Optimizable Design Schemes in Communication Systems for Providing Anonymity and Confidentiality

Chouchang Yang

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Professor Radha Poovendran, Chair

Professor Payman Arabshahi

Professor Archis Gate

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Optimizable Design Schemes in Communication Systems for Providing Anonymity and Confidentiality

Chouchang Yang

Chair of the Supervisory Committee:
Professor Radha Poovendran
Electrical Engineering

In applications requiring anonymity such as electronic mail, evading website censorship, and file sharing, the identities of source-destination pairs for each data flow should be untraceable. To achieve this goal, anonymous 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. Therefore, how to choose relays as covert and how to select routes among multiple paths should be studied. In this thesis, we present analytical frameworks for relay assignment and route selection in multi-path anonymous wireless networks. We show that joint relay assignment and route selection can be formulated as a convex optimization problem which guarantees global optimum solution. Our frameworks also consider two special cases. The first case is the route selection alone by giving the relay configuration. The second case is choosing relays as covert or not given the route selection strategies.

Given a subset of nodes is chosen to act as covert relays to hide timing information from unauthorized observers. We propose route selection 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. When sources and destinations are independent to each other, each source allocates route independently as well. When sources destination pairs are dependent, we investigate how to allocate routes for each source-destination pair to maximize anonymity with packet-loss rate as a constraint. 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 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 optimal relay configuration for given fixed route section can be derived from the information theoretic anonymity metric of joint relay and route selection. We showed that the problem of optimal relay assignment based on the trade-off between anonymity and throughput in a multiple wireless network can be solved by re-deriving rate-distorting frameworks. Our framework guarantees efficient computation of the global optimum.

For RFID system, to ensure security and privacy of passive RFID have notoriously been a difficult problem. In a passive RFID system, each tag has power and computational constraints. As a result, providing privacy protection can be a very challenging task. In this thesis, we propose a physical layer privacy protection scheme termed Varying Readers Transmitted Amplitude (VRTA). This scheme provides the data confidentiality in the uplink direction. i.e., tag to reader. By doing so, the users privacy is assured against the passive eavesdroppers. This scheme requires no modifications on the tag and the existing protocols as well as no pre-shared secrets. It has minimal impact on the existing system. We also perform analysis to show our scheme is theoretically secure against one and multiple passive eavesdroppers with properly chosen system parameters. Then, by applying VRTA system, the RFID tag can establish key agreement with reader for encryption or authentication purpose. Finally, we implement our scheme using USRP to validate our theoretical results and to further test the performance of our proposed scheme.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, I would like to express my sincere appreciation to my advisor, Professor Radha Poovendran, for his guidance and support in my PhD study. He has helped me keep pursuing my research goal and given me advices whenever I encountered difficulties.

I would like to thank the members in my PhD committee, Professor Payman Arabshahi, Professor Arvind , and Professor Archis Ghate, for their valuable suggestions and career advices. Also, many thanks to Dr. Huai-Rong Shao and Dr. Alanson Sample for providing the insightful discussions during my internship works. Especially thanks to Prof. Liu Hui, Prof. Basel Almoair and Prof Guang Gong for their supporting and advise for during my PhD study.

I am grateful to my colleagues and lab mates, Dr. Andrew Clark, Phillip Lee, Elisabeth Senmarti Robla, Xuhang Ying, Zhipeng Liu, Kalikinkar Mandal, Hossein Hosseini, Laila Abudahi, Sean Rice, and Fei Huo, who have provided huge help and friendship to me. Without their kindness and great help, I could not settle down my life in US and focus on my research.

I also thank all my friends whom I have met in Seattle. It is they who provide great friendship so that I feel energetic in my life. I would like to thank my wife Hwayoung Chae. She has always been around me during my good and bad times. I will never forget the valuable suggestion and encouragement she has given to me.

I will forever be thankful to my parents and my sister. They have been supportive during my PhD study. Thanks to their understanding and constant encouragement, I am able to pursue the degree without worries. I clearly know they will always be there whenever I need help and suggestion. I am the luckiest person because of my dear family

DEDICATION

To my parents.

Chapter 1

INTRODUCTION

A communication system provides connectivity and services between various senders and recipients for data transfer and individual communication. As the users are becoming increasing dependent on such systems, the anonymity, security and privacy also became very important for users. Since sensitive data operations and the storage of personal information would be processed during communication sessions, malicious attackers may affect those properties by intruding communication systems. Therefore, to provide reliable and safe communication against malicious attackers, the system design needs to consider not only system performance and capability but also security, privacy and anonymity aspects. The thesis in here focus on two specific problems in this area, namely, anonymous networking and data confidentiality in RFID.

Computer networks recently have become the truly lifeblood of business. As the success of network services become increasingly intertwined, speed, remote management, wireless networking and reliability of networked devices are all concerns that have been studied. However, recently it also have raised new problems concerning anonymity and privacy, since a large amount of personal information passes through computer networks each day. To preserve an individual's valuable information against adversaries when communicating, cryptographic approaches have been considered to a large extent in computer networks [33, 43]. Cryptography alone, however, is insufficient to prevent an adversary from determining the identities of the communicating parties [18]. Since network systems are designed to aim at high speed and low-latency for various data-driven services, those packets forwarded by low-latency relays expose the potential sources and destinations. As the results, the users identity can be exposed by applying traffic analysis [4, 11, 20], thus breaking anonymity. Although adding randomly delay in forwarded packets can thwart traffic analysis to preserve anonymity, those techniques result result in network performance degradation [50]. A unified mechanism for optimal anonymous network must incorporate the trade-off between anonymity and network performance, which remains open.

Radio frequency identification (RFID) has gained tremendous popularity in the recent years. Due it simple, cheap and low-energy, RFID applications are widely used these days such as passport, tracking goods and door access. However, the tiny, inexpensive RFID tags are also limited to have strong computation ability. Therefore, RFID tags especially passive tags are not able to implement strong cryptographic primitives [24]. Without such primitives, the adversary could easily eavesdrop on the communication session and potentially gain all the secrets about the tag. In doing so, the secrecy of the messages and the privacy of the tag is violated. The new defense approach should also consider the inherent nature of RFID systems to provide data confidentiality against malicious adversaries.

The need to address anonymity and security while maintain system performance motivate the development of network and RFID system. The traditional network system designs focus on the performance and reliability while our approach not only address the performance but also investigate how to improve the anonymity. In general, problems of computer networks and RFID do not address the solution in terms of anonymity and security [23, 27]. Identifying problem structure that leads to efficient solution algorithms is a key step towards achieving optimal system designs and performance in networked and RFID systems.

1.1 Contributions of this Thesis

The contribution of this thesis is a universal optimization framework for anonymous networked and RFID data confidentiality systems (Figure 1.1). By formulating anonymous networked systems as rate-distortion problems, the proposed algorithms can maximize anonymity based on different throughput performance. Moreover, three different network structures have been studied based on each type of networks [55–57]. Specifically, the first is joint route and relay selection system, while the second and third types are route and relay selection alone system. For RFID data security, we show how to optimize RFID system design for data confidentiality by utilizing the baseband property in wireless receiver to secure the data transmission from RFID tag. The performance results and optimized approaches are published in [22]. In general, problems of some daily communication system do not consider anonymity and security aspects which lead to be addressed and investigated for efficient solution during this whole manuscripts.

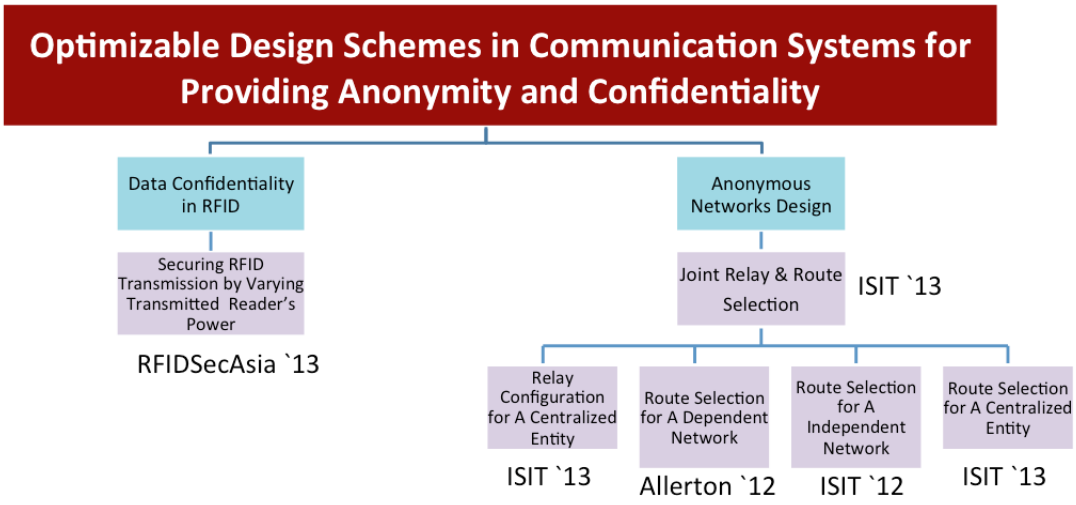


Figure 1.1: Organization of this thesis.

1.1.1 Optimal Designs for Anonymous Networks

In applications requiring anonymity such as electronic mail, evading website censorship, and file sharing, the identities of source-destination pairs for each data flow should be untraceable. In general,anonymity is an essential requirement for many communication systems to prevent unauthorized observers from determining the identities of the communicating parties. To achieve this goal, source and destination pairs in each communication session need to be unknown. However, by eavesdropping the timing of packet delivery, the attackers can use timing information to deduce possible source-destination pairs [21].

When relays directly forward received packets, there is a time correlation between incoming and outgoing packets. Hence, an eavesdropper can use timing-based traffic analysis to identify network flows such that the flow of packets can be traced back to the original sources and associated to possible destinations. Therefore, each source-destination pair will be discovered through the traffic analysis and the anonymity is hurt.

To thwart timing-based traffic analysis, the relays add random delay to forwarded packets. While such mechanisms increase anonymity, they cause performance degradation from delaying packets [49]. The route selection chosen by sources among multiple paths composed by various relays

with or without delays will result in various anonymity. Moreover, different route selections result in different packet-loss through link quality as well as packets dropping by relays due to buffer overflow from random delays. In general, there are two components that affect the anonymity and performance of a wireless network, namely, (i) relay configuration by applying randomly delay or not and (ii) the routing paths chosen by relay source nodes. In considering these two parameters under fixed network topologies, the optimized routes and relays selection should consider both network performance and anonymity. However, having low-latency networks will result in low anonymity while high-latency networks composed of randomly delaying relays have high anonymity but low throughput. Therefore, how to maximize the anonymity while still guaranteeing network performance should be investigated. By formulating joint design of route selection and relay configuration in multi-path anonymous networks as a rate-distortion problem, the optimal relay and route selection can be obtained. To do so, we first proved that this rate-distortion problem is convex. To develop various anonymous networks via rate-distortion problem, we consider different types of networks. We first consider a wireless network where routes and relays are jointly selected to maximize anonymity while maintaining performance. Then, two special cases are further investigated. The first special case is optimizing the anonymity via route selection alone while relay configuration is given. The second case is a wireless network with given route selection, while relays choose configuration assignments alone. For each of them, I give the algorithms in terms of relay assignment as well as route selections obtained by solving convex optimization problems. The preliminary results have verified the proposed algorithms can achieve maximal anonymity while the network performance as packet-loss is under constraints.

Route Selection Alone for Anonymous Wireless Networks

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.

In additional, based on control mechanism, we consider three different type of networks. The first network is (i) a centralized entity of sources choose the route with globe information. The second and third type is a network with given relay configuration, while a decentralized entity of sources select route individually with (ii) independent source-destination pairs or (iii) dependent source-destination pairs. For independent sources-destination pairs, each source choose destination independently. Therefore, the route selection will only depend on itself information. Opposed to the case of independent source-destination pairs when sources choose destination dependently, 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. 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 sources 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.

Relay Assignment Alone for Anonymous Wireless Networks

The idea of hiding source and destination pairs from eavesdroppers is very essential for anonymous networks. Although covert relay can prevent eavesdropper deduce possible source destination pairs via traffic analysis, this approach also result in low network throughput. Therefore, given fixed route selection strategies, how to optimize anonymity while maintain network performance would be decided by the relay configuration. Under this scenario, a mathematical notion for anonymity of network routes using Shannons equivocation is introduced to measure anonymity. By assuming a global passive eavesdropper who observes transmission schedules of all nodes in the network, the relay assignment as covert or not is obtained through a convex optimization. In addition, the relationship between overall network throughput and the anonymity is obtained by drawing a connection to the rate-distortion trade-off in information theory.

1.2 Data Confidentiality in RFID

Radio Frequency Identification (RFID) is a promising technology which can perform automated and unique identification of objects. Because of this attractive property, RFID since its invention has found many applications in various industries. This includes supply chain management, inventory control and passports/driver's license. As RFID technology becomes more ubiquitous, popular and widespread, there have been many attacks on the RFID systems [25, 34, 35]. Many security and privacy concerns has been raised as a result. One of major security concerns is the data confidentiality. Many passive Radio Frequency Identification (RFID) systems do not incorporate data confidentiality protections. In the entire communication session, the tag's replies are sent in plain text. Due to the nature of wireless channels, the open communication is susceptible to eavesdropping. The tag's ID and contents are easily compromised and recovered by the attacker. This can lead to further malicious attacks such as tag cloning and tracking. Therefore, it is imperative that tag to reader data confidentiality to be protected.

To secure data transmission between tags and readers, I have proposed a physical layer standalone system called Varying Reader's Transmitted Amplitude (VRTA). It provides data confidentiality protection to passive RFID systems against passive eavesdroppers. In our system, the reader generates and transmits a randomly varying amplitude waveform. This waveform can be successfully removed by the legitimate reader and is seen as interference in the view of the attacker. Hence, the reader and tag do not require pre-shared secrets for secure communication. In addition, the VRTA system utilizes only one transmitter and can be applied to all current commercial readers. Moreover, this proposed scheme is able to establish light-weight encryption for securing transmission from reader to tag direction against eavesdropping. Therefore, our proposed VRTA scheme is able to provide data confidentiality for RFID system.

1.3 Related Work

In this section, we give an overview of the related work in each of the areas considered in this thesis.

1.3.1 Optimal Designs for Anonymous Wireless Networks

Anonymity is an essential requirement for many communication systems, such as electronic mailing systems, evading website censor-ship, file sharing applications, hiding the content and users information from unauthorized parties. Therefore, the content and identity of source-destination pairs must remain anonymous. In order to prevent unauthorized adversaries knowing content and header associated sources and destinations from packets, cryptography have been applied for encryption packets [16, 54]. Anonymous networking protocols use cryptography to hide packet header, since the identities of source-destination pairs should remain unknown [9, 29].

Onion routing is a application layer protocol that provides low-latency, bidirectional and anonymous connection [10]. Onion routing uses cryptography to hide header and content data, and removing identification information from the header [44]. Onion routing decouples network, transport and application layers. Users can virtually execute any protocol, because onion routing wraps all content in small cells that cannot be eavesdropped by the censor; that is why it is the preferred mechanism to circumvent censors Onion routing. Although onion routing can preserve privacy encrypting the content against eavesdropping, source-destination pairs are still possible to exposed due to its low latency property [21]. That is because the transmitted packets of flows, 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 as depicted, leading to possible exposure of source-destination pairs [20].

The problem of decreasing relays timing correlation to prevent timing-based traffic analysis has attracted research attention these days. In [7], 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 [36]. In [18], 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.

By following this concept, the covert relay is proposed in [49] transmitting packets according

to a pre-specified probabilistic schedule, sending fake traffic when no real packets are available. This pre-specified probabilistic schedule is able to remove timing correlation at the relay nodes. 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 [18]. On the other hand, non-covert (visible) relays do not provide any anonymity. Moreover, different routes composed by various covert and visible relays result in different anonymity.

Therefore, the anonymity is decided by both relay configuration as covert or visible, and the selection of routing paths. 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 [50]. However, a unified approach for relay assignment and route selection that guarantees optimal anonymity, while maintaining network performance, is currently lacking. In general, a universal framework which can maximize anonymity but also maintaining network performance respect to different types of networks should be investigated.

1.3.2 Data Confidentiality in RFID System

Radio Frequency Identification (RFID) is a promising technology which can perform automated and unique identification of objects, e.g. supply chain management [3], object tracking [15] and environment sensing [51]. The tiny, inexpensive RFID tags can be easily attached to objects for seamless identification. A typical RFID system composed of RFID readers and tags. Each tag is applied backscatter technique for data communication. The RFID tags hold a memory that stores their identification number and data.

Because of its cheap cost and simple structure, RFID since its invention has found many applications in various industries. This includes supply chain management, inventory control and passports/driver's license [52]. As RFID technology becomes more popular, there have been many attacks on the RFID systems [25, 34, 35]. Many security and privacy concerns has been raised as a result. One of major security concerns is the data confidentiality.

However, one glaring weakness of RFID tags, especially passive RFID tags is its lack of capability for implementing strong crypto primitives for security purposes. When no or a weak crypto primitive is implemented, the adversary could easily eavesdrop to the communication session between the reader and the tag, he can potentially gain all the secrets about the tag. In doing so, the secrecy of the messages and the privacy of the tag is violated.

There have been a number of efforts to ensure the data confidentiality protection. This most common method is using the conventional cryptography encryption approach. In addition, physical layer signalling scheme such as frequency hopping [48, 53], jamming interference [1, 46] and direct sequence spread spectrum [42] have also been investigated.

Using the conventional encryption approach to secure RFID systems and achieve data confidentiality protection typically requires the tag to have sufficient memory as well as extensive computational power [14, 19]. This may be accomplished with more expensive and powerful active RFID tags, but it may be very challenging with power and memory constrained passive tags [26]. In addition, under limit resource from inherent hardware aspect, strong cryptographic operation is hard to afford it such that the encrypted tag's ID may still be exposed by intercepting messages from multiple communication sessions [2, 31]. Moreover, not only encryption but also authentication process in RFID system compromise the key agreement due to those simple hardware limitation [41].

Therefore, cryptography may not be suitable solution in providing the data confidentiality protection under RFID device. Frequency hopping is one popular approach used in the physical layer to ensure the system security in the wireless communication systems [48, 53]. Although frequency hopping can mitigate the passive eavesdropping attack by randomly and continuously changing its central carrier frequency, it contains two major drawbacks: 1). It potentially occupies a wider frequency spectrum, so the spectrum efficiency is decreased. 2). It can limit the tag's data rate [45].

Using jamming to secure RFID systems have been proposed in [46] and [1]. In these methods, a jamming (noise) signal is broadcasted by a separate transmitting antenna along with reader's continuous waves. At present, however, techniques for designing an optimal waveform to maximize the security as data confidentiality are not available. In addition, the separate transmitter for jamming signal will increase the cost of reader hardware.

The paper in [22] demonstrated that it is possible to use one single transmitter to perform the jamming by applying the property of baseband in varying amplitude while supplying sufficient

power to the tag. In this case, the additional transmitter for broadcasting noise is no longer needed. However, no theoretical analysis and system design parameters are given in the paper. Moreover, although physical layer security approaches are able to secure data from tag to readers, how to maintain data confidentiality from reader to tag is still unknown. One challenge is due to that the receiver of RFID tag is low-cost design hard to afford physical layer approach. Most existing works choose lightweight cipher with bitwise operation to encrypt data [40]. Similarly, the RFID tag authentication applied bitwise operation with secret key encryption to verify reader [37, 38]. However, how to deliver key agreement need to be addressed such that the readers and tags can use secret agreement for data encryption or mutual authentication purpose. Moreover, several literature have shown that the key agreement and tag's ID can be exposed via the mutual authentication protocols. Therefore, we will show how to apply physical security to set up a secure channel for data transmission as well as secret key establishment.

Chapter 2

OPTIMAL DESIGNS FOR ANONYMOUS NETWORK

Anonymity is an important requirement for Internet services such as web browsing, financial transactions, and file sharing. Users can use anonymous network to access internet service without revealing themselves identity. The existing frame work of anonymous network is to route traffic through mix relays, which encrypt packets and forward received packets in a random order. Although mix relays provide anonymity against eavesdroppers, they also have poor throughput performance due to delaying of forwarded packets.

Existing works [49] show that by assigning relays as mix relays or not in single path network, the balance of network throughput and anonymity can be solved. However, most networks have multiple paths between each source and destination pair. Therefore, the route selection will also affect the anonymity and network throughput. For example, we consider two different paths for the same source and destination pair where one route is composed of mix relays while the other is composed of normal relays. When source choose the route composed of mix relays, the anonymity regarding the source-destination pair is achieved but with poor throughput network performance. On the other hand, if source choose the route composed of normal relays, source destination pair can be found but with high throughput performance. In here we denote mix relays as covert relays while normal relays as visible relays.

In general, there are two design parameters that affect anonymity and throughput performance of a network. The first parameter is relay configuration as covert relays which provide higher anonymity with lower throughput, while visible relays provide no anonymity and high throughput. The second parameter is route selection between covert and visible relays, since packets traversing covert relays will minimize information leakage of source-destination pairs but will experience additional delay result in low throughput. In this work, we consider the reduction in throughput due to link-quality and packet-dropping from mix nodes, since mix nodes have longer latency which results in congestion and packet loss. By formulating the joint relay configuration and route se-

lection as a convex optimization problem, we are able to maximize the anonymity while satisfying network throughput constraints. Our algorithms enable efficient computation of both relay assignment and routing schedules for each source.The goal of this work is to design networks that achieve anonymity while also satisfying throughput constraints by joint assigning relay configuration and route selection.

