This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2008 proceedings.

Throughput Anonymity Trade-off in Wireless Networks under Latency Constraints

Parvathinathan Venkitasubramaniam and Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca,
NY, 14850
Email: {pv45,lt35}@cornell.edu

***Abstract*—Providing anonymity to routes in a wireless ad hoc network from passive eavesdroppers is considered. Using Shannon's equivocation as an information theoretic measure of anonymity, scheduling strategies are designed for wireless nodes using receiver directed signaling. The achievable rate region for multiaccess relays are characterized under constraints on average packet latency. The relationship between overall network throughput and the route anonymity is obtained by drawing a connection to the rate-distortion tradeoff in information theory. A decentralized implementation of the relaying strategy is proposed, and the corresponding performance analyzed.**

I. Introduction

A. Motivation

Eavesdropping of transmissions in a network can reveal vital information about the network operation. The transmission times of nodes alone can be used to determine source destination pairs and the routes of traffic flow. Such unauthorized information retrieval, known as a *traffic analysis attack*, compromises user privacy and also makes it possible to launch powerful attacks such as jamming and denial of service. While cryptography can be used to obfuscate the contents of communication, hiding the *act of communication* requires a fundamental redesign of networking protocols.

The challenge in the design of anonymous protocols is to hide the routing information from eavesdroppers without violating constraints imposed by the network. In this regard, the wireless medium presents its own advantages and disadvantages. On the one hand, it is difficult for eavesdroppers to ascertain the transmitting or receiving nodes of an encrypted wireless transmission, especially when different traffic streams are multiplexed at a single relay. On the other hand, the shared medium is band limited and susceptible to fading and interference, thereby constraining the network designer.

In this work, we are interested in designing *anonymous* transmission and relaying protocols in wireless networks to prevent the timing based inference of routes. We consider traffic flows where the *average per packet delay* is bounded. It is evident that modifying transmission schedules would result in loss of network performance. We are interested in the tradeoff between network performance, measured by throughput, and the level of anonymity that can be provided. Delay limitations on traffic are necessary in time sensitive applications such as media transmission, and in sensor networks, where node duty cycles are too sparse to store packets for long periods. In general, a bounded packet delay ensures stability and prevents congestion at any node in the network.

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.

B. Related Work

The idea of hiding routing information from eavesdroppers is classical, although with a few exceptions [1], [2], it has primarily been applied to Internet traffic over a wired network. Most Internet applications provide anonymity using a concept known as Mixing, pioneered by Chaum [3]. A Mix is a special node or server that collects packets from multiple users and transmits them after modifying the contents and random delaying such that, it is impossible to match an incoming and outgoing packet at a Mix. Since a single Mix stands a chance of being compromised, a (possibly random) sequence of Mixes are interposed between sources and destinations to protect against active means of gaining inference.

Subsequent to Chaum's contribution, many improved batching strategies [4], [5] have been designed to handle different types of traffic analysis attacks [6]. While the Mix based approach is useful for Internet applications such as anonymous remailers and web browsing [7], a study of flow correlation attacks [8] showed that when long streams of packets with latency constraints are forwarded through Mixes, it is possible to correlate incoming and outgoing streams almost perfectly.

In wireless networks, an alternative solution to Mixing is the idea of cover traffic [9], [10], where, irrespective of the active routes, the transmission schedules of all nodes are fixed apriori. If a node does not have any data packets, the transmission schedule is maintained by transmitting dummy packets. The fixed scheduling strategy, analyzed in [9] provides complete anonymity to the routes at all times. Constraints on traffic latency have however, not been considered. Furthermore, a fixed scheduling strategy requires synchronization across all nodes and assumes a constant network topology, which is not practical in ad hoc wireless networks.

978-1-4244-2026-1/08/$25.00 © 2008 IEEE

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

C. Main Contributions

In this work, we propose an information theoretic approach towards providing anonymity to traffic flows in a multihop wireless network. In particular, we quantify the route anonymity using Shannon's equivocation [11], and design network protocols that are adaptable to any desired level of anonymity. Equivocation measures the uncertainty of hidden information with respect to the eavesdropper's observation. It has primarily been used to quantify the secrecy of messages transmitted over channels such as wiretapped [12] and broadcast channels [13]; the goal was to characterize the optimal tradeoff between information rate and secrecy. We use equivocation to quantify the anonymity of network routes, and characterize the tradeoff between network throughput and anonymity. Previously, in [14], we considered a transmitter directed signaling network with strict delay constraints on the traffic, and derived the tradeoff between throughput and anonymity. The achievability of the throughput in [14] required centralized knowledge of the network routes.

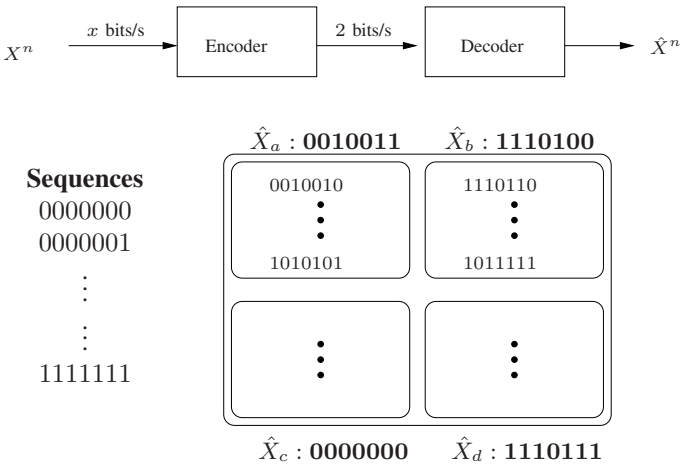
In this work, we consider a receiver directed physical layer model with average delay constraints, and propose a decentralized scheduling strategy to provide anonymity. We propose transmission and relaying strategies to hide the relaying operation of individual wireless nodes. These strategies, due to the latency constraints, result in a reduction in achievable relay rates at the nodes. Therefore, we selectively reveal portions of the network so that network throughput is maximized for the desired level of anonymity. A key intuition for this maximization comes from the rate-distortion tradeoff in information theory, which is explained as follows.

