Anonymous Networking with Minimum Latency in Multihop Networks

Parvathinathan Venkitasubramaniam and Lang Tong
Electrical and Computer Engineering
Cornell University, Ithaca, NY
{pv45,lt35}@cornell.edu†

Abstract

The problem of security against timing based traffic analysis in multihop networks is considered in this work. In particular, the relationship between the level of anonymity provided and the quality of service, as measured by network latency, is analyzed theoretically. Using an information theoretic measure of anonymity of routes in eavesdropped networks is considered, and packet scheduling strategies are designed to guarantee any desired level of anonymity. In particular, for individual relays, scheduling strategies based on mixing are designed so that the incoming and outgoing transmission epochs do not reveal any information. The proposed strategies utilize a limited fraction of dummy transmissions, and a significant reduction in packet latency at individual relays is demonstrated analytically for Poisson distributed arrivals. To minimize overall network latency, a randomized selection strategy is considered to choose the set of relays that use the designed scheduling strategies. The random selection is optimized for the desired level of anonymity using a well known distortion rate optimization in information theory. The tradeoff between overall network latency and anonymity in the network is characterized for centralized and decentralized scheduling strategies.

†This work is supported in part by the National Science Foundation under awards CCF-0635070 and CCF-0728872, and the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011.

1 Introduction

Anonymous networking refers to communication on a network without revealing the source-destination pairs or the paths of traffic flow. While contents of a message can be protected using encryption, hiding the act of communication requires a redesign of underlying network protocols. Changes in communication protocols can affect the quality of service in a network and it is necessary to minimize the loss in network performance while providing anonymity. In particular, protection against information retrieval from arrival and departure times of packets requires modification to the transmission schedules of nodes, which in turn increases network latency. In this paper, we consider the design of node transmission schedules that provide anonymity to network routes with minimum increase in network latency. In particular, we are interested in characterizing the tradeoff between the level of anonymity that can be provided and the average latency incurred in a multihop network.

Anonymous communication systems have typically been designed using Chaum Mixes [1]. A Mix is a node or server that collects packets from multiple users and outputs them in a manner that makes it infeasible to correlate an outgoing packet with a unique incoming packet. Specifically, a Mix performs re-encryption and packet padding to obfuscate the contents of each packet. Further, the Mix also changes the timing pattern of arrived packets by reordering and batching packets from multiple users together. According to the original batching strategy as proposed by Chaum, the Mix waits until at least one packet arrives from n different users before transmitting them all together in a batch. As is evident, the delay incurred by this Mix is potentially unbounded. Although improved batching strategies have been designed,

978-0-7695-3168-7 /08 $25.00 © 2008 IEEE
DOI 10.1109/SP.2008.18

notably by maintaining a pool of packets and flushing a fraction of them periodically [2, 3], the delay incurred due to the mixing strategies has not been analyzed or optimized. In fact, many of the known batching strategies were found to be vulnerable to *flow correlation* when long streams of packets were transmitted under strict delay constraints [4].

A common technique used in low latency anonymous communication is the transmission of dummy packets to "cover" the actual flow of traffic. Systems that use this approach such as ISDN Mixes [5] and Web-Mixes [6] require users to maintain a constant transmission rate of packets irrespective of whether they have actual data to communicate or not. This ensures that, to an external eavesdropper, the observed pattern of traffic is fixed irrespective of the routes of communication. A similar approach was also considered for a wireless multihop network in [7], where bounds were derived on the efficiency of using a fixed transmission schedule. Although fixed scheduling ensures complete anonymity, the high rate of dummy transmissions required makes it energy inefficient and unattractive for large networks. In the context of bandwidth constrained multihop networks, the following questions are yet to be addressed, particularly from a theoretical perspective. If the fraction of dummy transmissions were to be fixed, what is the minimum delay incurred at a Mix? If overall network latency were to be bounded, what is the maximum anonymity that can be achieved? More generally, what is the relationship between the achievable anonymity and network latency?

In this work, we address these issues using a theoretical foundational approach with emphasis on wireless multihop networks. The key to answering these questions is to quantify the anonymity achievable in a multihop network. Metrics of anonymity that have been proposed [8, 9] in the context of Mix networks are typically based on *anonymity sets* of individual packets. The anonymity set refers to the collection of all possible source-destination pairs of an observed packet. While these metrics quantify the anonymity provided by Mixes to individual packets, they do not apply to streams of packets and cannot be used to measure the overall anonymity of routes in the network. The approach we adopt is motivated by information-theoretic secrecy pioneered by Shannon through the concept of equivocation [10]. Equivocation has subsequently been used to measure the secrecy of messages transmitted over channels such as wiretap channel [11] and broadcast channels [12], where the goal was to maximize reliable information rate while providing a given level of secrecy. We use equivocation to measure the anonymity of the routes in a network, and the problem we address is to minimize network latency while guaranteeing a given level of anonymity.

1.1 Main Contributions

The goal of this work is to design scheduling strategies that minimize overall network latency for any desired level of anonymity. Assuming a passive omniscient adversary who has access to packet headers and transmission epochs* of all nodes in the network, we measure the anonymity of the network routes using information-theoretic equivocation. We then propose scheduling strategies for intermediate relays in the routes, and characterize the relationship between achievable anonymity and the incurred network latency. Our mathematical model and approach is based on our previous work in [13] where we considered optimizing the throughput-anonymity tradeoff in wireless networks under specific medium access models.

Our scheduling strategy has two important design components. First, for any individual node that relays packets from multiple users, we propose modifications to existing batching strategies of Mixes by adding a fixed rate of dummy transmissions; we demonstrate a significant reduction in average packet delay without compromising anonymity of the relay. In particular, when arrival schedules are distributed as Poisson processes, we provide an analytical characterization of the relationship between the incurred latency at a relay and the rate of dummy transmissions required. Second, depending on the level of anonymity required, we choose a random subset of relays to use the designed batching strategies, while the remaining relays do not modify their schedules and hence do not increase network latency. We optimize the distribution of the random subset of relays such that the overall network latency is minimized while achieving the desired level of anonymity. In particular, this optimization is shown to be equivalent to a distortion-rate optimization in information theory.

*The terms transmission epoch or transmission time refers to the time point of transmission of a packet.

1.2 Other Related Work

Subsequent to the original work by Chaum, the concept of Mixing has been successfully utilized in designing anonymous remailers such as Mixmaster and Mixminion [2, 14], and anonymous low-latency systems such as Tor [15]. From a system design perspective, Mix based anonymous systems broadly fall under two categories: mix-cascades and peer-to-peer systems. In a Mix-cascade, a dedicated set of servers are employed to mix traffic flows, and every packet is transmitted through the predefined set of Mix servers until it reaches the intended destination. Examples of Mix-cascade systems include JAP [6] and Reliable [16]. A peer-to-peer system does not have dedicated Mix servers, and every user independently mixes incoming traffic. The routes are therefore not predetermined at sources. Freenet [17] and Tarzan [18] are examples of peer-to-peer anonymizing systems. The approach we adopt is similar to peer-to-peer systems although we do not consider the optimal design of routes to maximize anonymity. The advantages of one approach versus the other are well summarized in [19]. Some anonymizing systems that do not use Mixes include DC-nets [20] and Crowds [21].

2 System Model

We represent the multihop network using a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes in the network and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed links. If $(A, B) \in \mathcal{E}$, then node B can receive transmissions from node A. A sequence of nodes $P = (V_1, \cdots, V_n) \in \mathcal{V}^*$ is a *valid path* in $\mathcal{G}$ if $(V_i, V_{i+1}) \in \mathcal{E}$, $\forall i < n$.

We assume that during any network observation by the eavesdropper, a subset of nodes communicate using a fixed set of paths. We call this set of paths $\mathbf{S} \in (\mathcal{V}^*)^*$ a network *session*. We use the notation $|\mathbf{S}|$ to denote the number of paths in session $\mathbf{S}$, and $|P|$ to denote the number of nodes (including the source and destination) in path P. The information that is to be hidden from the eavesdropper is the network session $\mathbf{S}$. Let $\mathcal{S}$ denote the set of all possible sessions which is a subset of $(\mathcal{V}^*)^*$. For example. consider the network $\mathcal{G}_1$ shown in Figure 1. Let S_1, S_2 always be the sources and D_1, D_2 the destinations. Further, let S_1, S_2 always communicate with distinct destinations. In that case, the set of all possible sessions is given by

$$\mathcal{S} = \Big\{ \begin{array}{l} \{(S_1, B, D_1), (S_2, B, D_2)\} \\ \{(S_1, B, D_2), (S_2, B, D_1)\} \Big\}. \end{array}$$

We model $\mathbf{S}$ as an i.i.d. random variable $\mathbf{S} \sim p(\mathbf{S})$. During any period of observation by an eavesdropper, she observes a random session (drawn i.i.d from $p(\mathbf{S})$) where the paths do not change. We assume that the eavesdropper is aware of the prior $p(\mathbf{S})$, which would aid in determining the session. In the example network $\mathcal{G}_1$, if the probability of one session were much higher than the other, then the uncertainty of the eavesdropper would be considerably smaller than if each session were equally likely. This intuition will be reflected in our definition of anonymity.

