privacy in network communication extends beyond the protection of communicated data; hiding the identities of communicating parties is equally essential. Knowledge of source-destination pairs or routes of information flow in a network not only compromises user anonymity, but also provides vital information for adversaries to jam a flow or launch a denial of service attack. On the Internet, different software applications are available that enable anonymous communication. Most of these applications are based on the concept of Chaum mixes [1]. A mix is a relay node or proxy server that collects packets from multiple users and outputs them in such a way that an external eavesdropper cannot link an outgoing packet to the corresponding sender. Specifically, a mix obfuscates the contents of every received packet using encryption techniques, and modifies the packet timing by delaying and reordering the packets. As expected, modifications to timing increase the latency of transmitted packets. Alternatively, if packets are subjected to strict latency constraints, the capabilities of the mix are restrained, thereby reducing the *anonymity* of outgoing packets. This gives rise to interesting questions: How do we measure the anonymity of a mixing strategy? If packets arriving to a mix are delay-constrained, what is the maximum achievable anonymity? In this paper, we address these questions from an information-theoretic perspective.

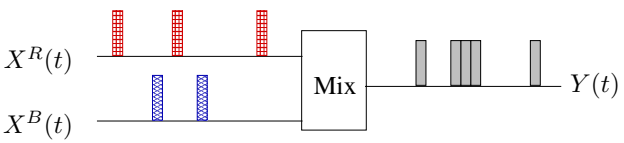

Fig. 1. Chaum mix: $X_R(t)$ and $X_B(t)$ are the two arrival processes, and $Y(t)$ is the departure process as observed by the eavesdropper.

Consider the setup in Figure 1. Each source transmits packets to a mix according to an independent Poisson process. Every received packet can be delayed by the mix up to a maximum of T seconds. We assume that the mix uses encryption to perfectly decorrelate the contents of incoming and outgoing packets, so the eavesdropper can only observe a single departure process. Using the packet timing in the arrival and departure processes, the eavesdropper's goal is to identify the source of each departing packet. We measure the anonymity of a mixing strategy using the normalized entropy of the a posteriori probability distribution of possible input-output pairings based on the eavesdropper's observation. The goal is to characterize the maximum achievable anonymity as a function of the arrival rate and delay constraint. In this paper, we provide lower and upper bounds for the maximum achievable anonymity.

In the original design by Chaum, for n inputs, the mixing strategy is to wait until packets are received from n different users, and send all the packets out in one batch. Since packets within a batch can be ordered arbitrarily, the eavesdropper can never identify the source of any packet. This strategy has been subsequently improved upon to address delay constraints [2], and also extended to networks of mixes [3]. Theoretical analyses of the anonymity of mixing are, however, very limited. In [4], an information theoretic metric of anonymity using equivocation [5] was proposed, and simulations were used to evaluate the anonymity of known mixing strategies. The approach in [4] treats every departing packet independently and does not take into account the delay constraints or the traffic statistics. In this paper, we consider the complete observation of the eavesdropper in our model, and our goal is to characterize the anonymity as a function of traffic load and the delay constraint. A related problem is that of jammers on a timing channel analyzed by Giles and Hajek in [6]; while the task of the jammer in [6] is to modify the packet timings so that communication through timing is prevented, the task of a mix is to modify the timing patterns such that the identities of transmitting sources are not revealed.

The rest of this paper is organized as follows. In Section II, we describe the mathematical formulation of the problem. In Sections III and IV, we characterize upper and lower bounds on the maximum achievable anonymity respectively. The comparison between bounds is presented in Section V.

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

II. Problem Setup

Let $\{X_R(t)\}$, $\{X_B(t)\}$ be two independent Poisson processes on $\mathbb{R}$ with equal rates λ. For purposes of this presentation, we assume that arrival rates are equal; unequal rates can be handled using similar techniques. $X_R(t)$ and $X_B(t)$ represent the arrival processes of packets from two sources to the mix (see Figure 1). For ease of presentation, we shall henceforth use the colours red and blue to refer to packets in $X_R(t)$ and $X_B(t)$ respectively.

