Multipath Flow Allocation in Anonymous Wireless Networks with Dependent Sources

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*— Protecting the identities of source-destination pairs against eavesdroppers is essential in anonymous wireless networks. In this paper, we consider multipath wireless networks with a pre-defined subset of covert relays and study the problem of flow allocation between correlated source-destination pairs to maximize anonymity under given packet-loss constraints. By formulating the problem within a rate-distortion framework, we derive optimal flow allocations that maximize anonymity based on the side information available to each source. Our results are demonstrated through simulation study.**

I. Introduction and Related Work

In an anonymous wireless network, the identities of communicating source-destination pairs must remain hidden. Due to the open wireless medium, however, eavesdroppers can observe the timing of packet delivery. When relays directly forward received packets, there is a time correlation between incoming and outgoing packets. Hence, with knowledge of the network topology, an eavesdropper can use timing-based traffic analysis to identify network flows and determine possible source-destination pairs.

The problem of decreasing relays' timing correlation to prevent timing-based traffic analysis has attracted research attention for the past few decades. In [1], the concept of mixing, where relays reorder the forwarding times of packets received from different sources in order to thwart timing-based traffic analysis, was introduced. To further decrease timing correlation between received and forwarded packets, inserting dummy packets into outgoing traffic in relays was proposed in [3]. In [4], it was analytically proven that a series of relays can completely break the timing correlation in wireless networks if their forwarding time is independent of the receiving time, given that some received packets can be dropped and dummy packets can be transmitted at sufficient rates. Relays with independent forwarding schedule are denoted "covert relays," while other normal relays are denoted "visible relays" [6].

Although eavesdroppers cannot trace packets through covert relays, packets dropped by covert relays reduce the network throughput [6]. In [5] and [6], relays are assigned as covert or visible by considering the anonymity and throughput, when each source-destination pair uses a single route. However, in practical wireless networks, each source has multiple routes to the same destination. Since link-quality varies from one route to another, given relays as covert and visible, different routes for the same source-destination pair will have different packet-loss and anonymity. Hence, route selection based on anonymity alone may lead to unacceptably high packet-loss. Moreover, as opposed to the case of independent source-destination pairs studied in [9], when source-destination pairs are dependent, the routes chosen by each source will not only affect that sources anonymity, but will also affect the anonymity of other source-destination pairs. The route selection by each source should therefore also consider other sources' actions. To the best of our knowledge, the problem of route selection for dependent source-destination pairs to maximize anonymity while incorporating packet-loss as a constraint has not been studied to date.

As opposed to the works in [5] and [6], where the goal was to determine the set of nodes acting as covert for a given network with pre-selected source-destination routes, in this paper, given a multipath wireless network with predetermined covert and visible relays, we investigate route selection for each source-destination pair to maximize anonymity with packet-loss rate as a constraint, when source-destination pairs are dependent. Since each source may have incomplete knowledge of which destination and routes other sources choose due to packet encryption and radio range, we consider three different cases depending on each source's knowledge of other sources. In each case, we show how to split flows among multiple paths to maximize anonymity under packet-loss constraint by considering the optimization as a rate-distortion problem.

The rest of the paper is organized as follows. In Section II, we define the system and adversary models. In Section III, we propose three different flow allocation algorithms based on the information available to each source. Section IV illustrates our simulation results. Section V concludes the paper.

II. Preliminaries and System Model

A. Adversary Model

We assume passive attacks by global eavesdroppers who can observe all network nodes and have complete knowledge of the network topology [6]. Since packet headers are encrypted, eavesdroppers can only deduce the packet flows by timing-based traffic analysis. The goal of the adversary is

This research was partially supported by the ARO MURI Grant W911NF-07-1-0287, ARO Grant W991NF-12-1-0448 and also by KACST

to determine the source-destination pairs using the network topology and timing analysis.

B. Relay Definitions and Notations

In this section, we adopt the definitions of covert and visible relays as in [6], and specify the notations used in this paper.

Definition 1 (Visible Relays): Visible relays forward packets in the order they are received. When a sender node n_i selects visible relay r_j to forward packets, we denote this event by $r_j^{n_i}$.
The notation $r_j^{n_i}$ reflects the fact that packets forwarded via visible relay r_j can be traced back to the sender n_i.