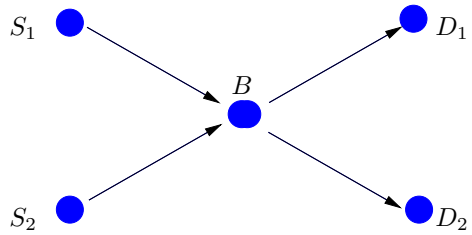


Figure 1: Two Node Switching Network: $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E})$, $\mathcal{V} = \{S_1, S_2, B, D_1, D_2\}$, $\mathcal{E} = \{(S_1, B), (S_2, B), (B, D_1), (B, D_2)\}$.

In this work, we do not consider the design of routes between sources and destinations. We assume that the set of possible sessions are given, and design transmission strategies for nodes so that an eavesdropper cannot determine the session $\mathbf{S}$. Note that although we assume a fixed set of paths in each session, this does not discount the possibility of having multiple paths between nodes.

Eavesdropper Observation: We consider a global passive eavesdropper who observes the packet transmission times of each node during a session. We assume that packet headers are decoded by the eavesdropper, and for each observed packet, the headers specify the transmitting and receiving nodes [†]. On each link (A, B), the eavesdropper observes a sequence of transmission times:

$$\tau_{A,B} = (T_{A,B}(1), T_{A,B}(2), \cdots),$$

where $T_{A,B}(i)$ is the transmission time of the i^{th} packet

[†]We assume a system similar to Mix networks where layered encryption would ensure headers only reveal links and not end-to-end routes

from node A to node B. Since we assume a global adversary, she has access to all the transmission schedules:

$$\tau = \{\tau_{A,B} : (A,B) \in \mathcal{E}\}.$$

Let $\mathbf{L}$ denote the set of all links observed by the eavesdropper. It is easy to see that $\mathbf{L}$ is a deterministic function of the session $\mathbf{S}$.

Note that $\mathbf{L}$ provides a minimum amount of information about the session, unless all sessions have an identical set of active links. To make any further inference about the end-to-end routes in the session, an eavesdropper would need to correlate transmission schedules in τ across multiple links. We model τ as a sequence of random variables with conditional distribution $q(\tau|\mathbf{S})$. The goal is to design $q(\tau|\mathbf{S})$ such that the eavesdropper obtains minimum information about the session $\mathbf{S}$ by observing τ.

2.1 Anonymity Measure

We define anonymity using equivocation [10], which is the conditional uncertainty of the information we wish to hide ($\mathbf{S}$) given the observation of the eavesdropper ($\tau, \mathbf{L}$).

Definition 1 *A distribution $q(\tau|\mathbf{S})$ is defined to have anonymity α if*

$$\frac{H(\mathbf{S}|\tau,\mathbf{L})}{H(\mathbf{S}|\mathbf{L})} \geq \alpha,$$

where $H(X|Y)$ is the entropy of random variable X given random variable Y:

$$H(X|Y) = -\mathbb{E}(\log p(X|Y)).$$

The normalized metric lies between $[0,1]$; a value of $\alpha = 0$ denotes no anonymity and $\alpha = 1$ denotes maximum anonymity. To understand the physical meaning, consider the maximum value $\alpha = 1$, which implies that

$$H(\mathbf{S}|\tau,\mathbf{L}) = \mathbf{H}(\mathbf{S}|\mathbf{L}).$$

$H(\mathbf{S}|\mathbf{L})$ is the uncertainty in the session if the eavesdropper ignored the knowledge of transmission schedules, and was to guess the session using the set of observed links $\mathbf{L}$ and the prior probability $p(\mathbf{S})$. Therefore, $\alpha = 1$ implies that knowledge of the schedule τ does not provide any additional information about the routes in the session.

For a general α, the physical interpretation comes from Fano's Inequality [22], which shows that the error probability is lower bounded by a monotonic function of equivocation. Specifically, if the error probability of the eavesdropper in identifying the session $\mathbf{S}$ is denoted by P_e, then,

$$P_e \geq \frac{H(\mathbf{S}|\tau,\mathbf{L}) - \mathbf{1}}{\log|\mathcal{S}|} \geq \frac{\alpha H(\mathbf{S}|\mathbf{L}) - \mathbf{1}}{\log|\mathcal{S}|},$$

where $\mathcal{S}$ is the set of all possible sessions. Since $H(\mathbf{S}|\mathbf{L})$ and $\log|\mathcal{S}|$ are constants, a higher level of anonymity implies a higher error probability for the eavesdropper.

Conditional entropy has been used previously to measure anonymity [8, 9] in the context of Mix networks. In [8, 9], the authors used equivocation to measure the uncertainty of source-destination pairs of individual packets. Since the measure did not cater to streams of packets where inter-packet timing reveals significant information, it cannot be extended to measure the overall anonymity of the routes in the network. To the best of our knowledge, ours is the first definition of anonymity of network routes based on streams of transmitted packets. The defined metric assumes an omniscient adversary and also does not take into account the possibility of active compromising of nodes. In Section 4.1, we discuss possible methods to extend this model to incorporate constrained adversaries and compromised nodes.

2.2 Relaying Strategy

The set of transmission schedules τ at most specifies when packets were transmitted between successive nodes in the path and do not indicate which packets were relayed from source to destination. In fact, some of the transmission times would represent dummy transmissions. Therefore, in addition to τ, we provide a set of schedules τ^R that indicate the transmission times of the relayed data packets. Note that τ^R would be a subset of τ and is not available to the eavesdropper. τ^R is indexed using the set of routes in the session and will be referred to as the relaying strategy.

Consider a session $\mathbf{S} = \{P(1), \cdots, P(|\mathbf{S}|)\}$, and where $P(i) = (A(i,1), A(i,2), \cdots, A(i,|P(i)|))$. The relaying strategy is denoted by $\tau^R = \{\tau^R_{i,j}\}$ where $1 \leq i \leq |\mathbf{S}|$ and for every i, $1 \leq j < |P(i)|$. Each sequence $\tau^R_{i,j} = (T^R_{i,j}(1), T^R_{i,j}(2), \cdots)$ represents the transmission times of *data packets* between nodes $A(i,j)$ and $A(i,j+1)$ on path $P(i)$. Therefore, for a path $P(i)$, the

sequence $\{T^R_{i,1}(n), T^R_{i,2}(n), \cdots, T^R_{i,|P(i)|-1}(n)\}$ denote the times when the n^{th} packet in the stream was transmitted by the nodes $\{A(i,1), \cdots, A(i,|P(i)|-1)\}$ respectively.

Given transmission schedule τ and session $\mathbf{S}$, a relaying strategy τ^R is valid iff it satisfies the following conditions:

1. $\forall i \leq |\mathbf{S}|, 1 < j \leq |P(i)|, \tau^R_{i,j} \subseteq \tau_{A(i,j),A(i,j+1)}$.
2. If $(A(i,j), A(i,j+1)) = (A(l,m), A(l,m+1))$, then $\tau^R_{i,j} \cap \tau^R_{l,m} = \phi$.
3. For every $i \leq |\mathbf{S}|$, $\{\tau^R_{i,j} : j < |P(i)|\}$ satisfy

$$T^R_{i,j+1}(n) - T^R_{i,j}(n) \geq 0. \quad (1)$$

Condition 2 states that if two paths share a common pair of nodes, then the relaying strategy for each of the paths should pick mutually exclusive subsets of τ. In other words, each element of the transmission schedule represents a single transmitted packet[‡]. Condition 3 is a causality condition; it ensures that a packet cannot be relayed by a node prior to its arrival time.

Since we allow relaying nodes to transmit dummy packets, the set of schedules τ^R would be a proper subset of τ. The transmission epochs in τ that are not in τ^R represent dummy transmissions. For a given transmission schedule τ, and a relaying strategy τ^R, we characterize the fraction of dummy packets transmitted by each relay and the average latency of packets from source to destination as follows.

