Anonymity under light traffic conditions using a network of mixes

Parv 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 analysis of a multi-source single-destination network of mixes is considered under strict latency constraints at each mix. Mixes are relay nodes that accept packets arriving from multiple sources and release them after variable delays to prevent an eavesdropper from perfectly identifying the sources of outgoing packets (also, the contents of the packets are encrypted to prevent these from being used to correlate the arrivals to the mix with its departures). Using an entropy-based measure to quantify anonymity, the anonymity provided by such a single-destination network of mixes is analyzed, with the focus on light traffic conditions. A general upper bound is presented that bounds the anonymity of a single-destination mix network in terms of a linear combination of the anonymity of two-stage networks. By using a specific mixing strategy, a lower bound is provided on the light traffic derivative of the anonymity of single-destination mix networks. The light traffic derivative of the upper bound coincides with the lower bound for the case of mix-cascades (linear single-destination mix networks). Thus, the optimal light traffic derivative of the anonymity is characterized for mix cascades.**

I. Introduction

Mix networks, first proposed by David Chaum [1], are used extensively on the Internet to facilitate anonymous communication in applications such as e-mail and web browsing. Conceptually, mixes are relay nodes or proxy servers that accept packets from multiple nodes and output them in a manner that makes it infeasible for an external observer to determine the originating sources of transmitted packets perfectly. Specifically, a mix uses encryption techniques, random delaying, and reordering of packets to minimize the information available to an eavesdropper. As expected, delaying and reordering would increase the latency of transmitted packets. Alternatively, if mixes are subjected to hard delay constraints, then the *anonymity* achievable decreases. In [2], we studied the anonymity of a single mix under a hard delay constraint, and provided inner and outer bounds on the fundamental trade-off between latency and anonymity. In this work, we analyze the anonymity of a single-destination network of delay-constrained mixes, with the focus being on light traffic conditions.

Consider the example network as shown in Figure 1, where a set of source nodes are connected to a single destination through an *in-tree* network of mixes. The sources transmit packets according to independent Poisson processes. Each mix is allowed to delay packets arbitrarily subject to a maximum delay constraint, and the strategies of the mixes in the network are designed jointly. The eavesdropper, Eve, observes the transmission times of packets between every pair of nodes in the network. Since packets are encrypted, she cannot use the contents to determine the path of any packet except for the link it was observed on. Using the timing information in the observed point processes and her knowledge of the mixing strategy, Eve's goal is to determine the originating source of every packet arriving at the destination node.

Based on the analytical model formulated in [2], we quantify anonymity of mix networks using the entropy of the a posteriori distribution (from Eve's perspective) of originating sources of the packets arriving at the destination. Our goal is to investigate the maximum achievable anonymity as a function of the network topology and the delay constraints of the individual mixes. In this work, we analyze the anonymity of general single-destination mix networks, with the focus being on light traffic conditions.

Subsequent to the original design of a mixing strategy by Chaum, different low latency mix networks were designed for delay-limited applications [3]. However, a timing analysis of incoming and outgoing processes [4] exposed the vulnerability of such low-latency systems. Theoretical analyses of the anonymity of mix networks or the fundamental trade-off between anonymity and latency in mix networks are limited. The information-theoretic metric of anonymity of mix networks proposed in [5] treats every departing packet independently and does not take into account delay limitations and traffic statistics. In [2], we quantified anonymity in a single mix system using the complete observation of the eavesdropper, and provided bounds on the maximum achievable anonymity as a function of traffic load and the delay constraint. In this work, we analyze the general class of single-destination mix networks. We provide an upper bound for the anonymity achievable in such a network in terms of the anonymity of related two-stage networks, and provide a lower bound on the light traffic derivative of the anonymity for such networks by analyzing a specific mixing strategy. For the case of mix cascades (linear single-destination mix networks) the light traffic derivative of our upper bound coincides with our lower bound. A slightly related problem is the study of timing

channels with jammers [6] or spurious departures [7]. While the goal in [6], [7] is to analyze the timing information that can be relayed between a single source-destination pair in the presence of an adversary, our task is to obfuscate the source information of packets by multiplexing transmissions from multiple nodes.

The remainder of the paper is organized as follows. In Section II, we provide the mathematical formulation of the problem, and provide a brief recap of the light traffic derivative of single mix systems. In Section III, we present an upper bound on the anonymity of general single-destination mix networks. In Section IV, we provide a lower bound on the light traffic derivative of anonymity of general single-destination mix networks and prove its optimality for the class of linear networks (mix cascades). Some concluding remarks are made in Section V.

II. Problem Setup

A *single-destination mix network* is defined by a 3-tuple $\mathcal{M} = (G, D, \Lambda)$. $G = (V, E)$ is an *in-tree* directed graph, where the set of nodes V can be divided into a set of leaf nodes $\mathbf{S} = \{S_1, \cdots, S_s\}$ denoting the sources, a set of intermediate nodes $\mathbf{M} = \{M_1, \cdots, M_m\}$ denoting the mixes, and the root node R that represents the final destination. Without loss of generality, we let $M_m \in \mathbf{M}$ be the only node in the graph connected to R. $D = (d_1, \cdots, d_{|\mathbf{M}|})$ and $\Lambda = (r_1, \cdots, r_{|\mathbf{S}|})$ are vectors of positive integers such that $r_i\lambda$ denotes the arrival rate from source S_i, and d_jT denotes the delay constraint of mix M_j. λ and T are positive real constants. We partition the set of edges as $E = E_s \bigcup E_m \bigcup E_r$ where

$$\begin{aligned} E_s &= \{(A,B) \in E : A \in \mathbf{S}\}, \\ E_m &= \{(A,B) \in E : A, B \in \mathbf{M}\}, \\ E_r &= \{(M_m, R)\}. \end{aligned}$$

An example of a single-destination mix network is shown in Figure 1.