Definition 2 (Covert Relays): Covert relays forward packets according to an independent pre-specified transmission schedule. When there are no packets to be forwarded at the next scheduled transmission time, the covert relay transmits a dummy packet. Otherwise, the covert relay forwards the first packet in its buffer. In addition, covert relays drop packets which occupy the buffer for more than δ time units to minimize latency and congestion. When sender node n_i selects covert relay r_j to forward a packet, we denote this event as r_j.
Unlike the case of visible relays, the notation r_j is used with no superscripts to emphasize that the adversary cannot associate a sender node with the observed forwarded packets.

Definition 3 (Super-node): A set of $k \geq 2$ covert relays within one hop of a sender node n_i is denoted as a super-node. The event that any one of the k covert relays in the super-node forwards packets from n_i is denoted by $r_{i_1 i_2 \cdots i_k}$
When the sender transmits packets to one of the covert relays comprising the super-node, the remaining covert relays will send dummy packets. An eavesdropper is therefore unable to recognize which covert relay forwards the packets at this hop. We use a super-node notation $r_{i_1 i_2 \cdots i_k}$ to represent the events that sender n_i sends packets to a covert relay r_{i_j} with $1 \leq j \leq k$.

C. Wireless Network Model

We assume a wireless network with dependent source-destination pairs consisting of N sources and M destinations. As opposed to independent source-destination networks, where each source chooses its destination independent of the other sources, we assume that the sources choose destinations according to a joint probability distribution. Therefore, each source's route selection can reveal not only its own destination, but also other source-destination pairs due to the dependency assumption. For example, in sensor networks, the sensors (sources) choose destinations based on observed events. When two or more sensors observe the same event, they will report the event to the same destination. Hence, if a source-destination pair is revealed via traffic analysis, an eavesdropper can also deduce other possible source-destination pairs.

We define $\mathrm{Z_i}$ to be a random variable representing source i taking one of the possible M destinations $\{D_1, D_2, ..., D_M\}$. Then, let $\mathbf{Z} = [\mathrm{Z_1}, \mathrm{Z_2}, \cdots, \mathrm{Z_N}]$ be the vector of random variables representing all possible source-destination pairs in the network. (For dependent source-destination pair networks, the $\mathrm{Z_i}$'s are dependent.)

We denote the route comprised by $\tilde{N}_i$ intermediate relay nodes corresponding to source-destination pair $\mathrm{Z_i}$ as $\mathbf{R}^{\mathrm{Z_i}} = [R_1, R_2, \cdots, R_{\tilde{N}_i}]$, where $R_l \in \{r_j^{n_i}, r_j, r_{i_1 i_2 \cdots i_k}\}$ represents the l^{th} forwarding relay, which can be either visible, covert or super-node. Let $\mathbf{R} = [\mathbf{R}^{\mathrm{Z_1}}, \mathbf{R}^{\mathrm{Z_2}}, \cdots, \mathbf{R}^{\mathrm{Z_N}}]$ represent the routes used by the N sources in the network. Throughout the paper, the transmission behavior of a source i is determined by its destination $\mathrm{Z_i}$ and the selected route $\mathbf{R}^{\mathrm{Z_i}}$.

We assume that packet-loss can be caused by two factors: link-quality and packet dropping by covert relays. Define $P_l = L(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})$ and $P_c = C(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}}, \delta)$ as the packet-loss rate caused by link-quality and covert relays' dropping in route $(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})$, respectively. (More discussion of the dropping rate function can be found in [5].)

For each source i with transmission rate η_i and packet-loss rate $P_e(i)$, we define the average packet-loss rate of the system as $\sum_{i=1}^{N} \zeta_i P_e(i)$, where $\zeta_i = \eta_i / \sum_{i'=1}^{N} \eta_{i'}$ represents the fraction of total network throughput originating at source i. Let $H(\mathbf{Z})$ represent the entropy of all source-destination pairs and $H(\mathbf{Z}|\mathbf{R})$ represent the entropy of the source-destination pairs after timing-based traffic analysis. As in [2], the normalized network anonymity under timing-based attacks is written as

$$\text{Anonymity Degree} = \frac{H(\mathbf{Z}|\mathbf{R})}{H(\mathbf{Z})} \tag{1}$$

When all relays are covert, $\mathbf{R}$ gives no information about $\mathbf{Z}$ [6] and, therefore, $H(\mathbf{Z}|\mathbf{R}) = H(\mathbf{Z})$, resulting in the maximum anonymity of one.

