# Optimized Flow Allocation for Anonymous Communication in Multipath Wireless Networks

Chouchang Yang*, Basel Alomair†, and Radha Poovendran*
*Network Security Lab, EE Dept, University of Washington, Seattle, WA, 98195, USA
†Computer Research Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

***Abstract*—In anonymous networks, a subset of nodes is chosen to act as covert relays to hide timing information from unauthorized observers. While such covert relays increase anonymity, they cause performance degradation by delaying or dropping packets. In this paper, we propose flow allocation methods that maximize anonymity for multipath wireless networks with predetermined covert relay nodes, while taking into account packet-loss as a constraint. Using a rate-distortion framework, we show how to assign probabilities which split the flows from source to destination among all possible routes and show that selecting routes according to the assigned probabilities achieves maximum anonymity given the packet-loss constraint.**

## I. Introduction

In applications requiring anonymous wireless communication, the identities of source-destination pairs for each data flow should be untraceable. Due to the open wireless medium, however, an eavesdropper can record timing information about transmitted packets. If intermediate relays forward messages as soon as they are received, the timing of the outgoing and incoming messages at each relay will be correlated. An eavesdropper can then use traffic analysis to infer source-destination pairs.

Designing reliable relays against timing-based traffic analysis is necessary for anonymous networks [1]-[3]. In [4], the authors showed that, by designing a series of network relays to embed the packet to be forwarded within a series of dummy packets (called "*chaff packets*"), the forwarding time of the received packet can be randomized in a way that prevents timing-based traffic analysis. Relays with randomized forwarding times are denoted *covert relays*, while other relays are denoted *visible relays* [5]. While eavesdroppers are able to trace packets through visible relays, they are unable to trace packets through covert relays. Furthermore, in order to minimize congestion and delays, covert relays may drop received packets [5]. Consequently, routing packets through covert relays preserves the anonymity of the source-destination pairs, but at the cost of reduced throughput and higher packet-loss.

In [5] and [8], the authors introduced the problem of selecting relays as covert and visible based on throughput and anonymity, under the assumption that each source-destination pair uses a single route. In typical wireless networks, however, each source has more than one possible route to the same destination.

Given a set of covert and visible relays, different routes will yield different packet-loss as well as different level of anonymity for the same source-destination pair. To date, the problem of selecting routes with maximum anonymity while minimizing packet-loss for a set of source-destination pairs is yet to be studied.

In this work, given a multipath wireless network with covert and visible relays, we investigate how to analytically choose routes for each source-destination pair in order to offer maximum anonymity while maintaining a packet-loss constraint. We consider two sources of packet-loss: link quality, determined by transmission power and the distance between nodes, and packets dropped by covert relays. We formulate the route selection problem within a rate-distortion framework, in which the fraction of flow allocated to each route is chosen to maximize the network anonymity without violating packet-loss constraints.

The rest of this paper is organized as follows. In Section II, we define the system and adversary models. In Section III, we formulate route selection as a rate-distortion problem and present solution algorithms. Section IV illustrates our results through simulation study. Section V concludes the paper.

## II. Preliminaries and System Model

### A. Adversary Model

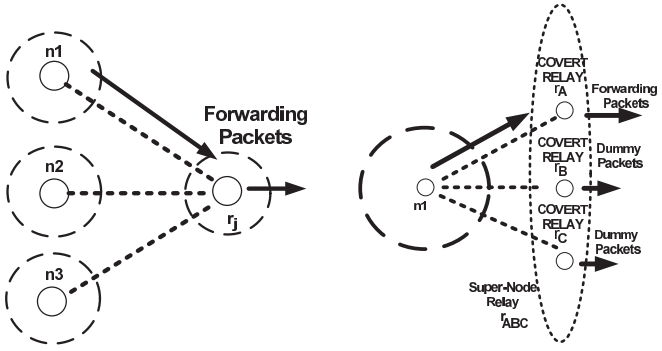
As in [5], we consider global eavesdroppers with knowledge of the network topology and ability to observe all network nodes. We assume that packets are encrypted so that eavesdroppers cannot expose source-destination pairs by reading packet headers. The goal of the eavesdroppers is to identify source-destination pairs by using timing-based traffic analysis.

