Optimized Relay-Route Assignment for Anonymity in Wireless Networks

Chouchang Yang, Basel Alomair, and Radha Poovendran

***Abstract*—Anonymous wireless networks use covert relays to prevent unauthorized entities from determining communicating parties through traffic timing analysis. In a multipath anonymous network, the choice of which relay nodes should be covert, as well as the route selection by the network nodes, affect both the anonymity and network performance. Although assigning relays as covert and selecting routes composed of covert relays can provide higher anonymity, the selection of these two parameters will increase the packet dropping rate of the network. In this paper, we introduce an analytical framework for joint relay assignment and route selection in multi-path anonymous wireless networks. The main contributions of this work are two-fold. First, we show that joint relay assignment and route selection can be formulated as a convex optimization problem which guarantees global optimum solution. Second, as special cases of our formulation, we derive solutions for the problem of route selection to maximize anonymity when the relay configuration is given, as well as the problem of relay configuration for a given route selection.**

I. Introduction

Anonymity is an essential requirement for many communication systems to prevent unauthorized observers from determining the identities of the communicating parties. Practical applications in which anonymity is essential include, but are not limited to, electronic mail, evading website censorship, and file sharing. In a network with multiple source and destination nodes, anonymous networking protocols use cryptography to prevent eavesdroppers from identifying source-destination pairs by reading packet headers [1,2]. Transmitted packets of flows, however, carry additional information that cannot be disguised by means of cryptography, namely, their transmission time (or the time-stamp of transmitted packets) and routing path. If intermediate nodes forward packets in a "first come, first serve" manner, the timing of the received and forwarded packets can be correlated, leading to possible exposure of source-destination pairs [3].

In order to improve the anonymity of source-destination pairs, existing works have investigated the idea of designing relays to be secure against timing-based attacks. In [4], a mixing method, which collects and reorders packets from multiple senders before forwarding was introduced. Mixing relays that add fake traffic (dummy packets) [5] and drop packets [6] have also been proposed as mechanisms to further increase anonymity.

C. Yang and R. Poovendran are with the Network Security Lab, Department of Electrical Engineering, University of Washington, Seattle, WA, USA, 98195. Email: {ccjack, rp3}@uw.edu.

B. Alomair is with the Computer Research Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia. Email: alomair@uw.edu.

In [8], the authors introduce the notion of a covert relay, which transmits packets according to a pre-specified probabilistic schedule, sending fake traffic when no real packets are available. This pre-specified probabilistic schedule is shown to remove timing correlation at the relay nodes [7,8]. While covert relays prevent eavesdroppers from tracing forwarded packets back to the sender, the use of a pre-specified schedule reduces throughput, leading to increased latency and packet drops due to buffer overflows [7]. On the other hand, non-covert (visible) relays do not provide any anonymity.

There are two components that affect the anonymity of a wireless network, namely, the configuration of relays as covert or visible, and the routing paths chosen by nodes. While these two components play pivotal roles in determining the anonymity of the network, they also affect network performance. Hence, an analytical approach is needed to quantify and maximize the anonymity of a given network when performance constraints are enforced. Relay assignment based on the trade-off between anonymity and throughput in a single-path wireless network was studied in [8]. For a network with fixed covert and visible relays, a rate-distortion approach to allocate flows among multipath in order to maximize anonymity subject to packet-loss constraints was introduced in [9,10]. However, a unified approach for relay assignment and route selection that guarantees optimal anonymity, while maintaining network performance, is currently lacking.

In this paper, we propose an optimization framework for joint relay configuration and route selection in multipath anonymous wireless networks. Within this framework, the optimal relay configuration and route selection need to be chosen to maximize a given information-theoretic anonymity metric, subject to a packet-loss constraint. We prove that this optimization problem is convex, leading to efficient algorithms which guarantee global optimality.

As special cases, we consider route selection when the relay configuration is given, as well as relay configuration for a given route selection. For both cases, we derive the optimal route selection and relay configuration using our framework.