III. Problem Formulation and Proposed Flow Allocation Algorithm

Since we want the uncertainty $H(\mathbf{Z}|\mathbf{R})$ to approach $H(\mathbf{Z})$, the problem of maximizing the anonymity is equivalent to optimizing the following rate-distortion function

$$R(D) = \min_{d(\mathbf{Z},\mathbf{R}) \leq D} [H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R})], \tag{2}$$

where the distortion function $d(\mathbf{Z}, \mathbf{R})$ is the average packet-loss rate when using routes $\mathbf{R}$ with source-destination pairs $\mathbf{Z}$, and D is the maximum average packet-loss rate that the network can afford.

Since the $\mathrm{Z_i}$'s are dependent, the optimization problem (2) for each source $\mathrm{Z_i}$ must take into account other sources' transmission behaviors. However, due to radio range constraints and packet encryption, each source may lack information regarding the other sources' transmission behaviors. We therefore divide the problem into three cases. In Case I, we assume that each source, i, has complete information about the transmission behaviors of other sources and knows the values of $\mathrm{Z_j}$ and $\mathbf{R}^{\mathrm{Z_j}}$ for all $j \neq i$ when choosing its routes $\mathbf{R}^{\mathrm{Z_i}}$. In Case II, we assume that each source, i, knows the partial information about other sources' transmission

behaviors, where partial information refers to the probability distributions of $\mathrm{Z_j}$ and $\mathbf{R}^{\mathrm{Z_j}}$ for all $j \neq i$, but does not know the exact values of any of those parameters when choosing its routes $\mathbf{R}^{\mathrm{Z_i}}$. In Case III, we assume that each source has no information about the other sources' transmission behaviors. Note that the local knowledge of transmission behaviors in Case III is a subset of the knowledge of transmission behaviors in Case II, which in turn is a subset of complete knowledge of transmission behaviors in Case I.

A. Case I: Full information regarding other sources

Define $\mathrm{P}_{\mathbf{Z}}(\mathbf{Z})$ to be the joint probability distribution of all source-destination pairs and $Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})$ to be the joint probability distribution of the route selections $\mathbf{R}$ conditioned on the source-destination pairs $\mathbf{Z}$. Then, using chain rule, $Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})$ can be written as

$$Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) = Q_{\mathrm{Z_1}}(\mathbf{R}^{\mathrm{Z_1}}|\mathbf{Z}) \prod_{i=2}^{N} Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathbf{Z}, \mathbf{R}^{\mathrm{Z_1}}, \cdots, \mathbf{R}^{\mathrm{Z_{i-1}}}), \tag{3}$$

where $Q_{\mathrm{Z_i}}$ is the conditional probability of route selection for source i, which represents choosing route $\mathbf{R}^{\mathrm{Z_i}}$ depending on the sources' current destinations, $\mathbf{Z}$, and routes already chosen $\mathbf{R}^{\mathrm{Z_1}}, \ldots, \mathbf{R}^{\mathrm{Z_{i-1}}}$.[1] In addition, we define $F_{\mathrm{Z_i}}$ as the packet-loss rate of the route chosen by source i, calculated by $F_{\mathrm{Z_i}} = 1 - (1 - P_c)(1 - P_l)$, where P_c is the dropping rate of covert relay and P_l is the packet-loss rate caused by link-quality (note that $F_{\mathrm{Z_i}}$ is not a linear function since P_c is a function of $Q_{\mathbf{Z}}$ [6], as will be discussed in more detail later). Since packet-loss from link-quality and covert relay dropping occur in two different layers, namely the physical layer and the network layer, we assume they are independent. We have

$$R(D) = \min_{d(\mathbf{Z},\mathbf{R}) \leq D} [H(\mathbf{R}) - H(\mathbf{R}|\mathbf{Z})] \tag{4}$$
$$= \min_{d(\mathbf{Z},\mathbf{R}) \leq D} \left[\sum_{\mathbf{R}} \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) \log \frac{Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})}{\sum_{\mathbf{R}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})} \right].$$

The optimization problem (4) can be written with Lagrange multiplier as

$$J = \sum_{\mathbf{R}} \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) \log \frac{Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})}{\sum_{\mathbf{Z}} P_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})} \tag{5}$$
$$+ \sum_{\mathbf{Z}} v_{\mathbf{Z}}(\mathbf{Z}) \sum_{\mathbf{R}} Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})$$
$$- s \sum_{\mathbf{R}} \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) \left(\sum_{i=1}^{N} F_{\mathrm{Z_i}}(\mathbf{Z}, \mathbf{R}, \delta, Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})) \zeta_i \right).$$

In (5), the multiplier $v_{\mathbf{Z}}$ is from the condition $\sum_{\mathbf{R}} Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) = 1$ and s is from the condition $d(\mathbf{Z}, \mathbf{R}) \leq D$.