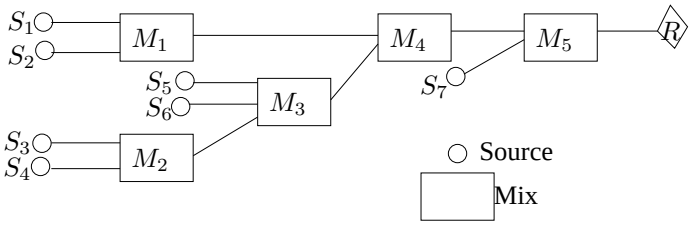

Fig. 1. A Mix Network:
$E_s = \{(S_1, M_1), (S_2, M_1), (S_3, M_2), (S_4, M_2), (S_5, M_3), (S_6, M_3), (S_7, M_5)\}, E_m = \{(M_1, M_4), (M_2, M_3), M_3, M_4), (M_4, M_5)\}$. $n_1 = 2, n_2 = 2, n_3 = 3, n_4 = 2, n_5 = 2$, $s = 7, s_1 = 2, s_2 = 2, s_3 = 4, s_4 = 6, s_5 = 7$

$\{\mathcal{Y}_e(t) : e \in E_s\}$, which represent the arrival processes, are modeled as independent stationary Poisson processes with their corresponding rates as specified by Λ. During the operation of the network, on each edge $(A, B) \in E$, a stream of packets is transmitted by node A to B, which is denoted by a point process $\mathcal{Y}_{(A,B)}(t)$. This point process need not be simple (i.e. batch transmissions are permitted).

Mix: Mix $M_i \in \mathbf{M}$ observes the processes $\{\mathcal{Y}_{(A,M_i)}(t) : (A, M_i) \in E\}$, which is the set of incoming streams of packets to the mix. The packets on any individual stream (A, M_i) have identical headers, and the contents do not reveal any information about the path of the packet prior to arriving at node A. Each mix has exactly one outgoing stream (as is evident from the tree structure of the network). Each arriving packet on $\{\mathcal{Y}_{(A,M_i)}(t) : (A, M_i) \in E\}$ may be delayed by mix M_i using a randomized strategy subject to causality and a maximum delay constraint of d_iT. A mix is allowed to transmit multiple packets in a single batch, in which case the order of packets within a batch does not matter. We assume that the mixes do not share any common randomness, but the strategies of the mixes can be jointly designed given complete knowledge of the network topology. Let $\Psi(\mathcal{M})$ denote the set of all valid mixing strategies for the network $\mathcal{M}$.

Eavesdropper: The eavesdropper, Eve, observes every individual point process in $\{\mathcal{Y}_e(t), e \in E\}$. As in the case of the mixes, the individual packets on each stream are indistinguishable to her. She is aware of the topology of the network and the mixing strategies of all the mixes, but does not have access to the realization of the private randomness used by each mix to implement its randomized mixing strategy. Using her complete knowledge, Eve's goal is to determine the original sources of the departed packets on the stream $\mathcal{Y}_{(M_m,R)}(t)$.

A. Anonymity

Consider the joint arrival process $\mathcal{Y}(t) = \bigcup_{e \in E_s} \mathcal{Y}_e(t)$ to the network. $\mathcal{Y}(t)$ is a stationary Poisson process of rate $\lambda \sum_i r_i$. We know that each mix is connected to the destination through exactly one directed path. For a mix M_k, let $M_k, M_{k_1}, \cdots, M_{k_n}, M_m$ be the sequence of mixes on the directed path from M_k to the destination. Then let $l_k = d_k + \sum_{i=1}^{n} d_{k_i} + d_m$. Note that l_kT is the maximum delay that can be experienced by a packet from its arrival time at mix M_k to its departure from the final mix M_m. Define

$$l_{max} \stackrel{\triangle}{=} \sup_{k \le m} l_k.$$

A packet transmitted by any source can be delayed in the mix-network by at most $l_{max}T$ seconds. Accordingly, the joint arrival process is divided into cycles of observation. Each cycle is initiated by the arrival of a packet on the stream $\mathcal{Y}(t)$ after a period of at least $l_{\max}T$ seconds of no arrivals. The cycle continues until the first time when there have been no arrivals for at least $l_{\max}T$ seconds, up until the time the next cycle is initiated by the arrival of a packet. Because of the delay constraints on the mixes, every packet that arrived in the cycle will depart the network before the cycle ends. Furthermore, the joint arrival process within each cycle is i.i.d, with respect to the Palm distribution relative to the starting time of cycles. (For the basic facts about Palm theory, see e.g. [8], [9].) Without loss of generality, we restrict ourselves to mixing strategies

that are Palm stationary from cycle to cycle. Then the joint realization of all the point processes $\{\mathcal{Y}_e(t), e \in E\}$ restricted to a cycle is Palm stationary from cycle to cycle. In this Palm stationary view, we focus on the cycle starting at time 0, i.e. the cycle initiated by the arrival of a packet at time 0 to the mix-network that has had no arrivals for at least $l_{max}T$ seconds (and is hence empty). The joint realization of all the point processes $\{\mathcal{Y}_e(t), e \in E\}$ restricted to this cycle is precisely the complete observation available to Eve over this cycle, and will be denoted by Θ. The underlying sample space on which Θ is defined is the one supporting the arrival processes and the individual private sources of randomness used by the mixes in implementing their randomized strategies. Let this sample space be denoted $(\Omega, \mathcal{F}, \mathbb{P}^0)$. We have used the notation $\mathbb{P}^0$ for the probability distribution to remind ourselves that we are talking about a the Palm distribution over cycles.

Let $N(\Theta)$ denote the total number of arrivals in the cycle starting at time 0, of which $N_i(\Theta)$ packets belong to source S_i. The number of departing packets on $\mathcal{Y}_{(M_m,R)}(t)$ during the cycle is thus also $N(\Theta)$. For each edge $e \in E$ we choose an ordering for the packets travelling over this edge during the cycle, i.e we choose a way to index the realization of $\{\mathcal{Y}_e(t), e \in E\}$ over this cycle. The choice of this indexing system is irrelevant; since the mixes are allowed to transmit packets in batches, we may also consider any particular ordering of packets departing in any batch over this edge. Given the realization in the underlying sample space (i.e. the realization of the arrival processes and the private randomness of the mixes) and a mixing strategy $\psi \in \Psi(\mathcal{M})$, the action of the mixes over this cycle can be viewed as determining Θ and then a sequence of bipartite matchings starting from the points corresponding to the source transmissions in the cycle through the points on the intermediate processes generated by the mixes until the points on the destination process $\mathcal{Y}_{M_m,R}(t)$ (restricted to this cycle)[1]. Given Θ, we define random variables $X_1, \cdots, X_{N(\Theta)}$, each taking values in $\{1, \ldots, s\}$, by working our way backwards through these bipartite matchings from the destination process $\mathcal{Y}_{M_m,R}(t)$ (restricted to this cycle), that is to say, since this process has $N(\Theta)$ points indexed in some way, we let $X_k(\Theta)$ denote the originating source of the k-th of these $N(\Theta)$ departing packets. The eavesdropper cannot, of course, determine the realization of $X_1, \cdots, X_{N(\Theta)}$. However, the joint distribution of $X_1, \cdots, X_{N(\Theta)}$, conditioned on Θ, under $\mathbb{P}^0$, is precisely the a posteriori distribution that the eavesdropper has over the originating sources of the departing packets over this cycle, conditioned on her available information Θ.

