

COST-AWARE ROUTING PROTOCOLS FOR LOCATION-PRIVACY AND EFFICIENCY IN  
WIRELESS SENSOR NETWORKS

By

Leron J. Lightfoot

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Electrical Engineering – Doctor of Philosophy

2015

## ABSTRACT

### COST-AWARE ROUTING PROTOCOLS FOR LOCATION-PRIVACY AND EFFICIENCY IN WIRELESS SENSOR NETWORKS

By

Leron J. Lightfoot

Wireless sensor networks (WSNs) can provide the world with a technology for real-time event monitoring for both military and civilian applications. One of the primary concerns that hinders the successful deployment of wireless sensor networks is how to provide adequate source and destination nodes location privacy. The privacy of the location is vital and highly jeopardized by the usage of wireless communications. While message content privacy can be ensured through message encryption, it is much more difficult to adequately address the location privacy issue. For WSNs, location privacy service is further complicated by the fact that sensors consist of low-cost and energy efficient radio devices. Therefore, using computationally intensive cryptographic algorithms (such as public-key cryptosystems) and large scale broadcasting-based protocols are not suitable for WSNs.

Many protocols have been proposed to provide location privacy but most of them are based on public-key cryptosystems, while others are either energy inefficient or have certain security flaws. After analyzing the security weaknesses of the existing schemes, we propose several creative and secure energy-aware routing protocols that can address the location privacy issue in WSNs. For source-location privacy, we propose 3 schemes. The first scheme routes each message to a randomly selected intermediate node (RSIN) before it is transmitted to the SINK node. However, this scheme can only provide local source-location privacy. In the second scheme, a network mixing ring (NMR) is proposed to provide network-level source-location privacy. The third scheme achieves network-level source-location privacy through a technique we call the Sink Toroidal Region (STaR) routing. For destination-location privacy, we propose the Bubble routing protocol and a series of R-STaR routing protocols. For each of these routing schemes, both security analysis us-

ing quantitative measurements and simulation results show that the proposed protocols are secure and energy-efficient.

While providing location privacy is vital, prolonging the lifetime of the network can be a very essential component as well. In this dissertation, we propose a cluster-based energy-aware routing scheme, called Quad-Region Cluster-Head Selection (Q-ReCHS), which will prolong the network lifetime by evenly distributing the energy load among all the nodes. Our extensive simulation results on cluster-based routing demonstrates that our proposed Q-ReCHS scheme can outperform many of the existing schemes.

Dedicated to my family

## ACKNOWLEDGEMENTS

First and foremost, I would like to give a heartfelt, special thanks to my advisor, Dr. Jian Ren. He has not only provided me with guidance for several years throughout my graduate studies at Michigan State University, he has also been motivating, encouraging and a true friend during the process. With his patience and willingness to provide continuous support, Dr. Ren made this journey towards a Ph.D. possible for me.

I would like to gratefully and sincerely acknowledge my mentor, friend, and role model, Dr. Percy Pierre. Serving as one my committee members, Dr. Pierre has been there for me since the start of my graduate career. From academics guidance to personal development, Dr. Pierre has been a constant supporter and I will be forever in his debt. Also, I would like to express my gratitude to the rest my dissertation committee members, Dr. Subir Biswas and Dr. Jonathan Hall for their comments, suggestions, and time commitment during this entire process.

Lastly, I would like to thank my parents and siblings for their love and continuous support during my long academic career. Special thanks to my brother, Dr. Leonard Lightfoot, for paving the way and constantly pushing me to greater heights. I would like to thank my beautiful daughter, Ariann Lightfoot, for reminding me that the battle for greatness is bigger than myself. Also, I would like give major thanks my labmates and friends, past and current, for always providing me with advice and feedback on research topics, life lessons and for making my time at MSU a memorable and an enjoyable experience. Last and certainly not least, I would like to thank my beautiful, intelligent, encouraging, motivating, loving, supportive fiancée, Dr. Ashley Johnson, for her love and for always keeping me on track to accomplish my goals by picking me up when I did not always have the strength to do it myself.

# TABLE OF CONTENTS

LIST OF TABLES . . . . .	ix
LIST OF FIGURES . . . . .	x
KEY TO ABBREVIATIONS . . . . .	xiii
<b>CHAPTER 1 INTRODUCTION . . . . .</b>	<b>1</b>
1.1 Wireless Sensor Networks . . . . .	1
1.1.1 What are Wireless Sensor Networks? . . . . .	1
1.1.2 Issues with Wireless Sensor Networks . . . . .	2
1.1.3 Location Privacy in Wireless Sensor Networks . . . . .	3
1.1.3.1 Why is Source-Location Privacy Important in Wireless Sensor Networks? . . . . .	4
1.1.3.2 Why is Destination-Location Privacy Important in Wireless Sensor Networks? . . . . .	4
1.2 Overview of the Dissertation . . . . .	4
1.2.1 Major Contributions . . . . .	4
1.2.2 Structure . . . . .	5
<b>CHAPTER 2 SOURCE-LOCATION PRIVACY PROTECTION . . . . .</b>	<b>6</b>
2.1 Limitations with Existing Solutions . . . . .	6
2.1.1 Limitations with Existing Solutions for Source-Location Privacy . . . . .	6
2.1.2 Summary of Major Limitations for Location Privacy . . . . .	8
2.2 Network Models and Design Goals . . . . .	8
2.2.1 The System Model . . . . .	8
2.2.2 Adversarial Model . . . . .	9
2.2.3 Design Goals . . . . .	9
2.3 Proposed Research Directions for SLP . . . . .	10
2.3.1 Directions For Source-Location Protection . . . . .	10
2.4 Source-Location Privacy using Randomly Selected Intermediate Nodes (RSIN) . . . . .	11
2.4.1 Constrained RSIN Scheme . . . . .	11
2.4.2 Security Analysis for RSIN . . . . .	14
2.4.3 Totally Random RSIN Scheme . . . . .	15
2.4.4 Security Analysis for Totally Random RSIN . . . . .	26
2.4.5 Simulation Results and Performance Comparison . . . . .	27
2.5 Source-Location Privacy with Mixing Ring . . . . .	29
2.5.1 Constrained RSIN . . . . .	29
2.5.2 Network Mixing Ring (NMR) . . . . .	29
2.5.3 Forwarding to the SINK . . . . .	33
2.5.4 Security Analysis for Mixing Ring Routing . . . . .	33

2.5.5	Performance Analysis and Simulation Results . . . . .	36
2.6	Source-Location Privacy using STaR Routing . . . . .	42
2.6.1	Preliminary: Source-Location Privacy Evaluation Model . . . . .	42
2.6.2	STaR Routing Scheme . . . . .	46
2.6.3	Security Analysis for STaR Routing . . . . .	49
2.6.4	Performance Analysis and Simulation Results . . . . .	53
CHAPTER 3 DESTINATION-LOCATION PRIVACY PROTECTION . . . . .		57
3.1	Limitations with Existing Solutions for Destination-Location Privacy . . . . .	57
3.2	Network Models and Design Goals . . . . .	58
3.2.1	The System Model . . . . .	58
3.2.2	The Adversaries Model . . . . .	59
3.2.3	Design Goals . . . . .	60
3.3	Proposed Research Directions for Destination-Location Privacy . . . . .	60
3.3.1	Directions for Destination-Location Privacy . . . . .	60
3.4	Preliminary: Destination-Location Privacy Evaluation Model . . . . .	60
3.5	Destination-Location Privacy using Bubble Routing . . . . .	64
3.5.1	Security Analysis for Bubble Routing . . . . .	67
3.6	Destination Location Privacy using R-STaR Routing Schemes . . . . .	69
3.6.1	R-STaR Routing Protocol . . . . .	69
3.6.2	R-STaR Routing Protocol with Fake Destination Nodes . . . . .	74
3.6.3	R-STaR Routing Protocol mix with Cost-Aware Routing (CAR) . . . . .	76
3.6.4	Security Analysis for R-STaR Routing Schemes . . . . .	77
3.6.4.1	First Routing Phase . . . . .	77
3.6.4.2	Second Routing Phase . . . . .	78
3.6.5	Performance Analysis and Simulation Results . . . . .	82
CHAPTER 4 Q-RECHS CLUSTER-BASED ENERGY EFFICIENT ROUTING IN WIRE- LESS SENSOR NETWORKS . . . . .		85
4.1	Introduction . . . . .	85
4.2	Related Work . . . . .	85
4.3	Energy Consumption Model . . . . .	87
4.4	Proposed Q-ReCHS Cluster-Based Routing Scheme . . . . .	87
4.4.1	Q-ReCHS Network Model . . . . .	88
4.4.2	Initialization Stage . . . . .	89
4.4.3	Q-ReCHS Cluster-Head Selection Process . . . . .	90
4.4.4	Systems Analysis . . . . .	92
4.5	Simulation Results . . . . .	92
CHAPTER 5 CONCLUSION AND FUTURE RESEARCH . . . . .		103
5.1	Conclusion for Source-Location Privacy . . . . .	103
5.2	Conclusion for Destination-Location Privacy . . . . .	103
5.3	Conclusion for Cluster-Based Routing . . . . .	104
5.4	Future Work . . . . .	104

BIBLIOGRAPHY . . . . . 105



## LIST OF TABLES

Table 3.1	Example destination node location table with $n_f = 4$ . . . . .	76
Table 3.2	Table of variables and descriptions for analyzing security of the proposed R- STaR schemes. . . . .	81

## LIST OF FIGURES

Figure 2.1	Illustration of RSIN . . . . .	13
Figure 2.2	Intermediate nodes distribution for constrained RSIN scheme . . . . .	15
Figure 2.3	Message forwarding through intermediate node(s) . . . . .	17
Figure 2.4	Performance for single-intermediate node: Power consumption for different packet lengths . . . . .	18
Figure 2.5	Performance for single-intermediate node: Power consumption for different packet generation intervals . . . . .	19
Figure 2.6	Performance for single-intermediate node: Message latency for different packet lengths . . . . .	20
Figure 2.7	Performance for single-intermediate node: Message latency for different packet generation intervals . . . . .	21
Figure 2.8	Performance for single-intermediate node: Message delivery ratio for different packet lengths . . . . .	22
Figure 2.9	Performance for single-intermediate node: Message delivery ratio for different packet generation intervals . . . . .	23
Figure 2.10	Performance for single-intermediate node: Power consumption for different length of random path . . . . .	24
Figure 2.11	Performance for single-intermediate node: Message latency for different length of random path . . . . .	25
Figure 2.12	Performance for single-intermediate node: Message delivery ratio for different length of random path . . . . .	26
Figure 2.13	Grids Formation . . . . .	28
Figure 2.14	Illustrate of the first two phases routing . . . . .	30
Figure 2.15	Message transmission in the ring . . . . .	32
Figure 2.16	Ring selection in simulation setup . . . . .	35
Figure 2.17	Mixing Ring: Power consumption of normal nodes . . . . .	37

Figure 2.18	Mixing Ring: Power consumption of ring nodes . . . . .	38
Figure 2.19	Mixing Ring: Message latency . . . . .	39
Figure 2.20	Mixing Ring: Message delivery ratio . . . . .	40
Figure 2.21	Distribution of the STaR area . . . . .	47
Figure 2.22	Routing illustration of the STaR protocol . . . . .	50
Figure 2.23	The source location analysis of STaR routing scheme . . . . .	52
Figure 2.24	Performance of STaR routing: Power consumption . . . . .	54
Figure 2.25	Performance of STaR routing: Message latency . . . . .	55
Figure 2.26	Performance of STaR routing: Message delivery ratio . . . . .	56
Figure 3.1	Illustration of the Bubble Routing . . . . .	64
Figure 3.2	Illustration of the routing for messages in the DEEP scheme . . . . .	68
Figure 3.3	Distribution of the R-STaR area . . . . .	70
Figure 3.4	Circle Grid Layout Network Environment with radius $R_N$ . . . . .	79
Figure 3.5	Illustration for analyzing the security strength using R-STaR routing . . . . .	80
Figure 3.6	Message Latency . . . . .	83
Figure 3.7	Energy Consumption . . . . .	84
Figure 4.1	Q-ReCHS Network Regions: 100 nodes, 100m x 100m, BS located (50m,50m)	89
Figure 4.2	Simulation Network Environment: 100 nodes, 100m x 100m, BS located (50m,50m) . . . . .	93
Figure 4.3	100 Nodes in 100m x 100m Environment: First-To-Die . . . . .	94
Figure 4.4	100 Nodes in 100m x 100m Environment: Half-To-Die . . . . .	95
Figure 4.5	100 Nodes in 100m x 100m Environment: Last-To-Die . . . . .	96
Figure 4.6	150 Nodes in 150m x 150m Environment: First-To-Die . . . . .	97
Figure 4.7	150 Nodes in 150m x 150m Environment: Half-To-Die . . . . .	98
Figure 4.8	150 Nodes in 150m x 150m Environment: Last-To-Die . . . . .	99

Figure 4.9	200 Nodes in 200m x 200m Environment: First-To-Die . . . . .	100
Figure 4.10	200 Nodes in 200m x 200m Environment: Half-To-Die . . . . .	101
Figure 4.11	200 Nodes in 200m x 200m Environment: Last-To-Die . . . . .	102

## KEY TO ABBREVIATIONS

DDI	Destination-location Disclosure Index
DLP	Destination-Location Privacy
DSI	Destination-location Space Index
NDSI	Normalized Destination-location Space Index
NMR	network mixing ring
NSSI	Normalized Source-location Space Index
Q-ReCHS	Quad-Region Cluster Head Selection
R-STaR	Reverse - Sink Toroidal Region
RSIN	Routing to a Single Intermediate Node
SDI	Source-location Disclosure Index
SLP	Source-Location Privacy
SSI	Source-location Space Index
STaR	Sink Toroidal Region
TTL	Time-To-Live
WSNs	Wireless Sensor Networks

# CHAPTER 1

## INTRODUCTION

### 1.1 Wireless Sensor Networks

#### 1.1.1 What are Wireless Sensor Networks?

A wireless sensor network (WSN) is a network of sensor nodes (devices) that are connected through shared wireless communication mediums. These sensor nodes can be used to monitor environmental activities, such as temperature, movement, sound, pressure, etc. A sensor node is made up of many parts including a radio transceiver, a microcontroller, and an energy source, typically a battery or an embedded energy harvesting technique. The size, cost, and capabilities of sensor nodes can vary and all have trade-offs. For instance, if you want to use relatively cheap sensor nodes in the network, there will be a trade-off in capabilities, such as computational speed, memory, transmitting range, and security (e.g. cryptographic-system "cryptosystem"). Alternatively, if you want to use sensor nodes with many capabilities, there will be a trade-off in size and cost, such as using a larger battery and a pricer microcontroller.

A sensor node in the network that has detected an activity in the environment and has a message (packet) to transmit, we refer to that node as the source node, and the node in the network that the message is intended for, we refer to that node as the sink node, destination node or base station (BS). The source node forwards messages to the sink node wirelessly through the network using multi-hop routing techniques. Some of the most popular multi-hop routing techniques are flooding-based routing and shortest-path routing algorithms. In flooding-based routing algorithm, the source node will forward the message to each neighbor node (nodes within the transmission range) and each receiving node will forward the message to each of its neighbors until the message reaches all connected nodes in the network, hence flooding the network with the message. In shortest-path

routing algorithm, the source node forwards the message on the shortest-path (in terms of hop counts) to the sink node.

### **1.1.2 Issues with Wireless Sensor Networks**

Wireless sensor (WSNs) rely on wireless communications, which is by nature a broadcast medium and is more vulnerable to security attacks than its wired counterpart due to lack of a physical boundary. Jamming or radio interference attacks can be performed on the network by an adversary emitting signals to jam the shared wireless channel of the network.

Sensors in the network are meant to be low-cost and energy efficient devices. They are designed to be deployed in environments where they can be damaged or destroyed; thus, the cost of these sensor nodes should be at a minimum. Clients can simply deploy many wireless sensor nodes into an environment and monitor the activities in the environment from one central location. Sensor nodes are also built to be placed in environments where they can be unattended for lengthy periods of time. These sensors may be deployed in areas where it is impractical for humans to attend and maintain the sensors, thus, changing or recharging the batteries in the sensor devices is infeasible. For the purpose of preserving battery life, using intensive cryptographic algorithms, such as public-key cryptosystems, and the usage of powerful transmitters are not suitable for WSNs.

One of the focuses of this dissertation is cluster-based energy-aware routing protocols to maximize network lifetime. This focus is strictly on energy-efficient routing. One of the usage of WSN's is to provide real-time event monitoring functionality. To do its job, the device energy must be a high priority for these devices to perform its functionality.

The other focus of this dissertation is on security for location-privacy using energy-efficient routing techniques. Privacy has been an extensively studied topic in wireless sensor networks. One of the major and unsettled issues in privacy of WSNs is how to provide adequate routing-based location-privacy. Privacy in a network consists of not only the privacy of the message content but also the privacy of the nodes location privacy.

### 1.1.3 Location Privacy in Wireless Sensor Networks

Security attacks, such as, privacy threats have been an extensively studied topic in Wireless Sensor Networks (WSNs). Privacy threats can be classified into two categories: (i) Content-based privacy threats, and (ii) Context-based privacy threats. Content-based privacy threats relate to protecting the content of the message and can be protected using cryptographic encryption algorithms. Context-based privacy threats relate to monitoring the transmission of the data (i.e. Eavesdropping and localization attacks). For instance, an adversary may be able to analyze the traffic patterns by monitoring the transmission of the data using radio frequency localization techniques. Unfortunately, using cryptographic techniques does not provide protection for context-based threats and presents a much greater challenge to solve. In this dissertation, we will focus on one particular context-based threat, routing-based location privacy.

When messages are transmitted wirelessly in the open air, any compatible receivers within the transmission range of the sender is able to intercept the traffic. An adversary may be well-equipped with powerful transceivers to analyze the traffic patterns. They may be able to intercept traffic from one or multiple locations within the network environment. Without an adequate protection of the routing paths, an adversary may be able to perform context-based attacks on the network. An adversary may be able to determine the a node location by using radio frequency (RF) localization techniques to locate the source and/or destination node location in a hop-by-hop approach. The confidentiality of the message content can be protected by encryption and authentication [1–12, 12–25] but the location of the source and/or sink can be exposed in routing patterns. To be more concise, there may be different types of information besides the message content that are linked with a message transmission. Therefore, even if a powerful encryption algorithm is used to protect nodes identities, the adversary may still be able to determine the location of the source and/or sink node by monitoring the traffic patterns and routing paths.

In developing routing schemes to address the issue for location privacy, the routing schemes must be energy-efficient due to the limited energy supply on the sensor devices. Therefore, energy



consumption along with location privacy are two very vital components for the successful deployment of wireless sensor networks. In this dissertation, we propose routing schemes that address the location privacy issue by using unique routing protocols.

### **1.1.3.1 Why is Source-Location Privacy Important in Wireless Sensor Networks?**

The Panda-Hunter Game that was introduced in [26, 27], created interest for providing a secure source-location privacy in WSNs. In the Panda-Hunter Game, a wireless sensor network is deployed in a habitat to monitor the location of a panda. The sensors are used to locate the general area of the panda. As soon as the panda is discovered, the corresponding source node will observe and report data periodically to the SINK node. However, the source location should be kept secure from illegal hunters who may try to track and locate the panda. The goal is to make it infeasible for the hunters to determine the location of the panda by analyzing the traffic patterns in the network.

### **1.1.3.2 Why is Destination-Location Privacy Important in Wireless Sensor Networks?**

Wireless Sensor Networks can be deployed for military intelligence networks. On a battlefield, soldiers can be equipped with sensor devices, in which messages are routed to a destination base station that is located within the battlefield. To destroy this network, enemies of soldiers can simply destroy the destination base station. For the safety of the soldiers that are monitoring the destination base station, the location of the destination must remain unexposed as messages are routed through the network. Therefore, in military intelligence networks, both the destination-location and the message content must be protected.

## **1.2 Overview of the Dissertation**

### **1.2.1 Major Contributions**

The major contributions of this dissertation are the following:

- We develop to protect the source-location privacy through a two-phase routing process.
  - Source-location privacy through routing to a single Randomly Selected Intermediate Node (RSIN).
  - Source-location privacy through using STaR routing.
- We develop source-location privacy through a network-level mixing ring.
- We devise network implementation criteria for source node privacy protection in WSNs.
- We develop to protect the destination-location privacy through a two-phase routing process.
  - Destination-location privacy using Bubble routing.
  - Destination-location privacy using R-STaR routing.
  - Destination-location privacy using R-STaR with Fake Destination Nodes.
  - Destination-location privacy using R-STaR mix with Cost-Aware Routing.
- We designed cluster-based energy-aware routing scheme using Q-ReCHS algorithm.
- We provide extensive simulation results under ns-2 and MATLAB for the proposed routing protocols.

## 1.2.2 Structure

The dissertation is structured as follows:

- **Chapter II** introduces the schemes for source-location privacy protection.
- **Chapter III** introduces the schemes for destination-location privacy protection.
- **Chapter IV** introduces the cluster-based routing schemes.
- **Chapter V** summarizes the dissertation.

## CHAPTER 2

### SOURCE-LOCATION PRIVACY PROTECTION

#### 2.1 Limitations with Existing Solutions

##### 2.1.1 Limitations with Existing Solutions for Source-Location Privacy

In the past two decades, originated largely from Chaum's mixnet [28] and DC-net [29], a number of source-location privacy communication protocols have been proposed [27, 30–59]. The mixnet family protocols use a set of "mix" servers that blends received packets to make the communication source (including the sender and the recipient) ambiguous. To accomplish the ambiguity between the senders and recipients, the schemes uses statistical properties of background traffic. The DC-net family protocols [29, 32, 33] utilize secure multiparty computation techniques. However, both approaches require public-key cryptosystems and are not suitable for WSNs.

Multiple schemes were proposed to provide destination location privacy. In [36, 37], base station location privacy based on multi-path routing and fake messages injection was proposed. In this scheme, every node in the networks has to transmit messages at a constant rate. Another base station location privacy scheme was introduced in [60], which involves location privacy routing and fake message injection. In this dissertation, we will address the source-location privacy in wireless sensor networks.

In [38, 39], source-location privacy is provided through broadcasting that mixes valid messages with dummy messages. The main idea is to have each node transmit messages consistently. In other words, whenever there is no valid message (or no event detected), the node has to transmit dummy messages. The transmission of dummy messages not only consumes significant amount of sensor energy, but also increases the networks collisions and decreases the packet delivery ratio. Therefore, these schemes are not quite suitable for large scale sensor networks.

Dynamic two-phase routing based protocols can also provide source-location privacy to make it infeasible for an adversary to trace back to the source-location through traffic monitoring and analysis. The main idea is to, first, route the message to a random node that is away from the actual message source node during the first phase. Then forward the message to the SINK node using single path routing during the second phase. However, both theoretical and practical results demonstrate that if the message is routed randomly for  $h$  hops, then the message will be largely within  $h/5$  hops away from the actual source,

To solve this problem, several approaches have been proposed. In phantom routing protocol, introduced in [27,40], the message from the actual source will be routed to a phantom source along a designed directed walk through either sector-based approach or hop-based approach. Take the sector-based directed walk as an example, the source node first randomly determines a direction that the message will be sent to. The direction information is stored in the header of the message, in which, every forwarder on the random walk path will forward this message to a random neighbor in the same direction determined by the source node. In this way, the phantom source can be away from the actual source. Unfortunately, once the message is captured on the random walk path, the adversaries will be able to get the direction information stored in the header of the message. Therefore, the exposure of direction information decreases the complexity for adversaries to trace back to the actual message source in the magnitude of  $2^h$ . Random walk from both the source node and the SINK node was also proposed in [41]. In this scheme, Bloom Filter was proposed to store the information of all the visited nodes in the networks for each message to prevent the messages from hopping back. However, this design allows the adversaries to recover significant routing information from the received messages. In fact, this design is impractical for large scale sensor networks. In this chapter, we will propose several dynamic routing based protocols that solves the vulnerabilities of the discussed existing schemes.

### **2.1.2 Summary of Major Limitations for Location Privacy**

- Public-key based schemes are not suitable for location protection in WSNs due to the high computation and communication overhead.
- Broadcasting based location protection schemes are energy inefficient.
- Dynamic routing based source-location protection schemes are the most promising but many vulnerabilities exposed in the existing schemes.

## **2.2 Network Models and Design Goals**

To get a better understanding of the network, in this section we will provide the system model and adversarial model to capture the relevant features of WSNs and potential adversaries in location privacy applications.

### **2.2.1 The System Model**

The following assumptions are made about the system:

- The network is divided into grids. The sensor nodes in each grid are fully connected. In each grid, there is one header node responsible for communicating with other nearby header nodes. The whole network is fully connected through multi-hop communications [61–64].
- Every node in the network can become a source node on a detection of an event. On detecting an event, a sensor source node will generate and send messages to the destination node through a multi-hop routing.
- Each message will include a unique node ID where the event was generated. The SINK (destination) node can only determine source node location based off the node ID.
- The sensor nodes are assumed to know their relative location and destination node location. We also assume that each sensor node has the knowledge of its adjacent neighboring nodes.

The information about the relative location of the sensor domain may also be broadcasted through this network for routing information update [65,66].

- The key management, including key generations, key distribution and key update, is beyond the scope of this dissertation. However, the interested readers are referred to references such as [8,24,67,68].

### 2.2.2 Adversarial Model

The adversaries have the following characteristics:

- **Well-equipped:** The adversary does not need to worry about the energy consumption. The adversary also has adequate computation capability. On detecting an event message, he could determine the immediate sender of this message by analyzing the strength and direction of the signal he received. He is able to move to this sender's location without much delay. If needed, the adversary can compromise sensor nodes in the network.
- **Passive:** The adversaries carry out some passive attacks, which only involve eavesdropping work.
- **Traffic-monitoring:** We assume that the adversary is unable to monitor the entire network from one central location. The adversary is able to monitor the traffic in an area that is within one-hop transmission range.

### 2.2.3 Design Goals

Our design goals can be summarized as follows:

- The adversaries should not be able to determine source location information by analyzing the traffic pattern.

- The adversaries should not be able to get the location of the source even if they are able to monitor a certain area of the sensor network and compromise a few network nodes.
- Only the sink node is able to identify the source-location with the messages received. The recovery of the source-location from the received message should be very efficient.
- The length of each message should be as short as possible to save the previous sensor node power. This is because that on average, transmission of one bit consumes about as much power as executing 800-1000 instructions [69].

## **2.3 Proposed Research Directions for SLP**

### **2.3.1 Directions For Source-Location Protection**

In this dissertation, we propose to address the source-location privacy through dynamic routing schemes. There are three schemes introduced in this dissertation.

In the first routing scheme, the message source randomly selects an intermediate node (RSIN) in the sensor domain, and then transmits the message to the randomly selected intermediate node before the message is transmitted to the SINK node. The intermediate node is expected to be away from the source node area in the sensor domain. Our analysis shows that this scheme can provide great local source-location privacy. However, it may not be able to provide adequate global source-location privacy.

To further improve the performance of global security, the second routing scheme using a three-phase routing process. In the first phase, each message is transmitted to a randomly selected intermediate node. This phase provides the local source-location privacy. In the second routing phase, the data packet will be routed from the randomly selected intermediate node to a network mixing ring (NMR). This phase offers network-level (global) source-location privacy. In the last phase, the data packet will be forwarded to the SINK node from specific nodes within the mixing ring.

The third scheme achieve global source-location privacy by routing through an intermediate node from a pre-determined region located around the SINK node. We call this region the Sink Toroidal Region (STaR). From the random intermediate node, the message will be routed to the SINK node through the shortest path routing. The STaR routing method is performed for every message the source node generate and send to the SINK node. For each of the routing schemes, both security analysis and simulation results are provided.

## 2.4 Source-Location Privacy using Randomly Selected Intermediate Nodes (RSIN)

In this section, we will describe the propose RSIN scheme. Also, we will provide analytical performance of the scheme through simulations.

### 2.4.1 Constrained RSIN Scheme

The RSIN routing protocol will use a two-phase routing technique. In the first phase, the source node will route each message through an intermediate node, which will be selected randomly within the network. To solve some of the vulnerabilities of the existing schemes, the intermediate node should be a minimum distance from the source node based on its relative location within the network environment. We refer to this minimum distance as  $d_{min}$ . We refer to the distance between the source node and the randomly selected intermediate node as  $d_{rand}$ . Therefore, for the intermediate node to be minimum distance of  $d_{min}$  from the source node, we will need  $d_{rand} \geq d_{min}$ .

For a source node to generate  $d_{rand}$ , it first generates a random variable  $x$  that normally distributed with mean 0 and variance  $\sigma^2$ , i.e.,  $X \sim N(0, \sigma)$ . With the random variable  $x$  determined, the source node can calculate  $d_{rand}$  as follows:

$$d_{rand} = d_{min} \times (|x| + 1).$$



Therefore, the probability [70] that  $d_{rand}$  is located in the interval  $[d_{min}, \rho d_{min})$  is:

$$2\varphi_{0,\sigma^2}(\rho - 1) - 1 = 2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\rho-1)^2}{2\sigma^2}} - 1 = 2\varphi\left(\frac{\rho-1}{\sigma}\right) - 1,$$

where  $\rho$  is a parameter larger than 1,  $\varphi_{0,\sigma^2}$  is the probability density function of Gaussian distribution [71]. If we choose  $\sigma$  to be 1.0, then the probability that  $d_{rand}$  falls within the interval  $[d_{min}, 2d_{min})$  will be  $2\Phi(\frac{1}{1}) - 1 = 0.6827$ . The probability that  $d_{rand}$  is in the interval  $[d_{min}, 3d_{min})$  will be  $2\Phi(\frac{2}{1}) - 1 = 0.9545$ .

After  $d_{rand}$  is determined, the source node randomly generates an intermediate node located at  $(x_d, y_d)$  that satisfies:

$$d_{rand} = \sqrt{(x_d - x_0)^2 + (y_d - y_0)^2} \geq d_{min}.$$

Since we assume that each sensor node only has knowledge of its adjacent nodes, the source node may not have accurate information of sensor nodes multiple hops away. In particular, the randomly selected intermediate node location  $(x_d, y_d)$  may not even exist in the network environment. However, the knowledge of relative location guarantees that the message packet will be forwarded to an intermediate node located with minimum distance  $d_{min}$  away from the source node. According to our assumption, the last node in the routing path adjacent to the intermediate node will be able to tell whether such a randomly selected intermediate node exists or not. In the case that such a node does not exist, this node will become the intermediate node, in which, this node will then route the received message to the SINK node. Upon receiving a data message, the intermediate node forwards the message to the SINK node.

In the explanatory example given in Figure 2.1,  $S_1, S_2$  denote two source nodes in the sensor network,  $D$  represents the SINK node and  $I_1, \dots, I_6$  are six randomly selected intermediate nodes that meet the constrained requirement. The selection of  $d_{rand}$  guarantees that none of the intermediate nodes will be located in the shaded areas. The nodes  $I_1, \dots, I_6$  will forward the messages  $M_1, \dots, M_6$  to the SINK node, respectively.

