Relay Secrecy in Wireless Networks with Eavesdropper

Parvathinathan Venkitasubramaniam, Ting He and Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email : {pv45, th255, lt35}@cornell.edu

***Abstract*— Anonymous monitoring of transmissions in a wireless network by eavesdroppers can provide critical information about the data flows in the network. It is, therefore, necessary to design network protocols that maintain secrecy of routes from eavesdroppers. In this work, we present a mathematical formulation of route secrecy when eavesdroppers observe transmission epochs of nodes. We propose a scheduling technique to provide complete secrecy of routes, and based on that, characterize achievable rate regions for two hop source destination pairs with a common relay under two PHY models : transmitter directed and receiver directed spread spectrum signaling. The results are also extended to the case when an additional constraint on packet loss is imposed.**

***Index Terms* - Network Security, Secrecy, Multiple Access, Packet Loss.**

I. Introduction

Wireless networks are prone to anonymous monitoring by eavesdroppers, who wish to gain valuable network information, namely source-destination pairs and data flows. Equipped with this knowledge, it is possible for malicious adversaries to then target specific routes for intrusion or jamming. Active intrusion attacks can be countered by sophisticated intrusion detection mechanisms. On the other hand, passive monitoring does not affect the network operation, and is not possible to detect. It is, therefore, necessary to modify the network protocols, so that information about data flows or source-destination pairs are not traceable by eavesdroppers monitoring node transmissions.

The inference of routing information from monitored transmissions, known as traffic analysis attack, is done in a variety of ways. The eavesdropper can identify a flow of traffic by correlating packet contents, packet lengths or transmission epochs across multiple nodes. Encrypting and random padding of bits are some measures adopted to remove the correlation of contents and lengths of packets across nodes. In this work, we are interested in the design of secure transmission schedules to prevent inference of routes by traffic analysis.

This work is supported in part by TRUST (The Team for Research in Ubiquitous Secure Technology), which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: Cisco, ESCHER, HP, IBM, Intel, Microsoft, ORNL, Qualcomm, Pirelli, Sun and Symantec, and the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

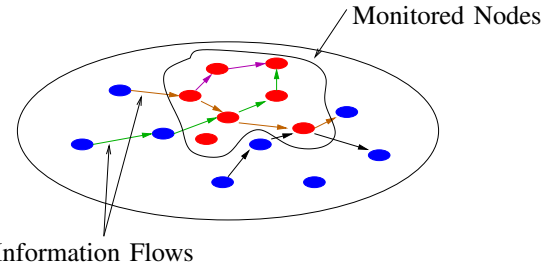

Fig. 1: Wireless Network with Eavesdropper

In the absence of an eavesdropper, the transmission schedule of relaying nodes are dependent on the arrival of packets, subject to the delay requirements. When transmissions are monitored, however, it may be necessary to decouple the transmission schedule of the nodes from the actual traffic flow to prevent flow correlation. For delay sensitive traffic, this may not be possible without reducing network performance. In particular, the design of such schedules would require transmission of dummy packets and could also result in packet drops. It is, therefore, necessary to characterize the achievable network performance under secrecy constraints.

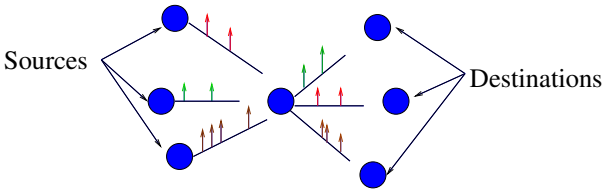

Fig. 2: $m \times 1$ Relay

In this work, we consider the problem of secrecy for a one-hop multiplex relay as shown in Figure 2. In particular, we characterize the set of achievable relay rates, when packets are subjected to a strict delay criterion under two PHY models : transmitter directed and receiver directed spread spectrum signaling. We provide transmission schemes to prevent flow correlation and show that as the delay increases, the achievable rate region converges to the optimal region. Furthermore, we also present achievable rate

regions when an additional constraint on packet loss is imposed.

A. Related Work

A countermeasure to traffic analysis attacks was first provided through the notion of Mix-net by Chaum[1]. A Mix is an intermediate node that re-encrypts and reorders packets from multiple sources to prevent matching of source and destination streams. The idea has been used effectively in providing anonymous communication for internet applications[2], [3], [4], [5].

For wireless networks without delay constraints, independent random transmission schedules were used in[6] to prevent flow correlation. The use of randomized routes as a countermeasure to traffic analysis attacks has been considered in [7], [8]. For low latency networks, it has been shown in [9] that simple mixing techniques are not effective to prevent correlation of transmission epochs. Their proposed solution utilized the idea of transmitting dummy packets to make departure epochs identical irrespective of the flows. The idea of having fixed transmission schedules independent of routes has also been considered in [10], where the authors give bounds on the performance loss incurred due to the secrecy constraints.