The objective of a rate-distortion optimization is to map a set of source sequences to a smaller set of reconstruction sequences such that the average distortion between the source and reconstructed sequences is minimized. The idea is to divide the set of source sequences into bins (Figure 1.a)), and generate a reconstruction sequence for each bin. The compression rate determines the total number of allowed bins, and the binning and reconstruction are performed such that overall distortion is minimized. A classical result in information theory characterizes the optimal distortion-rate trade-off as:

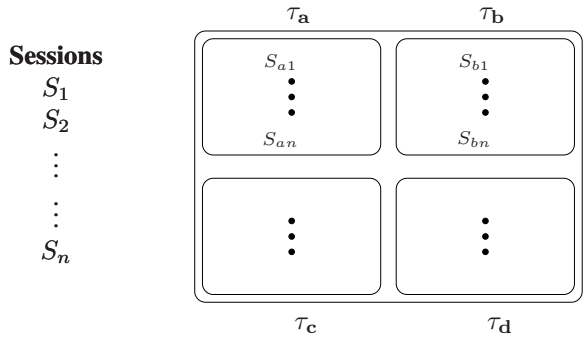
$$D(r) = \min_{q(\hat{S}|S):I(S,\hat{S})\leq r)} d(S,\hat{S}), \qquad (1)$$

where S is the source alphabet, $\hat{S}$ is the reconstruction alphabet and $d(S,\hat{S})$ is the distortion measure.

In the anonymous networking setup, let the set of active routes at any given time be referred to as a network session. The key idea is to divide the set of all possible network sessions into bins (Figure 1.b)) such that, for each bin, there exists a scheduling strategy that would make the sessions within that bin indistinguishable to an eavesdropper. The level of anonymity required determines the number of bins, and the optimal scheduling strategy plays the role of the reconstruction sequence by minimizing the performance loss across sessions within a bin. In this work, we show that the throughput-anonymity relation can be equated to a rate-distortion function.



(a) Rate-Distortion: Any sequence in bin a corresponds to reconstruction sequence $\hat{X}_a$. Reconstruction sequences $\hat{X}$ are chosen to minimize distortion within corresponding bins



(b) Anonymous Networking: For any network session in bin a, eavesdropper observes τ_a, and cannot decide from $S_{a1}\cdots S_{an}$. τs are designed to minimize performance loss within corresponding bins.

Fig. 1. Connection between rate distortion and anonymous networking.

The remainder of this paper is organized as follows. In Section II, the anonymity model and the formal problem setup are described. In Section III, the scheduling algorithms to hide the operation of individual relays are presented along with the characterization of achievable relay rates for Poisson distributed schedules. In Section IV, the extension of the strategies to multihop routes in a network are described. The characterization of throughput-anonymity tradeoff and a decentralized implementation are presented in Section V.

II. Problem Setup

Let the network be represented by a directed graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subset \mathcal{V}\times\mathcal{V}$ is the set of directed links. (A,B) is an element of $\mathcal{E}$ iff 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 $(V_i,V_{i+1}) \in \mathcal{E},\ \forall i < n$. The set of all possible paths is given by $\mathcal{P}(\mathcal{G})$.

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 $\mathbf{S} \in 2^{\mathcal{P}(\mathcal{G})}$ a network *session*.

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

The set of all possible sessions $\mathcal{S}$ is typically a strict subset of $2^{\mathcal{P}(\mathcal{G})}$. We model $\mathbf{S}$ as an i.i.d. random variable $\mathbf{S} \sim p(\mathbf{S})$. The information that we wish to hide from the eavesdropper is the network session $\mathbf{S}$. We assume that the prior $p(\mathbf{S})$ on sessions is available to the eavesdropper.

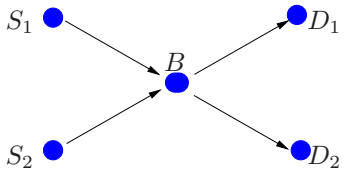


Fig. 2. Two Node Switching Network: $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E})$, $\mathcal{V} = \{S_1, S_2, B, D_1, D_2\}$, $\mathcal{E} = \{(S_1, B), (S_2, B), (B, D_1), (B, D_2)\}$.

For example, consider the simple network $\mathcal{G}_1$ as shown in Figure 2. Let S_1, S_2 be the source nodes and D_1, D_2 the destinations. Let S_1, S_2 always communicate with distinct destinations. For this network,

$$\mathcal{P}(\mathcal{G}_1) = \{ \quad (S_1, B), (S_1, B, D_1), (S_1, B, D_2), (S_2, B), \\ (S_2, B, D_1), (S_2, B, D_2), (B, D_1), (B, D_2) \ \}.$$

However, since destinations are always distinct,

$$\mathcal{S} = \Big\{ \quad \{(S_1, B, D_1), (S_2, B, D_2)\} \\ \{(S_1, B, D_2), (S_2, B, D_1)\} \Big\}.$$

For the purposes of obtaining an analytical characterization of the throughput-anonymity tradeoff, we have considered an abstraction that deviates from the real network operation. It is our hope that the insights obtained in this setting will provide design guidelines for real applications.

Transmission Schedules The eavesdroppers' observation in a session comprises of the packet transmission times of all the nodes. We assume that packet headers are encrypted, and hence, undecodable by the eavesdropper. Therefore, merely detecting a transmission on the wireless medium cannot provide the eavesdropper information about the transmitter or receiver. However, we consider a receiver directed physical layer model, where the eavesdropper can use knowledge of spreading sequences to determine the receiving node of every transmitted packet.

Receiver Directed Signaling: All packets received by a particular node are required to be modulated using the same spreading sequence, and each *receiving node* is associated with a unique orthogonal spreading sequence. Under this scheme, an eavesdropper would be able to "tune" his detector to a spreading sequence and detect transmission times of packets sent to the corresponding node. Note that since headers are not available, the identity of the transmitting node is hidden.

Observable Scheduling Let τ_A represent the transmission times of packets received by node A. The schedule τ_A is given by a point process,

$$\tau_A = \{T_A(1), T_A(2), \cdots\},$$

where $T_A(i)$ represents the transmission time of the i^{th} packet sent to node A. Since we cannot determine which nodes are being monitored, the eavesdroppers' complete observation is assumed to be $\tau = \{\tau_A : A \in \mathcal{V}\}$.

We model τ as a sequence of random variables with conditional distribution $q(\tau|\mathbf{S})$. The idea is to design $q(\tau|\mathbf{S})$ such that eavesdroppers obtain minimum information about the session $\mathbf{S}$ by observing τ. Note that τ only contains transmission times of packets received by each node, and does not indicate the routes in the network. The eavesdropper would therefore need to correlate transmission schedules across nodes to determine the actual flow of traffic.

A. Anonymity Measure

In Mix-net analyses, anonymity has been defined [4], [5] using the *anonymity set* (set of possible source-destination pairs) of an observed packet. Although anonymity sets provide useful metrics for Internet applications, in wireless ad hoc or sensor networks, it is equally important to hide the routes of data flow as it is to hide source destination information. Furthermore, to measure the overall anonymity in a network it is imperative to consider entire streams of packets, as significant information is inferable from the inter-packet timings [8].

We define anonymity using the notion of equivocation [11], which measures the uncertainty of the information we wish to hide ($\mathbf{S}$) given the complete observation of the eavesdropper (τ).