2.1 Model and Preliminaries

We first describe our adversary model and show how to model our problem as convex optimization framework for joint relay configuration and route selection. Then, we define relays as covert and visible by introducing their mechanism and function as follows. All transmitted packets by relays are encrypted at each hop using pairwise shared keys.

Adversary model

We consider passive attacks by global eavesdroppers who are able to observe all transmissions at each network nodes and have knowledge of the whole network topology. The eavesdroppers are 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 prevent the adversaries from associating packets with their sources and destinations from the content or header. The eavesdroppers can only identify source-destination pairs by using timing-based traffic analysis with the knowledge of the network topology.

Visible Relays

A visible relay forwards packets in first -in, fist-out order without adding any delay. Therefore, an eavesdropper can apply timing-based traffic analysis to trace the forwarded packets to the corresponding sender and deduce possible destination nodes by using the knowledge of network topology.

As illustrated in Figure 2.1, one of the sender $n1$ is sending the packet to the relay r_j and then forward to node $n4$. When the relay is setting as visible, the time stamp at transmitting at node $n1$ and the forwarding at relay r_j has correlation such that eavesdropper can deduce the forwarding packet is from the sender $n1$.

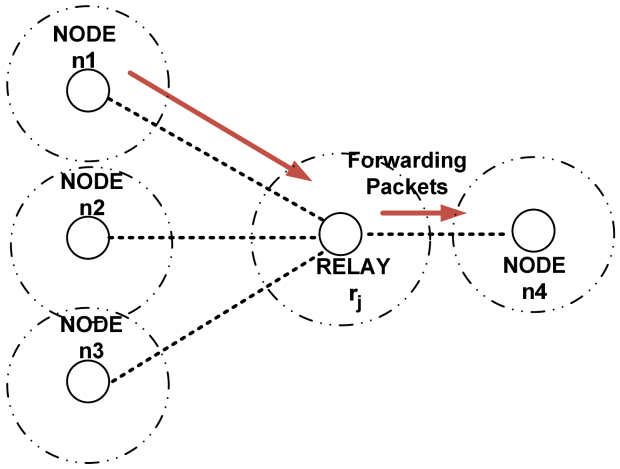


Figure 2.1: Example of a wireless network topology with senders and a relays to illustrate covert and visible relays.

Covert Relays

A covert relay transmits packets according to a pre-specified probabilistic schedule. The covert relay transmits dummy packets, when there are no packets to be forwarded. Fake packets transmitted by covert relays are identified at the next hop and be dropped. By applying the independent transmission schedule with real and fake packet along as well as encryption of the entire payload (packet and its headers), the covert relay prevent eavesdroppers from associating the forwarded packets with incoming flows.

By assumption, adversaries are unable to trace packets by observing covert relays. As shown in Figure 2.1, if relay r_j is covert, from the eavesdroppers viewpoint, the current outgoing packet could be coming from node $n1$, node $n2$, or node $n3$. Therefore, unlike the case of visible relays in previous example, the adversary cannot associate a source node with the observed packets. In the example of Figure 2.2, when node n1 sends a packet to covert relay rA, eavesdroppers are unable to determine (with certainty) which one of those three outgoing packets is from node n1. In other words, assuming node n1 chose its next hop relay with equal probability, the three events that node n1 is sending packets to covert relay rA, covert relay rB, or covert relay r are indistinguishable. .

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 [18]. Given that the buffer size $K > 1$, the packet-dropping rate for covert relays with various incoming rate is

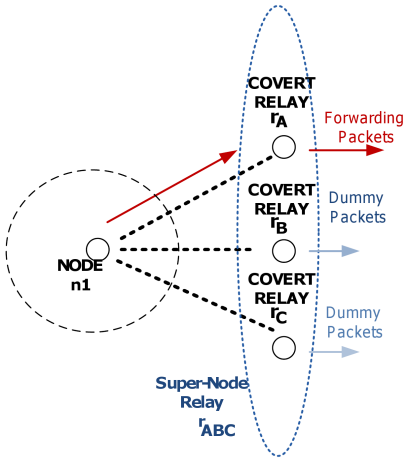


Figure 2.2: Example of a wireless network topology with senders and a relays to illustrate the mechanism of multiple covert relays .

$$P_e(\eta_{in}, \eta_{out}, K) = (\gamma^K - \gamma^{K+1})/(1 - \gamma^{K+1}), \tag{2.1}$$

where $\gamma = \eta_{in}/\eta_{out}$.

Since the delay in visible relay is much shorter than covert, the buffer overflows at visible relays are assumed to be negligible.

Network Structure

We consider a wireless network composed of N sources, M destinations and α intermediate nodes as covert or visible relays. The sources are $\{S_1, S_2,...,S_N\}$, destinations $\{D_1, D_2,...,D_M\}$ and relays $\{G_1, G_2,..., G_\alpha\}$ as depicted in Figure 2.3. Each source S_i has transmission rate η_i and chosen destination nodes randomly. 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. A random variable $\mathrm{R_i}$ refers to the route used by source

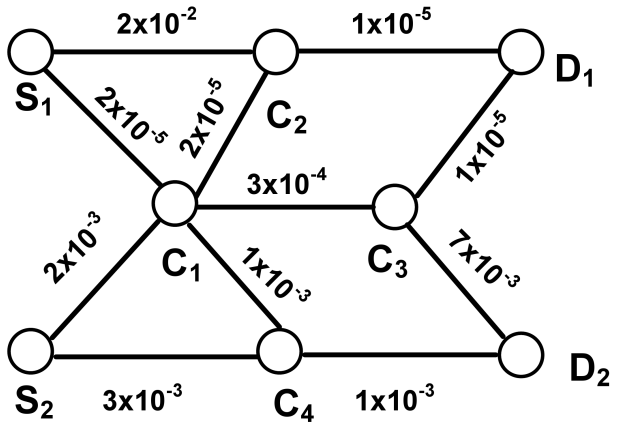


Figure 2.3: Example of a wireless network topology with two sources and two destination with four intermediate nodes for route and relays selection while the numbers refer to the link quality of each link.

S_i. Then, we denote $\mathbf{R} = [\mathrm{R}_1, \mathrm{R}_2, \cdots, \mathrm{R}_\mathrm{N}]$ 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 C_i is covert and visible, respectively. Then, the $\mathbf{C} = [\mathrm{C}_1, \mathrm{C}_2, \cdots, \mathrm{C}_\alpha]$ is 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 $\mathbf{Z}$.

The packet-loss in here is assumed from two factors, link-quality at the physical layer and buffer overflow at the network layer. The fraction of packets from source S_i that are lost at relay G_j due to link quality for route selection R_i is denoted 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 (2.1). Since these two factors are from different layers, the packet-loss are assumed to be independent to each other as well.

Anonymity Measurement Metric

To measure the anonymity, we introduce the entropy concept as measuring metric. The $H(\mathbf{Z})$ refer to 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 after timing-based traffic analysis is defined as $H(\mathbf{Z}|\mathbf{R}, \mathbf{C})$. As in [11], 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 traffic analysis.

2.2 Problem Formulation and Proposed Algorithms

Since adversaries can observe the whole routes chosen by sources, maximizing anonymity is equivalent to minimizing the source-destination pairs information available to the eavesdroppers. The information obtained by eavesdropper quantifies the loss in anonymity as $H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R}, \mathbf{C})$. Therefore, minimizing the anonymity loss by considering network constraints is

$$\min \left[H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{R}, \mathbf{C}) \right] \text{ s.t. } d(\mathbf{Z}, \mathbf{R}, \mathbf{C}) \leq D \tag{2.2}$$

The function $d(\cdot)$ is average packet-loss rate arising from the source-destination pairs $\mathbf{Z}$ with the route vector $\mathbf{R}$ and relay configuration $\mathbf{C}$. Here, D is the maximal packet-loss rate that network can afford. The minimization of anonymity loss is achieved by minimizing reduction in entropy of source destination pairing after traffic analysis. The packet-loss can be viewed as a penalty from the assignments in terms of routes and relays. Therefore, this minimizing anonymity loss is equivalent as a rate-distortion optimization problem. In next chapter, 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.

2.2.1 Convexity of 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

$$
\begin{aligned}
E_{\mathbf{Z}} &= \frac{\text{total packet-loss rate at } \alpha \text{ relays and } M \text{ destinations}}{\text{total transmission rate from N sources}} \\
&= \frac{1}{\sum_{i'=1}^{N} \eta_{i'}} \Bigg[\sum_{i=1}^{\alpha} \Bigg\{ \sum_{\mathbf{R},\mathbf{C}:[\mathrm{C}_i]=1} \eta_{in}(i) * P_e(\eta_{in}(i), \eta_{out}, K) \\
&+ \sum_{\mathbf{Z},\mathbf{C},\mathbf{R}} \sum_{j:\mathrm{R}_j \in G_i} \eta_j * P(\mathbf{Z}) Q(\mathbf{C}) Q(\mathbf{R}|\mathbf{Z}, \mathbf{C}) L_{j,i}(\mathrm{R}_j) \Bigg\} \\
&+ \sum_{l=1}^{M} \sum_{\mathbf{Z},\mathbf{C},\mathbf{R}} \sum_{j':\mathrm{R}_{j'} \in G_i} \eta_{j'} * P(\mathbf{Z}) Q(\mathbf{C}) Q(\mathbf{R}|\mathbf{Z}, \mathbf{C}) L_{j',D_l}(\mathrm{R}_{j'}) \Bigg], \qquad (2.3)
\end{aligned}
$$

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 (2.3) 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

$$\min_{Q(\mathbf{C}),Q(\mathbf{R}|\mathbf{Z},\mathbf{C}):E_{\mathbf{Z}}\leq D}[H(\mathbf{Z})-H(\mathbf{Z}|\mathbf{R},\mathbf{C})], \tag{2.4}$$

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 (5.1) 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

$$\min_{Q\in\mathbf{\Lambda}}\sum_{\mathbf{Z},\mathbf{R},\mathbf{C}}P(\mathbf{Z})Q(\mathbf{R},\mathbf{C}|\mathbf{Z})\log\frac{Q(\mathbf{R},\mathbf{C}|\mathbf{Z})}{q(\mathbf{R},\mathbf{C})}, \tag{2.5}$$

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

$$\mathbf{\Lambda}=\begin{cases} E_{\mathbf{Z}}(Q(\mathbf{R},\mathbf{C}|\mathbf{Z}))\leq D,\ \forall\,\mathbf{Z} & \text{(4a)},\\ \sum\limits_{\mathbf{R},\mathbf{C}}Q(\mathbf{R},\mathbf{C}|\mathbf{Z})=1,\ Q(\mathbf{R},\mathbf{C}|\mathbf{Z})\geq 0,\ \forall\,\mathbf{Z} & \text{(4b)},\\ \sum\limits_{\mathbf{R}}Q(\mathbf{R},\mathbf{C}|\mathbf{Z})=\sum\limits_{\mathbf{R},\mathbf{Z}}P(\mathbf{Z})Q(\mathbf{R},\mathbf{C}|\mathbf{Z}),\ \forall\,\mathbf{Z} & \text{(4c)}. \end{cases}$$

Lemma 2.1. *Let $Q^*(\mathbf{R},\mathbf{C}|\mathbf{Z})$ denote the optimal solution of (2.5). The optimal solution to (5.1), 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 (5.1). We then prove that they are optimal under (5.1). To check feasibility of $Q^*(\mathbf{C})$, $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$, note that (4a) implies that $E_{\mathbf{Z}}(Q^*(\mathbf{C}),Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C}))\leq D$.

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

$$\sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z})\hat{Q}(\mathbf{C})\hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C}))\log\frac{\hat{Q}(\mathbf{C})\hat{Q}(\mathbf{R}|\mathbf{Z},\mathbf{C}))}{q(\mathbf{R},\mathbf{C})} \tag{2.6}$$
$$< \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z})Q^*(\mathbf{C})Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C}))\log\frac{Q^*(\mathbf{C})Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C}))}{q(\mathbf{R},\mathbf{C})}.$$

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, $E_{\mathbf{Z}}(Q(\mathbf{R},\mathbf{C}|\mathbf{Z})) \leq D$ and (4c) holds by assumption. Furthermore, by (2.6)

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

This, however contradicts the assumption that Q^* is optimal, implying that (2.6) cannot hold and $Q^*(\mathbf{C}), Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is the optimal solution pairs to (5.1). □

The following theorem establishes convexity of problem (2.5).

Proposition 2.1. *The optimization problem presented in (2.5) 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 (2.5) is convex as a function of $Q(\mathbf{R},\mathbf{C}|\mathbf{Z})$. We denote the summand of the first term in (2.3) as f_1^i

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

and $P_e(x,\eta_{out},K)$ satisfies $\frac{dP_e}{dx} \geq 0$ and $\frac{d^2P_e}{dx^2} \geq 0$, 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 (2.3) 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 (2.3) 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 (2.5) is convex, when the buffer size $K > 1$. $\square$

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 show how to obtain the optimal relay configurations and route selection to maximal anonymity while maintain network performance.

2.3 Optimized Relay-Route Assignment for Anonymous Networks

Using the condition in (2.5) with system average packet loss rate not exceeding D, the optimization equation with Lagrange multiplier can be rewritten as

$$J = \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z})Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) \log \frac{Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})}{\sum_{\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})}, + \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} v(\mathbf{Z})Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) \tag{2.7}$$
$$+ \sum_{\mathbf{C}} \lambda(\mathbf{C}) \Bigg(\sum_{\mathbf{R}} Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) - \sum_{\mathbf{R},\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) \Bigg) - sE_{\mathbf{Z}}(Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}))$$

By solving Lagrange multiplier function from (2.7) from $\frac{\partial J}{\partial Q(\mathbf{R},\mathbf{C}|\mathbf{Z}} = 0$, the globally optimal solution $Q^*(\mathbf{R}\, mathbf{C}|\mathbf{Z})$ can be obtained. Since the packet loss is also decided by $Q(\mathbf{R}\, mathbf{C}|\mathbf{Z})$ for a finite buffer size in cover relays, the Lagrange multiplier function is non-linear in $Q(\mathbf{R}|\mathbf{Z}, \mathbf{C})$. To obtain optimal $Q^*(\mathbf{R}|\mathbf{Z}, \mathbf{C})$, Newton's method is applied to solve non-linear equation iteratively.

When buffer size in covert relays is increasing, the dropping rate is decreasing. Hence, the average packet-loss rate will be dominated by link-quality. As the result, the equation can be viewed as linear equation to obtain Q^* approximately.

To illustrate how to obtain Q^*, we consider two different conditions based on the buffer size in covert relays. In the first case, 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 case, we incorporate packet dropping by covert relays with finite buffers.

2.3.1 *Packet-loss from link-quality alone*

By setting buffer size K toward to infinite, the optimal $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ is

$$Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) = \frac{q(\mathbf{R}, \mathbf{C}) \sum_{\mathbf{R}} q(\mathbf{R}, \mathbf{C}) \exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}{\sum_{\mathbf{R}} q(\mathbf{R}, \mathbf{C}) \exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}, \tag{2.8}$$

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

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

Once the $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ is obtained, then we can find the relay configuration by using $\sum_{\mathbf{R},\mathbf{Z}} P(\mathbf{Z}) Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) = Q(\mathbf{C})$ and route selection via $\frac{Q(\mathbf{R},\mathbf{C}|\mathbf{Z})}{Q(\mathbf{C})} = Q(\mathbf{R}|\mathbf{Z}, \mathbf{C})$. The relay configuration $Q(\mathbf{C})$ indicates the probability distribution of relays as covert and visible while the route selection is $Q(\mathbf{R}|\mathbf{Z}, \mathbf{C}$ that each source choose the routes based on current relay configuration and source-destination pair.

The parameter s is the values that determine the trade-off between anonymity and packet-loss rate. When s is increasing, this result in less anonymity but better network performance via less packet-loss rate. Similarly, when s is decreasing, the network has higher anonymity but high packet-loss rate.

2.3.2 *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, $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ can be obtained by solving the function derived via Lagrange multipliers function.

$$f(\mathbf{R}, \mathbf{C}) : Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) - \frac{q(\mathbf{R}, \mathbf{C}) \sum_{\mathbf{R}} q(\mathbf{R}, \mathbf{C}) \exp[sT_{\mathbf{Z}}]}{\sum_{\mathbf{R}} q(\mathbf{R}, \mathbf{C}) \exp[sT_{\mathbf{Z}}]}, = 0, \tag{2.9}$$

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

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

Since equation (2.9) is non-linear in $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$, we use Newton's method to obtain $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ iteratively. Letting $\mathbf{Q}^{(n)}$ denote the value of $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ at the n iteration. $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z},)$ is obtained

by

$$\mathbf{Q}^{(n+1)} = \mathbf{Q}^{(n)} - \left[[\nabla \underline{f}]^{-1} \underline{f} \right]|_{\mathbf{Q}^{(n)}}, \tag{2.10}$$

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 (5.5) for all flows set $\mathbf{a}_i \subset \mathbf{R}, \mathbf{C}$ given source-destination pair $\mathbf{b} \subset \mathbf{Z}$.

Similarly, when the $Q^*(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ is obtained from Newton approach and then by using Q^*, we can obtain relay configuration $Q^*(\mathbf{C})$ and route selection $Q^*(\mathbf{R}|\mathbf{C}, \mathbf{Z})$.

2.3.3 Iteration Algorithm

In previous section we have showed how to obtain the optimal $Q^*(\mathbf{R}, \mathbf{C}|\mathbf{Z})$. Here we gave the iteration procedure in detail to indicate how to calculate the optimal route selections for centralized entity.

An iterative algorithm for obtaining $Q^*(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ is given as follows.

The Δ refer to the resolution of $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z})$ value. The iteration procedure calculate the value Q and update it toward to the optimal Q^* time. Then, the difference of each time Q will be getting smaller. We can set Δ as threshold such that when the difference between each iteration is smaller than that, the iterations procedure will be stopped and return the final value.

2.4 Result and Performance Discussion

To verify joint relay and route selection performance, we consider a network topology shown in Figure 2.3 where each source has multiple routes to destination nodes. The relays C_1, C_2, C_3 and C_4 are either visible or covert.

The transmission rate η for each node is 5, while buffer size K in covert relay is 5. Each source chooses one of the two destination D_1 or D_2 with equal probability. We show the network system performance for joint selection calculated by the above iteration procedure with single route selection and relay configuration performance comparisons.

In Figure 2.4, we set relays C_1 and C_3 as covert relays for route selection alone while giving uniformly route selection for relay configuration. The joint relay and route selection in Figure 2.4 has better performance than route selection or relay configuration alone, since joint selection approach can maximize anonymity via two parameters. Unlike joint selection, the relays or routes

Iteration Procedure : Algorithm for route selection

Input: Resolution Δ, Buffer size K, Slope value s, Iteration number, I Source-destination pair, $\mathrm{Z_i} = \beta$

All routes and relay configuration for this source-destination pair β : $\underline{\alpha_1}, \underline{\alpha_2}...\underline{\alpha_l}$

Output: Probability for each route $Q(\underline{\alpha_j}|\beta),\ j = 1, 2..l$

$q(\underline{\alpha_j}) \leftarrow \frac{1}{l}$, for $j = 1, 2..l$

while $n < I$

  if $\delta = \infty$

    $Q(\underline{\alpha_j}|\beta) \leftarrow$ eq. (2.9) for $j = 1, 2..l$

  else

    while $m_j > \Delta$ for $j = 1, 2...l$

      $m_j \leftarrow Q(\underline{\alpha_j}|\beta)$

      Compute $Q(\underline{\alpha_j}|\beta) \leftarrow$ by eq.(2.10)

      $m_j \leftarrow |m_j - Q_{\mathrm{Z_i}}|$ 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 $\{Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)\}_{j=1}^{l}$

optimization approach only use one parameter to maximize anonymity such that the performance is limited comparing with joint selection. In addition, the route selection has better performance than relay configuration. The reasons is because the uniformly route selection split the traffic to covert relay such that the packet-loss rate is reduced.

2.5 Conclusion

In this chapter, 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 joint selection can be derived and

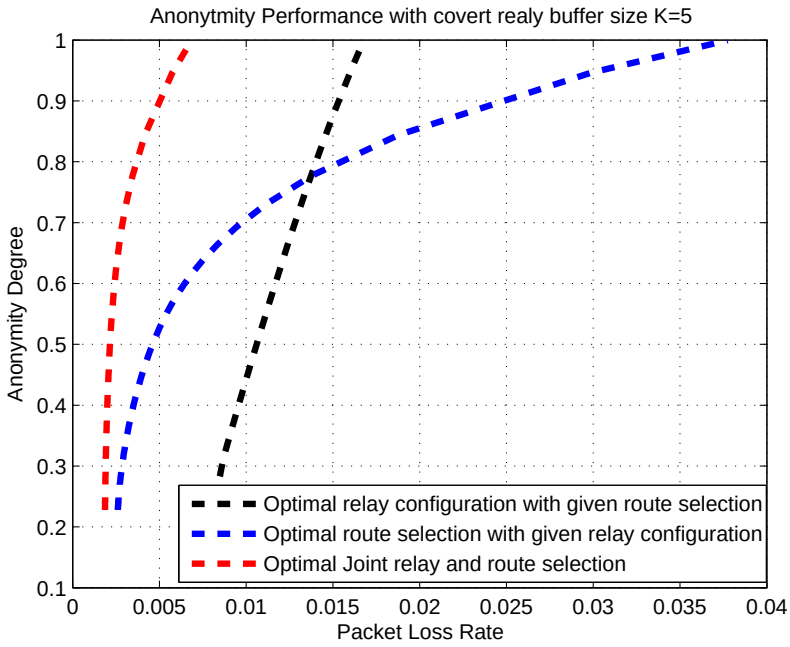


Figure 2.4: Comparison between joint relay-route selection, route selection alone and relay configuration alone. The joint relay-route selection has the best performance among the others

solved by using our rate-distorting framework. Our framework guarantees efficient computation of the global optimum. In the next section, we will study the performance of our approach in route selection and relay configuration alone.