### B. Relay Definitions and Notations

In this section, we give the definition of covert and visible relays introduced in [5] and specify the notations that will be used for the rest of the paper.

**Definition 1** (Visible Relays)**.** *A visible relay is a relay in which adversaries are able to trace the forwarded packets back to the senders. When a sender node $n_i$ selects visible relay $r_j$ to forward packets, we denote this event by $r_j^{n_i}$.*

As illustrated in Figure 1(a), when visible relay $r_j$ is forwarding packets from $n_1$, the adversary is able to deduce that these packets came from node $n_1$. Hence, we use the superscript $n_1$ in $r_j^{n_1}$ to emphasize the adversary's knowledge of the source node.




$(a)$ $(b)$

Fig. 1. The visible and covert relay assignments

**Definition 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. When sender node $n_i$ selects covert relay $r_j$ to forward a packet, we denote this event as $r_j$.*

By assumption, adversaries are unable to trace packets by observing covert relays. As shown in Figure 1(a), if relay $r_j$ is covert, from the eavesdropper's viewpoint, the current outgoing packet could be coming from node $n_1$, node $n_2$, or node $n_3$. Therefore, unlike the case of visible relays in Definition 1, the notation $r_j$ is used with no superscripts to emphasize that the adversary cannot associate a source node with the observed packets.

**Definition 3** (Super-node)**.** *If there are $k$ covert relays within one hop of a sender node $n_i$, the set of such $k$ nodes is called a super-node. The event that any one of the $k$ covert relays in the super-node is forwarding packets from $n_i$ is denoted by $r_{i_1 i_2 .. i_k}$*

In the example of Figure 1(b), when node $n_1$ sends a packet to covert relay $r_A$, eavesdroppers are unable to determine (with certainty) which one of those three outgoing packets is from node $n_1$. In other words, assuming node $n_1$ chose its next hop relay with equal probability, the three events that node $n_1$ is sending packets to covert relay $r_A$, covert relay $r_B$, or covert relay $r$ are indistinguishable. Thus, we use the notation $r_{ABC}$ to represent the event that one of the covert relays $r_A$, $r_B$, or $r_C$ is forwarding a packet from node $n_1$.

### C. Network System Model

We consider a network with $N$ sources and $M$ destinations. Each source $i$ chooses one of the $M$ destinations to send packets independently of other sources, with transmission rate $\eta_i$. We use $\mathrm{Z_i}$, which is a random variable taking a value from the destination set $\{D_1, D_2, ...D_M\}$, to represent the source-destination pair of source $i$. We define $P_{\mathrm{Z_i}}(D_j)$ to be the probability of source $i$ choosing destination $D_j$. Let

$$\mathbf{Z} = [\mathrm{Z_1}, \mathrm{Z_2}, \cdots, \mathrm{Z_N}] \tag{1}$$

where the $\mathrm{Z_i}$'s are independent of each other[1]. For source $i$ with $\tilde{L}^{th}$ hop transmission, let the series of intermediate relay nodes selected by $\mathrm{Z_i}$ be

$$\mathbf{R}^{\mathrm{Z_i}} = [R_1, R_2, \cdots, R_{\tilde{L}-1}], \tag{2}$$

where $R_l \in \{r_j^{n_i}, r_j, r_{i_1 i_2 .. i_k}\}$ represents the $l^{th}$ forwarding relay, which can be either visible, covert or super-node, and the destination node at the $\tilde{L}^{th}$ hop. Define

$$\mathbf{R} = [\mathbf{R}^{\mathrm{Z_1}}, \mathbf{R}^{\mathrm{Z_2}}, \cdots, \mathbf{R}^{\mathrm{Z_N}}] \tag{3}$$

to be the routes used by the $N$ sources in this network. We assume that each source chooses its route independently of other sources. Define packet-loss rate due to link-quality to be $P_l = L(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})$. Furthermore, to reduce latency, if packets stay in covert relays more than $\delta$ time units, they will be dropped [5]. For each route, define packet-loss due to packets dropped by covert relays to be $P_c = C(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}}, \delta)$. (Different dropping strategies cause different dropping rates; interested readers can refer to [8] for more discussion about dropping rate function.)