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

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

When $\alpha = 1$, $H(\mathbf{S}|\tau) = H(\mathbf{S})$, and the distribution $q(\tau|\mathbf{S})$ is defined to have *perfect anonymity*. In other words, the schedule τ does not provide any additional information about the routes than the available prior $p(\mathbf{S})$. For a general α, the physical interpretation comes from Fano's Inequality [15]: Let the error probability of the eavesdropper in decoding the session $\mathbf{S}$ be P_e. Then,

$$P_e \geq \frac{H(\mathbf{S}|\tau) - 1}{\log|\mathcal{S}|} \geq \frac{\alpha H(\mathbf{S}) - 1}{\log|\mathcal{S}|} \triangleq f(\alpha).$$

Furthermore, if $\mathcal{S}$ is a large set with uniform prior $\{p(\mathbf{s}) = \frac{1}{|\mathcal{S}|}, \forall \mathbf{s} \in \mathcal{S}\}$, then $f(\alpha) \approx \alpha$, which implies that the probability of error is lower bounded by the anonymity.

B. Network Constraints and Throughput

The design of schedule distribution $q(\tau|\mathbf{S})$ is subject to network constraints on medium access and latency. Our goal is to design $q(\tau|\mathbf{S})$ for any desired level of anonymity α such that network performance is maximized under the given constraints. In this work, network performance is measured using throughput and the medium access and delay constraints are described as follows.

Medium Access Constraints We consider long streams of packets, and measure the transmission rate to node A as:

$$\lambda_A = \lim_{n\to\infty} \frac{n}{T_A(n)}. \tag{2}$$

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

Owing to orthogonal receiver directed signaling, every λ_A is bounded independently by a constant C_A, which depends on medium characteristics of the medium and the reception capability of node A. If $\lambda_A \leq C_A$, every packet is successfully received at node A. We assume that the network operates in full duplex mode, where nodes can transmit and receive packets simultaneously as long as the transmission rates are within the specified bounds. Therefore, τ is a *valid network schedule* if and only if $\lambda_A \leq C_A$ for every node A.

Latency Constraint: In general, each relay is allowed to reencrypt packets, delay and reorder arrived packets, and transmit dummy packets. We consider time-sensitive traffic where the average per-packet delay at a node A is bounded by a constant Δ_A.

The schedule τ does not indicate which packets actually travel from source to destination on each route of a session. Further, some of the transmission times in τ could correspond to dummy packets. Therefore, the schedules need to be supplemented with a relaying strategy. The relaying strategy $\mathcal{Z}$ is a set of subsequences of τ that depends on the routes in $\mathbf{S}$, and contains only the transmission times of packets that are relayed from sources to destinations.

Definition 2: Let a session $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$, where each $P(i) = (A(i,1), \cdots, A(i,|P_i|))$ is a valid path. A set of transmission schedules $\mathcal{Z}(\mathbf{S},\tau) = \{\mathcal{Z}_{i,j} : i \leq |\mathbf{S}|, 1 < j \leq |P(i)|\}$ is a *valid relaying strategy* for the pair $\mathbf{S},\tau$ if:

1) $\forall i \leq |\mathbf{S}|, 1 < j \leq m(i)$, $\mathcal{Z}_{i,j} \subset \tau_{A(i,j)}$.
2) For every $i \leq |\mathbf{S}|$, $\{\mathcal{Z}_{i,j} : j < m(i)\}$ satisfy

$$\begin{aligned} Z_{i,j+1}(n) - Z_{i,j}(n) &\geq 0, \\ \lim_{n\to\infty} \sum_{m=1}^{n} \frac{Z_{i,j+1}(m) - Z_{i,j}(m)}{n} &\leq \Delta_{A(i,j)}. \end{aligned} \quad (3)$$

3) If $(A(i,j), A(i,j+1)) = (A(l,m), A(l,m+1))$, then $\mathcal{Z}_{i,j} \cap \mathcal{Z}_{l,m} = \phi$.

In the above definition, condition 2 guarantees that the streams of relayed packets satisfy the delay constraint at every intermediate relay. Condition 3 is required to ensure that, if any pair of nodes is common to multiple routes, the subsequences picked from the transmission schedules are mutually exclusive.

Note that the set of subsequences $\mathcal{Z}$ could be a strict subset of the transmission schedule τ. The transmission times in $\tau/\mathcal{Z}$ would represent dummy packet transmissions. In Section III, we also consider relaying strategies in which data packets can be dropped, so that higher relay rates are achievable. In that case, some of the transmission times in $\tau/\mathcal{Z}$ would correspond to the dropped packets.

The rates of packets relayed from sources to destinations can be determined using $\mathcal{Z}$. Specifically, the relay rates in session $\mathbf{S}$ are denoted by a vector $\mathcal{L}(\mathbf{S},\mathcal{Z}) = (\lambda_r(1), \cdots, \lambda_r(|\mathbf{S}|))$, where

$$\lambda_r(i) = \lim_{n\to\infty} \frac{n}{Z_{i,1}(n)}, \ \forall i.$$

Note that since all the subsequences on a route have same length, it is sufficient to use $\mathcal{Z}_{i,1}$ to compute the relay rate.

C. Performance Metric

The performance metric, *throughput*, is defined as the expected sum-rate of packets relayed from the sources to the destinations per session.

Definition 3: R is defined to be an *achievable throughput with anonymity* α if $\exists q(\tau|\mathbf{S})$ with anonymity α such that

1) For every session $\mathbf{S} = \{P(1), \cdots, P(|\mathbf{S}|)\}$, every realization of τ is a valid network schedule.
2) For every realization of $(\mathbf{S},\tau)$, there exists a valid relaying strategy $\mathcal{Z}(\mathbf{S},\tau)$ such that

$$\mathbb{E}\left(|\mathcal{L}(\mathbf{S},\mathcal{Z})|_1\right) \geq R, \quad (4)$$

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

The goal is to characterize the maximum achievable $R(\alpha)$. In this work, we design transmission and relaying strategies and characterize a lower bound on the maximum achievable $R(\alpha)$.

The maximum achievable $R(0)$ is easily computed given the medium access constraints. When $\alpha = 0$, the maximum sum-rate in a session $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$ can be obtained using the max-flow in $\mathbf{S}$ that satisfies medium access constraints. Let $\mathcal{L}_r^0(\mathbf{S}) = (\lambda_r^0(1), \cdots, \lambda_r^0(|\mathbf{S}|))$ represent the vector of achievable relay rates on the paths in session $\mathbf{S}$ when $\alpha = 0$, and let $\Lambda_r^0(\mathbf{S})$ be the maximum sum-rate. Then,

$$\Lambda_r^0(\mathbf{S}) = \max(\lambda_r^0(1) + \cdots + \lambda_r^0(|\mathbf{S}|)), \quad (5)$$

$$\sum_{i:B\in P(i)} \lambda_r^0(i) \leq C_B, \ \forall B \in \mathcal{V}. \quad (6)$$

The maximum network throughput when anonymity $\alpha = 0$ is then given by the expected sum-rate (expectation over $p(S)$)

$$R(\alpha = 0) = \mathbb{E}(\Lambda_r^0(\mathbf{S})).$$

In the following sections, we design schedules τ and relaying strategies $\mathcal{Z}$ for a general α, and characterize an achievable $R(\alpha)$.