The theoretical framework for secrecy in this work is motivated by the notion of equivocation developed by Shannon in [11]. The secrecy constraint we consider is a special case of Shannon's equivocation, known as maximum secrecy[12], wherein the observations provide zero information about the source.

II. Problem Setup

A. Definitions

Let the network be represented by a graph $G = (V, E)$, where V is the set of nodes and E is the set of links between pairs of nodes. A link (A, B) belonging to E denotes that node B can listen to the transmissions from A and vice-versa. Let $\mathcal{Y}_A = \{Y_A(1), Y_A(2), \cdots\}$ denote the time instants (known as *departure epochs*) at which A transmits packets. The *transmission rate* T_A of a node A is defined as the average number of packets per unit time transmitted by A. In other words,

$$T_A = \lim_{n\to\infty} \frac{n}{Y_A(n)}.$$

Packets from source nodes are routed through intermediate relays before they reach the destination. In general, the tasks carried out by a relay can be multi various; it can choose to decode and re-encode blocks of packets, it can relay unaltered packets after a random delay or it can re-order the packets before transmission. Re-encryption and packet padding occur at every node/relay to prevent any content based correlation. We are concerned with the kind of traffic, wherein each packet needs to be relayed within a fixed delay constraint Δ. We restrict the tasks of a relay to packet-reordering and timing perturbation. Depending on its transmission schedule, a relay picks departure epochs for the arriving packets such that the delay constraint is satisfied. A packet that is not relayed within Δ time units after arrival is dropped. A formal definition of a relay function is given as follows.

Let $\mathcal{Y}_A = \{Y_A(1), Y_A(2), \cdots, Y_A(n)\}$ represent the departure epochs of packets from A and $\mathcal{Y}_B = \{Y_B(1), Y_B(2), \cdots, Y_B(n)\}$ represent the departure epochs of packets from B. A 1×1 *relay map* is an algorithm that picks a subsequence $\mathcal{Y}_A^s$ of $\mathcal{Y}_A$ and an equal length subsequence $\mathcal{Y}_B^s$ of $\mathcal{Y}_B$ such that $\forall i, 0 \le Y_B^s(i) - Y_A^s(i) \le \Delta$.

If $|\mathcal{Y}_A| = n$ and $|\mathcal{Y}_A^s| = k(n)$, then the *relay rate* $\lambda(\mathcal{M})$ of the 1×1 relay map $\mathcal{M}$ is given by

$$\lambda = \lim_{n\to\infty} \frac{k(n)}{Y_A^s(k(n))}.$$

The rate of a relay map is dependent on the transmission rates of the nodes.

A single node can serve as a relay to multiple sources. An $m \times 1$ relay map is an algorithm that picks subsequences $\mathcal{Y}_{A_1}^s, \mathcal{Y}_{A_2}^s, \cdots, \mathcal{Y}_{A_m}^s$ from departure epochs of m nodes $A_1, \cdots, A_m$ and a subsequence Y_B^s from the departure epoch of the relay node B such that

1) $|\mathcal{Y}_B^s| = \sum_{i=1}^m |\mathcal{Y}_{A_i}^s|$.
2) Let $\mathcal{Y}^s$ be the sequence formed by the concatenating $\mathcal{Y}_{A_1}^s, \cdots, \mathcal{Y}_{A_m}^s$ and ordering the epochs in ascending order. Then,

$$\forall i \le |\mathcal{Y}^s|, 0 \le Y_B^s(i) - Y^s(i) \le \Delta.$$

An $m \times 1$ relay map is associated with a relay rate vector $\lambda(\mathcal{M}) = (\lambda_1, \cdots, \lambda_m)$ which is given by

$$\lambda_i = \lim_{n\to\infty} \frac{k_i(n)}{Y_{A_i}^s(k_i(n))},$$

where $k_i(n) = |\mathcal{Y}_{A_i}^s|$.

B. Medium Access Constraints

Nodes in a wireless network share a common channel and transmissions are susceptible to fading and interference. Depending on the PHY model, the rates of transmission are subjected to some medium access constraints specified by a region of rate vectors $\mathcal{C}$. If the transmission rates of the nodes belong to $\mathcal{C}$, the packets are received successfully at the receiving node. To this extent, we consider two different spread spectrum signaling models : *transmitter directed* and *receiver directed* spreading sequences.