The paper is organized as follows. In Section II, we describe the adversarial model, relay and network models, as well as the anonymity metric to be used in this work. In Section III, we present our convex optimization framework for joint relay configuration and route selection, and provide derivation for the special cases described above. Section IV illustrates our simulation results. Section V concludes the paper.

II. Preliminaries and System Model

A. Adversary Model

We assume a passive global eavesdropper with knowledge of the network topology. The adversary is also assumed to be capable of observing and recording timing information of transmitted packets at all network nodes. All the packets and packet headers are encrypted to ensure that the adversary cannot read packets to determine their source and destination from the content or header. Instead, the adversary identifies source-destination pairs by using timing-based traffic analysis and knowledge of the network topology.

B. Relay Definitions

Visible and covert relays are defined as follows:

1) Visible Relays: A visible relay is a first in, first out intermediate node which forward packets without adding delay. In particular, if a routing path from a source to a destination consists of only visible relays, then an eavesdropper can trace the forwarded packets to the sender with probability one.

2) Covert Relays: A covert relay transmits packets according to a pre-specified probabilistic schedule. When there are no packets to be forwarded, the covert relay transmits a dummy packet; otherwise, the covert relay forwards the first packet in its buffer. Fake packets transmitted by covert relays are identified at the next hop using cryptographic mechanism based on shared pairwise keys and dropped. The design of the real and fake packet scheduling along with the encryption of the entire payload (packet and its headers) enable a covert relay to prevent eavesdroppers from associating the forwarded packets with incoming flows.

By assuming that packets arrive as a Poisson process with rate η_{in}, and are transmitted according to a pre-determined Poisson schedule with rate η_{out}, the packet-dropping in covert relays with buffer size K can be modeled as a M/M/1/K queue [7]. We assume that the buffer size $K > 1$. Hence, the packet-dropping rate for covert relays with various incoming rate is

$$P_e(\eta_{in}, \eta_{out}, K) = (\gamma^K - \gamma^{K+1})/(1 - \gamma^{K+1}), \quad (1)$$
where $\gamma = \eta_{in}/\eta_{out}$.

Buffer overflows at visible relays are assumed to be negligible.

C. Network Model

We consider a wireless network with fixed topology composed of N sources, α relays and M destinations where sources are $\{S_1, S_2,...,S_N\}$, destinations $\{D_1, D_2,...,D_M\}$ and relays $\{G_1, G_2,..., G_\alpha\}$. Each source S_i has transmission rate η_i. We define a random variable $\mathrm{Z_i} \in \{D_1, D_2, ..., D_M\}$ as the destination for source S_i, and we denote $\mathbf{Z} = [\mathrm{Z_1}, \mathrm{Z_2}, \cdots, \mathrm{Z_N}]$ as the vector of random variables representing the source-destination pairs in network. In addition, we define random variable $\mathrm{R_i}$ as the route used by source S_i. Then, we denote $\mathbf{R} = [\mathrm{R_1}, \mathrm{R_2}, \cdots, \mathrm{R_N}]$ as the vector of random variables representing the routes used by N sources. We define $\mathrm{C_i}$ as a binary random variable with $\mathrm{C_i} = 1$ and $\mathrm{C_i} = 0$, when G_i is covert and visible, respectively. Then, we denote $\mathbf{C} = [\mathrm{C_1}, \mathrm{C_2}, \cdots, \mathrm{C_\alpha}]$ as the vector of random variables representing the configuration of α relays. To model the fact that the relays in an anonymous network only forward packets to the next hop and are not aware of the source and destination of forwarded packets, we assume that the relay configuration $\mathbf{C}$ is independent of the source-destination pairs $\mathbf{Z}$.

We assume packet-loss is caused by two factors, link-quality at the physical layer and buffer overflow at the network layer. We denote the fraction of packets from source S_i that are lost at relay G_j due to link quality for route selection R_i as $L_{i,j}(\mathrm{R}_i)$. The dropping rate due to buffer overflows at covert relays is given by $P_e(\eta_{in}, \eta_{out}, K)$ in (1).