III. Covert Relaying

Our approach to designing schedules and relay strategies derives its motivation from Mix networks, but differs in several key aspects due to properties of multihop wireless networks. First, owing to encrypted packet headers, if incoming and outgoing schedules at a particular node are uncorrelated, an eavesdropper would not be able to detect the flow of traffic through that node. Therefore, traffic from multiple sources are not always required to hide the relaying operation of a relay. Second, it may not be necessary to hide every link of communication in the network. It is possible to reveal certain portions of the routes to the eavesdropper without violating the anonymity requirement.

Using these observations, we propose the following strategy. In every session $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$, we divide the set of relays into two categories, *covert relays* and *visible relays*, which are defined as follows.

Covert Relays: A covert relay B uses an outgoing transmission schedule that is statistically independent of the schedules

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

of all nodes occurring previously in paths that contain B. This would make it impossible for an eavesdropper to correlate the schedule of packets received by B and that received by any subsequent node in the paths that contain B, effectively hiding the relay operation.

Visible Relays: A *visible* relay B generates its transmission schedule depending on the arrival times of packets at B. Specifically, every received packet is immediately relayed by B (processing delays are assumed to be negligible). It is evident that a relay operating under this highly correlated schedule is easily detected by an eavesdropper.

Due to the independent transmission schedule, a covert relay would need to transmit dummy packets, and therefore, achieve lower relay rates than a visible relay. Since the loss in relay rates at every covert relay can reduce overall network throughput, it is necessary, to choose the covert relays optimally to maximize network performance. The choice of covert relays depends on the routes of a session and the desired level of anonymity. Further, we allow randomization of the relay selection to increase the eavesdroppers' confusion.

In the remainder of this section, we describe the relaying strategy for a covert relay and characterize the set of achievable relay rates under the delay and medium access constraints. In Section IV, we present the random selection strategy and using the characterization of covert relay rates, derive the relationship between network throughput and anonymity.

A. Covert Relay Rate Regions

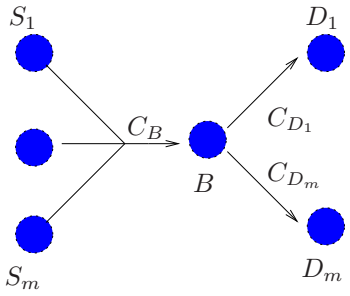


Fig. 3. $m \times 1$ Relay

Consider an $m \times 1$ relay, as shown in Figure 3, where the relay B forwards packets from sources $S_1, \cdots, S_m$ to distinct destinations $D_1, \cdots, D_m$. We shall use the notation $\tau_{S_i,B}$ to denote the schedule of packets transmitted from S_i to B. Since all transmissions to B use the same spreading sequence, $\tau_B = \bigcup \tau_{S_i,B}$. Let $\lambda(S_i, B)$ denote the rate of process $\tau_{S_i,B}$. Due to the medium access constraints,

$$\lambda(B) = \sum \lambda(S_i, B) \leq C_B.$$

Relay B generates an outgoing schedule τ_{B,D_i} ($\subset \tau_{D_i}$) apriori for every pair S_i, D_i. We assume that D_i receives packets only from node B, and hence the maximum rate allocated to τ_{B,D_i} is C_{D_i}. If multiple nodes are transmitting packets to D_i, C_{D_i} can be replaced by the corresponding rate allocated to node B.

The goal is to pick subsequences $\{\mathcal{Z}_{S_i,B}\}, \{\mathcal{Z}_{B,D_i}\}$ from $\{\tau_{S_i,B}\}$,$\{\tau_{B,D_i}\}$ respectively, such that the relaying strategy is valid. Note that, since we are analyzing a single covert relay, $\mathcal{Z}$ is indexed using the node identities. In the general multihop network model, each $\mathcal{Z}_{S_i,B}$ would be represented as $\mathcal{Z}_{i,j}$ depending on which paths in the session contain B.

We are interested in characterizing the set of achievable relay rates $\{\lambda_r(S_i, D_i)\}$ where

$$\lambda_r(S_i, D_i) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{n}{Z_{S_i,B}(n)}$$

is the rate of relayed packets from S_i to D_i through B.

If B were a visible relay, any set of rates satisfying the medium access constraints would be achievable. Specifically, any set of rates that satisfy

$$\lambda_r(S_i, D_i) \leq C_{D_i}, \quad \sum_{i=1}^{m} \lambda_r(S_i, D_i) \leq C_B. \tag{7}$$

are achievable. To characterize the achievable rate region for a covert relay, we use a technique from [16], called the Bounded-Greedy-Match (BGM) algorithm (see Table I). The BGM algorithm, proposed in the context of chaff insertion in stepping stone attacks, is used to match incoming and outgoing transmission times within a strict delay constraint Δ, such that the number of packets dropped are minimized. Under a strict delay constraint, every received packet needs to be forwarded within Δ time units or otherwise dropped. In [17], we used the BGM algorithm and characterized achievable rates for a pair of independent Poisson schedules, when traffic is subjected to a strict delay constraint.

Let $T_{S_i,B}(n), T_{B,D_i}(n)$ represent the arrival time of the n^{th} packet from S_i and departure time of n^{th} packet from B.
1. Initialize $i = 1, j = 1, k = 1$.
2. Let $t = \min\{T_{S_i,B}(i), T_{B,D_i}(j)\}$.
3. If $t = T_{B,D_i}(j)$, then
   i. B transmits a dummy packet at time $T_{B,D_i}(j)$.
   ii. $j = j + 1$.
   else if $T_{B,D_i}(j) - T_{S_i,B}(i) \leq \Delta$
   i. B transmits the i^{th} packet from S_i at $T_{B,D_i}(j)$.
   ii. $Z_{S_i,B}(k) = T_{S_i,B}(i)$, $Z_{B,D_i}(k) = T_{B,D_i}(j)$.
   iii. $i = i + 1, j = j + 1, k = k + 1$.
   else
   i. Drop the i^{th} packet that arrived from S_i.
   ii. $i = i + 1$.
4. Repeat Step 2,3 until the end of the streams.