[1] Without loss of generality, we assume that the sources are indexed such that source i is the i^{th} source to select a route $\mathbf{R}^{\mathrm{Z_i}}$. Hence, the first source $\mathrm{Z_1}$ selects the routes $\mathbf{R}$ only based on the source-destination pairs $\mathbf{Z}$, since there are no existent routes already chosen by other sources.

We analyze the optimization problem (5) under two scenarios. In the first scenario, we assume that the covert relays never drop packets (i.e., $\delta = \infty$), so that packet loss only occurs due to link-quality. In the second scenario, we assume that packet loss is caused by both link-quality and packet dropping by covert relays. We consider these scenarios separately because in the second case, the packet-loss rate is a function of $Q_{\mathbf{Z}}$ which makes the problem tractable only via numerical methods. Once we obtain $Q_{\mathbf{Z}}$, as will be shown later in this section, each source's route selection $Q_{\mathrm{Z_i}}$ can be calculated by using Bayes' rule.

1) No packet dropping by covert relays: If packet loss occurs only due to link-quality, the average packet-loss rate $d(\mathbf{Z}, \mathbf{R})$ is independent of $Q_{\mathbf{Z}}$. We can obtain $Q_{\mathbf{Z}}$ by solving for the root of $\frac{\partial J}{\partial Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})} = 0$, which, as shown in [8,10], can be written as

$$Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z}) = \frac{q_{\mathbf{Z}}(\mathbf{R}) \exp\left[s\big(\sum_{i=1}^{N} L(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})\zeta_i\big)\right]}{\sum_{\mathbf{R}} q_{\mathbf{Z}}(\mathbf{R}) \exp\left[s\big(\sum_{i=1}^{N} L(\mathrm{Z_i}, \mathbf{R}^{\mathrm{Z_i}})\zeta_i\big)\right]}, \tag{6}$$

where $q_{\mathbf{Z}}(\mathbf{R}) = \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})$, $\mathrm{Z_i} \subset \mathbf{Z}$, and $\mathbf{R}^{\mathrm{Z_i}} \subset \mathbf{R}$.

2) Packet dropping by covert relays: If packet loss is caused by covert relays' dropping and link-quality, $Q_{\mathbf{Z}}$ is obtained by solving the equation $f_j = 0$, where f_j is defined as

$$f_j : Q_{\mathbf{Z}}(\mathbf{a}_j|\mathbf{b}) - \frac{q_{\mathbf{Z}}(\mathbf{a}_j) \exp\left[s(\sum_{i=1}^{N} T_{\mathrm{Z_i}}\zeta_i)\right]}{\sum_j q_{\mathbf{Z}}(\mathbf{a}_j) \exp\left[s(\sum_{i=1}^{N} T_{\mathrm{Z_i}}\zeta_i)\right]}, \tag{7}$$

where $\mathbf{a}_j \in \mathbf{R},\ \mathbf{b} \in \mathbf{Z},\ \ q_{\mathbf{Z}}(\mathbf{a}_j) = \sum_{\mathbf{b}} \mathrm{P}_{\mathbf{Z}}(\mathbf{b}) Q_{\mathbf{Z}}(\mathbf{a}_j|\mathbf{b})$,

$$T_{\mathrm{Z_i}} = F_{\mathrm{Z_i}}(\mathbf{b}, \mathbf{a}_j, \delta, Q_{\mathbf{Z}}(\mathbf{a}_j|\mathbf{b})) + Q_{\mathbf{Z}}(\mathbf{a}_j|\mathbf{b}) F'_{\mathrm{Z_i}}(\mathbf{b}, \mathbf{a}_j, \delta, Q_{\mathbf{Z}}(\mathbf{a}_j|\mathbf{b})).$$

Since, in equation (7), $T_{\mathrm{Z_i}}$ contains the root $Q_{\mathbf{Z}}$, we use Newton's method to solve (7) numerically by the iteration procedure

$$\mathbf{Q}_{\mathbf{Z}}^{n+1} = \mathbf{Q}_{\mathbf{Z}}^{n} - \big[\,[\nabla \mathbf{F}]^{-1} \mathbf{F}\,\big]|_{\mathbf{Q}_{\mathbf{Z}}^{n}}, \tag{8}$$

where $\mathbf{Q}_{\mathbf{Z}}^{n} = [Q_{\mathbf{Z}}^{n}(\mathbf{a}_1|\mathbf{b}), Q_{\mathbf{Z}}^{n}(\mathbf{a}_2|\mathbf{b}), ..., Q_{\mathbf{Z}}^{n}(\mathbf{a}_l|\mathbf{b})]^T$ are the values at the n^{th} iteration, and $\mathbf{F} = [f_1, f_2, ..., f_l]^T$, for f_j in (7).