For source $i$ with transmission rate $\eta_i$, and packet-loss rate $P_e(i)$, let the maximum average packet-loss rate the total network can afford be $D = \sum_{i=1}^{N} \zeta_i P_e(i)$, where $\zeta_i = \eta_i / \sum_{i'=1}^{N} \eta_{i'}$ represents the fraction of total network throughput originated at source $i$. We use $H(\mathbf{Z})$ to represent the uncertainty of all source-destination pairs. Since eavesdroppers can observe the timing information $\mathbf{R}$, the uncertainty of the source-destination pairs to the eavesdropper after timing-based traffic analysis is defined as $H(\mathbf{Z}|\mathbf{R})$. As in [7], we express the network's anonymity degree by normalizing the uncertainty of all source-destination pairs as

$$\text{Anonymity Degree} = \frac{H(\mathbf{Z}|\mathbf{R})}{H(\mathbf{Z})}. \tag{4}$$

If all the relays in the network are covert, then the timing of packet transmissions gives the eavesdropper no information about the source-destination pairs. Hence $H(\mathbf{Z}|\mathbf{R}) = H(\mathbf{Z})$, resulting in the maximum anonymity of one.

## III. Problem Formulation and Proposed Algorithm

Consider the example of Figure 2 with $R_B$ as covert relay. The source-destination pairs, ($S_1$,$D_1$), ($S_2$,$D_1$) and ($S_3$,$D_3$), all have multiple routes. The problem considered in this paper is to find the probabilities with which a flow from a source must be allocated across different possible routes between that source and a specific destination in order to maximize anonymity while satisfying a packet-loss constraint.

Since adversaries can eavesdrop and observe the routes chosen by sources, denoted by $\mathbf{R}$, maximizing anonymity is equivalent to minimizing the source-destination pairs information available to eavesdroppers, given by $H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R})$ with the packet-loss rate as the penalty. Hence, the route selection

[1]We leave the case where the $\mathrm{Z_i}$'s are not independent for future work.

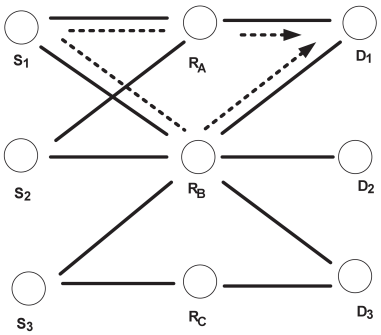



Fig. 2. Example of a wireless network topology with three sources, three relay nodes (which can be designated as visible or covert by the network), and three destinations. The two possible paths between source $S_1$ and destination $D_1$ are shown as dotted lines.

is equivalent optimizing the rate-distortion function subject to the distortion D, which is the maximum average packet-loss rate system can admit:

$$R(D) = \min_{d(\mathbf{Z},\mathbf{R})\leq D} [H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R})]. \quad (5)$$

Here, $d(\mathbf{Z},\mathbf{R})$ denotes distortion function defined as the system average packet-loss rate arising from the route vector $\mathbf{R}$ with source-destination pairs $\mathbf{Z}$.