Let $\Gamma^\psi(\Theta)$ denote the entropy of the joint distribution of $(X_1, \cdots, X_{N(\Theta)})$.

Definition 1: The *anonymity* achieved by a mixing strategy $\psi \in \Psi(\mathcal{M})$ is defined as:

$$A^\psi_{\mathcal{M}}(\lambda T) = \frac{\mathbb{E}^0(\Gamma^\psi(\Theta))}{\mathbb{E}^0(N(\Theta))},$$

where $\mathbb{E}^0$ is the Palm expectation corresponding to the Palm distribution $\mathbb{P}^0$.

It is a standard fact in Palm theory that this quantity, even though defined in terms of the Palm distribution corresponding to a particular way of decomposing the dynamics into cycles, is really an ergodic quantity, i.e. other ways of decomposing the dynamics into cycles (as long as one can describe the a posteriori distribution of the eavesdropper over ways of associating departing packets to originating sources as a function of cycles) will result in the same quantity. For details, see the discussion on transfer formulas in [8]. This observation will be used in one of the later results.

We have given a very elaborate sample space description of the system prior to defining the anonymity of a policy, in order to clarify the subsequent discussion. This definition is identical to the one that was made for single mixes in [2], where it was not necessary to be so elaborate.

The maximum achievable anonymity in the mix-network $\mathcal{M}$ is given by:

$$\mathcal{A}_{\mathcal{M}}(\lambda T) = \sup_{\psi \in \Psi(\mathcal{M})} A^\psi_{\mathcal{M}}(\lambda T).$$

It is easy to see $0 \leq \mathcal{A}_{\mathcal{M}}(\lambda T) \leq \log(s)$.

The focus of this work is on understanding the maximum anonymity of $\mathcal{M}$ under light traffic conditions, which we discuss via the light traffic derivative:

$$\Delta_0(\mathcal{M}) = \lim_{\lambda T \to 0} \frac{d}{d\lambda T} \mathcal{A}_{\mathcal{M}}(\lambda T).$$

To be concrete, we will think of T as fixed and let $\lambda \to 0$.

In the following subsection, we briefly recap the optimal light traffic performance of a single-mix system from [2].

B. Recap: Single Mix

In [2], we analyzed the anonymity of a single mix with two arrival processes with equal rate λ. Note that this is a special case of a mix-network. From Theorem 3 of [2], we know that the light traffic derivative of a single mix system $\mathcal{M}_1$ with two equal rate sources is given by:

$$\lim_{\lambda T \to 0} \frac{d}{d\lambda T} \mathcal{A}_{\mathcal{M}_1}(\lambda T) = 1.$$

Using the same strategy and techniques used in [2], following is a straightforward generalization of the result to a single mix with k sources with unequal arrival rates.

Theorem 1: Let $\mathcal{M}_1^k = (G_1^k, D_1^k, \Lambda_1^k)$ be a mix-network with k sources transmitting packets at rates $r_1\lambda, \cdots, r_k\lambda$ respectively, and a single mix with delay constraint dT. Then,

$$\begin{aligned} \lim_{\lambda T \to 0} \frac{d}{d\lambda T} \mathcal{A}_{\mathcal{M}_1^k}(\lambda T) &= \frac{d}{\sum r_k} \left(\left(\sum_{j=1}^k r_k \right)^2 - \sum_{j=1}^k r_k^2 \right) \\ &\triangleq \Delta_0^1(\Lambda_1^k, d). \end{aligned}$$

[1]The requirement that the mixing strategy be causal and constructed by each mix based on its own private randomness of course restricts both Θ and the kinds of matchings that can result.

We now proceed to present our results on single-destination mix networks.

III. General Upper Bound

In this section we give an upper bound on the maximum achievable anonymity of a single-destination mix network, as a linear combination of the anonymity of smaller subnetworks. We will later use this upper bound to characterize the light traffic derivative for linear networks (mix cascades).

Let n_k denote the number of incoming edges to mix M_k, i.e., $n_k = |\{(A, M_k) : (A, M_k) \in E\}|$. Each packet stream observed by M_k contains packets from some set of sources, and these are mutually disjoint. Let $\mathcal{S}_{k,1}, \cdots, \mathcal{S}_{k,n_k}$ represent the respective sets of originating sources of packets on each incoming stream. Let the number of sources connected to M_k through the i^{th} incoming edge be $|\mathcal{S}_{k,i}| = s_{k,i}$ and the total number of sources connected to M_k be $\sum_{i=1}^{n_k} s_{k,i} = s_k$.

For each mix $M_k \in \mathbf{M}$ in network $\mathcal{M}$, we define an *auxiliary network* $\mathcal{M}_k = (G_k, D_k, \Lambda_k)$. G_k is obtained from G after the following modifications:

1. From graph G, remove all mixes that do not relay packets from sources[2] in $\bigcup_{i=1}^{n_k} \mathcal{S}_{k,i}$.
2. From graph G, remove all mixes that occur in the forward path from M_k to R and connect M_k to R by a single edge.
3. From graph G, for every incoming edge $e_{k,i}$ to M_k, replace the subgraph[3] that connects the sources in $\mathcal{S}_i$ to M_k by a single mix $M'_{k,i}$.
4. For every i, remove all the sources in $\mathcal{S}_{k,i}$ and their source edges, and replace them by a single source $S'_{k,i}$. If a source that was replaced was directly connected to M_k in $\mathcal{M}$, then the new source replacing it is also connected directly to M_k. Otherwise, $S'_{k,i}$ is connected to $M'_{k,i}$.