B. Case II: Partial information regarding other sources

In this case, we assume that each source i knows the probability distribution of the other sources' transmission behaviors. For source i, the maximum anonymity under partial information is achieved when $H(\mathrm{Z_i}|\mathbf{R})$ is as close to $H(\mathrm{Z_i})$ as possible. The route selection probability $Q_{\mathrm{Z_i}}$ is

obtained by

$$\begin{aligned}&\min_{d(\mathbf{Z},\mathbf{R})\le D}\left[H(\mathrm{Z_i})-H(\mathrm{Z_i}|\mathbf{R})\right] \\ &=\min_{d(\mathbf{Z},\mathbf{R})\le D}\left[H(\mathbf{R})-H(\mathbf{R}|\mathrm{Z_i})\right] \\ &=\min_{d(\mathbf{Z},\mathbf{R})\le D}\Bigg[\sum_{\mathbf{R}}\sum_{\mathrm{Z_i}}\mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i})\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})\times \\ &\quad\log\frac{\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})}{\sum_{\mathrm{Z_i}}\mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i})\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})}\Bigg],\end{aligned} \tag{9}$$

where $\mathbf{R}^{/\mathrm{Z_i}}=[\mathbf{R}^{\mathrm{Z}_1},\ldots,\mathbf{R}^{\mathrm{Z}_{i-1}},\mathbf{R}^{\mathrm{Z}_{i+1}},\ldots,\mathbf{R}^{\mathrm{Z}_N}]$ is the set of vectors representing all routes except $\mathbf{R}^{\mathrm{Z_i}}$ and $\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})$ is the route selection of other sources conditioned on $\mathrm{Z_i}$.

Note that the solution to (9) results in the maximum-uncertainty flow allocation for source i individually, but does not achieve the global optimum of (2). $Q_{\mathrm{Z_i}}$ can be obtained by solving (9) using Lagrange multipliers with the conditions $d(\mathbf{Z},\mathbf{R})\le D$ and $\sum_{\mathbf{R}^{\mathrm{Z_i}}}Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})=1$. As in Case I, we solve for $Q_{\mathrm{Z_i}}$ by considering two scenarios, based on whether the covert relays drop packets.

1) No packet dropping by covert relays: If packets loss occurs only due to link-quality, then similar to Case I, we obtain $Q_{\mathrm{Z_i}}$ as [8,10]:

$$Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})=\frac{\prod_{\mathbf{R}^{/\mathrm{Z_i}}}q_{\mathbf{Z}}(\mathbf{R})^{\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})}\exp[sL(\mathrm{Z_i},\mathbf{R}^{\mathrm{Z_i}})\zeta_i]}{\sum_{\mathbf{R}^{\mathrm{Z_i}}}\left(\prod_{\mathbf{R}^{/\mathrm{Z_i}}}q_{\mathbf{Z}}(\mathbf{R})^{\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i})}\exp[sL(\mathrm{Z_i},\mathbf{R}^{\mathrm{Z_i}})\zeta_i]\right)}, \tag{10}$$

where $q_{\mathbf{Z}}(\mathbf{R})=\sum_{\mathbf{Z}}\mathrm{P}_{\mathbf{Z}}(\mathbf{Z})[\prod_{i=1}^{N}Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})]$.