Dummy transmission rate: If $A(i,j)$ represents the i^{th} node of path $P(j)$ in session $\mathbf{S}$, then the dummy transmission rate from node V_1 to node V_2 during session $\mathbf{S}$ is given by:

$$d_{V_1V_2} = \frac{|\tau_{V_1,V_2}| - \sum_{i,j:(A(i,j),A(i,j+1))=(V_1,V_2)} |\tau^R_{ij}|}{|\tau_{V_1,V_2}|}.$$

Latency Overhead: If the length of the packet stream on path $P(i)$ is n packets, then the average latency overhead on path $P(i)$ is the sum of delays incurred at each relay in $P(i)$ due to the scheduling strategy, given by:

$$\begin{aligned} \delta(P(i)) &= \frac{1}{n} \sum_{j=2}^{|P(i)|-1} \sum_{k=1}^{n} (T^R_{ij}(k) - T^R_{i,j-1}(k)) \\ &= \frac{1}{n} \sum_{k=1}^{n} (\tau^R_{i,|P(i)|-1}(k) - \tau^R_{i,1}(k)). \quad (2) \end{aligned}$$

[‡]If multiple packets are transmitted together in a batch, then they need to be represented as distinct elements in τ with an ϵ.

Note that the latency overhead does not include transmission delays and only measures the latency incurred due to the scheduling and relaying strategies. The overall latency would include transmission delays and is taken into consideration in the definition of the quality of service metric.

2.3 Network Latency

The performance metric we wish to optimize in this work is the average network latency (average end-to-end delay per session). If each node were to transmit a high rate of dummy transmissions, then anonymity can be provided with minimum increase in network latency. Such a strategy is however impractical in bandwidth constrained network. We therefore define our metric as the minimum latency incurred subject to a constraint on the maximum allowed fraction of dummy packets.

Definition 2 *Let $\Delta_t(P(i))$ denote the latency incurred on path $P(i)$ due to transmission delays. Then, $\Delta(\alpha, \lambda)$ is defined to be an* **achievable latency with anonymity α and dummy transmission rate** *λ if there exists $q(\tau|\mathbf{S})$ with anonymity α such that*

1. *For every realization of $(\mathbf{S}, \tau)$, there exists a valid relaying strategy τ^R that satisfies*

$$\begin{aligned} \forall (A,B) \in \mathcal{E},\ d_{AB} &\leq \lambda, \quad (3) \\ \mathbb{E}\left(\sum_{i=1}^{|\mathbf{S}|} \delta(P(i)) + \Delta_t(P(i))\right) &\leq \Delta, \quad (4) \end{aligned}$$

where the expectation is over the joint pdf of τ and $\mathbf{S}$.

A parameter that we have not considered so far is the rate of transmission of packets by each node. Since different routes could have different rates of transmission, it may be possible for an eavesdropper to use the rates of transmission to infer additional information about the session. To rectify this problem, we fix a minimum rate of transmission and impose a restriction that the rate on each route be an integral multiple of the minimum rate. Further, if the rate on a particular route is k times the minimum rate, we treat them as k separate routes. Therefore, every route according to the redefinition would have equal rate. Note that since this division would be reflected in the definition of the sessions, the anonymity condition does not need to be modified.

In the subsequent sections, we design scheduling and relaying strategies assuming unit arrival rate on all paths, and optimize the achievable network latency Δ for any desired level of anonymity α and dummy transmission rate λ.

3 Scheduling Strategy

Our approach to designing scheduling algorithms for multihop networks is motivated by anonymous peer-to-peer systems [18], where each node, apart from transmitting its own data packets acts as an intermediate relay that mixes incoming traffic from other nodes. In the mixing approach, every intermediate node in a route would use batching strategies to modify the timing pattern of arriving packets, thereby adding to the overall network latency. However, this would not be necessary, and depending on the level of anonymity required, it will be sufficient for a smaller subset of nodes to modify transmission schedules using batching strategies while the remaining nodes relay packets as and when they arrive. In other words, it is possible to "reveal" some portions of the routes without violating the anonymity constraint. This is a key intuition that we exploit in leveraging latency for anonymity.

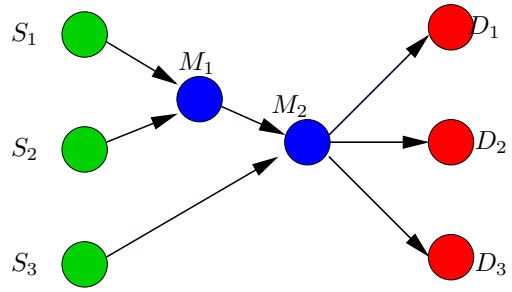


Figure 2: Example: Sources S_i transmit packets to destinations D_j through M_1, M_2

Consider the example in Figure 2, where sources S_1, S_2, S_3 are equally likely to transmit to destinations D_1, D_2, D_3. If both the intermediate relays M_1, M_2 acted as Mixes by employing batching algorithms, the network would have maximum anonymity, since packet timing would not reveal any information about source-destination pairs. Since M_1 and M_2 modify the timing pattern of the packet streams, the net overhead in latency for the routes from S_1, S_2 would be the sum of batching delays at M_1 and M_2. It is however easy to see that since M_2 mixes the flows from all three sources, allowing node M_1 to relay packets without modifying transmission schedule would not compromise the anonymity. In that case, the total latency overhead can be reduced (since only M_2 contributes to the delay). In more general terms, by choosing the optimal set of relays to modify their transmission schedules (henceforth referred to as *covert relays*), overhead in latency can be minimized without reducing anonymity.

Our strategy involves two fundamental design problems: design of scheduling strategy for a covert relay and the optimal selection of relays to be covert in a session. The scheduling strategy designed for a covert relay should ensure that given an outgoing stream of packets, every incoming stream is equally likely to have been the source of packets. The design is however limited due by the fraction of dummy transmissions allowed. The optimal selection of covert relays depends on the routes of the session, the level of anonymity required, and the delay incurred at each covert relay. We propose a randomized selection strategy, where the set of covert relays are chosen as a random function of the session and the desired level of anonymity. We then optimize the random distribution to obtain minimum latency for the desired level of anonymity.

In the remainder of this section, we describe the scheduling strategy for a covert relay and characterize the delay incurred at a single relay given the fraction of dummy transmissions allowed. In Section 4, we optimize the selection strategy and characterize the relationship between anonymity and the achievable network latency.

3.1 Covert Relaying

The task of the covert relay is to obfuscate the departure times of arriving packet streams, so that by analyzing arrival and departure times of packets, an eavesdropper is incapable of identifying a particular input-output pair accurately. Consider a relay as shown in Figure 3. Given the transmission times of packets on the links $\{(S_i, B)\}$ and $\{(B, D_i)\}$, every path $\{(S_i, B, D_j)\}$ should be equally likely. From the definition of the relaying strategy, we know that the design is subject to the following conditions:

1. The relaying strategy should be causal (as given in (1)).
2. Data packets cannot be dropped.
3. The maximum dummy transmission rate is λ.

The strategy we propose is a modification of the stan-

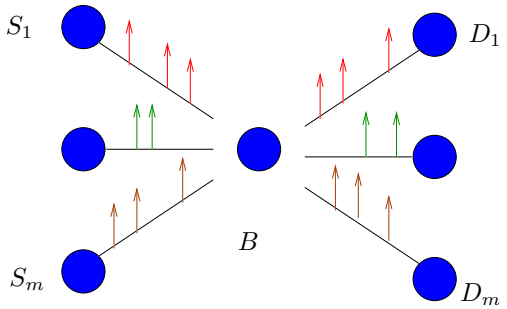


Figure 3: $m \times 1$ Relay Node: Sources S_i transmits packets to D_i through B

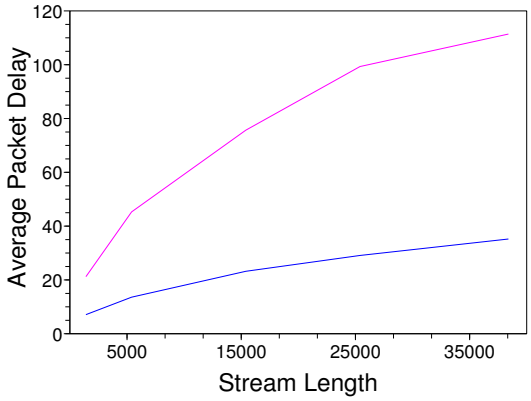


Figure 4: Average per packet delay for a relay node using simple threshold mixing strategy on two packet streams without dummy transmissions.

dard batching strategy of Mixes using a fixed additional rate of dummy transmissions. The need for introducing dummy packets can be illustrated using the following example. Consider a mix using the standard batching strategy on packets arriving from two sources; the mix waits until one packet arrives from both sources before transmitting them together. If arrivals are distributed as independent Poisson processes, the departure process for each stream is equivalent to that of an $M/M/1$ queue with arrival rate λ and service rate λ. For an $M/M/1$ queue, the mean waiting time is given by $\frac{1}{\lambda-\mu}$ where λ is the arrival rate and μ is the service rate. Therefore, when packets from two sources arrive at equal rates, the expected delay of transmitted packets would increase indefinitely as the length of the packet stream increases. This can be observed in Figure 4, where the average packet delay versus the length of the packet stream is plotted for Poisson and Pareto distributed schedules, when the standard batching strategy is applied.