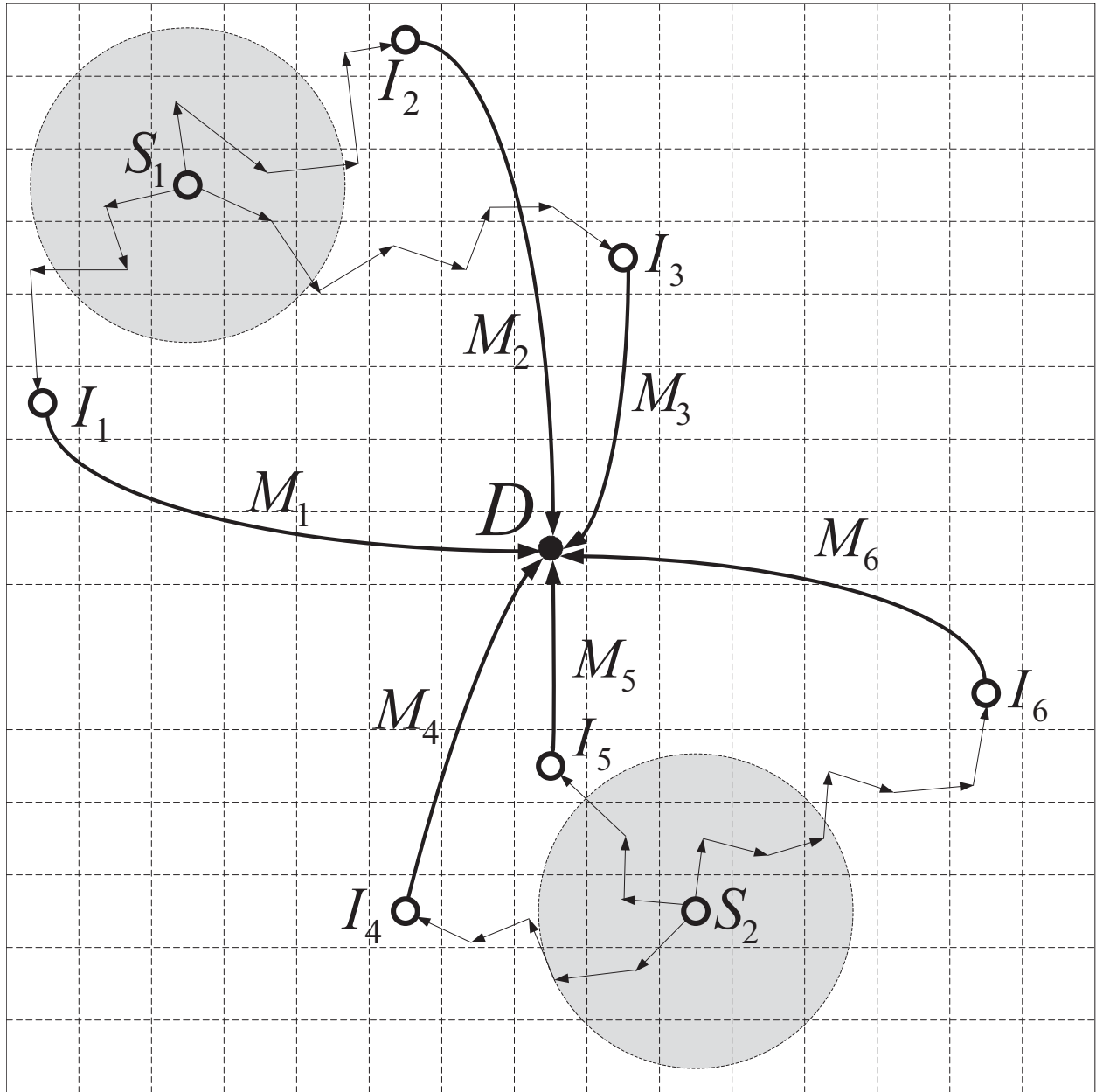


Figure 2.1 Illustration of RSIN

## 2.4.2 Security Analysis for RSIN

In the constrained RSIN scheme, the intermediate node location is randomly determined by the source node based on its relative location within the sensor domain. From probability point of view, every node outside of the minimum distance  $d_{min}$  radius from the source node can be selected as the intermediate node. However, since we assume that the source node does not have full knowledge of sensor nodes more than one hop away from itself, the intermediate node selected by the source node may not even exist.

It is impossible for the adversary to trace or identify the real message source node based on an individual traffic monitoring. This is because (i) this message is equally likely to be generated by many possible sources, and (ii) the probability for multiple events from the same source to use repeated routing path is very low for large scale sensor networks.

If an adversary tries to trace the source-location from a message packet in the route through which the packet is being transmitted, then the adversary will be led to the randomly selected intermediate node, instead of the real message source. Since the intermediate node is randomly selected for each data message, the probability that the adversaries will receive the messages from one source node continuously is virtually zero in a large scale network.

As shown in Figure 2.1, if an adversary receives  $M_2$  forwarded from  $I_2$ , the adversary will be lured to the direction of  $I_2$ , which is quite away from the actual source node  $S_1$ . While for message  $M_3$  transmitted from the intermediate node  $I_3$ , since it is far away from  $I_2$ , the probability for the adversary to receive  $M_3$  is close to zero according to the assumption.

This example shows that even if the location of one intermediate node is discovered by an adversary, the source-location is still at least  $d_{min}$  away from the real source node. Therefore, if  $d_{min}$  is appropriately selected, the source-location can still be well protected.

Unlike the directed walk in phantom routing, our protocol does not leak side information in the header to the adversaries. Since the intermediate node is determined before each data message is transmitted by the source, the data message carries no observable side information of the message

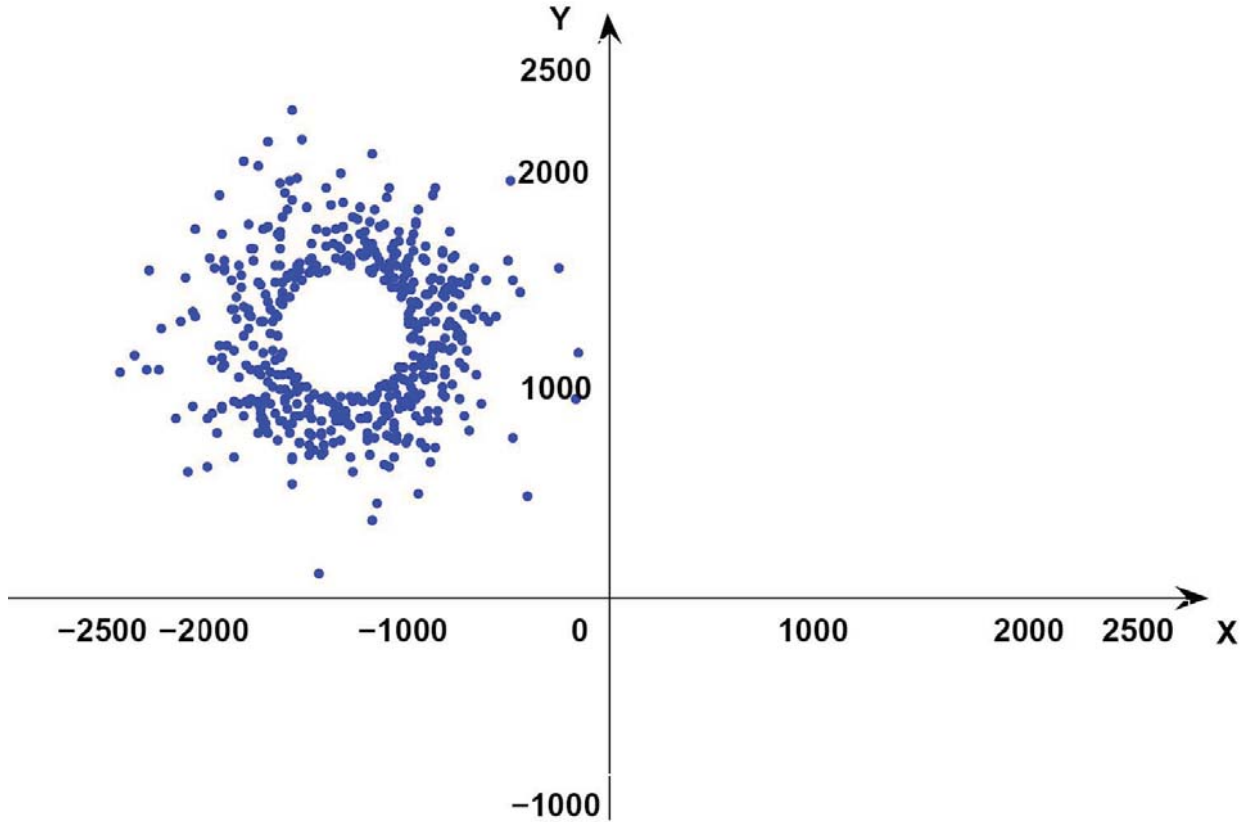


Figure 2.2 Intermediate nodes distribution for constrained RSIN scheme

source-location in its content due to message content encryption. Therefore, our proposed protocol can protect source-location privacy.

### 2.4.3 Totally Random RSIN Scheme

Although the constrained RSIN scheme works well in some scenarios, its limitation is that the probability for an intermediate node to be selected is proportional to the distance from the source node. Therefore, for large scale sensor networks, the intermediate nodes tend to be relatively close to the source node. In other words, the intermediate nodes are highly likely to be concentrated in an area surrounding the source node, but with minimum distance  $d_{min}$  from the source, as illustrated in Figure 2.2. In this simulation example, we have a network environment size of  $5000 \times 5000$  meters. We set the source node is located at  $(-1250, 1250)$  and the SINK node located at  $(0, 0)$ .

Also, we set the minimum distance to be  $d_{min} = 250$ . We randomly selected 500 intermediate nodes location according to the constrained RSIN scheme as described in Section 2.4.1 with  $\sigma$  equal to 1. It can be seen that all intermediate nodes are distributed around the source node. When these nodes forward messages to the SINK node. The adversaries may very likely be able to locate the sub-area of intermediate nodes since all messages are routed from the sub-area as shown in Figure 2.2. Therefore, for large scale sensor networks, the constrained RSIN scheme may not be able to provide adequate global source-location privacy but does provide local source-location privacy.

In order to provide global location privacy over the sensor networks, the selection of intermediate nodes has to be totally random, i.e., every sensor node in the networks should be equally likely to be selected as an intermediate node by all possible source nodes. On the other hand, if the selection is totally random, some intermediate nodes can be near the real source node. Fortunately, the probability for this is very low for large scale sensor networks. Nevertheless, to prevent this from happening, in totally random RSIN scheme, the intermediate nodes is requires to be at least  $d_{min}$  away from the real source node.

Although totally random RSIN scheme can achieve global location privacy, it also has some limitations:

- The length of a routing path tends to be too long. For instance, in Figure 2.3,  $S, D, I$  are the source node, SINK node and intermediate node, respectively. The distance between  $S, D$  and  $I, D$  are  $d$  and  $b$ , respectively. Therefore, if a message is transmitted through  $I$ , the total length of the routing path is nearly  $d + 2b$ , which is much longer than  $d$ . As a result, this routing may consume too much energy.
- The message delivery ratio may decrease due to increase of the routing length.
- A long single routing path may allow adversaries to deduce information of the source-location. Take the routing path from  $S$  to  $I$  in Figure 2.3 as an example, once a packet is

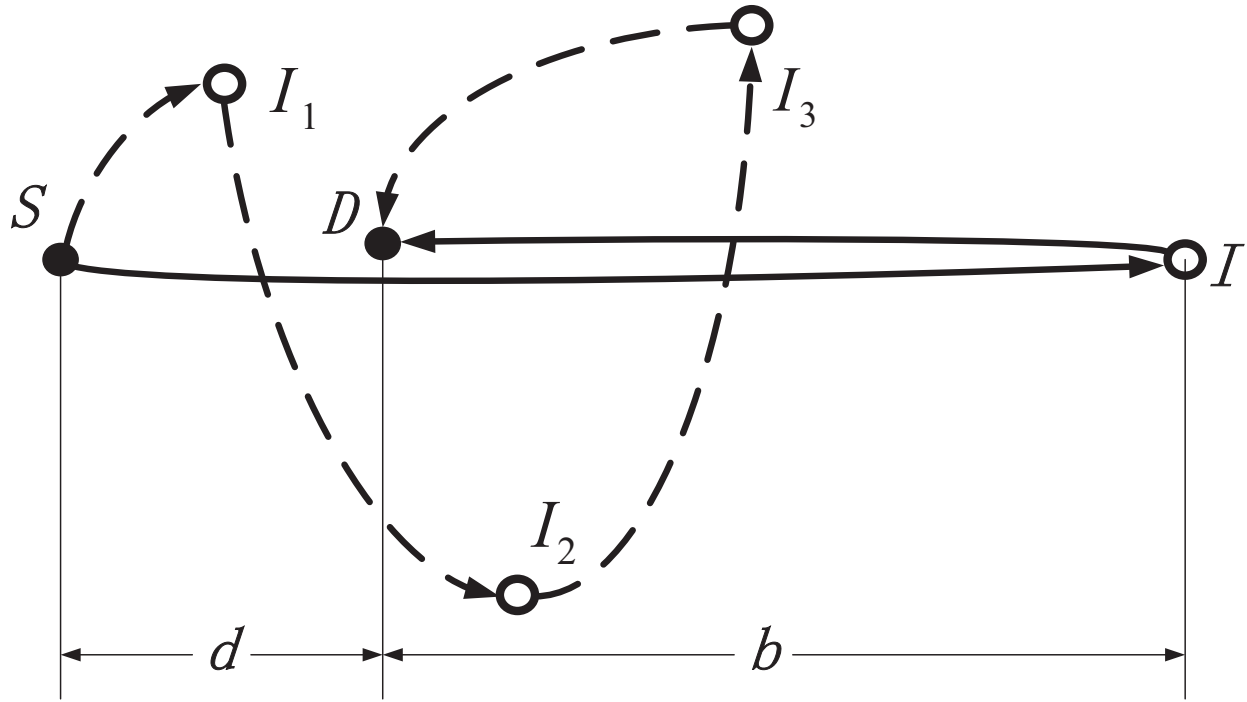


Figure 2.3 Message forwarding through intermediate node(s)

captured by adversaries en-route, the adversaries may get the direction of the source-location according to transmission direction of the captured packet.

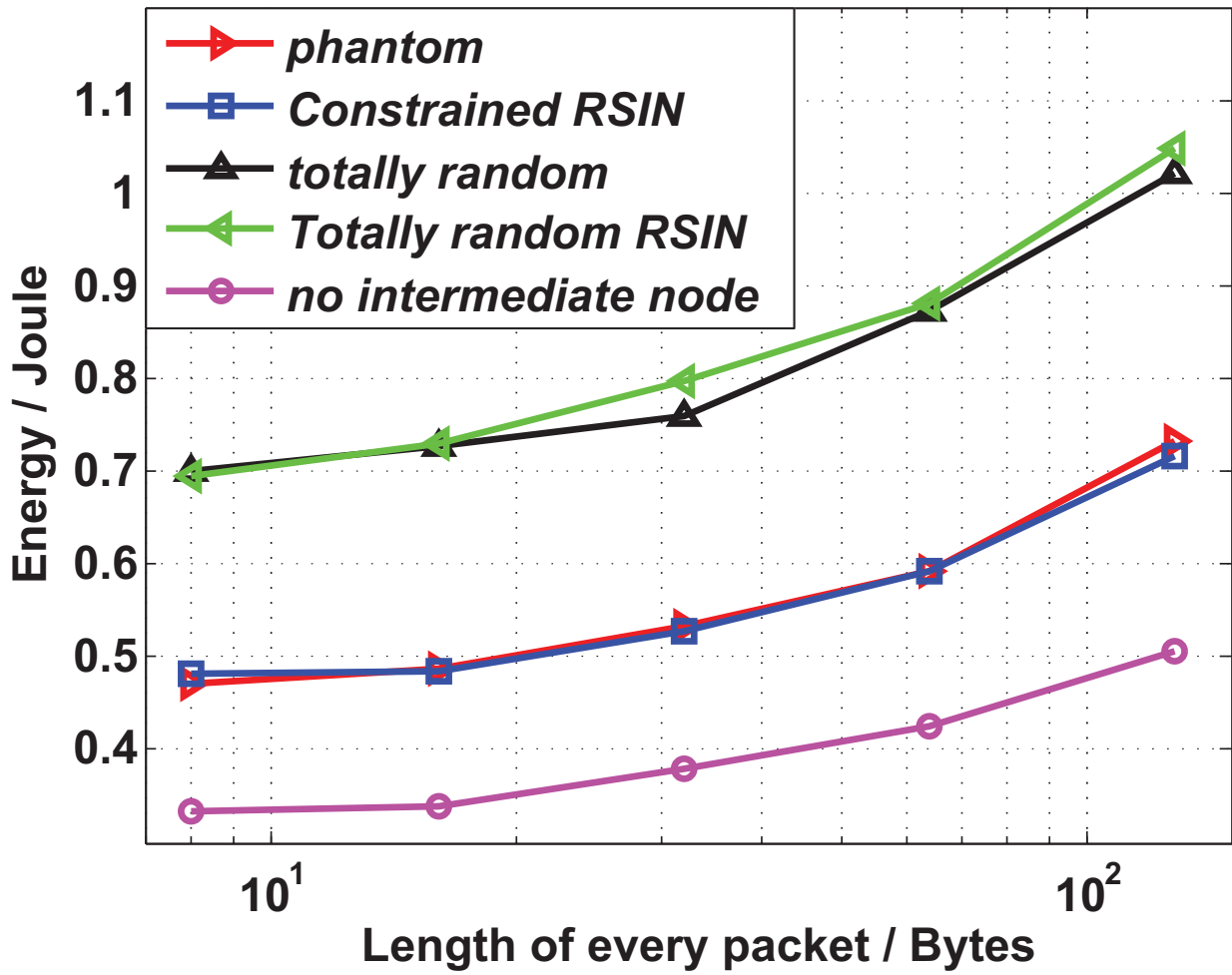


Figure 2.4 Performance for single-intermediate node: Power consumption for different packet lengths

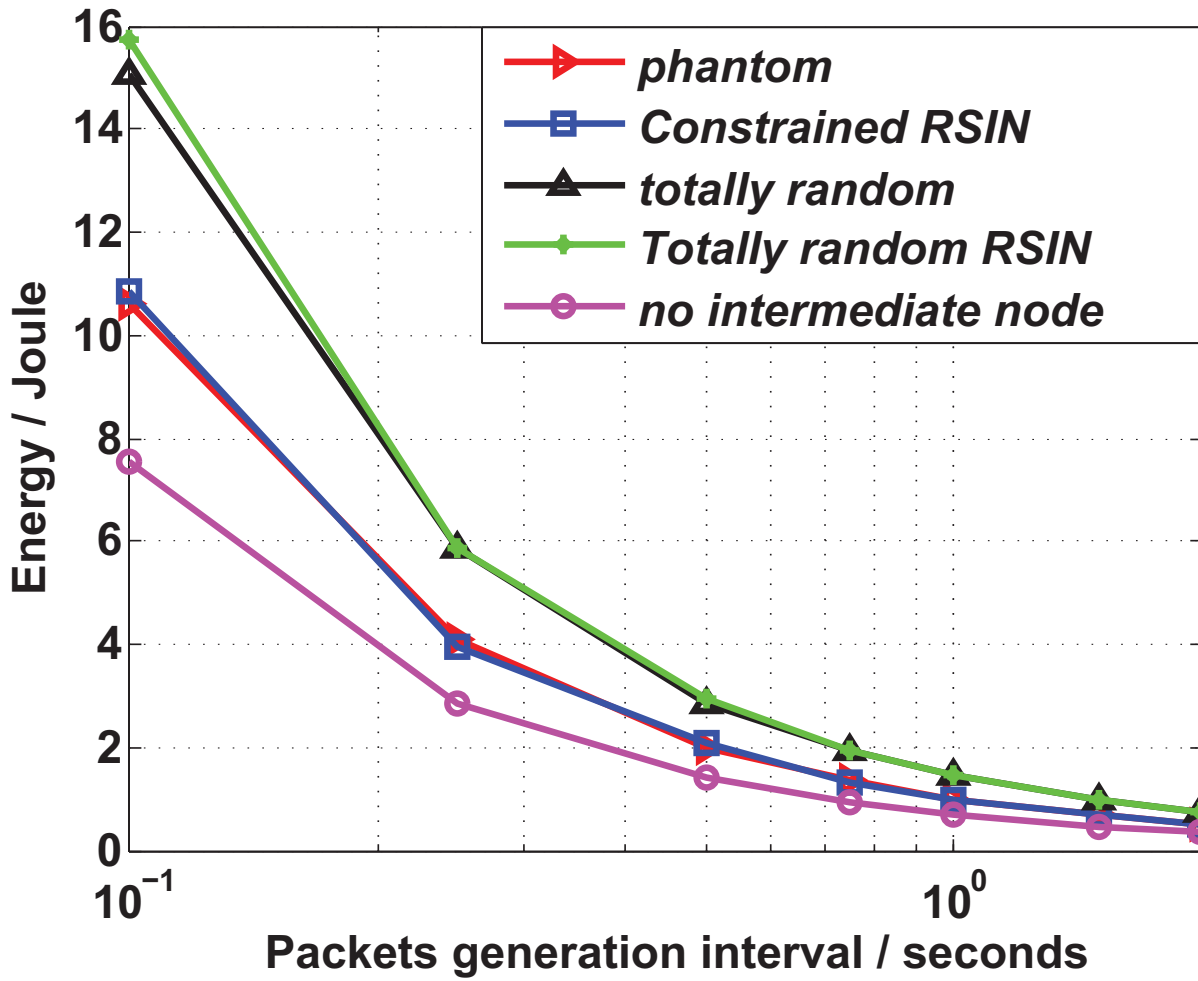


Figure 2.5 Performance for single-intermediate node: Power consumption for different packet generation intervals



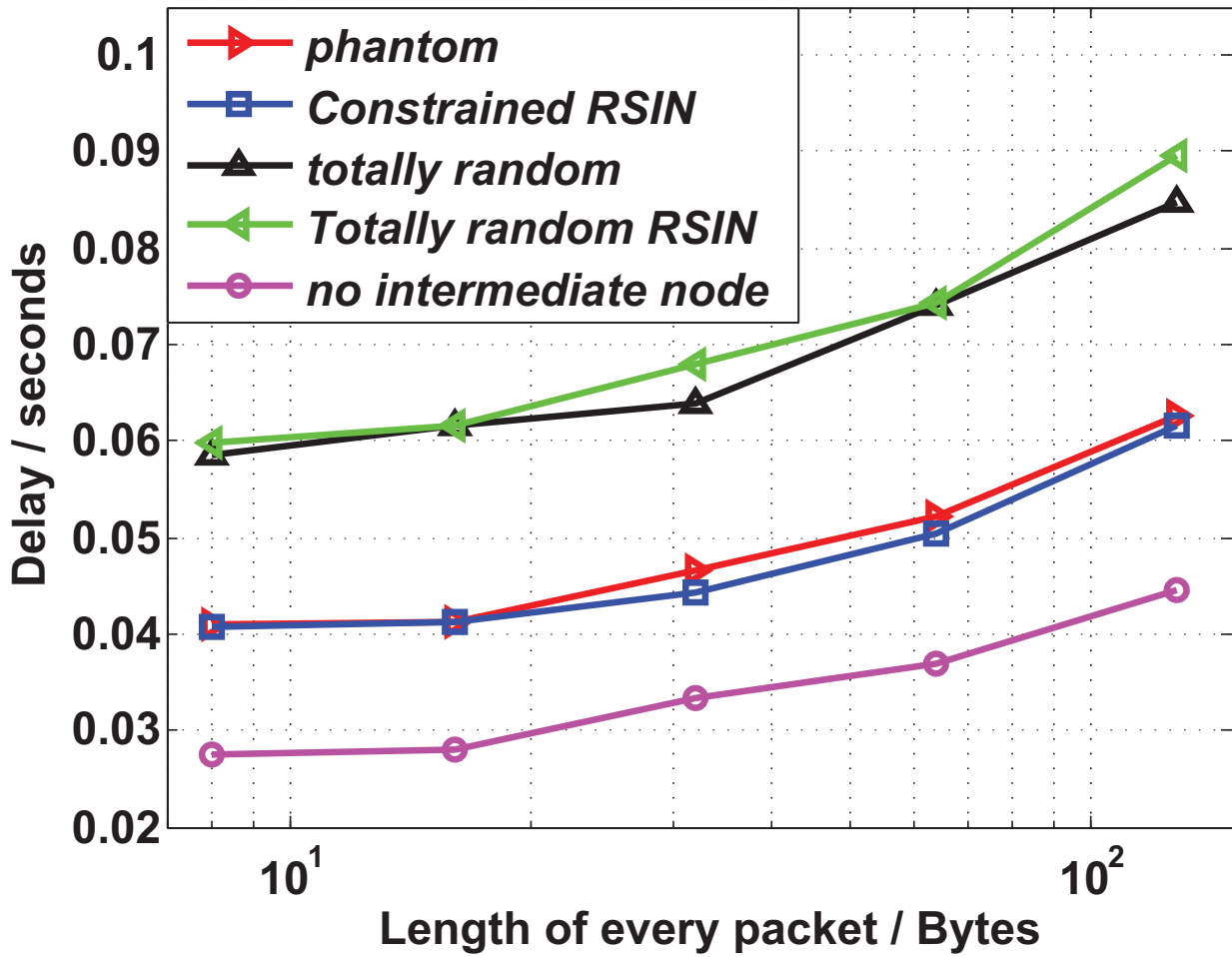


Figure 2.6 Performance for single-intermediate node: Message latency for different packet lengths

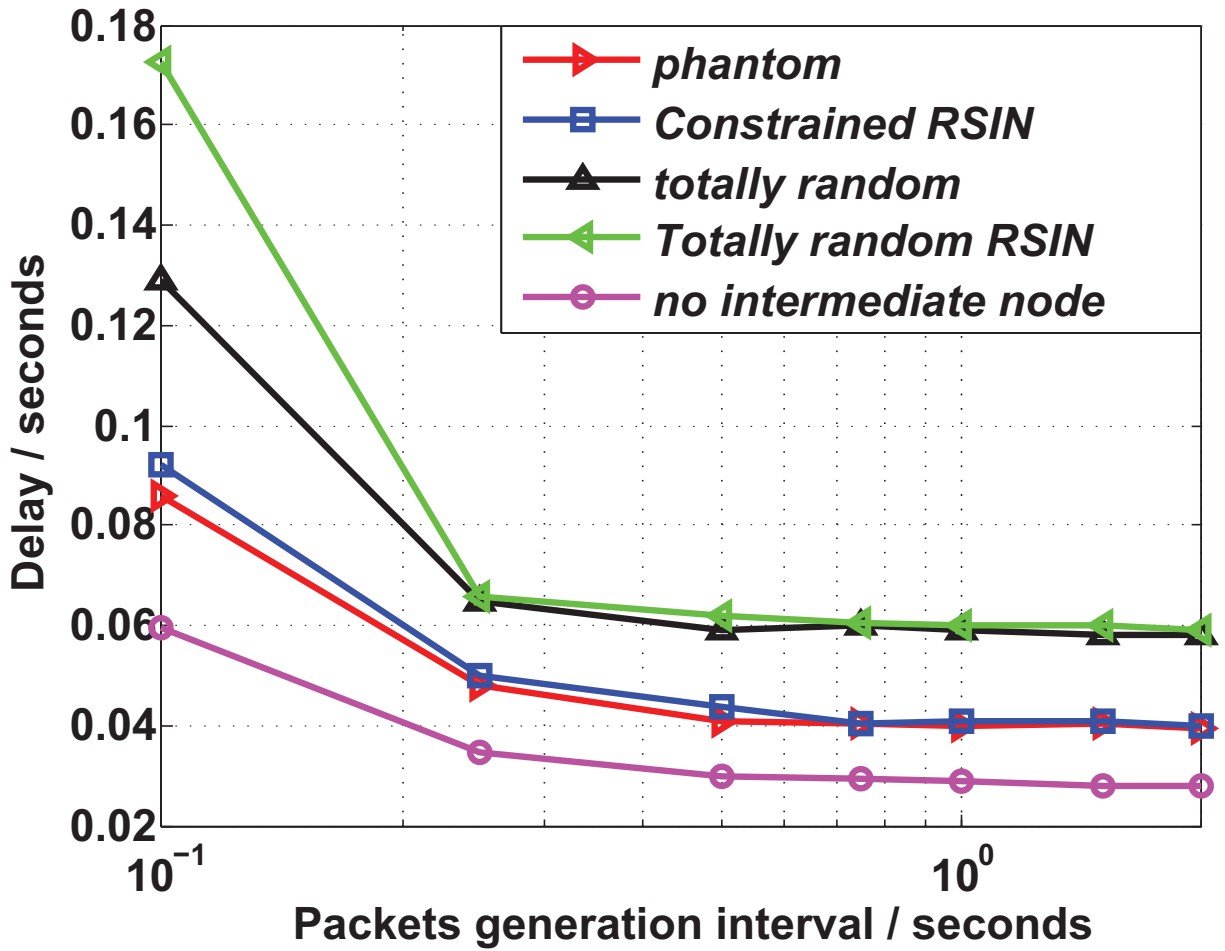


Figure 2.7 Performance for single-intermediate node: Message latency for different packet generation intervals

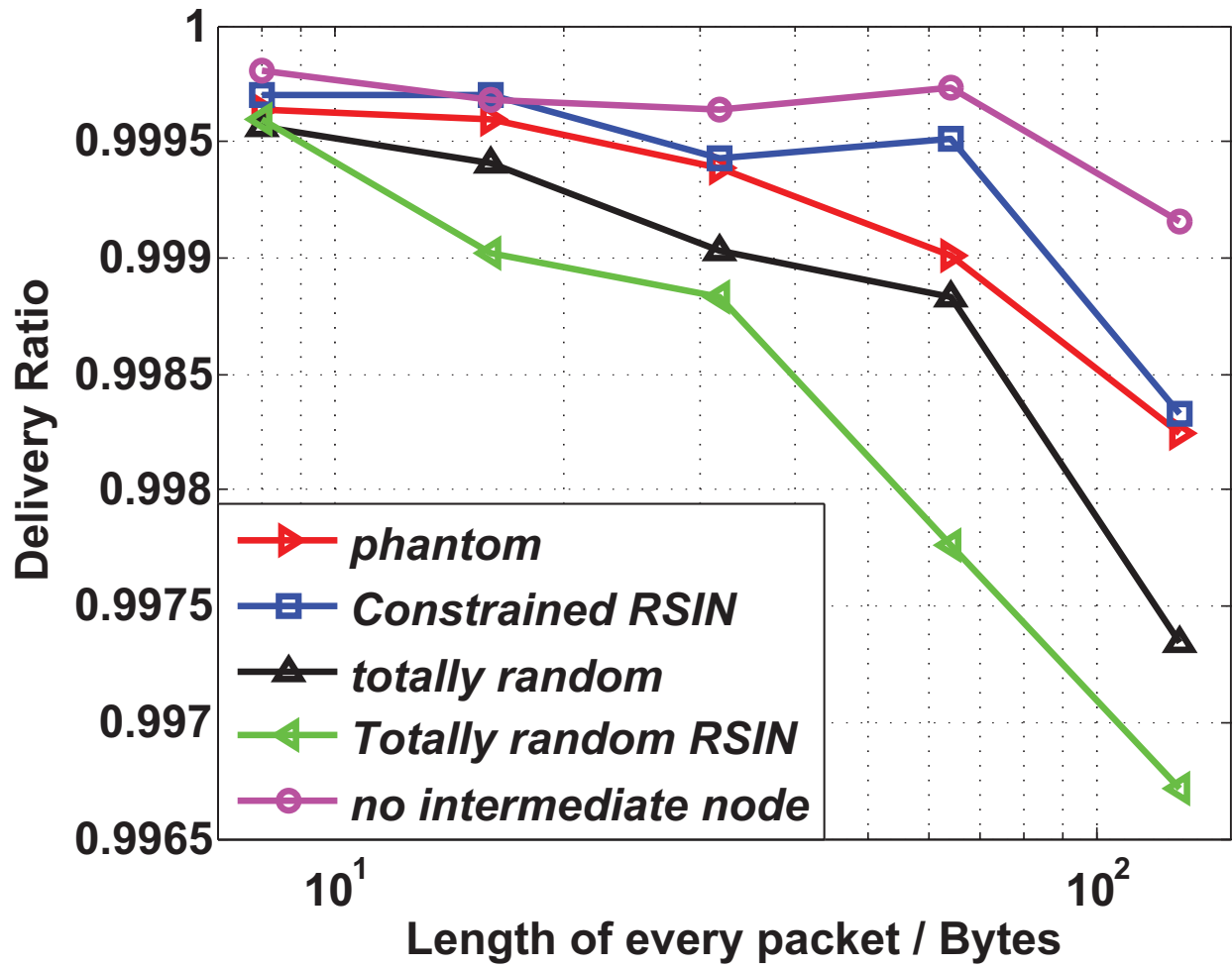


Figure 2.8 Performance for single-intermediate node: Message delivery ratio for different packet lengths