2) Packet dropping by covert relays: When packet loss is caused by both link-quality and covert relays' dropping, then $Q_{\mathrm{Z_i}}$ can be obtained using the Newton's method in (8) numerically with the following function f_j, defined as

$$f_j: Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)-\frac{\prod_{\mathbf{R}^{/\mathrm{Z_i}}}q_{\mathbf{Z}}(\mathbf{R})^{\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|b)}\exp[sT_{\mathrm{Z_i}}\zeta_i]}{\sum_{j}\left(\prod_{\mathbf{R}^{/\mathrm{Z_i}}}q_{\mathbf{Z}}(\mathbf{R})^{\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|b)}\exp[sT_{\mathrm{Z_i}}\zeta_i]\right)}, \tag{11}$$

where $\mathbf{a}_j\in\mathbf{R}^{\mathrm{Z_i}}, b\in\mathrm{Z_i}$, $q_{\mathbf{Z}}(\mathbf{R})$ is as in(10), and

$$T_{\mathrm{Z_i}}=F(b,\mathbf{a}_j,\delta,Q_{\mathrm{Z_i}}(\mathbf{a}_j|b))+ Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)F'(b,\mathbf{a}_j,\delta,Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)).$$

C. Case III: No information regarding other sources

In this case, we assume that each source has no information regarding the other sources' transmission behaviors. Since source i cannot compute the global objective function of (2), source i instead selects the routes $\mathbf{R}^{\mathrm{Z_i}}$ such that $H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})$ is as close to $H(\mathrm{Z_i})$ as possible. The route selection probability $Q_{\mathrm{Z_i}}$ can be obtained by:

$$\begin{aligned}&\min_{d(\mathbf{Z},\mathbf{R})\le D}\left[H(\mathrm{Z_i})-H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})\right] \\ &=\min_{d(\mathbf{Z},\mathbf{R})\le D}\left[H(\mathbf{R}^{\mathrm{Z_i}})-H(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})\right] \\ &=\min_{d(\mathbf{Z},\mathbf{R})\le D}\Bigg[\sum_{\mathbf{R}^{\mathrm{Z_i}}}\sum_{\mathrm{Z_i}}\mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i}) \\ &\qquad\times\log\frac{Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})}{\sum_{\mathrm{Z_i}}\mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})}\Bigg].\end{aligned} \tag{12}$$

Equation (12) results in the maximum uncertainty $H(\mathrm{Z_i}|\mathbf{R}^{\mathrm{Z_i}})$, conditioned on the route selection $\mathbf{R}^{\mathrm{Z_i}}$ for source i, but does not achieve the global optimum in (2).

$Q_{\mathrm{Z_i}}$ can be obtained from (12) using Lagrange multipliers with the conditions $d(\mathbf{Z},\mathbf{R})\le D$ and $\sum_{\mathbf{R}^{\mathrm{Z_i}}}Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})=1$. As in Case I and Case II, we consider the two different scenarios of covert relay behavior below.

1) No packet dropping by covert relays: When packet loss occurs due to link-quality alone, we have, similar to Cases I and II [8,10],

$$Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})=\frac{q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}})\exp[sL(\mathrm{Z_i},\mathbf{R}^{\mathrm{Z_i}})\zeta_i]}{\sum_{\mathbf{R}^{\mathrm{Z_i}}}q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}})\exp[sL(\mathrm{Z_i},\mathbf{R}^{\mathrm{Z_i}})\zeta_i]}, \tag{13}$$

where $q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}})=\sum_{\mathrm{Z_i}}\mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i})Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})$.

2) Packet dropping by covert relays: If packet loss is caused by both link-quality and packet dropping from covert relays, $Q_{\mathrm{Z_i}}$ is obtained using the Newton's method (8) to solve $f_j=0$, defined by

$$f_j: Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)-\frac{q_{\mathrm{Z_i}}(\mathbf{a}_j)\exp[sT_{\mathrm{Z_i}}\zeta_i]}{\sum_j q_{\mathrm{Z_i}}(\mathbf{a}_j)\exp[sT_{\mathrm{Z_i}}\zeta_i]}, \tag{14}$$

where $\mathbf{a}_j\in\mathbf{R}^{\mathrm{Z_i}}, b\in\mathrm{Z_i}$, $q_{\mathrm{Z_i}}(\mathbf{a}_j)=\sum_b\mathrm{P}_{\mathrm{Z_i}}(\mathrm{b})Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)$,

$$T_{\mathrm{Z_i}}=F(b,\mathbf{a}_j,\delta,Q_{\mathrm{Z_i}}(\mathbf{a}_j|b))+ Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)F'(b,\mathbf{a}_j,\delta,Q_{\mathrm{Z_i}}(\mathbf{a}_j|b)).$$

D. Algorithm for Flow Allocation

An algorithm to compute the flow allocation $Q_{\mathrm{Z_i}}$ is given as **Flow Allocation Procedure** below.