TABLE I
BOUNDED GREEDY MATCH ALGORITHM

For an average delay constraint, consider the following. After generating the independent outgoing schedule, node B picks epochs from $\tau_{S_i,B}$ and τ_{B,D_i} according to the BGM algorithm, ignoring the delay constraint. In other words, the strict delay constraint is treated as infinite. In order to ensure that every data packet is relayed, node B is required to transmit a strictly positive rate of dummy transmissions. Although, all sources utilize identical spreading sequences, the relay can decode packet headers and distinguish the individual processes. The overall relaying strategy is therefore a parallel application of the BGM algorithm on each pair of incoming and outgoing processes. The following theorem provides an

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

analytical characterization of the achievable rate region, when the schedules are independent Poisson processes.

Theorem 1: If $\{\tau_{S_i,B}\}_{i=1}^m$ and $\{\tau_{B,D_i}\}_{i=1}^m$ are independent Poisson processes, any set of rates $\{\lambda_r(S_i,D_i)\}_{i=1}^m$ that satisfy

$$\lambda_r(S_i,D_i) \le C_{D_i} - \frac{1}{\Delta_B}, \forall i, \quad \sum_{i=1}^m \lambda_r(S_i,D_i) \le C_B \qquad (8)$$

are achievable by covert relay B.

Proof: Refer to Appendix

From the theorem, it is clear that the rates achievable by a covert relay are less than that of a visible relay. Specifically, the relay B transmits additional dummy packets at rate of $\frac{1}{\Delta_B}$ for every source-destination pair to maintain independence in schedules.

B. Achievable Rate Regions with Packet Drops

If the relay B is allowed to drop data packets, then the achievable relay rate region can be further improved. The strategy is as follows. Let the sources and the relay transmit packets at the maximum rates possible. The relay B uses the BGM algorithm on each pair $\tau_{S_i,B}, \tau_{B,D_i}$ with a finite strict delay Δ_i^*. The Δ_i^* is chosen such that for the pair of transmission rates $\lambda(S_i,B), \lambda(B,D_i)$, the average per packet delay is bounded by Δ_B. Since $\Delta_i^* < \infty$, we know from [17] that the BGM algorithm results in a non-zero rate of dropped packets for independent Poisson processes, and the achievable relay rate would be strictly less than the source transmission rate $\lambda(S_i,B)$.

Theorem 2: The set of relay rates $\{\lambda(S_i,B)\}$ are achievable for an $m\times 1$ covert relay with packet drops if $\exists\{\lambda(S_i,B)\}$ s.t.

$$\begin{aligned} \lambda_r(S_i,B) &\le \lambda(S_i,B)\frac{C_{D_i}\left(e^{-\Delta_i^*(\lambda(S_i,B)-C_{D_i})}-1\right)}{C_{D_i}e^{-\Delta_i^*(\lambda(S_i,B)-C_{D_i})}-\lambda(S_i,B)}, \\ &\triangleq \lambda(S_i,B)(1-\epsilon_B(S_i,D_i)) \\ \sum_i \lambda(S_i,B) &\le C_B, \end{aligned}$$

where Δ_i^* is the solution to

$$\frac{1+e^{\Delta_i^*(C_{D_i}-\lambda(S_i,B))}\left[\Delta_i^*(C_{D_i}-\lambda(S_i,B))-1\right]}{(\lambda(S_i,B)-C_{D_i})\left[1-e^{\Delta_i^*(C_{D_i}-\lambda(S_i,B))}\right]} = \Delta_B.$$

Proof: Refer to Appendix

$\epsilon_B(S_i,D_i)$ in Theorem 2 denotes the fraction of packets dropped by B for the packet stream from source S_i. When $\Delta_B \ge \frac{1}{C_{D_i}-C_B}, \forall i$, the packet drop rate vanishes and the covert and visible relay rate regions are identical.

Figure 4 provides a comparison of the different rate regions. The figure clearly demonstrates that allowing packet losses increases the set of achievable rates at a covert relay. At the (linear) portion of the regions, where the boundaries coincide, the transmission rates $\lambda(S_1,B), \lambda(S_2,B)$ satisfy the conditions $C_{D_i}-\lambda(S_i,B) \ge \frac{1}{\Delta_B}, i=1,2$. In that case, covert relaying does not reduce the relay rates (albeit requires dummy transmissions by B).

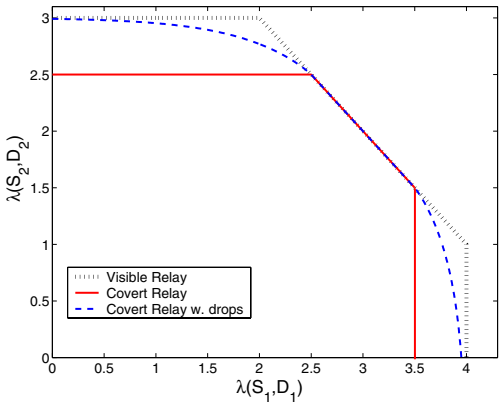


Fig. 4. 2×1 rate region: $C_B=5, C_{D_1}=4, C_{D_2}=3, \Delta=2$

An important feature in the algorithms presented is that the relays do not require prior knowledge about transmission schedules of the source nodes. The decision to transmit any packet is based on events occurring between its arrival time and the subsequent departure time. This makes it particularly attractive for a decentralized implementation of the scheduling, where nodes are unaware of transmission schedules of other nodes in the paths.

IV. Covert Relay Selection

Using the characterized rate regions in Sections III-A and III-B, we now proceed to design the strategy to select relays in each session to be covert. In any session $\mathbf{S}$, if a subset of relays $\mathbf{B}$ are chosen to be covert, then the schedules τ and the relaying strategy $\mathcal{Z}$ can be obtained from the algorithms described previously.

We model the set of covert relays $\mathbf{B}$ as a random variable with a conditional probability mass function $\{q(\mathbf{B}|\mathbf{S}) : \mathbf{B} \in 2^{\mathcal{V}}, \mathbf{S}\in\mathcal{S}\}$. The eavesdropper is assumed to have knowledge of the distribution $q(\mathbf{B}|\mathbf{S})$, however, he is unaware of the realization of the randomness. The goal is to optimize the conditional p.m.f $\{q(\mathbf{B}|\mathbf{S})\}$ so that network throughput is maximized for a given level of anonymity α. In order to perform this optimization, we need to characterize the achievable network throughput and the eavesdropper's "observed session" for any given pair $\mathbf{S},\mathbf{S}$.

A. Eavesdropper's Observed Session

When a relay is visible, we assume that the eavesdropper perfectly correlates the transmission times of packets to the relay and those of the subsequent node in a route containing the relay. As a result, the eavesdropper would perfectly detect certain portions of the routes in the session. The perfectly detected portion is denoted by a set of paths $\hat{\mathbf{S}} \in 2^{\mathcal{P}(\mathcal{G})}$, which can be expressed as a deterministic function of the actual session $\mathbf{S}$ and the set of covert relays $\mathbf{B}$.