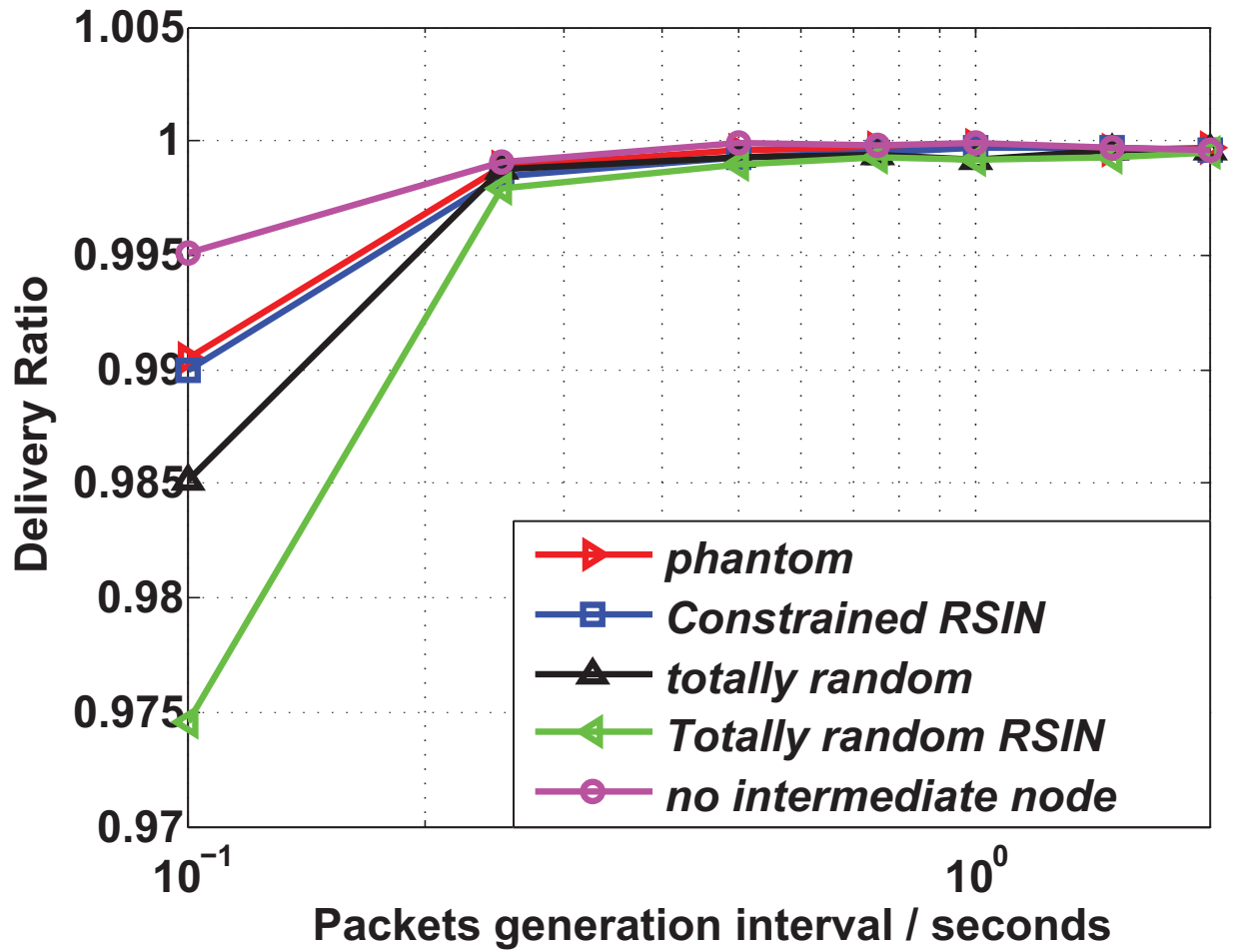


Figure 2.9 Performance for single-intermediate node: Message delivery ratio for different packet generation intervals

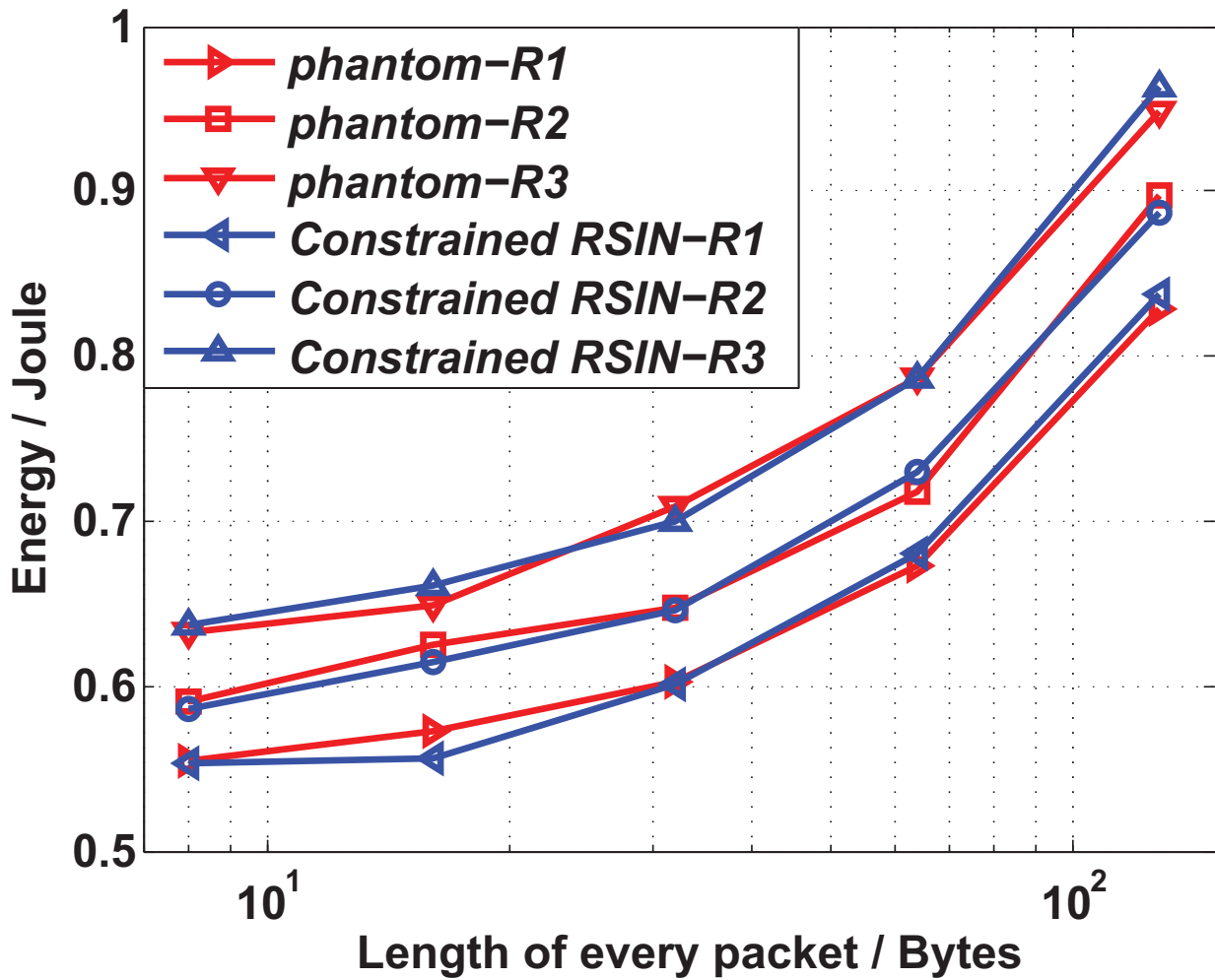


Figure 2.10 Performance for single-intermediate node: Power consumption for different length of random path

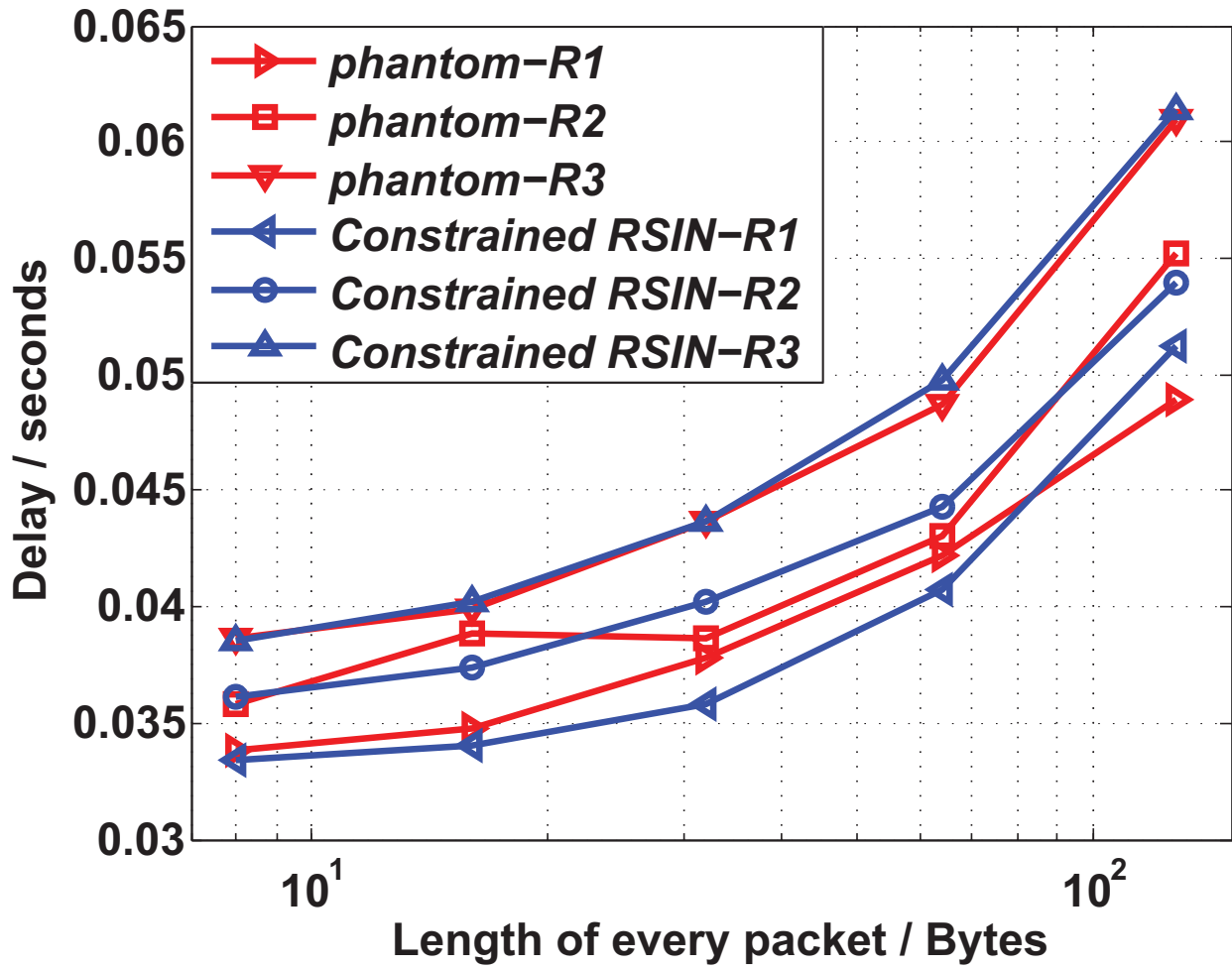


Figure 2.11 Performance for single-intermediate node: Message latency for different length of random path

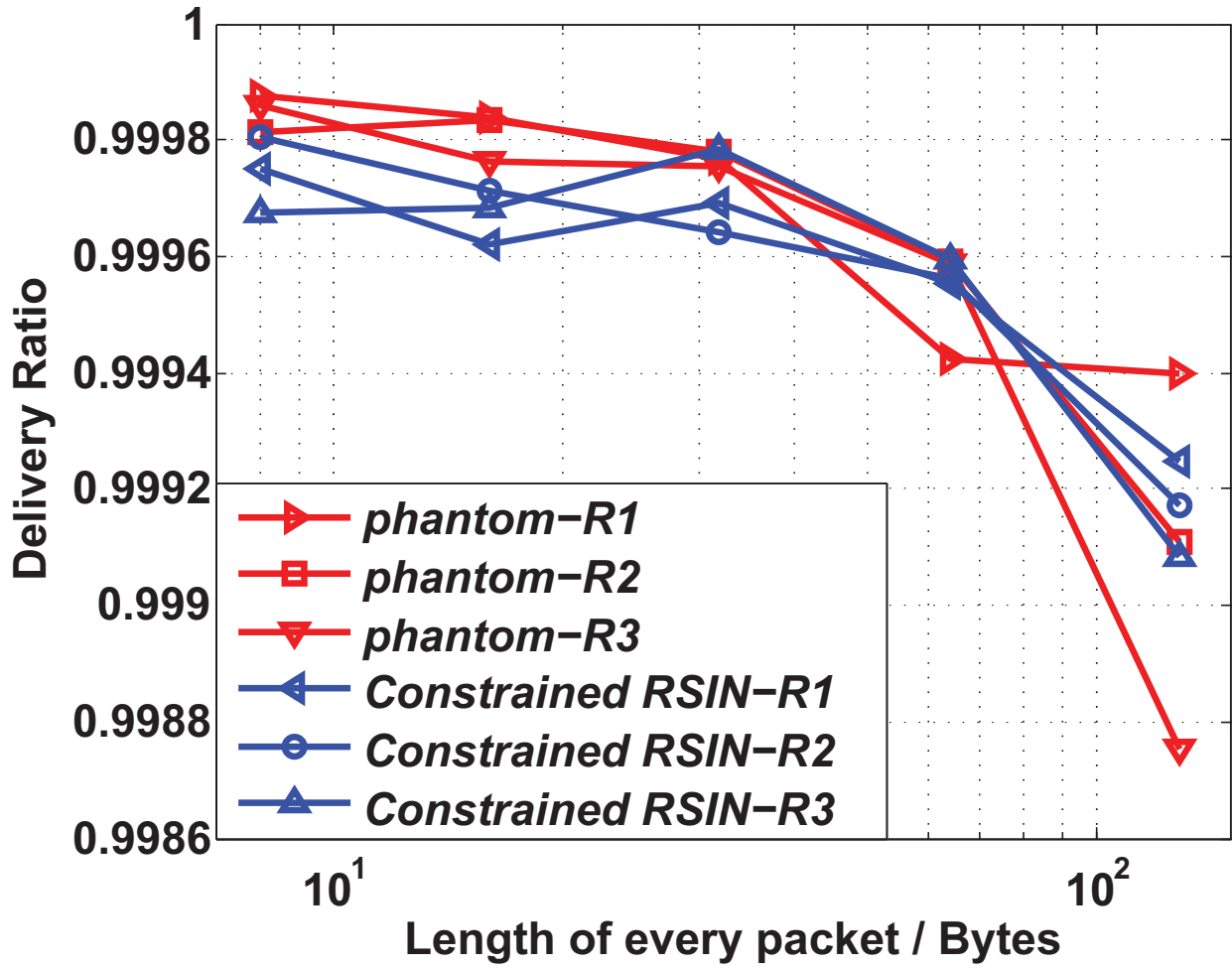


Figure 2.12 Performance for single-intermediate node: Message delivery ratio for different length of random path

#### 2.4.4 Security Analysis for Totally Random RSIN

Unlike constrained RSIN scheme, in totally random RSIN scheme, the intermediate nodes that are selected in the sensor network are evenly distributed. Every node outside of the  $d_{min}$  in the networks has the same possibility of being selected as the intermediate node. The messages can be forwarded to the SINK node from all possible directions. Even if the location of one intermediate node is successfully identified, the source-location is still at least  $d_{min}$  distance away. Therefore, the global location privacy is achieved.

### 2.4.5 Simulation Results and Performance Comparison

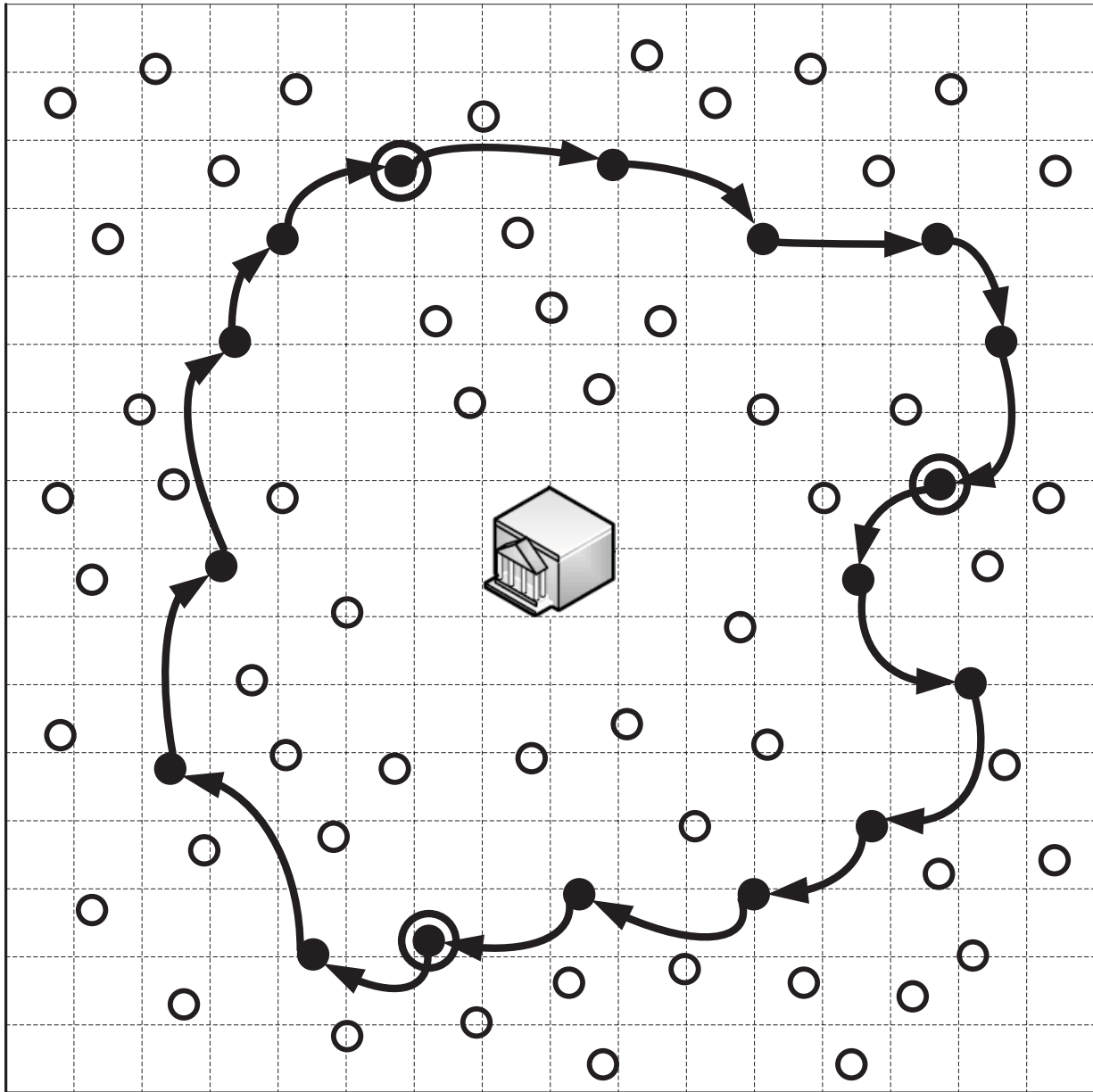
To evaluate the performance of our propose 'constraint RSIN' and 'totally randomly RSIN', we conduct simulations using ns-2 on RedHat Linux system. In the simulation, 400 nodes are distributed in an area of size  $3360 \times 3360$  meters. The SINK node is located at the center of the network.

Simulation results are provided in Figure 2.4, 2.6, 2.8, 2.10, where (a), (c), (e) illustrate the relationship between performance and the packet lengths, (b), (d), (f) show performance with different packet generation intervals, (g), (h), (i) show performance with different length of random path. For the simulation figures, 'totally random' represent a scheme that is similar to totally random RSIN but without restriction to  $d_{min}$  for intermediate node location selection. Let 'phantom' represent the phantom routing discussed in the related work section.

For simulation results from (a) to (f), we set  $d_{min} = 480$  meters for constrained RSIN and totally random RSIN schemes. The simulation shows that after four hops, the average distance between the phantom source node and the real source node for phantom routing is 526.12 meters. While for constrained RSIN scheme, the average distance between the intermediate node and the source is 529.14 meters. For simulation results from (g) to (i),  $R1$ ,  $R2$ ,  $R3$  for phantom routing corresponds to 526.12 meters, 783.60 meters, and 1042.20 meters on average between the phantom source node and the real source node, respectively. For constrained RSIN,  $R1$ ,  $R2$ ,  $R3$  corresponds to 529.14 meters, 786.51 meters, and 1049.46 meters on average between the intermediate node and the source node, respectively.

Through analysis and simulation results, we have 2 findings: (i) Direct routing without intermediate node has the best performance; (ii) Constrained RSIN and phantom routing have comparable performance, while constrained RSIN scheme provides better location privacy protection due to no direction exposure in the header.





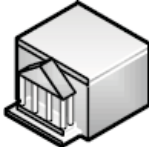
- Normal Node
- Normal Ring Node
- ⊙ Relay Ring Node
-  SINK

Figure 2.13 Grids Formation

## 2.5 Source-Location Privacy with Mixing Ring

In this section, we propose a three-phase routing protocol to provide source-location privacy. The first phase (constrained RSIN), which has been introduced in the last section, provides local source-location privacy. The second phase (NMR) offers the network-level source-location privacy. The last phase forwards the message to the SINK node.

After the formation of all the grids, a large ring, called the *mixing ring*, is generated in the WSN to provide network-level traffic mix. The mixing ring is composed of multiple header nodes, which are named *ring nodes*. The ring nodes are further divided into *relay ring nodes* and *normal ring nodes*. The messages that will be transmitted in the mixing ring are referred to as *vehicle messages*. Vehicle messages will be transmitted in the ring in the clockwise direction, called *ring direction*. Only relay ring nodes can generate vehicle message. We also define the grids containing ring node as *ring grids*, the grids without ring nodes as *normal grids*. The sensor nodes in normal grids are defined as *normal nodes*, the messages sent by the normal nodes are referred as *data messages*.

### 2.5.1 Constrained RSIN

In this phase, the message will be forwarded to an intermediate node in the same way as the constrained RSIN introduced in the last section. Then the message will be forwarded to the nearest ring node by this intermediate node.

An example is given in Figure 2.14, where  $S$  indicates a source node in the network and  $I_1, I_2, I_3$  are three intermediate nodes. The selection of  $d_{rand}$  guarantees that none of the intermediate nodes will be in the shaded area. Then  $I_1, I_2, I_3$  will forward these messages  $M_1, M_2, M_3$  to the ring nodes  $R_1, R_2, R_3$ , respectively.

### 2.5.2 Network Mixing Ring (NMR)

In the second routing phase, the messages will be forwarded hop-by-hop in the ring. The message can hop along the ring direction for a random number of times before it is being transmitted to the

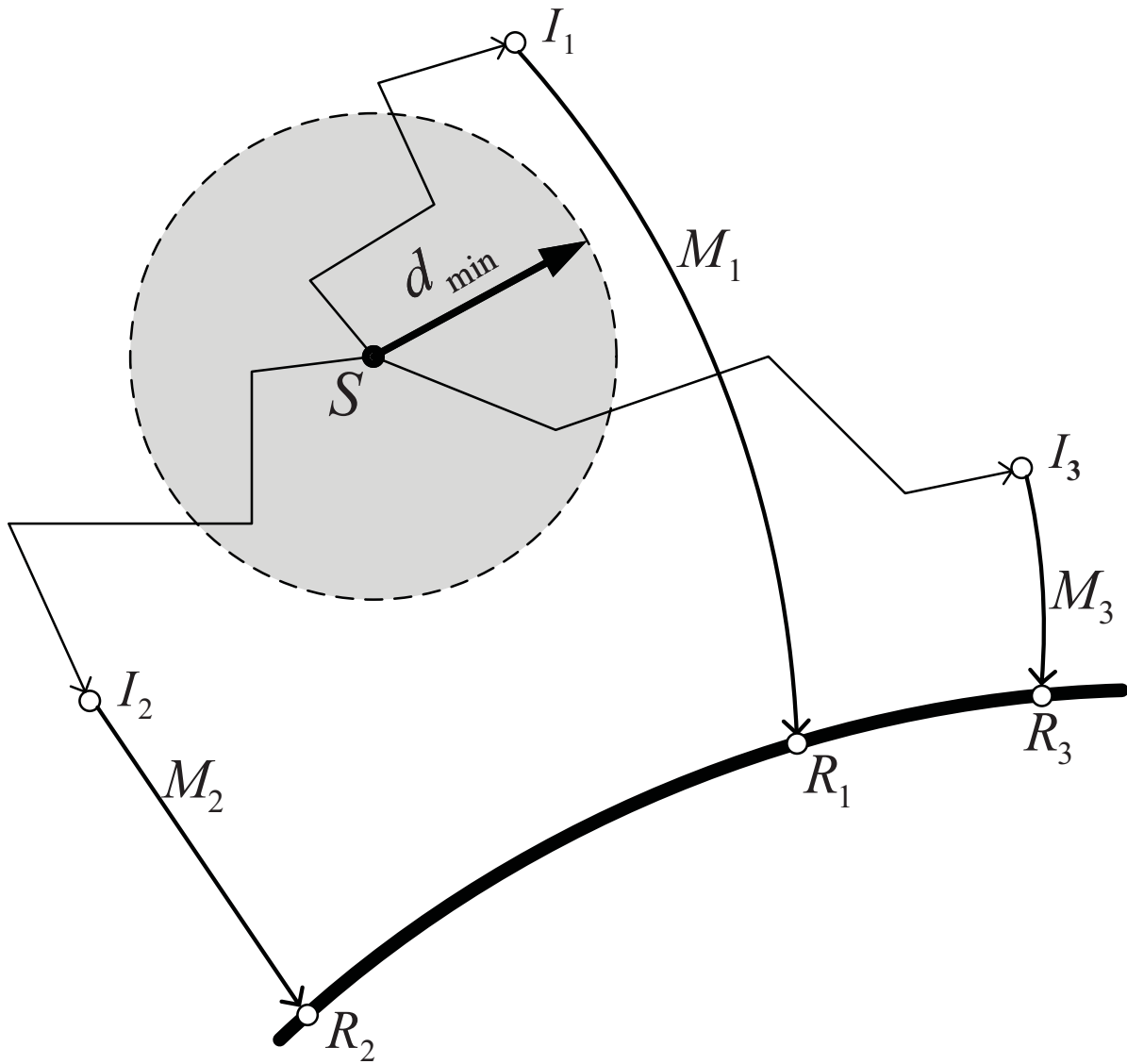


Figure 2.14 Illustrate of the first two phases routing

SINK node.

This routing process provides source-location privacy that resembles the airport terminal transportation system. The message transmission in the ring acts as a network-level mix. As long as it is infeasible for an adversary to distinguish the message initiator from the message forwarder in the mixing ring, then it would be infeasible for the adversaries to identify the real message source-location. Therefore, our goal is to design security mechanisms such that it is infeasible for anyone to distinguish the message source node from the message forwarding node.

Relay ring nodes generate vehicle messages to be transmitted in the mixing ring. The normal ring nodes can store data messages received from the normal nodes. The vehicle messages may contain several data units. These units are left unused initially. If a unit in the vehicle message is not used, we name this unit as *dummy unit*, composed of any fixed data structure, such as all 0s. The length of a unit is the same as the data message sent by a normal node. Upon receiving a vehicle message, if a normal ring node has a real data message received and there is still a dummy unit in the vehicle message, it can replace this dummy unit with the data message. The updated vehicle message will then be forwarded to its successor ring node. If this normal ring node has not received any data messages from the normal nodes, or there is no dummy units left in the vehicle message, it simply forwards this vehicle message. The vehicle message should be sent at the rate which could ensure that all the data messages could be embedded in vehicle messages and forwarded to the SINK with minimum delay.

In our scheme, to thwart message source analysis, the message transmission in the ring is encrypted. Each ring node shares a secret key with its predecessor ring node and a secret key with its successor ring node. As an example, in Figure 2.15, ring node  $B$  shares a key  $K_{AB}$  with ring node  $A$  and a key  $K_{BC}$  with ring node  $C$ . When node  $B$  receives a packet  $M_1$  from node  $A$ , it first decrypts  $M_1$  using the share secret key  $K_{AB}$ . Let  $m_1 = D_{K_{AB}}(M_1)$ . Upon decryption, node  $B$  will be able to find the dummy unit(s) in  $m_1$  and replace the dummy unit(s) with the data message(s) that it received from the normal nodes. Denote the updated message as  $\{D_{K_{AB}}(M_1)\}$ . The updated vehicle message will be encrypted using the shared secret  $K_{BC}$  before it is transmitted to the node

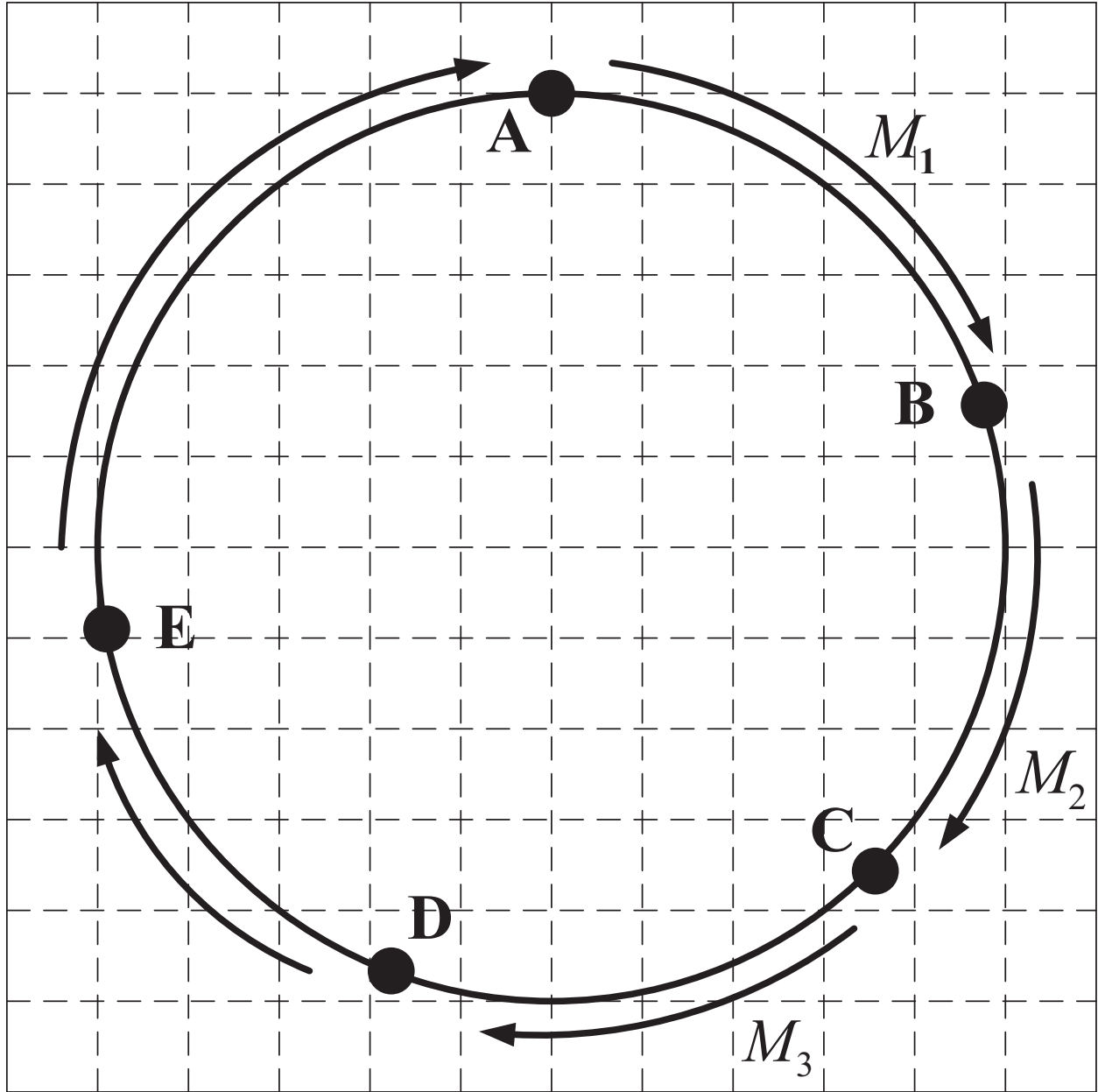


Figure 2.15 Message transmission in the ring

C. Denote the message that generated in node  $B$  as  $M_2$ , then we have

$$M_2 = E_{K_{BC}}(\{D_{K_{AB}}(M_1)\}). \quad (2.1)$$

When DES or AES encryption algorithm is being used to provide message encryption, then it is computationally infeasible to find the correlation between  $M_1$  and  $M_2$ .