In the auxiliary network $\mathcal{M}_k$, Λ_k is an n_k length vector and the vector D_k has length no greater than $n_k + 1$. The arrival rate from source $S'_{k,i}$ is $s_{k,i}\lambda$. The delay constraint of mix $M'_{k,i}$ in $\mathcal{M}_k$ is the maximum total delay on a directed path in the replaced subgraph. The delay constraint of mix M_k in $\mathcal{M}_k$ is l_k. It is easy to see that $\mathcal{M}_k$ is also a single-destination mix network. Its anonymity, denoted $\mathcal{A}_{\mathcal{M}_k}(\lambda T)$, is given using Definition 1. See Figure 2 for an example of an auxiliary network.

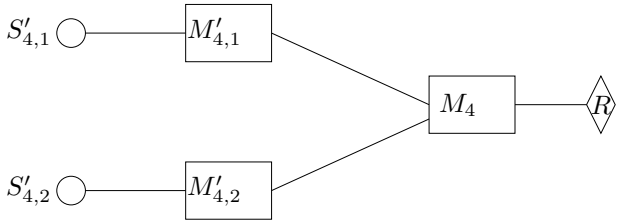

Fig. 2. Auxiliary network for mix M_4 in Figure 1

[2]Such a mix would not be connected to M_k in the unique path from it to the root.

[3]Note that, due to the in-tree nature of G, the corresponding subgraphs for the incoming edges to M_k are mutually exclusive

Theorem 2: For a single-destination mix network $\mathcal{M}$,

$$\mathcal{A}_{\mathcal{M}}(\lambda T) \leq \sum_{k=1}^{m} \mathcal{A}_{\mathcal{M}_k}(\lambda T)\frac{s_k}{s},$$

where $\mathcal{M}_k$ is the auxiliary network for mix M_k in $\mathcal{M}$.

Proof: Consider the set of incoming edges into the final mix M_m in $\mathcal{M}$, denoted by $e_{m,1}, \cdots, e_{m,n_m}$. For each edge $e_{m,i}$ that does not originate at a source, we define a *residual network* $\mathcal{M}_{m,i} = (G_{m,i}, D_{m,i}, \Lambda_{m,i})$ as follows: $G_{m,i}$ is the subgraph that connects the sources in $\mathcal{S}_{m,i}$ to M_m in G, such that $G_{m,i}$ includes $\mathcal{S}_{m,i}$ and node M_m. This subgraph $G_{m,i}$ would also be an in-tree graph with M_m as the root node. Let $M_{m,i}$ denote the final mix[4] in this residual network. The delay constraint of every mix in $\mathcal{M}_{m,i}$ is identical to that in $\mathcal{M}$, except for $M_{m,i}$, whose delay constraint in $\mathcal{M}_{m,i}$ is the sum of its delay constraint in $\mathcal{M}$ and d_m. The arrival rates of sources in $\mathcal{M}_{m,i}$ are identical to their arrival rates in $\mathcal{M}$. See Figure 3 for an example.

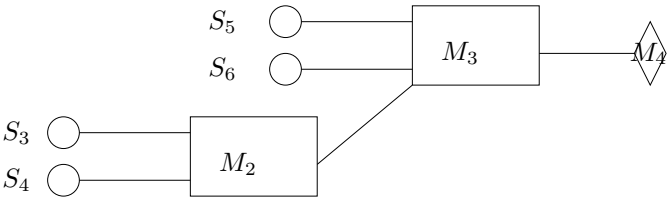

Fig. 3. Residual network for edge (M_3, M_4) in Figure 1

Lemma 1: For a single-destination mix network $\mathcal{M}$,

$$\mathcal{A}_{\mathcal{M}}(\lambda T) \leq \mathcal{A}_{\mathcal{M}_m}(\lambda T) + \sum_{i=1}^{n_m} \frac{s_{m,i}}{s}\mathcal{A}_{\mathcal{M}_{m,i}}(\lambda T).$$

where $\mathcal{M}_{m,k}$ is the residual network for edge $e_{m,k}$ in $\mathcal{M}$. (If $e_{m,i}$ originates at a source the expression $\mathcal{A}_{\mathcal{M}_{m,i}}(\lambda T)$ is interpreted as 0.)

Proof: As in the discussion leading up to the definition of anonymity, consider the cycle started by the arrival of a packet in $\mathcal{Y}(t)$ after a duration of length at least $l_{max}T$ without any arrivals. Given any mixing strategy $\psi \in \Psi(\mathcal{M})$, assumed Palm stationary without loss of generality, let Θ denote the observations of the eavesdropper over this cycle, with $N(\Theta)$ arrivals, of which $N_i(\Theta)$ packets are transmitted by source S_i. Let $X_1, \cdots, X_{N(\Theta)}$ be as defined in Section II-A; each X_k takes values in $\{1, \ldots, s\}$ and denotes the originating source of the k-the departure from the network, where the departures have been listed in some arbitrary manner. By definition, we have

$$A^{\psi}_{\mathcal{M}}(\lambda T) = \frac{\mathbb{E}^0(\Gamma^{\psi}(\Theta))}{\mathbb{E}^0(N(\Theta))}.$$

where $\Gamma^{\psi}(\Theta)$ may also be written as $H^{\psi}(X_1, \cdots, X_{N(\Theta)})$. It is important to note that in the latter expression the entropy is being calculated *after* conditioning on Θ, namely this expression denotes a random variable (which is a function of Θ).

[4]The final mix of $G_{m,i}$ is the node in G that is connected to M_m through edge $e_{m,i}$

Consider the final mix M_m in $\mathcal{M}$. Each edge $e_{m,i}$ to M_m contains packets from an exclusive set of sources $\mathcal{S}_{m,i}$, such that $\bigcup_{i=1}^{n_m} \mathcal{S}_{m,i}$ is the set of all sources $\mathbf{S}$. Define random variables $Z_1, \cdots, Z_{N(\Theta)}$ as deterministic functions of $X_1, \cdots, X_{N(\Theta)}$:

$$Z_i = j \text{ if } X_i \in \mathcal{S}_{m,j}. \tag{1}$$

Since Z_i is a deterministic functions of X_i we have

$$\begin{aligned} & H^{\psi}(X_1, \cdots, X_{N(\Theta)}) \\ = \; & H^{\psi}(Z_1, \cdots, Z_{N(\Theta)}) \\ & + H^{\psi}(X_1, \cdots, X_{N(\Theta)} | Z_1, \cdots, Z_{N(\Theta)}) \ , \end{aligned}$$

where this equation should be interpreted as holding after conditioning on Θ, i.e. it is actually an equality between random variables.