**Lemma 1.** *Let* $R(D) = \min_{d(\boldsymbol{Z},\boldsymbol{R})\leq D}[H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R})]$. *Then*

$$R(D) = \sum_{i=1}^{N} \min_{d(\boldsymbol{Z},\boldsymbol{R})\leq D} [H(\mathrm{Z_i}) - H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})]. \quad (6)$$

*Proof*: Since each source chooses a route and destination independently of other sources, $H(\mathbf{Z}) = \sum_{i=1}^{N} H(\mathrm{Z_i})$, $H(\mathbf{Z}|\mathbf{R}) = \sum_{i=1}^{N} H(\mathrm{Z_i}|\mathbf{R})$, and $H(\mathrm{Z_i}|\mathbf{R}) = H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})$ for $1 \leq i \leq N$. This implies

$$\begin{aligned} R(D) &= \min_{d(\mathbf{Z},\mathbf{R})\leq D} \big[H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R})\big] \\ &= \sum_{i=1}^{N} \min_{d(\mathbf{Z},\mathbf{R})\leq D} \big[H(\mathrm{Z_i}) - H(\mathrm{Z_i}|\mathbf{R})\big] \\ &= \sum_{i=1}^{N} \min_{d(\mathbf{Z},\mathbf{R})\leq D} \big[H(\mathrm{Z_i}) - H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})\big]. \quad \square \end{aligned}$$

Lemma (1) shows that the problem of optimizing the network's overall anonymity is equivalent to optimizing the anonymity of each of the N independent sources individually. That is, finding the route for each source-destination pair that maximizes the anonymity of a given source-destination pair is independent of all other pairs in the network.

We define the distortion function for the formulation studied in this article by averaging the packet-loss rate from each route. Let $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}} = \underline{\alpha}|\mathrm{Z_i} = \beta)$ denote the probability that source $i$ selects route $\underline{\alpha}$ given the corresponding destination is $\beta$. Assuming the packet losses due to link-quality and packet-dropping by covert relays are independent, the packet-loss rate function can be expressed as

$$F(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}}, \delta, Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Zi}}|\mathrm{Z_i})) = 1 - (1 - P_c)(1 - P_l). \quad (7)$$

Then, we have [2]

$$\begin{aligned} R(D) &= \sum_{i=1}^{N} \min_{d(\mathbf{Z},\mathbf{R})\leq D} [H(\mathrm{Z_i}) - H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})], \\ &= \sum_{i=1}^{N} \min_{Q_{\mathrm{Z_i}}(j|k)\in\mathbf{Q}} [H(\mathbf{R}^{\mathrm{Z_i}}) - H(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})], \\ &= \sum_{i=1}^{N} \min_{Q_{\mathrm{Z_i}}(j|k)\in\mathbf{Q}} \Bigg[ \sum_j \sum_k \mathrm{P_{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k) \\ &\times \log \frac{Q_{\mathrm{Z_i}}(j|k)}{\sum_k \mathrm{P_{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k)} \Bigg], \end{aligned}$$

$$\text{where } \mathbf{Q} = \begin{cases} \sum_{i=1}^{N} \sum_j \sum_k P_{\mathrm{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k) \\ \times F(k,j,\delta,Q_{\mathrm{Z_i}}(j|k))\zeta_i \leq D,\ Q_{\mathrm{Z_i}}(j|k) \geq 0, \end{cases}$$

$$\text{and } 1 \leq i \leq N,\ j \in \mathbf{R}^{\mathrm{Z_i}},\ k \in \mathrm{Z_i}, \zeta_i = \eta_i \Big/ \sum_{i'=1}^{N} \eta_{i'}.$$

Using the condition $\sum_j Q_{\mathrm{Z_i}}(j|k) = 1$ and system average packet loss rate not exceeding $D$, the optimization function with Lagrange multiplier can be written as

$$\begin{aligned} J = \sum_{i=1}^{N} \Bigg[ & \sum_j \sum_k \mathrm{P_{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k) \log \frac{Q_{\mathrm{Z_i}}(j|k)}{\sum_k P_{\mathrm{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k)} \\ & + \sum_k v_{\mathrm{Z_i}}(k) \sum_j Q_{\mathrm{Z_i}}(j|k) \ + \\ & - s \sum_j \sum_k \mathrm{P_{Z_i}}(k) Q_{\mathrm{Z_i}}(j|k) F(k,j,\delta,Q_{\mathrm{Z_i}}(j|k))\zeta_i \Bigg]. \quad (8) \end{aligned}$$

We consider the optimization of equation (8) under two different scenarios. In the first scenario, covert relays do not drop packets. In the second scenario, we incorporate packet-dropping by covert relays. In what follows, we discuss how to obtain $Q_{\mathrm{Z_i}}$ for these two scenarios.

