Relay Secrecy in Wireless Networks with Eavesdroppers

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 scheduling techniques to provide complete secrecy of routes, and characterize achievable rate regions for a multiplex relay under transmitter directed spread spectrum signaling. Further, we extend the results to the case when an additional constraint on packet loss is imposed.**

***Index Terms* - Network Security, Traffic Mix, Scheduling, Packet Coding.**

I. Introduction

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

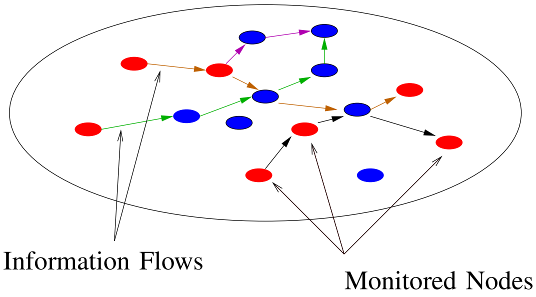

Fig. 1: Wireless Network with Eavesdroppers

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. We are interested in the design of secure schedules to prevent the inference of routes based on transmission epochs.

In general, the transmission schedule of relaying nodes are dependent on the arrival of packets, subject to the delay requirements. To elude eavesdroppers, 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, however, this may not be possible without affecting 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 optimize the achievable network performance while maintaining route secrecy.

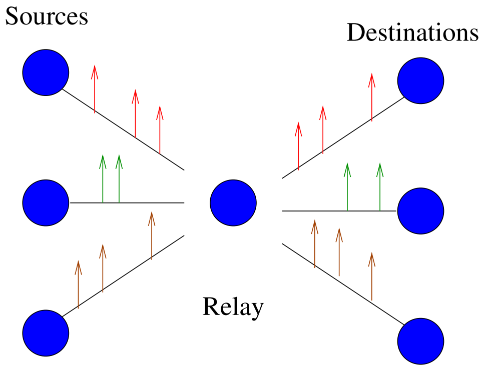

Fig. 2: Two Hop Network

In this work, we consider the problem of secrecy for a two-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 a transmitter directed physical layer signaling. We provide transmission schemes to prevent flow correlation and show that as the delay increases, the achievable rate region converges to the best possible region. Furthermore, we also present achievable rate regions when an additional constraint on packet loss is imposed.

his work is supported in part by the National Science Foundation under award CCF-0635070, the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011, and TRUST (The Team for Research in Ubiquitous Secure Technology) sponsored by the National Science Fundation under award CCF-0424422.

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 MIX concept is ideal for delay insensitive traffic and the has been used effectively in providing anonymous communication for Internet applications [2], [3], [4], [5]. In [6], the authors propose STOP-and-GO MIXes wherein the MIX, apart from re-encryption and packet padding, also incorporates a random delay to avoid timing correlation. For wireless networks, random delaying was used in [7] to prevent flow correlation.

For low latency networks, it was shown in [8] 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 [9], where the authors give bounds on the performance loss incurred due to the secrecy constraints. The use of randomized routes as a countermeasure to traffic analysis attacks has been considered in [10], [11].

In the techniques discussed above, the secrecy provided was not characterized analytically. The theoretical framework for secrecy in this work is motivated by the notion of equivocation developed by Shannon in [12]. The secrecy constraint we consider is a special case of Shannon's equivocation, known as maximum secrecy [13], wherein the observations provide zero information about the source.

The paper is organized as follows: In Section II we explain the analytical framework of the route secrecy problem. The basic results on the achievable rate regions are presented in Section III. Coding schemes and packet loss constraints are discussed in Section IV. Some concluding remarks and possible future extensions are given in V.

II. Problem Setup

A. Definitions

In this work, we propose techniques to hide the presence of a two-hop relay (Fig. 2) from an eavesdropper. 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 and it can re-order the packets before transmission. It is assumed that re-encryption and packet padding occur at every 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.

Let the network be represented by a directed 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. 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,

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

The relay function is defined 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.

The map for a node relaying multiple flows can be defined analogously. 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 arranging the epochs in ascending order. Then,

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

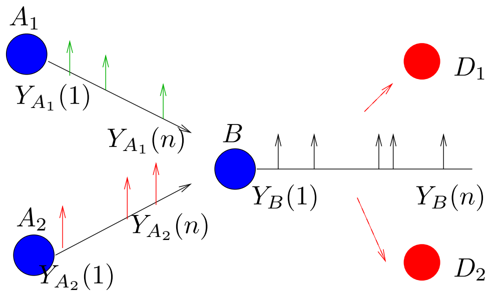

Fig. 3: $m \times 1$ Relay Map

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 tx. 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 a transmitter directed spread spectrum signaling model.