Lemma 2: For any mixing strategy ψ in $\mathcal{M}$,

$$\frac{\mathbb{E}^0(H^{\psi}(Z_1, \cdots, Z_{N(\Theta)}))}{\mathbb{E}^0(N(\Theta))} \leq \mathcal{A}_{\mathcal{M}_m}(\lambda T).$$

Proof: Refer to the Appendix □

Consider the term $H^{\psi}(X_1, \cdots, X_{N(\Theta)} | Z_1, \cdots, Z_{N(\Theta)})$. Since Z_is are deterministic functions of $X_1, \cdots, X_{N(\Theta)}$, any realization of the variables $Z_1, \cdots, Z_{N(\Theta)}$ would divide $X_1, \cdots, X_{N(\Theta)}$ into mutually exclusive sets $\{\{X_j : j \in I_i(\Theta)\} : i = 1 \cdots n_m\}$, such that $\forall j \in I_i(\Theta), Z_j = i$. Then, for any strategy ψ

$$\begin{aligned} & H^{\psi}(X_1, \cdots, X_{N(\Theta)} | Z_1, \cdots, Z_{N(\Theta)}) \\ \leq \; & \sum_{i=1}^{n_m} H^{\psi}(\{X_j, j \in I_i(\Theta)\} | Z_1, \cdots, Z_{N(\Theta)}) \ , \end{aligned}$$

where again this inequality should be thought of as holding after conditioning on Θ.

Lemma 3: For any mixing strategy ψ in $\mathcal{M}$

$$\frac{\mathbb{E}^0(H^{\psi}(\{X_j, j \in I_i(\Theta)\} | Z_1, \cdots, Z_{N(\Theta)}))}{\mathbb{E}^0(N(\Theta))} \leq \frac{s_{m,i}}{s} \mathcal{A}_{\mathcal{M}_{m,i}}(\lambda T).$$

Proof: Refer to the Appendix □

Using Lemmas 2 and 3, Lemma 1 is proved. □

The proof of Theorem 2 follows from a recursive application of Lemma 1 on the terms $\mathcal{A}_{\mathcal{M}_{m,i}}(\lambda T)$. □

IV. Light Traffic Derivative

Mix-Cascade: A *mix-cascade* is a special case of a single-destination mix network, where all mixes lie on the same directed path to the destination (see example in Figure 4).

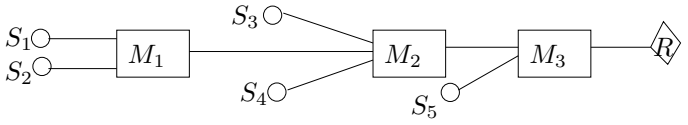

Fig. 4. Example of a Mix cascade

Theorem 3: For any single-destination mix network $\mathcal{M}$,

$$\Delta_0(\mathcal{M}) \geq \sum_{k=1}^{m} \frac{l_k}{s} \left(\left(\sum_{j=1}^{n_k} s_{k,j} \right)^2 - \sum_{j=1}^{n_k} s_{k,j}^2 \right).$$

If $\mathcal{M}$ is a mix-cascade:

$$\Delta_0(\mathcal{M}) = \sum_{k=1}^{m} \frac{l_k}{s} \left(\left(\sum_{j=1}^{n_k} s_{k,j} \right)^2 - \sum_{j=1}^{n_k} s_{k,j}^2 \right).$$

For a mix-cascade, note that the optimal light traffic derivative is equivalently expressible in the form:

$$\Delta_0(\mathcal{M}) = \sum_{k=1}^{m} \frac{s_k}{s} \Delta_0^1(\Lambda_k, l_k),$$

where Δ_0^1 is the single mix light traffic derivative (see Theorem 1), and Λ_k is the set of sources in the auxiliary network for M_k. Observe that the above linear form of the light traffic anonymity is identical to that in the upper bound of Theorem 2.

The theorem is proved by specifying a mixing strategy and characterizing its light traffic derivative. Since we are free to choose the strategy in this approach, we work with a strategy ψ that does not explicitly depend on λ, but does depend on T. In discussing the light traffic derivative, think of T as fixed and $\lambda \to 0$, so ψ is unambiguously defined.

Proof: Lower bound for general mix-networks For any strategy ψ, let Θ denote the information available to the eavesdropper over the cycle starting at time 0 in the Palm stationary view with respect to cycles. We have

$$A^{\psi}_{\mathcal{M}}(\lambda T) = \frac{\mathbb{E}^0(\Gamma^{\psi}(\Theta))}{\mathbb{E}^0(N(\Theta))} = \frac{\sum_{n=2}^{\infty} \mathbb{P}^0(N=n)\mathbb{E}^0(\Gamma^{\psi} | N=n)}{\mathbb{E}^0(N)}.$$

For a Poisson arrival process, it is easily shown that cycles with more than 2 packets do not contribute to the light traffic derivative, and $\Delta_0(\mathcal{M})$ is lower bounded as:

$$\begin{aligned} \Delta_0(\mathcal{M}) \; & \geq \; \lim_{\lambda T \to 0} \frac{d}{d\lambda T} \frac{\mathbb{E}^0(\Gamma^{\psi}(\Theta))}{\mathbb{E}^0(N(\Theta))} \\ & = \; \lim_{\lambda T \to 0} \frac{d}{d\lambda T} \left[\frac{\mathbb{P}^0(N(\Theta)=2)\mathbb{E}^0(\Gamma^{\psi}(\Theta) | N(\Theta)=2)}{\mathbb{E}^0(N(\Theta))} \right]. \end{aligned} \tag{2}$$

Our lower bound for the light traffic derivative is achieved by the following strategy, denoted by ψ_l. Mix M_k in $\mathcal{M}$ waits for an arrival after an idle period of at least $l_{\max} T$ seconds. All packets that arrive in the $d_k T-$second period following this arrival are transmitted along with this arrival in a single batch at the end of the $d_k T-$second period. During the $(l_k - d_k)T$ second period following this batched transmission, all packets that arrive to M_k are transmitted without any delay. At this point (*i.e.* $l_k T$ seconds following the initiating arrival), the mix resets and waits for a new arrival to restart this process. Note that the initial wait for $l_{\max} T$ seconds was merely an initialization step in the strategy.