In the following exposition we design scheduling strategies which demonstrate that by appropriately including dummy transmissions, the average delay can be reduced significantly and the maximum packet delay can be bounded, even for an infinite stream of packets.

2×1 Relay: Consider a relay node forwarding packets from 2 sources (Figure 3 with $m = 2$). If a packet from source 1 arrives to an empty relay, it waits until a packet arrives from source 2 for a maximum of Δ^* seconds. If a packet arrives from source 2 before the Δ^*-second period expires, then the two packets are randomly reordered and transmitted together in a batch. If no packet arrives from source 2 before the Δ^*-second period expires, then a dummy packet is generated and transmitted along with the queued packet in a batch. For the remainder of the route, the generated dummy packet is treated as if it arrived from source 2 and is transmitted until the destination node. During the waiting period, if another packet arrived from source 1, then a parallel Δ^*-second waiting period for that packet is started instantaneously. This ensures that there is no queuing delay and the maximum delay incurred by any packet is bounded by Δ^* seconds. The strategy is similar if a packet from source 2 arrived to an empty queue. It is easy to see that every transmission by the relay is a batch of two packets, one for each destination. Therefore, the schedules of both outgoing streams from the relay are identical, and it is impossible for an eavesdropper to identify the input-output pair even for long streams of packets.

The rate of dummy transmissions cannot exceed λ, and the delay Δ^* is chosen so that the constraint is satisfied. Although the maximum waiting period for any packet is Δ^* seconds, the average overhead in latency would be strictly less than Δ^*. When the input processes are Poisson distributed, the following theorem characterizes the value of Δ^* and the average latency overhead, given the rate of dummy transmissions λ.

Theorem 1 *For a 2×1 relay B, if sources transmit packets according to unit-rate Poisson processes, then for a given rate of dummy transmissions λ, the average delay $\delta(B, \lambda)$ incurred due to batching by relay B is given by:*

$$\delta(B, \lambda) \quad = \quad \frac{1 - \lambda^2}{8\lambda}$$

and the maximum delay incurred by any packet at B is:

$$\Delta^* = \frac{1-\lambda}{2\lambda}$$

Proof: Refer to Appendix

As is evident from the Theorem, as $\lambda \to \infty$, the average overhead in latency becomes negligible. In other words, a high rate of dummy transmissions incurs no cost in delay. The strategy can be generalized to more than 2 sources as follows.

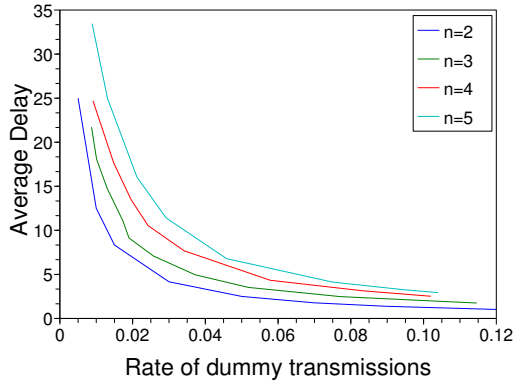
$m \times 1$ Relay: When a packet from one source arrives to the relay node, it waits (for a maximum of Δ^* seconds) until at least one packet arrives from each of the other sources, otherwise dummy packets are generated in place of packets from the sources that did not arrive within Δ^* seconds. Every transmission by the relay contains a batch of m packets, some of which would be dummy packets. Therefore, all outgoing streams from the relay will have identical transmission schedules thereby making all input-output pairs at the relay equally likely. The exact characterization for the average delay as a function of λ becomes exceedingly cumbersome for more than 2 sources. However, the following theorem provides an upper bound on the average delay incurred for a given λ.

Theorem 2 *When M sources transmit packets according to Poisson processes of unit rate, the average delay δ incurred is upper bounded as:*

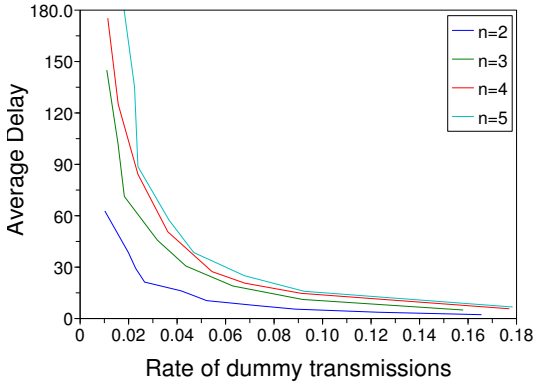
$$\delta(B,\lambda) \leq \frac{(M-1)-\lambda^2}{8\lambda}$$

Proof: Refer to Appendix

Note that, as the number of sources increases, the scheduling strategy provides higher anonymity at the cost of higher latency. Although the theorems only consider Poisson distributed schedules, the strategy can be applied to any general distribution of schedules. For Poisson and Pareto distributed schedules, Figure 5 plots the relationship between average delay and the dummy transmission rate for a relay node, when $m = 2, 3, 4, 5$. For $m > 2$, the figure plots the simulated average delay, which is strictly less than the bound in Theorem 2. As is evident from Figure 4 and Figure 5, the average delay for Pareto distributed processes is considerably higher than Poisson schedules. This is due to the higher burstiness of arrivals in Pareto distributed schedules. As the mean number of packets that arrive in a burst increases, the fraction of dummy packets required to achieve the same delay also increases.



(a) Poisson source schedules



(b) Pareto source schedules

Figure 5: Average delay versus the fraction of dummy transmissions.

The analytical and numerical results presented thus far demonstrate that by appropriately inserting a limited fraction of dummy transmissions it is possible to reduce average latency significantly. It is important to note that the dummy packets generated need to be relayed until the corresponding destination. If the generated dummy packets were to be dropped at a subsequent relay node, the eavesdropper would be able to correlate this transmission schedule with that of the original source transmission, thereby revealing information about the route.

3.2 Covert Relay Selection

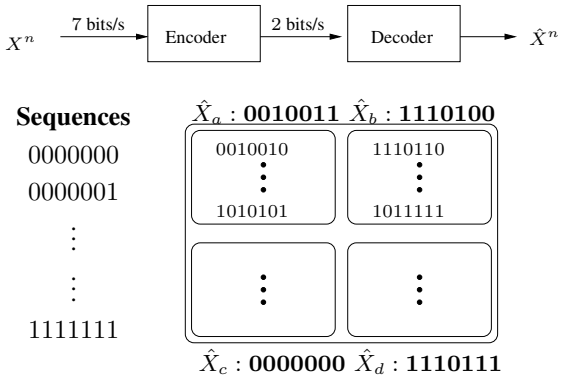
The key idea to minimizing network latency for any given level of anonymity is to select the set of covert relays in each session. To that extent, we consider a random selection strategy, where given the session $\mathbf{S}$ and anonymity level α, we select a subset $\mathbf{B} \subset \mathcal{V}$ of relays to be covert with probability $q_\alpha(\mathbf{B}|\mathbf{S})$ such that

$$\forall \alpha, \mathbf{S} \sum_{\mathbf{B} \in 2^{\mathcal{V}}} q_\alpha(\mathbf{B}|\mathbf{S}) = 1.$$

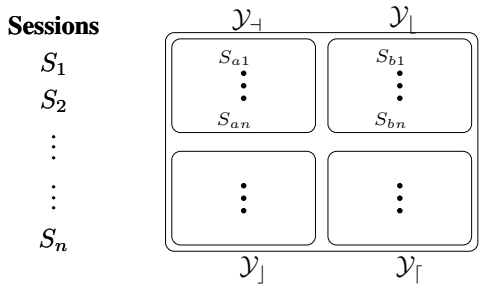
The probability mass function $q_\alpha(\mathbf{B}|\mathbf{S})$ is chosen such that an anonymity of α is guaranteed with minimum overall network latency. The problem of designing the optimal distribution has two fundamentally contrasting paradigms. On the one hand, increasing the number of covert relays would result in a higher level of anonymity. On the other hand, reducing the number of covert relays is beneficial in terms of latency. To obtain the optimal distribution under these constraints, we draw a connection to a well known distortion rate optimization in information theory.