Apparently, the energy drainage for the relay ring nodes will be faster than the normal ring nodes. To balance the energy consumption, the normal ring nodes can take turns to be the relay ring nodes. Similarly, since the energy drainage for the ring nodes will be faster than the regular grid nodes, the nodes in the selected ring grid can take turns to be the ring node.

### 2.5.3 Forwarding to the SINK

After a vehicle message arrives at a relay ring node, it will be forwarded to the SINK by this relay ring node with certain probability  $p$ . Here  $p$  is a parameter related to the number of relay ring nodes on the mixing ring. If this vehicle message is not forwarded to the SINK by the relay ring node, it will be forwarded to the next ring node until another relay ring node is reached.

### 2.5.4 Security Analysis for Mixing Ring Routing

We will first analyze that the proposed routing to a random intermediate node (RSIN) in phase one can provide local source-location privacy. Unlike phantom routing, which has no control over the phantom source without leaking significant side information, in the proposed RSIN scheme, the intermediate node is determined before each data message is transmitted by the source-location, the data message carries no observable side information of the message source-location in its content. Therefore, it does not have the security drawbacks of phantom routing discussed before. It is also impossible for the adversary to trace back and identify the real message source based on an individual traffic monitoring. This is because the probability for multiple events from the same source to use the same routing path and intermediate node is very low for large sensor networks.

If an adversary tries to trace back the source-location from the message packet in the route through which the packet is being transmitted to the mixing ring, then the adversary will be led to the randomly selected intermediate node instead of the real message source. Since the intermediate node is randomly selected for each data message, the probability that the adversaries will receive the messages from one source node continuously is pretty small. As shown in Figure 2.14, if an adversary receives  $M_2$  forwarded by  $I_2$ , it would be led to  $I_2$ . However, the next intermediate node  $I_3$  is far from  $I_2$ , so the adversaries could not receive  $M_3$ .

Even if one intermediate node's location is discovered by the adversaries, the source-location is still well protected because the locations of the intermediate nodes are at least  $d_{min}$  away from the real source node. Therefore, the proposed protocol can provide the local source-location privacy.

As shown in Figure 2.14, the intermediate nodes  $I_1, I_2, I_3$  forward messages to ring nodes  $R_1, R_2, R_3$ , respectively. This means that messages generated from one source node will not be forwarded to a specific ring node. Conversely, the data messages received from one ring node could also be transmitted from many different source nodes in the network.

The routing in the mixing ring is the second phase routing. This phase aims at providing network-level source-location privacy. This is achieved by hop-by-hop message encryption. Without hop-by-hop message encryption, by comparing the vehicle message that a node received and transmitted, the adversary can determine whether a data message has been loaded into the updated vehicle message. However, once the hop-by-hop message encryption is implemented, it is computationally infeasible for an adversary to distinguish the message initiator and message forwarder in the mixing ring. The messages across the network are totally mixed up. As shown in Figure 2.14, a data message received by ring node  $B$  could be sent to the SINK node from a completely different ring node, maybe node  $E$  for instance. Therefore, the network-level source-location privacy is achieved.

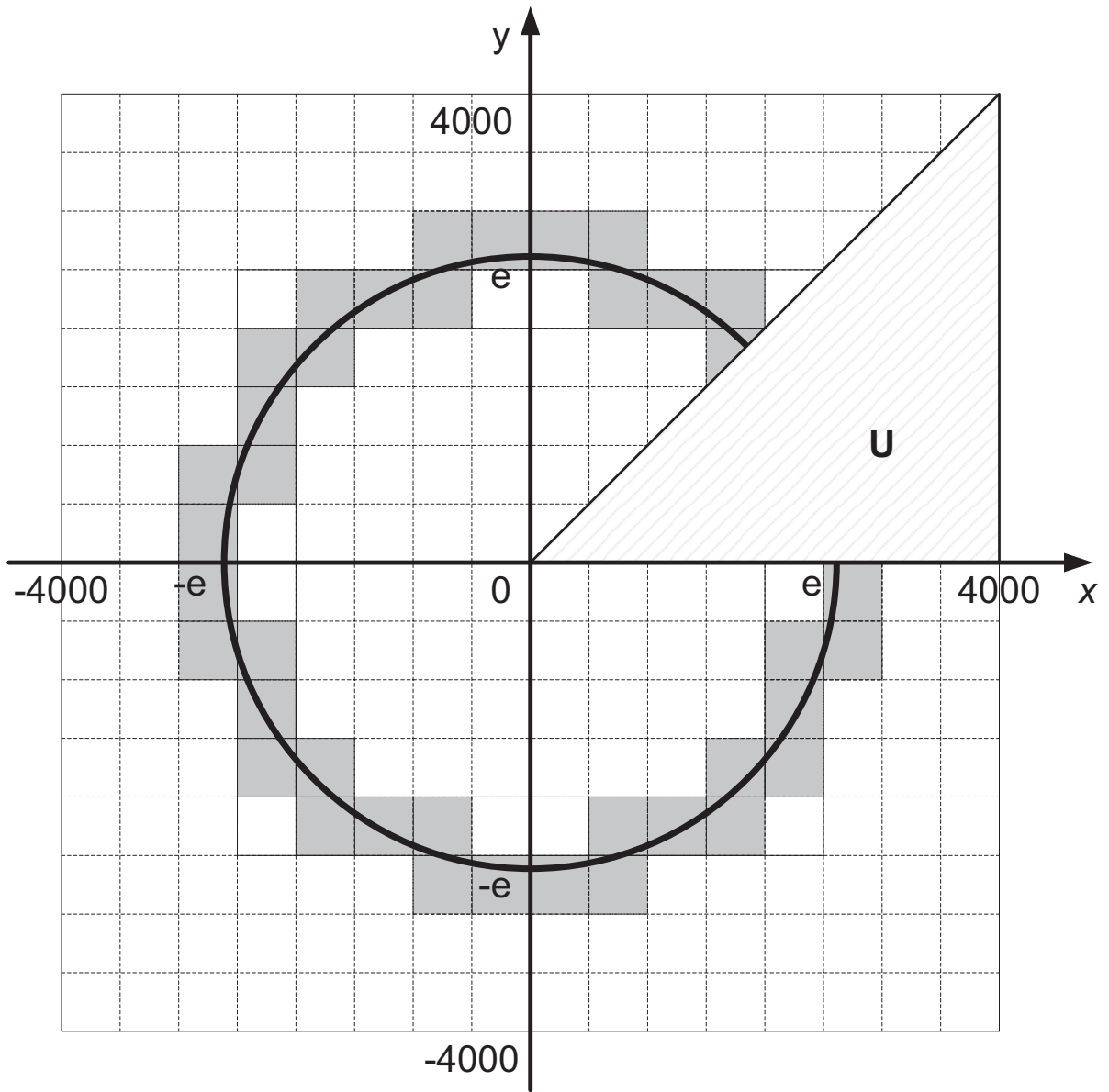


Figure 2.16 Ring selection in simulation setup



### 2.5.5 Performance Analysis and Simulation Results

In our design, all data messages will be delivered to the SINK node through the mixing ring. While providing network-level source-location privacy, the location of the ring should be selected to ensure that the overall energy consumption and latency for message transmission to be lowest for the normal nodes to complete these operations. We assume that each sensor node in the network has complete knowledge of its relative location in the sensor network and also some ring nodes. We also assume that the energy drainage for each transmission is proportional to the square of the distance, i.e.

$$\mathcal{E} = \alpha \times d^2,$$

where  $\mathcal{E}$  denotes the energy consumption,  $\alpha$  is a constant parameter and  $d$  is the distance of the transmission. Figure 2.16 gives an example of a target area of size  $8000 \times 8000$  meters. The shaded grids are selected as the ring grids. The line in the middle of the shaded area is indicated by the solid line. If the density of the sensor nodes in the sensor network is  $\lambda$ , then the total energy consumption for each sensor in this area to transmit one message to a ring node can be calculated as follows:

$$\begin{aligned} \mathcal{E}_{total} &= 8\mathcal{E}_U \\ &= 8\alpha\lambda \int_0^{\pi/4} \int_0^{4000/\cos\theta} (r-e)^2 r dr d\theta, \end{aligned}$$

where  $\mathcal{E}_U$  is the energy consumption for area  $U$  as demonstrated in Figure 2.16. It can be calculated that when  $e = 3061$ , the overall power consumption  $\mathcal{E}_{total}$  achieves the minimum. In this way, we get the optimal ring location.

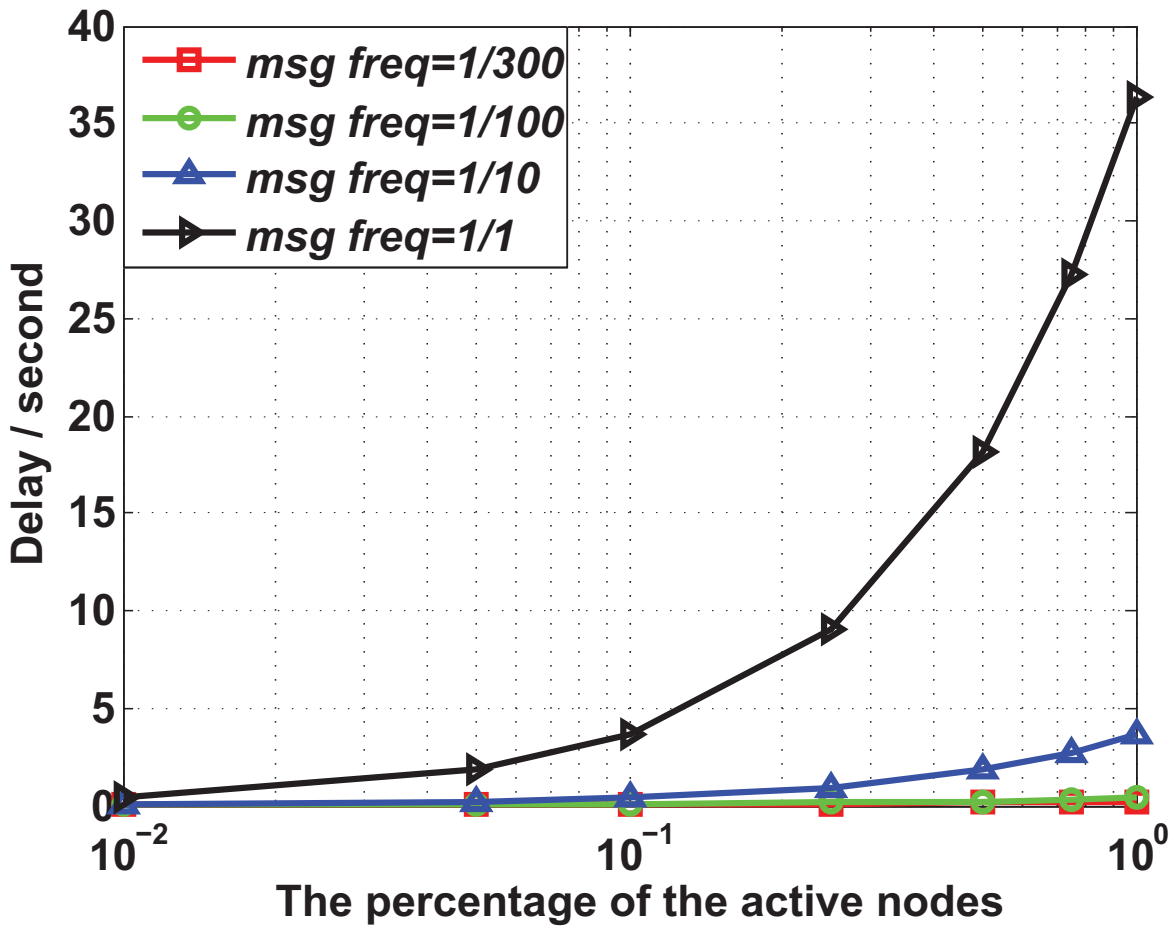


Figure 2.17 Mixing Ring: Power consumption of normal nodes

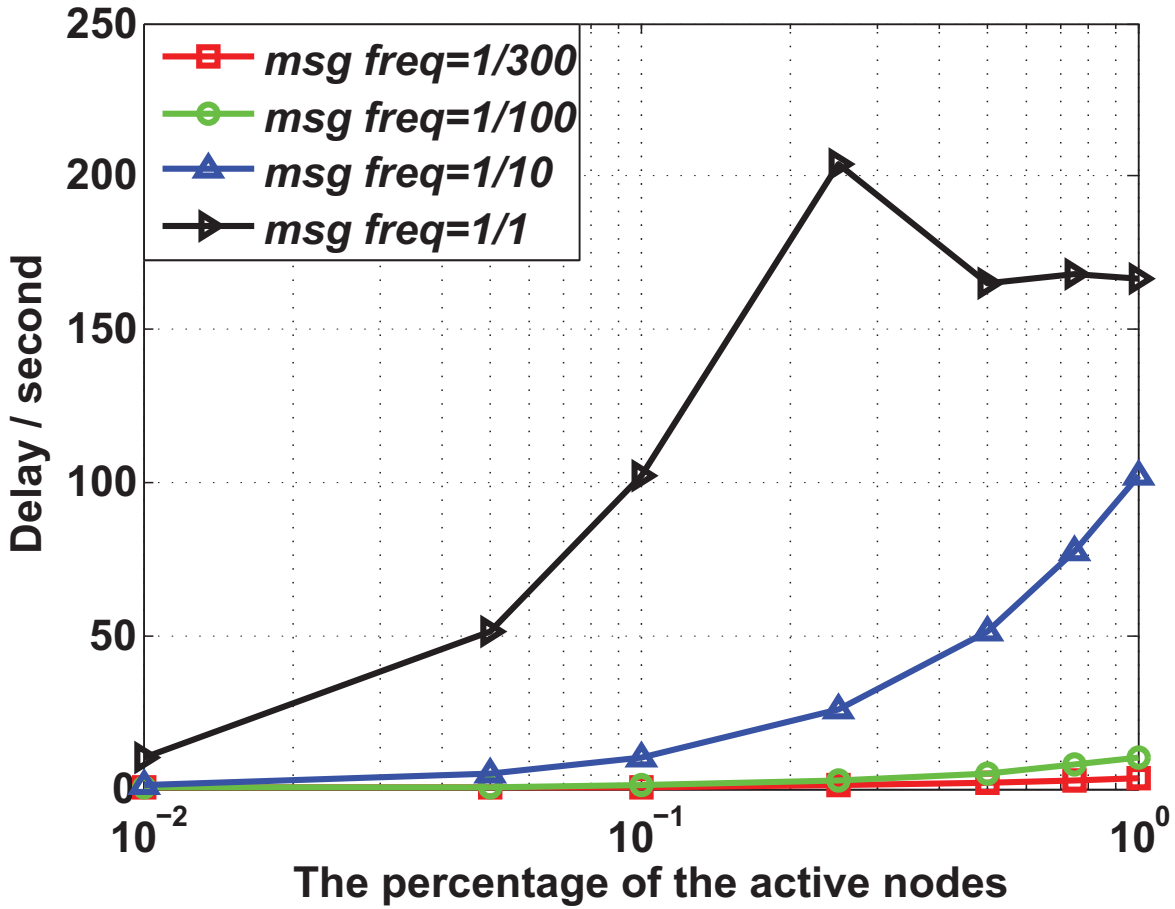


Figure 2.18 Mixing Ring: Power consumption of ring nodes

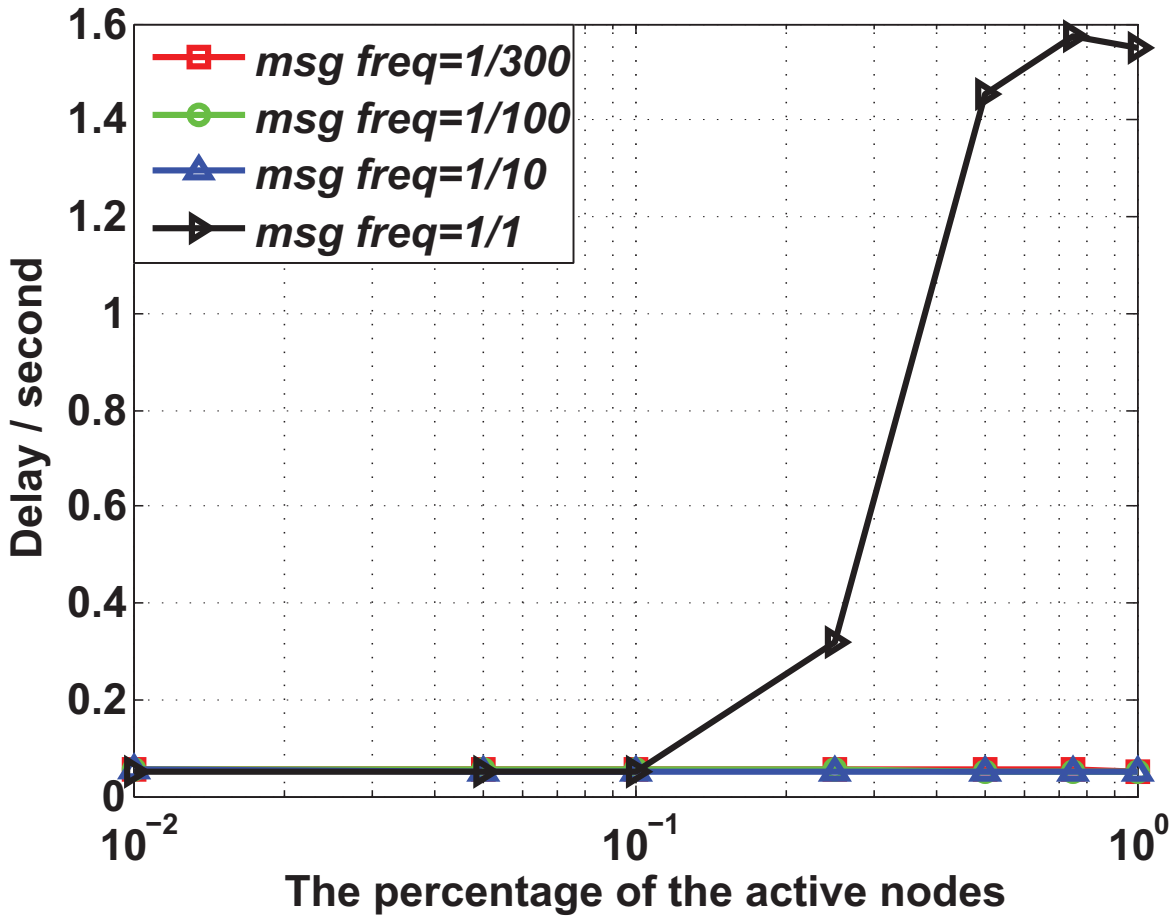


Figure 2.19 Mixing Ring: Message latency

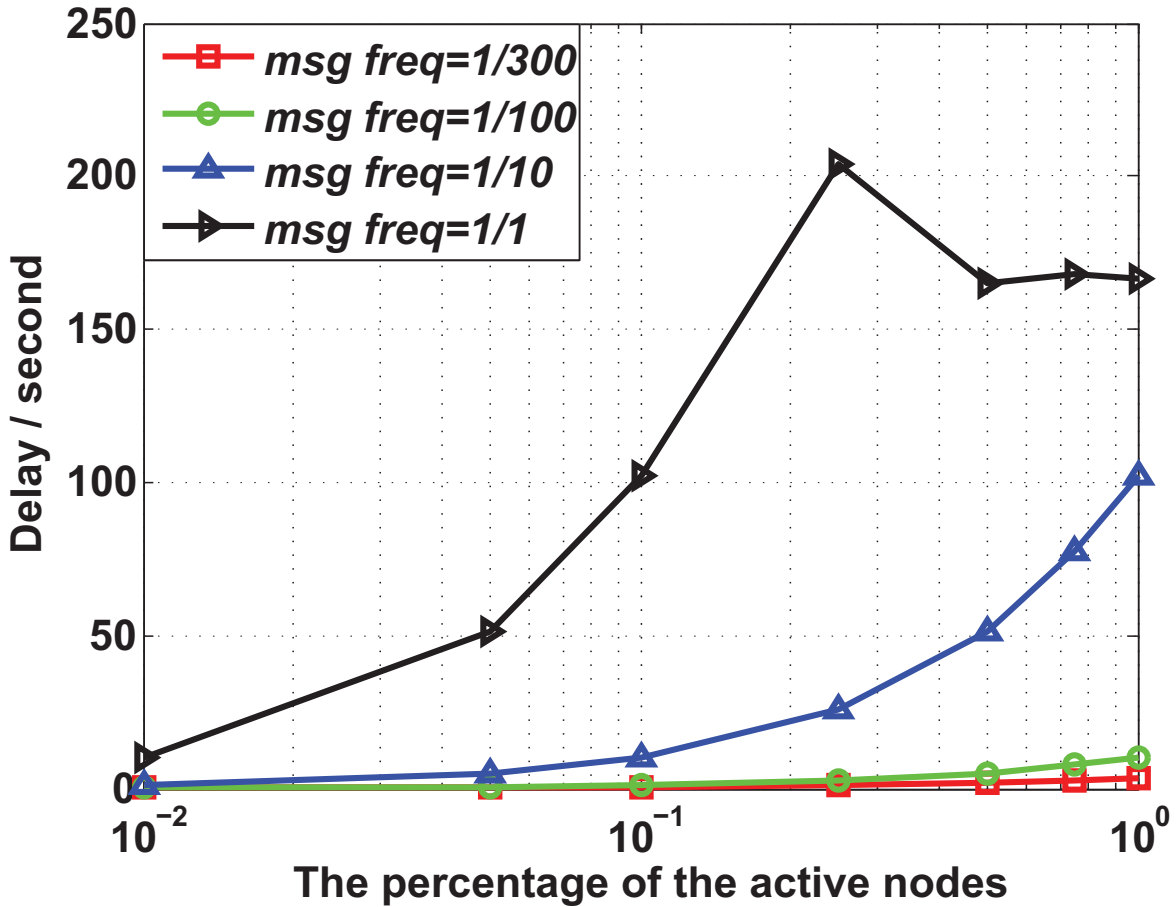


Figure 2.20 Mixing Ring: Message delivery ratio

In practical application, for large sensor network, usually only a small fraction of the sensor nodes in the network has events to report. We name these nodes as *active nodes*. We also define two parameters in the simulation:  $\tau$ , the number of data messages a normal node generates in each second, and  $a$ , active nodes ratio.

Assume the network is composed of  $g$  normal nodes, and the ring consists of  $r$  ring nodes. On average, one ring node should be responsible for delivering the data messages from  $g/r$  normal nodes. Assume data messages are  $l$ -bit long, then on average, in each second, a ring node will receive:

$$\gamma = \frac{g}{r} \times l \times a \times \tau = \frac{gla\tau}{r},$$

messages.

If vehicle messages are  $L$ -bit long, the number of vehicle messages generated by a ring node in one second is:

$$\frac{gl\alpha\tau}{r} \times \frac{1}{L} = \frac{gl\alpha\tau}{rL}.$$

Since only the relay ring nodes on the ring can generate vehicle messages. If there are  $n$  relay ring nodes on the ring, then each relay ring node needs to generate at least

$$\frac{gl\alpha\tau}{rL} \times \frac{r}{n} = \frac{gl\alpha\tau}{nL},$$

vehicle messages each second.

Simulation results are provided in Figure 2.17, 2.19 to demonstrate the power consumption for both normal nodes and ring nodes, message latency and message delivery ratio of the proposed scheme. Our simulation was performed using NS2 on Linux system. In the simulation, the target area is a square field of size  $8000 \times 8000$  meters. We partition this field into 2400 normal grids/nodes. The mixing ring is composed of 80 grids, i.e,  $r = 80$ . There are four relay ring nodes in the mixing ring, i.e,  $n = 4$ . We assume that the randomly selected intermediate node is at least 600 meters away from the real message source. The data messages are 8-bit long, i.e,  $l = 8$ . The vehicle messages are 16-bit long, i.e,  $L = 16$ .

From the Figure 2.17.(a) and (b), we can see that ring nodes consume more energy than normal nodes. To solve this problem, the nodes in ring grids can take turns to be the ring nodes. It is also noticed that the delivery ratio drops exponentially when the traffic volume increases. It is primarily because of the traffic collisions and packet losses caused by the increased traffic volume. For a large sensor network, it is usually not necessary for all the sensor nodes to be active at the same time. In practice, the percentage of active nodes might be very low. The transmission frequency also tends not to be very high. In other words, the traffic volume may be low. In this scenario, we can ensure almost 100% delivery ratio, as shown in Figure 2.19.(d). The simulation results demonstrate that the proposed scheme is very efficient and can be used for practical applications.

## 2.6 Source-Location Privacy using STaR Routing

### 2.6.1 Preliminary: Source-Location Privacy Evaluation Model

In [72], security analysis of source-location privacy based on quantitative measurement on information leakage of the source-location has been proposed. The quantitative measurement divides information leakage analysis into three categories:

1. *Correlation-based source identification attack:* Correlation-based attack is an ID based source node determination. When an adversary receives a message with an ID whose location is already known, the location of this node is also known.
2. *Routing traceback attack:* Routing traceback is an attack that when an adversary captures a message, he can identify the immediate message sender and quickly move to it. For fixed path routing of length  $n$ , if the adversary can capture  $n$  messages from this source, then he is able to locate the message source node.
3. *Reducing source space attack:* Reducing source space attack refers to the attack that the adversary can limit the source node to a proper subset/area in the networks when a message is captured. When multiple messages are captured, the subset/area may be further reduced so that the source-location can be limited to a subset/area that may lead to a relative easy or complete source identification.

To prevent correlation-based source identification, a dynamic ID based approach can be used to prevent adversaries from relating messages transmitted from each source [72, 73]. This can be done by requiring that each node in the network be preloaded with an *ID-hash-chain* so that a different and uncorrelated ID is attached to each message. The adversaries are no longer able to get any useful information about the source node through correlation-based source identification.

For routing traceback and reducing source node space analysis, three criteria have been defined in [72].

**Definition 1 (Source-location Disclosure Index (SDI))** *SDI measures, from an information entropy point of view, the amount of source-location information that one message can leak to the adversaries.*

For a routing scheme, if we assume the total privacy for a source node  $S$  is 1, and the  $SDI$  is fixed, then the adversary only needs to receive  $\lceil \frac{1}{SDI} \rceil$  messages initiated from  $S$  in order to successfully locate  $S$ . Therefore, for a good SLP scheme,  $SDI$  should be as small as possible.

**Definition 2 (Source-location Space Index (SSI))** *SSI is defined as the set of possible network nodes, or area of the possible network domain, that a message can be transmitted from.*

For a routing scheme, if  $SSI$  is large, it means that the message may be transmitted by many possible source nodes. On the contrary, if  $SSI$  is small, then the adversary can limit the possible source nodes to a small group. Therefore, for a SLP scheme,  $SSI$  should be as large as possible so that the complexity for an adversary to perform an exhaustive search of the message source is maximized.

**Definition 3 (Normalized Source-location Space Index (NSSI))** *NSSI is defined as the ratio of the SSI area over the total area of the network domain. Therefore,  $NSSI \in [0, 1]$ , and we always have  $NSSI = 1 - \delta$  for some  $\delta \in [0, 1]$ . The  $\delta$  is called the local degree.*

It is clear that the scheme with the local degree 0 provides the highest degree of SLP.

Based on these definitions, we have derived in [72] that fixed path routing schemes are the least secure SLP schemes.

**Lemma 1** *Suppose there is a fixed routing path between the source node  $S$  and the destination node  $D$  of length  $L$  hops.  $A$  is an adversary who can detect all messages transmitted to  $D$ . Then, after receiving  $L$  messages,  $A$  will be able to trace back to the source node  $S$ , i.e.,*

$$SDI = \frac{1}{L}.$$



If there are  $n$  disjoint routing paths between the source node  $S$  and destination node  $D$ , and the length of the  $n$  paths are:  $L_1, L_2, \dots, L_n$ , respectively. For each message, the source node  $S$  will send it along path  $L_i$  with probability  $p_i$ , where

$$\sum_{i=1}^n p_i = 1.$$

For path  $i$ , we have  $SDI_i = \frac{p_i}{L_i}, i = 1, \dots, n$ . Define the overall SDI as

$$SDI = \sum_{i=1}^n p_i \cdot SDI_i.$$

We will then have the following result [72].

**Theorem 1** *Suppose there are  $n$  disjoint routing paths between the source node  $S$  and the SINK node  $D$ . The lengths of the  $n$  routing paths are  $L_1, L_2, \dots, L_n$ . Let  $p_i$  be the probability that messages will be transmitted along the path  $L_i$ , then when  $p_i = \frac{L_i}{L_1 + L_2 + \dots + L_n}, i = 1, 2, \dots, n$ , the SDI is minimized, which is*

$$SDI = \frac{1}{L_1 + L_2 + \dots + L_n}.$$

**Corollary 1** *Suppose there are  $n$  disjoint routing paths between the source node  $S$  and the destination node  $D$ . The length of the  $n$  routing paths are  $L_1, L_2, \dots, L_n$ , respectively. The adversary then needs to receive on average*

$$\frac{1}{SDI} = L_1 + L_2 + \dots + L_n$$

*messages to fully determine the location of the source node, i.e., trace back to the source node.*

As a result, to provide SLP in wireless sensor networks, we have to increase the total number of possible routing paths between the destination node and the source node. However, for a practical network configuration, the number of routing paths cannot be increased without limitation. This means that we will always have  $SDI > 0$ .

We can summarize the two defects of the SLP schemes through fixed routing paths as follows:

- *Non-zero SDI*: For fixed path routing, no matter how dedicated the scheme is designed, *SDI* is always larger than 0. In other words, for each message sent out by one source node, from a probability point of view, there is always a fraction of source information to be leaked to the adversaries. So, no matter how small the *SDI* is, when enough messages are received, the adversaries are always able to locate the source node.
- *Limited SSI*: Because the routing paths are fixed for the source node, the correlation between the messages transmitted on a particular path and the source node is high. In other words, *SSI* is small compared to the overall sensor network size.

Phantom routing is a dynamic routing scheme. In this scheme, the message is first routed to a phantom source through a random path before it is forwarded to the actual destination node. To make sure that the phantom source is away from the actual source node, the direction information must be stored in the message's header. In this way, the intermediate nodes on the routing path are able to select the next forward node on the routing path along the same direction.