Each packet may be delayed by the mix using a randomized strategy subject to causality and a maximum delay constraint of T. We let Ψ_T denote the set of all valid mixing strategies. $Y^R(t)$ and $Y^B(t)$ denote the corresponding departure processes of red and blue packets. In either or both of these processes, multiple packets may depart simultaneously. The eavesdropper, Eve, unaware of the colours of departing packets, observes the net departure process $Y(t) = Y^R(t)+Y^B(t)$. In the rest of this paper, the term *departure process* refers to the process $\{Y(t)\}$. She also observes the arrival processes $\{X^R(t)\}$ and $\{X^B(t)\}$. Using her observation and knowledge of the mixing strategy, Eve's goal is to guess the colours of departing packets. Note that she does not have access to the realization of random variables used by the mix.

A. Figure of Merit

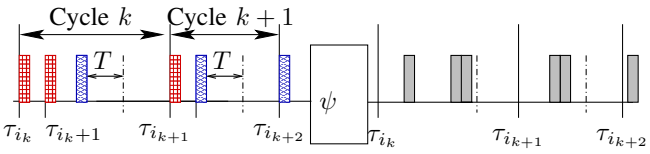

Fig. 2. Division of arrival process into cycles

The joint arrival process $X(t) = X_R(t) + X_B(t)$ can be viewed as a concatenation of cycles (see Figure 2). If the points of $X(t)$ are denoted by $\{\tau_k\}$, then consider a subsequence $\{\tau_{i_k}\}$ of $\{\tau_k\}$ such that

$$\begin{aligned} \tau_{i_k} - \tau_{i_k-1} &> T, \ \forall k, \\ \tau_n - \tau_{n-1} &\leq T, \ \forall k, n : i_{k-1} < n < i_k. \end{aligned}$$

Let $\tau_{i_0} \leq 0 < \tau_{i_1}$. The k^{th} cycle is defined by the time period of observation $[\tau_{i_k}, \tau_{i_{k+1}})$. Each cycle is preceded by a period of T seconds, where there have been no arrivals. Therefore all packets that arrived to the mix prior to τ_{i_k} would have departed before τ_{i_k}. Each cycle continues until the first time when there have been no arrivals to the mix for a period of at least T seconds. Therefore all arrivals during a cycle would have departed during the cycle. Using stationarity and memorylessness of Poisson processes, it is easily verified that for the Palm distribution with respect to the starting time of cycles, each cycle is independent and identically distributed.

During the k^{th} cycle, namely $[\tau_{i_k}, \tau_{i_{k+1}})$, let N_k^R, N_k^B denote the number of red and blue packets respectively that arrived to the mix. The total number of arrivals during the cycle is denoted by $N_k = N_k^R + N_k^B$, which is equal to the total number of departures during the cycle. We enumerate the departures in cycle k in increasing order from 1 through N_k. Note that a mixing strategy can send multiple packets at the same time, in which case, it does not matter how the departures are ordered within such *batches*. Consider any colouring of the numbers 1 through N_k with the colours red or blue, such that the colour red (resp. blue) is used exactly N_k^R (resp. N_k^B) times. If the k^{th} packet is coloured red, it indicates that the k^{th} departure in the cycle is a red packet. Given the realization of the arrival processes and departure process, the mixing strategy $\psi \in \Psi_T$ results in a probability distribution on the set of all such colourings. This may be thought of as resulting from a maximum a posteriori calculation done by Eve, who has knowledge of the randomized strategy of the mix, but does not have access to the realization of the random variables used by the mix. The entropy of this probability distribution, denoted by Λ_k^ψ, is a real valued measurable function defined on the space of realizations of the arrival and departure processes. However, since the arrival process is a random object and the departure process is random conditioned on the arrivals (involving the randomized strategy of the mix), Λ_k^ψ is a random variable (as are N_k, N_k^R, and N_k^B).