Define function $t : 2^{\mathcal{P}(\mathcal{G})}\times\mathcal{V} \to 2^{\mathcal{P}(\mathcal{G})}$ such that for a set of paths $\mathbf{P}$, $t(\mathbf{P},B)$ contains the observed paths when only node B is covert. In a receiver directed signaling network with encrypted headers, it is not possible to detect the source

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

node in any route. Therefore, $t(\mathbf{P}, \phi)$ is obtained by removing source nodes from every path in $\mathbf{P}$.

When $B \neq \phi$, a path $P \in \mathcal{P}(\mathcal{G})$ belongs to $t(\mathbf{P}, B)$ if and only if it satisfies one of the following conditions:
1. $\exists P' = (A_1, \cdots, A_k, B, A_{k+1}, \cdots, A_n) \in \mathbf{P}$, such that $P = (A_1, \cdots, A_k, B)$ or $P = (A_{k+1}, \cdots, A_n)$.
2. $P \in \mathbf{P}$ and $B \notin P$.

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

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

$$\hat{\mathbf{S}} = t(\cdots(t(t(\mathbf{S}, \phi), B_1)\cdots), B_m) \stackrel{\triangle}{=} \mathbf{T}(\mathbf{S}, \mathbf{B}). \tag{9}$$

$\hat{\mathbf{S}}$ denotes the portion of the session perfectly detected by the eavesdropper. In Section V, it will be shown that $\hat{\mathbf{S}}$ is a sufficient statistic for the eavesdropper in detecting $\mathbf{S}$.

B. Throughput Function

The relaying strategies in Section III were designed to maximize achievable rates at a single covert relay. Extending those results to multihop routes, we characterize the loss in sum-rate when a subset of relays $\mathbf{B}$ are covert in session $\mathbf{S}$.

The loss in sum-rate depends on the delay requirement at each covert relay in $\mathbf{B}$. Let $\mathcal{L}_r(\mathbf{S}, \mathbf{B}) = (\lambda_r^{\mathbf{B}}(1), \cdots, \lambda_r^{\mathbf{B}}(|\mathbf{S}|)$ represent the achievable relay rates from sources to destinations for the session $\mathbf{S} = (P(1), \cdots, P(|\mathbf{S}|))$, when nodes in $\mathbf{B}$ are covert, and let $\Lambda_r(\mathbf{S}, \mathbf{B}) \stackrel{\triangle}{=} \sum_{i=1}^{|\mathbf{S}|} \lambda_r^{\mathbf{B}}(i)$ denote the achievable sum-rate.

Consider the relaying strategies without packet loss (Section III-A). We know that, at a covert relay B, the delay constraint of Δ_B incurs a reduction in relay rate of $\frac{1}{\Delta_B}$. Therefore,

$$\lambda_r^{\mathbf{B}}(i) = \lambda_r^0(i) - \sum_{B \in \mathbf{B} \cap P(i)} \frac{1}{\Delta_B}.$$

When packet drops are allowed, then using the relaying strategy described in Section III-B, if $A(i,j)$ represents the j^{th} node in path $P(i)$,

$$\frac{\lambda_r^{\mathbf{B}}(i)}{\lambda_r^0(i)} = \prod_{j: A(i,j) \in \mathbf{B} \cap P(i)} \left(1 - \epsilon_{A(i,j)}(A(i,j-1), A(i,j+1))\right).$$

where $\epsilon_B(A, C)$ represents the fraction of data packets transmitted by node A to C, that are dropped by covert relay B. Using the techniques in Section III, $\epsilon_B(A, C)$ can be characterized as a function of the transmission rates of nodes A, B. Note that Theorem 2 provides an analytical characterization of rates at the first covert relay following a source node. When a path contains multiple covert relays, the schedules of data packets after the first covert relay are no longer Poisson distributed, and hence the results of the Theorem do not directly apply. Analytical characterization of achievable rates using multiple covert relays is, in general, cumbersome, but a numerical evaluation can be easily performed.

V. Performance Characterization

A. Throughput-Anonymity Tradeoff

Using the eavesdropper estimate and the throughput characterization, we now proceed to optimize $\{q(\mathbf{B}|\mathbf{S})\}$. For a desired α, the optimal distribution $q(\mathbf{B}|\mathbf{S})$ can be obtained using a brute force search over a large dimensional probability simplex. The procedure would be computationally intensive, and impractical for large networks. The following result, however, proves the duality of this optimization to information theoretic rate-distortion function, which can then be used to obtain the optimal strategy efficiently and characterize the maximum throughput $R(\alpha)$ analytically.

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

$$d(\mathbf{S}, \hat{\mathbf{S}}) = \begin{cases} \min_{\mathcal{B}(\mathbf{S}, \hat{\mathbf{S}})} \Lambda_r^0(\mathbf{S}) - \Lambda_r(\mathbf{S}, \mathbf{B}) & \mathcal{B}(\mathbf{S}, \hat{\mathbf{S}}) \neq \phi \\ \infty & \text{o.w.} \end{cases} \tag{10}$$

where $\mathcal{B}(\mathbf{S}, \hat{\mathbf{S}}) = \{\mathbf{B} : \hat{\mathbf{S}} = \mathbf{T}(\mathbf{S}, \mathbf{B})\}$. Then, a throughput R is achievable with $\alpha-$anonymity if

$$R(0) - R(\alpha) \geq D\left(H(\mathbf{S})(1-\alpha)\right),$$

where $D(r)$ is the *Distortion-Rate* function defined as

$$D(r) = \min_{q(\hat{\mathbf{S}}|\mathbf{S}): I(\mathbf{S}; \hat{\mathbf{S}}) \leq r} \mathbb{E}(d(\mathbf{S}, \hat{\mathbf{S}})). \tag{11}$$

Proof: Refer to Appendix.

The above theorem characterizes $R(\alpha)$ using the single letter representation of a rate-distortion function. The loss function $d(\mathbf{S}, \hat{\mathbf{S}})$ in (10) represents the throughput reduction due to covert relaying. Although the loss function parameters do not explicitly include the set of covert relays $\mathbf{B}$, it is shown in the proof of Theorem 3 that given $\mathbf{S}, \hat{\mathbf{S}}$, the set of covert relays $\mathbf{B}$ is unique (the minimization in (10) is trivial). Therefore, the distribution $q(\mathbf{B}|\mathbf{S})$ to chose covert relays is equivalent to the distortion minimizing distribution in (11). As a result, the Blahut-Arimoto algorithm [18] provides an efficient iterative technique to obtain $q(\mathbf{B}|\mathbf{S})$ and the achievable network throughput $R(\alpha)$.