In [72], each sensor node is assumed to have a unique ID that corresponds to a physical location. The problem of this design is that it makes it possible for the adversaries to monitor and link all messages from the same source node together, which may help the adversaries to identify the source-location since the IDs correspond to the grids' locations. Therefore, we have

$$SDI > 0.$$

Whenever the adversaries discover a message sent from a grid with an ID that they already know, they can use this message to move closer to the message source. Fortunately, this problems can be easily solved if dynamic ID is assigned for each message. In this case, the correlation between the source node and the message received in the random path can be viewed as zero, i.e.,

$$SDI \simeq 0.$$

However, the direction information stored in the message header can facilitate the adversary to narrow the possible area of the source node. Take the section-based random walk as an example.

Once a message is corrupted by an adversary on the random path, the adversary can determine to which direction of the current location the actual source node is located. Therefore, we have

$$NSSI < 1.$$

When multiple adversaries collaborate in the target area  $T$ , the  $NSSI$  can be further reduced and the SLP is no longer well protected.

### 2.6.2 STaR Routing Scheme

The STaR routing scheme, which was introduced in [74], provides source-location privacy through a two-phase routing protocol. In the first phase, the source node routes the message to a randomly selected intermediate node located in a pre-determine region around the SINK node. We call this region the Sink Toroidal Region (STaR). The random intermediate node services as a fake source when the message is forwarded to the SINK node. The intermediate node will forward the message to the SINK node by single-path routing in the second phase. The combination of these two phases guarantees the local degree to be small, therefore, providing a high degree of SLP. In STaR SLP scheme, to prevent the adversaries from getting any useful source-location information through *correlation-based source identification*, a dynamic ID proposed in [73] should be assigned for each message.

In our scheme, the network is evenly divided into small grids [73]. We assume that the sensor nodes in each grid are all within the direct communication range of each other. In each grid, the header node coordinates the communication with other header nodes nearby. We assume that the whole network is fully connected through the multi-hop communications.

The goal of the proposed scheme is to provide local and global source-location privacy with adequate energy-efficient routing. Local privacy is obtained by the fact that the intermediate node is expected to be neither too close, nor too far from the real source, for most cases. The STaR area would be a large area with at least a minimum radius distance  $r$  from the SINK node to provide global privacy. Also, the STaR area guarantees that the intermediate node is at most a maximum

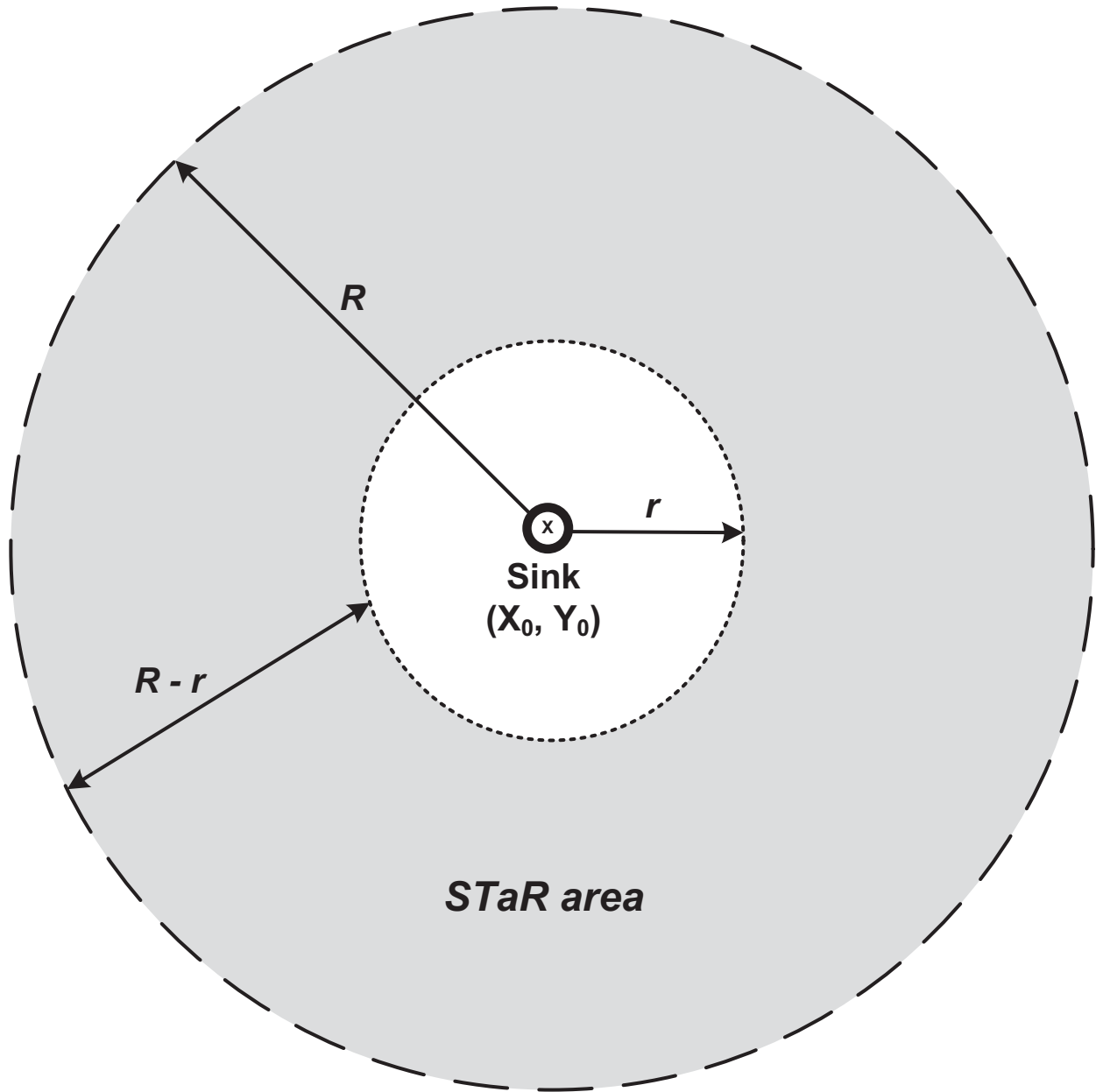


Figure 2.21 Distribution of the STaR area

distance  $R$  from the SINK node to limit the energy consumption in the routing paths. This routing scheme is designed to give the illusion that the source node is sending messages to the SINK node from all the possible directions. In this way, the STaR creates an effect that is similar to the totally random RRIN scheme [75] but with less energy consumption and shorter delays.

We assume that each sensor node only has knowledge of its adjacent nodes and has no accurate information of the sensor nodes more than one hop away. We also assume that each node has knowledge of the parameters that are shown in Figure 2.21. The description of the parameters are as follows:

- $x_0, y_0$ : The corresponding X and Y coordinates of the SINK node location,
- $R$ : The pre-determined radius from the SINK to the outer-edge of the STaR area,
- $r$ : The pre-determined radius from the SINK node to the inner-edge of the STaR area.

From these parameters,  $\{x_0, y_0, R, r\}$ , the source nodes are able to generate random points within the STaR area. Since we assume that the SINK node is located at the relative location  $(x_0, y_0)$ , the source node selects the random location  $(x, y)$  according to the following two steps:

1. Randomly select  $d$  uniformly from  $[r, R]$ .
2. Randomly select  $\theta$  uniformly from  $[0, 2\pi]$ .

In this way, we can calculate the coordinate of the intermediate node as  $(x, y) = (x_0 + d \cos(\theta), y_0 + d \sin(\theta))$ .

After obtaining the random location  $(x, y)$ , the message can then be routed towards the grid at location  $(x, y)$ . Since each node only knows its adjacent neighbor nodes' relative location, it can determine the direction that the message should be routed. Once the message is within the desired grid of the random location, the message is routed to the header node of the grid. The header node then becomes the random intermediate node. If the desired grid does not contain any nodes, then the last node in the routing path would become the desired location and the header node in that grid

would become the intermediate node. The intermediate node then routes the received message to the SINK node using single-path routing.

### 2.6.3 Security Analysis for STaR Routing

We will analyze the SLP for STaR routing scheme. We assume the adversary is unable to monitor the entire sensor area of the source node, since otherwise it can monitor the actual event directly.

In the proposed STaR routing scheme, a random intermediate node is selected for each message from the STaR area shown in Figure 2.21. Unlike the directed walk of the phantom routing scheme, our protocol does not leak direction information to the adversaries.

**Theorem 2** *For the proposed STaR scheme, if assume that the STaR area is large enough so that the probability for multiple messages to be routed using the same intermediate node is negligible. Then the amount of source-location information that can be leaked from one message is negligible, i.e.,*

$$SDI \simeq 0,$$

*and SLP with local degree 0.*

**Proof 1** *For the proposed STaR scheme, the source-location disclosure index can be analyzed in two scenarios based on the location of the adversarial attacks: (i) the adversary monitors traffic between the randomly selected intermediate node in STaR area and the SINK node, and (ii) the adversary monitors traffic between the source node and the randomly selected intermediate node in the STaR area.*

*For scenario (i), first, the STaR area is at least an  $r$  distance away from the SINK node. Second, every node within the STaR area from the SINK node has equal probability in being selected as the intermediate node for all messages by all possible source nodes. Therefore, the messages are being routed from the STaR area to SINK from all directions with equal probability, as shown in Figure 2.22. Therefore, the messages to be transmitted from the STaR area to the SINK node area*

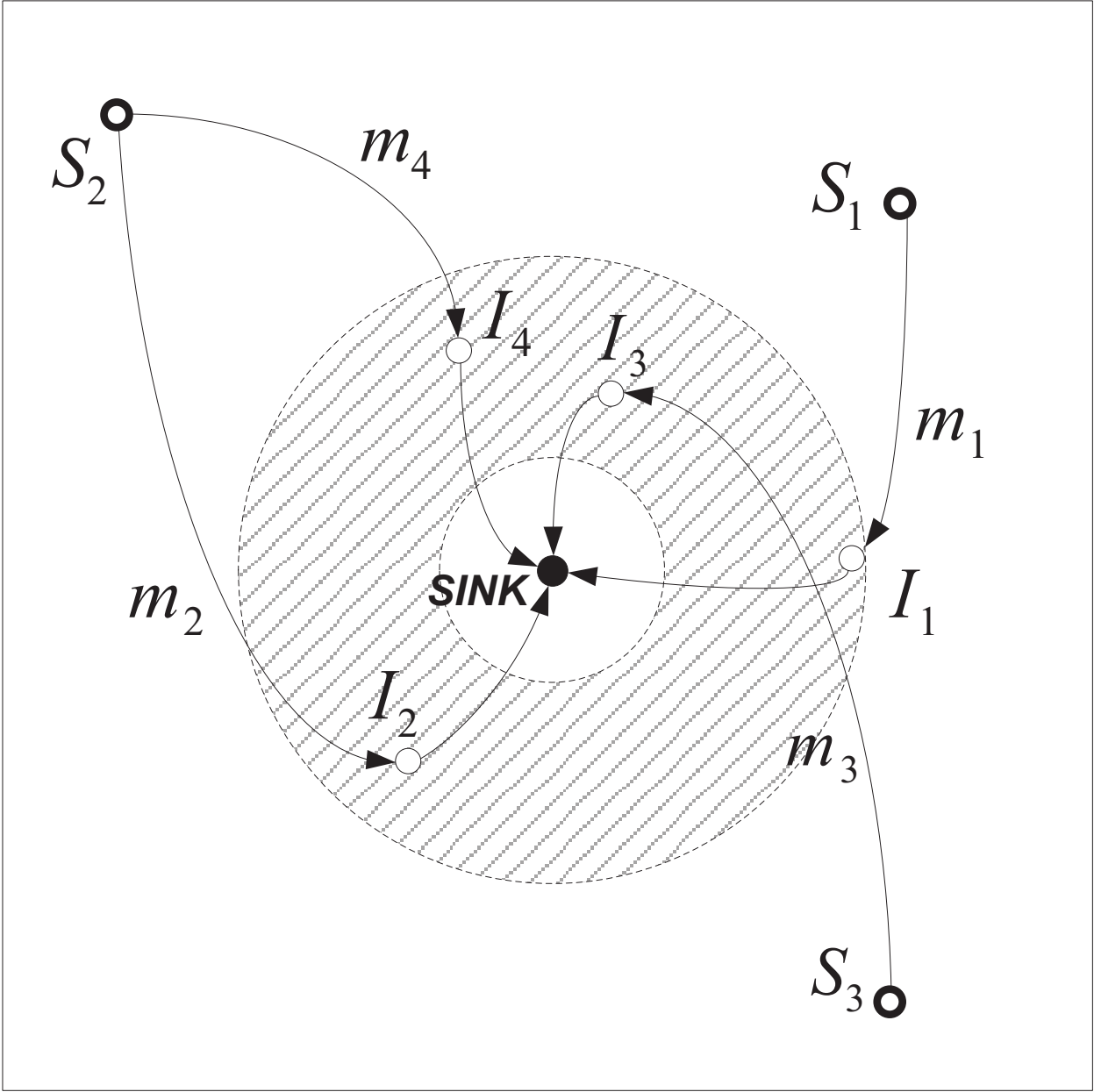


Figure 2.22 Routing illustration of the STaR protocol

bear no correlation with the actual message source and it is impossible for an adversary to gain any information of source-location for the message source.

It is also impractical for the adversary to perform routing traceback to figure out the source location by only monitoring and analyzing traffic patterns around the SINK node. In this scenario, the global source location privacy can be assured and hence source location privacy with a low local degree can be guaranteed.

For case (ii), the message source may be located anywhere in the network domain and the intermediate node is expected to be far from the real source for most cases. The probability for the adversary to intercept a message is very low for large wireless sensor networks. The probability for an adversary intercept multiple messages from the same source in the same location is negligible. In fact, if a dynamic ID is used for each message, then even if an adversary is able to intercept one message, he is still unable to link them together.

When only one message is intercepted, the adversary can only determine the immediate previous node based on our assumption. The adversary can neither determine the direction of the next previous hop source node, nor the direction of the actual source node. In other words, the source-location disclosure index equals to 0 with an negligible exception for the nodes in the area of the adversary. Therefore, we have

$$SDI \simeq 0,$$

and the SLP local degree is 0.

**Theorem 3** In the proposed STaR scheme, for any received message, the source node is either located in the area that the adversary can monitor, or the adversary has no information of the actual message source. That is

$$NSSI \simeq 1.$$

**Proof 2** Similar to Theorem 2, the analysis can be divided into two scenario: (i) the adversary monitors traffic between the randomly selected intermediate node in STaR and the SINK node, and



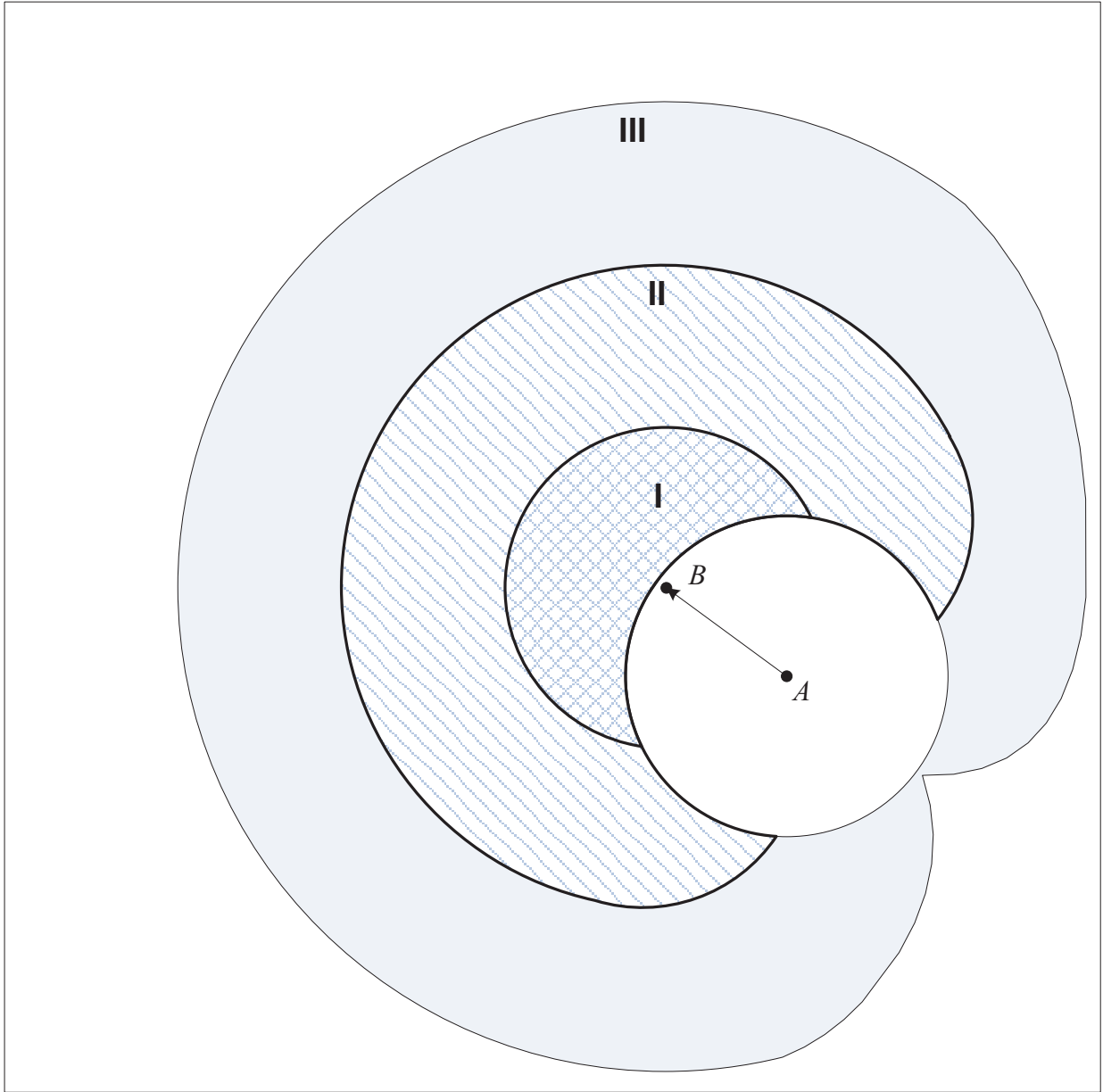


Figure 2.23 The source location analysis of STaR routing scheme

(ii) the adversary monitors traffic between the source node and the randomly selected intermediate node in the STaR area.

For case (i), when an adversary receives a message, if he is unable to find the message source in his area, then the source node can be in any direction and hop distance away from the location where the message is received.

For case (ii), as shown in Figure 2.23, when a message is received in location A, based on our assumption, the adversary can find the immediate message node, say B. However, the only information that the adversary can get for the nodes prior to B are located in the shaded area I, I + II, and I + II + III, and so on. Since the adversary can neither determine the direction, nor the hop distance of the actual source node, the actual message source node can be located anywhere of the sensor domain except the area centered at A. In this way, we have

$$NSSI \simeq 1.$$

#### 2.6.4 Performance Analysis and Simulation Results

To evaluate the performance of the schemes proposed, extensive simulations have been conducted using ns-2 on RedHat Linux system. The results of the simulations are shown in Fig. 2.24, 2.26. In the simulation, 400 nodes are randomly distributed in a square target area of size  $3360 \times 3360$  meters, while the SINK node is located at the center of the network. We set hop count of directed walking of phantom routing to be four, which on average the phantom source was found to be 526.12 meters away from the real source. For RRIN scheme, the minimum distance between the source node and the intermediate nodes was set to 480 meters, and the average distance turned out to be 529.14 meters. We also illustrate the performance of the totally randomly selected intermediate nodes. For STaR routing, the inner radius,  $r$ , was set to 480 meters, while the outer radius,  $R$ , was set to 640 meters.

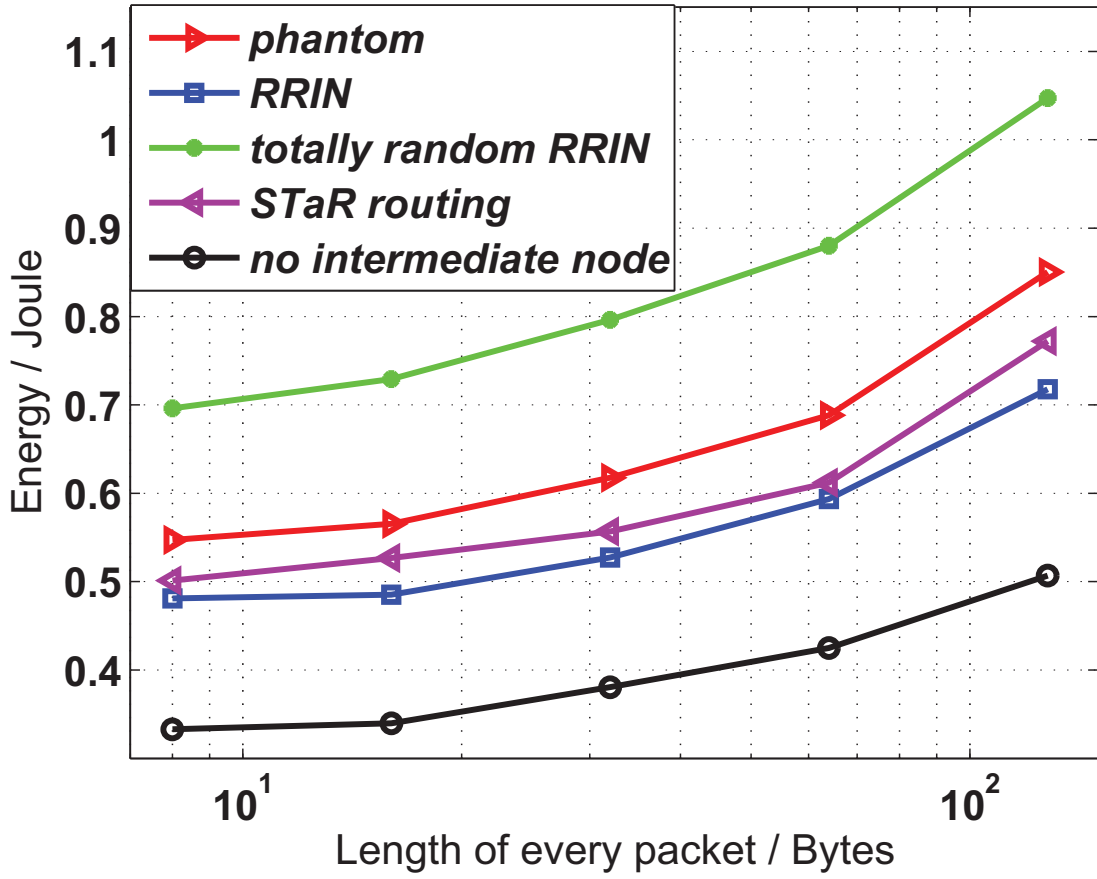


Figure 2.24 Performance of STaR routing: Power consumption

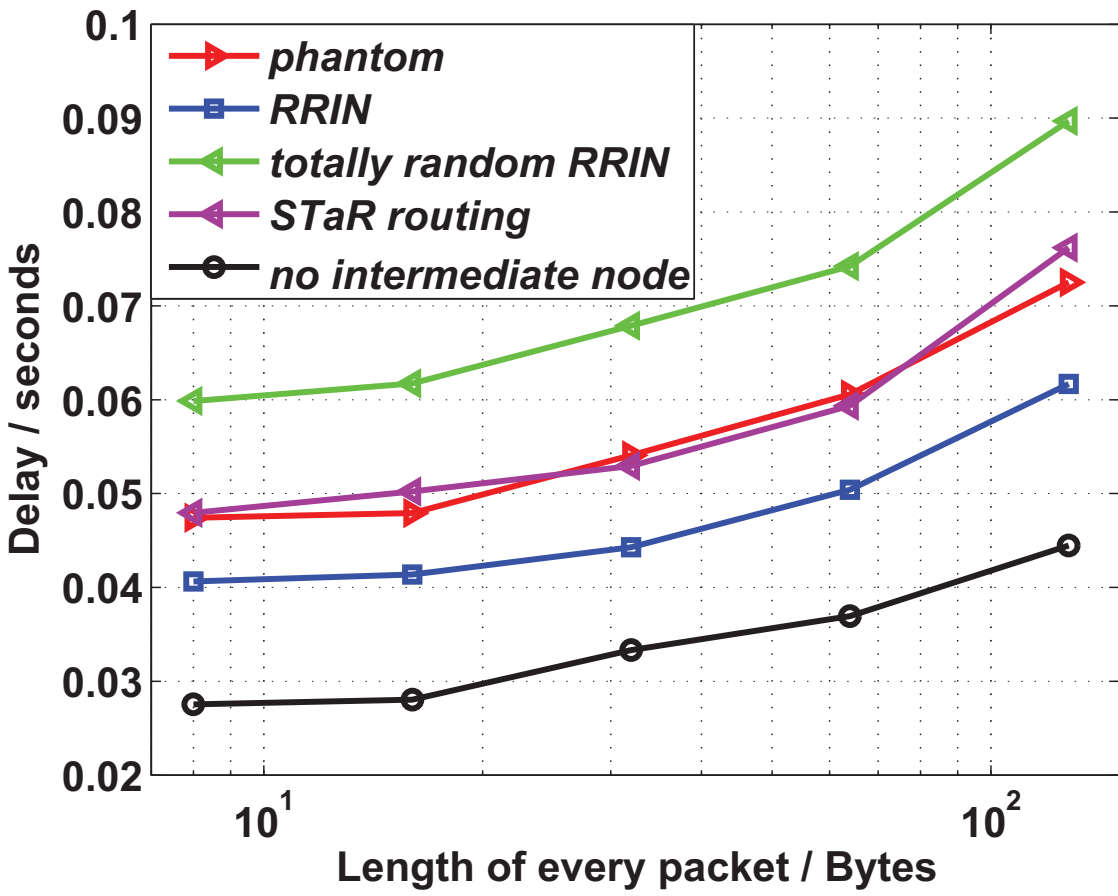


Figure 2.25 Performance of STaR routing: Message latency

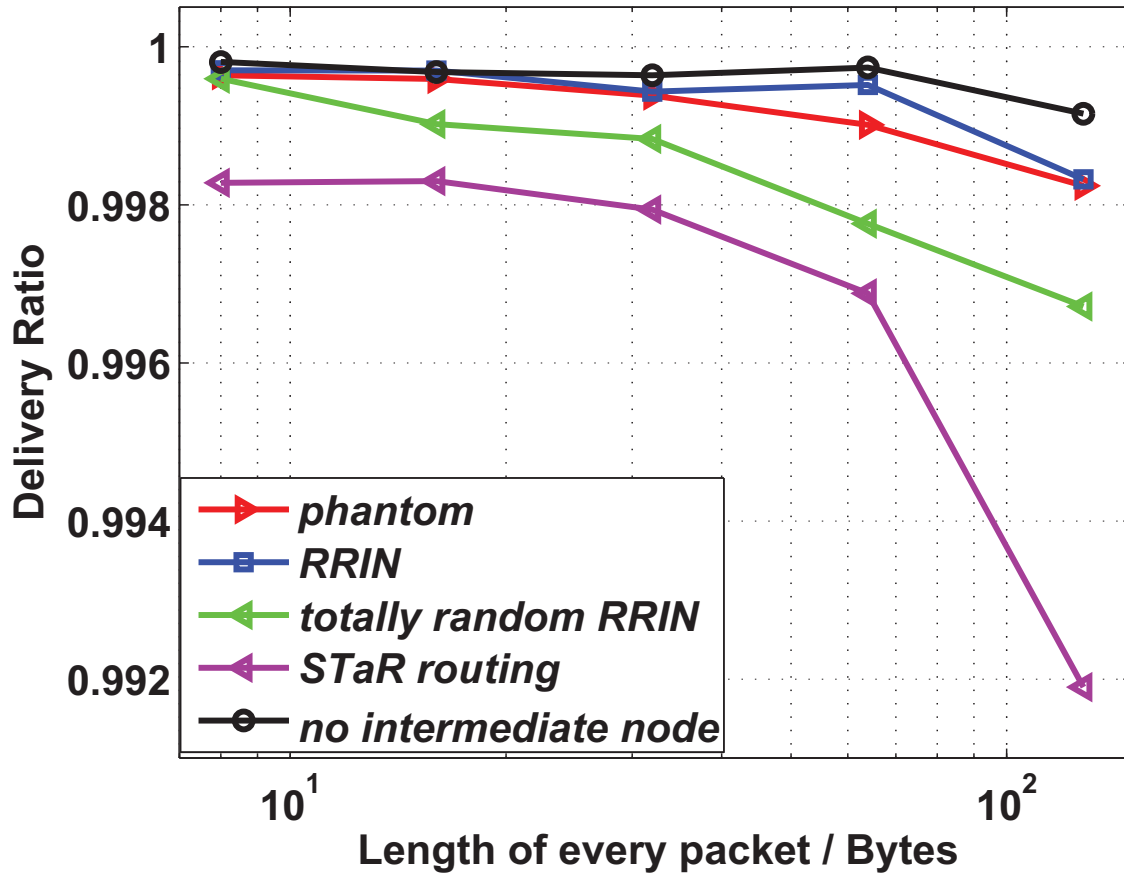


Figure 2.26 Performance of STaR routing: Message delivery ratio

Through analysis and simulation results, we find that direct routing without intermediate node has the best performance while totally random RRIN has the worse performance. The performance of the RRIN scheme is better than phantom routing for comparable security since the average routing paths in phantom routing is longer than the RRIN due to the more curved routing paths. The performance of STaR is between the totally random RRIN and constrained RRIN. The delivery ratio for STaR is slightly lower than the two RRIN schemes due to the possible higher collisions ratio.

## CHAPTER 3

### DESTINATION-LOCATION PRIVACY PROTECTION

#### 3.1 Limitations with Existing Solutions for Destination-Location Privacy

The onion routing protocol, discussed in [34], provides anonymous communication in a network. As a message is passed in the network, onion routers repeatedly encrypts the message between routers. This technique provides privacy of the identity of the sender, destination and the message content. Onion routing prevents an adversary from eavesdropping on the message content and protect against traffic-analysis attacks by making the sender and destination anonymous. This protocol was designed to protect communication over the internet network and uses cryptosystems, which make it not suitable for WSNs.

Broadcasting-based schemes provide location privacy by mixing the real messages with fake messages so that they become indistinguishable to the adversaries. In [37], the DEEP (Differential Enforced Fractal Propagation) routing protocol is introduced. In this scheme, the source node send the real message to the destination using a random walk routing in the direction on the destination and nodes in the routing path injects fake messages into the network to create hot spots to help protect the destination location. The LPR (Location-Privacy Routing) with fake message injection is proposed in [76]. This scheme is similar to the DEEP routing protocol but does not create hot spots and try to solve the vulnerability of the DEEP routing. The LPR protocol randomizes the routing paths so that the forwarding direction of the real message is not always towards the destination. In section 2.6.3, we will analysis the security vulnerability of these two schemes in providing destination-location protection schemes.

Providing location privacy through dynamic routing is, in our opinion, one of the most feasible approaches in WSNs [27, 40]. The main idea is to prevent the adversaries from monitoring the traffic to the location of a source or destination node. A representative example of a routing-

based protocol is the phantom routing protocol, which involves two phases: a random or direct walk phase and a subsequent flooding/single path routing phase. In the random walking phase, the message from the actual source will be routed to a phantom source along a random path or a designed directed path. The phantom source is expected to be far away from the actual source. With sector-based directed walk, the source node first randomly determines a direction that the message will be sent. Every forwarder on the direct walk path will forward this message to a random neighbor in the same direction to ensure that the phantom source will be away from the actual source. This scheme was design to protect the source node location and not the destination node location. The strength of this scheme is that the source node will first send the message to a phantom source that may or may not be in the direction of the destination node location. The weakness is when the phantom source route the message to the destination, the message will always be routed in the direction of the destination node.