Definition 1: The **anonymity** A^ψ of a mixing strategy ψ is defined as

$$A^\psi(\lambda T) = \mathbb{E}\left\{ \liminf_{K\to\infty} \frac{\sum_{k=-K}^{K} \Lambda_k^\psi}{\sum_{k=-K}^{K} N_k} \right\}, \qquad (1)$$

where the expectation is over arrival and departure processes.

We assume without loss of generality that the mixing strategy is Palm-stationary from cycle to cycle, so the limit can be assumed to exist. By the pointwise ergodic theorem:

$$A^\psi(\lambda T) = \frac{\tilde{\mathbb{E}}^0[\Lambda_0^\psi]}{\tilde{\mathbb{E}}^0[N_0]},$$

where $\tilde{\mathbb{E}}^0$ denotes the Palm expectation with respect to the starting times of cycles.

For a two source system, A^ψ takes values in $[0, 1]$. $A^\psi = 0$ implies that Eve can accurately identify the colour of every departing packet, while $A^\psi = 1$ implies that every departing packet is equally likely to be red or blue from her perspective. Intuitively, the larger A^ψ is, the harder it is for Eve to determine if any departing packet is red or blue.

We wish to characterize the maximum achievable anonymity:

$$\mathcal{A}(\lambda T) = \sup_{\psi\in\Psi_T} A^\psi(\lambda T).$$

In the subsequent sections we provide lower and upper bounds for $\mathcal{A}(\lambda T)$. Note that since arrival processes have identical rates, the defined system can be equivalently treated as a system with arrival rate λT and delay constraint 1. The anonymity is therefore expressed as a function of λT.

III. Upper Bound

The maximum anonymity achievable within the class of causal mixing strategies is at most equal to that achievable

within the class of non-causal strategies. We therefore derive an upper bound for $\mathcal{A}(\lambda T)$ by relaxing the causality constraint. Specifically, we assume that the mix has complete knowledge of the arrival times within each cycle prior to generating the packet departure times for that cycle.

Let $\Upsilon_R = (\tau_R^1, \cdots, \tau_R^{N_R})$ and $\Upsilon_B = (\tau_B^1, \cdots, \tau_B^{N-N_R})$ denote the set of arrival times of red and blue packets respectively, and $\Upsilon_R^T = (\tau_R^1 + T, \cdots, \tau_R^{N_R} + T)$ and $\Upsilon_B^T = (\tau_B^1 + T, \cdots, \tau_B^{N-N_R} + T)$ represent the respective deadlines of red and blue packets in a cycle. Let $\Upsilon = (t_1, \cdots, t_{2N})$ be the ordered union of $\Upsilon_R, \Upsilon_B, \Upsilon_R^T$ and Υ_B^T.

Let m_i denote the number of packets that depart in the interval $I_i = (t_i, t_{i+1}]$. Since there are no arrivals or expiring deadlines within each I_i except perhaps at t_{i+1}, and the mix is assumed to have non-causal knowledge of all arrival times in the cycle, it suffices for the mix to transmit all m_i packets in a single batch. We measure the uncertainty of Eve for the cycle using the logarithm of the cardinality of the number of colourings of the N departures that are consistent with Eve's observations, which is an upper bound on the entropy achieved by the non-causal mix in the cycle. Accordingly, we define the task of the (non-causal) mix to be the design the *departure sequence* $\mathbf{m} = \{m_1, m_2, \cdots, m_{2N-1}\}$ given Υ_B, Υ_R such that the uncertainty as measured by this logarithm is maximized. Since the design of the departure sequence only depends on the order of arrivals and deadlines in a cycle, it is sufficient to view each cycle as a sequence of arrivals and deadlines, irrespective of the individual times Υ.