Transmitter Directed Spreading Sequences : Each transmitting node in a shared channel uses an orthogonal spreading code to transmit its packets. The constraints on transmission rates for the nodes are therefore independent. In other words, for a set of

nodes $A_1, \cdots, A_n$, the medium access region is given by

$$\mathcal{C} = \{(T_{A_1}, \cdots, T_{A_n} : T_{A_i} \leq C_{A_i},\ i = 1, \cdots, n\}.$$

Receiver Directed Spreading Sequences : The nodes transmitting to a common node/relay use the same spreading sequence. Each receiving node has an independent rate constraint. The sum-rate of nodes transmitting to a single receiver is therefore bounded by a maximum value. If nodes $A_1, \cdots, A_n$ transmit packets to node B, the region $\mathcal{C}$ is given by

$$\mathcal{C} = \{(T_{A_1}, \cdots, T_{A_n} : \sum T_{A_i} \leq C_B\}.$$

C. Secrecy

As mentioned earlier, by correlating transmission epochs from multiple nodes, the eavesdropper can obtain information about routes within the network. The goal is, therefore, to schedule transmissions so as to maximize the secrecy of the routes with respect to the eavesdropper.

Secrecy can be formally defined as follows. Let $\mathcal{A} = \{A_1, A_2, \cdots, A_k\}$ be a set of nodes and $\mathcal{F} \subset 2^{\mathcal{A}}$ denote the set of all ordered node-pairs in $\mathcal{A}$ ($|\mathcal{F}| = |\mathcal{A}|(|\mathcal{A}| - 1)$). Since transmissions from nodes not physically connected can be correlated to infer a flow, it is necessary to consider all possible node-pairs. During a given session, the set of node-pairs in $\mathcal{F}$ that require non-zero relay rate is denoted by the flow vector $F \subset 2^{\mathcal{F}}$. We define $\mathcal{A}$ to have *complete relay secrecy* if the transmission epochs of the nodes in $\mathcal{A}$ and F are independent. In other words, the conditional distribution

$$p(\mathcal{Y}_{A_1}, \mathcal{Y}_{A_2}, \cdots, \mathcal{Y}_{A_k}|F) = p(\mathcal{Y}_{A_1}, \mathcal{Y}_{A_2}, \cdots, \mathcal{Y}_{A_k}). \quad (1)$$

If for any flow vector F, the joint distribution of transmission epochs is unaltered, then it is impossible to infer the flow to any degree of accuracy.

D. Achievable Rates

A rate vector $\mathbf{R} = (R_1, \cdots, R_m)$ for a set of node-pairs with common relay $\{(A_1, B), (A_2, B), \cdots, (A_m, B)\}$ is an achievable rate vector, if there exists a conditional distribution $p(\mathcal{Y}_{A_1}, \mathcal{Y}_{A_2}, \cdots, \mathcal{Y}_{A_m}|F)$ and an $m \times 1$ relay map such that following conditions are satisfied

1) The transmission rate $\{T_{A_1}, T_{A_2}, \cdots, T_{A_m}, T_B\}$ satisfy the medium access constraints (4).
2) For every realization $(\mathcal{Y}_{A_1}, \cdots, \mathcal{Y}_{A_m})$,
$$\lambda_i(\mathcal{M}) \geq R_i, i = 1, \cdots, m.$$
3) $\{A_1, \cdots, A_m, B\}$ have complete relay secrecy.

In the following section, we present an achievable rate region for the special case of providing relay secrecy for an $m \times 1$ multiplex relay (Fig. 3), where a single node relays packets from m nodes. The results

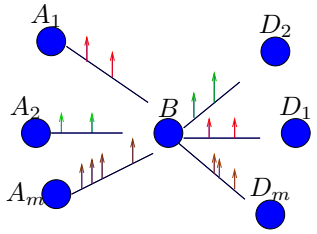

Fig. 3: $m \times 1$ Relay

are presented for the two PHY models discussed in Section II-B.

III. Rate Region

In the absence of an eavesdropper, the flow-rates achievable in a network can be obtained purely from the topology and medium access restrictions. In the presence of eavesdropper, however, the secrecy condition imposes additional constraints which can lower the achievable rates.

The secrecy condition in (1) indicates that the distribution of transmission epochs are independent of the flows. A special case of this condition is when, the transmission schedule of each node is drawn from an independent distribution and the marginal distributions are not dependent on the flows. In other words, for a set of nodes $A_1, \cdots, A_k$

$$p(\mathcal{Y}_{A_1}, \cdots, \mathcal{Y}_{A_k}|F) = p(\mathcal{Y}_{A_1})p(\mathcal{Y}_{A_2}) \cdots p(\mathcal{Y}_{A_k}).$$

Statistical independence of departure epochs is a sufficient condition to ensure relay secrecy. In general, it may be possible to design schedules such that the transmission epochs are not independent and yet guarantee relay secrecy. The notion of flow independent schedules has also been considered in [10], [7], wherein the transmission schedules were fixed apriori irrespective of the data flows.