### *A. Route Selection Without Packet-Dropping by Covert Relays*

In this section, we consider the optimization problem of equation (8) assuming link-quality is the only source for packet loss. That is, we set the time constraint of the covert relays to $\delta = \infty$. In this case, the packet-loss rate function in equation (7) reduces to $F(k,j,\delta,Q_{\mathrm{Z_i}}(j|k)) = L(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})$. From [6,10], the roots of $\frac{\partial J}{\partial Q_{\mathrm{Z_i}}(j|k)} = 0$ are

$$Q_{\mathrm{Z_i}}(j|k) = \frac{q_{\mathrm{Z_i}}(j) \exp[sL(k,j)\zeta_i]}{\sum_j \ q_{\mathrm{Z_i}}(j) \exp[sL(k,j)\zeta_i]}, \quad (9)$$

where $q_{\mathrm{z_i}}(j) = \sum_k \mathrm{P_{Z_i}(k)} Q_{\mathrm{Z_i}}(j|k)$ and $L(k,j)$ represents the route's link-quality. As Equation (9) shows, $Q_{\mathrm{Z_i}}$ is computed by giving the input $s$. Each value of the parameter $s$ corresponds to a different choice of $D$, so that varying $s$ allows us to compute the anonymity-packet loss curve $R(D)$ [6,10].

[2] Due to covert relay super-node effect, for multiple routes map to the same $\mathrm{Z_i}$ and $\mathbf{R}^{\mathrm{Z_i}}$, we use the route with minimum packet-loss rate for the packet-loss rate function F. This is because those multiple routes have the same anonymity degree.

### B. Packet-Loss Rate from Link Quality and Packet-Dropping by Covert Relays

When we apply time constraint to covert relays, covert relays drop the packets staying in buffer more than $\delta$ time units [5]. This increase packet-loss rate. However, the dropping rate in here is also dependent on the flow allocation, which is denoted by $Q_{\mathrm{Z_i}}$ as described in [8]. Therefore, by solving $\frac{\partial J}{\partial Q_{\mathrm{Z_i}}(j|k)} = 0$, where $J$ is as given in equation (8), we obtain

$$f_j : Q_{\mathrm{Z_i}}(j|k) - \frac{q_{\mathrm{Z_i}}(j)\exp\left[sT_{\mathrm{Z_i}}\zeta_i\right]}{\sum_j\ q_{\mathrm{Z_i}}(j)\exp\left[sT_{\mathrm{Z_i}}\zeta_i\right]} = 0, \quad \forall j \in \mathbf{R}_{\mathrm{Z_i}}, \tag{10}$$

where

$$T_{\mathrm{Z_i}} = F(k, j, \delta, Q_{\mathrm{Z_i}}(j|k)) + Q_{\mathrm{Z_i}}(j|k)F'(k, j, \delta, Q_{\mathrm{Z_i}}(j|k)),$$
$$q_{\mathrm{z_i}}(j) = \sum_k \mathrm{P_{Z_i}(k)}Q_{\mathrm{Z_i}}(j|k) \text{ and } 1 \le i \le N,$$

and $F'$ denotes the first derivative of $F$ with respect to $Q_{\mathrm{Z_i}}(j|k)$. In equation (10), $\exp[sT_{\mathrm{Z_i}}\zeta_i]$ contains the term $Q_{\mathrm{Z_i}}$ that needs to be estimated. Hence the root cannot be evaluated directly from $Q_{\mathrm{Z_i}}(j|k) = \frac{q_{\mathrm{Z_i}}(j)\exp[sT_{\mathrm{Z_i}}\zeta_i]}{\sum_j\ q_{\mathrm{Z_i}}(j)exp[sT_{\mathrm{Z_i}}\zeta_i]}$, since it is a nonlinear polynomial. We present two numerical methods for obtaining $Q_{\mathrm{Z_i}}$ as follows.