D. Anonymity Metric

We use $H(\mathbf{Z})$ to represent the uncertainty of all source-destination pairs before timing-based traffic analysis. Since eavesdroppers observe the timing information from the transmitted packets in each route, the packets in $\mathrm{R_i}$ from source i, for $1 \leq i \leq N$ with relays configuration $\mathbf{C}$ can reveal possible source-destination pairs. Thus, the uncertainty of the source-destination pairs deduced by eavesdropper with timing-based traffic analysis is defined as $H(\mathbf{Z}|\mathbf{R}, \mathbf{C})$. As in [12], we normalize by the uncertainty of all source-destination pairs before timing-based traffic analysis attacks to measure anonymity degree as

$$\text{Anonymity Degree} = \frac{H(\mathbf{Z}|\mathbf{R}, \mathbf{C})}{H(\mathbf{Z})} = \frac{H(\mathbf{Z}) - I(\mathbf{Z}; \mathbf{R}, \mathbf{C})}{H(\mathbf{Z})}.$$

Anonymity degree varies between 0 and 1. If all the relays are chosen as covert in $\mathbf{C}$, then the timing of packet transmission in routes $\mathbf{R}$ gives the eavesdropper no information about the associated source-destination pairs. Hence $H(\mathbf{Z}|\mathbf{R}, \mathbf{C}) = H(\mathbf{Z})$, resulting in the maximum anonymity of one. Since $H(\mathbf{Z})$ is constant, maximizing anonymity degree is equivalent to minimizing the mutual information $I(\mathbf{Z}; \mathbf{R}, \mathbf{C})$. The mutual information $I(\mathbf{Z}; \mathbf{R}, \mathbf{C})$ quantifies the loss in anonymity after passive attacks.

III. Problem Formulation

In this section, we present our convex optimization framework for joint relay configuration and route selection. We then derive solutions for two special cases of this framework, namely, route selection for given relay configuration and relay configuration for a fixed set of routes.

A. Joint Relay and Route Selection

In a wireless network, there are two inter-related design components that affect anonymity, namely, relay configuration and route selection. Moreover, these two components also affect the packet-loss rate, since each route has different link-quality and relay configuration (recall that covert relays drop packets, while visible relays do not).

We let $Q(\mathrm{C_1} = c_1, \mathrm{C_2} = c_2, ..., \mathrm{C_\alpha} = c_\alpha)$ denote the probability that the relay configuration is given by $(c_1, c_2, ..., c_\alpha)$. $Q(c_1, c_2, ..., c_\alpha)$ is interpreted as the fraction of time as covert and visible when relay configuration is given by $(c_1, c_2, ..., c_\alpha)$. $Q(\mathrm{R_1}{=}r_1, \mathrm{R_2}{=}r_2, ..., \mathrm{R_N}{=}r_N|\mathbf{C}, \mathbf{Z})$ denotes the conditional probability that the route selection is given by $(r_1, r_2, ..., r_N)$,

when the source-destination pairs are $\mathbf{Z} = (z_1, z_2, ..., z_N)$ and the relay configuration $\mathbf{C} = (c_1, c_2, ..., c_\alpha)$. We denote the average packet-loss rate as $E_{\mathbf{Z}}$, calculated by

$$E_{\mathbf{Z}} = \frac{\text{total packet-loss rate at } \alpha \text{ relays and } M \text{ destinations}}{\text{total transmission rate from N sources}}$$
$$= \frac{1}{\sum_{i'=1}^{N} \eta_{i'}} \Bigg[\sum_{i=1}^{\alpha} \Bigg\{ \sum_{\mathbf{R},\mathbf{C}:[\mathrm{C}_i]=1} \eta_{in}(i) * P_e(\eta_{in}(i), \eta_{out}, K) + \sum_{\mathbf{Z},\mathbf{C},\mathbf{R}} \sum_{j:\mathrm{R}_j \in G_i} \eta_j * P(\mathbf{Z}) Q(\mathbf{C}) Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) L_{j,i}(\mathrm{R}_j) \Bigg\} + \sum_{l=1}^{M} \sum_{\mathbf{Z},\mathbf{C},\mathbf{R}} \sum_{j':\mathrm{R}_{j'} \in G_i} \eta_{j'} * P(\mathbf{Z}) Q(\mathbf{C}) Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) L_{j',D_l}(\mathrm{R}_{j'}) \Bigg], \tag{2}$$

where $\eta_{in}(i) =$
$$\sum_{j:\mathrm{R}_j \in G_i} \sum_{\mathbf{Z}} \eta_j * P(\mathbf{Z}) Q(\mathbf{C}) Q(\mathbf{R}|\mathbf{Z},\mathbf{C})(1 - L_{j,i}(\mathrm{R}_j)).$$