We assume that the sources generate packets at Poisson time points which determine the schedules of the source nodes. In order to satisfy the secrecy condition, the relay nodes generate departure epochs from independent Poisson processes. To an eavesdropper monitoring the nodes, it is impossible to decipher the actual flows by observing time points, since at all times, the schedules are statistically independent. However, due to the delay constraint, the secrecy condition leads to a reduced rate region, which is characterized in the following sections.

A. Receiver Directed Signaling

As mentioned earlier, in order to ensure complete relay secrecy for an $m \times 1$ multiplex relay, all the nodes involved have statistically independent transmission schedules. The schedules for the source nodes is determined by the source packet generation process, while the relay generates an independent Poisson point process. When the spreading sequences are receiver

directed, the constraints on transmission rates are independent for different receiving nodes. When characterizing the achievable rates for an $m \times 1$ relay, we assume that the destination nodes for different sources are distinct. Therefore, the constraint on the rates of the relay node are independent for each destination node.

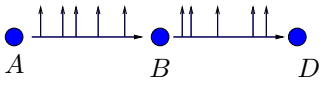

Fig. 4: 1×1 Relay

To characterize the achievable rates for a 1×1 relay map, we use the BOUNDED-GREEDY-MATCH (BGM) algorithm proposed in [13] that optimally maps Point processes with the least packet drops. Since epochs are generated according to independent Poisson processes, the strict delay constraint makes it impossible to relay all transmitted packets. The relay rate is, therefore, strictly less than the transmission rates of the nodes.

Let node A be the source node sending packets to destination D through relay B (see Fig. 4). The algorithm is as follows; When a packet arrives at B, if there exists a departure epoch within Δ of the arrival instant and has not been matched to any previous arrival, it is assigned to the arrived packet. Otherwise, the packet is dropped. The transmission schedule of A is obtained from the generation times of packets while node B generates an independent Poisson process of a fixed rate and uses the algorithm to map arrival epochs to the generated schedule.

Theorem 1: If the maximum transmission rates allowed to nodes B and D are C_B and C_D respectively, respectively, the maximum achievable relay rate R between (A, B), when $\mathcal{Y}_A, \mathcal{Y}_B$ are independent Poisson processes is obtained when $T_A = C_B, T_B = C_D$ and is given by

$$R = \begin{cases} C_B \frac{C_D\left(e^{-\Delta(C_B - C_D)} - 1\right)}{C_D e^{-\Delta(C_B - C_D)} - C_B} & C_B \neq C_D \\ \frac{C_B^2 \Delta}{1 + C_B \Delta} & C_B = C_D \end{cases}. \quad (2)$$

Proof: Refer to Appendix.

As is evident from the expression in Theorem 1, as $\Delta \to \infty$, the maximum relay rate is given by $\min\{C_B, C_D\}$. Similarly as $C_B \to \infty$, the maximum rate is C_D for any finite Δ and vice-versa. Clearly, when Δ is finite, the transmission rates T_A, T_B of the nodes are strictly greater than the achievable information relay rate, thereby resulting in packet drops. Packet losses can, however, be countered if the source employs forward error correcting (FEC) schemes.

Since the PHY layer is a receiver directed signaling scheme, for an $m \times 1$ relay, each outgoing stream from the relay would be an independent Poisson process which depends on the transmission rate allowed to the particular destination node. If there is no delay constraint, $\Delta = \infty$, then the achievable rate region is identical to the rate region without any secrecy constraint; the relay node can store incoming packets and regenerate them independently using a Poisson scheduler [6]. The rate region is then determined solely based on medium access constraints. The achievable relay rate region when nodes $A_1, \cdots, A_m$ send packets through relay B to destinations $D_1, \cdots, D_m$ respectively is given by

$$\sum_i R_i \leq C_B; R_i \leq C_{D_i}. \quad (3)$$

We assume the relay node decodes the packet headers and can distinguish packets arriving from multiple sources. Hence, the relay can treat the incoming packets from each source as a separate stream, although the streams have the same spreading code. For each source-destination pair, the relay node uses the BGM algorithm to map the packets in each arrival process to the corresponding outgoing stream. The following theorem characterizes the achievable rate region of this $m \times 1$ relay map.

Theorem 2: Let C_B be the maximum Tx. rate allowed to the relay B and C_{D_i} the maximum allowed rate to destination node D_i. An achievable rate region $\mathcal{R}^r$ for the $m \times 1$ relay is given as follows.