Since mutual information functions with constraints as described in (4), (9), and (12) are convex functions, the method of Lagrange multipliers is known to converge to the optimal value [10]. Therefore, convergence of the above algorithm is guaranteed.

Due to all initial values of $q_{\mathbf{Z}}(\mathbf{R})$ or $q_{\mathbf{Z}}(\mathbf{R}^{\mathrm{Z_i}})$ converging 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 for each source i, represented by $Q_{\mathrm{Z_i}}$. By varying the parameter s, a range of points on the anonymity-packet loss rate-distortion curve can be obtained for different condition of knowledge in transmission behavior.

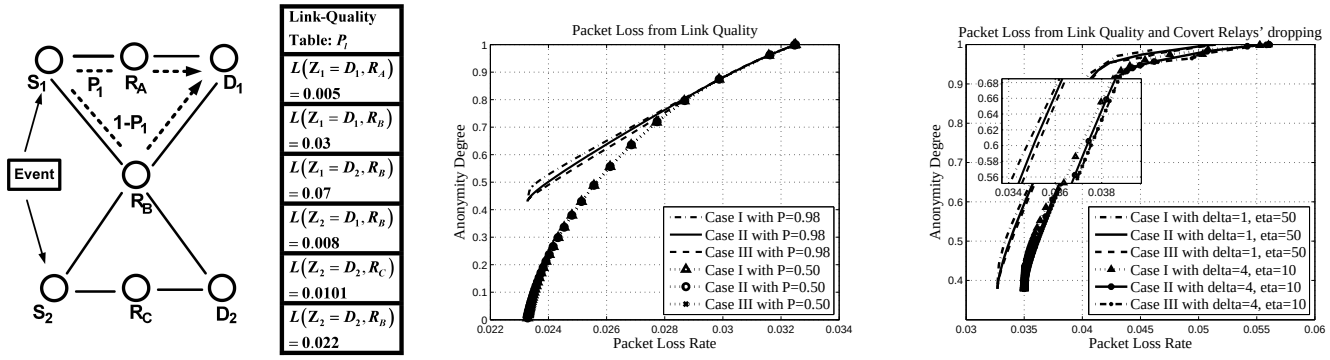

Fig. 1. (a) Example of a wireless network of dependent source-destination pairs with Link-Quality Table. (b) The anonymity degree achieved as a function of packet-loss constraint, when packet losses are due to link quality alone. Three cases of source information are considered under different levels of correlation between Z_1 and Z_2. (c) The anonymity degree achieved as a function of the packet-loss constraint when packet losses are due to both link quality and covert relays' dropping. For each case of source information, different values of dropping time δ and transmission rate η are considered. The correlation between Z_1 and Z_2 was set to $P = 0.98$.

Flow Allocation Procedure: Algorithm for flow allocation

Input: Dropping time δ, Slope value s, Iteration number I
Output: Probability for each route : $\{Q_{\mathrm{Z_i}}\}_{i=1}^{N}$
Initialize: **If** Case I or Case II: $q_{\mathbf{Z}}(\mathbf{R}) \leftarrow \frac{1}{\|\mathbf{R}\|}$
If Case III: $q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}) \leftarrow \frac{1}{\|\mathbf{R}^{\mathrm{Z_i}}\|}$ for i=1..N
while $n < I$
If $\delta = \infty$
If Case I: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathbf{Z}, \mathbf{R}^{\mathrm{Z}_1}, ..\mathbf{R}^{\mathrm{Z_{i-1}}}) \leftarrow$ (6) for i=1..N
If Case II: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i}) \leftarrow$ (10) for i=1..N
If Case III: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i}) \leftarrow$ (13) for i=1..N
else
If Case I: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathbf{Z}, \mathbf{R}^{\mathrm{Z}_1}, ..\mathbf{R}^{\mathrm{Z_{i-1}}}) \leftarrow$ (7) for i=1..N
If Case II: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i}) \leftarrow$ (11) for i=1..N
If Case III: $Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i}) \leftarrow$ (14) for i=1..N
end
If Case I: $q_{\mathbf{Z}}(\mathbf{R}) \leftarrow \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) Q_{\mathbf{Z}}(\mathbf{R}|\mathbf{Z})$
If Case II: $q_{\mathbf{Z}}(\mathbf{R}) \leftarrow \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}) [\prod_{i=1}^{N} Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})]$
If Case III: $q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}) \leftarrow \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}) Q_{\mathrm{Z_i}}(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i})$ for i=1..N
n $\leftarrow$ n+1
end while and **return** $\{Q_{\mathrm{Z_i}}\}_{i=1}^{N}$

IV. Simulation Results

For illustration purpose, we consider a network of two source-destination pairs with three intermediate nodes. We emphasize, however, that the algorithm is valid for multi-hop networks of arbitrary size.