In order to achieve the throughput of Theorem 3, it is necessary for every relay to be aware of the entire session $\mathbf{S}$ and to use an identical randomizer. From a practical perspective, this could be achieved, if nodes exchange local messages with their neighbours such that they reach a distributed consensus about the session. Since total number of sessions is finite, perfect convergence can be reached in finite time, assuming perfect transmissions. However, there may be network applications where each node is only aware of its adjacent nodes in the paths. This is especially important to prevent information retrieval by compromising nodes in the network. Under such circumstances, we propose a decentralized alternative, where nodes are not required to exchange messages.

B. Decentralized Implementation

Define function $l : \mathcal{V} \times \mathcal{S} \mapsto 2^{\mathcal{V} \times \mathcal{V}}$, where $l(A, \mathbf{S})$ denotes the information available to node A in session $\mathbf{S}$. Then,

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

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

In other words, $l(B,\mathbf{S})$ is the set of node pairs $(A(i,j-1), A(i,j+1))$ such that node B relays packets from $A(i,j-1)$ to $A(i,j+1)$ on route $P(i)$ of $\mathbf{S}$.

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. Further, we do not assume any common randomness available to the nodes, and hence, the decisions of multiple nodes are conditionally independent (conditioned on the session). Accordingly, we define function

$$q_c : \mathcal{V} \times 2^{\mathcal{V}\times\mathcal{V}} \mapsto [0,1],$$

where $q_c(B, l(A,\mathbf{S}))$ is the probability that node B is covert in session $\mathbf{S}$. Owing to conditional independence, the probability that nodes in a subset $\mathbf{B}$ are covert in session $\mathbf{S}$ is given by:

$$q(\mathbf{B}|\mathbf{S}) = \prod_{B\in\mathbf{B}} q_c(B, l(B,\mathbf{S})) \prod_{B\notin\mathbf{B}} (1 - q_c(B, l(B,\mathbf{S}))). \quad (12)$$

Let Q^* represent the set of all conditional probability mass functions $q(\mathbf{B}|\mathbf{S})$, such that there exists covert probability function $q_c(\cdot,\cdot)$ which satisfies (12) for every $(\mathbf{B},\mathbf{S})$. From Theorem 3, we know that the pairs of variables $(\mathbf{S},\mathbf{B})$ and $(\mathbf{S},\hat{\mathbf{S}})$ have a one-one correspondence. Therefore, Q^* corresponds to an equivalent set Q^{**} of conditional probabilities $q(\hat{\mathbf{S}}|\mathbf{S})$.

Theorem 4: A throughput $R(\alpha)$ that satisfies

$$R(0) - R(\alpha) \geq D'\left(H(\mathbf{S})(1-\alpha)\right),$$

is achievable with a decentralized strategy where

$$D'(r) = \min_{q(\hat{\mathbf{S}}|\mathbf{S})\in Q^{**}: I(\mathbf{S};\hat{\mathbf{S}})\leq r} \mathbb{E}(d(\mathbf{S},\hat{\mathbf{S}})). \quad (13)$$

Proof: Since the minimizing distribution $q(\hat{\mathbf{S}}|\mathbf{S})$ is an element of Q^{**}, it corresponds to a conditional distribution $q(\mathbf{B}|\mathbf{S})$ that is expressible in the form (12), which in turn provides the decentralized strategy through the covert probability function $q_c(\cdot|\cdot)$. The achievability of $R(\alpha)$ then follows from the proof of Theorem 3. □

Note that the minimization in (13) is over a subset of the probability simplex, 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. This is illustrated in the following example.

C. Example

Consider the switching network example shown in Figure 5. During any network session, each source S_i picks a distinct destination D_j, and for each pairing $\{S_i, D_j\}$ there is a fixed set of paths. The set of possible sessions, $\mathcal{S}$, contains 24 elements (distinct $\{S_i, D_j\}$ pairings) which are assumed equiprobable. For this setup, Figure 6 plots the throughput-anonymity region for the different strategies.

As can be seen, when all relays are visible, the maximum sum-rate is achieved with a strictly positive anonymity level.

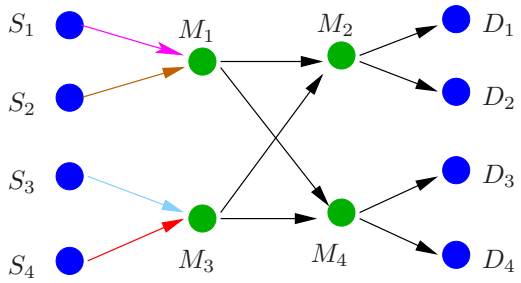


Fig. 5. Switching Network: $\{S_i\}$ transmit to $\{D_i\}$ through relays $\{M_i\}$.

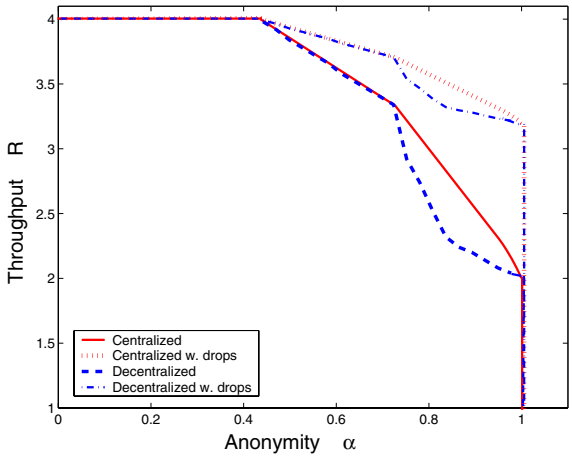


Fig. 6. Throughput-Anonymity Region for 4×4 switching network with $C_{M_i} = C_{D_i} = 1 \ \ \forall i, \Delta_{M_1} = \Delta_{M_3} = 4, \Delta_{M_2} = \Delta_{M_4} = 2$.

This is because, receiver directed signaling induces some uncertainty about the source nodes. Note that the throughput-anonymity curves resulting from the centralized strategy are convex; this is due to the average nature of the metrics, namely equivocation and expected sum-rate. The figure also illustrates the gain in throughput when relaying strategies are allowed to drop data packets.

An interesting observation is that the difference between centralized and decentralized strategies is significant only at higher anonymity levels. This is because, a higher level of anonymity would require multiple relays in every route to be covert, and the performance is therefore affected by the lack of common randomness. The non-convexity of the decentralized throughput can be attributed to the disparate knowledge of the session at different nodes; without common knowledge of the session, it is not possible to time-share multiple strategies.

VI. Conclusions

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 wireless network. Based on the metric, we designed scheduling and relaying strategies to maximize network performance with a guaranteed level of anonymity. Although we consider specific constraints on delay and medium access, the ideas of covert relaying and the randomized selection are quite general, and apply to arbitrary multihop wireless networks.