Chapter 3

ROUTE SELECTION I: A CENTRALIZED ENTITY NETWORK

In previous chapter, we have investigated how to jointly assign relay and route selection for anonymous networks. However, not all networks are able to have joint selection. In fact, some networks where relays configuration might have been given only can maximize anonymity via route selection. Therefore, we consider a special case in here that a centralized entity of all source-destination pairs select routes jointly while relay configurations are given.

Then, the anonymity and network performance only depend on route selection schedule. 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 $Q(\mathbf{R}, \mathbf{C}|\mathbf{Z}) = Q(\mathbf{R}|\mathbf{Z}, \mathbf{C} = c_1, c_2, ..., c_\alpha)$. Take Figure 2.3 as example, two source choose destinations jointly with given relay configuration to maximize anonymity while considering packet-loss rate of network performance. In next section, we will show how to formulate the convex optimization to achieve our goal.

3.1 Problem Formulation for A Centralized Entity of Source-Destination Pairs

By considering a multipath wireless network with given relay configuration, the anonymity degree under this scenario is

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

Since our goal is to maximize anonymity while maintaining network performance, then the optimal joint route selection problems can be formulated as

$$\min \left[H(\mathbf{Z}|\mathbf{C}) - H(\mathbf{Z}|\mathbf{R}, \mathbf{C}) \right] \text{ s.t. } d(\mathbf{Z}, \mathbf{R}, \mathbf{C}) \leq D \tag{3.1}$$

The (3.1) is the special case from (2.2) given $\mathbf{C}$. Hence, a convex optimization with Lagrange multiplier is written as

$$J = \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} P(\mathbf{Z},\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) \log \frac{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}{\sum_{\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R},\mathbf{C}|\mathbf{Z})},$$
$$+ \sum_{\mathbf{Z},\mathbf{R},\mathbf{C}} v(\mathbf{Z},\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) - sE_{\mathbf{Z}}(Q(\mathbf{R},\mathbf{C}|\mathbf{Z}))$$

By solving Lagrange multiplier function from (3.2), the globally optimal solution $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ can be obtained. However, for a finite buffer size of covert relays, packet-loss will depend on the incoming traffic rate in covert relays as well as route selection among multi-paths. Then, Lagrange multiplier function is non-linear in $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$. To obtain optimal $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$, one can apply Newton's method to solve non-linear equation iteratively.

When buffer size in covert relays is increasing, the dropping rate is decreasing. Therefore, the average packet-loss rate will approach to the rate by link-quality. Namely, the packet-loss will be dominated by link-quality. Therefore, the equation can be viewed as linear equation to obtain Q^* approximately. For example, if we have infinite buffer in covert relays, the $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is

3.2 Proposed Route Selection for A Centralized Entity

Therefore, 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.

3.2.1 Packet-loss from link-quality alone

By solving $\frac{\partial J}{\partial Q_{\mathbf{F}}(\mathbf{F}|\mathbf{Z},\mathbf{R})} = 0$, where J is as given in equation (2.5). We can obtain

$$Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) = \frac{q(\mathbf{R}|\mathbf{C}) \exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}{\sum_{\mathbf{R}} q(\mathbf{R}|\mathbf{C}) \exp[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)]}, \tag{3.2}$$
$$\text{where } q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}),$$
$$L_q(\mathrm{R}_i) = \sum_{j:\{G_j \cap \mathrm{R}_i\} \cup \{\mathrm{D}_j \cap \mathrm{R}_i\}} L_{i,j}(\mathrm{R}_i) * (\eta_i / \sum_{i'=1}^{N} \eta_{i'}).$$

As mentioned earlier, $s < 0$ is a parameter that determines the trade-off between anonymity and packet-loss rate. The choice of parameter s determines the trade-off between anonymity and packet-loss. Decreasing the value of s increases the value of the achieved anonymity, at the cost of higher packet-loss rate [6].

3.2.2 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 derived via Lagrange multipliers function.

$$f(\mathbf{R}) : Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) - \frac{q(\mathbf{R}|\mathbf{C})\exp[sT_{\mathbf{Z}}]}{\sum\limits_{\mathbf{R}} q(\mathbf{R}|\mathbf{C})\exp[sT_{\mathbf{Z}})]} = 0, \tag{3.3}$$

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

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

Since equation (3.3) 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

$$\mathbf{Q}^{(n+1)} = \mathbf{Q}^{(n)} - \big[\,[\nabla \underline{f}]^{-1}\underline{f}\,]|_{\mathbf{Q}^{(n)}}, \tag{3.4}$$

where $\mathbf{Q}^{(n)}$=$[Q^{(n)}(\mathbf{a}_1|\mathbf{b},\mathbf{C}), Q^{(n)}(\mathbf{a}_2|\mathbf{b},\mathbf{C}), ..., Q^{(n)}(\mathbf{a}_l|\mathbf{b},\mathbf{C})]^T$ and $\underline{f} = [f(\mathbf{a}_1), f(\mathbf{a}_2), ..., f(\mathbf{a}_l)]^T$ are the column vector of (3.3) for all flows set $\mathbf{a}_i \subset \mathbf{R}$ given source-destination pair $\mathbf{b} \subset \mathbf{Z}$.

3.2.3 Iteration Procedure

In previous section we have showed how to obtain the optimal $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$. Here we gave the iteration procedure in detail to indicate how to calculate the optimal route selections for centralized entity. An iterative algorithm for obtaining $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is list in next page.

3.3 Result and Performance Discussion

The simulated network topology consists of two sources, two destinations, and four relays with buffer size $K = 5$ (Figure 2.3). Each source has transmission rate 5. Covert relays are assumed to

Iteration Procedure : Algorithm for route selection

Input: Resolution Δ, Slope value s, Iteration number, I Source-destination pair, $\mathrm{Z_i} = \beta$

All routes for this source-destination pair β : $\underline{\alpha_1}, \underline{\alpha_2}...\underline{\alpha_l}$

Output: Probability for each route $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta),\ j = 1, 2..l$

$q(\underline{\alpha_j}) \leftarrow \frac{1}{l}$, for $j = 1, 2..l$

while $n < I$

  if $\delta = \infty$

    $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta) \leftarrow$ eq. (3.2) 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 eq. (3.4)

      $m_j \leftarrow |m_j - Q_{\mathrm{Z_i}}|$ 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 $\{Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta)\}_{j=1}^{l}$

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 2.3 was chosen uniformly at random from the interval [10^{-5}, 10^{-2}].

The choice of parameter s determines the trade-off between anonymity and packet-loss. Decreasing the value of s increases the value of the achieved anonymity, at the cost of higher packet-loss rate. Figure 3.1 shows the anonymity of the optimal route selection obtained from (7) for each given relay configuration and packet-loss constraint. By considering various covert relay configurations as (C_1,C_3), (C_1,C_2,C_3), (C_1,C_2,C_3,C_4) in Figure 2.3, the results shows different 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.

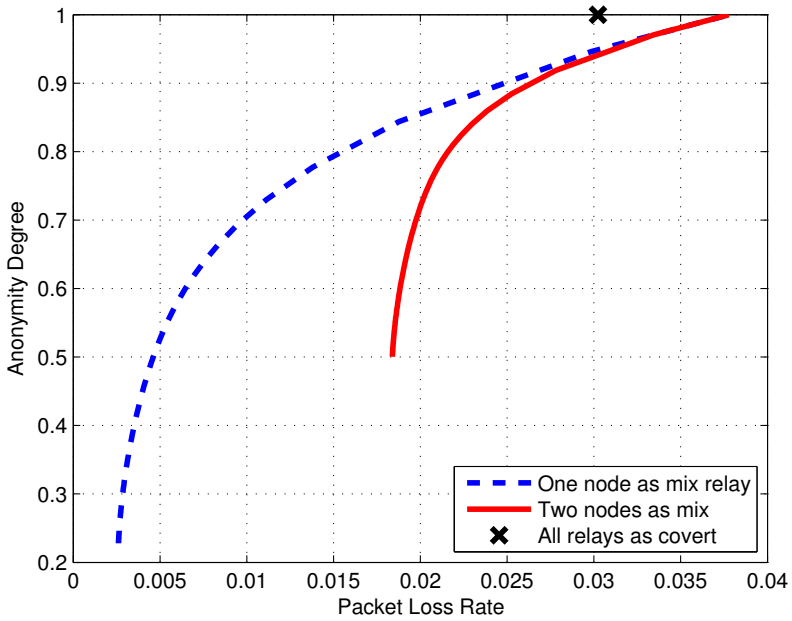


Figure 3.1: Results for different relay configuration with buffer size K=5 for covert relays. Varying route selection among multipath result in different anonymity and packet-loss.

3.4 Conclusion

In this chapter, we studied the problem of maximizing anonymity subject to packet-loss constraints through joint route selection of a centralized entity in multipath anonymous networks. For an information theoretic anonymity metric, we formulated joint relay selection as a convex optimization problem. We showed that the problem of optimal route selection for given relay configuration 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 sourcedestination pairs, and compare our approach with independent optimization of relay configuration and route selection.

Chapter 4

ROUTE SELECTION II: A DECENTRALIZED ENTITY FOR INDEPENDENT NETWORKS

For a multipath wireless network, we have investigated joint relay and route selection and a centralize route selection given relay configuration. In addition to those types of networks, we consider a de-centralized entity for individual route selection. To further address the optimal design of decentralized networks, we consider two different properties of networks. The first is independent source-destination pairs while the other is dependent source-destination pair.

In this chapter, we consider a independent sources and destinations property to maximize anonymity. Especially, most computer networks are designed to be de-centralized for scalar purpose such that each source choose route individually. Since the relay configuration is given in advanced, the anonymity and network performance depend on individual route selection schedule. In here, we will show that for independent source, the individual route selection will only depend on itself condition rather than others. Then, we propose individual route selection 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

4.1 Problem Formulation for A Decentralized Entity of Independent Network

Consider the example of Figure (4.1) 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. Each source chooses destination independently and the link quality is shown in table (4.1). When relay configuration is given for de-centralized networks, each source select route independently to maximize anonymity while maintaining network performance. For example, R_B is setting as cover, the source S_1 can choose the route pass through

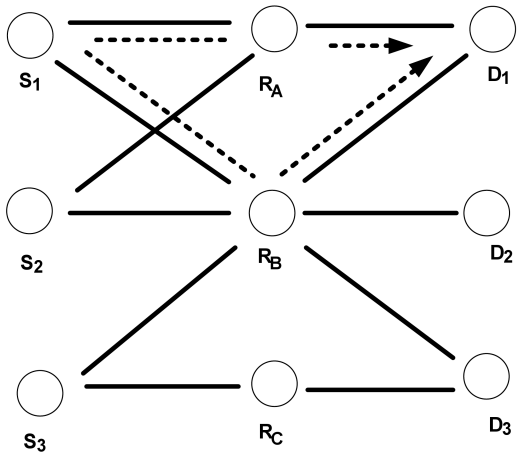


Figure 4.1: Example of a wireless network topology with three sources, three relay nodes (which can be designated as visible or covert by the network), and three destinations. The two possible paths between source S_1 and destination D_1 are shown as dotted lines

relay R_A or R_B to have different anonymity and network performance. The problem considered in this section 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.

Table 4.1: 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)	2×10^{-2}	(S_1, D_1, R_B)	3×10^{-2}
(S_2, D_2, R_B)	2.2×10^{-2}	(S_3, D_2, R_B)	3.7×10^{-2}
(S_1, D_2, R_B)	7×10^{-2}	(S_2, D_2, R_A)	8×10^{-3}
(S_3, D_3, R_B)	2×10^{-1}	(S_1, D_3, R_B)	10^{-1}
(S_2, D_3, R_B)	3×10^{-2}	(S_3, D_3, R_C)	10^{-3}

4.2 Proposed Route Selection for Independent Source-Destination Pair

By giving the relay configuration, the rate distortion function is rewritten as

$$\min_{E_{\mathbf{Z}} \leq D} [H(\mathbf{Z}|\mathbf{C}) - H(\mathbf{Z}|\mathbf{R}, \mathbf{C})]. \tag{4.1}$$

Here, $E_{\mathbf{Z}}$ denotes distortion function defined as the system average packet-loss rate arising from the route vector $\mathbf{R}$ with source-destination pairs $\mathbf{Z}$ given relay configuration $\mathbf{C}$. Later we will show how to split the route selection into each source selection independently.

Lemma 4.1. *Let* $R(D) = \min_{E_{\mathbf{Z}} \leq D} [H(\mathbf{Z}|\mathbf{C}) - H(\mathbf{Z}|\mathbf{R}, \mathbf{C})]$. *Then,*

$$R(D) = \sum_{i=1}^{N} \min_{E_{\mathbf{Z}} \leq D} [H(\mathrm{Z_i}|\mathbf{C}) - H(\mathrm{Z_i}|\mathrm{R_i}, \mathbf{C})]. \tag{4.2}$$

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

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

Lemma (4.1) 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 as E_{Z_i} where $E_{\mathbf{Z}} = \sum_{i=1}^{N} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) E_{\mathrm{Z_i}}(Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}))$. Let $Q(\mathrm{Z_i} = \underline{\alpha}|\mathrm{Z_i} = \beta, \mathbf{C})$ denote the probability that source i selects route $\underline{\alpha}$ given the corresponding destination is β.

Then, we have

$$R(D) = \sum_{i=1}^{N} \min_{E_{\mathbf{Z}} \le D} [H(\mathrm{Z_i}) - H(\mathrm{Z_i}|\mathrm{R_i}, \mathbf{C})] = \sum_{i=1}^{N} \min_{Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) \in \mathbf{Q}} [H(\mathrm{R_i}|\mathbf{C}) - H(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})],$$

$$= \sum_{i=1}^{N} \min_{Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) \in \mathbf{Q}} \left[\sum_{\mathrm{R_i}} \sum_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \log \frac{Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})}{\sum\limits_{\mathrm{Z_i}} \mathrm{P}(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{Z_i}|\mathrm{R_i}, \mathbf{C})} \right],$$

where $\mathbf{Q} = \left\{ \sum\limits_{i=1}^{N} \sum\limits_{\mathrm{R_i}} \sum\limits_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) E_{\mathrm{Z_i}}(Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \le D,\ Q(\mathrm{R_i}|\mathrm{Z_i}) \ge 0, \right.$

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

$$J = \sum_{i=1}^{N} \left[\sum_{\mathrm{R_i}} \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(k|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \log \frac{Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})}{\sum\limits_{\mathrm{Z_i}} P\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})} + \sum_{\mathrm{Z_i}} v(\mathrm{Z_i}) \sum_{\mathrm{R_i}} Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \ + \right.$$

$$\left. - s \sum_{\mathrm{R_i}} \sum_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) E_{\mathrm{Z_i}}(Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \right]. \tag{4.3}$$

We consider the optimization of equation (6.2) under two different scenarios. In the first scenario, covert relays do not drop packets by considering infinite buffer size. In the second scenario, we incorporate packet-dropping by covert relays. In what follows, we discuss how to obtain $Q(\mathrm{R_i}|\mathrm{Z_i})$ for these two scenarios.

4.2.1 Route Selection Without Packet-Dropping by Covert Relays

In this section, we consider the optimization problem of equation (6.2) assuming link-quality is the only source for packet loss. That is, we set the time constraint of the covert relays to $= \infty$. In this case, the packet-loss rate function in equation (2.3) reduces to $E_{\mathrm{Z_i}}(Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})) = L_q(\mathrm{R_i})$. From [6,10], the roots of $\frac{\partial J}{\partial Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})} = 0$ are

$$Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) = \frac{q(\mathrm{R_i}|\mathbf{C}) \exp[sL_q(\mathrm{R_i})]}{\sum_j \ q(\mathrm{R_i}|\mathbf{C}) \exp[sL_q(\mathrm{R_i})]}, \tag{4.4}$$

where $q(\mathrm{R_i}|\mathbf{C}) = \sum_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ and $L_q(\mathrm{R_i})$ represents the route's link-quality where source $\mathrm{Z_i}$ choose route $\mathrm{R_i}$. As Equation (4.4) shows, $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ 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 $R(D)$.

4.2.2 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 K time units [5]. This increase packet-loss rate. However, the dropping rate in here is also dependent on the route selection, which is denoted by $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ as described in [8]. Therefore, by solving $\frac{\partial J}{\partial Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})} = 0$, where J is as given in equation (6.2), we obtain

$$f_j : Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) - \frac{q(\mathrm{R_i}|\mathbf{C}) \exp\left[sT_{\mathrm{Z_i}}\right]}{\sum_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C}) \exp\left[sT_{\mathrm{Z_i}}\right]} = 0, \tag{4.5}$$

where

$$T_{\mathrm{Z_i}} = \frac{\partial E_{Z_i}}{\partial Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})} * (1/P(\mathrm{Z_i})), q(\mathrm{R_i}|\mathbf{C}) = \sum_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i})|\mathrm{Z_i}, \mathbf{C})$$

. In equation (4.5), $\exp[sT_{\mathrm{Z_i}}]$ contains the term $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ that needs to be estimated. Hence the root cannot be evaluated directly from $Q(j|k, \mathbf{C}) = \frac{q(j|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]}{\sum_j q(j|\mathbf{C}) exp[sT_{\mathrm{Z_i}}]}$, since it is a nonlinear polynomial. We present Newton method for obtaining $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ as follows.

Newton-Raphson Method

Under the Newton-Raphson method for solving nonlinear systems of equations [9], we have

$$\mathbf{Q}^{n+1} = \mathbf{Q}^n - \left[\, [\nabla \mathbf{F}]^{-1} \mathbf{F} \,\right]|_{\mathbf{Q}^n} \tag{4.6}$$

where $\mathbf{Q}^n = [Q^n(\underline{\alpha_1}|\beta, \mathbf{C}), Q^n_{\mathrm{Z_i}}(\underline{\alpha_2}|\beta, \mathbf{C}), ...Q^n(\underline{\alpha_l}|\beta, \mathbf{C})]^T$, and $\mathbf{F} = [f_1, f_2, ...f_l]^T$, for f_j in equation (4.5). $Q^n(\underline{\alpha_j}|\beta, \mathbf{C})$ represents the value at n^{th} iteration of each route $\underline{\alpha_j}$, given the destination β via Newtons-Raphson method. After assigning arbitrary probability values for $\mathbf{Q}^0$ as initial vectors, all possible initial vectors converge to the same result by recursive equation (5.5).

4.2.3 Iteration Procedure

In previous section we have showed how to obtain the optimal $Q^*(\mathbf{R}|\mathbf{Z}, \mathbf{C})$. Here we gave the iteration procedure in detail to indicate how to calculate the optimal route selections for centralized entity. An iterative algorithm for obtaining $Q^*(\mathbf{R}|\mathbf{Z}, \mathbf{C})$ is given as follows.

Iteration Procedure : Algorithm for route selection

Input: Resolution, Δ, Buffer size, δ, Slope value, s Iteration number, I

Source-destination pair, $\mathrm{Z_i} = \beta$

All routes for this source-destination pair β : $\underline{\alpha_1}, \underline{\alpha_2}...\underline{\alpha_l}$

Output: Probability for each route $Q_{\mathrm{Z_i}}(\underline{\alpha_j}|\beta, \mathbf{C}),\ j = 1, 2..l$

$q(\underline{\alpha_j}|\mathbf{C}) \leftarrow \frac{1}{l}$, for $j = 1, 2..l$

while $n < I$

  if $\delta = \infty$

    $Q(\underline{\alpha_j}|\beta, \mathbf{C}) \leftarrow$ eq. (4.4) for $j = 1, 2..l$

  else

    while $m_j > \Delta$ for $j = 1, 2...l$

      $m_j \leftarrow Q(\underline{\alpha_j}|\beta, \mathbf{C})$

      Compute $Q(\underline{\alpha_j}|\beta, \mathbf{C}) \leftarrow$ by eq. (4.6)

      $m_j \leftarrow |m_j - Q(\alpha_j|\beta, \mathbf{C})$ for $j = 1, ..l$

    end

  end

  $q(\underline{\alpha_j}|\mathbf{C}) \leftarrow \sum_k \mathrm{P(k|C)} Q\underline{\alpha_j}|k, \mathbf{C})$ for $j = 1, 2..l$

end while

return $\{Q(\underline{\alpha_j}|\beta, \mathbf{C})\}_{j=1}^{l}$

4.3 Result and Performance Discussion

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 4.1, with node B acting as a covert relay without time constraint, i.e., $\delta = \infty$, and link-quality in Table (4.1), we compare our proposed method with all possible probability distributions $\mathbf{Q}$ in Figure 4.2. The figure shows that our approach is optimal for this network topology.

Figure 4.3 shows the performance with different covert relay configurations, including (R_B),

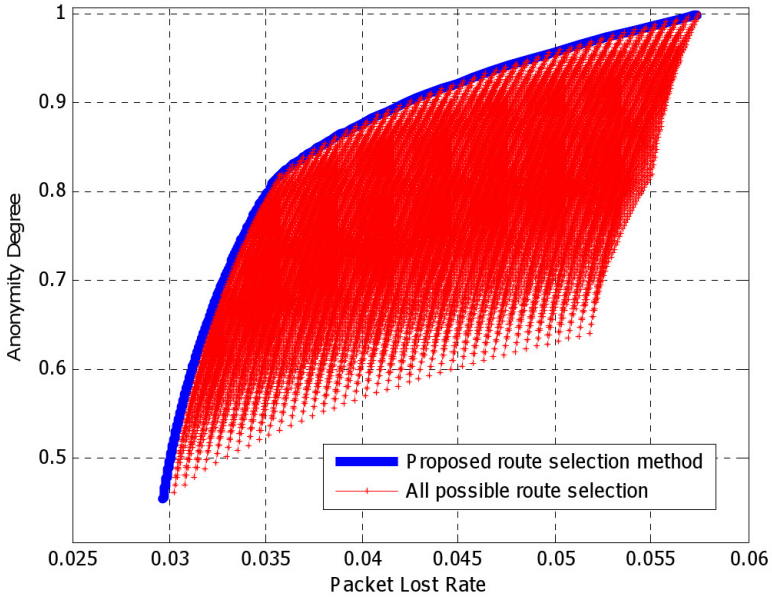


Figure 4.2: Comparison between manual assignment and the proposed rate-distortion method. We compare the proposed method performance with manual assign method which run all possible probability for each route.