Transmitter Directed Signaling : 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

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

C. Secrecy

An eavesdropper, by correlating transmission epochs from multiple nodes, 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.

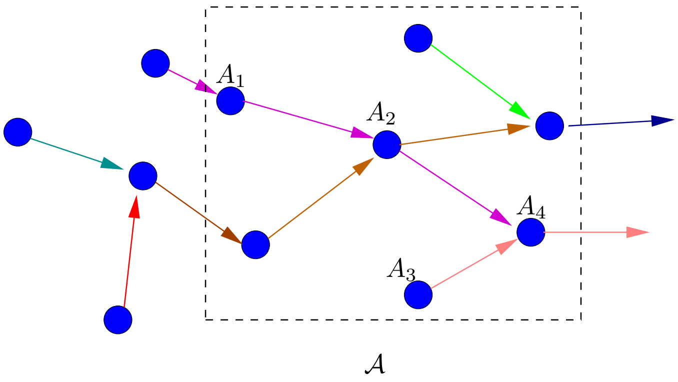

Fig. 4: Network Flows: Secrecy for a subset of nodes $\mathcal{A}$

Secrecy can be formally defined as follows. Let $\mathcal{A} = \{A_1, A_2, \cdots, A_k\}$ be a subset 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 $F \subset \mathcal{F}$. We define $\mathcal{A}$ to have *perfect relay secrecy* if for any $F \subset \mathcal{F}$, the transmission epochs of the nodes in $\mathcal{A}$ and F are independent. In other words, for every $F \in \mathcal{F}$ the conditional distribution

$$p(\mathcal{Y}_{A_1}, \mathcal{Y}_{A_2}, \cdots, \mathcal{Y}_{A_k}|F) = p(\mathcal{Y}_{A_1}, \mathcal{Y}_{A_2}, \cdots, \mathcal{Y}_{A_k}). \quad (2)$$

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. This condition is a special case of equivocation[12], [13], known as maximum secrecy, wherein

$$H(F|\mathcal{Y}) = H(F).$$

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 (1).
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 perfect relay secrecy.

In the following section, we present achievable rate regions for the special case of providing relay secrecy for an $m \times 1$ multiplex relay (Fig. 2), where a single node relays packets from m nodes. The results are presented for the PHY model discussed in Section II-B.

III. Rate Region

In the absence of eavesdropping concerns, 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 (2) 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,

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

Similar ideas have been considered in literature [9], [10], wherein the transmission schedules were deterministic irrespective of the flows. 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.

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. Single Relay Achievable Rate

To characterize the achievable rates for a 1×1 relay map, we use the Bounded-Greedy-Match (BGM) algorithm proposed in [14] that optimally maps Point processes with the least packet drops. Since epochs are generated according to independent Poisson processes, the delay constraint makes it impossible to relay all transmitted packets. Hence, the relay rate is strictly less than the transmission rates of the nodes.

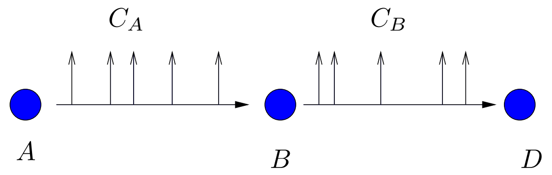

Fig. 5: 1×1 Relay

Let node A be the transmitting node and B the relay. 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 for nodes A, B are C_A, C_B respectively, the maximum achievable relay rate R between (A, B), when $\mathcal{Y}_A, \mathcal{Y}_B$ are independent Poisson processes is obtained when $T_A = C_A, T_B = C_B$ and is given by

$$R = \begin{cases} C_A \frac{C_B\left(e^{-\Delta(C_B - C_A)} - 1\right)}{C_B e^{-\Delta(C_B - C_A)} - C_A} & C_A \neq C_B \\ \frac{C_A^2 \Delta}{1 + C_A \Delta} & C_A = C_B \end{cases} \tag{3}$$

Proof: Refer to Appendix.

A special case of this result, when nodes have equal transmission rates was obtained in [15] under a different context. As the delay constraint Δ increases, it is easy to see that the relay rate converges to $\min\{C_A, C_B\}$ which is the optimal rate under no secrecy constraint. The minimization is due to the stability requirement in the network. Furthermore, the convergence of the relay rate with Δ can be shown to be exponential. The optimal rate can be obtained for a finite Δ when at least one of the transmission constraints C_A, C_B becomes infinite. Intuitively, this is easy to see; since one of the nodes has an infinite transmission rate, every transmitted packet can be matched perfectly.

Clearly, when Δ, C_A, C_B are 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. Therefore, the source needs to employ forward error correcting (FEC) schemes in order to deliver the information to the destination reliably. A detailed exposition of the packet loss and coding is discussed in section IV.