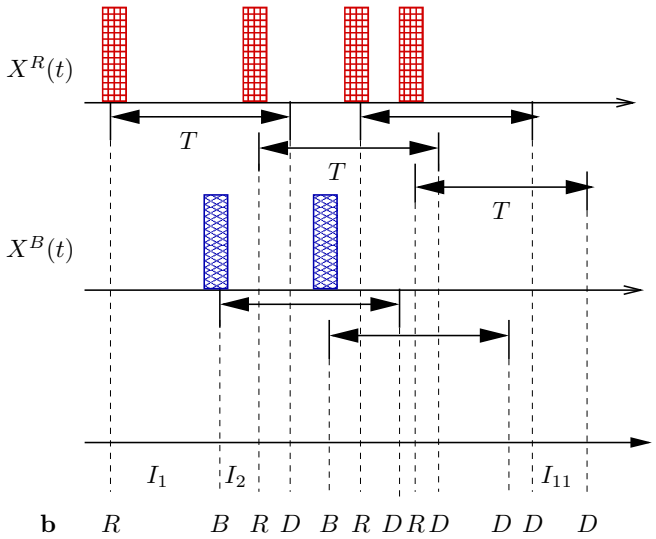

Fig. 3. Example of arrival sequence in a cycle with 6 arrivals. Arrival and deadline segments: $\underbrace{RBR}_{l_1}\ \underbrace{D}_{k_1}\ \underbrace{BR}_{l_2}\ \underbrace{D}_{k_2}\ \underbrace{R}_{l_3}\ \underbrace{DDDD}_{k_3}$. $S = 3, l_1 = 3, l_2 = 2, l_3 = 1,\ k_1 = 1, k_2 = 1, k_3 = 4$.

We represent a cycle with N arrivals as a vector $\mathbf{b} = [b(1)b(2)\cdots b(2N)]$ such that each $b(i) \in \{B, R, D\}$. B and R denote arrivals of red and blue packets respectively, and D denotes the deadline of a packet (see Figure 3). Consider the unique division of $\mathbf{b}$ into segments, wherein each segment is a contiguous sub-vector in $\mathbf{b}$ that contains only arrivals or only deadlines, and the segments containing arrivals and deadlines alternate in $\mathbf{b}$. Let S denote the total number of arrival segments (which is equal to the number of deadline segments). Let l_s be the length of the s^{th} arrival segment and k_s be the length of the s^{th} deadline segment. It is easily verified that any vector $\mathbf{b}$ of length $2N$ satisfying the conditions

$$l_s, k_s > 0,\ 1 \le s \le S, \tag{2}$$

$$\sum_{u=1}^{s} l_u > \sum_{u=1}^{s} k_u,\ 1 \le s < S, \tag{3}$$

$$l_1 + \cdots + l_S = k_1 + \cdots + k_S = N, \tag{4}$$

corresponds to an $N-$arrival cycle. Let $\mathcal{B}_N$ denote the set of all $\mathbf{b} \in \{B, R, D\}^{2N}$ that satisfy (2)-(4), and let $\mathcal{B} = \bigcup_N \mathcal{B}_N$. We refer to elements of $\mathcal{B}$ as *interlacing patterns*.

For every interlacing pattern $\mathbf{b} \in \mathcal{B}_N$, the mix generates a departure sequence $\mathbf{m} \in (\mathbb{Z}^+)^{2N-1}$. A departure sequence $\mathbf{m}$ is *feasible* for an interlacing pattern $\mathbf{b}$ if and only if:

$$\sum_{i=1}^{2N-1} m_i = N,$$

$$\forall i, |k : k \le i, b(k) = D| \le \sum_{j=1}^{i-1} m_j \le |k : k < i, b(k) \ne D|.$$