$(R_1, \cdots, R_m) \in \mathcal{R}^r$ if $\exists\ T_{A_1}, \cdots, T_{A_m}$ such that

$$R_i \leq T_{A_i} \frac{C_{D_i}\left(e^{-\Delta(T_{A_i} - C_{D_i})} - 1\right)}{C_{D_i} e^{-\Delta(T_{A_i} - C_{D_i})} - T_{A_i}}, \quad \sum_i T_{A_i} \leq C_B$$

Proof: Refer to Appendix.

T_{A_i} represents the transmission rate from source node A_i to the relay. Since the BGM algorithm has been proven to minimize the packet loss[14], this strategy provides the best achievable rates, when the transmission schedules are drawn from independent Poisson processes. It is easily seen from the theorem, that as $\Delta \to \infty$, all rates that satisfy the medium access constraints (3) are achievable by this technique. The achievable rate region for a 2×1 relay with receiver directed signaling is shown in Figure 5.

B. Transmitter Directed Signaling

When the signaling is transmitter directed, the constraint on the transmission rates are independent for each source node and the relay. Moreover, since the transmission rate constraint for the relay is independent of the number of destinations, the following results hold even if multiple source nodes share a common destination.

If there is no delay constraint, $\Delta = \infty$, then the achievable rate region is identical to the rate region without any secrecy constraint; the rate region is then determined solely based on medium access constraints, $i.e.$, ,

$$R_i \leq C_{A_i}; \sum_i R_i \leq C_B. \quad (4)$$

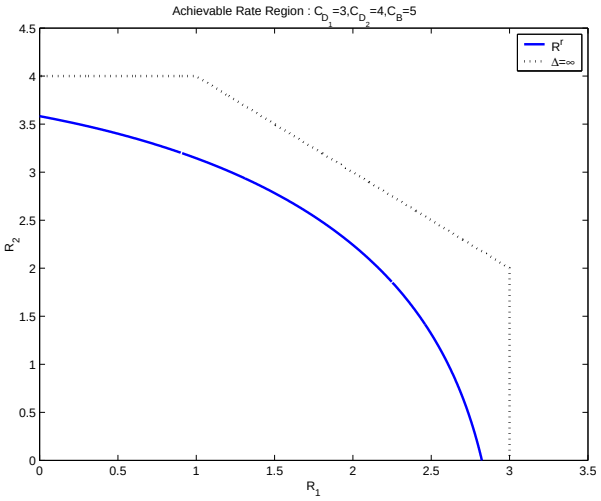

Fig. 5: Achievable Regions for 2×1 relay with Receiver directed signaling : $\Delta = 1$

An achievable rate region when Δ is finite can be obtained through a direct extension of the single source relay case considered in the previous section. The relay node ignores the origin of the packets and executes the BGM algorithm on the joint traffic from all the nodes. This strategy, which we refer to as *homogenous relay map* results in an achievable rate region $\mathcal{R}_H$ given by

Theorem 3: $(R_1, \cdots, R_m)$ belongs to $\mathcal{R}_H$ iff $\exists\ T_{A_i} \in [0, C_{A_i}],\ i = 1, \cdots, m$ s.t

$$R_i = T_{A_i} \frac{C_B(e^{-\Delta(\ \ _j T_{A_j} - C_B)} - 1)}{C_B e^{(-\Delta(\ \ _j T_{A_j} - C_B))} - \sum_j T_{A_j}}.$$

Proof : Since the relay ignores the source of the packets, it applies BGM algorithm on the joint arrival process of transmission rate $\sum_j T_{A_j}$. The proof follows from Theorem 1. □

It is easily shown that as Δ increases, the region $\mathcal{R}_H$ converges to the optimal rate region given by (3). Similarly, as discussed in the single relay case, as $C_B \to \infty$, it is possible to achieve all rate vectors satisfying the medium access constraints.

The region in Theorem 2 can be significantly improved if origin of packets are taken into consideration. The algorithm we propose is the following. The nodes transmitting to the relay are assigned unique indices from 1 to m such that the node with a higher index is given more priority when in contention. Every subset of nodes $S \subset 2^{\mathcal{A}}$ is assigned a priority value $\alpha(S) \in [0, 1]$. As long as there is no contention between packets from different sources for a particular departure epoch, the relay functions as a homogenous relay map. If packets from any subset of nodes S contend for the same departure epoch, the relay generates a Bernoulli random variable $Z \sim \mathcal{B}(\alpha(S))$. Let A_i be the node in S with the highest index. If $Z = 1$, then the packet from A_i is assigned that epoch. If $Z = 0$, then the packet that arrived earliest is assigned that epoch. By considering all possible index assignments and priority values, the rate region is obtained. The algorithm for 2 nodes is formally stated in Table assuming A_1 has index 2 and priority value α.