The connection to the distortion-rate optimization can be explained using the following intuition. The objective of the rate-distortion problem is to compress a set of source sequences into a smaller set of codewords, such that the average distortion between each sequence and the corresponding codeword is minimized. The idea is to divide the set of source sequences into bins (Figure 6.a)) such that for each bin, one codeword is generated and all sequences in the bin are mapped to it. The total number of bins (or codewords) is determined by the required level of compression (compression rate). The binning strategy and codewords are designed such that average distortion between sequences and codewords are minimized. A classical result in information theory characterizes the optimal tradeoff between the rate of data compression and the minimum achievable distortion achievable [22].

In our problem setup, the goal is to choose covert relays in each session such that the eavesdropper obtains minimum information about the session. The key idea is to divide the set of all possible network sessions into bins such that, for each bin, there exists one set of covert relays that would make the sessions within that bin indistinguishable to an eavesdropper (Figure 6.b)). Here, the level of anonymity required determines the total number of bins. The binning strategy and the distribution of covert relays are designed to minimize the average latency across sessions within each bin. In the subsequent analysis, we utilize this intuition in optimizing the distribution of covert relays.



(a) Rate-Distortion: Any sequence in bin a is mapped to codeword $\hat{X}_a$. Codewords $\hat{X}$ are chosen to minimize distortion within corresponding bins



(b) Anonymous Networking: For any network session in bin a, eavesdropper observes τ_a, and cannot distinguish $S_{a1} \cdots S_{an}$. τs is designed to minimize latency within corresponding bin.

Figure 6: Connection between rate distortion and anonymous networking.

Note that the strategy as specified by the distribution $q_\alpha(\mathbf{B}|\mathbf{S})$ suggests a centralized implementation, since the decision to choose covert relays requires complete knowledge of routes in the session. An alternative to a centralized implementation is to facilitate exchange of information between nodes at the start of any session so that a consensus is reached on the choice of covert relays. Since the number of sessions and nodes are finite, the convergence of decisions is guaranteed in finite time. For applications where message exchange is not possible, we describe a decentralized approach to select covert relays in Section 5 which achieves the same anonymity at the cost of higher latency.

3.3 Eavesdropper Observation

Given the set of covert relays $\mathbf{B}$, the eavesdropper can detect portions of the routes in the session. Specifically, if a relay is not covert, the eavesdropper can perfectly match the incoming and outgoing schedules at the relay thereby revealing a two-hop connection between pairs of nodes. By analyzing the correlation of schedules across multiple relays in the network, the eavesdropper obtains a partial observation of the session. The remaining portion is completely indiscernible since covert relays perfectly anonymize their schedules.

The eavesdropper's observation is denoted by a set of paths $\hat{\mathbf{S}} \in (\mathcal{V}^*)^*$. $\hat{\mathbf{S}}$ can be expressed as a deterministic function of the actual session $\mathbf{S}$ and the set of covert relays $\mathbf{B}$. Let $t : (\mathcal{V}^*)^* \times \mathcal{V} \rightarrow (\mathcal{V}^*)^*$ represent this function when exactly one relay is covert. More generally, for any set of paths $\mathbf{P} \in (\mathcal{V}^*)^*$, $t(\mathbf{P}, B)$ is the set of observed paths by the eavesdropper when only node B is covert.
$t(\mathbf{P}, B)$ is defined as the set:

$$\{P \in \mathcal{V}^* : P \text{ satisfies one of the following conditions:}$$
$$1.\ \exists P' = (A_1, \cdots, A_k, B, A_{k+1}, \cdots, A_n) \in \mathbf{P},\ \text{s.t.}$$
$$P = (A_1, \cdots, A_k, B) \text{ or } P = (B, A_{k+1}, \cdots, A_n)$$
$$2.\ P \in \mathbf{P} \text{ and } B \notin P.\}$$

Condition 1 states that, when a path in $\mathbf{P}$ contains a covert relay, the eavesdropper would observe two independent paths, one terminating at B and the other originating from B. Condition 2 states that a path that does not contain a covert relay is fully observed.

When a subset $\mathbf{B} = (B_1, \cdots, B_m)$ of relays are covert, then $\hat{\mathbf{S}}$ is obtained by repeated application of $t()$:

$$\hat{\mathbf{S}} = t(\cdots(t(t(\mathbf{S}, B_1), B_2)\cdots), B_m) \stackrel{\triangle}{=} \mathbf{T}(\mathbf{S}, \mathbf{B}). \quad (5)$$

The eavesdropper observes $\hat{\mathbf{S}}$ and tries to estimate the actual session $\mathbf{S}$. Since each covert relay perfectly anonymizes the incoming traffic, and $\hat{\mathbf{S}}$ is derived assuming all non-covert multihop connections are perfectly decoded, the actual transmission times τ do not reveal any additional information about $\mathbf{S}$ than $\hat{\mathbf{S}}$. A formal proof of this statement is provided in the appendix as part of the proof of Theorem 3.

3.4 Network Latency Function

When anonymity $\alpha = 0$, the minimum average delay in a session $\mathbf{S}$ is incurred when none of the relays are covert. This minimum delay for $\mathbf{S}$ is the average transmission delay on the routes of the session, since all nodes merely forward packets immediately upon arrival. For a given session $\mathbf{S}$, we denote this quantity by $\mathbf{\Delta}_t(\mathbf{S})$. According to definition 2 the overall network latency when anonymity $\alpha = 0$ is given by the expected delay over sessions:

$$\Delta(\alpha = 0) = \mathbb{E}(\mathbf{\Delta}_t(\mathbf{S})) = \mathbb{E}(\sum_i \Delta_t(P(i))).$$

When the relays in a subset $\mathbf{B}$ are covert, the increase in latency depends on the delay incurred at each covert relay in $\mathbf{B}$ due to the scheduling strategy, which in turn, depends on the number of paths that contain the relay. Let $\mathbf{\Delta}^c(\mathbf{S}, \mathbf{B}) = (\Delta_1^c(\mathbf{S}, \mathbf{B}), \cdots, \Delta_{|\mathbf{S}|}^c(\mathbf{S}, \mathbf{B}))$ represent the increase in average delays from sources to destinations for the paths in session $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$, when nodes in $\mathbf{B}$ are covert. Therefore

$$\Delta(\mathbf{S}, \mathbf{B}) \stackrel{\triangle}{=} \mathbf{\Delta}_t(\mathbf{S}) + \sum_{i=1}^{|\mathbf{S}|} \Delta_i^c(\mathbf{S}, \mathbf{B})$$

is the total latency in the session. From (2), we know that

$$\Delta_i^c(\mathbf{B}) = \sum_{B \in \mathbf{B} \bigcap P(i)} \delta(B, \lambda), \quad (6)$$

where $\delta(B, \lambda)$ is the average packet delay at covert relay B which is obtained using the strategies in Section 3.1.

4 Latency Anonymity Tradeoff

Given the values of α and λ, using the characterization of network latency and eavesdropper inference, the distribution $q_\alpha(\mathbf{B}|\mathbf{S})$ can be optimized using a brute force search over the (finely discretized) probability simplex. However, this procedure is computationally intensive, and impractical to perform for large networks. The following result characterizes the optimizing distribution and minimum network latency as a function of α, using a well known distortion-rate optimization in information theory.