(R_A, R_B), and (R_A, R_B, R_C) from Figure 4.1. When all relays are covert relays, there is no trade-off between packet-loss rate and anonymity. Figure 4.3 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.

Figure 4.4 compares the Newton-Raphson algorithms for solving (4.5), with the buffer size K set to 5, 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 there is a trade-off between anonymity and performance.

4.4 Conculsion

In this chapter, 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

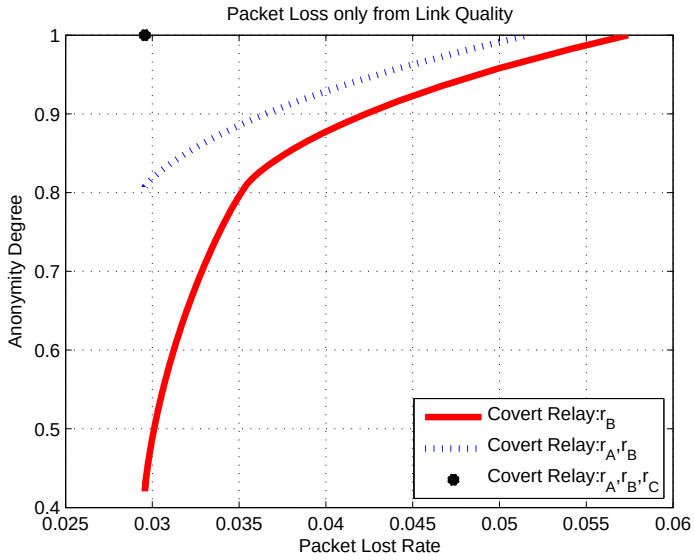


Figure 4.3: The anonymity degree with packet-loss rate caused only by link-quality under different relay configurations.

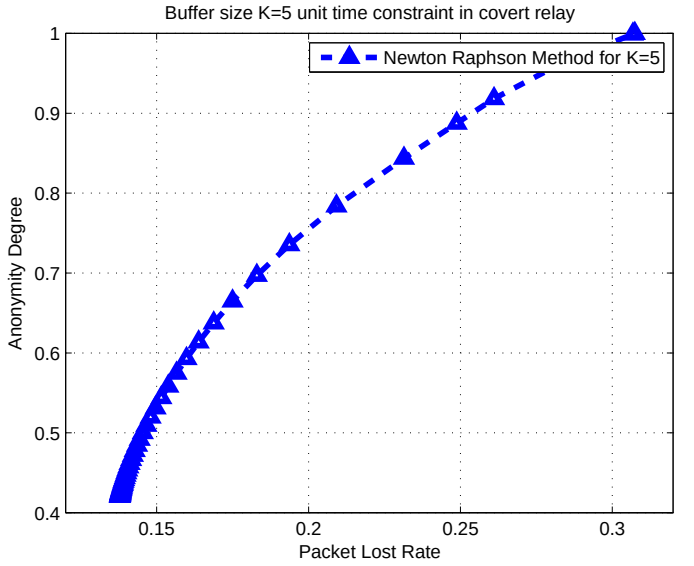


Figure 4.4: The anonymity degree with packet-loss rate caused by link-quality and covert relays dropping by setting R_B as cover relay.

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 4.1, the formulation is applicable to wireless networks with multiple intermediate hops with prespecified covert relays.

After using rate-distortion function to optimize route selection, we have shown that our proposed method can find the optimum probability for each route to get maximum anonymity by considering the packet-loss rate criteria. In addition, our proposed method to select routes only depend on itself transmission schedule. For each source,namely user, they choose relays only depend on itself source-destination pair $\mathrm{Z_i}$. Thus, our method offer a flexible way to select relays respect to the optimum probability $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ in order to have maximum anonymity degree without exceeding certain amount packet-loss rate.

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.

Chapter 5

RELAY SELECTION III: A DECENTRALIZED ENTITY FOR DEPENDENT NETWORKS

We have consider independent network in previous chapter. In here we consider 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 sources 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.

Therefore, each source require the knowledge of the other source's action to minimize the penalty from the other sources which compromise itself possible source and destination pairs. However, each source may have incomplete knowledge of which destination and routes other sources choose due to packet encryption and radio range. Therefore, we consider three different cases depending on each sources knowledge of other sources. In Case I, we assume that each source, i, has complete information about the transmission behaviors of other sources and knows the values of Z_j and R_j for all $j \neq= i$ when choosing its routes R_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 Z_j and R_j for all $j \neq= i$, but does not know the exact values of any of those parameters when choosing its routes R_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.

In each case, we show how to choose routes among multiple paths to maximize anonymity under packet-loss constraint by considering the optimization as a rate-distortion problem.

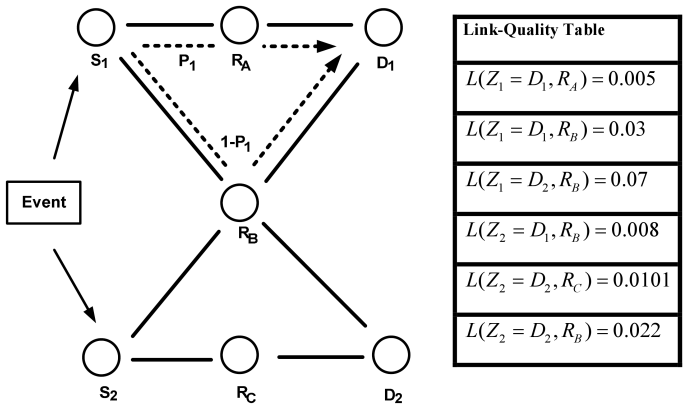


Figure 5.1: Example of a wireless network of dependent source-destination pairs with Link-Quality Table.

5.1 Problem Formulation for A Decentralized Entity of Dependent Network

Given relay configuration, the problem of maximizing the anonymity is equivalent to reduced the anonymity loss as following

$$R(D) = \min_{d(\mathbf{Z},\mathbf{R},\mathbf{C}) \leq D} [H(\mathbf{Z}|\mathbf{C}) - H(\mathbf{Z}|\mathbf{R},\mathbf{C})], \tag{5.1}$$

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

Since the Z_i's are dependent, the optimization problem (5.1) for each source 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 as mentioned earlier.

5.2 Case I: Full Information Regarding Other Sources

In this section, we assume each source has full information regarding others. Therefore, the convex optimization problem is the same as eq(5.1). By solving the convex optimization problem, we have

$$R(D) = \min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\leq D}[H(\mathbf{R}|\mathbf{C}) - H(\mathbf{R}|\mathbf{Z},\mathbf{C})] \tag{5.2}$$
$$= \min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\leq D}\Big[\sum_{\mathbf{R}}\sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C})\log\frac{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}{\sum_{\mathbf{R}}\mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}\Big].$$

The problem formulation in (5.2) is the same as the centralized entity given relay configuration. Therefore, the optimal $Q^*(\mathbf{R}|\mathbf{Z},\mathbf{C})$ are

No packet dropping by covert relays

If packet loss occurs only due to link-quality, the average packet-loss rate $d(\mathbf{Z},\mathbf{R},\mathbf{C})$ is independent of $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$. We can obtain $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ by solving for the root of $\frac{\partial J}{\partial Q(\mathbf{R}|\mathbf{Z},\mathbf{C})} = 0$, which, as shown in [6, 8], can be written as

$$Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) = \frac{q(\mathbf{R}|\mathbf{C})\exp\big[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)\big]}{\sum_{\mathbf{R}} q(\mathbf{R}|\mathbf{C})\exp\big[s\big(\sum_{i=1}^{N} L_q(\mathrm{R_i})\big)\big]}, \tag{5.3}$$

where $q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}}\mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$, $\mathrm{Z_i}\subset\mathbf{Z}$, and $\mathrm{R_i}\subset\mathbf{R}$.

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

$$f_j : Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) - \frac{q(\mathbf{R}|\mathbf{C})\exp\big[sT_{\mathbf{Z}}\big]}{\sum_{\mathbf{R}} q(\mathbf{R}|\mathbf{C})\exp\big[sT_{\mathbf{Z}}\big]}, \tag{5.4}$$
$$\text{where } q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}}\mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}),\ \ T_{\mathbf{Z}} = \frac{\partial E_{\mathbf{Z}}}{\partial Q(\mathbf{R}|\mathbf{Z},\mathbf{C})} * (1/P(\mathbf{Z})).$$

Since, in equation (5.4), $T_{\mathbf{Z}}$ contains the root $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$, we use Newton's method to solve (5.4) numerically by the iteration procedure

$$\mathbf{Q}^{n+1} = \mathbf{Q}^n - \big[\,[\nabla\mathbf{F}]^{-1}\mathbf{F}\,\big]|_{\mathbf{Q}^n}, \tag{5.5}$$

where $\mathbf{Q}^n = [Q^n(\mathbf{a}_1|\mathbf{b},\mathbf{C}), Q^n(\mathbf{a}_2|\mathbf{b},\mathbf{C}), ..., Q^n(\mathbf{a}_l|\mathbf{b},\mathbf{C})]^T$ are the values at the n^{th} iteration, and $\mathbf{F} = [f_1, f_2, ..., f_l]^T$, for f_j in (5.4).

Once we have optimal $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$, the individual route selection can be obtained. We first denote $P_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})$ as the joint probability distribution of all source-destination pairs and $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ to be the joint probability distribution of the route selections $\mathbf{R}$ conditioned on the source-destination pairs $\mathbf{Z}$. Then, using chain rule, $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ can be written as

$$Q(\mathbf{R}|\mathbf{Z},\mathbf{C}) = Q(\mathrm{R}_1|\mathbf{Z},\mathbf{C}) \prod_{i=2}^{N} Q(\mathrm{R_i}|\mathbf{Z},\mathrm{R}_1,\cdots,\mathrm{R}_{i-1},\mathbf{C}), \tag{5.6}$$

where $Q(\mathrm{R_i}|\cdot,\mathbf{C})$ is the conditional probability of route selection for source i, which represents choosing route $\mathrm{R_i}$ depending on the sources' current destinations, $\mathbf{Z}$, and routes already chosen $\mathrm{R}_1,\ldots,\mathrm{R}_{i-1}$. Without loss of generality, we assume that the sources are indexed such that source i is the i^{th} source to select a route $\mathrm{R_i}$. Hence, the first source 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. In addition, we define $F_{\mathbf{Z}}$ as the packet-loss rate of the route chosen by source i, calculated by $F_{\mathbf{Z}} = 1-(1-P_c)(1-P_l)$, where P_c is the dropping rate of covert relay and P_l is the packet-loss rate caused by link-quality (note that $F_{\mathbf{Z}}$ is not a linear function since P_c is a function of $Q(\mathbf{R}|\mathbf{Z},\mathbf{C})$ [6], as will be discussed in more detail later).

5.3 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. By considering this scenario, the uncertainty each source after traffic analysis is denoted as $H(\mathrm{Z_i}|\mathbf{R},\mathbf{C})$. Therefore, for source i, the maximum anonymity under partial information is achieved when $H(\mathrm{Z_i}|\mathbf{R},\mathbf{C})$ is as close to $H(\mathrm{Z_i})$ as possible. The route selection

probability for Z_i is obtained by

$$\begin{aligned}
&\min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\leq D} [H(\mathrm{Z_i}|\mathbf{C}) - H(\mathrm{Z_i}|\mathbf{R},\mathbf{C})\,] \qquad (5.7)\\
=&\min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\leq D} [H(\mathbf{R}|\mathbf{C}) - H(\mathbf{R}|\mathrm{Z_i},\mathbf{C})\,]\\
=&\min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\leq D} \Bigg[\sum_{\mathbf{R}} \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C})\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})\times\\
&\log \frac{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})Q(\mathbf{R}^{\mathrm{Z_i}}|\mathrm{Z_i},\mathbf{C})}{\sum\limits_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C})\mathrm{P}(\mathbf{R}^{/\mathrm{Z_i}}|\mathrm{Z_i},\mathbf{C})Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})} \Bigg],
\end{aligned}$$

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

Note that the solution to (5.7) results in the maximum-uncertainty route selection for source i individually, but does not achieve the global optimum of (5.1). $Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})$ can be obtained by solving (5.7) using Lagrange multipliers with the conditions $d(\mathbf{Z},\mathbf{R},\mathbf{C}) \leq D$ and $\sum_{\mathrm{R_i}} Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) = 1$. As in Case I, we solve for $Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})$ by considering two scenarios, based on whether the covert relays drop packets.

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 :

$$Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) = \frac{\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sL_q(\mathrm{R_i})]}{\sum\limits_{\mathrm{R_i}} \big(\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sL_q(\mathrm{R_i})] \big)}, \qquad (5.8)$$

where $q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})[\prod_{i=1}^{N} Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})]$.

Packet dropping by covert relays

When packet loss is caused by both link-quality and covert relays' dropping, then $Q(\mathrm{R_i}|\mathrm{R_i},\mathbf{C})$ can be obtained using the Newton's method in (5.5) numerically with the following function f_j, defined

as

$$f_j : Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) - \frac{\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})}\exp[sT_{\mathrm{Z_i}}]}{\sum\limits_{\mathrm{R_i}}\big(\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})}\exp[sT_{\mathrm{Z_i}}]\big)} = 0, \tag{5.9}$$

where $\mathbf{a}_j \in \mathbf{R}^{\mathrm{Z_i}}, b \in \mathrm{Z_i},\ q_{\mathbf{Z}}(\mathbf{R})$ is as in(5.8), and

$$T_{\mathrm{Z_i}} = \frac{\partial E_{Z_i}}{\partial Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})} * (1/P(\mathrm{Z_i})), q(\mathbf{R}|\mathbf{C}) = \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})[\prod_{i=1}^{N} Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})].$$

5.4 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 (5.1), source i instead selects the routes $\mathrm{R_i}$ such that $H(\mathrm{Z_i}|\mathrm{R_i},\mathbf{C})$ is as close to $H(\mathrm{Z_i}|\mathbf{C})$ as possible. The route selection probability $Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})$ can be obtained by:

$$\begin{aligned}
&\min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\le D} [H(\mathrm{Z_i}|\mathbf{C}) - H(\mathrm{Z_i}|\mathrm{R_i},\mathbf{C})\,] \qquad (5.10)\\
&= \min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\le D} [H(\mathrm{Z_i}|\mathbf{C}) - H(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})\,]\\
&= \min_{d(\mathbf{Z},\mathbf{R},\mathbf{C})\le D} \Bigg[\sum_{\mathrm{R_i}}\sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})\\
&\qquad\qquad \times \log \frac{Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})}{\sum\limits_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})} \Bigg].
\end{aligned}$$

Equation (5.10) results in the maximum uncertainty $H(\mathrm{Z_i}|\mathrm{R_i},\mathbf{C})$, conditioned on the route selection $\mathrm{R_i}$ for source i, but does not achieve the global optimum in (5.1).

$Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})$ can be obtained from (5.10) using Lagrange multipliers with the conditions $d(\mathbf{Z},\mathbf{R},\mathbf{C}) \le D$ and $\sum_{\mathrm{R_i}} Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) = 1$. As in Case I and Case II, we consider the two different scenarios of covert relay behavior below.

No packet dropping by covert relays

When packet loss occurs due to link-quality alone, we have, similar to Cases I and II [6, 8],

$$Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) = \frac{q(\mathrm{R_i}|\mathbf{C})\exp[sL_q(\mathrm{R_i})]}{\sum_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C})\exp[sL_q(\mathrm{R_i})]}, \tag{5.11}$$

where $q(\mathrm{R_i}|\mathbf{C}) = \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C})$.

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 (5.5) to solve $f_j = 0$, defined by

$$f_j : Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) - \frac{q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]}{\sum_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]} = 0, \tag{5.12}$$

$$\text{where } T_{\mathrm{Z_i}} = \frac{\partial E_{Z_i}}{\partial Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})} * (1/P(\mathrm{Z_i})), q(\mathrm{R_i}|\mathbf{C}) = \sum_{\mathrm{Z_i}} P(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$$

5.4.1 Iteration Algorithm

An algorithm to compute the flow allocation $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ is given as **Flow Allocation Procedure** below.

Due to all initial values of $q(\mathbf{R}|\mathbf{C})$ or $q(\mathrm{R_i}|\mathbf{C})$ 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. 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.

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.

5.5 Independent Source for Three Case

If sources are independent, each source will reach the same performance. That is because the other sources' information do not compromise the source itself status. We here will show how each case approach will reach the same result

Route Selection Procedure: Algorithm for Route Selection

Input: Buffer size δ, Slope value s, Iteration number I

Output: Probability for each route : Q for R_1, R_2,...,R_N

Initialize: **If** Case I or Case II: $q(\mathbf{R}|\mathbf{C}) \leftarrow \frac{1}{\|\mathbf{R}\|}$

If Case III: $q(\mathrm{R_i}|\mathbf{C}) \leftarrow \frac{1}{\|\mathrm{R_i}\|}$ for i=1..N

while $n < I$

If $\delta = \infty$

If Case I: $Q(\mathrm{R_i}|\mathbf{Z}, \mathrm{R}_1, ..\mathrm{R}_{i-1}, \mathbf{C}) \leftarrow$ (5.3) for i=1..N

If Case II: $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \leftarrow$ (5.8) for i=1..N

If Case III: $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \leftarrow$ (5.11) for i=1..N

else

If Case I: $Q(\mathrm{R_i}|\mathbf{Z}, \mathrm{R}_1, ..\mathrm{R}_{i-1}, \mathbf{C}) \leftarrow$ (5.4) for i=1..N

If Case II: $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \leftarrow$ (5.9) for i=1..N

If Case III: $Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C}) \leftarrow$ (5.12) for i=1..N

end

If Case I: $q(\mathbf{R}|\mathbf{C}) \leftarrow \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C}) Q(\mathbf{R}|\mathbf{Z}, \mathbf{C})$

If Case II: $q(\mathbf{R}|\mathbf{C}) \leftarrow \sum_{\mathbf{Z}} \mathrm{P}_{\mathbf{Z}}(\mathbf{Z}|\mathbf{C})[\prod_{i=1}^{N} Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})]$

If Case III: $q(\mathrm{R_i}|\mathbf{C}) \leftarrow \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(\mathrm{Z_i}|\mathbf{C}) Q(\mathrm{R_i}|\mathrm{Z_i}, \mathbf{C})$ for i=1..N

n $\leftarrow$ n+1

end while and **return** Q

5.5.1 CASE I → CASE III

By considering independent source property, the source i route selection in case I is $Q(\mathrm{R_i}|\mathbf{C}, \mathbf{Z}, \mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R}_{i-1})$. Then, we can have

$$Q(\mathrm{R_i}|\mathbf{C}, \mathrm{Z}_1, \mathrm{Z}_2, ..., \mathrm{Z_N}, \mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R}_{i-1}) = \frac{Q(\mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R_N}|\mathbf{C}, \mathrm{Z}_1, \mathrm{Z}_2, ..., \mathrm{Z_N})}{\sum_{\mathrm{R_i}} Q(\mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R_N}|\mathbf{C}, \mathrm{Z}_1, \mathrm{Z}_2, \mathrm{Z_N})} \tag{5.13}$$

$$= \frac{\prod_{i=1}^{N} Q(\mathrm{R_i}|\mathbf{C}, \mathrm{Z_i})}{\sum_{\mathrm{R_i}} \prod_{i=1}^{N} Q(\mathrm{R_i}|\mathbf{C}, \mathrm{Z_i}}) = Q(\mathrm{R_i}|\mathbf{C}, \mathrm{Z_i})$$

Therefore, the case I can has the same performance as case III, when the source and destination

pairs are independent.