TABLE I: BOUNDED-PRIORITY-MATCH (α-BGM).

BOUNDED-PRIORITY-MATCH($\mathcal{Y}_{A_1}$, $\mathcal{Y}_{A_2}$, $\mathcal{Y}_B$, α, Δ):

$m_1 = m_2 = n = 1$;
while ($m_1 \le |\mathcal{Y}_{A_1}|$ or $m_2 \le |\mathcal{Y}_{A_1}|$) and $n \le |\mathcal{Y}_B|$
  if $Y_B(n) - Y_{A_1}(m_1) < 0$ and $Y_B(n) - Y_{A_2}(m_1) < 0$
    At $Y_B(n)$ Tx dummy packet; $n = n + 1$;
  else if $Y_B(n) - Y_{A_1}(m_1) \le \Delta$, $Y_B(n) - Y_{A_2}(m_2) > \Delta$
    $\mathcal{Y}^s_{A_1} = \mathcal{Y}^s_{A_1} \cup Y_{A_1}(m_1), \mathcal{Y}^s_B = \mathcal{Y}^s_B \cup Y_B(n)$;
    Drop $Y_{A_2}(m_2)$; Increment m_1, m_2, n by 1 ;
  else if $Y_B(n) - Y_{A_1}(m_1) > \Delta$, $Y_B(n) - Y_{A_2}(m_2) \le \Delta$
    $\mathcal{Y}^s_{A_2} = \mathcal{Y}^s_{A_2} \cup Y_{A_2}(m_2), \mathcal{Y}^s_B = \mathcal{Y}^s_B \cup Y_B(n)$;
    Drop $Y_{A_1}(m_2)$; Increment m_1, m_2, n by 1 ;
  else if $Y_B(n) - Y_{A_1}(m_1) \le \Delta$, $Y_B(n) - Y_{A_2}(m_2) \le \Delta$
    Generate random variable $Z \sim \mathcal{B}(\alpha)$
      If $Z = 0$, $i = \arg\min\{Y_{A_1}(m_1), Y_{A_2}(m_2)\}$
      else $i = 1$
    $\mathcal{Y}^s_{A_i} = \mathcal{Y}^s_{A_i} \cup Y_{A_i}(m_i), \mathcal{Y}^s_B = \mathcal{Y}^s_B \cup Y_B(n)$;
    $m_i = m_i + 1$, $n = n + 1$
  else
    Drop $Y_{A_1}(m_1), Y_{A_2}(m_2)$; Increment m_1, m_2 by 1 ;
    end
  end
end

Let $\mathcal{R}_P$ denote the set of all rate vectors achievable by using the priority relay map (all priority assignments). The following theorem provides bounds for $\mathcal{R}_P$.

Theorem 4:

$$\begin{aligned} \text{Let } \mathcal{R}_{out} &= \{(R_1, R_2) : \\ R_i &\le f(C_{A_i}, C_D), \\ \sum_i R_i &\le f(\sum C_{A_i}, C_D)\}, \end{aligned} \tag{5}$$

$$\text{where } f(a, b) = a \frac{b(e^{-\Delta(a-b)} - 1)}{be^{(-\Delta(a-b))} - a}. \tag{6}$$

Then, $\mathcal{R}_H \subseteq \mathcal{R}_P \subseteq \mathcal{R}_{out}$.

Proof: Refer to Appendix

The outer bound $\mathcal{R}_{out}$ described in the theorem is a tighter bound (than (3)) to the achievable rate region when we restrict the epochs to be independent Poisson transmissions. Figure 6 plots an example of the different regions for a 2×1 relay. As can be seen, the achievable rate region of the priority relay map $\mathcal{R}_P$ nearly coincides with the outer bound. As Δ increases, the regions $\mathcal{R}_H, \mathcal{R}_P$ and $\mathcal{R}_{out}$ converge to the optimal region given by (4).