Like destination-location privacy schemes, source-location privacy [72,77] is also a vital component to context-based attacks. Many of the techniques used for protection of the source node can not be applied for protection of the destination node. Source-location privacy is beyond the scope of this paper but we can apply some of the attributes of source location privacy schemes for destination location privacy, such as using fake message injection and two-phase routing techniques. In this paper, we focus on providing routing-based destination-location privacy.

## **3.2 Network Models and Design Goals**

### **3.2.1 The System Model**

The following assumptions are made about the system:

- The network is divided into grids. The sensor nodes in each grid are fully connected. In each grid, there is one header node responsible for communicating with other nearby header nodes. The whole network is fully connected through multi-hop communications.

- Every node in the network can become a source node on a detection of an event. On detecting an event, a sensor source node will generate and send messages to the destination node through a multi-hop routing.
- Each message will include a unique node ID where the event was generated. The SINK node can only determine source node location based off the node ID.
- The sensor nodes are assumed to know their relative location and destination node location. We also assume that each sensor node has the knowledge of its adjacent neighboring nodes. The information about the relative location of the sensor domain may also be broadcasted through this network for routing information update.
- The key management, including key generations, key distribution and key update, is beyond the scope of this paper. However, the interested readers are referred to references such as [8].

### 3.2.2 The Adversaries Model

In this paper, the adversary has the following characteristics:

- **Well-equipped:** The adversary does not need to worry about the energy consumption and has adequate computation capability. On detecting a transmitted message, the adversary could determine the receiver node by waiting to see which neighbor node retransmit the message. The adversary is able to move to this receiver's location without much delay and has enough memory to store any useful information. If needed, the adversary could compromise some sensor nodes in the network.
- **Passive:** The adversaries carry out some passive attacks, which only involve eavesdropping work.
- **Traffic-monitoring:** The adversary is able to monitor the traffic in an area and receive all messages in this area. However, we assume that the adversary is unable to monitor the entire network.



### **3.2.3 Design Goals**

Our design goals can be summarized as follows:

- The adversaries should not be able to get the destination-location information by analyzing the traffic pattern.
- The adversaries should not be able to get the destination-location information even if they are able to monitor a certain area of the sensor network and compromise a few network nodes.
- The length of each message should be as short as possible to save the previous sensor node power.

## **3.3 Proposed Research Directions for Destination-Location Privacy**

### **3.3.1 Directions for Destination-Location Privacy**

In this section, we propose a unique routing technique, called bubble routing, that can provide strong destination-location privacy with low tradeoff in the energy overhead. In our proposed scheme, the source node randomly selects an intermediate node from pre-determined region located around the destination node, which we refer to as the bubble region. The bubble region would be large enough to make it infeasible for an adversary to monitor the entire area. Also, in this scheme, we will mix real messages with fake messages to add to the security strength in providing destination-location privacy.

## **3.4 Preliminary: Destination-Location Privacy Evaluation Model**

In this section, we will provide security analysis of destination-location privacy based on quantitative measurement on information leakage of the destination-location. The quantitative measurement divides information leakage analysis into three categories:

1. *Correlation-based destination identification attack*: Correlation-based attack is an ID based destination node determination. When an adversary receives a message with an ID whose location is already known, the location of this node is also known.
2. *Routing forward attack*: Routing forward is an attack that when a message is forward to the next hop node, an adversary captures a message and identify the immediate message receiver and quickly move to it. For fixed path routing of length  $n$ , if the adversary can capture  $n$  messages from the source, then an adversary can locate the message destination node with  $n$  captured messages.
3. *Reducing destination space attack*: Reducing destination space attack refers to the attack that the adversary can limit the destination node to a proper subset/area in the networks when a message is captured. When multiple messages are captured, the subset/area may be further reduced so that the destination-location can be limited to a subset/area that may lead to the destination node.

To prevent correlation-based destination identification, a dynamic ID based approach can be used to prevent adversaries from relating messages transmitted from each source [72,73]. This can be done by requiring that each node in the network be preloaded with an *ID-hash-chain* so that a different and uncorrelated ID is attached to each message. The adversaries are no longer able to get any useful information about the destination node through correlation-based source identification.

For routing traceforward and reducing destination node space analysis, can be defined in three criteria as follow:

**Definition 4 (Destination-location Disclosure Index (DDI))** *DDI measures, from an information entropy point of view, the amount of destination-location information that one message can leak.*

For a routing scheme, if we assume the total privacy for a destination node  $D$  is 1, and the  $DDI$  is fixed, then the adversary only needs to receive  $\lceil \frac{1}{DDI} \rceil$  messages routed to  $D$  in order to successfully locate  $D$ . Therefore, for a good DLP scheme,  $DDI$  should be as small as possible.

**Definition 5 (Destination-location Space Index (DSI))** *DSI is defined as the set of possible network nodes, or sub-area of the network domain, that can contain the destination node.*

For a routing scheme, if *DSI* is large, means that practically the entire network can contain the destination node. On the contrary, if *DSI* is small, then the adversary can limit the network to a small-sub areas of nodes that possible contain the destination node. To provide adequate destination-location privacy, *DSI* should be as large as possible.

**Definition 6 (Normalized Destination-location Space Index (NDSI))** *NDSI is defined as the ratio of the DSI area over the total area of the network domain. Therefore,  $NDSI \in [0, 1]$ , and we always have  $NDSI = 1 - \delta$  for some  $\delta \in [0, 1]$ . The  $\delta$  is called the local degree.*

It is clear that the scheme with the local degree 0 provides the highest degree of DLP.

Based on these definitions, we have derived in [72] that fixed path routing schemes are the least secure DLP schemes.

**Lemma 2** *Suppose there is a fixed routing path between the source node  $S$  and the destination node  $D$  of length  $L$  hops. Then, after receiving  $L$  messages, an adversary will be able to trace forward to the destination node  $D$ , i.e.,*

$$DDI = \frac{1}{L}.$$

If there are  $n$  disjoint routing paths between the source node  $S$  and destination node  $D$ , and the length of the  $n$  paths are:  $L_1, L_2, \dots, L_n$ , respectively. For each message, the source node  $S$  will send it along path  $L_i$  with probability  $p_i$ , where

$$\sum_{i=1}^n p_i = 1.$$

For path  $i$ , we have  $DDI_i = \frac{p_i}{L_i}, i = 1, \dots, n$ . Define the overall DDI as

$$DDI = \sum_{i=1}^n p_i \cdot DDI_i.$$

**Theorem 4** Suppose there are  $n$  disjoint routing pathes between the source node  $S$  and the SINK node  $D$ . The lengths of the  $n$  routing pathes are  $L_1, L_2, \dots, L_n$ . Let  $p_i$  be the probability that messages will be transmitted along the path  $L_i$ , then when  $p_i = \frac{L_i}{L_1 + L_2 + \dots + L_n}$ ,  $i = 1, 2, \dots, n$ , the SDI is minimized, which is

$$DDI = \frac{1}{L_1 + L_2 + \dots + L_n}.$$

**Corollary 2** Suppose there are  $n$  disjoint routing paths between the source node  $S$  and the destination node  $D$ . The length of the  $n$  routing paths are  $L_1, L_2, \dots, L_n$ , respectively. The adversary then needs to receive on average

$$\frac{1}{DDI} = L_1 + L_2 + \dots + L_n$$

messages to fully determine the location of the destination node, i.e., trace forward to the destination node.

As a result, to provide DLP in wireless sensor networks, we have to increase the total number of possible routing paths between the destination node and the source node. However, for a practical network configuration, the number of routing paths cannot be increased without limitation. This means that we will always have  $DDI > 0$ .

We can summarize the two defects of the DLP schemes through fixed routing pathes as follows:

- **Non-zero DDI:** For fixed path routing, no matter how dedicated the scheme is designed,  $DDI$  is always larger than 0. In other words, for each message sent out by one source node, from a probability point of view, there is always a fraction of destination information to be leaked to the adversaries. So, no matter how small the  $DDI$  is, when enough messages are received, the adversaries are always able to locate the destination node.
- **Limited DSI:** Because the routing paths are fixed to the destination node, the correlation between the messages transmitted on a particular path and the destination node is high. In other words,  $DSI$  is small compared to the overall sensor network size.

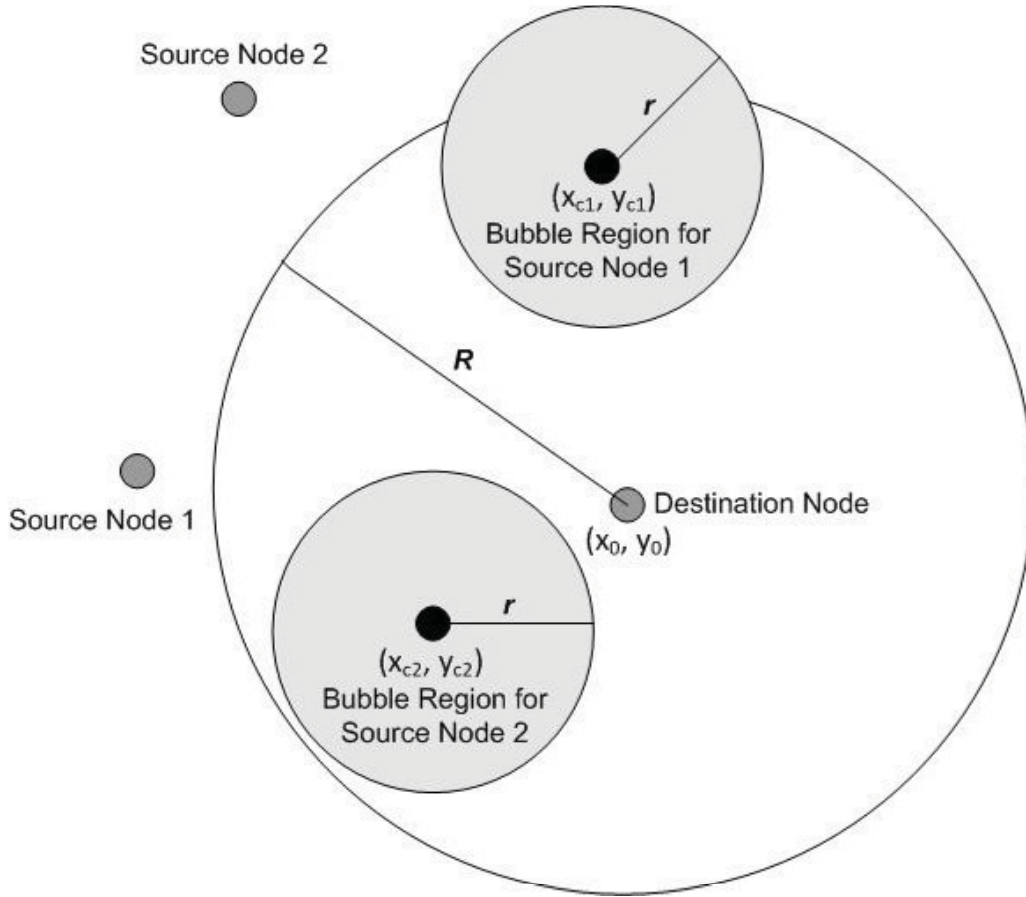


Figure 3.1 Illustration of the Bubble Routing

### 3.5 Destination-Location Privacy using Bubble Routing

In [78], we propose two phase routing scheme to protect the destination location called bubble routing. In the first phase, the source node randomly selects an intermediate node from a pre-determine sub-region, which we will refer to as the bubble region. We assume that the bubble region is large enough to make infeasible for an adversary to monitor the entire bubble region. In the second phase, the intermediate node will then route the message to the destination node.

For the source node to establish an intermediate node, it must do the following steps. The first step is to randomly select a location within a circular area of radius  $R$  around the destination node. We assume that each sensor node only has knowledge of its adjacent nodes relative location and the destination node relative location. Let  $x_0; y_0$  represent the relative location of the destination

node. From these perimeters,  $x_0; y_0; R$ , the source node is able to generate a random point within the pre-determine area. Since we assume that the destination node is located at the relative location  $(x_0; y_0)$ , the source node selects the random location  $(x_c; y_c)$  according to the following procedures:

1. Randomly select  $d_c$  uniformly from  $[0, R]$ .
2. Randomly select  $\theta_c$  uniformly from  $[0, 2\pi]$ .

In this way, we can calculate the coordinate of the random selected point as  $(x_c, y_c) = (x_0 + d_c \cos(\theta_c), y_0 + d_c \sin(\theta_c))$ .

After obtaining the random point  $(x_c; y_c)$ , in the second step, the source node will generate another random point located within a circular area of radius  $r$  around the random point  $(x_c; y_c)$ . We will call the point  $(x_c; y_c)$  the center point for pre-determine sub-region. We will refer this pre-determine sub-region with radius  $r$  as the bubble region. Also, we assume that  $r$  is smaller than  $R$ .

Now with the new perimeters,  $x_c; y_c; r$ , the source node can randomly generate a point  $(x_i, y_i)$  to determine an intermediate node. We can compute the random location  $(x_i, y_i)$  according to the following procedures:

1. Randomly select  $d_i$  uniformly from  $[0, r]$ .
2. Randomly select  $\theta_i$  uniformly from  $[0, 2\pi]$ .
3. Calculate  $(x_i, y_i) = (x_c + d_i \cos(\theta_i), y_c + d_i \sin(\theta_i))$ .

After calculating the random selected point,  $(x_i, y_i)$ , the source node will then route the message towards the grid that contains the location  $(x_i, y_i)$ . Since each node knows its adjacent neighbor nodes's relative location, it can determine the direction that the message should be routed to. Once the message is within the desired grid of the random location, the message is routed to the header node of the grid. The header node then becomes the random intermediate node. If the desired grid does not contain any nodes, then the last node in the routing path would become

the desired location and the header node in that grid would become the intermediate node. We assume that when node detects an event, it will send messages periodically. For every message that is created from a source node, the intermediate node location will be generated using the parameters  $x_c; y_c; r$  and source node will calculate a new point for each message. In other words, the source node will send each message to a different intermediate node located in the same bubble area. Also, each source node in the network will determine its own bubble region by using the steps described above. Figure 3.1 illustrates the Bubble routing scheme. When source node 1 sends a message, it will forward the message to the bubble region 1 and source node 2 will forward messages to bubble region 2. As you can see, the message routes from the source node will not lead an adversary directly to the destination node. Once the intermediate node receives the message, then the message is forwarded to the destination node.

To increase the security on protecting the destination-location, we will inject fake messages into the network. As the real message is being routed in the network by a hop-by-hop approach, every node in the routing path will have the probability,  $p_{fake}$ , to generate a fake message. The real message and the fake messages will be indistinguishable to the adversaries. By mixing real messages with fake messages will make it more difficult for an adversary to determine the direction of the routing path of the real message. A node that generates a fake message will randomly select a neighbor node to send the fake message to and also will forward the real message to the next node in the routing path. The fake message will have a pre-determined time-to-live ( $TTL_{fake}$ ) parameter associated with it. As the fake message passes in the network, for each hop it takes the parameter  $TTL_{fake}$  decreases by 1. The node that receives the fake message with  $TTL_{fake} = 0$  will discard the fake message. Also, when nodes receive fake messages, they can create another fake message with probability  $p_{fake}$ . The purpose of  $TTL_{fake}$  parameter is to limit the energy consumption used in injecting the network with fake messages. We refer to this routing scheme as the bubble routing scheme.

In bubble routing scheme, the security strength in protecting the destination-location relies on the parameters,  $r; R; p_{fake}; TTL_{fake}$ . To increase the security of this scheme you can simply increase

any of the parameters  $r; R; p_{fake}; TTL_{fake}$ . Increasing any of these parameters will also increase the energy consumption. For any destination-location privacy schemes, there will be a tradeoff in security and energy cost. Our goal is to show that a small increase in the energy cost can cause a huge increase in security protection of the destination-location. In the next section, we will analysis the security of our proposed bubble routing scheme and other well known existing schemes.

### 3.5.1 Security Analysis for Bubble Routing

In this section, we will analyze that the proposed bubble routing scheme can provide destination-location privacy. Also, we will analysis other destination-location privacy schemes and determine its vulnerabilities for different adversary attacks.

**Theorem 5** *For the proposed bubble scheme, if assume that the bubble area is large enough so that the probability for multiple messages to be routed using the same intermediate node is negligible. Then the amount of destination-location information that can be leaked from one message is negligible, i.e.,*

$$DDI \simeq 0,$$

*and DLP with local degree 0.*

DEEP routing [37] and LPR routing with fake packet injection [76] have similar adversary attack vulnerabilities. We assume that an adversary have full knowledge of the routing protocol that's being used in the WSNs. With the DEEP protocol, the vulnerability is that the real message is always forward towards the direction of the destination node. In other words, the real message will always be forward on a random shortest path to the destination. The traffic on these random shortest paths will be heavier than the fake messages routing paths. Also, the number nodes involved in forwarding the real messages, from source to destination, will be small due to the small number paths the real message can take. Figure 3.2 illustrate the DEEP scheme routing a real message to the destination node. From the figure, the shortest random path for the message to take from source to destination is 6 hops. As you can see, the real message can only take a small



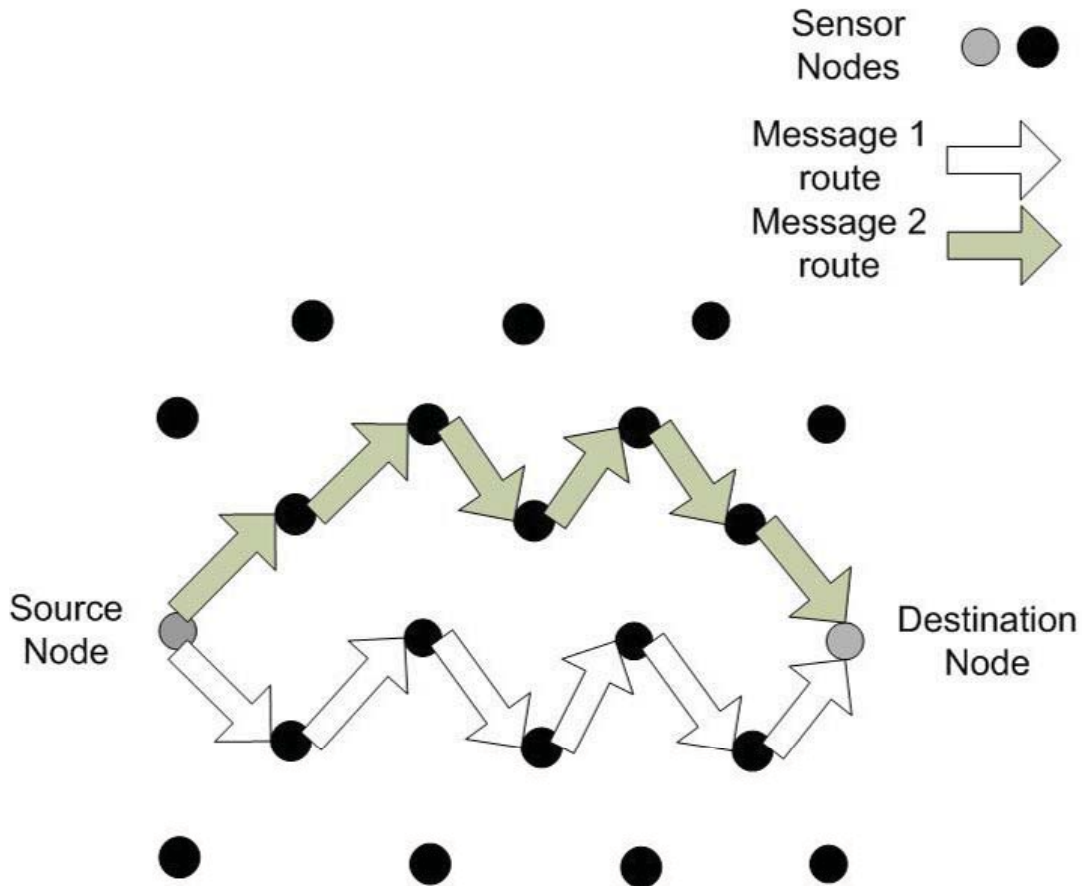


Figure 3.2 Illustration of the routing for messages in the DEEP scheme

number of random paths to reach the destination in 6 hops. The LPR protocol with fake packet injections attempts to solve the vulnerability of the DEEP routing. The LPR protocol randomizes the routing paths so that the forwarding direction of the real message is not always towards the direction of the destination node. The vulnerability of the LPR scheme is that an adversary can use to its advantage the knowledge that more than 50 percent of the real packets will be forward towards the destination node and all of the fake packets will be forward away from the destination. Therefore, if two packets are forward in the similar direction from the same node, the adversary will know that destination node is not in that direction. The vulnerabilities of the DEEP and LPR schemes will be solved with our proposed scheme.

Our bubble routing scheme addresses the vulnerabilities of the DEEP and LPR routing schemes. In our scheme, random intermediate nodes are selected from a bubble region. We assume that the

bubble region is large enough that it would be unpractical for an adversary to monitor entirely. From the probability point of view, for a large network, the chances that the messages will be routed using the same path and the same intermediate node are extremely low. The increase in the number routing paths to the intermediate nodes in the bubble region solves the vulnerabilities of the DEEP routing scheme. In addition to the increase of routing paths for the real messages, our scheme also injects fake messages into the network to make more difficult for an adversary to discover the destination-location. To solve the vulnerability of the LPR scheme, the fake messages are forward to any random neighbor node. Therefore, based off transmission of fake messages, an adversary can not eliminate an area in the network of where the destination node may not be located. Also, if an adversary discovers an intermediate node, he will be located in the bubble region. For most cases, the bubble region will not contain the destination node. After discovering the bubble region, an adversary's work is only half way done. The security analysis shows that our proposed bubble routing scheme can provide much better security in protecting the destination-location privacy in comparison to the DEEP and LPR routing schemes.

## **3.6 Destination Location Privacy using R-STaR Routing Schemes**

### **3.6.1 R-STaR Routing Protocol**

The Reverse-STaR DLP scheme is similar to the STaR routing scheme discussed in the previous chapter. Instead of the STaR area being located around the SINK node, the STaR area is located around the source node, which we call it the Reverse-STaR or R-STaR area. R-STaR routing scheme provides destination-location privacy through a two-phase routing protocol. In the first phase, the source node routes the message to a randomly selected intermediate node located in a pre-determine region around the source node. We call this region the Reverse Sink Toroidal Region (R-STaR). Opposite of the STaR routing scheme, the random intermediate node act as a fake destination node. In the second phase, the intermediate node in the R-STaR area will forward the message to the SINK node using single-path routing.

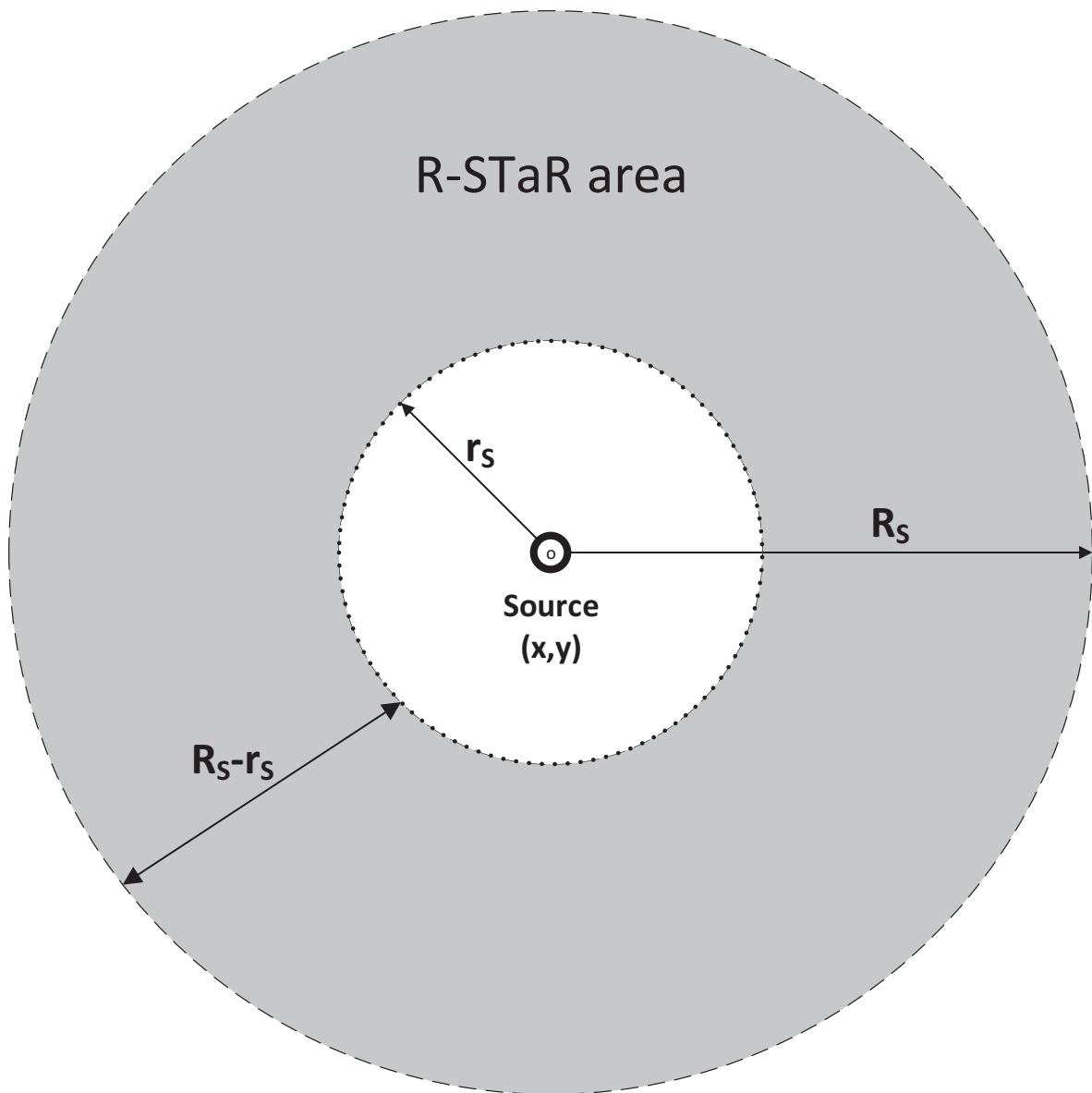


Figure 3.3 Distribution of the R-STaR area

In our scheme, the network is evenly divided into small grids [73]. We assume that the sensor nodes in each grid are all within the direct communication range of each other. In each grid, the header node coordinates the communication with other header nodes nearby. We assume that the whole network is fully connected through the multi-hop communications.

The goal of the proposed scheme is to provide destination-location privacy with adequate

energy-efficient routing. Destination-location privacy is obtained by using this scheme because the source node will not leak any information of the destination node location as the message is routed to the intermediate nodes in the first routing phase. Since the intermediate node serves as a fake location of the destination node in the first routing phase and can be located any direction of the source node with equal probability, an adversary will be unable to determine the destination node sub-area location by just monitoring messages transmitted near the source node. Therefore, the R-STaR routing scheme can provide adequate global destination-location privacy

Global privacy is obtained by the fact that the messages from the source node to the intermediate node will not provide any information as to a any sub-area where the destination node may be located. We assume that the R-STaR area would be a large area with at least a minimum radius distance  $r$  from the source node. Also, the R-STaR area guarantees that the intermediate node is at most a maximum distance  $R$  from the source node to limit the energy consumption in the routing paths in the first phase. This routing scheme is designed to give the illusion that the destination node can located in all the possible directions. In this way, the R-STaR creates an effect that is similar to the totally random RRIN scheme [75] but with less energy consumption and shorter routes.

We assume that each sensor node only has knowledge of its adjacent nodes and has no accurate information of the sensor nodes more than one hop away. We also assume that each node has knowledge of the parameters that are shown in Figure 3.3. The description of the parameters are as follows:

- $x_S, y_S$ : The corresponding X and Y coordinates of the source node location,
- $R_S$ : The pre-determined radius from the source to the outer-edge of the R-STaR area,
- $r_S$ : The pre-determined radius from the source node to the inner-edge of the R-STaR area.

From these parameters,  $\{x_S, y_S, R_S, r_S\}$ , the source nodes are able to generate random points within the R-STaR area. Since we assume that the source node is located at the relative location

$(x_S, y_S)$ , the source node determines intermediate node location  $(x_i, y_i)$  according to the following steps:

1. Randomly select  $d_i$  uniformly from  $[r_S, R_S]$ .
2. Randomly select  $\theta_i$  uniformly from  $[0, 2\pi]$ .
3. Calculate the coordinate of the intermediate node as  $(x_i, y_i) = (x_S + d_i \cos(\theta_i), y_S + d_i \sin(\theta_i))$ .

After obtaining the random location  $(x_i, y_i)$ , the message can then be routed towards the grid that contains the location  $(x_i, y_i)$ . Since each node only knows its adjacent neighbor nodes' relative location, it can determine the direction that the message should be routed. Once the message is within the desired grid of the random location, the message is routed to the header node of the grid. The header node then becomes the random intermediate node. If the desired grid does not contain any nodes, then the last node in the routing path would become the desired location and the header node in that grid would become the intermediate node. The intermediate node then routes the received message to the destination node using single-path routing.

To provide additional security for the second phase routing path, we will have the destination node inject fake messages into the network. When the real message is received by the destination node, the destination node will generate a fake message into the network to give the illusion that the message is continued to be forward in the network to hide the identity of real destination node. Without this additional security feature, an adversary is able to locate the real destination node location by simply determining where the last hop node of the message routing path. But by adding the security feature of having the destination node forward a fake message, upon receiving the real message, along the same path trajectory of the real message, an adversary will not be able to distinguish destination node from any of the forwarding nodes in the routing path. This additional security feature will provide local network privacy for the destination-location.

For conserving the energy of the network, the fake message that is generated by the real destination node will have certain Time-To-Live (TTL) value for the number of hops the it is forward in

the network. We will denote the TTL value for the fake message as  $N_{TTL}$  and can be determined by the following equation,

$$N_{TTL} = \lceil \alpha \cdot N_{hop} \rceil,$$

where  $\alpha$  is a random generated value from  $[0, 1]$  and  $N_{hop}$  is the hop count of the real message from the source node to the destination node. If the source node is located within 10 hops from the destination node,  $N_{TTL}$  will be derived as,

$$N_{TTL} = \lceil (\alpha \cdot N_{hop}) + 10 \rceil, N_{hop} < 10,$$

to ensure that the fake message will forward at least 10 hops away from the destination node when the source node is located is near the destination node.

The fake generated message must take the same path trajectory as the real message as it was routed from the intermediate node to the destination node. And to do so, we must determine the following values:  $d_f, \theta_f, X_f, Y_f$ . Where  $(X_f, Y_f)$  are the coordinates of a random point on the trajectory path for the fake message route. To determine this random point, we must first determine  $d_f$ , which is the distance from the destination node to the random fake point  $(X_f, Y_f)$ . The destination node will determine  $d_f$  as follow,

$$d_f = T_r \cdot N_{TTL},$$

where  $T_r$  is the transmission range (in meters) of wireless sensor devices. This equations ensure that the random fake point,  $(X_f, Y_f)$ , will be at least  $N_{TTL}$  hops from the destination node.

Second, we must determine the  $\theta_f$ , which is the angle of message path trajectory from the destination node to the fake random point. To do so, we must determine the slope of trajectory path of the real message from the intermediate node to the destination node, which we will denote this slope as  $m_{slope}$ . Let  $(X_i, Y_i)$  and  $(X_d, Y_d)$  represent the coordinates of the intermediate node and destination node, respectively. Therefore,  $\theta_f$  can be determine as follow,

$$\theta_f = \tan^{-1}(m_{slope}) = \tan^{-1}((Y_d - Y_i)/(X_d - X_i)).$$

With  $d_f$  and  $\theta_f$ , we can determine the random point  $(X_f, Y_f)$  as follow,

- $X_f = \cos(\theta_f) \cdot d_f + X_d$
- $Y_f = \sin(\theta_f) \cdot d_f + Y_d$ .