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.

REFERENCES

[1] J. Deng, R. Han, and S. Mishra, "Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks," in *IEEE International Conference on Dependable Systems and Networks (DSN)*, (Florence, Italy), pp. 594–603, June 2004.

[2] S. Jiang, N. H. Vaidya, and W. Zhao, "A mix route algorithm for mix-net in wireless mobile ad hoc networks," in *Proceedings of IEEE Mobile Sensor and Ad-hoc and Sensor Systems*, pp. 406–415, October 2004.

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

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

[5] 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.

[6] A. Serjantov, R. Dingledine, and P. Syverson, "From a trickle to a flood: Active attacks on several MIX types," in *Proceedings of the Fifth International Workshop on Information Hiding (IH'02), Lecture Notes in Computer Science*, vol. 2578, (Noordwijkerhout, The Netherlands), pp. 36–52, October 2002.

[7] 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.

[8] 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.

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

[10] R. E. Newman, I. S. Moskowitz, P. Syverson, and A. Serjantov, "Metrics for traffic analysis prevention," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2003)*, Springer-Verlag, LNCS 2760, April 2003.

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

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

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

[14] 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.

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

[16] A. Blum, D. Song, and S. Venkataraman, "Detection of Interactive Stepping Stones: Algorithms and Confidence Bounds," in *Conference of Recent Advance in Intrusion Detection (RAID)*, (Sophia Antipolis, French Riviera, France), September 2004.

[17] P. Venkitasubramaniam, T. He, and L. Tong, "Networking with secrecy constraints," in *Proc. of 2006 IEEE Military Communications Conference*, (Washington D.C), Oct. 2006.

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

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

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

APPENDIX

A. Proof of Theorems 1 and 2

Consider the application of the BGM algorithm with finite delay constraint Δ_i^*. For this analysis, we adopt a technique used in [19]. Consider the two point processes $\tau_{S_i,B}, \tau_{B,D_i}$. If a packet in $\tau_{S_i,B}$, say at time t, is dropped by the BGM algorithm, we insert a virtual packet at time $t+\Delta_i^*$ in τ_{B,D_i}. Similarly, if a packet at time t in τ_{B,D_i} is designated as dummy packet, we insert a virtual packet at time t in $\tau_{S_i,B}$. From [19], we know that the difference process $\mathcal{W} = \{T_{B,D_1}(i) - T_{S_1,B}(i)\}$ is a random walk with two absorbing barriers at 0 and Δ. Therefore, the average delay incurred by the BGM algorithm is equal to the expected step size of the process $\mathcal{W}$. Following the exposition in ([20], Page 67), the cumulative distribution of step size x in $(0,\Delta)$ is given by

$$H(x) = \frac{1 - \frac{\lambda(S_i,B)}{C_{D_i}} \exp(\Delta_i^* + x)(\lambda(S_i,B) - C_{D_i})}{1 - \frac{\lambda(S_i,B)^2}{C_{D_i}^2} \exp(\Delta_i^*(\lambda(S_i,B) - C_{D_i}))}. \quad (14)$$

Using the expression above, the average delay Δ for the BGM algorithm with strict delay Δ_i^* can be evaluated as:

$$\mathbb{E}(x) = \frac{1 + \exp(\Delta_i^*(\lambda(S_i,B) - C_{D_i}))\,[\Delta(\lambda(S_i,B) - C_{D_i}) - 1]}{(C_{D_i} - \lambda(S_i,B))\,[1 - \exp(\Delta_i^*(\lambda(S_i,B) - C_{D_i}))]}.$$

Therefore, for any pair of transmission rates $\lambda(S_i,B), C_{D_i}$, the BGM algorithm with strict delay Δ_i^* that solves $\mathbb{E}(x) = \Delta_B$, would incur an average delay of Δ_B. A finite strict delay constraint, however, results in non-zero packet loss [17].

In order to design relaying strategies that incur zero packet loss, we set $\Delta_i^* = \infty$ and reduce $\lambda(S_i,B)$ such that the average delay constraint is satisfied. Specifically, as $\Delta_i^* \to \infty$,

$$\begin{aligned}\mathbb{E}(x) &= \frac{1 + \exp(\Delta_i^*(\lambda(S_i,B) - C_{D_i}))\Delta_i^*(\lambda(S_i,B) - C_{D_i})}{(C_{D_i} - \lambda(S_i,B))\,[1 - \exp(\Delta_i^*(\lambda(S_i,B) - C_{D_i}))]} \\ &= \frac{1}{C_{D_i} - \lambda(S_i,B)}.\end{aligned}$$

If $\lambda(S_i,B) \le C_{D_i} - \frac{1}{\Delta_B}$, then the BGM algorithm with $\Delta_i^* = \infty$ satisfies the average constraint with no packet drops. □

B. Proof of Theorem 3

Consider the distribution $q^*(\hat{\mathbf{S}}|\mathbf{S})$ that minimizes (11). From the definition of $d(\mathbf{S},\hat{\mathbf{S}})$, it is easy to see that if $\nexists \mathbf{B}\ s.t.\ \hat{\mathbf{S}} \neq \mathbf{T}(\mathbf{S},\mathbf{B})$, then $q^*(\hat{\mathbf{S}}|\mathbf{S}) = 0$. Given $\mathbf{S},\hat{\mathbf{S}}$, we now show that the set of covert relays $\mathbf{B}$ is uniquely determined.

Suppose $\exists \mathbf{B}_1 \neq \mathbf{B}_2$ such that $\hat{\mathbf{S}}(\mathbf{S},\mathbf{B}_1) = \hat{\mathbf{S}}(\mathbf{S},\mathbf{B}_2)$. Then without loss of generality, 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

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

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 $\hat{\mathbf{S}}(\mathbf{S},\mathbf{B}_1)$ would contain a path that ends in B_{11}, whereas $\hat{\mathbf{S}}(\mathbf{S},\mathbf{B}_2)$ cannot contain such a path. Therefore, we have 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, $I(\mathbf{S};\hat{\mathbf{S}}) \le (1-\alpha)H(\mathbf{S})$ ensures that an anonymity α is guaranteed. Further, for every $\mathbf{B}$, the function $d(\mathbf{S},\hat{\mathbf{S}})$ evaluates the difference in achievable sum-rates $\Lambda_r^0(\mathbf{S})$ and $\Lambda_r(\mathbf{S},\mathbf{B})$. Therefore, taking expectation over $q^*(\mathbf{B}|\mathbf{S})$, the throughput is shown to be achievable. □

Authorized licensed use limited to: Cornell University. Downloaded on October 7, 2008 at 15:30 from IEEE Xplore. Restrictions apply.