5.5.2 CASE II → CASE III

In case II, we have

$$f_j : Q(\mathrm{R_i}|\mathrm{Z_i},\mathbf{C}) - \frac{\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}]}{\sum\limits_{\mathrm{R_i}} \big(\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}] \big)} = 0, \tag{5.14}$$

By applying independent source, we have

$$\begin{aligned}
&\frac{\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}]}{\sum\limits_{\mathrm{R_i}} \big(\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathbf{R}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}] \big)}, \\
&= \frac{\Big[\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R}_1|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \times \prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R}_2|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \times \prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R_N}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \Big]}{\Big[\prod\limits_{\mathbf{R}^{/\mathrm{Z_i}}} q(\mathrm{R}_1|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \times \prod\limits_{\mathbf{R}^{/\mathrm{Z_i}}} q(\mathrm{R}_2|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \times \prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R_N}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \Big]} \\
&\times \frac{\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R_i}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}]}{\sum\limits_{\mathrm{R_i}} \big(\prod\limits_{\mathbf{R}^{/\mathrm{R_i}}} q(\mathrm{R_i}|\mathbf{C})^{\mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}] \big)} \\
&= \frac{q(\mathrm{R_i}|\mathbf{C})^{\sum\limits_{\mathbf{R}^{/\mathrm{R_i}}} \mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}]}{\sum\limits_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C})^{\sum\limits_{\mathbf{R}^{/\mathrm{R_i}}} \mathrm{P}(\mathbf{R}^{/\mathrm{R_i}}|\mathrm{Z_i},\mathbf{C})} \exp[sT_{\mathrm{Z_i}}]} \\
&= \frac{q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]}{\sum_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]}
\end{aligned} \tag{5.15}$$

Therefore, the optimal $Q(\mathrm{R_i}|\mathrm{Z_i})$ in case II is to solve the (5.16) which is the same as the case III. Therefore, for independent case, the case II is the same as case III.

$$Q(\mathrm{R_i}|\mathrm{Z_i}) - \frac{q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]}{\sum_{\mathrm{R_i}} q(\mathrm{R_i}|\mathbf{C}) \exp[sT_{\mathrm{Z_i}}]} = 0 \tag{5.16}$$

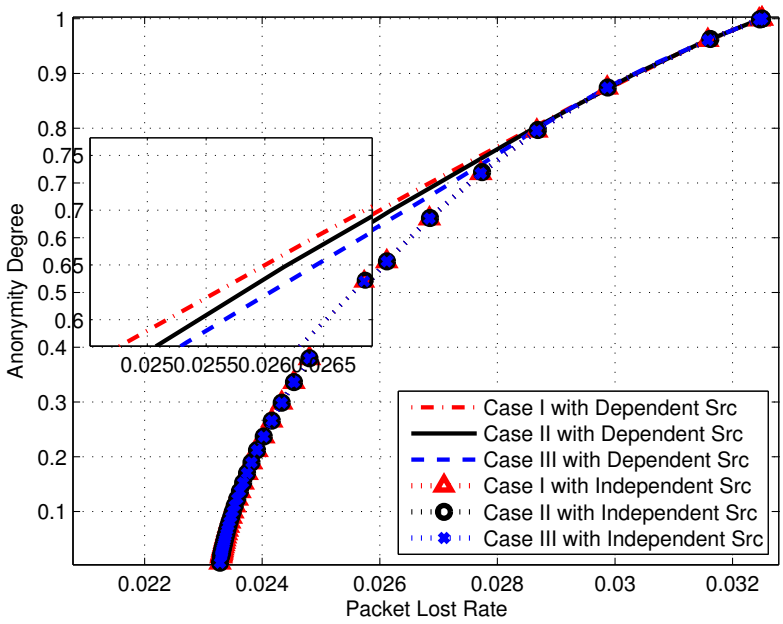


Figure 5.2: The anonymity degree achieved as a function of packet-loss constraint, when packet losses are due to link quality alone. Three cases of source information are considered under different levels of correlation between Z_1 and Z_2.

5.6 Result and Performance Discussion

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 5.1, 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 $Pr(Z_1 = D_1) = Pr(Z_1 = D_2) = \frac{1}{2}$; 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 with different values of P and constraints on the average packet loss rate.

In Figure 5.2, we set $\delta = \infty$ for R_B (i.e., no dropping of packets). When $P = 0.98$, the net-

work experiences the best performance under Case I, since source S_i knows the other source's full transmission behavior and can select $\mathrm{R_i}$ 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 $\mathrm{P_1}$ shown in Figure 5.1 for S_1 as the following. ($\mathrm{P_1}$ 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, $\mathrm{P_1}$ 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, $\mathrm{P_1}$ 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 $\mathrm{Z_2}$ and $\mathrm{Z_1}$ 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 5.1, we use the dropping rate function investigated in previous chapter with different transmission rates η and buffer sizeδ. 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 5.1(c), Case I outperforms Case II, due to the additional information available to the sources, while Case II outperforms Case III.

5.7 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-

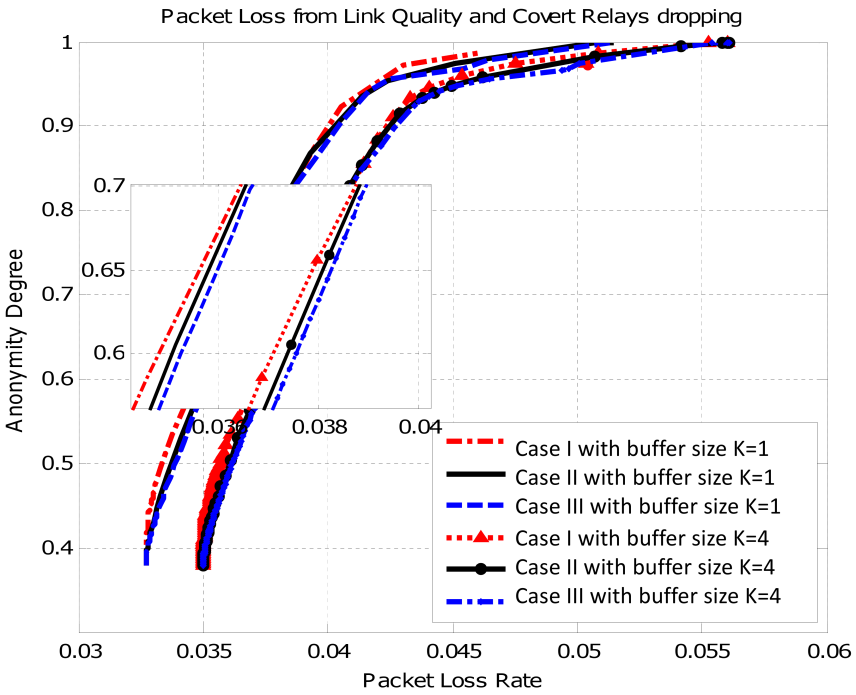


Figure 5.3: The anonymity degree achieved as a function of the packet-loss constraint when packet losses are due to both link quality and covert relays' dropping. For each case of source information, different values of buffer size δ and transmission rate η are considered. The correlation between Z_1 and Z_2 was set to $P = 0.98$.

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.

Chapter 6

RELAY CONFIGURATION FOR GIVEN ROUTE SELECTION

In anonymous network, route selection and relay configuration both affect network performance and anonymity. We have discussed the various route selection based on different entities given relay configuration. To further investigate the anonymous networks, we consider relay configuration assignment given route selection to maximize anonymity in here.

How to optimize relay configuration for anonymous networks was first addressed by []. However, they studied the relay configuration under single path networks where each source does not need to do route selection. In here, we consider multipath network scenario which also cover the single path case in the existing work. In this type of network, we maximize anonymity by choosing relay configuration with given route selection. Namely, the routing selection probability distribution $P(\mathbf{R}|\mathbf{Z},\mathbf{C})$ is given in advance, this work is generalization of [Venkitasubramaniam et al '08] to multi-path network where each source-destination pair has a single path and can be defined as $Q(\mathbf{R}=[\mathrm{R}_1,\mathrm{R}_2,...,\mathrm{R}_\mathrm{N}|\mathbf{Z},\mathbf{C}])=1$. Our relay configuration distribution is denoted as $Q(\mathbf{C})$ and our goal is to optimize anonymity while maintaining performance.

6.1 Problem formulation for Relay Configuration

Given the routing selection probability distribution $Q(\mathbf{R}=[\mathrm{R}_1,\mathrm{R}_2,...,\mathrm{R}_\mathrm{N}]|\mathbf{Z},\mathbf{C}])=P(\mathbf{R}|\mathbf{Z},\mathbf{C})$, the minimization of anonymity loss with given relay configuration is

$$\min_{Q(\mathbf{C})} \; [H(\mathbf{Z}|\mathbf{R})-H(\mathbf{Z}|\mathbf{R},\mathbf{C})\,]\; s.t\; d(\mathbf{Z},\mathbf{R},\mathbf{C})\le D \tag{6.1}$$

Therefore, after rearrange the (6.1), we have

$$R(D) = \min_{E_{\mathbf{Z}} \le D} [H(\mathbf{Z}|\mathbf{R}) - H(\mathbf{Z}|\mathbf{R},\mathbf{C})] = \min_{Q(\mathbf{C}))\in\mathbf{Q}} [H(\mathbf{C}|\mathbf{R}) - H(\mathbf{C}|\mathbf{R},\mathbf{Z})],$$
$$= \min_{Q(\mathbf{C})\in\mathbf{Q}} \Big[\sum_{\mathbf{R},\mathbf{Z}} \sum_{\mathbf{C}} P(\mathbf{Z})P(\mathbf{R}|\mathbf{Z},\mathbf{C})Q(\mathbf{C}) \log \frac{P(\mathbf{R}|\mathbf{C},\mathbf{Z})Q(\mathbf{C})}{\sum_{\mathbf{Z}} \mathrm{P}(\mathbf{Z})P(\mathbf{R}|\mathbf{C},\mathbf{Z})Q(\mathbf{C}} \Big],$$
$$\text{where } \mathbf{Q} = \Big\{ \sum_{\mathbf{R},\mathbf{Z}} \sum_{\mathbf{C}} P(\mathbf{Z})P(\mathbf{R}|\mathbf{C})Q(\mathbf{C})E_{\mathbf{Z}}(P(\mathbf{R}|\mathbf{Z},\mathbf{C})Q(\mathbf{C})) \le D,\ Q(\mathbf{C}) \ge 0,$$

Using the condition $\sum_{\mathbf{C}} Q(\mathbf{C}) = 1$ and system average packet loss rate not exceeding D, the optimization function with Lagrange multiplier can be written as

$$J = \sum_{i=1}^{N} \Big[\sum_{\mathrm{R_i}} \sum_{\mathrm{Z_i}} \mathrm{P}_{\mathrm{Z_i}}(k)Q(\mathrm{R_i}|\mathrm{Z_i}) \log \frac{Q(\mathrm{R_i}|\mathrm{Z_i})}{\sum_{\mathrm{Z_i}} P\mathrm{Z_i})Q(\mathrm{R_i}|\mathrm{Z_i})} + \sum_{\mathrm{Z_i}} v(\mathrm{Z_i}) \sum_{\mathrm{R_i}} Q(\mathrm{R_i}|\mathrm{Z_i}) \ +$$
$$- s \sum_{\mathrm{R_i}} \sum_{\mathrm{Z_i}} P(\mathrm{Z_i})Q(\mathrm{R_i}|\mathrm{Z_i})E_{\mathrm{Z_i}}(Q(\mathrm{R_i}|\mathrm{Z_i})) \Big]. \tag{6.2}$$

Therefore, by using Lagrange multiplier to reformulate the minimization problem, we have the optimal relay assignment $Q(\mathbf{C})$

$$Q(\mathbf{C}) = \frac{\prod_{\mathbf{R},\mathbf{Z}} \big[\frac{q(\mathbf{R},\mathbf{C})}{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}\big]^{P(\mathbf{Z},\mathbf{R}|\mathbf{C})} \exp[s\Phi(\mathbf{C})]}{\sum_{\mathbf{C}} \big(\prod_{\mathbf{R},\mathbf{Z}} \big[\frac{q(\mathbf{R},\mathbf{C})}{Q(\mathbf{R}|\mathbf{Z},\mathbf{C})}\big]^{P(\mathbf{Z},\mathbf{R}|\mathbf{C})} \exp[s\Phi(\mathbf{C})]\big)}, \tag{6.3}$$
$$\text{where } q(\mathbf{R},\mathbf{C}) = \sum_{\mathbf{C},\mathbf{Z}} P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C})Q(\mathbf{C}), P(\mathbf{Z},\mathbf{R}|\mathbf{C}) = P(\mathbf{Z})Q(\mathbf{R}|\mathbf{Z},\mathbf{C}), \Phi(\mathbf{C}) = \frac{\partial E_{\mathbf{Z}}}{\partial Q(\mathbf{C})}.$$

6.2 Result and Performance Discussion

Figure (6.1) shows the anonymity-packet loss trade-off for choosing relay assignments with a given route selection and set of source-destination pairs. In here, I illustrated two different route selection schedules, which are fixed and uniform selection used in Figure (2.3). Under fixed route selection, source $S1$ 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

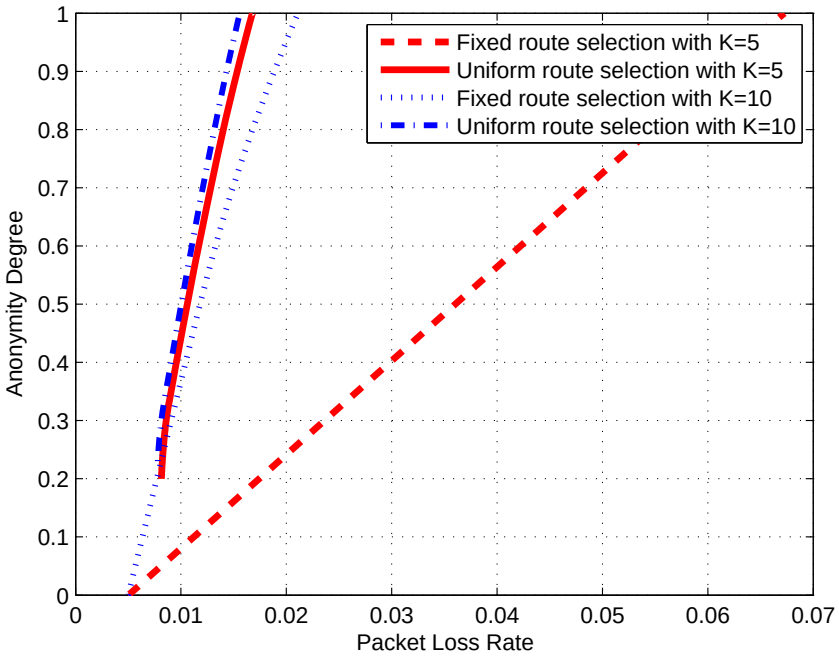


Figure 6.1: Results for given different route selection strategies with covert relay buffer size K=5 and K=10. Varying the relay configuration between covert and visible yields different anonymity for each packet-loss constraint.

selection divides flow equally among the relays, leading to shorter average buffer occupancy and hence fewer packets drops.

6.3 Conclusion

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

Chapter 7

SECURING RFID TRANSMISSION BY VARYING TRANSMITTED READERS AMPLITUDE

Radio frequency identification (RFID) has gained tremendous popularity and research attention in the recent years. There are two distinct advantages with RIFD systems [28]: First, RFID provides unique identifications of each object. Each RFID tag contains an unique identification (UID) number that distinguishes itself from all other tags. Second, the reading range could potentially go up to tens of meters while no line of sight requirement is needed. Since its invention, the RFID technology has found a wide range of applications. This includes passport, drivers license, building access control and supply chain management just to name a few.

Based on the power source that drives communications between the reader and the tag, RFID tags can be classified into active, semi-active and passive three classes [17].. active and semi-passive tags all have on-board batteries, which provide them with reasonable computational capability. Passive RFID tags on the other hand can only harvest power from the reader. The first two classes of tags are powerful, implementations of secure cryptographic primitives on these tags are possible. Many RFID applications such as inventory control require the tag to be inexpensive, in these scenarios, only the passive RFID tags are applicable. However, passive tags are computationally constrained, their memory is generally also very limited. Therefore, one glaring weakness of the passive RFID system is its security and privacy concerns due to these constraints.

Different RFID standards employee different protocols for communications between the reader and the tag. However, in general, this process can be roughly depicted as shown in Figure 1. When a reader tries to communicate with a tag, it first sends a query to the tag. Upon receiving the query, the tag replies with its response. The response may include tags secret information. The connection between them is also established. The reader then sends instructions to the tag to perform various tasks. It is imperative that the sensitive information such as a tags UID to be protected.

There have been numerous attempts made in securing the transmissions between the reader and

the tag. The most straightforward method is through the use of encryption. The computational complexity of public key cryptography is too high. Presently it is not feasible to passive tags. Generally, symmetric key cryptography is implemented. Due to the memory constraint, the key length is usually shorter than 80 bits, which is the presently accepted level for a symmetric key cryptography to be considered secure. Thus it cannot be considered to be secure. For example, the key length of popular EPCglobal Class 1 Gen 2 standard is only 32 bits [3]. This can easily be broken with the exhaustive search attack, and hence does not offer protection against a computationally powerful adversary. Consequently, other alternatives have been sought.

Motivated by the increasing concerns over the privacy issue of RFID systems, we consider the problem of keeping tag's response to the reader secure in the presence of the eavesdroppers. More specifically, we present a physical layer scheme to provide the tag's data confidentiality protection without any need of pre-sharing secrets between tags and readers. In doing so, the user privacy is ensured and the property of tag's universal application is still preserved which can be used by different applications in different countries without any common secrets in advance.

Our proposed frameworks are summarized below:

- **VRTA system design**: We propose a physical layer stand-alone system called propose a Varying Reader's Transmitted Amplitude (VRTA). It provides data confidentiality protection to the passive RFID systems against passive eavesdroppers. In our system, the reader generates and transmits a randomly varying amplitude waveform. This waveform can be successfully removed by the legitimate reader and is seen as interference in the view of the attacker. Hence, the reader and tag do not require pre-shared secrets for secure communication. In addition, the VRTA system utilizes only one transmitter and can be applied to all current commercial readers.

- **Optimal system design parameters**: We conduct theoretical analysis by considering the optimal strategy for the attackers. We show theoretically with the proper selection of system parameters, our scheme is resistant to our adversarial models.

- **VRTA system implementation and performance**: We implement the VRTA prototype using software define radio N210 and USRP1. We verify our system performance through experi-

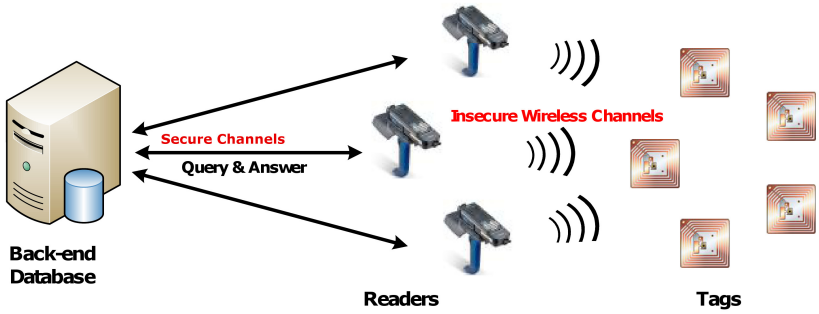


Figure 7.1: RFID system composed of back-end database, readers and tags.

ments. Our experimental results confirm with our theoretical results. They further show the decoding bit error rate (BER) for the single eavesdropper is very close to 0.5, which is perfect secrecy [47]. This also holds for two colluding eavesdroppers if the reader's Tx is placed close to the tag. Thus, the tag to reader data confidentiality are ensured with our VRTA system.

- **Key establishment protocols without pre-sharing secrets**: We leverage our proposed VRTA system to establish secure channel between reader and tags for key establishment protocols. The key establishment is lightweight and no need of pre-sharing secrets. Each key settlement does not rely on previous session key information which increase the robustness

7.1 System and Adversarial Models

In this section, we introduce the system and adversarial models as well as the assumptions our work is based on.

7.1.1 System Model

We consider the most common passive RFID system which is consisted of three components: A back-end database, one or multiple readers and one or multiple passive RFID tags as shown in Figure 7.1. The connection from the reader to the database is assumed through the secure channel. The communication between readers and tags are through insecure wireless channels. Since one reader can only communicate with one tag at a time, and the communication between a tag and a

reader is independently secured, the system model of our interest can be reduced one reader and one passive tag.

The reader is assumed to be full-duplex. At the baseband, the reader can generate N equally spaced levels of amplitude from A_0 to A_{N-1}. We define the step size ΔA to be the difference between two consecutive amplitudes. i.e., $\Delta A = A_{k+1} - A_k$ for $k = 0, \cdots, N-2$. Each time, the reader randomly chooses among N levels of amplitude for transmission. Furthermore, each amplitude duration is T.

The passive RFID tag communicates with the reader via backscattering modulation. It switches its impedance either to low or high to denote a data bit of 1 or 0. We denote the gain coefficient when the tag sets its impedance to high (bit 1) and high (bit 0) to be T_0 and T_1 respectively. The tag's symbol time is T_s.

7.1.2 Adversarial Model

In our adversarial model, the adversaries are passive eavesdroppers whose goal is to deduce the data contents transmitted from the tag to the reader.

We first consider the single passive eavesdropper case. The eavesdropper can listen to and intercept the communication between the reader and the tag. In addition, the eavesdropper is assumed to be mobile. He can freely move around or stay put as he chooses. He also has the complete knowledge of all the protocols and frequencies used for communications between the reader and the tag. Moreover, we assume while the adversary has the knowledge about the time varying nature of the reader's transmitted waveform, he does not have the specific values chosen for the "random" amplitude at any given time. However, the adversary knows the minimum amplitude A_0, maximum amplitude A_{N-1}, the number of steps N and step size ΔA. In other words, the adversary has full knowledge of the system design parameters. We further assume the two coefficients of the tag's impedance T_0, T_1, as well as all channel gains including reader to tag, tag to eavesdropper, reader to eavesdropper are known to the eavesdropper. In this case, the eavesdropper is rather powerful. In practise, it would be very difficult to obtain the impedance gain coefficients and channel gains. Nevertheless, if our designed system can withstand this adversarial model, we can conclude our system can withstand all the weaker adversarial models.

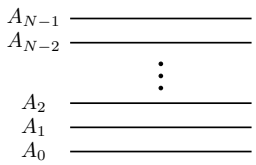


Figure 7.2: N input amplitude levels shown in baseband.

We then test our system by considering two colluding eavesdroppers. In addition to the single eavesdropper's assumptions, we further assume the two eavesdroppers' signal are perfectly synchronized, implying that there would be no relative delays between the two received signals. Finally, we generalize our framework to an arbitrary number of eavesdroppers.

7.2 VRTA Scheme

In this section, we present the high level overview of our proposed scheme. We also show the decoding procedure for the legitimate reader.

7.2.1 Model Formulation

The total input amplitude range is divided into N equally spaced steps. The minimum and maximum reader's transmitted amplitudes are A_0 and A_{N-1} respectively. This is shown in Figure 7.2. In the VRTA scheme, ΔA and N are our system design parameters.