B. Multiplex Relay Region

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 and stability constraints,

$$R_i \leq C_{A_i}; \sum_i R_i \leq C_B. \tag{4}$$

A straightforward 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 2: $(R_1, \cdots, R_m)$ belongs to $\mathcal{R}_H$ iff $\exists\ T_{A_i} \in [0, C_{A_i}],\ i = 1, \cdots, m$ s.t

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

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 (4). Similarly, when Δ is finite and $C_B \to \infty$, it is also possible to achieve all rate vectors satisfying the medium access constraints. The homogenous map region (denoted by R_H) is shown in Figure 6.

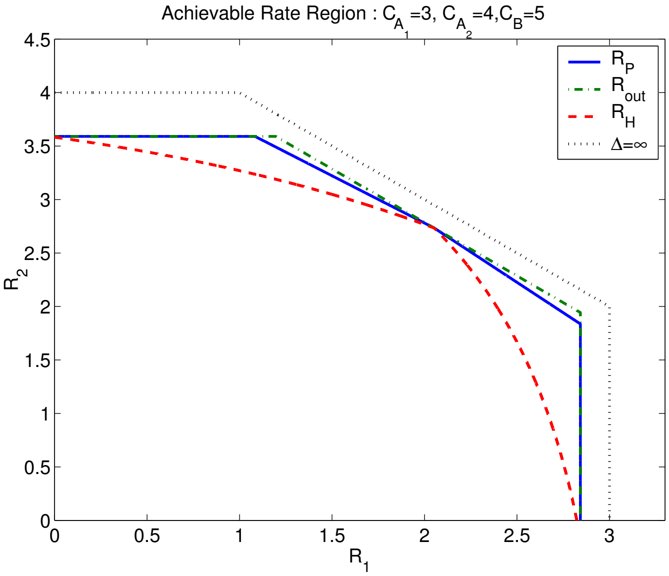

Fig. 6: Achievable Regions for 2×1 relay with Transmitter directed signaling : $\Delta = 1$

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 earlier 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 I (refer to Appendix) assuming A_1 has index 1 and priority value α.

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 3:

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

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

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

Proof: Refer to Appendix

The piecewise linear shape of the achievable rate region is due to the two basic components of the algorithm : priority encoding and time sharing. For example, in the two node case, the three vertices of the polygon represent the achievable rates when priority 1 is provided to either A_1 or A_2 and the maximum 0 priority sum-rate. The convexity of the achievable rate region is an outcome of the time-sharing strategy adopted in the algorithm. The parameter α represents the fraction of time the relay provides priority to node A_1 (or A_2).

Although the vertices of the piecewise linear region do not have a closed form analytical expression, the inner and outer bounds provided are asymptotically tight. The outer bound $\mathcal{R}_{out}$ described in the theorem is a sum-rate bound to the achievable rate region which can be obtained using the optimality of the BGM algorithm discussed in the previous section. The inner bound is obtained by using only priority 0, wherein the region reduces to that of the homogenous relay map.

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 (1).

IV. Packet Loss and Coding

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 recovery of packets at the destination. Coding for packet recovery has been addressed in literature[16], [17]. In particular, in [16], the authors propose coding schemes to recover packets when transmissions result in packet erasures. Since packets can be appended with a sequence number, the erasure positions are known to the receiver. For every block of information packets, parity packets are transmitted such that, for every i, the ith bit from every packet arranged in sequence forms a codeword from an erasure correcting codebook.

It can be shown that the erasures arising for an independent Poisson schedule using Greedy algorithm and Priority encoding are Markovian. Moreover, due to the memoryless nature of the Poisson process, the marginal probability of erasure for a source destination pair (S, D) is given by $R_{S,D}/T_S$ where $R_{S,D}$ is the achieved relay rate and T_S is the Tx. rate of the source. Hence, as the block length increases, it is possible to obtain an end-to-end information packet rate of $1-\epsilon$[18], where ϵ is the fraction of packets dropped.

For a fixed block length, the information packet rate reliably delivered would be strictly less than the capacity of the erasure channel. However, as the block length of packets considered increases, it is possible to design codes with rates arbitrarily close to capacity. 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 4: 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

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

Proof: Refer to Appendix

The packet loss for the greedy algorithm can be shown to be a monotonic function of the sum-rate of transmissions. Since the packet loss constraints are identical for the nodes, the rate region in Theorem 4 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. An example plot for the two node case is plotted in Fig. 7.

V. 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 perfect secrecy and characterized achievable rates for a multiplex relay in Poisson traffic. Achievable rate regions when the medium access constraints are based on receiver directed signaling is considered in [19].