We consider the wireless network topology and link quality table shown in Figure 1(a), in which R_B is a covert relay while R_A and R_C are visible relays with the same transmission rate. The two sources choose their destinations based on an observed event. We let $Pr(\mathrm{Z}_1 = D_1) = Pr(\mathrm{Z}_1 = D_2) = \frac{1}{2}$; i.e., source S_1 selects one of the two destinations D_1 or D_2 with equal probability. We assume that, based on the event, S_2 chooses the same destination as S_1 with probability P and chooses the other destination with probability $(1 - P)$. We show the network system performance for each case calculated by the above **Flow Allocation Procedure** with different values of P and constraints on the average packet loss rate.

In Figure 1(b), we set $\delta = \infty$ for R_B (i.e., no dropping of packets). When $P = 0.98$, the network experiences the best performance under Case I, since source S_i knows the other source's full transmission behavior and can select $\mathbf{R}^{\mathrm{Z_i}}$ accordingly. Similarly, the network performance in Case II is superior to Case III, since each source can utilize the available partial information for route selection. As an example, when the average packet-loss rate constraint is 0.026, the anonymity in Case I, Case II and Case III is 0.647, 0.6344 and 0.62 respectively. To achieve this anonymity, we show how to assign the flow allocation P_1 shown in Figure 1(a) for S_1 as the following. (P_1 and $1-\mathrm{P}_1$ is the probability to select the route R_A and R_B respectively when $\mathrm{Z}_1 = D_1$.) In Case I, P_1 is 0.7698 when S_2 sends packets to D_1, and 0.57 when S_2 sends packets to D_2. In Case II and Case III, P_1 is 0.773 and 0.8543 respectively without considering S_2's actions.

When $P = 0.5$, the probabilities for S_2 to choose the destination nodes D_1 and D_2 are both 0.5 which result in Z_2 and Z_1 becoming independent. Hence, each source S_i's transmission source cannot reveal the other source's destination node, and therefore knowing the other source's transmission source does not improve performance. Thus, Case I, II and III have the same performance when $P = 0.5$.

In Figure 1(c), we use the dropping rate function investigated in [5] with different transmission rates η and time constraints δ. We apply $\delta = 1$ with $\eta = 50$ and $\delta = 4$ with $\eta = 10$ in R_B to evaluate the performance under full, partial, and no information when $P = 0.98$. As shown in Figure 1(c), Case I outperforms Case II, due to the additional information available to the sources, while Case II outperforms Case III.

V. Conclusion

In this paper, we considered multipath wireless networks with a pre-defined subset of covert relays and studied

the problem of flow allocation between correlated source-destination pairs to maximize anonymity under given packet-loss constraints. We formulated the problem of maximizing anonymity through flow allocation as rate-distortion optimization. Making use of side information available to each source in the form of correlation among sources, we considered flow allocation under three different cases of information exchange among the sources: 1) Full information exchange, 2) Partial information exchange and 3) No information exchange. For each case, we formulated Lagrange optimization problems from the constrained rate-distortion problems and presented closed-form solutions when packet losses are due to link quality alone and numerical procedures when packet-losses are caused by link-quality and packet-dropping by covert relays. We also illustrated our formulation via simulation study.

The main focus of this work was to find the optimal route selection for a wireless network with a given set of covert and visible relays in the presence of passive eavesdroppers. In our future work, we will consider the problem of joint optimization of route selection and assignment of relays as covert or visible in order to maximize the anonymity under packet-loss constrains.

REFERENCES

[1] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 8488, February 1981

[2] A. Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," *Proceedings of Privacy Enhancing Technologies (PET)*, Springer-Verlag, 2002.

[3] A. Pfitzmann, B. Pfitzmann, and M. Waidner, "ISDN-MIXes: Untraceable communication with very small bandwidth overhead", *Proceedings of the GI/ITG Conference*, vol. 267, pp. 451-463, Feb, 1991.

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

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

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

[9] C. Yang, B. Alomair, and R. Poovendran, "Optimized Flow Allocation for Anonymous Communication in Multipath Wireless Networks," *Proceedings of the 2012 IEEE International Symposium on Information Theory*.

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