C. Fixed Packet Loss

As mentioned in Section III-A, the finite delay constraint imposed on the transmission schedule results in packet loss. Hence, it is necessary for the source to use a forward error correction scheme to ensure reliable

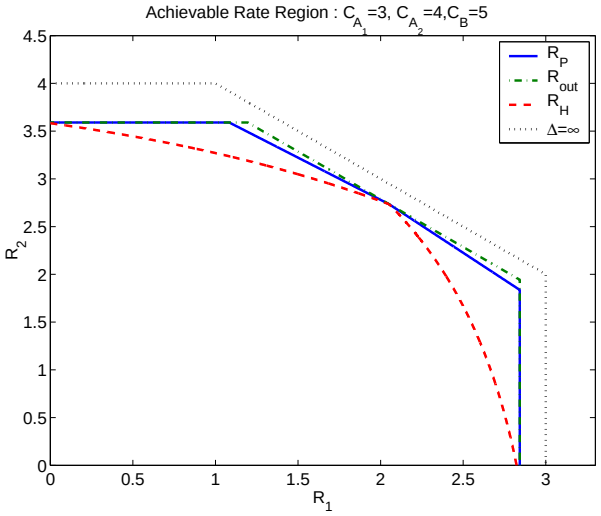

Fig. 6: Achievable Regions for 2×1 relay with Transmitter directed signaling : $\Delta = 1$

recovery of packets at the destination. In practice, it may be necessary to design strategies for a fixed packet drop fraction ϵ depending on the nature of data and availability of good codes. The following theorem characterizes an achievable rate region for the $m \times 1$ relay, such that the packet drop fraction is less than a fixed ϵ.

Theorem 5: 1. The achievable relay rate region $\mathcal{R}_\epsilon$ for the $m \times 1$ relay with packet loss constraint ϵ for *transmitter directed signaling* is given by $\mathcal{R}_\epsilon = \mathcal{R}_H \cap \mathcal{S}_\epsilon$, where

$$\mathcal{S}_\epsilon = \left\{ (R_1, \cdots, R_m) : \sum_i R_i \leq x(1-\epsilon) \right\},$$

and x is the solution of

$$\epsilon = \frac{C_B - x}{C_B \exp(-\Delta(x - C_B)) - x}. \tag{7}$$

2. The achievable relay rate region $\mathcal{R}_\epsilon^r$ for the $m \times 1$ relay with packet loss constraint ϵ for *receiver directed signaling* is given by $\mathcal{R}_\epsilon^r = \mathcal{R}^r \cap \mathcal{S}_\epsilon$, where

$$\mathcal{S}_\epsilon = \{(R_1, \cdots, R_m) : R_i \leq x_i(1-\epsilon),\ i = 1, \cdots, m\},$$

and x_i is the solution of

$$\epsilon = \frac{C_{D_i} - x_i}{C_{D_i} \exp(-\Delta(x_i - C_{D_i})) - x_i}. \tag{8}$$

Proof: Refer to Appendix

The rate region in Theorem 5 is obtained by using the homogenous relay map scheme described in SectionIII-B coupled with the constraint on sum-transmission rate due to the packet loss fraction ϵ.

IV. Conclusions

In this work, we formally defined the problem of hiding data flows from eavesdroppers observing transmission epochs. We proposed a possible solution for providing complete secrecy and characterized achievable rates for a multiplex relay in Poisson traffic. Allowing relays to perform re-encoding is a worthwhile extension to pursue. Although we have considered only a single relay system, the basic ideas are extendable to longer routes also.

References

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

[2] Q. Sun, D. R. Simon, Y. Wang, W. Russell, V. N. Padmanabhan, and L. Qiu, "Statistical identification of encrypted web browsing traffic," in *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, (Berkeley, California), p. 19, May 2002.

[3] C. Gulcu and G. Tsudik, "Mixing e-mail with babel," in *Proceedings of the Symposium on Network and Distributed System Security*, pp. 2–19, February 1996.

[4] 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.

[5] 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.

[6] X. Hong, P. Wang, J. Kong, Q. Zheng, and J. Liu, "Effective Probabilistic Approach Protecting Sensor Traffic," in *Military Communications Conference, 2005*, (Atlantic City, NJ), pp. 1–7, Oct. 2005.

[7] 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.

[8] J. Kong and X. Hong, "Anodr: Anonymous on demand routing with untraceable routes for mobile ad-hoc networks," in *ACM Intemational Symposium on Mobile Ad Hoc Nefworking and Computing*, (Annapolis, MD), June 2003.

[9] 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.

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

[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] 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.

[14] T. He and L. Tong, "Detecting Stepping-stone Traffic in Chaff: Fundamental Limits and Robust Algorithms," Tech. Rep. ACSP-TR-06-06-01, Cornell University, June 2006. `http://acsp.ece.cornell.edu/pubR.html`.