Let $\mathcal{M}(\mathbf{b})$ denote the set of feasible departure sequences for an interlacing pattern $\mathbf{b}$. Let $\gamma(\mathbf{b}, \mathbf{m})$ denote Eve's uncertainty of packet colourings given she observes a departure sequence $\mathbf{m} \in \mathcal{M}(\mathbf{b})$. Then, the maximum uncertainty achievable by a non-causal mixing strategy for the arrival sequence $\mathbf{b}$ is:

$$\Gamma^*(\mathbf{b}) = \sup_{\mathbf{m} \in \mathcal{M}(\mathbf{b})} \gamma(\mathbf{b}, \mathbf{m}).$$

Lemma 1: For an interlacing pattern $\mathbf{b}$, let $\mathcal{D}(\mathbf{b}) = \{d_1, \cdots, d_S\}$ where

$$d_s = l_s + \sum_{u=1}^{s-1} (l_u + k_u),\ s = 1 \cdots S,$$

and let $\mathcal{M}'(\mathbf{b}) = \{\mathbf{m} \in \mathcal{M}(\mathbf{b}) : m_i > 0 \text{ only if } i \in \mathcal{D}(\mathbf{b})\}$. Then,

$$\Gamma^*(\mathbf{b}) = \sup_{\mathbf{m} \in \mathcal{M}'(\mathbf{b})} \gamma(\mathbf{b}, \mathbf{m}).$$

Proof: Available in [7]. □

Lemma 1 states that the mix needs to consider only a subset $\mathcal{M}'(\mathbf{b})$ of feasible departure sequences in order to maximize uncertainty. Specifically, it suffices for the mix to transmit packets only in the intervals in $\{I_{d_s}, d_s \in \mathcal{D}(\mathbf{b})\}$, which is the set of all intervals that immediately follow arrival segments.

Consider the set of interlacing patterns that satisfy the following additional property:

$$\sum_{u=1}^{s} k_u \ge \sum_{u=1}^{s-1} l_u,\ s = 1 \cdots S. \tag{5}$$

We shall refer to interlacing patterns that satisfy the above condition as *restricted interlacing patterns*. According to (5),

every packet that arrived in the s^{th} arrival segment has a deadline that expires prior to the $(s+2)^{th}$ arrival segment. Using Lemma 1, it is easy to see that for every packet that arrived in the s^{th} arrival segment, the mix has at most two choices: transmit the packet in interval I_{d_s} or transmit the packet in interval $I_{d_{s+1}}$. This reduction enables us to compute an upper bound on $\Gamma^*(\mathbf{b})$ by restricting our attention to the set of restricted interlacing patterns.

Although the set of restricted interlacing patterns do not exhaustively cover $\mathcal{B}$, any interlacing pattern can be transformed to a restricted pattern by delaying deadlines of packets until the condition in (5) is met. Since delaying deadlines can only increase the achievable uncertainty, the transformed pattern can be used to evaluate an upper bound on $\Gamma^*(\mathbf{b})$ for all $\mathbf{b} \in \mathcal{B}$. Define the following notation:

$$\begin{aligned}
\mathbf{b}_i^j &: b(i)b(i+1)\cdots b(j) \\
n_R(\mathbf{b}) &: |\{i : b(i) = R\}| \\
n_B(\mathbf{b}) &: |\{i : b(i) = B\}| \\
m_{min}(\mathbf{b}) &: \min(n_R(\mathbf{b}_{k_1+1}^{l_1}), n_B(\mathbf{b}_{k_1+1}^{l_1})) \\
m_{max}(\mathbf{b}) &: \max(n_R(\mathbf{b}_{k_1+1}^{l_1}), n_B(\mathbf{b}_{k_1+1}^{l_1})) \\
r_l(\mathbf{b}, m) &: \max(0, m - n_B(\mathbf{b}_{k_1+1}^{l_1})) \\
r_u(\mathbf{b}, m) &: \min(m, n_R(\mathbf{b}_{k_1+1}^{l_1}))
\end{aligned}$$