The packet-loss rate in (2) represents summing the packet-loss in each node by considering the packet-loss from buffer overflow and link-quality. The first term represents the packet-loss from buffer overflows when relay G_i is covert, the second term represents the packet-loss due to link-quality at both covert and visible relays, and the third term represents the packet-loss due to link-quality at the destination.

Both anonymity and packet loss are functions of $Q(\mathbf{C})$ and $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$. Hence, the optimization problem of selecting the relay configuration distribution $Q(\mathbf{C})$ and the flow allocation $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ in order to minimize the anonymity loss while satisfying a packet-loss constraint is formulated as

$$\min_{Q(\mathbf{C}),Q(\mathbf{R}|\mathbf{Z},\mathbf{C}):E_{\mathbf{Z}} \le D} [H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R},\mathbf{C})], \tag{3}$$

The average packet-loss rate $E_{\mathbf{Z}}$ is a function of route selection $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ and relay configuration $Q(\mathbf{C})$. We denote by D the maximum average packet-loss of the network. However, $E_{\mathbf{Z}}$ is bilinear as a function of the optimization variables $Q(\mathbf{C})$ and $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$. Hence, a polynomial-time solution of (3) is not guaranteed.

Making use of the fact that $Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = Q(\mathbf{C}) \times Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$, we formulate an equivalent problem with $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$ as the optimization variable. Since the relay configuration $\mathbf{C}$ and the set of source-destination pairs $\mathbf{Z}$ are independent, we include the constraint $Q(\mathbf{C}|\mathbf{Z}) = Q(\mathbf{C})$. Then the equivalent formulation is given by

$$\min_{Q \in \mathbf{\Lambda}} \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z}) Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) \log \frac{Q(\mathbf{R},\mathbf{C}|\mathbf{Z})}{q(\mathbf{R},\mathbf{C})}, \tag{4}$$
$$\text{where } q(\mathbf{R},\mathbf{C}) = \sum_{\mathbf{Z}} P(\mathbf{Z}) Q(\mathbf{R},\mathbf{C}|\mathbf{Z}),$$
$$\mathbf{\Lambda} = \begin{cases} E_{\mathbf{Z}}(Q(\mathbf{R},\mathbf{C}|\mathbf{Z})) \le D, \ \forall \mathbf{Z} & \text{(4a)}, \\ \sum_{\mathbf{R},\mathbf{C}} Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = 1, \ Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) \ge 0, \ \forall \mathbf{Z} & \text{(4b)}, \\ \sum_{\mathbf{R}} Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = \sum_{\mathbf{R},\mathbf{Z}} P(\mathbf{Z}) Q(\mathbf{R},\mathbf{C}|\mathbf{Z}), \ \forall \mathbf{Z} & \text{(4c)}. \end{cases}$$

Lemma 1. *Let $Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z})$ denote the optimal solution of (4). The optimal solution to (3), denoted $Q^*(\mathbf{C})$, $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ can be obtained by*

$$Q^*(\mathbf{C}) = \sum_{\mathbf{R},\mathbf{Z}} Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z}), \ Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C}) = \frac{Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z})}{Q^*(\mathbf{C})}.$$

Proof: The proof is in two steps. First, we show that the values $Q^*(\mathbf{C})$, $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ are feasible under (3). We then prove that they are optimal under (3). To check feasibility of $Q^*(\mathbf{C})$, $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$, note that (4a) implies that $E_{\mathbf{Z}}(Q^*(\mathbf{C}), Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})) \le D$.