When the tag starts replying to the reader's command, the reader instead of transmitting a constant amplitude as the current approach does, it uniformly and randomly selects one of N amplitude levels from A_0 to A_{N-1} and transmits that amplitude to the tag for a predefined time duration T. In the subsequent time durations, the reader repeats this process till the end of the communication session.

The system with one present eavesdropper is depicted in Figure 7.3. h_{rt}, h_{tr}, h_{te}, h_{re} and h_{rr} denote channel gains of reader's Tx to tag, tag to reader's Rx, tag to eavesdropper, reader's Tx to eavesdropper and reader's Tx to its Rx respectively.

We now formulate the mathematical expressions for the communication between the reader and tag as well as the eavesdropper's intercepted signals. Let f_{ti}, g_{ti}, r_{ti} and m_{ti} be the reader's trans-

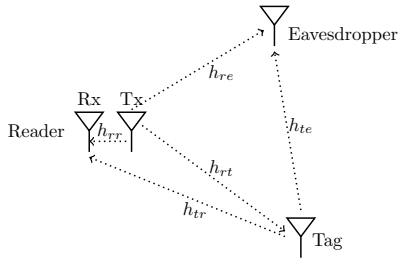


Figure 7.3: System diagram with one eavesdropper

mitted signal, tag's replied signal, reader's received signal and eavesdropper's intercepted signal respectively at time instance i, we have:

$$\begin{aligned} f_{ti} &= A_{ti}, \\ g_{ti} &= h_{rt}(T_{ri}A_{ti}), \\ r_{ti} &= h_{rr}A_{ti} + h_{rt}h_{tr}(T_{ri}A_{ti}), \qquad (7.1) \\ m_{ti} &= h_{re}A_{ti} + h_{rt}h_{te}(T_{ri}A_{ti}). \qquad (7.2) \end{aligned}$$

where $ti \in \{0, \cdots, N-1\}$. T_{ri} is the tag's replied signal at time i, $ri \in \{0, 1\}$ represents backscattered coefficient from different impedance for bit 0 and 1 respectively.

7.2.2 Reader Decoding

At any given time instance i, the reader's received signal r_{ti} from the tag's response T_{ri} would be the sum of reader's time varying amplitude and the tag's returned data with proper channel gain adjustments. When the reader performs decoding, since it can estimate the channel response h_{rr} from the synchronization sequence which we will discuss in the protocol design section, we assume h_{rr} is known to the reader's Rx. Moreover, it knows the input waveform A_{ti} for all i, by observing (7.1), the reader can successfully remove the interference term $h_{rr}A_{ti}$ and obtain s_{ti}.

$$\begin{aligned} s_{ti} &= \frac{r_{ti}}{h_{rr}} - A_{ti} \\ &= \frac{h_{rt}h_{tr}T_{ri}A_{ti}}{h_{rr}}. \end{aligned}$$

Suppose the length of the tag's reply contains M bits, the obtained waveforms after removing the interference becomes $\mathbf{s} = (s_{t0}, \cdots s_{t(M-1)})$. The magnitude of each element in $\mathbf{s}$ should be close to one of two levels, $\frac{h_{rt}h_{tr}T_0\bar{A}}{h_{rr}}$ or $\frac{h_{rt}h_{tr}T_1\bar{A}}{h_{rr}}$, corresponding to tag's reply of bit 0 or 1. $\bar{A} = \frac{A_0+A_{N-1}}{2}$ is the middle point of the input amplitude and $\frac{h_{rt}h_{tr}T_1\bar{A}}{h_{rr}} >> \frac{h_{rt}h_{tr}T_0\bar{A}}{h_{rr}}$. Then the reader's decoding procedure is as follows:

1. Define two sets $\mathcal{A}_0$ and $\mathcal{A}_1$, where $\mathcal{A}_0 = \{s_i | s_i \approx \frac{h_{rt}h_{tr}T_0\bar{A}}{h_{rr}}\}$ and $\mathcal{A}_1 = \{s_i | s_i \approx \frac{h_{rt}h_{tr}T_1\bar{A}}{h_{rr}}\}$.

2. Find a_0, a_1, where $a_0 = \frac{\sum_{s_i \in \mathcal{A}_0} s_i}{\#\mathcal{A}_0}$ and $a_1 = \frac{\sum_{s_i \in \mathcal{A}_1} s_i}{\#\mathcal{A}_1}$ are the average of low and high level signals respectively. Here $\#$ denotes the cardinality of the set.

3. The reader decodes the message bit y_i at time instance i with the following decision rule:

$$y_i = \begin{cases} 0, & s_{ti} \leq \frac{1}{2}(a_1 + a_0), \\ 1, & \text{otherwise.} \end{cases} \tag{7.3}$$

7.3 Security Analysis and Optimal System Parameter Design

In this section, we first discuss the reader's amplitude duration T relative to tag's symbol time T_s. Then we consider the single eavesdropper's attacking strategy. From this, we deduce the optimal system design parameters. We then show the VRTA scheme utilizing these parameters can withstand one eavesdropper. Finally, we consider the two colluding eavesdroppers scenario.

7.3.1 Selection of T and T_s

In our scheme, we require the tag's symbol time T_s to be an integer multiple of reader's time duration T. Equivalently, the amplitude varying rate should be an integer multiple of tag's data rate. If this condition is not met, the eavesdropper can perform differential decoding to recover message bits. This is demonstrated in Figure 7.4.

The black line in the figure represents the reader's varying amplitude. In the absence of the noise, the red line is the eavesdropper's received signals with tag's replies. Since the eavesdropper is assumed to know tag's symbol time T_s, the start point of the varying amplitude and the its duration

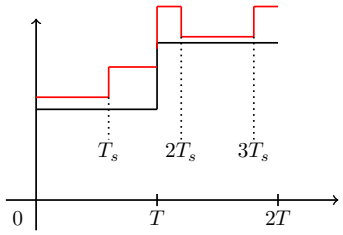


Figure 7.4: Attack exists when $T_s \neq kT$.

T, if T_s is not an integer multiple of T, i.e., $T_s \neq kT$, where k is an positive integer, then the eavesdropper can find some amplitude duration such that the tags response switch from bit 1 to bit 0 or from bit 0 to 1.

In this example, the tag's actual reply is (0,1,0,1). The eavesdropper sees a sudden change in the received signal level at T_s, which is less than the reader's amplitude duration T, he immediately identify the tag's first 2 bits reply are (0,1).

In general, let Δr be the average difference in amplitude between the tag's reply of 0 and 1, the eavesdropper decodes the tag's response y_{i+1} from the average of two tag's data received data intervals r_{i+1} and r_i by performing the differential decoding:

$$y_{i+1} = \begin{cases} y_i, & r_{i+1} - r_i \leq \frac{1}{2}\Delta r, \\ 1 - y_i, & \text{Otherwise.} \end{cases} \tag{7.4}$$

If initially the tags response does not change, the eavesdropper cannot immediate recover those bits. However, as soon as tag's reply flips, regardless from 0 to 1 or 1 to 0, the eavesdropper can immediate recover these two bits, and consequently recovering all preceding and succeeding message bits.

In addition, if the input amplitude A_0 or A_{N-1} is transmitted, the eavesdropper has a better chance of recovering the tag's reply. The reasons is that when reader sends A_0 or A_{N-1}, since the eavesdropper had the knowledge of the channel gain as well as A_0 and A_{N-1}, he can distinguish the tag's reply is 0 or 1 by examining the amplitude level of the received signal. In our system, we require $T_s = kT$. In tag's one symbol time, the probability of taking A_0 or A_{N-1} as the input amplitude increases with increased k. Thus the value of k should be kept small.

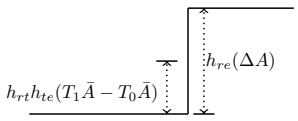


Figure 7.5: Attacking Scenario 1

7.3.2 Single Eavesdropper

In this section, we show two potential attacks which can be employed by the eavesdropper as well as the corresponding system parameter selections to mitigate these two attacks.

From (7.2), the eavesdropper's received signal at a given time i is:

$$m_{ti} \quad = \quad h_{re}A_{ti} + h_{rt}h_{te}(T_{ri}A_{ti}).$$

Here $h_{rt}h_{te}(T_{ri}A_{ti})$ contains tag's replied bit information while $h_{re}A_{ti}$ is the seen as the interference. The eavesdropper can potentially launch two attacks.

Case 1: $h_{rt}h_{te}(T_1A_{k+1} - T_0A_k) < h_{re}(A_{k+1} - A_k)$, where $k = 0, \cdots, N-2$.

Since all steps are equally spaced, we have $\Delta A = A_{k+1} - A_k$. Furthermore, tag's backscattered signal of 1 is much greater than 0, i.e., $T_1 >> T_0$, by choosing $A_0 >> \Delta A$, we can approximate the condition for case 1 as follows:

$$h_{rt}h_{te}(T_1\bar{A} - T_0\bar{A}) < h_{re}(\Delta A),$$

where $\bar{A} = \frac{A_{N-1}+A_0}{2}$ is the middle point of the input amplitude range.

Note that the identical attack still exists without the approximation. The only difference is that the decision region for decoding may be slightly different for different initial input amplitudes. Writing out all N decision regions for N different amplitudes are redundant and would not affect the outcome of the attack. Thus the approximation is used. The same is true for the attack case 2.

In this attack, the difference of two consecutive amplitudes is greater than the difference between tag's replied messages 0 and 1. This is pictorially shown in Figure 7.5. This implies the interference step is too great that the eavesdropper can immediately identify the interference level $h_{re}A_{ti}$. Thus at time instance i, the eavesdropper simply decodes the received bit y_i as follows:

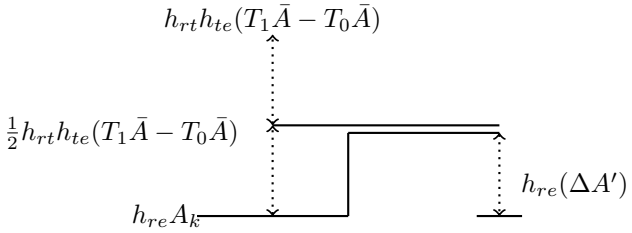


Figure 7.6: Attacking Scenario 2

$$y_i = \begin{cases} 0, & (m_{ti} - h_{re}A_{ti}) \leq \frac{1}{2}h_{rt}h_{te}(T_1\bar{A} + T_0\bar{A}), \\ 1, & \text{Otherwise.} \end{cases} \tag{7.5}$$

Therefore, to prevent the attack case 1, the following condition should be satisfied.

$$\Delta A \leq \frac{h_{rt}h_{te}(T_1\bar{A} - T_0\bar{A})}{h_{re}}, \tag{7.6}$$

ΔA in (7.6) determines the upper bound on the chosen step size to ensure a secure system.

Case 2: $h_{rt}h_{te}(T_1A'_{k+1} - T_0A'_k) > 2h_{re}(A'_{k+1} - A'_k)$, where $k = 0, \cdots, L-2$.
We have defined a new set of partitions for the input amplitudes which has a total of L steps. We denote the new step size $\Delta A' = A'_{k+1} - A'_k$. Using the same argument, we approximate the condition as

$$h_{rt}h_{te}(T_1\bar{A} - T_0\bar{A}) > 2h_{re}(\Delta A').$$

This is shown pictorially in Figure 7.6.

This case implies the tag's replied signal is much stronger than the step size $\Delta A'$ when taking into considerations of the channel gain. The eavesdropper picks two intervals m_{ti} and m_{tj} satisfying $|m_{ti} - m_{tj}| < |h_{rt}h_{te}(T_1\bar{A} - T_0\bar{A})|$. If the input A'_{ti} and A'_{tj} that corresponds to output m_{ti} and m_{tj} is separated by less than or equal to the step size, i.e., $|A'_{ti} - A'_{tj}| \leq \Delta A'$, then the eavesdropper

can identify the tag's replied 2-bit tuple as follows:

$$(y_i, y_j) = \begin{cases} (1,0)\ , & m_{ti} - m_{tj} > \frac{1}{2} h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A}), \\ (0,0) \text{ or } (1,1), & |m_{ti} - m_{tj}| \leq \frac{1}{2} h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A}), \\ (0,1)\ , & m_{ti} - m_{tj} < -\frac{1}{2} h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A}). \end{cases}$$

We see that half of the time the eavesdropper would be able to uniquely decode the 2-bit tuple. For the case where the eavesdropper cannot tell whether it's (0,0) or (1,1), he can maintain one of m_{ti} or m_{tj} and choose another interval m_{tk} that satisfies the constraint and repeat the same decoding procedure. The eavesdropper does this repeatedly until he is able to decode. Once the eavesdropper can decode one 2-bit tuple, he can then uniquely decode all the bits taken previously.

The smaller the step size, the more steps fall into the attacking region, the higher the probability of this attack can occur. Therefore, to best prevent this attack, the following condition should be satisfied.

$$\Delta A' \geq \frac{h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A})}{2 h_{re}}, \tag{7.7}$$

$\Delta A'$ in (7.7) determines the lower bound on the chosen step size.

Based on the conditions drawn from (7.6) and (7.7), we conclude our design parameter ΔA should be within the interval shown as follows:

$$\frac{h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A})}{2 h_{re}} \leq \Delta A \leq \frac{h_{rt} h_{te} (T_1 \bar{A} - T_0 \bar{A})}{h_{re}}. \tag{7.8}$$

In the RFID systems where the distance between reader's Tx to eavesdropper d_{re} is comparable to the distance between tag and eavesdropper d_{te}, then $|h_{re}| \approx |h_{te}|$. In the commercial readers, this condition is satisfied if the reader is placed close to the tag. Then (7.8) reduces to:

$$\frac{h_{rt} (T_1 \bar{A} - T_0 \bar{A})}{2} \leq \Delta A \leq h_{rt} (T_1 \bar{A} - T_0 \bar{A}).$$

In this case, the step size contains an upper and lower bound, it becomes simple to choose.

The system becomes a little more difficult to design if $|h_{re}| \approx |h_{te}|$ does not hold. Since the reader has no knowledge of the whereabouts of the eavesdropper, h_{re} and h_{te} are not known. Therefore, it is not guaranteed that one can find ΔA which can always satisfy (7.8) and consequently

thwart the attack. However, here we have assumed all factors to be ideal, in the real system with channel gain inconsistencies, non-linear channel gain and added noise, we expect it would be very difficult for the eavesdropper to perform the aforementioned attacks.

7.3.3 Two Eavesdroppers

In this section, we consider two colluding eavesdropper's case. Each eavesdropper intercepts its own set of signals independent from the other eavesdropper. The optimal strategy for the eavesdropper is to try to cancel the interference and then perform the decoding. We first show how this can be accomplished. Then we show the necessary condition to prevent this attack. Finally, we generalize our prevention method to arbitrary number of eavesdroppers.

Let m_{ti} and m'_{ti} be the two eavesdropper's received signal at time i, with a tag's response of 0. Using the same notation, we can write m_{ti} and m'_{ti} as follows:

$$\begin{aligned} m_{ti} &= h_{re}A_{ti} + h_{rt}h_{te}(T_0A_{ti}), \\ m'_{ti} &= h'_{re}A_{ti} + h_{rt}h'_{te}(T_0A_{ti}). \end{aligned}$$

When the tag's response is 0, the impedance is set to high, implying $T_0 \approx 0$. Therefore, the two eavesdroppers can estimate the channel gain ratio between them as follows:

$$\frac{|m_{ti}|}{|m'_{ti}|} \approx |\frac{h_{re}}{h'_{re}}|.$$

Suppose at time instance j, two eavesdroppers receive m_{tj} and m'_{tj} with a tag response of 1:

$$\begin{aligned} m_{tj} &= h_{re}A_{tj} + h_{rt}h_{te}(T_1A_{tj}), \\ m'_{tj} &= h'_{re}A_{tj} + h_{rt}h'_{te}(T_1A_{tj}). \end{aligned}$$

The two eavesdroppers try to cancel the interference as follows:

$$m'_{tj}\frac{|m_{ti}|}{|m'_{ti}|} - m_{tj} \approx h_{rt}T_1A_{tj}(h'_{te}\frac{h_{re}}{h'_{re}} - h_{te}). \tag{7.9}$$

$h_{rt}T_1A_{tj}(h'_{te}\frac{h_{re}}{h'_{re}} - h_{te})$ in (7.9) represents the value for tag's response of bit 1. A non-zero value implies the two eavesdroppers can successfully remove the interference. Following this, they can perform the identical decoding procedure as the legitimate reader as shown earlier.

This attack is not successful if (7.9) is zero. This occurs when the reader is placed close to the tag, then $|\frac{h_{te}}{h_{re}}| \approx |\frac{h'_{te}}{h'_{re}}|$. This implies in an attempt to cancel the interference by the two eavesdroppers, their respective channel gains cause the message contents also to be canceled out, leaving the attack unsuccessful.

This argument applies to any numbers of colluding eavesdroppers. Therefore, we conclude that by selecting the system parameters using (7.8), our VRTA scheme is secure against an arbitrary number of eavesdroppers if the reader is placed close to the tag.

7.4 RFID Secure Transmission Protocol

In Section 3, we have made the assumption that the reader can completely remove the interference A_{ti}. This requires the knowledge of channel gain h_{rr}. In this section, we first show how this is achieved. Then we present our protocol to ensure tag to reader data confidentiality.

In the commercial readers, the reader's Tx and Rx are placed very close together. In this case, $|h_{rr}| \approx 1$, $|h_{rr}|$ is treated as known in the view of the reader. Alternatively, if one wants to be very accurate, or concerned with non-linear gain over different amplitudes which we will discuss in the next section, one can always measure the channel gain and store it in a look up table.

The benchmark waveform for channel estimation is shown in Figure 7.7. The reader first sends out a pseudorandom (PN) sequence. This sequence is used to perform synchronization between Tx and Rx to identify the start of a communication session for the receiver. The sequence chosen is an m-sequence with length 63. The reason for choosing m sequence is because it has ideal 2-level autocorrelations, which is desired for receiver synchronization [13].

After sending the PN sequence, the reader's Tx starts sending a stair case function which is used to measure channel gains for different input amplitudes. The reason is to measure the non-linear gains and make the decoding less error prone. At the reader's Rx, it obtains the channel gain as follows:

1. The receiver synchronizes with the incoming signal by computing the correlations between the received signal and the locally generated PN sequence.

2. Once the signals are synchronized, the receiver identifies the start of the stair case function.

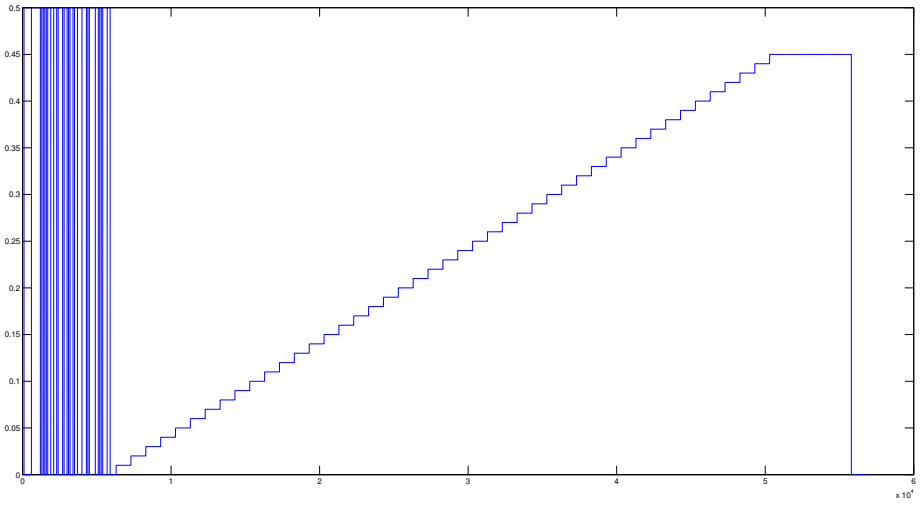


Figure 7.7: The benchmark waveform for channel gain estimation.

3. For each step, the receiver computes the channel gain by taking the average of the received signals for that step and then divide by the input amplitude.

4. Channel gains for different input amplitudes are stored in a look-up table.

Note that in most cases, this procedure needs to be applied only once, the channel gain is obtained and stored. In the subsequent communications, the reader's Rx immediately reads the channel gain from the look-up table, then it performs the proper gain adjustments to the received signals. In the rare event where the channel gain is no longer accurate, one can always repeat this procedure to update the channel gain.

7.4.1 Secure Transmission Protocol

In this section, we explain our protocols for ensuring tag to reader data confidentiality. Figure 7.8 shows the protocol command issued by the reader. The entire protocol works as follows:

1. Similarly to the case of channel estimation, the reader initiates the communication by sending the same PN sequence to the receiver for synchronization.

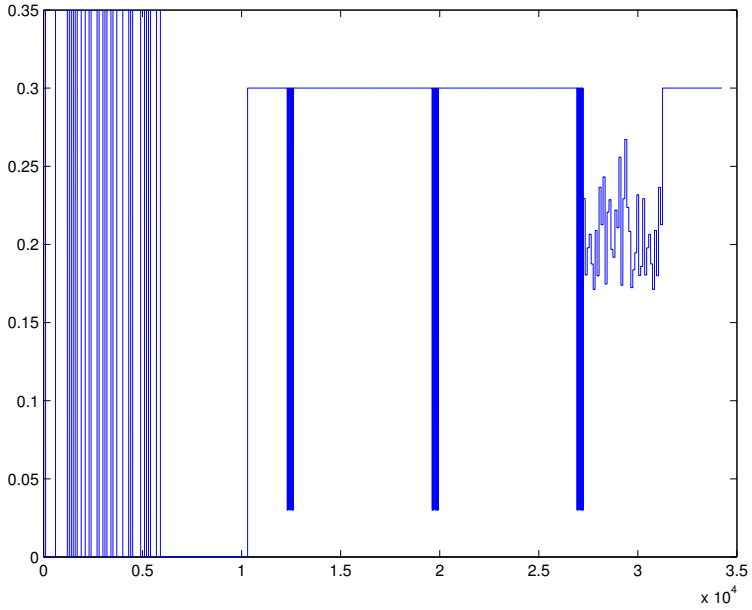


Figure 7.8: VRTA protocol composed of the PN sequence, three repeated commands and varying amplitude waveform.