Define function $F : \mathcal{B} \to \mathbb{R}$ such that $F(\phi) = 1$ and

$$F(\mathbf{b}) = \sup_{m_{min}(\mathbf{b}) \le m \le m_{max}(\mathbf{b})} \sum_{r=r_l(\mathbf{b},m)}^{r_u(\mathbf{b},m)} \binom{k_1+m}{n_R(b_1^{k_1})+r} F(W(\mathbf{b}, m, r)),$$

where $W(\mathbf{b}, m, r)$, defined for $0 \le m \le l_1 - k_1, r_l(\mathbf{b}, m) \le r \le r_u(\mathbf{b}, m)$, is the vector obtained from $\mathbf{b}$ after the following modifications:
1. remove the first k_1 arrivals.
2. remove the first $k_1 + m$ deadlines.
3. remove the first r Rs and $m - r$ Bs in $\mathbf{b}_{k_1+1}^{l_1}$.

Theorem 1: Define $\tilde{F}(\mathbf{b}) \stackrel{\triangle}{=} \inf_{\tilde{\mathbf{b}} \in \mathcal{B}_*(\mathbf{b})} F(\mathbf{b})$, where $\mathcal{B}_*(\mathbf{b})$ is the set of restricted interlacing patterns that can be obtained from $\mathbf{b}$ by delaying deadlines. Then

$$\mathcal{A}(\lambda T) \le \mathcal{A}^u(\lambda T) = \frac{\mathbb{E}\{\log_2 \tilde{F}(\mathbf{b})\}}{\mathbb{E}\{n_R(\mathbf{b}) + n_B(\mathbf{b})\}},$$

where the expectation is over interlacing patterns in a cycle.
Proof: Available in [7]. □

Although characterizing the infimum in the definition of $\tilde{F}$ may, in general, be intractable, it is possible to characterize weaker upper bounds by considering specific transformations. The following theorem characterizes one such bound using a transformation that delays all deadlines to the end of the cycle.

Theorem 2:

$$\mathcal{A}^u(\lambda T) \le 1 - \sum_{n=1}^{\infty} H(W_n)(1-\beta)^{n-1}\beta^2 \ ,$$

where W_n are Binomial $(n, \frac{1}{2})$ random variables, and $\beta = e^{-2\lambda T}$.

Proof: Available in [7]. □

The transformation used in Theorem 2 results in a restricted pattern that contains only one arrival segment and one deadline segment. Therefore, the uncertainty is expressible using a single combinatorial expression without recursions. In general, using transformations that result in n arrival segments and evaluating the corresponding $n-$step recursions, the bound can be progressively improved as n increases. Numerical evaluation of bounds using multistep recursions are illustrated in Section V.

IV. Lower Bound

We obtain a lower bound on the anonymity by specifying a mixing strategy and computing its anonymity. The mixing strategy is motivated by Lemma 1 and its application to restricted interlacing patterns. Specifically, consider the start of a cycle that was initiated by the arrival of a blue packet at time t. The mix divides the time following the arrival of this packet into slots of length $\frac{T}{2}$. The division of time slots is continued until the first slot $(t+(S-1)T/2, t+ST/2]$, $S > 1$, which contains less than two arrivals. Following this slot, the mix waits for a new arrival to reinitiate the slot division.

For every packet that arrives in the i^{th} slot we advance the deadline to time $t + (i+1)\frac{T}{2}$. This effectively transforms the sequence of arrivals and deadlines into a restricted interlacing pattern, and owing to Lemma 1, the times $\{t + i\frac{T}{2}\}$ are the departure points. Further, there are at most two possible departure points for every arrived packet.