Theorem 3 *Let* $d : 2^{\mathcal{P}} \times 2^{\mathcal{P}} \rightarrow \mathcal{R}$ *s.t*

$$d_\lambda(\mathbf{S}, \hat{\mathbf{S}}) = \begin{cases} \Delta^c(\mathbf{S}, \mathbf{B}) - \mathbf{\Delta}_t(\mathbf{S}) & \exists \mathbf{B} \text{ s.t. } \hat{\mathbf{S}} = T(\mathbf{S}, \mathbf{B}) \\ \infty & \text{o.w.} \end{cases}. \quad (7)$$

Then, a network latency $\Delta(\alpha, \lambda)$ is achievable with average anonymity α and dummy transmission rate λ if

$$\Delta(\alpha, \lambda) - \Delta(0) \geq D\left(H(\mathbf{S}|\mathbf{L})(\mathbf{1} - \alpha)\right),$$

where $D(r)$ is the Distortion-Rate function:

$$D(r) = \min_{q(\hat{\mathbf{S}}|\mathbf{S}): I(\mathbf{S};\hat{\mathbf{S}}|\mathbf{L}) \leq \mathbf{r}} \mathbb{E}(d_\lambda(\mathbf{S}, \hat{\mathbf{S}})). \tag{8}$$

Proof: Refer to Appendix.

The distortion-rate function in (8) is used in information theory to provide the minimum average distortion incurred in order to compress a set of source sequences. The theorem demonstrates the mathematical equivalence between the two optimizations described in the intuitive argument earlier. Specifically, the function $d_\lambda(\mathbf{S}, \hat{\mathbf{S}})$ in (7) characterizes the increase in latency in a given session $\mathbf{S}$, when the observed session is $\hat{\mathbf{S}}$. The function $d_\lambda(\mathbf{S}, \hat{\mathbf{S}})$ does not explicitly include the set of covert relays $\mathbf{B}$. However, in the proof of the theorem, we show that given $\hat{\mathbf{S}}$, the set of covert relays $\mathbf{B}$ is unique. As a result, the distribution $q_\alpha(\mathbf{B}|\mathbf{S})$ to chose covert relays is equivalent to the distortion minimizing distribution in (8).

4.1 Discussion

The consequence of the connection to rate-distortion extends beyond the idea of choosing covert relays; rate distortion is a field that has been studied for many decades [22], and the numerous models and techniques developed therein, could serve to design strategies for route anonymity. The form of the rate-distortion problem used in this work is a slight modification of classical rate-distortion, due to the presence of side information $\mathbf{L}$ provided by packet headers. However, $\mathbf{L}$ is a deterministic function of $\mathbf{S}$, and is available to the network designer as well. As a result, the Blahut-Arimoto algorithm [23] used in standard rate-distortion optimization provides an efficient iterative technique to characterize the achievable network latency $\Delta(\alpha, \lambda)$ and obtain the optimal scheduling strategy $q_\alpha(\mathbf{B}|\mathbf{S})$.

Our assumption of an omniscient adversary is very conservative, and typically an eavesdropper would only monitor carefully chosen portions of the network. We believe that our analytical approach can be extended to model such constrained eavesdroppers as well. Specifically, if the eavesdropper monitors a random subset of nodes, then her observation, currently represented using the pair $\hat{\mathbf{S}}, \mathbf{L}$, would correspond to a random function of $\hat{\mathbf{S}}, \mathbf{L}$ depending on the fraction of monitored nodes (fraction here only refers to number of nodes and not the actual set of nodes). A similar approach can be adopted to model active adversaries. If an eavesdropper were to compromise a subset of relays, thereby revealing two-hop information, then the inference thus obtained can be modeled as unknown side information available to the adversary. Analyzing these extended models is however not straightforward, since the set of monitored nodes could be chosen depending on the optimal distribution of covert relays.

Note that our approach of making transmission schedules statistically independent, assumes that the eavesdropper can detect even the slightest of correlation. In general, detecting dependencies across transmission schedules is a hard problem, especially when dumy transmissions are allowed. There has been significant ongoing effort in using information-theoretic methods for this purpose, in the context of detecting covert timing channels [24, 25]. Our approach while conservative, provides an achievable quality-of-service with provable anonymity in a network.

5 Decentralized Approach

In order to achieve the performance of Theorem 3, it is necessary that every relay be aware of the entire session $\mathbf{S}$. Further, since the strategy involves a random selection, all nodes need to share some common randomness. In network applications where centralized control is not possible and message exchanges across nodes are not allowed, the performance stated in Theorem 3 may not be achievable. In such networks each node would have partial information about the session, which contains only the immediate transmitting and receiving nodes. This is similar to Mix networks where layered encryption can ensure that each Mix only has knowledge of the neighbouring nodes in the routes. Therefore, for networks where nodes only have local information and do not share common randomness, we propose the following decentralized approach.

The local information available to a relay node in any session is a set of node pairs that represent the immediate sender-receiver pairs at the relay. We define function $l : \mathcal{V} \times \mathcal{S} \mapsto 2^{\mathcal{V}\times\mathcal{V}}$ where $l(B, \mathbf{S})$ denotes the information available to node B in session $\mathbf{S}$.

If $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$ and $A(i,j)$ represent the i^{th} node of path $P(j)$ in $\mathbf{S}$, then,

$$l(B, \mathbf{S}) = \{(A(i, j-1), A(i, j+1)) : A(i,j) = B\}. \quad (9)$$

In other words, $l(B, \mathbf{S})$ is the set of all node pairs $(A(i, j-1), A(i, j+1))$ such that node B relays packets from $A(i, j-1)$ to $A(i, j+1)$ in session $\mathbf{S}$. It is important to note that $l(B, \mathbf{S})$ is a many-to-one function from the set of sessions, and therefore, multiple sessions that result in identical $l(B, \mathbf{S})$ would be indistinguishable to the relay.

Since there are no message exchanges across nodes with regard to the session information, we require that each node makes a decision to be covert based on the local information function only. In other words, during any session, node B would make a decision to be covert depending only on the value of $l(B, \mathbf{S})$. Further, we do not assume any common randomness available to the nodes, and hence, the decisions of multiple nodes are independent in each session. Accordingly, we define a probability function for each node:

$$q'_\alpha : \mathcal{V} \times 2^{\mathcal{V} \times \mathcal{V}} \mapsto [0, 1],$$

where $q'_\alpha(B, l(B, \mathbf{S}))$ is the probability that node B is covert in session $\mathbf{S}$. Since each node makes an independent decision, the probability that a subset of nodes $\mathbf{B}$ have decided to be covert in session $\mathbf{S}$ is a product of decision probabilities of the individual nodes:

$$q(\mathbf{B}|\mathbf{S}) = \prod_{B \in \mathbf{B}} q'_\alpha(B, l(B, \mathbf{S})) \prod_{B \notin \mathbf{B}} (1 - q'_\alpha(B, l(B, \mathbf{S}))). \quad (10)$$

The above equation specifies a particular product form for the probabilities $q(\mathbf{B}|\mathbf{S})$, and every decentralized strategy would correspond to a set of $\{q(\mathbf{B}|\mathbf{S})\}$ which can be thus expressed. So let Q^* represent the set of all conditional probability mass functions $\{q(\mathbf{B}|\mathbf{S})\}$, such that there exists probabilities $q'_\alpha(\mathbf{B}, \mathbf{S})$ which satisfies (10) for every realization of $(\mathbf{S}, \mathbf{B})$. We know that $(\mathbf{S}, \mathbf{B})$ and $(\mathbf{S}, \hat{\mathbf{S}})$ have a one-one correspondence. Therefore, the set Q^* would have a one-one correspondence to a set Q^{**} of conditional probabilities $q(\hat{\mathbf{S}}|\mathbf{S})$.

Then, using the result of Theorem 3, we can obtain the achievable latency for the decentralized strategy as:

$$\Delta(\alpha, \lambda) - \Delta(0) \geq D'\left(H(\mathbf{S}|\mathbf{L})(\mathbf{1} - \alpha)\right),$$

where

$$D'(r) = \min_{q(\hat{S}|S) \in Q^{**} : I(S;\hat{S}|\mathbf{L}) \leq \mathbf{r}} \mathbb{E}(d_\lambda(S, \hat{S})). \quad (11)$$

Note that the minimization in (11) is over a subspace of the probability space as compared to Theorem 3, and could therefore result in a lower throughput than that of Theorem 3. Even if $l(B, \mathbf{S})$ uniquely identifies the session for all $B, \mathbf{S}$, the throughput may not reach the optimal value of Theorem 1 owing to lack of common randomness (independent decisions). This is illustrated in the following example.