*1) Newton-Raphson Method:* Under the Newton-Raphson method for solving nonlinear systems of equations [9], we have

$$\mathbf{Q}_{\mathrm{Z_i}}^{n+1} = \mathbf{Q}_{\mathrm{Z_i}}^{n} - \left[\,[\nabla\mathbf{F}]^{-1}\mathbf{F}\,\right]|_{\mathbf{Q}_{\mathrm{Z_i}}^{n}} \tag{11}$$

where $\mathbf{Q}_{\mathrm{Z_i}}^{n} = [Q_{\mathrm{Z_i}}^{n}(\underline{\alpha_1}|\beta), Q_{\mathrm{Z_i}}^{n}(\underline{\alpha_2}|\beta), ...Q_{\mathrm{Z_i}}^{n}(\underline{\alpha_l}|\beta)]^T$, and $\mathbf{F} = [f_1, f_2, ...f_l]^T$, for $f_j$ in equation (10). $Q_{\mathrm{Z_i}}^{n}(\underline{\alpha_j}|\beta)$ represents the value at $n^{th}$ iteration of each route $\underline{\alpha_j}$, given the destination $\beta$ via Newtons-Raphson method. After assigning arbitrary probability values for $\mathbf{Q}_{\mathrm{Z_i}}^{0}$ as initial vectors, all possible initial vectors converge to the same result by recursive equation (11).

*2) Taylor Series Expansion:* Another approximation to resolve equation (10) is using Taylor series expansion. The procedure is as follows. 1) In equation (10), set $\lambda = \sum_j\ q_{\mathrm{Z_i}}(j)\exp[sT_{\mathrm{Z_i}}\zeta_i]$ and, using an arbitrary probability as the initial value for $\acute{a}$, expanding equation (10) as:

$$q_{\mathrm{z_i}}(\underline{\alpha_j})\sum_{m=0}^{n}\left(\frac{exp(sT_{\mathrm{Z_i}}(\acute{a})\zeta_i)^{(m)}}{m!}\left[Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta) - \acute{a}\right]^m\right) - Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)\lambda = 0 \tag{12}$$

where $T_{\mathrm{Z_i}}(Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)) = F(\beta, \underline{\alpha_j}, \delta, Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)) + Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)F'(\beta, \underline{\alpha_j}, \delta, Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta))$

2) Solving the polynomial equation (12) for each $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)$ in terms of $\lambda$ by giving $n^{th}$ order approximation. Then all the roots can be described in terms of $\lambda$. 3) By using the property $\sum_j Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)) = 1$, we can solve for $\lambda$. Once we find $\lambda$, we can obtain $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)$. 4) Let new $\acute{a}$ equal to $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)$ obtained from 3), then repeat procedure 1), 2) and 3). Eventually, we can approach $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)$ using any initial value $\acute{a}$.

### C. Iteration Procedure

An iterative algorithm for obtaining $Q_{\mathrm{Z_i}}(j|k)$ is given as follows.