Let R_i, B_i denote the number of red and blue packets that arrived during the i^{th} slot. At the first departure point, the mix does not transmit any packets. Therefore, during the second slot, the mix contains R_1 red and $B_1 + 1$ blue packets in its queue. From the $R_2 + B_2$ packets that arrived in the second slot, the mix randomly chooses $M_2 = \lceil \frac{R_2+B_2}{2} \rceil$ packets. These M_2 packets are batched together with the R_1+B_1+1 packets in the queue and transmitted at the second departure point $(t + T)$. The M_2 packets can be chosen in $1 + \min\{R_2, B_2\}$ different ways, where the number of red packets varies between $\max\{0, M_2 - B_2\}$ and $\min\{R_2, M_2\}$. The mix randomizes the choice of M_2 packets such that the number of red packets chosen is uniformly distributed in $[\max\{0, M_2 - B_2\}, \min\{R_2, M_2\}]$.

The remaining $R_2 + B_2 - M_2$ packets are queued. At the third departure point, among the $R_3 + B_3$ newly arrived packets, $M_3 = \lceil \frac{R_3+B_3}{2} \rceil$ packets are similarly chosen and batched with the queued $R_2+B_2-M_2$ packets and transmitted as a batch. This continues until slot S, at which point the mix's queue empties and it waits for a new arrival to restart this process.

Define families of probability mass functions $\{P_{\lambda T}(R, B)\}$ and $\{P^*_{\lambda T}(R, B)\}$ as follows:

$$\begin{aligned}
P_{\lambda T}(r, b) &= e^{-\lambda T}\left(\frac{(\lambda T/2)^r}{r!}\right)\left(\frac{(\lambda T/2)^b}{b!}\right) \quad r, b \ge 0 \\
P^*_{\lambda T}(r, b) &= \frac{e^{-\lambda T/2}\frac{(\lambda T/2)^r}{r!}e^{-\lambda T/2}\frac{(\lambda T/2)^b}{b!}}{1 - e^{-\lambda T} - \lambda T e^{-\lambda T}} \quad r + b \ge 2.
\end{aligned}$$

$$f(R_1,B_1,R_2,B_2) = \sum_{r=\max(0,\lceil\frac{R_2+B_2}{2}\rceil-B_2)}^{\min(R_2,\lceil\frac{R_2+B_2}{2}\rceil)} \frac{1}{\min(R_2,B_2)+1}\log\left(\binom{R_1+B_1+1+\lceil\frac{R_2+B_2}{2}\rceil}{r+R_1}(\min(R_2,B_2)+1)\right)$$

$$g(R_1,B_1,R_2,B_2) = \frac{1}{\min(R_1,B_1)+1}\sum_{q_r=\max\{0,\lfloor\frac{R_1+B_1}{2}\rfloor-B_1\}}^{\min(R_1,\lfloor\frac{R_1+B_1}{2}\rfloor)} f\left(q_r,\lceil\frac{R_1+B_1}{2}\rceil-q_r-1,R_2,B_2\right)$$

Theorem 3: $\mathcal{A}(\lambda T) \geq \mathcal{A}^l(\lambda T)$ where

$$\mathcal{A}^l(\lambda T) = \frac{\rho\mathbb{E}(f(\mathbf{N}_1,\mathbf{N}_2)) + (1-\rho)\mathbb{E}(g(\mathbf{N}_1^*,\mathbf{N}_2))}{\rho(1+2\lambda T)+(1-\rho)\lambda T}, \quad (6)$$

$\mathbf{N}_2 \sim P_{\lambda T}$, $\mathbf{N}_1 \sim P_{\lambda T}$, $\mathbf{N}_1^* \sim P_{\lambda T}^*$ are independent random variables and $\rho = e^{-\lambda T}(1+\lambda T)$.

Proof: The lower bound is obtained by computing the anonymity of the mixing strategy described at the beginning of this section. After the arrival of a packet to an empty queue, the mix's queue would next become empty only when the number of arrivals in a slot is less than two. We shall refer to the time period starting from the arrival of a packet to an empty queue until the next time when a packet arrives to an empty queue as a *sub-cycle*. Each cycle (as defined in Section II-A) may contain multiple sub-cycles, but the start of a cycle always coincides with the start of a sub-cycle.