Owing to (2), we can restrict our analysis of the strategy ψ_l to cycles with $N(\Theta) = 2$ packets. The maximum achievable

entropy in a $2-$packet cycle is $\Gamma = 1$, which occurs when the two packets belong to different sources and eventually depart in a single batch from the final mix M_m. Consider the following events defined with respect to the cycle initiated by a packet arriving at time 0 after a duration with no arrivals of length at least $l_{max}T$:

E_2: $N(\Theta) = 2$.
$E^a_{i,j}$: There is exactly one arrival each from S_i and S_j, with the packet from S_i initiating the cycle.
$E^{\psi_l}_{i,j}$: A packet from S_i and a packet from S_j depart in a batch from mix M_m, when strategy ψ_l is used.

We can write

$$\begin{aligned}\mathbb{E}^0(\Gamma^{\psi_l}(\Theta)|E_2) &= \sum_{i=1}^{s}\sum_{j\neq i}\mathbb{P}^0\{E^a_{i,j}\bigcap E^{\psi_l}_{i,j}|E_2\} \\ &= \sum_{i=1}^{s}\sum_{j\neq i}\mathbb{P}^0\{E^a_{i,j}|E_2\}\mathbb{P}^0\{E^{\psi_l}_{i,j}|E^a_{i,j},E_2\}\ .\end{aligned}$$

Since all sources transmit at equal rate, for $i \neq j$,

$$\mathbb{P}^0\{E^a_{i,j}|E_2\} = \frac{1}{s^2}. \tag{3}$$

Let $l_{i,k}T$ and $l_{j,k}T$ denote the total delay experienced by the packets from S_i, S_j respectively until they reach M_k, where M_k denotes the first mix at which the paths from S_i and S_j to the root meet.

Lemma 4:

$$\begin{aligned}\mathbb{P}^0\{E^{\psi_l}_{i,j}|E^a_{i,j},E_2\} &= \frac{\max\{0,(l_k+l_{i,k}-l_{j,k})T\}}{l_{\max}T} \\ &\quad -\frac{\max\{0,(-l_k+l_{i,k}-l_{j,k})T\}}{l_{\max}T} \\ &\quad +o(\lambda T)\ .\end{aligned}$$

Proof: Refer to the Appendix. □

Combining the terms for (i,j) and (j,i) in Lemma 4 and using (3), we can write

$$\begin{aligned}&\mathbb{E}^0(\Gamma^{\psi_l}(\Theta)|E_2) \\ &= \frac{1}{s^2}\sum_{k=1}^{m}\sum_{i=1}^{n_k}\sum_{j=1,j\neq i}^{n_k} s_{k,i}s_{k,j}\left(\frac{l_k}{l_{\max}}+o(\lambda T)\right) \\ &= \frac{1}{s^2}\sum_{k=1}^{m}\frac{l_k}{l_{\max}}\left(\left(\sum_{j=1}^{n_k}s_{k,j}\right)^2-\sum_{j=1}^{n_k}s^2_{k,j}\right)+o(\lambda T) \\ &\triangleq \Gamma_2+o(\lambda T).\end{aligned}$$

Using the properties of $M/D/\infty$ queues, we know that

$$\begin{aligned}\mathbb{P}^0(E_2) &= (1-e^{-sl_{\max}T})e^{-sl_{\max}T}, \\ \mathbb{E}^0(N(\Theta)) &= e^{sl_{\max}T}.\end{aligned}$$

Therefore, using (2), we can write

$$\begin{aligned}\Delta_0(\mathcal{M}) &\geq \lim_{\lambda T\to 0}\frac{d}{d\lambda T}\frac{(1-e^{-sl_{\max}T})(\Gamma_2+o(\lambda T))}{e^{2sl_{\max}T}} \\ &= sl_{\max}\Gamma_2 \\ &= \sum_{k=1}^{m}\frac{l_k}{s}\left(\left(\sum_{j=1}^{n_k}s_{k,j}\right)^2-\sum_{j=1}^{n_k}s^2_{k,j}\right)\end{aligned}$$

Upper bound for Mix-cascades: Using Theorem 2,

$$\Delta_0(\mathcal{M}) \leq \sum_{k=1}^{m}\frac{s_k}{s}\lim_{\lambda T\to 0}\frac{d}{d\lambda T}\mathcal{A}_{\mathcal{M}_k}(\lambda T).$$

Therefore, to prove the theorem, it is sufficient to show that:

$$\lim_{\lambda T\to 0}\frac{d\mathcal{A}_{\mathcal{M}_k}(\lambda T)}{d\lambda T} \leq \frac{l_k}{s_k}\left(\left(\sum_{j=1}^{n_k}s_{k,j}\right)^2-\sum_{j=1}^{n_k}s^2_{k,j}\right).$$

In a mix-cascade, each mix M_k has at most one incoming packet stream that does not arrive directly from a source node. Therefore, the auxiliary network $\mathcal{M}_k$ for any $k \leq m$ would contain at most two mixes (see Figure 5). If $\mathcal{M}_k$ contains only one mix, then the statement of Lemma 5 follows from Theorem 1.

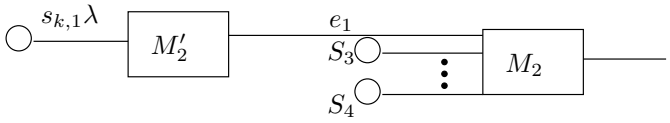

Fig. 5. Equivalent Network $\mathcal{M}_2$ for mix M_2 in Figure 2

When the auxiliary network $\mathcal{M}_k$ for mix M_k contains two mixes, let the penultimate mix be denoted by M'_k. The delay constraint of mix M'_k is $(l_{\max}T - l_kT)$, and that of mix M_k is l_kT. There is exactly one source connected to M'_k with transmission rate $s_{k,1}\lambda$, and there are n_k-1 sources connected directly to M_k with equal transmission rates λ. We label the sources $S'_1,\cdots,S'_k$ such that S'_1 is connected to M'_k.