It remains to show optimality. Suppose there exist $\hat{Q}(\mathbf{C})$ and $\hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C})$ such that

$$\sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z}) \hat{Q}(\mathbf{C}) \hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C})) \log \frac{\hat{Q}(\mathbf{C}) \hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C}))}{q(\mathbf{R},\mathbf{C})} \tag{5}$$
$$< \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z}) Q^*(\mathbf{C}) Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})) \log \frac{Q^*(\mathbf{C}) Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C}))}{q(\mathbf{R},\mathbf{C})}.$$

Then, without loss of generality, since $\mathbf{C}$ is independent of $\mathbf{Z}$, define $\hat{Q}(\mathbf{C}|\mathbf{Z}) = \hat{Q}(\mathbf{C})$ for all $\mathbf{Z}$ and let $\hat{Q}(\mathbf{R},\mathbf{C}|\mathbf{Z}) = \hat{Q}(\mathbf{C}) \hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C})$. Then, $E_{\mathbf{Z}}(Q(\mathbf{R},\mathbf{C}|\mathbf{Z})) \le D$ and (4c) holds by assumption. Furthermore, by (5)

$$\sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z}) \hat{Q}(\mathbf{R},\mathbf{C}|\mathbf{Z})) \log \frac{\hat{Q}(\mathbf{R},\mathbf{C}|\mathbf{Z}))}{q(\mathbf{R},\mathbf{C})}$$
$$< \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z}) Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z})) \log \frac{Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z}))}{q(\mathbf{R},\mathbf{C})}.$$

This, however contradicts the assumption that Q^* is optimal, implying that (5) cannot hold and $Q^*(\mathbf{C}), Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is the optimal solution pairs to (3). □

The following theorem establishes convexity of problem (4).

Theorem 1. *The optimization problem presented in (4) is convex, when the buffer size of a covert relay is $K > 1$.*

Proof: We first show the packet-loss rate function $E_{\mathbf{Z}}$ in (4) is convex as a function of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$. We denote the summand of the first term in (2) as f_1^i

$$f_1^i(x) = x * P_e(x, \eta_{out}, K), \quad 1 \le i \le \alpha,$$
$$\text{where } x = \sum_{j:\mathrm{R}_j \in G_i} \sum_{\mathbf{Z}} \eta_j * P(\mathbf{Z}) Q(\mathbf{R},\mathbf{C}|\mathbf{Z})(1 - L_{j,i}(\mathrm{R}_j)),$$

and $P_e(x, \eta_{out}, K)$ satisfies $\frac{dP_e}{dx} \ge 0$ and $\frac{d^2P_e}{dx^2} \ge 0$, when the buffer size is $K > 1$. Hence given $K > 1$, $P_e(x, \eta_{out}, K)$ is a convex and increasing function of x which implies $f_1(x)$ is a convex function of x. Since x is a linear combination of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$ where $P_{\mathbf{Z}}$, η_j, η_{out}, and $L_{j,i}$ are constants, $f_1^i(x)$ is also a convex function of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$.

The second and third terms of (2) are linear functions of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$, since link-quality $L_{j,i}(\mathrm{R}_j)$, source j transmission rate η_j and $P(\mathbf{Z})$ are constants. Hence, both are convex functions of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$. Since all three terms of (2) are individually convex, their sum is also a convex function. $E_{\mathbf{Z}}$ is therefore a convex function of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$.

In addition, $\sum_{\mathbf{R},\mathbf{C}} Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = 1$ and $\sum_{\mathbf{R}} Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = \sum_{\mathbf{R},\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$ are affine equality constraints which preserve convexity. Finally, the mutual information function $I(\mathbf{Z};\mathbf{R},\mathbf{C})$ is a convex function in terms of variable $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$ given $P(\mathbf{Z})$. Therefore, the optimization in (4) is convex, when the buffer size $K > 1$. □

Lemma 1 and Theorem 1 imply that efficient algorithms can be used to obtain the globally optimal route selection $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ and relay configuration $Q(\mathbf{C})$.

In what follows, we analyze the optimal route selection and relay configuration for two special cases. In the first case, we consider route selection in a network with given, fixed relay configuration. In the second case, we consider relay configuration in a network with given, fixed source-destination pairs and routes.