Now with the random point  $(X_f, Y_f)$ , the destination node can send the fake generated message on the same trajectory path as the real message. The fake message will only be forward in the the direction of the random  $(X_f, Y_f)$  point for  $N_{hop}$  hops. If the fake message reach a node on the edge of the network, this node in the forwarding process will discard the fake message and act as the destination node of the message.

### 3.6.2 R-STaR Routing Protocol with Fake Destination Nodes

To enhance the security for destination-location privacy using the R-STaR routing scheme, we will have the source node send messages to fake destination locations. The purpose is to provide global-network security for the second routing phase in the R-STaR routing scheme. In the second phase of the R-STaR routing scheme, the message transmits from the intermediate nodes to the destination using shortest/single path routing. If all messages are transmitted from the R-StaR area to the destination node directly, the second phase presents some security vulnerabilities. To help prevent some of these security vulnerabilities, we designed the R-STaR routing scheme with fake destination nodes.

The fake destination nodes will be located, relatively, the same distance from the source node as the real destination node. Let the real destination location coordinates be  $(X_D, Y_D)$  and the distance from the source node to the real destination node be denoted as  $d_D$ . The following steps will taken by the source node to determine a random fake location:

1. Randomly select  $\theta_f$  uniformly from  $[0, 2\pi]$ .
2. Calculate the coordinate of the fake destination location as  $(X_f, Y_f) = (X_S + d_D \cos(\theta_f), Y_S + d_D \sin(\theta_f))$ .

To increase security of this scheme, the source node will generate  $n_f$  fake destination locations to route the messages to. An increase in  $n_f$  is an increase in security in the second routing phase for the R-STaR routing scheme. Therefore, for each fake destination node location, the source node will have to generate  $n_f$  random coordinates as follow:

- Let  $i$  represent values  $[0, \dots, n_f]$ .
- Randomly select  $\theta_f^i$  uniformly from  $[0, 2\pi]$ .
- Calculate the coordinate of the  $i^{th}$  fake destination location as  $(x_f^i, y_f^i) = (x_S + d_D \cos(\theta_f^i), y_S + d_D \sin(\theta_f^i))$ .

The Source node will send messages to fake destination location  $(x_f^i, y_f^i)$  at probability rate, which will be denoted as  $p_D$ . Each fake destination node location will have even probability to be selected by the source node as the destination. Also, at the same probability rate  $p_D$ , the source node will have in selecting the real destination node as the location to route the message to. Therefore, we have

$$p_D = \frac{1}{n_f + 1}.$$

From this equation, we have the following findings:

- if  $n_f = 0$ ,  $p_D = 1$ , then every message will be routed to the real destination node (The original R-STaR routing scheme);
- if  $n_f = 1$ ,  $p_D = 0.5$ , then 50% of the messages will be routed to the real destination node and 50% will be routed to the fake destination node;
- if  $n_f = 2$ ,  $p_D = 0.333\dots$ , then 33.3% of the messages will be routed to the real destination node  $D$ , 33.3% will be routed to first fake destination node  $D_f^1$ , 33.3% will be routed to second fake destination node  $D_f^2$ .

We must keep in mind when a source node detect an event, the source node will transmit messages periodically as long as it is still detecting an event. Prior to detecting an event, the source node



will generate  $n_f$  random fake destination node locations, similar to example shown on Table 3.1. Each source node in the network will create its own distinguish destination node location table. For each message transmitted, the source node will generate a random intermediate node location within the R-STaR area for the first routing phase and from the list of  $n_f$  random fake destination node locations and the real destination node, the source node will randomly determine, for each message, if the message will be sent to the real destination node or one of the  $n_f$  fake destination nodes for the second routing phase. Each fake destination and real destination node location can be selected at an equal probability rate of  $p_D$ . With using the R-STaR routing scheme with fake destination node locations, we can provide routing security for both phases in the routing path without any additional energy consumption in comparison to the original R-STaR routing scheme.

<b>Dest. Nodes</b> (with $n_f = 4$ )	<b>Coordinates</b>	<b>Probability</b> ( $p_D$ )
$D$ (Real Dest. Node)	$(x_D, y_D)$	$p_D = 1/(n_f + 1) = 0.2$
$D_f^1$	$(x_f^1, y_f^1)$	$p_D = 0.2$
$D_f^2$	$(x_f^2, y_f^2)$	$p_D = 0.2$
$D_f^3$	$(x_f^3, y_f^3)$	$p_D = 0.2$
$D_f^{n_f} = D_f^4$	$(x_f^4, y_f^4)$	$p_D = 0.2$

Table 3.1 Example destination node location table with  $n_f = 4$

### 3.6.3 R-STaR Routing Protocol mix with Cost-Aware Routing (CAR)

The cost-aware routing (CAR) algorithm that was proposed in [79]. CAR was designed to provide an energy-efficient routing technique for source-location privacy. For the proposed R-STaR routing mix with CAR routing is a two-phase routing scheme to provide balanced energy-efficient routing algorithm for destination location privacy. The first phase routing is the same as the R-STaR routing where the message will be routed to a random intermediate node located in the R-STaR area but using CAR energy-efficient routing technique. In the second phase, the message will be routed to the real destination node or a fake destination node location using CAR energy-efficient routing technique. R-STaR mix with CAR routing can provide an additional security feature due to the

fact that CAR routing randomizes and increases the number of routing paths for the real message to take from source node to intermediate node and from the intermediate node to the destination node. Without CAR routing, in the first routing phase transmit the message to the intermediate node using shortest path routing and does the same in the second phase when transmitting from the intermediate node to the destination node.

### 3.6.4 Security Analysis for R-STaR Routing Schemes

In this section, we will analyze the security for the R-STaR routing, R-STaR routing with fake destination nodes and R-STaR mix with cost-aware routing. We assume that the adversary is unable to monitor the entire network and can only trace forward to the next node that is one hop away with each transmitted message in the network. We will analyze the security all three proposed schemes by analyzing the security of each routing phase.

#### 3.6.4.1 First Routing Phase

For R-STaR and R-STaR with fake destination node routing schemes, the first routing phase is the same. In the first routing phase, the source node route the message to a random intermediate node located in R-STaR area using the shortest routing path. For the R-STaR mix with Cost-Aware Routing (CAR), the first routing phase uses CAR algorithm when routing the message from the source node to the random intermediate node located in the R-STaR area. Besides the differences in the routing technique that is used in routing the message to the random intermediate node, all three schemes routes the message to an intermediate node in first routing phase that can be located in any direction not associated with the location of the destination node. Therefore, in the first routing phase for all three schemes, no information is leaked about the destination node location by analyzing traffic in the first routing phase.

**Theorem 6** *For the proposed schemes, if assume that the R-STaR area is large enough so that the probability for multiple messages to be routed using the same intermediate node is negligible*

and equal probability of selecting an intermediate node from all possible directions. Therefore, the amount of destination-location information that can be leaked from all messages in the first routing phase is negligible, i.e.,

$$DDI \simeq 0,$$

and DLP with local degree 0.

Also, DSI will include all nodes in the network as possible destination nodes. Therefore,

$$DSI = N,$$

where  $N$  represent the number nodes in the entire network environment, and

$$NDSI = \frac{DSI}{N} = 1.$$

### 3.6.4.2 Second Routing Phase

To analyze the security for the second routing phase for the R-STaR, R-STaR with fake destination node and R-STaR mix with cost-aware routing, we will assume that the adversary have identified the entire R-STaR area dimensions and is trying to monitor messages in the second routing phase. For simplicity, we will further assume that network environment is a circle region with a radius  $R_N$ , as seen in figure 3.4. To help analyze the security strength of the proposed routing techniques, we will break the entire network environment into 3 regions.

- Region 1: The source node region of the network. This region have a radius  $r$  with the source node being the center location of the region.
- Region 2: The R-STaR region of the network. This region have outer radius of  $R$  and inner radius of  $r$ . It is located around the source node location.
- Region 3: Outside the R-STaR region. This region is the remaining area of the network that is outside of region 1 and region 2.

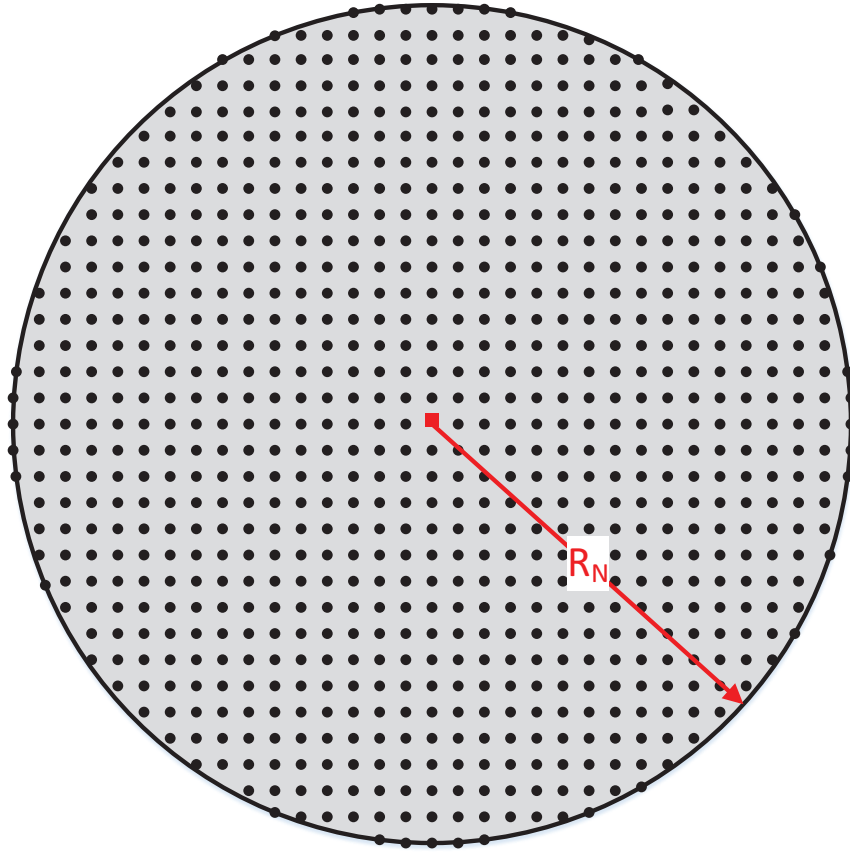


Figure 3.4 Circle Grid Layout Network Environment with radius  $R_N$

The second routing phase transmit the real message from the random intermediate node to a destination node. Recall, upon the destination node retrieving each real message, the destination node will generate and transmit a fake message into the network. The fake generated message will be sent in the network for random number of hops,  $N_{TTL}$ , on the same trajectory as the real message help conceal the destination node location and identity. To further help analyze the routing techniques, we will break the routing path into routing phases.

- Route Phase 1:  $S \rightarrow I \in m_r$  : The routing path of the real message,  $m_r$ , from the source node,  $S$ , to the random intermediate node,  $I$ .
- Route Phase 2:  $I \rightarrow D \in m_r$  : The routing path of the real message,  $m_r$ , from the random intermediate node,  $I$ , to the real destination,  $D$ .

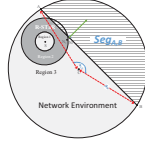


Figure 3.5 Illustration for analyzing the security strength using R-STaR routing

- Route Phase 3:  $D \rightarrow D_f \in m_f$  : The routing path of the fake message,  $m_f$ , from the real destination,  $D$ , to the fake destination node,  $D_f$ .

With the assumption that the adversary knows location and the dimensions of the R-STaR area, we will analyze how much information an adversary can gain from observing one message transmitted from node  $N_1$ , in region 2 (R-STaR area), to node  $N_2$ , in region 3. If an adversary observes a communication between a node in the R-STaR area and one outside the R-STaR area, he is able to assume that the message is transmitting on the second or third routing phase. For analyzing purposes, the adversary assume that the messages are being routed in the route phase 2:  $I \rightarrow D \in m_r$ . With the described assumptions, we must analyze how much information is leaked to adversary with observing one message, one hop outside the R-STaR area when using the proposed destination location privacy R-STaR routing techniques. Figure 3.5 helps illustrate the communication between node  $N_1$  and node  $N_2$ .

If an adversary is able to observe one message transmission from a node  $N_1$  and node  $N_2$  and assume that the message is routed on the second phase, an adversary can assume that the message will be routed on the same trajectory as the angle between the two nodes. With the information that the message will be routed on a shortest path to the destination, the an adversary can assume that the destination node would be located give or take 90 degrees from the trajectory path, providing that the layout of the network is grid as shown in figure 3.4. In figure 3.5, the arrow from  $N_2$  shows the assume trajectory of the message.  $Seg_{A,B}$  area represent the area that an adversary can assume where the destination node can be located. Table 3.2 defines the variables in figure 3.5 and these variables will be used to analyzed the security strength of the proposed DLP R-STaR routing schemes.

Variables	Descriptions
$R_N$	Radius of the Network Environment (N.E.)
$C$	Center point of the Network Environment
$N_1$	The node location one-hop inside the R-STaR area in Region 2
$N_2$	The node location point one-hop outside the R-STaR area in Region 3
$\theta_{A,B}$	Angle from center $C$ of the N.E. to points $A$ and $B$
$Seg_{A,B}$	Segment region between points $A$ and $B$

Table 3.2 Table of variables and descriptions for analyzing security of the proposed R-STaR schemes.

The segment area,  $Seg_{A,B}$ , can be determine using the following equation:

$$Seg_{A,B} = \frac{R_N^2}{2} \cdot \left( \frac{\pi}{180} \cdot \theta_{A,B} - \sin(\theta_{A,B}) \right).$$

When using the original R-STaR routing technique and observing the communication between  $N_1$  and  $N_2$ , the amount of information that can be leak when assuming the message is in route phase 2:  $I \rightarrow D \in m_r$ , can be determined as follow:

Let

$$DSI = Seg_{A,B}$$

and

$$NDSI = \frac{DSI}{TheEntireNetworkArea}.$$

Therefore,

$$DSI = \frac{R_N^2}{2} \cdot \left( \frac{\pi}{180} \cdot \theta_{A,B} - \sin(\theta_{A,B}) \right),$$

$$NDSI = \frac{Seg_{A,B}}{\pi \cdot R_N^2},$$

$$NDSI = \frac{1}{2 \cdot \pi} \cdot \left( \frac{\pi}{180} \cdot \theta_{A,B} - \sin(\theta_{A,B}) \right).$$

For R-STaR routing scheme with fake destination nodes, the second routing phase transmit the real message from the random intermediate node to a real or fake destination nodes using shortest path routing. If an adversary observes the communication between node  $N_1$  and node  $N_2$ , the adversary would not know if the message is being routed towards the real destination or one of the fake destinations. Therefore, the R-STaR routing scheme with fake destination nodes would provide a greater security than just the original R-STaR routing scheme. Increase in the number of fake destination nodes, will provide an increase in the security for the second routing phase using this scheme.

For R-STaR routing mix with cost-aware routing (CAR) scheme, the second routing phase transmit the real message from the random intermediate node to the real destination using cost-aware routing technique. Since this scheme does not use the shortest path routing during the second phase, an adversary can not assume the direction and/or sub-area of where the destination node may be located by only observing the communication between node  $N_1$  and node  $N_2$ . Therefore, the R-STaR routing mix with CAR scheme will provide greater security than the original R-STaR routing scheme.

To provide a greater security scheme, we can combine our R-STaR routing with fake destination nodes scheme and our R-STaR routing mix with CAR routing scheme. Using these combination of schemes, the second phase would transmit the real message to the real or fake destinations using CAR routing technique instead of the shortest path routing. With the combination of the two schemes can prove to be provide very strong destination location privacy scheme by eliminating any vulnerabilities that either scheme may present when used alone.

### **3.6.5 Performance Analysis and Simulation Results**

To evaluate the performance of the schemes proposed [80], extensive simulations have been conducted using ns-2 on RedHat Linux system. The results of the simulations are shown in Figure 3.6 and Figure 3.7. In the simulation, 400 nodes are evenly distributed in a square target area of size  $2100 \times 2100$  meters network environment. We illustrate the performance of the totally ran-

domly selected intermediate nodes (RSIN) and the R-STaR routing protocol. For RSIN scheme, the source node can select any node in the network as the intermediate node with equal probability outside a radius minimum distance of  $d_{min}$ . We set  $d_{min}$  to equal 395 meters. For R-STaR routing, the inner radius,  $r$ , was set to 395 meters, while the outer radius,  $R$ , was set to 605 meters.

Through analysis and simulation results, we find that R-STaR protocol provide the best results. R-STaR provides similar level of security as the RSIN scheme but with less energy consumption and shorter delays.

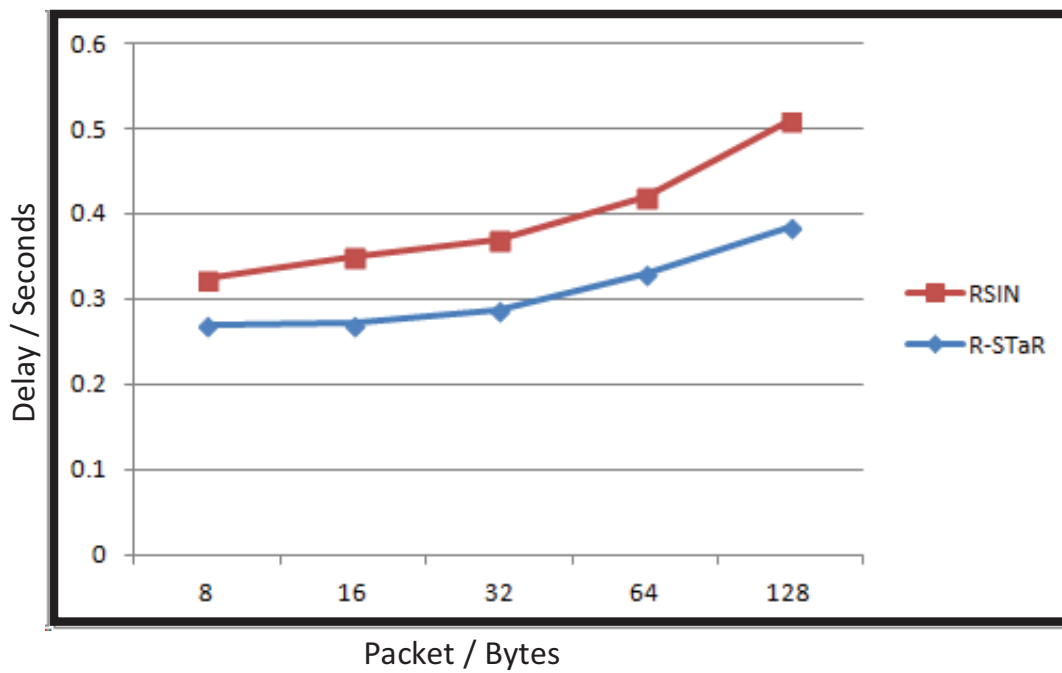


Figure 3.6 Message Latency



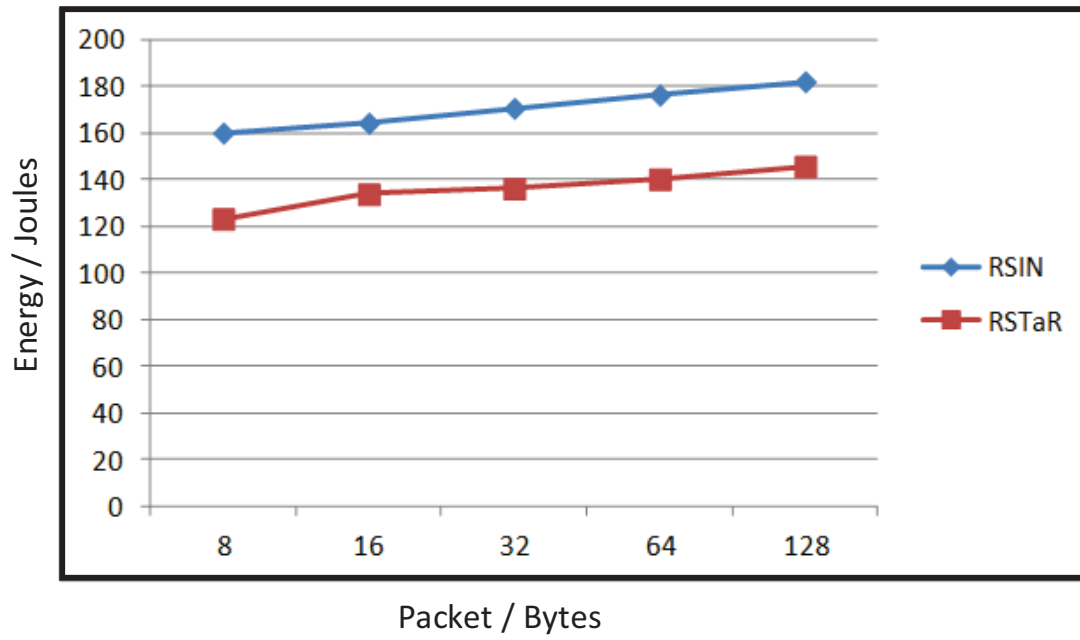


Figure 3.7 Energy Consumption

## CHAPTER 4

### Q-RECHS CLUSTER-BASED ENERGY EFFICIENT ROUTING IN WIRELESS SENSOR NETWORKS

#### 4.1 Introduction

With the popularity of wireless sensor networks, many issues have been raised over the past several years. Prolonging the lifetime of WSN's is critical concern. One of the usage of WSN's is to provide real-time event monitoring functionality. To do its job, the device energy must be a high priority for these devices to perform its functionality. In a network environment, having a portion of the network filled with dead devices does not serve much purpose, therefore we will define the network lifetime as the timescale (rounds) it takes for certain percentage of the network nodes to die. In this chapter, we proposed a cluster-based energy-aware routing scheme called Quad-Region Cluster-Head Selection (Q-ReCHS), that will prolong the network lifetime by evenly distributing the energy load among all the nodes. We will provide simulation results using defined metrics for measuring lifetime of the network: First-To-Die(FTD), Half-To-Die(HTD), and Last-To-Die(LTD). We will measure the number of packets sent to the Base Station (BS) and number rounds it take for schemes to reach the defined metrics.

#### 4.2 Related Work

Many different approaches have carried out to design feasible WSNs. Energy conservation is crucial to prolong the network lifetime of WSNs. Many approaches for energy efficient routing have been proposed to reduce energy consumption. One alternative approach to conserve energy is using clustering technique. In addition, when scalability is considered to be a major problem when network density is of hundreds and thousands of nodes then clustering is considered to be a useful technique. In various WSNs applications, routing efficiency is considered important for

energy efficiency, load balancing, and data fusion [81]. In this chapter, we are concern about Cluster-Head (CH) selection schemes and discuss some of the associated schemes.

Low Energy Adaptive Clustering Hierarchy (LEACH) [82] is well know clustering algorithm in which the CH in the cluster is periodically rotated among members to achieve energy balance. However, this scheme showed only partial success, it needs a new cluster formation process at every section. With cluster formation, in each cluster a new CH node is re-elected with random probability, and from the promising CH candidates. The optimal node should be adaptively optimized for minimum communication distances to the maximum number of one hop neighbors. This only produce worst suboptimal solution due to cluster re-election process, which results in the nodes to spend additional delay and energy. In addition, LEACH does not take in account the remaining energy of the node in the CH selection process.

In [64], the authors proposed a protocol called HEED, where a node uses its parameters communication cost between Cluster-Members (CM) and remaining energy to probabilistically elect itself to become a CH. HEED is a multi-hop clustering algorithm for WSNs. The remaining energy of each sensor node is used to probabilistically determine and choose the first set of CHs, as usually performed in other clustering techniques. In [64], communication cost between the Cluster-Members shows the node degree or nodeas proximity to the neighbor and the main parameter that decides whether to join the cluster. However, hot spot issue in HEED appears in areas that are close to the sink, as nodes in such areas need to relay incoming traffic from other parts of the network. Furthermore, knowledge of the whole WSNs is necessary to determine communication cost between Cluster-Members. HEED is a distributed clustering mechanism in which CH nodes are picked from the WSNs. HEEDs CH selection parameter is a hybrid of energy and communication cost.

### 4.3 Energy Consumption Model

In this paper, we use a radio model proposed in [82] as radio energy model to measure energy consumption. We assume nodes dissipate energy when either in the transmit mode  $T_x$  or the receiver mode  $R_x$ .

The energy consumed during the transmit mode consists of transmitter circuitry  $t_c$  and transmitter amplifier  $t_a$ . Therefore, we have transmit mode energy model as:

$$E_{tx} = (t_c \times k) + (t_a \times k \times d^2), \quad (4.1)$$

where  $k$  is the packet size in bits and  $d$  represent the distance of the transmit range. The energy consumed during the receiver mode consists of receiver circuitry  $r_c$ . Therefore we have receiver mode energy model as

$$E_{rx} = r_c \times k. \quad (4.2)$$

In our work, we use the similar assumption as [82], where the transmitter circuitry  $t_c$  and receiver circuitry  $r_c$  consumes  $50 \text{ nJ/bit}$  and the transmitter amplifier  $t_a$  is set to  $100 \text{ pJ/bit/m}^2$ . We also assume that all sensor nodes are sensing the environment at a fixed rate and always detects an event every round; thus always have data to send to the CH each round. Listening energy can be neglected in this paper since all nodes act as source nodes during every round.

### 4.4 Proposed Q-ReCHS Cluster-Based Routing Scheme

Our proposed cluster-based routing protocol is designed to prolong the lifetime of the wireless sensor network. We call our scheme Quad-Region Cluster-Head Selection (Q-ReCHS). Q-ReCHS will provide a better evenly distribution of energy load among the nodes in the network compare to the well known LEACH cluster-based routing scheme.

#### 4.4.1 Q-ReCHS Network Model

In this section, the network model of our proposed Q-ReCHS routing scheme is defined as follows:

- Nodes are randomly distributed throughout a rectangular network environment.
- Each node is assigned to one of four network regions as shown in fig. 4.1.
- All nodes have similar capabilities (Transmission Power Levels, Initial Energy). Each node can be a Cluster-Member (CM) or a Cluster-Head (CH).
- Nodes are left unattended, therefore, recharging or replacing batteries are not feasible.
- During each round, every alive node act as a source node and all transmit one packet to the Base Station (BS). After each round, a new set of CHs are selected
- We used defined metrics for measuring lifetime of the network: First-To-Die(FTD), Half-To-Die(HTD), and Last-To-Die(LTD).
- Nodes use TDMA schedule within each cluster.
- Each node can use 2 frequency channels to transmit:
  - Intra Cluster Communication - Unique frequency channel to transmit from CM to its respective CH.
  - Inter Cluster Communication - Unique frequency channel to communicate to other CHs and/or BS.
- Each node can be in any of the following modes:
  - *Cluster-Member (CM)*: Regular cluster node; When a Cluster-Member (CM) detect an event, the CM become the source node and transmit the packet directly to the its corresponded CH.

- *Cluster-Head (CH)*: Serve as the head of the cluster; CH will received packets from CM within it's cluster. After receiving a packet from CMs, CH can transmit compressed packets directly to the BS.

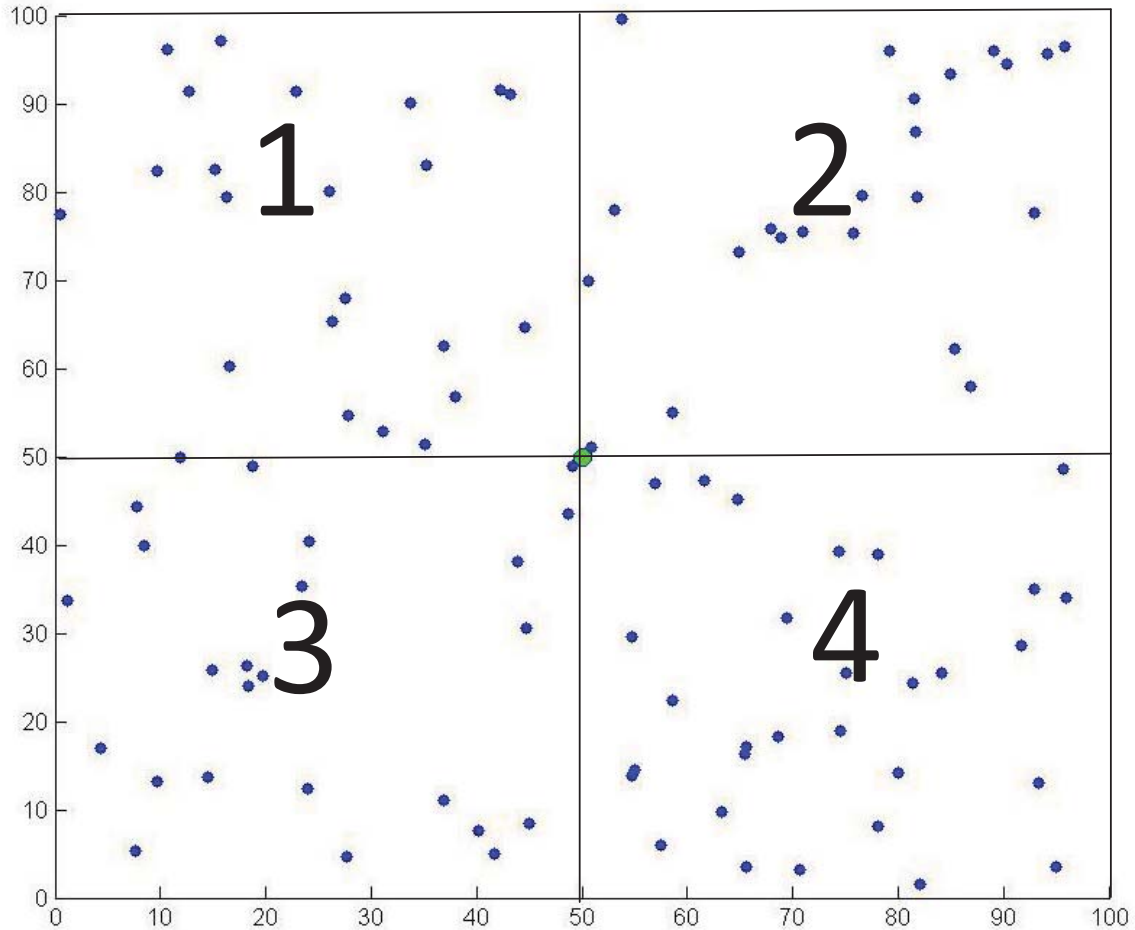


Figure 4.1 Q-ReCHS Network Regions: 100 nodes, 100m x 100m, BS located (50m,50m)

#### 4.4.2 Initialization Stage

During the network set up, the Base Station (BS) node will broadcast a beacon signal that covers the entire network. Each node receives the beacon signal and use the received signal strength to determine the distance to the BS. For prolonging the network lifetime, nodes using should minimize the transmit power to send packets to the BS. In this dissertation, we will define the

distance to the BS as  $d_i^{BS}$ , where  $i$  represent the  $i^{th}$  node in the network.

After the nodes establish its corresponding distance to the BS,  $d_i^{BS}$ , each node will establish its region location within the network. As mentioned in section 4.4.1, the network is divided into 4 equal regions. We assume that each node knows its relative location within the network using a GPS receiver or pre-programmed location coordinates for each node. Using its relative location, each node will assign its self to 1 of the 4 regions shown in Figure 4.1.

#### 4.4.3 Q-ReCHS Cluster-Head Selection Process

According to the study presented in [83], a network containing 100 nodes in a  $100\ m \times 100\ m$  network environment, it was found that the optimum number of clusters should be around 3-5 to minimize the average energy consumption per round. Therefore, to keep our scheme within the optimum range of CHs in the network, the BS will use the following condition to determine the number of CHs that should be assigned to each region:

$$C_\alpha = \left\lceil \frac{N_\alpha}{26} \right\rceil. \quad (4.3)$$

Let  $\alpha$  represent the region id, where  $\alpha \in 1, 2, 3, 4$  and let  $N_\alpha$  represent the number active nodes in the designated region. Therefore, we have

$$\sum_{\alpha=1}^4 N_\alpha = N \quad (4.4)$$

and

$$\sum_{\alpha=1}^4 C_\alpha = C, \quad (4.5)$$

where  $N$  represents the number of active nodes in the network and  $C$  represents the number of CHs in the network for the current round.