Consider a modified definition of cycles in the joint arrival process to analyze the anonymity of this network. A cycle starts following an idle period of *exactly* $l_{\max}T$ seconds, and continues until the first time an idle period of exactly $l_{\max}T$ seconds occurs after the arrival of at least one packet from any of the sources $S'_2,\cdots,S'_{n_k}$. We refer to the first packet that arrives in a cycle from any of these sources as the defining packet of the cycle. According to this definition of a cycle, all packets that arrive prior to the defining packet would be from source S'_1, while packets arriving after the defining packet could be from any source. We consider the Palm stationary situation with respect to such cycles and let Θ denote the observation of the eavesdropper over the cycle that starts at time 0 after an idle period of duration exactly $l_{max}T$ (note that there is now no packet at 0, but there is some packet at time $-l_{max}T$). Whatever the strategy ψ of the two mixes, all packets arriving during this cycle must leave before the end of the cycle, so if we let $\Gamma^{\psi}(\Theta)$ denote the entropy of the a posteriori distribution of the eavesdropper over the originating

sources of the departing packets over this cycle when the mixes M_k and M'_k use strategy ψ, then the optimal light traffic derivative is given by:

$$\Delta_0(\mathcal{M}_k) = \sup_{\psi} \lim_{\lambda T \to 0} \frac{d}{d\lambda T} \frac{\mathbb{E}^0(\Gamma^\psi(\Theta)}{\mathbb{E}^0(N(\Theta))}.$$

The form of the upper bound in the statement of the theorem is reminiscent of the formula for the light traffic derivative of the anonymity in a single mix with multiple sources that was present in Theorem 1. This analogy can be understood by considering the mix M_k as receiving input flows all but one of which are Poisson, while one of the inputs is the output from mix M'_k. With the current definition of cycles one can view the portion of Θ for a duration $l_{max}T$ after the defining packet of the cycle as a Poisson process with rate $s_k\lambda$, the approximation becoming increasingly accurate in light traffic, irrespective of the strategy used by mix M'_k, while the portion of Θ for a duration $l_{max}T$ prior to the defining packet can be viewed as a Poisson process of rate $s_{k,1}\lambda$ (comprised only of packets from the source to mix M'_k), the approximation becoming increasingly accurate in light traffic, irrespective of the strategy used by mix M'_k. In light traffic it is still true that the contribution to the anonymity is dominated by cycles containing exactly two packets, so we may assume that one or the other situation obtains: either there is a packet ahead of the defining packet or there is one before the defining packet. The result comes from summing the resulting individual contributions, each of which has a form similar to that in Theorem 1. The details are available in [10].

V. Concluding Remarks

The main result in this paper is the characterization of the optimal light traffic derivative for the anonymity achievable by mix-cascades (linear single-destination mix networks). We also provide a lower bound on the light traffic derivative of the anonymity for general single-destination mix networks. The strategy used to prove this bound is, however, not always optimal. This can be explained by the fact that the proposed strategy does not incorporate the common information about absolute time available to the mixes. As opposed to the case of mix-cascades, this common time reference can be exploited by mixes working in parallel in a general single-destination mix network to get a light traffic derivative strictly better than the presented lower bound. An explicit example illustrating this phenomenon is available in [10]. Finally, we also presented an upper bound for the anonymity achievable in any single-destination mix network in terms of simpler networks. This can be combined with any upper bound for the anonymity of single mixes with multiple inputs to give an explicit upper bound on the anonymity of single-destination mix networks. For instance, a technique similar to that used in [2] in the case of two inputs can be used to generate such explicit upper bounds.

Acknowledgment

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

Appendix

Proof of Lemma 2

Consider the network $\mathcal{M}_m$, where the subnetwork connected to incoming edge $e_{m,i}$ is replaced by a single mix $M'_{m,i}$. There is a single source $S'_{m,i}$ that transmits packets to $M'_{m,i}$ at a rate equal to the total arrival rate in the replaced subnetwork. The maximum delay allowed for mix $M'_{m,i}$ is equal to the maximum delay that can be experienced by a packet within the replaced subnetwork. As a result, any mixing strategy employed by the mixes of the replaced subnetwork can be simulated by the single mix $M'_{m,i}$ in $\mathcal{M}_m$. Specifically, $M'_{m,i}$ can use its randomness to thin the single arrival process from $S'_{m,i}$ into multiple independent Poisson processes to simulate the multiple sources in $\mathcal{S}_{m,i}$. $M'_{m,i}$ then simulates the actions of the mixes in the replaced subnetwork and generates the corresponding intermediate point processes. This constitutes a potentially suboptimal mixing strategy in the mix network $\mathcal{M}_m$, hence its associated anonymity is at most $\mathcal{A}_{\mathcal{M}_m}(\lambda T)$. Note that in the mix network $\mathcal{M}_m$ the problem of the eavesdropper is that of associating the departures with the corresponding aggregated arrival processes.

We now visualize the situation where a genie provides to Eve the realization of the intermediate processes within each simulated subnetwork (including the artificial arrival processes created by thinning). Then Eve's net observation would be no different than what she would have observed in the original network $\mathcal{M}$. Therefore, the a posteriori distribution, at the aggregated level, of sources of departing packets in $\mathcal{M}_m$ conditioned on the genie information (in addition to the usual observations in $\mathcal{M}_m$) is identical to the distribution of $(Z_1, \cdots, Z_{N(\Theta)})$, where Θ now represents the overall information of Eve (i.e. the genie-provided information and the usual observation in $\mathcal{M}_m$) which corresponds one-to-one to observations that would have occurred in the network $\mathcal{M}$ (note that the cycle structure was defined purely in terms of the overall arrival process, so it does not change, since $l_{max}T$ is the same in both $\mathcal{M}$ and $\mathcal{M}_m$). When calculating the numerator term for the anonymity in the network $\mathcal{M}_m$ in the genie-aided case, we may first average over the information provided by the genie. Thus the contribution in the genie-aided case stands in relation to that in the non-genie-aided case as conditional entropy does to entropy. Since conditioning can only reduce entropy we conclude that the numerator term $\mathbb{E}^0(H^\psi(Z_1, \cdots, Z_{N(\Theta)}))$ is no bigger the numerator term in the computation of the anonymity for the given potentially suboptimal mixing strategy in $\mathcal{M}_m$, and since this in turn is no bigger than the optimal numerator term in the computation of anonymity of $\mathcal{M}_m$, this completes the proof. □

Proof of Lemma 3

Consider the network $\mathcal{M}_{m,i}$, the residual network corresponding to edge $e_{m,i}$ in $\mathcal{M}$. The delay constraint of the final mix in $\mathcal{M}_{m,i}$ is the sum of its delay constraint in $\mathcal{M}$ and the delay constraint of M_m in $\mathcal{M}$. Therefore, any mixing strategy in the original network $\mathcal{M}$ can be simulated by the network

$\mathcal{M}_{m,i}$ as follows. All mixes common to networks $\mathcal{M}$ and $\mathcal{M}_{m,i}$, except the final mix of $\mathcal{M}_{m,i}$, use identical strategies. The final mix of $\mathcal{M}_{m,i}$ uses its available randomness to simulate the excluded arrival processes, and the strategies of the mixes removed from $\mathcal{M}$ (to obtain $\mathcal{M}_{m,i}$). Note that any arrivals from the simulated arrival processes are just dummies and do not have any real existence.