5.1 Example

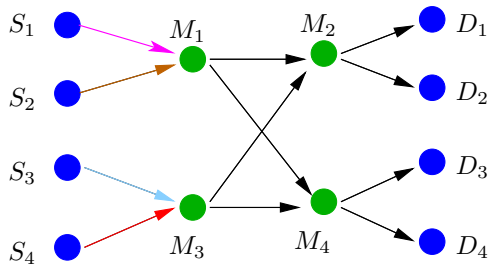
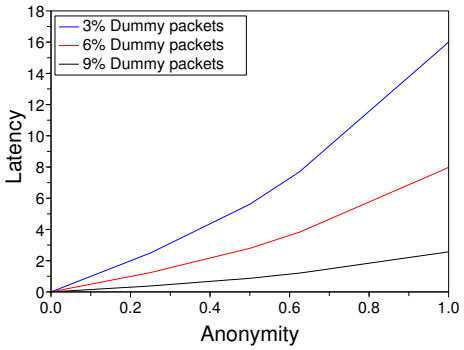


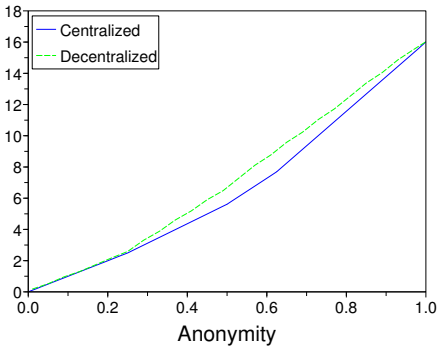
Figure 7: Switching Network: $\{S_i\}$ transmit to $\{D_i\}$ through relays $\{M_i\}$.

Consider the example of a switching network as shown in Figure 7. During any network session, each source S_i picks a distinct destination D_j, and for each pair S_i, D_j there is a fixed path through the intermediate relays. The set of possible sessions, $\mathcal{S}$, contains 24 elements (all possible pairings) which are assumed equiprobable. For this setup, Figure 8 plots the latency-anonymity region under different constraints on the rate of dummy transmissions.

As expected, as the fraction of dummy transmissions increase, the latency incurred due to anonymity reduces. Note that, the latency incurred is a convex U function of the anonymity. The reason for the convexity is that conditional entropy and network latency are average metrics. Therefore, time sharing of strategies (where the strategy for a session is chosen probabilistically from a set of strategies) would also yield a valid strategy. This is, however, not possible with the decentralized strategy as the nodes do not possess any shared randomness. The decentralized approach does not increase the latency significantly, and the performance is identical at the extreme values of α. This is because, when all the relays or none of them are covert, there is no necessity for centralized decisions. Although we have illustrated our ideas and the validity of the theoretical approach using a simple example, the applicability of the strategies extends to general multihop networks.



(a) Centralized Approach: Performance characterization for different fractions of dummy transmissions



(b) Comparison between centralized and decentralized approach.

Figure 8: Trade-Off between Anonymity and Latency: All paths have unit rate of transmission, and the 2×1 relaying strategy is used for the chosen covert relays.

6 Conclusions and Future work

One of our key contributions in this work is the theoretical model for anonymity against traffic analysis. To the best of our knowledge, this is the first analytical metric designed to measure the secrecy of *routes* in an eavesdropped multihop network. Based on the metric, we designed scheduling and relaying strategies to minimize network latency with a guaranteed level of anonymity. Although we consider specific constraints on dummy transmissions and the session models, the ideas of covert relaying and the randomized selection are quite general. An important future direction is to consider eavesdroppers who observe the network for long durations of time. This requires a dynamic session model, where it is important to maintain anonymity of routes under changes in sessions due to nodes ending or starting communications.

References

[1] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the ACM*, vol. 24, pp. 84–88, February 1981.

[2] L. Cottrell, "Mixmaster and Remailer Attacks," tech. rep., 1994.

[3] C. Díaz and A. Serjantov, "Generalizing mixes," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2003)*, Springer-Verlag, LNCS 2760, April 2003.

[4] Y. Zhu, X. Fu, B. Graham, R.Bettati, and W. Zhao, "On flow correlation attacks and countermeasures in mix networks," in *Proceedings of Privacy Enhancing Technologies workshop*, May 26-28 2004.

[5] A. Pfitzmann, B. Pfitzmann, and M. Waidner, "ISDN-MIXes: Untraceable communication with very small bandwidth overhead," in *Proceedings of the GI/ITG Conference: Communication in Distributed Systems, Informatik-Fachberichte*, vol. 267, (Mannheim, Germany), pp. 451–463, February 1991.

[6] O. Berthold, H. Federrath, and S. Kopsell, "Web MIXes: A system for anonymous and unobservable Internet access," in *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability, Lecture Notes in Computer Science*, vol. 2009, (Berkeley, CA), pp. 115–129, July 2000.

[7] B.Radosavljevic and B. Hajek, "Hiding traffic flow in communication networks," in *Military Communications Conference*, 1992.

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

[9] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)* (R. Dingledine and P. Syverson, eds.), Springer-Verlag, LNCS 2482, April 2002.

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

[11] A. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.

[12] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. on Information Theory*, vol. 24, pp. 339–348, May 1978.

[13] P. Venkitasubramaniam, T. He, and L. Tong, "Anonymous Networking amidst Eavesdroppers," *to appear IEEE Transactions on Information Theory: Special Issue on Information-Thoeretic Security*, 2008.

[14] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: design of a type iii anonymous remailer protocol," in *Proceedings of 2003 Symposium on Security and Privacy*, pp. 2–15, May 2003.

[15] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-generation Onion Router," in *Proc. USENIX Security Symposium.*, (San Diego, CA), 2004.

[16] R. Dingledine and P. Syverson, "Reliable MIX Cascade Networks through Reputation," *Financial Cryptography. Springer-Verlag, LNCS*, vol. 2357, 2002.

[17] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system,," in *Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability (2001)*, pp. 46–66. http://freenet.sourceforge.net.

[18] M. J. Freedman and R. Morris, "Tarzan: A peer-topeer anonymizing network layer," in *Proceedings of 9th ACM Conference on Computer and Communications Security*, (Washington, DC), Nov. 2002.

[19] R. Bohme, G. Danezis, C. Diaz, S. Kopsell, and A. Pfitzmann, "Mix cascades vs. peer-to-peer: Is one concept superior?," in *Privacy Enhancing Technologies (PET 2004)*, (Toronto, Canada), May 2004.

[20] D. Chaum, "The dining cryptographers problem : Unconditional sender and recipient untraceability," *Journal of Cryptography*, vol. 1, no. 1, pp. 65–75, 1988.

[21] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for Web transactions," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66–92, 1998.

[22] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[23] R. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," *IEEE Trans. Infor. Theory*, vol. IT-18, July 1972.

[24] S. Gianvecchio and H. Wang, "Detecting covert timing channels: an entropy-based approach," in *CCS '07: Proceedings of the 14th ACM conference on Computer and communications security*, pp. 307–316, 2007.

[25] T. He and L. Tong, "Detecting Information Flows: Fundamental Limits and Optimal Algorithms." submitted to IEEE Trans. on Information Theory, 2007.

[26] D. Cox and H. Miller, *The Theory of Stochastic Processes*. New York: John Wiley & Sons Inc., 1965.

Appendix

Proof of Theorem 1

To prove the theorem, we adopt the technique used in [25]. Consider the two point processes $\tau_{S_1,B}, \tau_{S_2,B}$. For a mix that uses the simple batching strategy, the delay of the j^{th} packet from S_1 would be $(T_{S_1,B}(j) - T_{S_2,B})^+$. Let $X_j = T_{S_1,B}(j) - T_{S_2,B}(j)$. Define

$$\begin{aligned} T_j^R &\triangleq X_j - X_{j-1} \\ &= (T_{S_1,B}(j) - T_{S_2,B}(j-1)) - (T_{S_1,B}(j) - T_{S_2,B}(j-1)). \end{aligned}$$

We see that T_j^R's are i.i.d. random variables; each T_j^R is the difference between two independent exponential random variables with mean 1. The process $\{X_j\}_{j=1}^{\infty}$ is a general random walk with step T_j^R [26].

Running the mixing algorithm is going to result in dummy transmissions at different points of time depending on which packet is waiting in the queue. For every dummy packet transmitted at t in place of a packet from source 2, we insert a virtual packet at $t + \Delta^*$ in $\tau_{S_1,B}$; for every dummy packet inserted at time s in place of a packet from source S_1, we insert a virtual packet at $s + \Delta^*$ in $\tau_{S_2,B}$.