[15] 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 [14]. Consider the two point processes $\mathcal{Y}_A, \mathcal{Y}_B$. If a packet in $\mathcal{Y}_A$, say at time t is designated as dummy packet by the BGM algorithm, we insert a virtual packet at the $t + \Delta$ in $\mathcal{Y}_B$. Similarly, if a packet at time t in $\mathcal{Y}_B$ is designated as dummy packet, we insert a virtual packet at time t in $\mathcal{Y}_A$. Now we consider the difference process $\mathcal{Z} = \{Y_B(i) - Y_A(i)\}$ between

the two processes. At every occurrence of a dummy packet, the difference process hits a reflecting barrier, either at 0 or at Δ. The net probability of chaff is, therefore, the probability of hitting either barrier.

If the transmission rates of node A and B are T_A and T_B respectively, from the analysis in [15], we know that the probability of hitting Δ is given by

$$\Pr\{Z(i) = \Delta\} = \frac{1 - \frac{T_A}{T_B}}{\frac{T_B}{T_A}e^{-\Delta(T_A - T_B)} - \frac{T_A}{T_B}}.$$

It is easy to see that the fraction of chaff in $\mathcal{Y}_A$ is

$$\epsilon_A = \frac{T_B \Pr\{Z(i) = \Delta\}}{T_A(1 - \Pr\{Z(i) = \Delta\})} = \frac{T_B - T_A}{T_B e^{-\Delta(T_A - T_B)} - T_A}.$$

Since the rate of relayed packets increases with the transmission rates of either nodes, the achievability of the theorem is proved. In [13], the authors have shown that the BGM algorithm inserts the least chaff fraction for any pair of point processes. Hence, for any (T_A, T_B), it is impossible to obtain a higher information relay rate than (2).

□

Proof of Theorem 2

Since the nodes use receiver directed signaling, the relay node generates an independent outgoing Poisson process for each source. If the source transmission rate is T_{A_i} and the maximum allowed rate to destination D_i is C_{D_i}, we know from theorem 1 that

$$R_i = T_{A_i} \frac{C_{D_i}(e^{-\Delta(T_{A_i} - C_{D_i})} - 1)}{C_{D_i} e^{(-\Delta T_{A_i})} - T_{A_i}}. \tag{9}$$

Since the maximum allowed transmission rate to the relay is C_B, the sum-rate must satisfy

$$\sum_i T_{A_i} \leq C_B. \tag{10}$$

Combining (9) and (10), the theorem is proved.

□

Proof of Theorem 4

The inner bound is trivially shown as the homogenous map is a special case of the priority map when $\alpha(S) = 0, \forall S$. The outer bound is obtained using the optimality of BGM algorithm. Let node A_i transmit at rates T_i. Then, the sum information relay rate obtained by using the homogenous map is given by:

$$\sum_i R_i = f\left(\sum_i T_i, C_B\right). \tag{11}$$

Since BGM inserts the least fraction of dummy packets[13], this is the maximum sum-rate achievable for the given transmission rates. It is easy to see that $\sum_i R_i$ in (11) is an increasing function of $\sum_i T_i$. Therefore, the maximum sum-rate possible (when transmissions are independent Poisson processes) is given by

$$(\sum_i R_i)_{\max} = f\left(\sum_i C_{A_i}, C_B\right). \tag{12}$$

The best rate for A_i is obtained when $T_i = 0, j \neq i$ is zero. By replacing $\sum_j C_{A_j}$ by C_{A_i} in (12), we can obtain the remaining conditions that specify $\mathcal{R}_{out}$. □

Proof of Theorem 5

1. We consider the homogenous relay map. From Theorem 1, we know that for a set of transmission rates of sources$(T_{A_1}, \cdots, T_{A_m})$ the least fraction of chaff in the incoming stream is given by

$$\epsilon = \frac{C_B - (\sum_i T_{A_i})}{C_B \exp(-\Delta((\sum_i T_{A_i}) - C_B)) - \sum_i T_{A_i}},$$

when the relay transmits at the highest rate.

It is easily shown that ϵ is an increasing function of $\sum_i T_{A_i}$. Hence, an upper bound on ϵ corresponds to an upper bound on the sum transmission rate $\sum_i T_{A_i}$. Therefore, for any rate vector that satisfies $\sum_i T_{A_i} \leq x$ where x is given by 7, the homogenous relay map guarantees that relay rates satisfy the packet loss constraint. □

2. In the case of receiver directed signaling, the outgoing streams to destinations are independent. The packet loss for any stream is dependent only on the transmission rate of the source node and the rate of that particular outgoing stream. If the maximum allowed rate to the destination D_i is C_{D_i} and ϵ is the allowed packet loss, then from Theorem 1, we know that the maximum allowed source transmission rate x_i satisfies

$$\epsilon = \frac{C_{D_i} - x_i}{C_{D_i} \exp(-\Delta(x_i - C_{D_i})) - x_i}.$$

Combining the above equation with the achievable rate region $\mathcal{R}^r$, we get the result.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.