B. Special Case 1: Route Selection for Given Fixed Relay Configuration

The case of a given fixed relay configuration $(c_1, c_2, ..., c_\alpha)$ can be represented by the distribution $Q(c_1, c_2, ..., c_\alpha) = 1$. We then have $Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = Q(\mathbf{R}|\mathbf{Z},\mathbf{C} = c_1, c_2, ..., c_\alpha)$. By solving Lagrange multiplier function from (4), the globally optimal solution $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ can be obtained. We consider the route selection under two different scenarios. In the first scenario, we assume that the buffer size of a covert relay is sufficiently large so that packet loss is due to link quality alone. In the second scenario, we incorporate packet dropping by covert relays with finite buffers.

1) Packet-loss from link-quality alone: Solving (4) via Lagrange multipliers yields

$$Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) = \frac{q(\mathbf{R}|\mathbf{C})\exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}{\sum_{\mathbf{R}} q(\mathbf{R}|\mathbf{C})\exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}, \quad (6)$$

$$\text{where } q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}),$$

$$L_q(\mathrm{R}_i) = \sum_{j:\{G_j\cap \mathrm{R}_i\}\cup\{\mathrm{D}_j\cap \mathrm{R}_i\}} L_{i,j}(\mathrm{R}_i) * (\eta_i / \sum_{i'=1}^{N} \eta_{i'}).$$

Here, $s < 0$ is a parameter that determines the trade-off between anonymity and packet-loss rate. Decreasing the value of s increases the value of the achieved anonymity, at the cost of higher packet-loss rate [11].

2) Packet-loss from link-quality and dropping by covert relays: If we assume a finite buffer size K for the covert relays, packets will be dropped when the buffer is full. Then, $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ can be obtained by solving the function

$$f(\mathbf{R}) : Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) - \frac{q(\mathbf{R}|\mathbf{C})\exp[sT_{\mathbf{Z}}]}{\sum_{\mathbf{R}} q(\mathbf{R}|\mathbf{C})\exp[sT_{\mathbf{Z}}]} = 0, \quad (7)$$

$$\text{where } q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}),$$

$$T_{\mathbf{Z}} = \frac{\partial E_{\mathbf{Z}}}{\partial Q(\mathbf{R}|\mathbf{Z},\mathbf{C})} * (1/P(\mathbf{Z})).$$

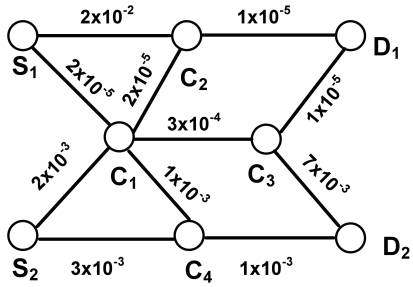


Fig. 1. Example of a wireless network topology with two sources and two destination with four intermediate nodes for joint route and relays selection.

Since equation (7) is non-linear in $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$, we use Newton's method to obtain $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ iteratively. Letting $\mathbf{Q}^{(n)}$ denote the value of $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ at the n iteration. $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is obtained by

$$\mathbf{Q}^{(n+1)} = \mathbf{Q}^{(n)} - \big[\,[\nabla \underline{f}]^{-1}\underline{f}\,]|_{\mathbf{Q}^{(n)}}, \quad (8)$$

where $\mathbf{Q}^{(n)} = [Q^{(n)}(\mathbf{a}_1|\mathbf{b}), Q^{(n)}(\mathbf{a}_2|\mathbf{b}), ..., Q^{(n)}(\mathbf{a}_l|\mathbf{b})]^T$ and $\underline{f} = [f(\mathbf{a}_1), f(\mathbf{a}_2), ..., f(\mathbf{a}_l)]^T$ are the column vector of (7) for all flows set $\mathbf{a}_i \subset \mathbf{R}$ given source-destination pair $\mathbf{b} \subset \mathbf{Z}$.