2. The reader sends a constant amplitude wave followed by a command. This is repeated three times. The constant amplitude wave supplies the tag with sufficient power to identify the command, while the three repeated commands can assist the tag in identifying the starting point for the incoming time varying waveforms and lead to more robust system design.

3. After observing the third command, the tag starts replying to the reader. At the same time, the reader starts sending out the time varying waveform. This waveform serves two functions: 1). To supply the tag with sufficient power for computations and replies via backscattering modulation. 2). To serve as interference signal to the eavesdropper, preventing him from being able to decode.

Note that in the most accurate case, the magnitude of the backscattered response changes slightly with different input amplitudes. This is due to the tag's response is essentially the input amplitude multiplied by a coefficient T_0 or T_1. Therefore, if the $N\Delta A$ is comparable with A_0. The assumption that we can replace instantaneous A_{ti} with the average amplitude $\bar{A}$ in Section 3 becomes invalid.

This may further cause decoding errors. In addition, if A_0 is very low, then the tag may not be able to harvest the power at all. Consequently, in our input amplitude range design, we choose the minimum and maximum amplitude to be 0.17 and 0.27 respectively.

7.5 Lightweight Key Establishment Application

The VART system is able to provide the data confidentiality from tags to reader, but not guarantee securing data when readers sending information to tag. Since most RFID tags are lack of powerful computation capability, many tags applied primitive bitwise operations such as AND, OR and XOR cipher to secure data [30] or authentication purpose [37, 38].

The Minimalist Mutual Authentication Protocol (MMAP) and Efficient Mutual Authentication Protocol (EMAP) were first proposed in [37, 38] to provide light-weight mutual authentication. Similarly, bitwise operation for encryption is proposed in [32]. Although bitwise operation is very simple to use, how to distribute and management the key agreement became challenging for traditional RFID system. In here, we applied the VART system to distribute the key agreement for encryption and authentication purpose. The VART system generate the one way secure channel from tag to reader for encryption and authentication purpose.

7.5.1 One Time Key Protection Transmission

In this section, we will describe how to generate key agreements. First, the RFID tags generate N-bit keys by using pseudo-random number generators such as [5, 39]. Then this key is transported securely using our VRTA scheme which is a secure channel. This allows the reader and the tag to share the same key. A secure channel prevent eavesdropper obtaining the key agreement between tags and readers. Hence the reader can securely communicate with this tag using the establishment key. The steps in our proposed one time key protection transmission is

1. The reader sends a hello message to the tags which power the tag and start to vary the amplitude for secure channel establishment.

2. The tag starts to generate the random bit sequence via pseudo-random number generator and send this bit sequence as secret key agreement back to reader.

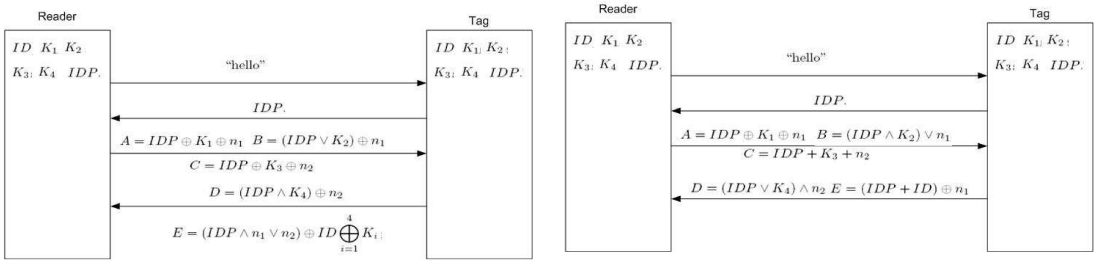
3. The reader receives the secret bit sequence from the tag via the secure channel established by VRTA system. Then, the reader applies the XOR bit operation to encrypt the data through secret bit sequence.

4. The tag receives the encrypted data from reader and uses bit operation XOR to process decryption. Then, if the reader asks for one more iteration, the tag generates another random sequence for another transmission.

When reader sent private data to tag, the key agreements of bit sequence protect the privacy by using XOR operation with data information. When tags send value data back to reader, VATA secure the information. This protocol only require random number generator for key establishment which minimize the computation loading for low cost tags.

7.5.2 Mutual Authentication Protocol Improvement

The steps of the EMAP and MMAP protocols both require reader and tag using their storage keys to verify each side. The tag also stores an identifier IDP, which is used by the reader to identify the tag, as well as a secret quantity ID. All keys and identifier are assumed to be bit strings of length k. All computation are performed via bitwise operation such as XOR, AND, and OR. The protocol steps is list in Figure (7.9) where the notations $\oplus$,$\wedge$, and $\vee$ denote bitwise XOR, AND, and OR, respectively.



(a) (b)

Figure 7.9: (a) EMAP mutual authentication protocol where RFID tag (b) MMAP mutual athenticaiton protocolnt.

Attacks on MMAP and EMAP

The drawback for EMAP and MMAP are the eavesdropper can see the binary strings for each iteration. The attacks on each protocol have bee addressed in the references [2, 31]. The goal of both attacks is to determine the secret quantity ID. An example of the attack on MMAP is as follows. When the adversary eavesdrops on a protocol instance, the adversary has access to the messages $B = (IDP \wedge K2) \vee n1$ and IDP. By the properties of bitwise OR and AND, if $(IDP)_i$= 0,then $(B)_i$= $(IDP \wedge K_2)i \vee (n_1)_i = (n_1)_i$. Hence any bits of n_1 corresponding to 0 bits of IDP will become known to the adversary. For example, suppose that the adversary observes that B = 011000 and IDP= 101100. Based on the above, the adversary has that n_1 =*1**00. Using the observed message E, the adversary then determines that ID is given by ID= $(E \oplus n_1) - IDP$ = *1**00 + 010100 = ****00. This reveals the two least significant bits of ID; the steps to recovering the remaining bits of E based on further protocol runs are described in [2] .The attack on EMAP is described as follows. Since the message D is given by D = $(IDP \wedge K_4) \oplus n_2$, whenever $(IDP)_i$= 0, $(n_2)_i = D_i$. Similarly, whenever $(IDP)_i$= 1,B_i is given by the complement of $(n_1)_i$. The remaining bits of n_1 and n_2 are obtain tampering with the message sent between the reader and tag, as described in [31].

The major drawbacks for those existing approaches is that the wireless open medium allow eavesdroppers to see the encrypted data and then able to decode it. In

Proposed Mutual Authentication Protocol

In here, we proposed using VART system to establish secure channel for the mutual authentication process. The steps in our proposed mutual authentication protocols are:

1. The reader send the ”hello” message to the tag and start to vary the amplitude for secure channel establishment.

2. The tag use pseudo-random number generators to generate the key agreement K and send back to reader with its IDP.

3. Based on the value of IDP, the reader looks up the values of his data memory to pick up the

mutual authentication secret agreement S. Then the reader encrypt the message $A = K \oplus S$ and send it back to the tag.

4. The tag computes the encrypted message $A \oplus S = K$ to obtain he mutual authentication secret agreement K. Then, the tag send back the new IDP' and mutual secret agreement S' back to the reader.

5. The reader update the IDP and the mutual secret agreement S'

By using secure channel established my VART system, we can provide not only data confidentiality from readers to tags but also robust mutual authentication.

7.6 VRTA Implementation and Experimental Results

We have implemented our scheme using the Intel WISP tag as the passive RFID tag, and USRPs as the reader and eavesdroppers. In this section, we first describe the implementation of secure transmission protocol and experimental setup with the selection of the system design parameters. Then using our design parameters, we show the waveforms after removing the interference for the legitimate reader, the single eavesdropper, and the two colluding eavesdroppers. We have verified our VRTA system performance with different tag's data rates and amplitude varying frequencies. Finally, performance comparisons are made in terms of BER among different parties and settings.

7.6.1 Experimental Setup

We use one USRP N210 with one RFX900 daughter board as the reader. The reader uses the linear vertical directional antennas with a gain of 2 dBi. The eavesdroppers are implemented using USRP 1 with two RFX900 daughter board. The two eavesdroppers use the circular polarized antennas with a gain of 6 dBi. Thus, in our experimental settings, the eavesdroppers are more powerful than the legitimate reader's Rx. We use Intel WISP tag as our passive RFID tag.

Both the eavesdropper and the reader's sampling rate is 1 MHz. Upon receiving the third command, the tag replies a 40-bit sequence. In our experiment, the 40-bit sequence is predefined so we can evaluate the performance of our system. Furthermore, we have determined the relationship between tag's data rate T_s and the amplitude duration T is $T_s = kT$, where k is an positive integer.

In this experiment, since tag's data rate is set to 10 kbps, we consider two amplitude varying rates at 10 kHz and 20 kHz. Or equivalently, $k = 1$ and $k = 2$. We want to see in practise if there is any impact on the performance of our system in terms of decoding BER by changing the amplitude varying rates. Furthermore, throughout our experiments, we take $\Delta A = 0.005$, the corresponding number of steps $N = 20$.

7.6.2 Reader's Decoding Performance

We first show the experimental results for the legitimate reader. The reader follows the interference cancelation and decoding procedure shown in Section III. Figures 7.10 and 7.11 display the resulting waveforms after removing the interference signals.

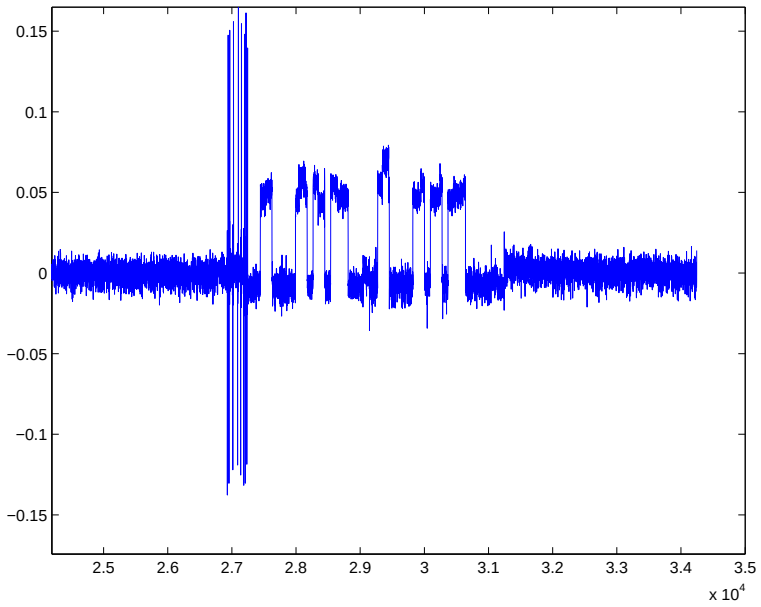


Figure 7.10: Results for reader's signal after removing interference when amplitude varying rate is 10 kHz.

We see in both cases, the signals are relatively clean, the reader can uniquely decode them to message bits. Therefore, we conclude that the reader's amplitude varying rate does not impact the reader's performance.

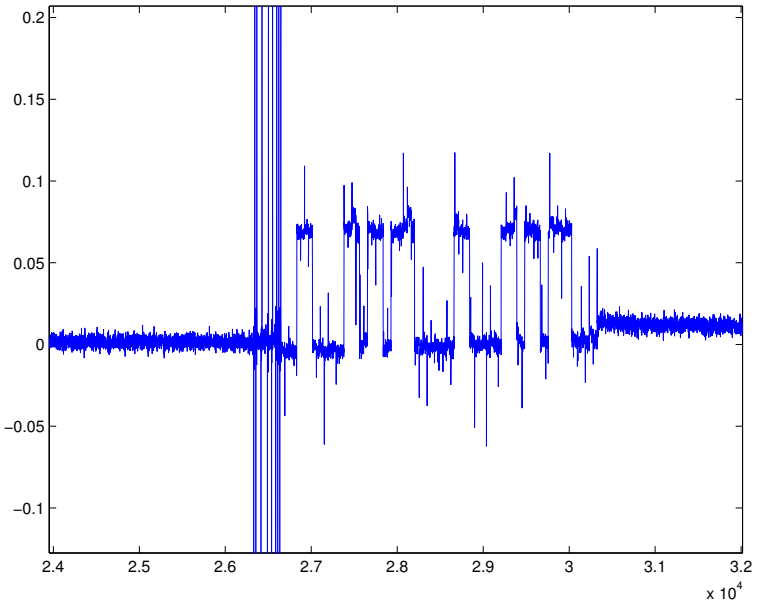


Figure 7.11: Results for reader's signal after removing interference when amplitude varying rate is 20 kHz

7.6.3 Single Eavesdropper Attack

In this section, we consider the single eavesdropper attack. The RFID reader, tag and the eavesdropper's locations are depicted in Figure 7.12. We choose the eavesdropper to be on the opposite side of the reader. This is because the eavesdropper wants to maximize the ratio of the received tag's signal to reader's time varying interference.

As our assumptions have stated, the eavesdropper has complete knowledge of the protocol. Therefore, he can perform the synchronization and identify the starting point for the time varying signal. The eavesdropper's received signals for the 10 kHz and 20 kHz amplitude varying rate are shown in Figure 7.13.

In the analysis section, we have theoretically shown by correctly choosing the system design parameters, the single eavesdropper cannot perform decoding on the tag's replies. Nevertheless, the single eavesdropper still tries to apply the two attacks discussed in Section IV.

The eavesdropper is assumed to know all system parameters and channel gains. In the first

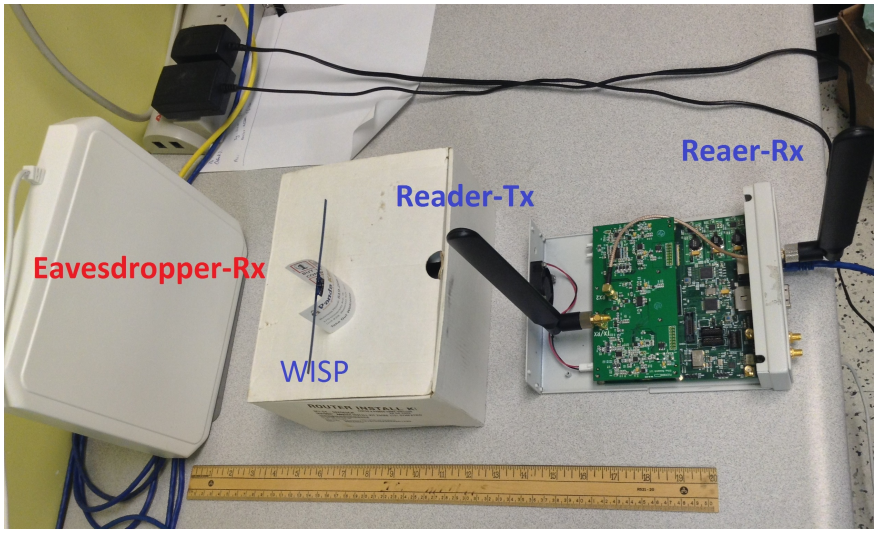


Figure 7.12: Single eavesdropper experiment setup

attack, he subtracts the received signal by the varying amplitude level that is immediate below. He repeats this procedure for all 40 tag's symbol time to obtain the waveform after removing the interference. In the second attack, the eavesdropper uses the first point in the first symbol time as the reference, and subtract all subsequent received signal points by the first point to remove the interference. Note that in the second attack, the eavesdropper can use any points as a reference. In fact, we selected the first point from each of 40 tag's symbol time as the reference point, the resulting BERs are not affected.

Now the eavesdropper can perform the decoding. He computes the average for each symbol time as well as the average for the entire waveform. If the magnitude of the signal in each symbol time is less than or equal to one half of the average of the entire waveform, then the eavesdropper decodes as bit 0. Otherwise, the eavesdropper decodes as bit 1.

The tag's reply bit patterns are shown in the first plot of Figure 7.14. Second and third plots in Figure 7.14 are the eavesdropper's decoded bits after applying the two aforementioned attacks. The BER in both cases are close 0.5. Therefore, our experiment results support our theoretical results in that by correctly choosing the system parameters, the single eavesdropper cannot decode the tag's replies with a higher successful rate than the random guessing.

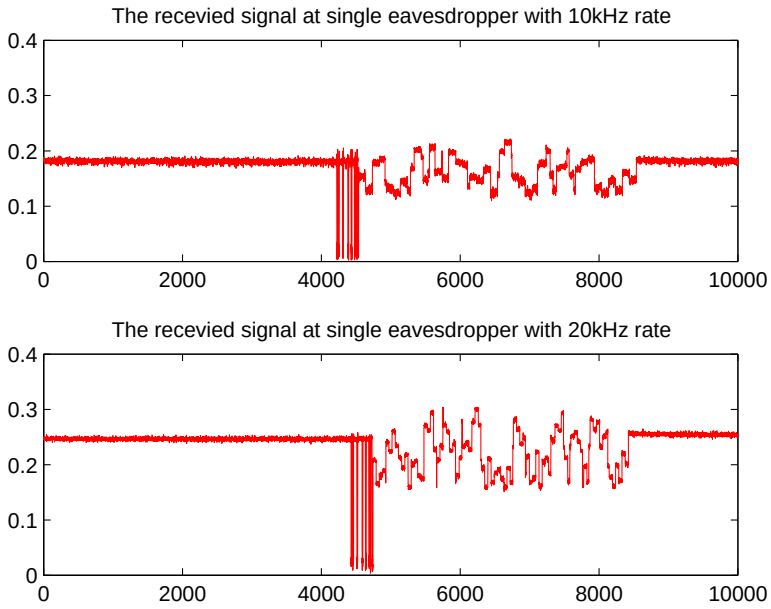


Figure 7.13: Results for received signals by the eavesdropper with 10 kHz and 20 kHz varying rate.

7.6.4 Two Colluding Eavesdroppers

In this section, we demonstrate the multiple colluding adversaries attack. As usual, we set the reader's varying amplitude at two different rate 10 kHz and 20 kHz. Our initial experimental setup is shown in Figure 7.15.

We initially choose the location such that one eavesdropper is close to the reader and the other one is closer to the tag, and the reader's Tx is placed distant from the tag. From our theoretical analysis earlier, this setup would allow the two colluding eavesdroppers to remove the interference. The two eavesdroppers received signals are shown in Figure 7.16, when varying rate is 20 kHz. Following the procedures described in Section IV, the waveform after removing the interference is shown in Figure 7.17. One can observe this waveform is very clean, the two eavesdropper can uniquely decode them into message bits.

Now we keep all parties in place and start moving the tag towards the reader's Tx, based on our analysis, the closer the tag is towards the reader's Tx, the harder it is for the two eavesdroppers to cancel the interference. Figure 7.18 shows the two eavesdroppers' received signal when the tag is

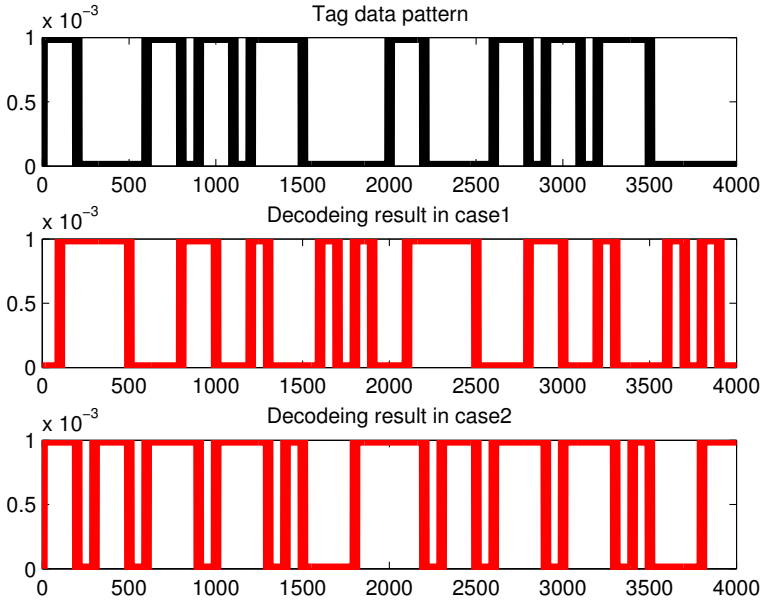


Figure 7.14: Decoding results of the eavesdropper by applying two attacks

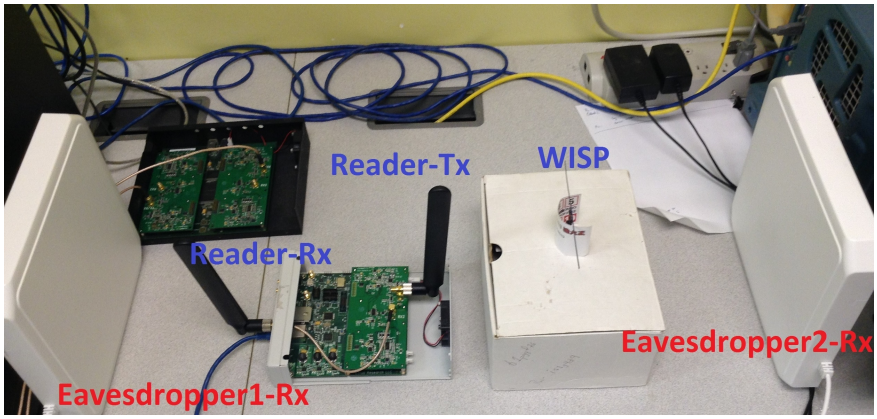


Figure 7.15: Two colluding eavesdroppers experiment setup.