We now imagine a genie which, under this simulated strategy, reveals to the eavesdropper the realizations of all the simulated excluded arrival processes and the simulated intermediate processes, and also, over the edge $(M_{m,i}, R)$, for each simulated departure, the identity of the simulated incoming link to the simulated mix M_m over which it arrived. We decide to analyze the network $\mathcal{M}_{m,i}$ under this strategy by using cycles defined in terms of an overall arrival process including the true arrivals to $\mathcal{M}_{m,i}$ and the simulated arrivals, with the minimal length of idle period determining the end of a cycle being $l_{max}T$, as in the original network $\mathcal{M}$. Then, for the cycle starting at time 0, in the Palm stationary view with respect to cycles, Eve's total observation (comprised of the information provided by the genie and the original observations in $\mathcal{M}_{m,i}$) would be no different from what she would have observed in the original network when strategy ψ was used, which we may represent by Θ (as in the original network), together with the information $(Z_1, \cdots, Z_{N(\Theta)})$.

Eve's problem, in the network $\mathcal{M}_{m,i}$, is that of associating to the points going over the edge $(M_{m,i}, R)$ their originating sources. When the potentially suboptimal simulation strategy above is used and Eve is genie-aided, then in computing this anonymity based on the above cycle structure, the corresponding numerator term is the Palm expectation of $H^{\psi}(\{X_j : j \in I_i(\Theta)\}|Z_1, \cdots, Z_{N(\Theta)})$. This stands in relation to the numerator term for the calculation of anonymity for this strategy with this cycle structure in the absence of the genie as conditional entropy does to entropy, and so is no bigger than the latter. The denominator term in both cases is $\mathbb{E}(N_i(\Theta))$, and so we get

$$\mathcal{A}_{\mathcal{M}_{m,i}}(\lambda T) \geq \frac{\mathbb{E}^0(H^{\psi}(\{X_j : j \in I_i(\Theta)\}|Z_1, \cdots, Z_{N(\Theta)})}{\mathbb{E}^0(N_i(\Theta))} \ , \tag{4}$$

where we have also observed that the proposed simulation based strategy is potentially suboptimal. Finally, since we assume all sources transmit packets according to independent Poisson processes of equal rate λ, each packet is equally likely to have arrived from any source. Therefore, the expected number of packets in a cycle which belong to a subset of sources with net rate $s_{m,i}\lambda$ is given by

$$\mathbb{E}^0(N_i(\Theta)) = \frac{s_{m,i}}{s}\mathbb{E}^0(N(\Theta)). \tag{5}$$

Combining (5) and (4), the lemma is proved. □

Proof of Lemma 4

Consider a two-packet cycle initiated at time 0 by a packet originating from S_i after an idle period of duration at least $l_{max}T$, and where the other packet is from S_j. Let M_k be the first mix where the paths from S_i and S_j to the destination node meet. Subsequent to mix M_k, their paths to the destination would be identical. In other words, M_k is the only common mix on their paths, where the packets from S_i and S_j arrive on different edges. According to ψ_l, in such a 2 packet cycle, the packets from S_i and S_j would be delayed by d_iT seconds at every mix M_i on their respective paths until they reach M_k. If and only if the delay between arrival times of the packets at M_k is within l_kT seconds, the packets would eventually depart in a batch from the final mix M_m.

Let $l_{i,k}T$ and $l_{j,k}T$ denote the total delay experienced by the packets from S_i, S_j respectively until they reach M_k. Let $\tau > 0$ denote the time of arrival of the second packet in the cycle, i.e. the packet from S_j. Then,

$$\begin{aligned} & \mathbb{P}^0\{E^{\psi_l}_{i,j}|E^a_{i,j}, E_2\} \\ = \ & \mathbb{P}^0\{|\tau + l_{j,k}T - l_{i,k}T| \leq l_kT \big| E^a_{i,j}.E_2\} \\ = \ & \frac{e^{-\max\{0,-l_kT+l_{i,k}T-l_{j,k}T\}} - e^{-\max\{0,l_kT+l_{i,k}T-l_{j,k}T\}}}{1-e^{-l_{\max}T}} \\ = \ & \frac{\max\{0,(l_k+l_{i,k}-l_{j,k})T\} - \max\{0,(-l_k+l_{i,k}-l_{j,k})T\}}{l_{\max}T} \\ & + o(\lambda T) \end{aligned} \tag{6}$$

This proves the lemma. □

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] P. Venkitasubramaniam and V. Anantharam, "On the anonymity of chaum mixes," in *2008 Proc. International Symposium on Information Theory*, Toronto, Canada, Jul. 2008.

[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] N. Levine, M. K. Reiter, C. Wang, and M. Wright, "Timing Attacks in Low-Latency Mix Systems," in *Proc. 8th International Conference on Financial Cryptography*, Key West, FL, Feb. 2004, pp. 251–265.

[5] C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*, San Francisco, CA, April 2002.

[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] R. Sundaresan and S. Verdu, "Capacity of Queues via Point Process Channels," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2697–2709, June 2006.

[8] F. Baccelli and P. Brémaud, *Elements of Queueing Theory*. New York: Springer, 2003.

[9] D. J. Daley and D. V. Jone, *An Introduction to the Theory of Point Processes*. New York: Springer, 1988.

[10] P. Venkitasubramaniam and V. Anantharam, "Anonymity under light traffic conditions of single-destination networks of mixes," 2008, in preparation.