**Iteration Procedure** : Algorithm for route selection

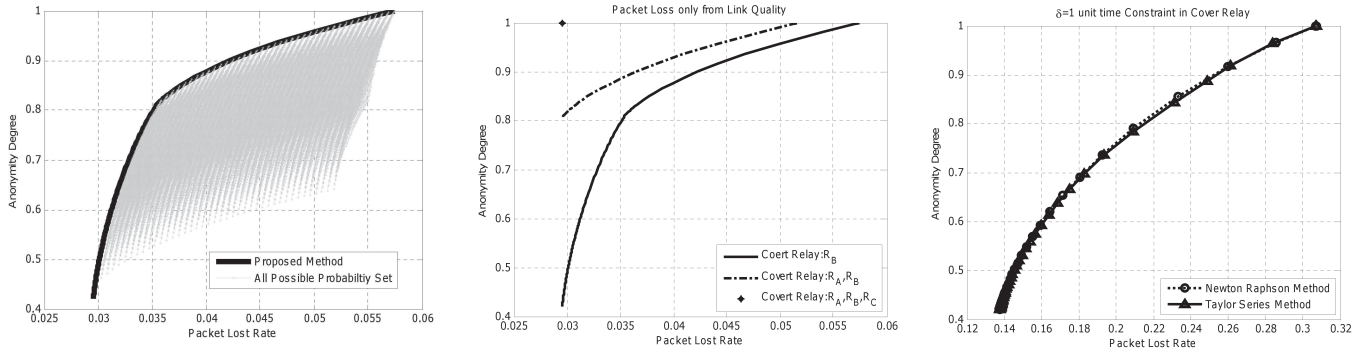
**Input**: Resolution, $\Delta$
Time constraint, $\delta$
Slope value, $s$
Iteration number, $I$
Source-destination pair, $\mathrm{Z_i} = \beta$
All routes for this source-destination pair $\beta$ : $\underline{\alpha_1}, \underline{\alpha_2}...\underline{\alpha_l}$
**Output:** Probability for each route $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta),\ j = 1, 2..l$
$q_{\mathrm{Z_i}}(\underline{\alpha_j}) \leftarrow \frac{1}{l}$, for $j = 1, 2..l$
**while** $n < I$
  **if** $\delta = \infty$
    $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta) \leftarrow$ eq. (9) for $j = 1, 2..l$
  **else**
    **while** $m_j > \Delta$ for $j = 1, 2...l$
      $m_j \leftarrow Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)$
      Compute $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta) \leftarrow$ by Newton-Raphson or Taylor Series method
      $m_j \leftarrow |m_j - Q_{\mathrm{Z_i}}|$ for $j = 1, ..l$
    **end**
  **end**
  $q_{\mathrm{Z_i}}(\underline{\alpha_j}) \leftarrow \sum_k \mathrm{P_{Z_i}(k)}Q_{\mathrm{Z_i}}(\underline{\alpha_j}|k)$ for $j = 1, 2..l$
**end while**
**return** $\{Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)\}_{j=1}^{l}$

Since any initial distribution value of $q_{\mathrm{Z_i}}(\underline{\alpha_j})$ will converge to the same result, we choose a uniform initial distribution. For a given slope $s$ and iteration number $I$, the above procedure returns the optimal flow allocation, represented by $Q_{\mathrm{Z_i}}(j|k)$. By varying the parameter $s$, a range of points on the anonymity-packet loss rate-distortion curve can be obtained.

## IV. Simulation Results

To verify that our proposed algorithm is optimal, we compare our results with all possible probability assignments. For simulation, we assume that each source has the same transmission rate and $\mathrm{Z_i}$ is uniformly distributed. For the network shown in Figure 2, with node $B$ acting as a covert relay without time constraint, i.e., $\delta = \infty$, we compare our proposed method with all possible probability distributions $\mathbf{Q}$ in Figure 3(a). The figure shows that our approach is optimal for this network topology.

Figure 3(b) shows the performance with different covert relay configurations, including ($R_B$), ($R_A$, $R_B$), and ($R_A$, $R_B$,$R_C$) from Figure 2. When all relays are covert relays, there is no trade-off between packet-loss rate and anonymity. Figure 3(b) also shows the maximum anonymity that can be achieved as a function of the packet-loss constraint $D$. For example, when only relay $R_B$ is covert, given $D$ is equal to $0.035$, the maximum anonymity is $0.788$. In order to achieve this maximum anonymity, the source-destination pair $(S_1, D_1)$ selects $R_A$ and $R_B$ to be 0.157 and 0.843, respectively.