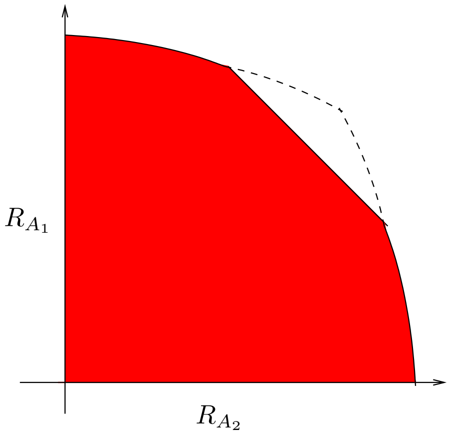

Fig. 7: Loss Limited Region for 2×1 relay with Transmitter directed signaling

Although we have considered only a single relay system, the basic ideas are extendable to longer routes also. As routes get longer, the packet loss fraction increases with every hop. Hence the perfect secrecy consideration may not be ideally suited. In such situations, the notion of equivocation lends to a partial secrecy metric, when $H(F|\mathcal{Y}) = \alpha H(F)$. Furthermore, allowing a node to perform block re-encoding is also an interesting direction to pursue.

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] 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.

[7] 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.

[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] 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.

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

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

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

[14] 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.

[15] T. He and L. Tong, "Detecting Traffic Flows in Chaff: Fundamental Limits and Robust Algorithms." submitted to IEEE Trans. on Information Theory, 2006.

[16] N. Shacham and P. McKenney, "Pakcet Recovery in High-Speed Networks using Coding and Buffer Management," in *Proc. IEEE INFOCOM*, pp. 124–131, 1990.

[17] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols," in *Proc. ACM SIGCOMM Computer Communication Review*, vol. 27, pp. 24–36, 1997.

[18] S. Boucheron and M. R. Salamatian, "About Priority Encoding Transmission," *IEEE Trans. Inform. Theory*, vol. IT-46, March 2000.

[19] P. Venkitasubramaniam, T. He, and L. Tong, "Networking with Secrecy Constraints." Accepted to 2006 IEEE Military Communications Conference, October 2006.

[20] 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.

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

APPENDIX

Bounded Priority Match Algorithm

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

```
BOUNDED-PRIORITY-MATCH($\mathcal{Y}_{A_1}$, $\mathcal{Y}_{A_2}$, $\mathcal{Y}_B$, $\alpha$, $\Delta$):

  $m_1 = m_2 = n = 1$;
  while ($m_1 \leq |\mathcal{Y}_{A_1}|$ or $m_2 \leq |\mathcal{Y}_{A_1}|$) and $n \leq |\mathcal{Y}_B|$
    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 $Y_B(n) - Y_{A_1}(m_1) \leq \Delta$, $Y_B(n) - Y_{A_2}(m_2) > \Delta$
      $\mathcal{Y}^s_{A_1} = \mathcal{Y}^s_{A_1} \cup Y_{A_1}(m_1), \mathcal{Y}^s_B = \mathcal{Y}^s_B \cup Y_B(n)$ ;
      Drop $Y_{A_2}(m_2)$ ; Increment $m_1, m_2, n$ by 1 ;

    else if $Y_B(n) - Y_{A_1}(m_1) > \Delta$, $Y_B(n) - Y_{A_2}(m_2) \leq \Delta$
      $\mathcal{Y}^s_{A_2} = \mathcal{Y}^s_{A_2} \cup Y_{A_2}(m_2), \mathcal{Y}^s_B = \mathcal{Y}^s_B \cup Y_B(n)$ ;
      Drop $Y_{A_1}(m_2)$ ; Increment $m_1, m_2, n$ by 1 ;

    else if $Y_B(n) - Y_{A_1}(m_1) \leq \Delta$, $Y_B(n) - Y_{A_2}(m_2) \leq \Delta$
      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
```

Proof of Theorem 1

To prove the theorem, we adopt the technique used in [20]. 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 [21], 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 [14], 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 (3).

□

Proof of Theorem 3

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:

$$\sum_i R_i = f\left(\sum_i T_i, C_B\right). \tag{8}$$

Since BGM inserts the least fraction of dummy packets[14], this is the maximum sum-rate achievable for the given transmission rates. It is easy to see that $\sum_i R_i$ in (8) is an increasing function of $\sum_i T_i$. Therefore, the maximum sum-rate possible (when transmissions are independent Poisson processes) is given by

$$(\sum_i R_i)_{\max} = f\left(\sum_i C_{A_i}, C_B\right). \tag{9}$$

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 (9), we can obtain the remaining conditions that specify $\mathcal{R}_{out}$. □

Proof of Theorem 4

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.

□

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.