Let N' be the total number of packets that arrived in a particular sub-cycle. The mixing strategy results in an a posteriori distribution on the possible colourings of the N' departed packets in the sub-cycle. The entropy of this distribution, denoted by Γ, is a real valued measurable function defined on the space of realizations of the arrival and departure processes. The mixing strategy reduces the arrivals in a sub-cycle to a restricted interlacing pattern. Further, the counts of departures $\{M_i\}$ are deterministically related to the number of red and blue arrivals in the slots $\{(R_i,B_i)\} \stackrel{\triangle}{=} \eta$. We therefore write Γ as a function of η.

Due to the random arrival process, η, $\Gamma(\eta)$ and N' are random variables. It is easy to see that, with respect to the Palm distribution conditioned on the starting time of sub-cycles, η is i.i.d in every sub-cycle. Further, the randomness used by the mixing strategy in each sub-cycle is assumed independent. Therefore $\Gamma(\eta)$ and N' are also $i.i.d$ and the anonymity of this strategy is given by:

$$A(\lambda T) = \frac{\tilde{\mathbb{E}}^0(\Gamma(\eta))}{\tilde{\mathbb{E}}^0(N')},$$

where $\tilde{\mathbb{E}}^0$ denotes the Palm expectation over η conditioned on the starting times of sub-cycles. The rest of the proof involves the derivation of the probability distribution of colourings and the subsequent evaluation of Γ which are provided in [7]. □

V. Comparison of Bounds

The lower and upper bounds asymptotically (as $\lambda T \to \infty$) converge to the maximum value 1, and their first derivatives in light traffic are identical.

Theorem 4:

$$\lim_{\lambda T\to 0}\frac{d\mathcal{A}^l(\lambda T)}{d\lambda T} = \lim_{\lambda T\to 0}\frac{d\mathcal{A}^u(\lambda T)}{d\lambda T} = 1.$$

Proof: Available in [7]. □

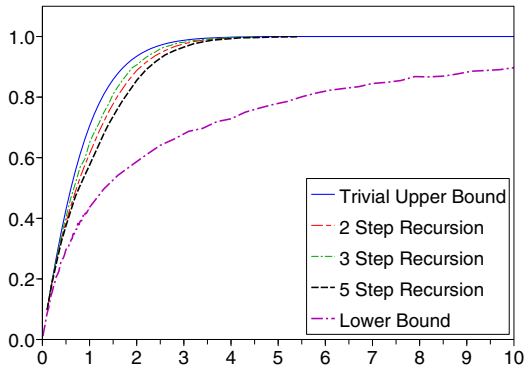

Fig. 4. Anonymity vs λT

Figure 4 plots the anonymity versus λT for the lower and upper bounds. The trivial upper bound is the bound corresponding to Theorem 2. The $n-$step recursion bounds are obtained by dividing each cycle into n slots, and delaying all deadlines within a slot to the end of the slot.

Acknowledgment

The research was partially supported by NSF grants CCF-0728872, CCF-0500023, CCF-0635372, and CNS-0627161.

References

[1] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 84–88, February 1981.

[2] D. Kesdogan, J. Egner, and R. Buschkes, "Stop-and-go MIXes providing probabilistic security in an open system," in *Second International Workshop on Information Hiding (IH'98), Lecture Notes in Computer Science*, vol. 1525, Portland, Oregon, April 1998, pp. 83–98.

[3] M. G. Reed, P. F. Syverson, and D. M. Goldschlag, "Anonymous connections and onion routing," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 482–494, May 1998.

[4] C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.

[5] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, 1949.

[6] J. Giles and B. Hajek, "An Information-Theoretic and Game-Theoretic Study of Timing Channels," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2455–2477, September 2002.

[7] http://acsp.ece.cornell.edu/members/Parv/ISIT08_Proofs.pdf.