$(a)$ $(b)$ $(c)$

Fig. 3. Results for the network in figure 2 using the packet loss given in Table I. (a) Comparison between manual assignment and the proposed rate-distortion method. We compare the proposed method performance with manual assign method which run all possible probability for each route. (b) The anonymity degree with packet-loss rate caused only by link-quality under different relay configurations. (c) The anonymity degree with packet-loss rate caused by link-quality and covert relay's dropping. The Newton-Raphson and Taylor approximation methods produce similar anonymity and packet-loss rate.

Figure 3(c) compares the Newton-Raphson and Taylor approximation algorithms for solving (10), with the time constraint $\delta$ set to 1, a maximum transmission rate of 3 for each node, and the dropping function described in [8] for $(R_B)$ as covert relay. We observe that both methods produce similar results, while the Taylor approximation does not require inverting a large matrix.

TABLE I
LINK-QUALITY TABLE FOR FIGURE 2 UNDER DIFFERENT (SOURCE, DESTINATION, RELAY) COMBINATIONS

| (S, D, R) | Link Quality | (S, D, R) | Link Quality |
|---|---|---|---|
| $(S_1, D_1, R_A)$ | $10^{-6}$ | $(S_2, D_1, R_A)$ | $10^{-4}$ |
| $(S_3, D_1, R_B)$ | $2 \times 10^{-2}$ | $(S_1, D_1, R_B)$ | $3 \times 10^{-2}$ |
| $(S_2, D_2, R_B)$ | $2.2 \times 10^{-2}$ | $(S_3, D_2, R_B)$ | $3.7 \times 10^{-2}$ |
| $(S_1, D_2, R_B)$ | $7 \times 10^{-2}$ | $(S_2, D_2, R_A)$ | $8 \times 10^{-3}$ |
| $(S_3, D_3, R_B)$ | $2 \times 10^{-1}$ | $(S_1, D_3, R_B)$ | $10^{-1}$ |
| $(S_2, D_3, R_B)$ | $3 \times 10^{-2}$ | $(S_3, D_3, R_C)$ | $10^{-3}$ |

## V. CONCLUSION

In this paper, we studied the problem of route selection in order to maximize anonymity in wireless networks subject to a constraint on packet loss. We formulated the problem within a rate-distortion framework, in which each node independently chooses a flow allocation among multiple routes. We introduced algorithms for optimal route selection under two cases, namely the case where covert relays drop packets in order to avoid congestion, and the case where covert relays do not drop packets. We evaluated the performance of our algorithms in both cases via simulation study. While the example for simulation used the network in Figure 2, the formulation is applicable to wireless networks with multiple intermediate hops with prespecified covert relays.

The main focus of this work was to find the optimal route selection for a given set of covert and visible relays with independent sources. In our future work, we will investigate flow allocation when the sources are dependent. We will also study the problem of joint optimization of both the route selection and the assignment of relays as covert or visible for a given network.

## REFERENCES


[1] C. Daz and A. Serjantov, "Generalizing mixes", in Proc. *Privacy Enhancing Technologies Workshop (PET)*. Berlin, Germany: Springer-Verlag, Apr. 2003.

[2] D. Kesdogan, J. Egner, and R. Buschkes, "Stop-and-go MIXes providing probabilistic anonymity in an open system", in Proc. *2nd Int.Workshop on Information Hiding (IH98)*, Portland, OR, Apr. 1998.

[3] D. Chaum, "Untraceable elextronic mail, return address and digital pseudonyms," *Commun. ACM* , vol. 24, pp. 84-88, Feb. 1981.

[4] T. He and L. Tong, "Detection of information flows," *IEEE Trans. Inform. Theory*, vol.54, no. 11, pp. 4925-4945, Nov. 2008.

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

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

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

[8] P. Venkitasubramaniam and L. Tong, "Throughput-anonymity tradeoff in wireless networks under latency constraints," *Proceedings of 2008 IEEE INFOCOM*, (Phoenix, AZ), pp. 241-245, April 2008.

[9] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, 1992.

[10] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Edition, Wiley, 2006.