C. Special Case 2: Relay Configuration for Given Route Selection

Given the route for each source-destination pair, $Q(\mathbf{R},\mathbf{C}|\mathbf{Z}) = Q(\mathbf{C}) \times Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ where route selection $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is given. By solving the Lagrange multiplier function from (4), the global solution $Q(\mathbf{C})$ can be obtained by

$$Q(\mathbf{C}) = \frac{\prod_{\mathbf{R},\mathbf{Z}} \big[\frac{q(\mathbf{R},\mathbf{C})}{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}\big]^{P(\mathbf{Z},\mathbf{R}|\mathbf{C})}\exp[s\Phi(\mathbf{C})]}{\sum_{\mathbf{C}}\big(\prod_{\mathbf{R},\mathbf{Z}} \big[\frac{q(\mathbf{R},\mathbf{C})}{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}\big]^{P(\mathbf{Z},\mathbf{R}|\mathbf{C})}\exp[s\Phi(\mathbf{C})]\big)}, \quad (9)$$

$$\text{where } q(\mathbf{R},\mathbf{C}) = \sum_{\mathbf{C},\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C})Q(\mathbf{C}),$$

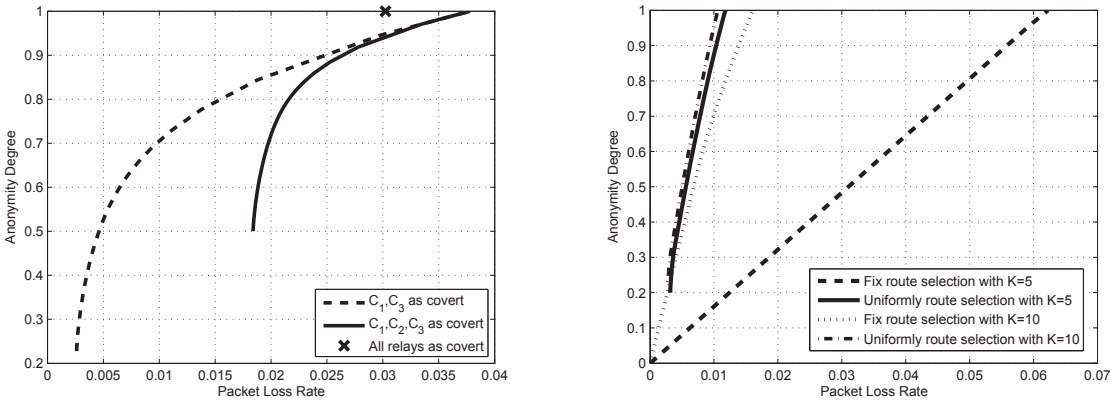
$$P(\mathbf{Z},\mathbf{R}|\mathbf{C}) = P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}), \Phi(\mathbf{C}) = \frac{\partial E_{\mathbf{Z}}}{\partial Q(\mathbf{C})}.$$

The $Q(\mathbf{C})$ obtained from (9) indicates the assignment of covert and visible relays.

IV. Simulation Results

The simulated network topology consists of two sources, two destinations, and four relays with buffer size $K = 5$ (Figure 1). Each source has transmission rate 5. Covert relays are assumed to have transmission rate 10. The source-destination pairs are chosen by $P(\mathrm{Z}_1 = D_i, \mathrm{Z}_2 = D_j) = 1/4$, for $i, j = 1, 2$ (i.e., all source-destination pairs are equally likely). The link quality of each link in Figure 1 was chosen uniformly at random from the interval $[10^{-5}, 10^{-2}]$.

Figure 2(a) shows the anonymity of the optimal route selection obtained from (7) for each given relay configuration and packet-loss constraint. We consider covert relay configurations (C_1, C_3), (C_1, C_2, C_3), (C_1, C_2, C_3, C_4), and show the



(a) (b)

Fig. 2. (a) Results for different relay configuration with buffer size K=5 for covert relays. Varying route selection among multipath result in different anonymity and packet-loss. (b) Results for given different route selection strategies with covert relay buffer size K=5 and K=10. Varying the relay configuration between covert and visible yields different anonymity for each packet-loss constraint.

resulting trade-offs between anonymity and packet-loss rate. When the required packet-loss rate is low, fewer covert relays should be used due to their increased packet dropping. On the other hand, when higher packet-loss is permissible, assigning relays to be covert increases the anonymity degree. We observe that adding the covert relay C_2 (solid line) actually reduces the achievable anonymity degree for each packet loss constraint. This illustrates the need for joint route and relay assignments.