Let the new packet delays $(T_{S_1,B}(j) - T_{S_2,B}(j))$ after the insertion of virtual packets be $\{X'_j\}_{j=0}^{\infty}$. It can be shown that $\{X'_j\}_{j=0}^{\infty}$ is also a random walk with step T_j^R, but it has two absorbing barriers at $-T$ and T, $i.e.$

$$X'_j = \min(\max(X'_{j-1} + T_j^R,\ -\Delta^*),\ \Delta^*).$$

Since it is almost surely impossible for $X'_{j-1} + T_j^R$ to be exactly equal to $-\Delta^*$ or Δ^*, each time $X'_j = -\Delta^*$ corresponds to a dummy transmission in place of source 2, and $X'_j = \Delta^*$ corresponds to a dummy transmission in place of source 1. From example 2.16 in [26], we know that the probability of $X'_j = \Delta^*$ is given by

$$\Pr\{X'_j = \Delta^*\} = \frac{1}{2(1+\Delta^*)} = \Pr\{X'_j = -\Delta^*\}.$$

Therefore, the fraction of dummy transmissions corresponding to source 1 alone is

$$\lambda = \frac{\Pr\{X'_i = \Delta\}}{(1 - \Pr\{X'_i = 0\})} = \frac{1}{1 + 2\Delta^*}. \tag{12}$$

To characterize the average delay, we derive the stationary distribution of the random walk between the barriers. Every $0 \leq X'_j \leq \Delta^*$ contributes to the delay of a source 2 packet, and every $-\Delta^* \leq X'_j \leq 0$ corresponds to the delay of a source 1 packet. The average delay incurred would be the mean of the stationary distribution within those respective limits (due to symmetricity it would be the same for source 1 or 2). Again, following the exposition in example 2.16 in ([26], Page 67), the cumulative distribution of the delay in the interval $(-\Delta^*, \Delta^*)$ is given by

$$\Pr(X'_i \leq x) = \frac{\Delta^* + 1 + x}{2(1+\Delta^*)}. \tag{13}$$

Using (12) and (13), $\delta(B, \lambda)$ can be evaluated as:

$$\begin{aligned} \delta(B,\lambda) &= \mathbb{E}\{X'_i | X'_i \in (-\Delta^*, \Delta^*)\} \\ &= \frac{(\Delta^*)^2}{2(1+2\Delta^*)}. \\ &= \frac{1-\lambda^2}{8\lambda}. \end{aligned}$$

□

Proof of Theorem 2

When more than 2 sources are being relayed, it is exceedingly cumbersome to characterize the exact delay. We provide an upper bound on the delay instead. We approximate the $m \times 1$ relaying strategy as follows. When a packet from source 1 arrives at an empty queue, it waits for a packet from source 2 for a time $\Delta^*/(M-1)$ seconds, if no packet arrives immediately a dummy packet is generated for source 2. This process is then repeated for the remaining $M-2$ sources. In this manner, every packet from source 1 has one distinct packet (data or dummy) from each of the sources within T seconds. The fraction of dummy packets would now correspond to a delay of $\Delta^*/(M-1)$ instead of T. Further the average delay would $M-1$ times the average delay incurred by each random walk. Plugging these factors into (12) and (13), we obtain the result in Theorem 2. □

Proof of Theorem 3

From (6), we know that $\mathbf{Delta}^c(\mathbf{S}, \mathbf{B})$ is an achievable latency vector when nodes in $\mathbf{B}$ are covert. It remains to be seen that the condition $H(\mathbf{S}|\hat{\mathbf{S}}, \mathbf{L}) \geq \alpha$ guarantees an anonymity α. For this purpose, it is sufficient to show that

$$H(\mathbf{S}|\tau, \mathbf{L}) \geq \mathbf{H}(\mathbf{S}|\hat{\mathbf{S}}, \mathbf{L}).$$

Let $\hat{\tau}$ be the schedules generated assuming $\hat{\mathbf{S}}$ was a session and none of the nodes were covert. The transmission rates of nodes in $\hat{\tau}$ are assumed identical to τ. For the nodes that are the sources in $\mathbf{S}$, the schedules are independent in τ and $\hat{\tau}$. Session $\hat{\mathbf{S}}$ has additional sources due to the broken paths, which also generate independent transmission schedules. The set of these additional sources is identical to the set of covert relays in $\mathbf{S}$. Therefore, the schedules are independent in τ as well. Since the remaining nodes relay all received packets within negligible processing delay, $p(\tau|\mathbf{S}) = p(\hat{\tau}|\mathbf{S})$. Then, using the data processing inequality $(\mathbf{S} - (\hat{\mathbf{S}}, \mathbf{L}) - (\hat{\tau}, \mathbf{L}))$

$$H(\mathbf{S}|\tau, \mathbf{L}) = \mathbf{H}(\mathbf{S}|\hat{\tau}, \mathbf{L}) \geq \mathbf{H}(\mathbf{S}|\hat{\mathbf{S}}, \mathbf{L}).$$

Consider the optimal solution $q^*(\hat{\mathbf{S}}|\mathbf{S})$ of the distortion rate problem,

$$D = \min_{q(\hat{\mathbf{S}}|\mathbf{S}): I(\mathbf{S};\hat{\mathbf{S}}|\mathbf{L}) \leq (\mathbf{1}-\alpha)\mathbf{H}(\mathbf{S})} \mathbb{E}(d(\mathbf{S}, \hat{\mathbf{S}})).$$

From the definition of $d(\mathbf{S}, \hat{\mathbf{S}})$, it is easy to see that if $\nexists \mathbf{B}\ s.t.\ \hat{\mathbf{S}} = \mathbf{T}(\mathbf{S}, \mathbf{B})$, then $q^*(\hat{\mathbf{S}}|\mathbf{S}) = 0$. Given $\mathbf{S}, \hat{\mathbf{S}}$, we can show that the set of covert relays $\mathbf{B}$ are uniquely determined, using the following argument:

Suppose $\exists \mathbf{B}_1 \neq \mathbf{B}_2$ such that $\mathbf{T}(\mathbf{S}, \mathbf{B}_1) = \mathbf{T}(\mathbf{S}, \mathbf{B}_2)$. Then, we can write $\mathbf{B}_1 = (\mathbf{B}, \mathbf{B}'_1), \mathbf{B}_2 = (\mathbf{B}, \mathbf{B}'_2)$ where $\mathbf{B}'_1 = (B_{11}, \cdots, B_{1m})$, $\mathbf{B}'_2 = (B_{21}, \cdots, B_{2n})$ and $\mathbf{B}'_1 \bigcap \mathbf{B}'_2 = \phi$. We know that

$$\begin{aligned} \hat{\mathbf{S}}(\mathbf{S}, \mathbf{B}_1) &= t(\cdots t(\mathbf{T}(\mathbf{S}, \mathbf{B}), B_{11}), \cdots), B_{1m}) \\ &= t(\cdots t(\mathbf{T}(\mathbf{S}, \mathbf{B}), B_{21}), \cdots), B_{2n}) = \hat{\mathbf{S}}(\mathbf{S}, \mathbf{B}_2). \end{aligned}$$

Suppose none of the paths in $\mathbf{T}(\mathbf{S}, \mathbf{B})$ contain $\mathbf{B}'_1 \bigcup \mathbf{B}'_2$, then it does not matter if those relays are covert or not, in which case the subset of covert relays would be $\mathbf{B}$.

If $\exists P \in \mathbf{T}(\mathbf{S}, \mathbf{B})$ that contains B_{11}, then $\mathbf{T}(\mathbf{S}, \mathbf{B}_1)$ would contain a path that ends in B_{11}, whereas $\mathbf{T}(\mathbf{S}, \mathbf{B}_2)$ cannot contain such a path, which is a contradiction.

The above argument shows that we can equivalently write $q^*(\hat{\mathbf{S}}|\mathbf{S}) = q^*(\mathbf{B}|\mathbf{S})$. Therefore, q^* specifies a valid selection strategy. Since $H(\mathbf{S})$ is fixed apriori, $I(\mathbf{S}; \hat{\mathbf{S}}|\mathbf{L}) \leq (\mathbf{1}-\alpha)\mathbf{H}(\mathbf{S}|\mathbf{L})$ ensures that an anonymity α is guaranteed. Further, for every $\mathbf{B}$, the function d evaluates the difference in achievable rate vectors $\bar{\mathbf{\Delta}}^0(\mathbf{S})$ and $\bar{\mathbf{\Delta}}^c(\mathbf{S}, B)$. Taking expectation over $q^*(\mathbf{B}|\mathbf{S})$, it is easy to see that the distortion D is achievable with $\alpha-$anonymity.

□