5cm away from the reader's Tx. Figure 7.19 shows the waveform after the interference cancellation. Comparing Figures 7.17 and 7.19, we see noticeable differences in the decoded waveform. One can clearly make decoding decisions given the waveform in Figure 7.17. However, when given the waveform in Figure 7.19, decoding becomes more difficult and decoding errors are expected to

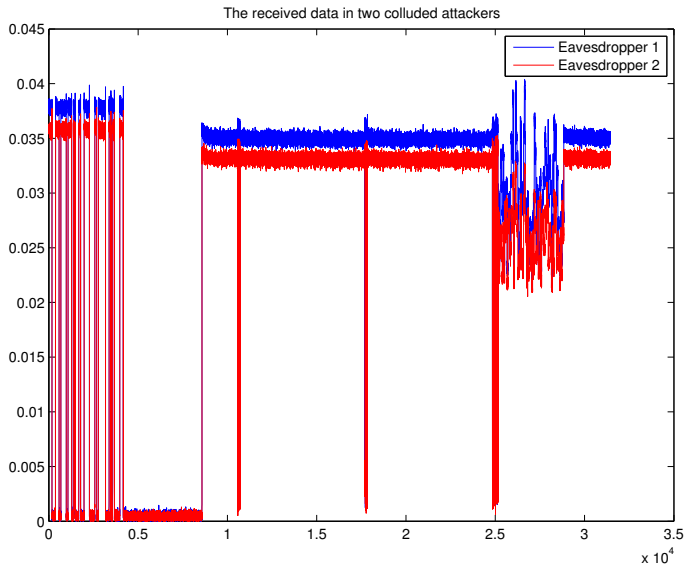


Figure 7.16: Two eavesdroppers received signals when tag is distant from reader's Tx

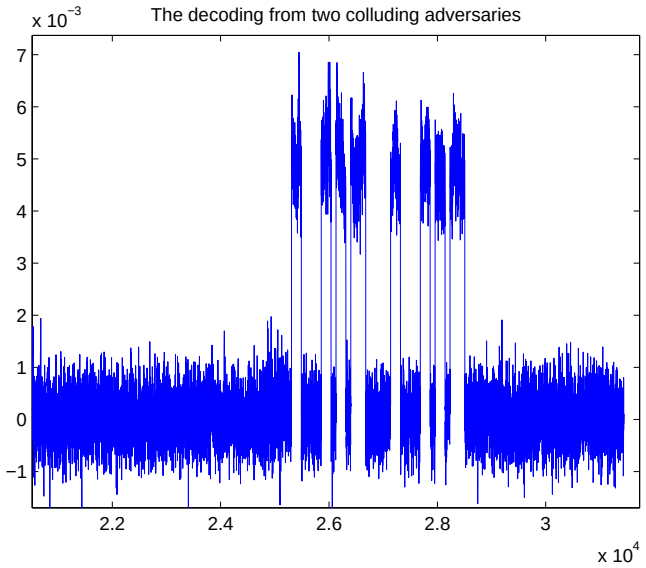


Figure 7.17: Waveform after interference cancellation when tag is distant from reader's Tx.

occur.

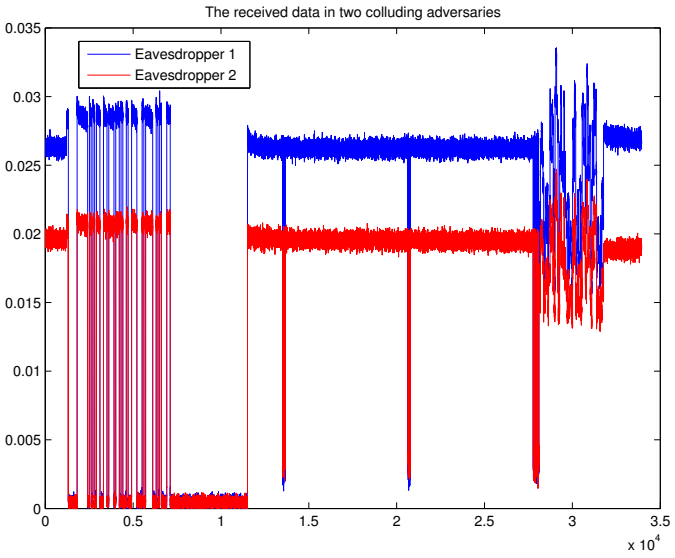


Figure 7.18: Two eavesdroppers received signals when tag is close to reader's Tx.

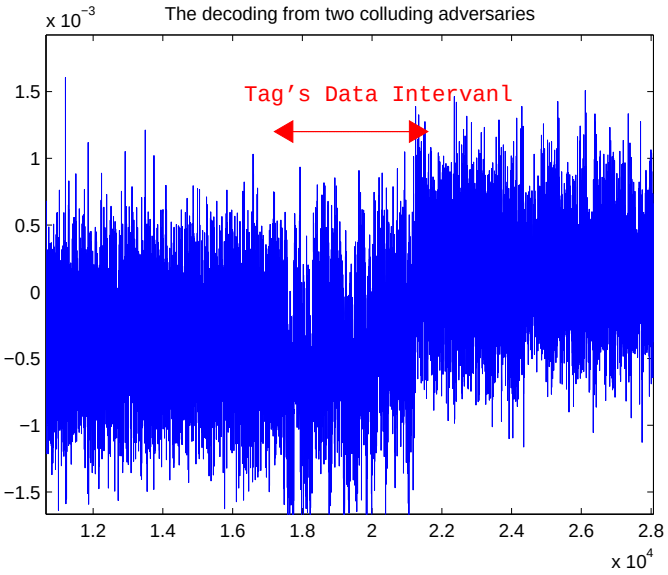


Figure 7.19: Waveform after interference cancellation when tag is close to reader's Tx.

7.6.5 Performance Comparisons

In this section, we present the overall system performance in terms of decoding BER for the different parties. The decoding BER is calculated from 1000 decoded bits. The distance is measured between

Table 7.1: VRTA system performances

Distance	25cm	10cm	5cm
BER at single adversaries(10 kHz)	0.50	0.50	0.48
BER at single adversaries(20 kHz)	0.51	0.49	0.50
BER at colluding adversaries(10 kHz)	0	0.19	0.51
BER at colluding adversaries(20 kHz)	0	0.24	0.49
BER at Reader (10 kHz)	0	0	0
BER at Reader (20 kHz)	0	0	0

the reader's Tx to the tag. The results are shown in Table 7.1.

Several conclusions are drawn from these results:

- The varying rate does not affect the reader's performance in terms of decoding tag's data.
- Single eavesdropper's decoding BER is nearly 0.5, implying he can do no better than guessing.
- Two colluding eavesdropper's decoding BER is low when the tag is distant from reader's Tx, as the tag moves closer to the reader's Tx, BER starts to increase till it reaches approximately 0.5 at 5cm.

These experimental results have confirmed with our theoretical analysis results. In addition, at the distance of 10 cm, the higher varying rate gives better protection against colluding eavesdroppers. This is contradictory to our theoretical results. There are 3 reasons: 1). We did not implement the attack by looking for input amplitude A_0 and A_{N-1}. This is because the accurate estimation of the channel is very difficult in the real system. Without this information, the eavesdroppers cannot identify A_0 and A_{N-1}. Thus, eavesdropper cannot improve his BER. 2). The input amplitude experience non-linear channel gains. In our experiments, the two eavesdroppers cancel the interference only using the estimated gain ratio. Therefore, the system is more error prone with the higher variation rate. 3). Analysis assumes the perfect synchronization. However, this assumption is very difficult to achieve in practise. Non-perfect synchronization would cause higher BER in a higher amplitude varying rate system.

7.7 Conclusions and Future Work

In this paper, we have introduced VRTA, an novel approach to ensure tag to reader data confidentiality. This scheme requires no modifications on the tag and the existing protocols, only the amplitude of the reader supplied waveform is varied. Therefore, it has minimal impact on the existing system. We have also shown the decoding procedure for the legitimate reader. Moreover, we have considered two adversarial models. The single passive eavesdropper and multiple colluding passive eavesdroppers. We have theoretically shown with proper selections of the system parameters, our scheme can withstand the single eavesdropper's attack. We have also shown using parameters selected for single eavesdropper adversarial model, our VRTA scheme is secure against any arbitrary number of eavesdroppers when the reader is placed close to the tag. In addition, we have implemented our scheme with USRP, both the single eavesdropper and two colluding eavesdroppers models are considered. Experimental performance confirms with our theoretical analysis. In addition, we apply VRTA approaches to distribute key agreement for securing reader's data transmission. Since most of RFID tags are not able to execute conventional key establishment approaches such as Diffie-Hellman and other public-key-based methods, VRTA is a good platform to distribute the keys system by leveraging the secure channel from tags to readers. We show how to apply our proposed system to delivery key or have key establishment without pre-sharing secrets between reader and tags. By establishing key agreements, the reader can secure data transmission to tags through encryption assistance and even use in mutual authentication purpose. Therefore, the computational requirements can be minimized and affordable for passive tags by utilizing VRTA secure channel property instead of extensive key establishments.

For future works, we would like consider the following problems: 1). Comparing and investigating our approach with other existing key establishment mechanisms in terms of computational resources. Our goal is to minimize computation requirement for key establishment without any need of pre-sharing secrets.

2). Implement VRTA within the EPCglobal Class 1 Gen 2 framework to further evaluate the performance of our proposed system. Since this standard has been widely used in many industries, by verifying the VRTA performance in EPCglobal framework we can be aware how much performance we can achieve to benefit industry product.

3). Backscattering modulation is widely used in battery-less wireless sensors. We can extend our system to other backscattering modulation systems for security enhancement.

BIBLIOGRAPHY

[1] F. Achard and O. Savry. A cross layer approach to preserve privacy in rfid iso/iec 15693 systems. In *RFID-Technologies and Applications (RFID-TA), 2012 IEEE International Conference on*, pages 85–90, Nov 2012.

[2] Basel Alomair, Loukas Lazos, and Radha Poovendran. Passive attacks on a class of authentication protocols for rfid. In Kil-Hyun Nam and Gwangsoo Rhee, editors, *Information Security and Cryptology - ICISC 2007*, volume 4817 of *Lecture Notes in Computer Science*, pages 102–115. Springer Berlin Heidelberg, 2007.

[3] Coral Atock. Where's my stuff?[supply chain management]. *Manufacturing Engineer*, 82(2):24–27, 2003.

[4] Adam Back, Ulf Mller, and Anton Stiglic. Traffic analysis attacks and trade-offs in anonymity providing systems. In IraS. Moskowitz, editor, *Information Hiding*, volume 2137 of *Lecture Notes in Computer Science*, pages 245–257. Springer Berlin Heidelberg, 2001.

[5] Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Pseudorandom functions revisited: The cascade construction and its concrete security. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 514–523. IEEE, 1996.

[6] R.E. Blahut. Computation of channel capacity and rate-distortion functions. *Information Theory, IEEE Transactions on*, 18(4):460–473, Jul 1972.

[7] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–90, February 1981.

[8] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[9] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.

[10] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. Technical report, DTIC Document, 2004.

[11] Claudia Daz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul Syverson, editors, *Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 54–68. Springer Berlin Heidelberg, 2003.

[12] EPCGlobal. *UHF Class 1 Gen 2 Standard, V 2.0*, Nov. 2013.

[13] Pingzhi Fan. *Sequence Design for Communications Applications*. Research Studies Press, 1996.

[14] Martin Feldhofer, Sandra Dominikus, and Johannes Wolkerstorfer. Strong authentication for rfid systems using the aes algorithm. In Marc Joye and Jean-Jacques Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004*, volume 3156 of *Lecture Notes in Computer Science*, pages 357–370. Springer Berlin Heidelberg, 2004.

[15] Emory A Fry and Leslie A Lenert. Mascal: Rfid tracking of patients, staff and equipment to enhance hospital response to mass casualty events. In *AMIA Annual Symposium Proceedings*, volume 2005, page 261. American Medical Informatics Association, 2005.

[16] Philippe Golle, Markus Jakobsson, Ari Juels, and Paul Syverson. Universal re-encryption for mixnets. In *Topics in Cryptology–CT-RSA 2004*, pages 163–178. Springer, 2004.

[17] Amit Grover and Hal Berghel. A survey of rfid deployment and security issues. *JIPS*, 7(4):561–580, 2011.

[18] Ting He and Lang Tong. Detection of information flows. *Information Theory, IEEE Transactions on*, 54(11):4925–4945, Nov 2008.

[19] D. Henrici and P. Muller. Hash-based enhancement of location privacy for radio-frequency identification devices using varying identifiers. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 149–153, March 2004.

[20] Nicholas Hopper, Eugene Y. Vasserman, and Eric Chan-TIN. How much anonymity does network latency leak? *ACM Trans. Inf. Syst. Secur.*, 13(2):13:1–13:28, March 2010.

[21] Nicholas Hopper, Eugene Y Vasserman, and Eric Chan-Tin. How much anonymity does network latency leak? *ACM Transactions on Information and System Security (TISSEC)*, 13(2):13, 2010.

[22] Fei Huo, Chouchang Yang, Guang Gong, and Radha Poovendran. A framework to securing rfid transmissions by varying transmitted reader's power. In *Proceedings of the RFIDSec'13 Asia*, Cryptology and Information Security Series, pages 57–68, Nov 2013.

[23] Matthew O Jackson. A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, pages 11–49, 2005.

[24] A. Juels. Rfid security and privacy: a research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):381–394, Feb 2006.

[25] A. Juels. Rfid security and privacy: a research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):381–394, Feb 2006.

[26] Ari Juels. Minimalist cryptography for low-cost rfid tags (extended abstract). In Carlo Blundo and Stelvio Cimato, editors, *Security in Communication Networks*, volume 3352 of *Lecture Notes in Computer Science*, pages 149–164. Springer Berlin Heidelberg, 2005.

[27] Ari Juels. Rfid security and privacy: A research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):381–394, 2006.

[28] Ari Juels. Rfid security and privacy: A research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):381–394, 2006.

[29] Jiejun Kong, Xiaoyan Hong, and M. Gerla. An identity-free and on-demand routing scheme against anonymity threats in mobile ad hoc networks. *Mobile Computing, IEEE Transactions on*, 6(8):888–902, Aug 2007.

[30] Mikko Lehtonen, Thorsten Staake, and Florian Michahelles. From identification to authentication–a review of rfid product authentication techniques. In *Networked RFID Systems and Lightweight Cryptography*, pages 169–187. Springer, 2008.

[31] Tieyan Li and R. Deng. Vulnerability analysis of emap-an efficient rfid mutual authentication protocol. In *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*, pages 238–245, April 2007.

[32] Yiyuan Luo, Qi Chai, Guang Gong, and Xuejia Lai. A lightweight stream cipher wg-7 for rfid encryption and authentication. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–6. IEEE, 2010.

[33] Ralph C. Merkle. Secure communications over insecure channels. *Commun. ACM*, 21(4):294–299, April 1978.

[34] A. Mitrokotsa, M.R. Rieback, and A.S. Tanenbaum. Classification of rfid attacks. Proceedings of the 2nd International Workshop on RFID Technology - Concepts, Applications, Challenges (IWRT'08), 10th International Conference on Enterprise Information Systems, pages 73–86, Barcelona, Spain, June 2008. INSTICC Press, Portugal, INSTICC Press, Portugal.

[35] David Molnar and David Wagner. Privacy and security in library rfid: Issues, practices, and architectures. In *Proceedings of the 11th ACM Conference on Computer and Communications Security*, CCS '04, pages 210–219, New York, NY, USA, 2004. ACM.

[36] Pai Peng, Peng Ning, Douglas S. Reeves, and Xinyuan Wang. Active timing-based correlation of perturbed traffic flows with chaff packets. In *Proceedings of the Second International Workshop on Security in Distributed Computing Systems (SDCS) (ICDCSW'05) - Volume 02*, ICDCSW '05, pages 107–113, Washington, DC, USA, 2005. IEEE Computer Society.

[37] Pedro Peris-Lopez, Julio Cesar Hernandez-Castro, Juan M Estevez-Tapiador, and Arturo Ribagorda. Emap: An efficient mutual-authentication protocol for low-cost rfid tags. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pages 352–361. Springer, 2006.

[38] Pedro Peris-Lopez, Julio Cesar Hernandez-Castro, Juan M Estevez-Tapiador, and Arturo Ribagorda. M2ap: A minimalist mutual-authentication protocol for low-cost rfid tags. In *Ubiquitous Intelligence and Computing*, pages 912–923. Springer, 2006.

[39] Pedro Peris-Lopez, Julio Cesar Hernandez-Castro, Juan M Estevez-Tapiador, and Arturo Ribagorda. Lameda prng for epc class-1 generation-2 rfid specification. *Computer Standards & Interfaces*, 31(1):88–97, 2009.

[40] Pedro Peris-Lopez, Julio Cesar Hernandez-Castro, Juan ME Tapiador, and Arturo Ribagorda. Advances in ultralightweight cryptography for low-cost rfid tags: Gossamer protocol. In *Information security applications*, pages 56–68. Springer, 2009.

[41] RC-W Phan. Cryptanalysis of a new ultralightweight rfid authentication protocolsasi. *Dependable and Secure Computing, IEEE Transactions on*, 6(4):316–320, 2009.

[42] Zhu Qiuling, Zhang Chun, Liu Zhongqi, Wang Jingchao, Li Fule, and Wang Zhihua. A robust radio frequency identification system enhanced with spread spectrum technique. In *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, pages 37–40, May 2009.

[43] N.K. Ratha, J.H. Connell, and R.M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[44] Michael G Reed, Paul F Syverson, and David M Goldschlag. Anonymous connections and onion routing. *Selected Areas in Communications, IEEE Journal on*, 16(4):482–494, 1998.

[45] SanjayE. Sarma, StephenA. Weis, and DanielW. Engels. Rfid systems and security and privacy implications. In BurtonS. Kaliski, etinK. Ko, and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2002*, volume 2523 of *Lecture Notes in Computer Science*, pages 454–469. Springer Berlin Heidelberg, 2003.

[46] O. Savry, F. Pebay-Peyroula, F. Dehmas, G. Robert, and J. Reverdy. Rfid noisy reader how to prevent from eavesdropping on the communication? In Pascal Paillier and Ingrid Verbauwhede, editors, *Cryptographic Hardware and Embedded Systems - CHES 2007*, volume 4727 of *Lecture Notes in Computer Science*, pages 334–345. Springer Berlin Heidelberg, 2007.

[47] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal, Vol 28, pp. 656–715*, Octomber 1949.

[48] Yi-Sheng Shiu, Shih-Yu Chang, Hsiao-Chun Wu, S.C.-H. Huang, and Hsiao-Hwa Chen. Physical layer security in wireless networks: a tutorial. *Wireless Communications, IEEE*, 18(2):66–74, April 2011.

[49] P. Venkitasubramaniam, Ting He, and Lang Tong. Anonymous networking amidst eavesdroppers. *IEEE Trans. Inf. Theor.*, 54(6):2770–2784, June 2008.

[50] P. Venkitasubramaniam and Lang Tong. Throughput anonymity trade-off in wireless networks under latency constraints. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages –, April 2008.

[51] Roy Want. Enabling ubiquitous sensing with rfid. *Computer*, 37(4):84–86, 2004.

[52] R. Weinstein. Rfid: a technical overview and its application to the enterprise. *IT Professional*, 7(3):27–33, May 2005.

[53] StephenA. Weis, SanjayE. Sarma, RonaldL. Rivest, and DanielW. Engels. Security and privacy aspects of low-cost radio frequency identification systems. In Dieter Hutter, Gnter Mller, Werner Stephan, and Markus Ullmann, editors, *Security in Pervasive Computing*, volume 2802 of *Lecture Notes in Computer Science*, pages 201–212. Springer Berlin Heidelberg, 2004.

[54] Stallings William and William Stallings. *Cryptography and Network Security, 4/E*. Pearson Education India, 2006.

[55] Chouchang Yang, Basel Alomair, and Radha Poovendran. Multipath flow allocation in anonymous wireless networks with dependent sources. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1395–1400. IEEE, 2012.

[56] Chouchang Yang, Basel Alomair, and Radha Poovendran. Optimized flow allocation for anonymous communication in multipath wireless networks. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 219–223. IEEE, 2012.

[57] Chouchang Yang, Basel Alomair, and Radha Poovendran. Optimized relay-route assignment for anonymity in wireless networks. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 76–80. IEEE, 2013.

Appendix A

LIST OF PUBLICATIONS

1. 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.

2. 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.

3. C. Yang, B. Alomair, and R. Poovendran, "Optimized Relay-Route Assignment for Anonymity in Wireless Networks," in *Proc. the IEEE International Symposium on Information Theory (ISIT)*, pp. 76-80, July 2013.

4. F. Huo, C. Yang, G. Gong, R. Poovendran, "A Framework to Securing RFID Transmissions by Varying Transmitted Reader's Power," in *Proc. the RFIDSec' 13 Asia*, pp. 57-68, Nov. 2013.

VITA

Chouchang Yang was born in Taipei, Taiwan. He received his BS degree in the Department of Electrical Engineering and MS degree in Institute of Computer and Communication Engineering from National Cheng Kung University in 2004 and 2006 respectively. In 2011, he joined Network Security Lab at University of Washington and pursued his Ph.D degree. He is at present working towards her Ph.D degree in the field of communications and networking. Her current work focuses on applying information theory to anonymous networks, and physical layer security in RFID. His Ph.D thesis is titled Optimizable Design Schemes in Communication Systems for Providing Anonymity and Confidentiality. He welcomes your comments to ccjack@uw.edu

.