Figure 2(b) shows the anonymity-packet loss trade-off for choosing relay assignments obtained from (9) with a given route selection and set of source-destination pairs. We consider two cases of route selection, which we denote as fixed and uniform. Under fixed route selection, source S_1 uses the route $\{S_1, C_2, D_1\}$ with probability 1, while S_2 uses the route $\{S_2, C_4, D_2\}$ with probability 1. Under uniform route selection, source S_1 (S_2) chooses each possible path to the destination D_1 (D_2) with equal probability. For covert relays with buffer size K=5, uniform route selection provides higher anonymity subject packet-loss constraints. This is because uniform route selection divides flow equally among the relays, leading to shorter average buffer occupancy and hence fewer packets drops.

V. Conclusion and Future Work

In this paper, we studied the problem of maximizing anonymity subject to packet-loss constraints through joint route selection and relay configuration in multipath anonymous networks. For an information theoretic anonymity metric, we formulated joint relay and route selection as a convex optimization problem. We showed that the problem of optimal route selection for given relay configuration, as well as optimal relay configuration for given fixed route selection can be derived and solved as special cases of our framework. Our framework guarantees efficient computation of the global optimum.

In the future work, we will study the performance of our approach in networks with additional relays and source-destination pairs, and compare our approach with independent optimization of relay configuration and route selection.

VI. Acknowledgements

We would like to thank Andrew Clark, Phillip Lee, and Ferney Maldonado, for helpful discussions on this paper.

References

[1] J. Kong, X. Hong, and M. Gerla,"An Identity-Free and On-Demand Routing Scheme against Anonymity Threats in Mobile Ad Hoc Networks". *IEEE Transactions on Mobile Computing*, Volume 6, Issue 8, pp. 888 - 902, Aug. 2007

[2] R. Dingledine, N. Mathewson, and P. Syverson. "Tor: The second generation onion router". in *Proc. the 13th USENIX Security Symposium*, August 2004.

[3] N. Hopper, E. Y. Vasserman, and E. Chan-Tin, "How much anonymity does network latency leak?" in *Proc. 14th ACM Conference on Computer and Communications Security*, ACM Press, 2007.

[4] D. Chaum, "Untraceable electronic mail, return address and digital pseudonyms," *Communication of the ACM*, vol. 24, pp. 84-88, Feb. 1981.

[5] P. Peng, P. Ning, D. Reeves, and X.Wang, Active timing-based correlation of perturbed traffic flows with chaff packets, in *Proc. 25th IEEE International Conference on Distributed Computing Systems Workshops*, Columbus, OH, Jun. 2005

[6] B. N. Levine, M. K. Reiter, C. Wang, and M. Wright. "Timing attacks in low-latency mix-based systems," in *Proc. of Financial Cryptography (FC 04)*. Springer-Verlag, LNCS 3110, February 2004.

[7] T. He and L. Tong, "Detection of information flows," *IEEE Transactions on Information Theory*, vol.54, no. 11, pp. 4925-4945, Nov. 2008.

[8] P. Venkitasubramaniam, T. He, and L. Tong, "Anonymous networking amidst eavesdroppers," *IEEE Transactions on Information Theory*, vol.54, no.6, pp. 2770-2784, Jun. 2008

[9] C. Yang, B. Alomair, and R. Poovendran, "Optimized Flow Allocation for Anonymous Communication in Multipath Wireless Networks," in *Proc. the IEEE International Symposium on Information Theory (ISIT)*, pp. 219-223, July 2012.

[10] C. Yang, B. Alomair, and R. Poovendran, "Multipath Flow Allocation in Anonymous Wireless Networks with Dependent Sources," in *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing*, October 2012.

[11] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol.18, no.4, pp. 460-473, July 1972.

[12] C. Diaz, S. Seys, J. Claessens, and B. Preneel. "Towards measuring anonymity".*Proceedings of Privacy Enhancing Technologies Workshop*. Springer-Verlag, LNCS 2482, April 2002.