To determine the cluster-heads (CHs) in the network, each active node (alive node) in the network will send its node ID, the current remaining energy level, and the region ID to the Base Station. Let  $E_i$  represent the current remaining energy level of the  $i^{th}$  node. After the BS receives

the information from all active nodes in the network, the BS will then compute the average region energy level for all four regions. We denote the average region energy level as  $E_{\alpha}^{avg}$ , where  $\alpha$  represents the region ID. The average region energy level will be computed as follow:

$$E_{\alpha}^{avg} = \frac{\sum_{i=1}^{N_{\alpha}} E_i}{N_{\alpha}}, \alpha \in 1, 2, 3, 4. \quad (4.6)$$

Among the active nodes, the BS will create a subset of nodes for each region that fit the following condition:

$$E_i \geq E_{\alpha}^{avg}. \quad (4.7)$$

From the subset of nodes that satisfy equation (4.7), the BS will randomly select  $C_{\alpha}$  cluster-heads for each region. The selected nodes will serve as the CHs only for the next round.

The clusters will be formulated by having each CH send out a network wide beacon signal to the network with its node ID. From the received signals, each node will cluster with the CH with the strongest signal strength (hence, closest CH) and become cluster-members (CM) with the closest cluster-head. Each node will align with the CH that would require the least amount energy to transmit to. Each cluster-member (CM) should use minimum transmit power to transmit a packet to a CH. Therefore, the CM will determine the distance to its corresponding CH from the CH beacon signal to minimize the transmit power required to reach the CH. We will denote the distance from CM to the CH as  $d_i^{CH}$ . To provide additional energy consumption awareness, a node can opt out becoming a CM if the node distance to the BS is less than the the distance to nearest CH. In other words, if

$$d_i^{BS} \leq d_i^{CH}, \quad (4.8)$$

then the node can save energy by transmitting directly to the BS rather than transmitting to the nearest CH. This approach will provide network wide energy savings because CHs will have less CM which will result in less energy consumption. After a CM detects an event, the CM will transmit a packet to its corresponding CH. For energy savings, the CH compresses all received packets from its CMs using lossy compression before transmitting to the BS.



#### 4.4.4 Systems Analysis

By having the CHs located in all 4 regions, this will solve some of the weaknesses of LEACH protocol, where CHs can be selected anywhere in the network without considering the location that would minimize the distance that CMs will have to transmit to CHs. Also, LEACH protocol does not use remaining energy as a criteria for selecting CHs, which is another weakness for the LEACH protocol, while Q-ReCHS only uses energy rich nodes as CHs since CHs will consume the most energy in each round.

### 4.5 Simulation Results

In this section, we evaluated the performance of our proposed Q-ReCHS scheme and existing schemes, using MATLAB. In our simulation, we assumed an error free physical layer and an ideal MAC layer. Also, we assumed that the energy consumption in the transmission and reception of data packets follows the energy consumption model defined in Section 4.3. We compare our scheme with the LEACH protocol and the protocol using direct transmission from source to the BS. We compare the simulation based on the amount of packets sent to BS vs. number of rounds (timescale) to measure the lifetime of the network: First-To-Die(FTD), Half-To-Die(HTD), and Last-To-Die(LTD).

In our simulation, the nodes are randomly distributed throughout the network, as shown in Figure 4.2. The initial energy for all nodes is set to  $0.5 J$ . For Q-ReCHS, the network is divided into 4 equal region areas, as shown in Figure 4.1. For LEACH protocol, we use the probability  $P = 0.05$  in the simulations. We use several network environments to test the performance of the schemes. The network environments are described below:

1.  $N = 100; X_{dim} = 100m; Y_{dim} = 100m,$
2.  $N = 150; X_{dim} = 150m; Y_{dim} = 150m,$
3.  $N = 200; X_{dim} = 200m; Y_{dim} = 200m.$

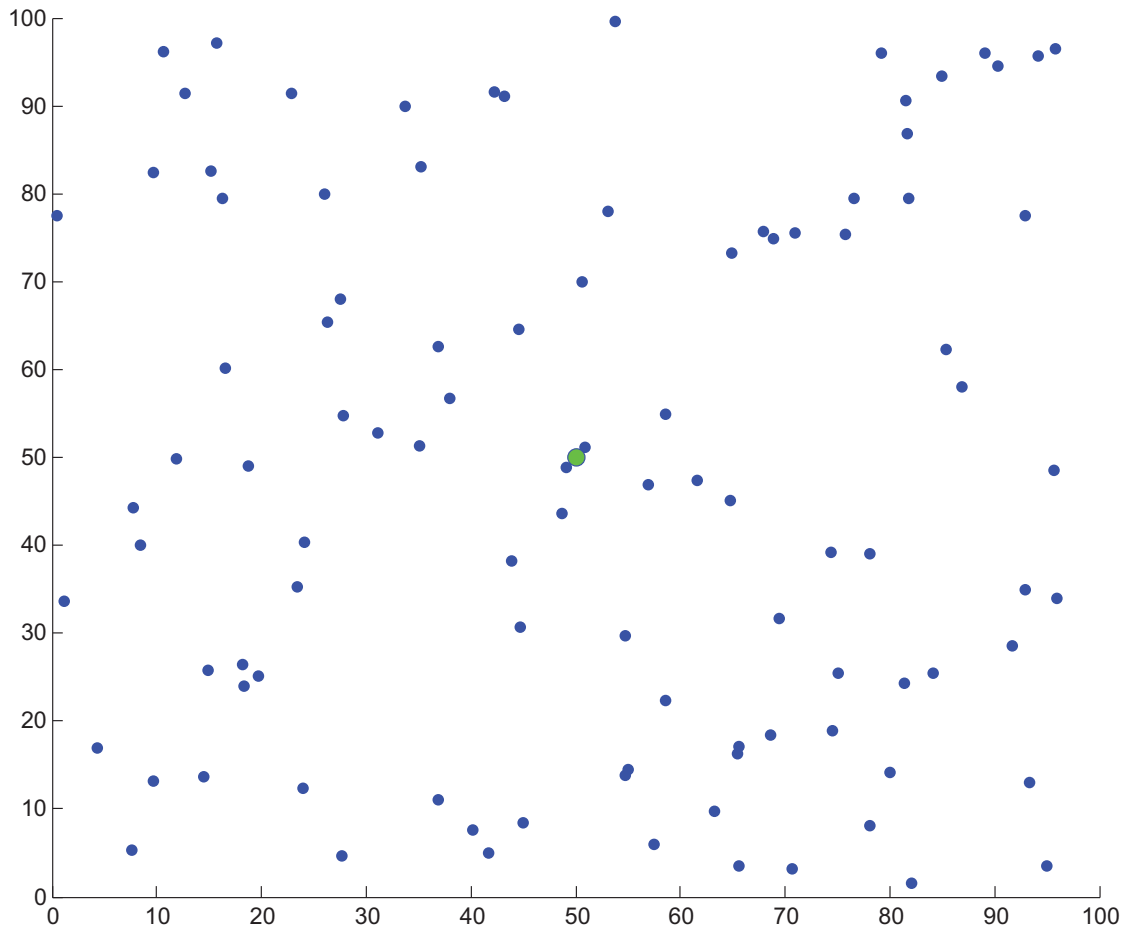


Figure 4.2 Simulation Network Environment: 100 nodes, 100m x 100m, BS located (50m,50m)

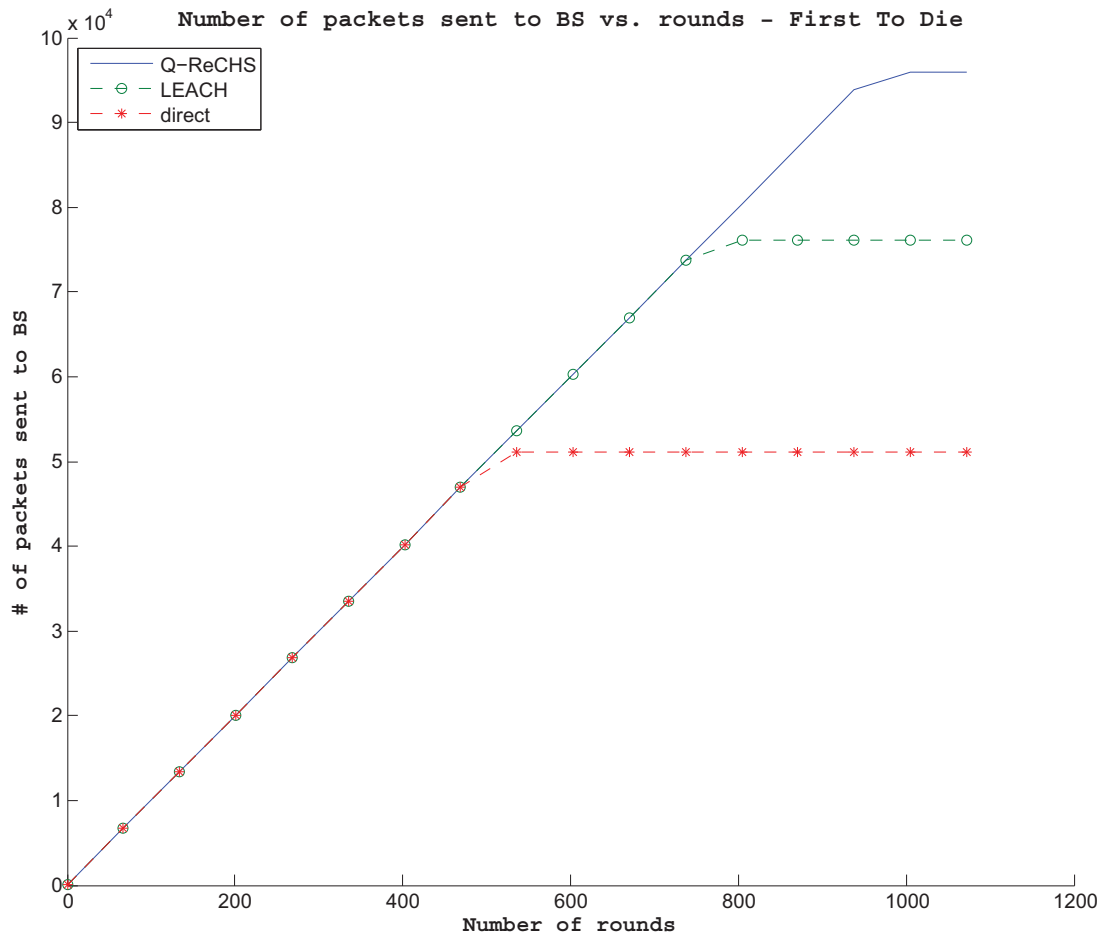


Figure 4.3 100 Nodes in 100m x 100m Environment: First-To-Die

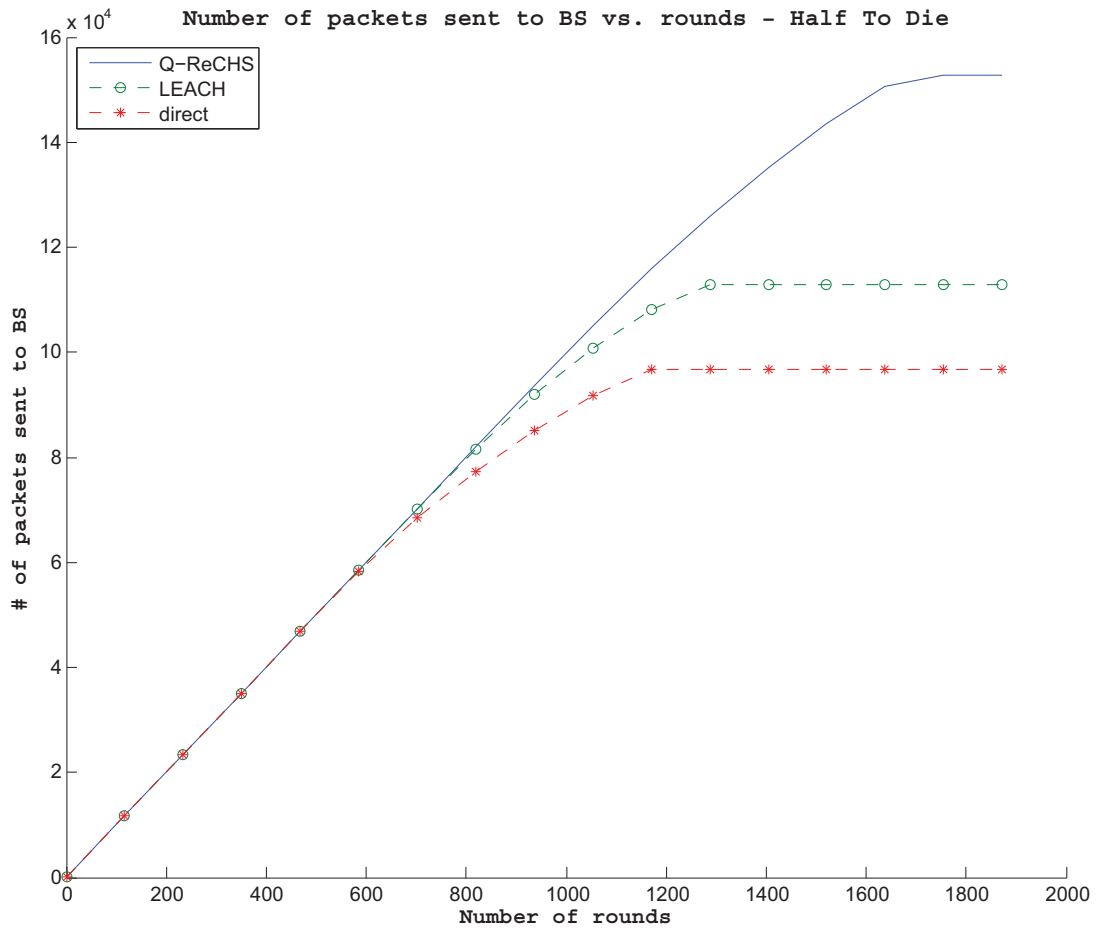


Figure 4.4 100 Nodes in 100m x 100m Environment: Half-To-Die

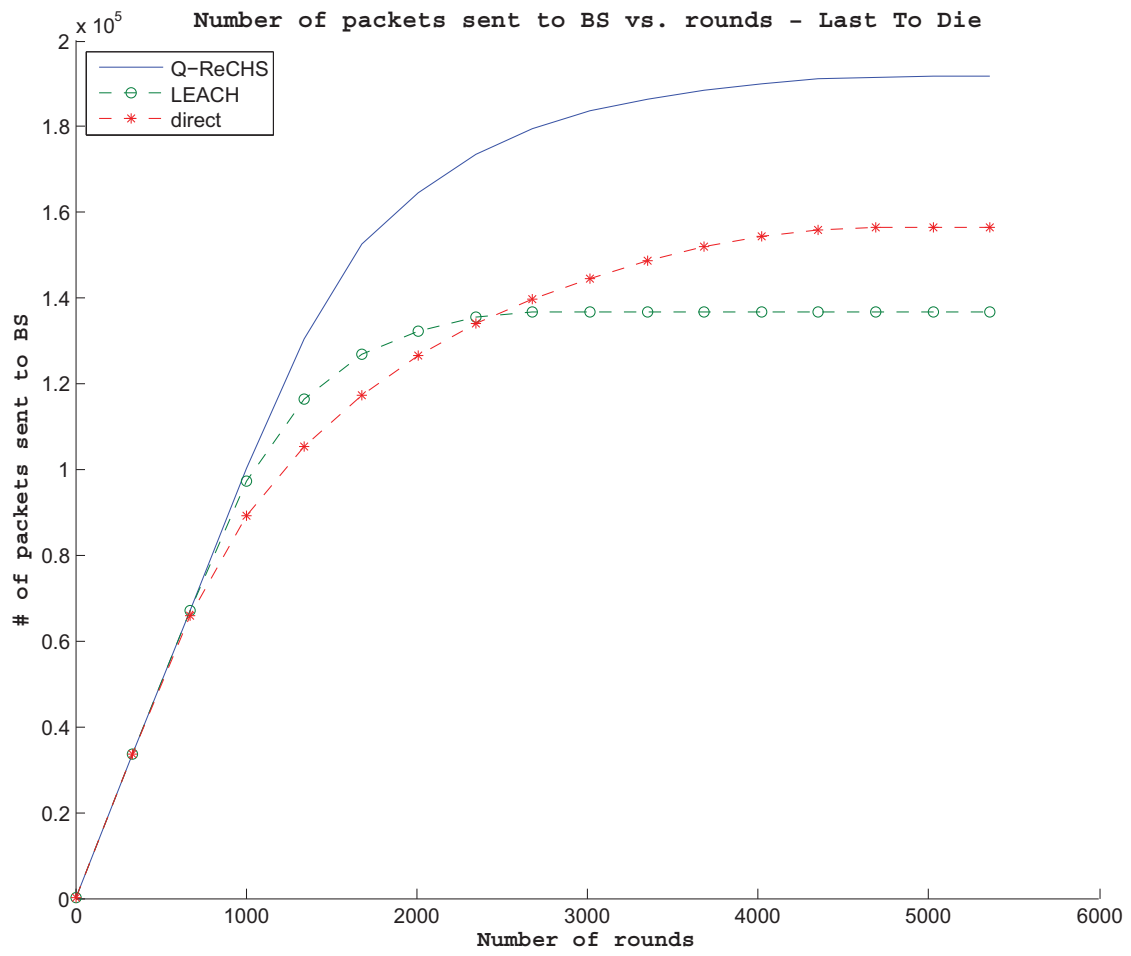


Figure 4.5 100 Nodes in 100m x 100m Environment: Last-To-Die

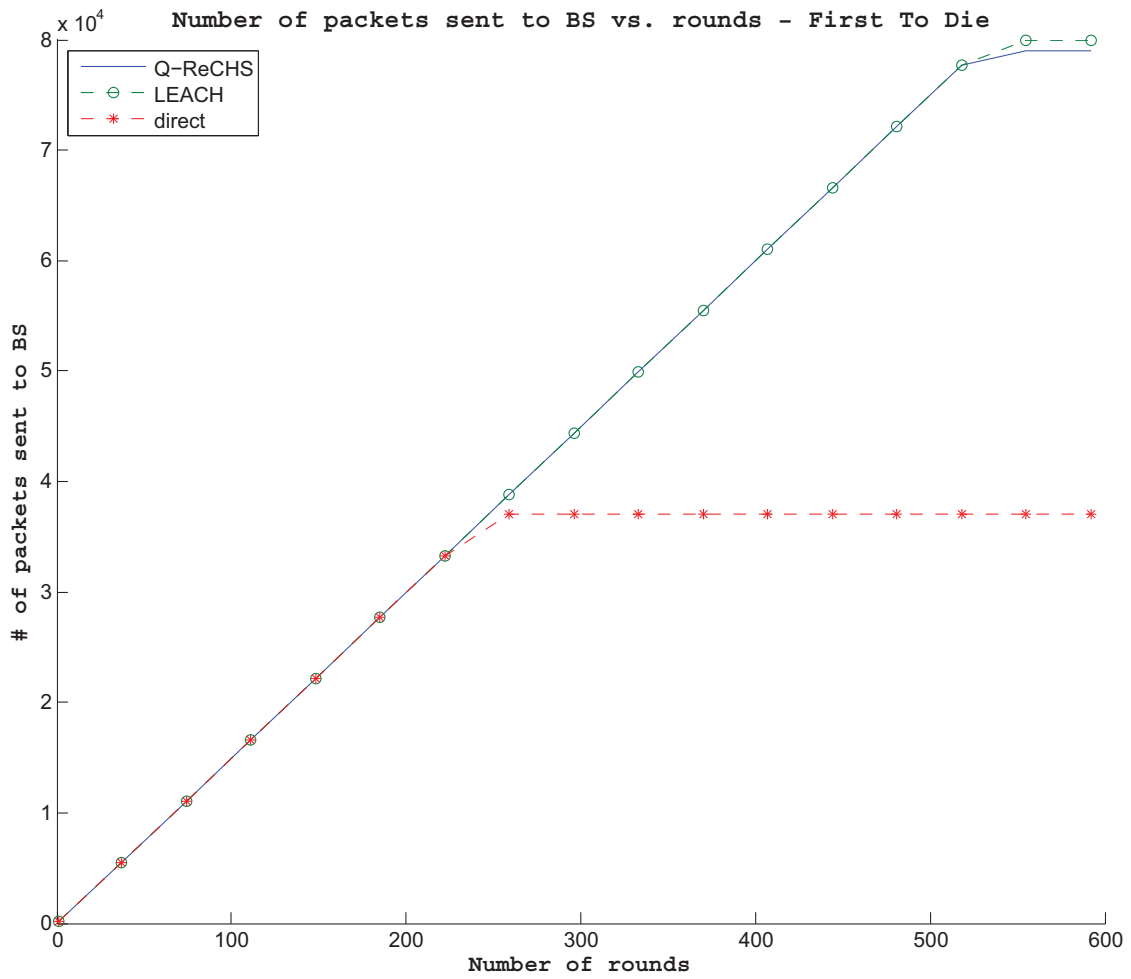


Figure 4.6 150 Nodes in 150m x 150m Environment: First-To-Die

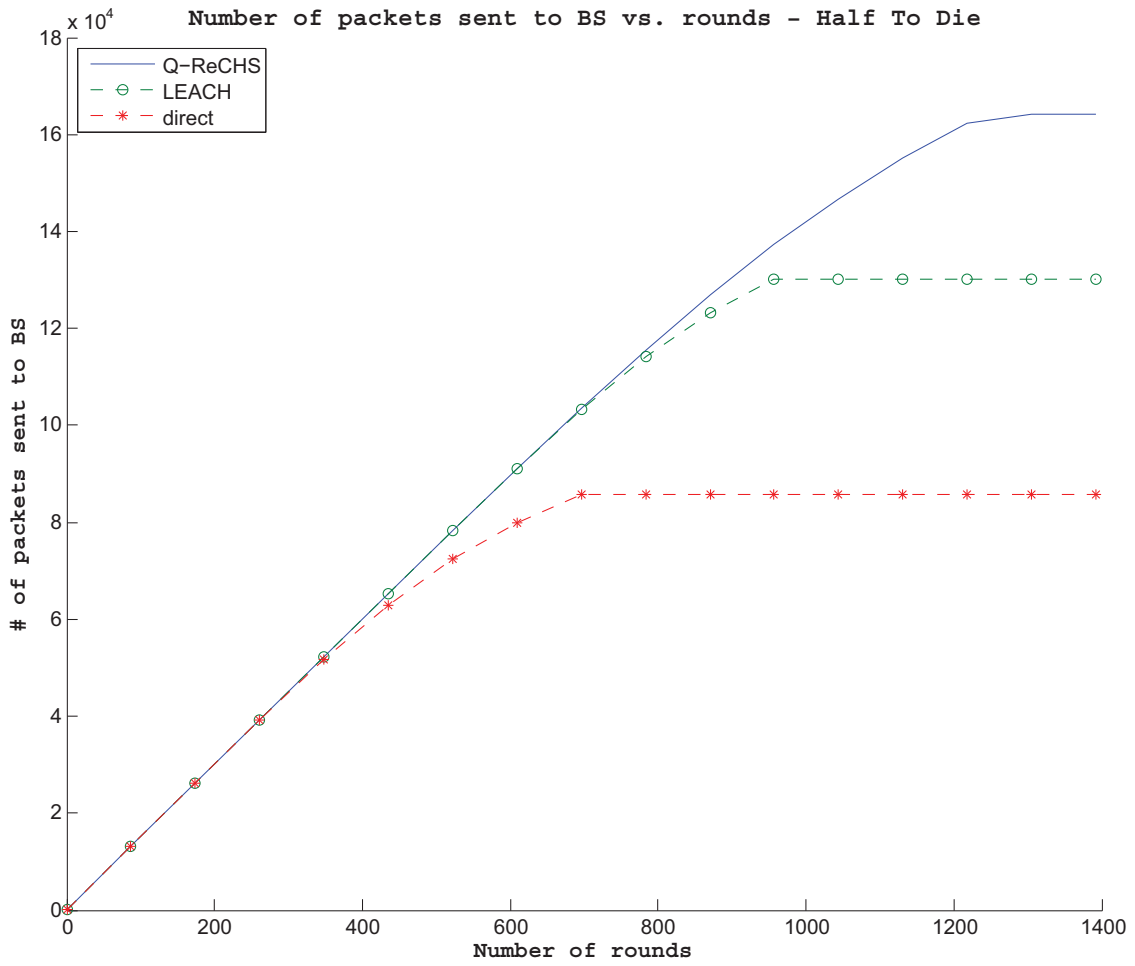


Figure 4.7 150 Nodes in 150m x 150m Environment: Half-To-Die

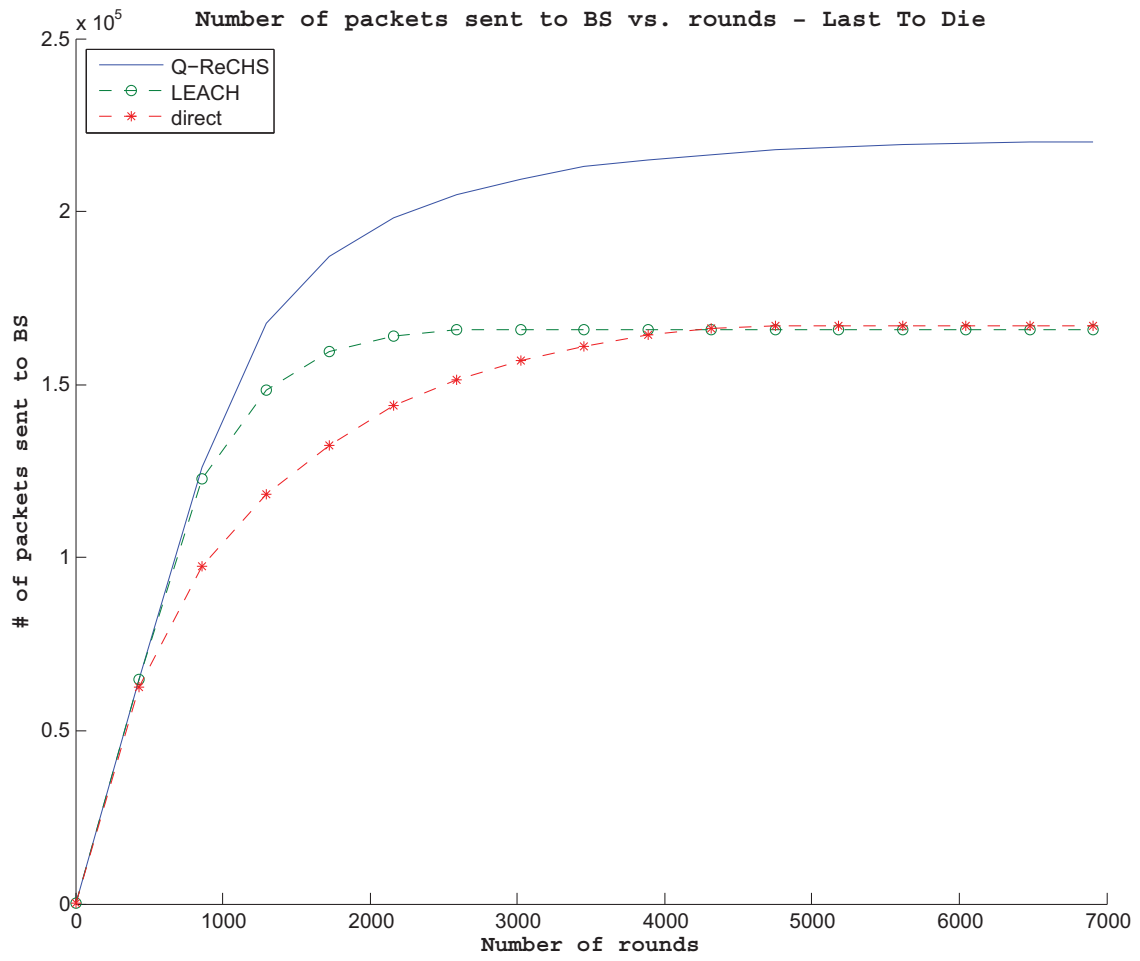


Figure 4.8 150 Nodes in 150m x 150m Environment: Last-To-Die



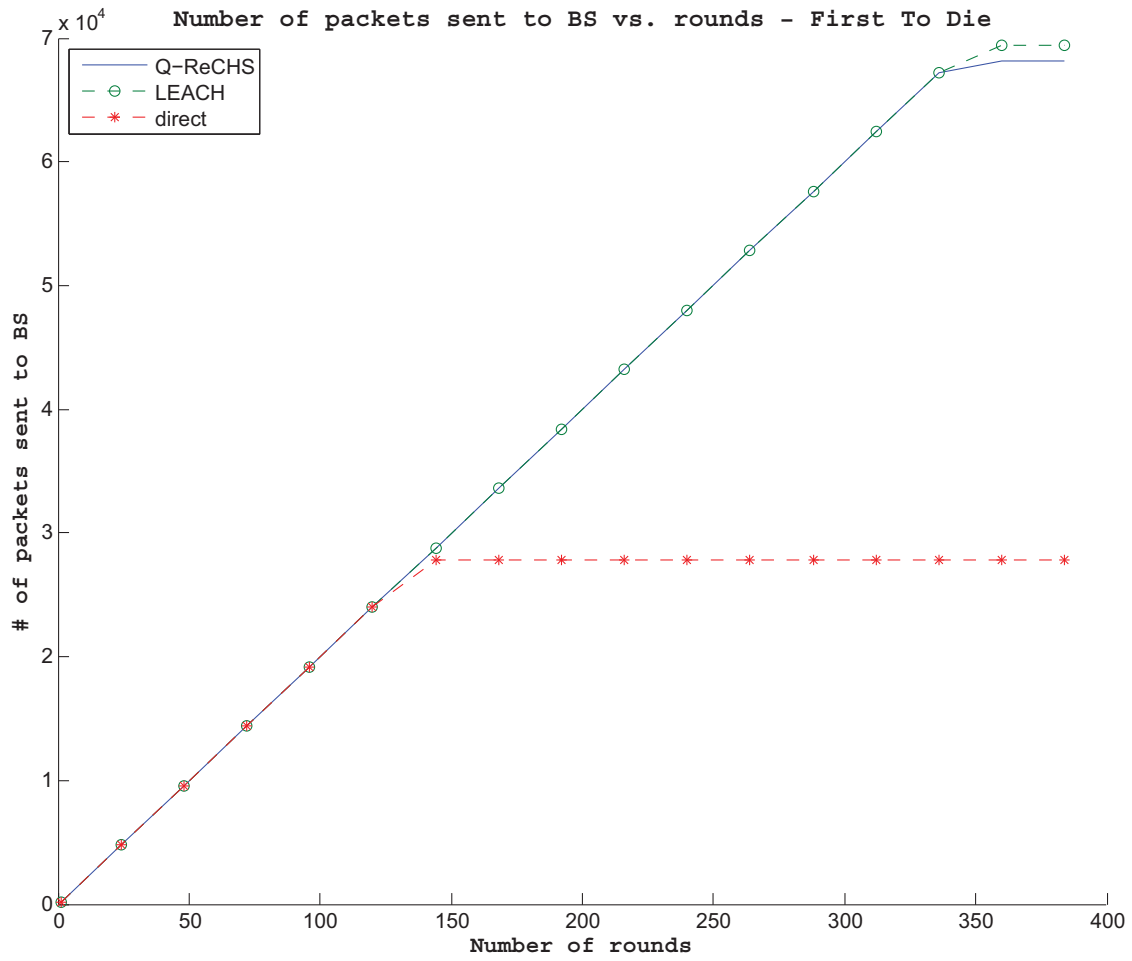


Figure 4.9 200 Nodes in 200m x 200m Environment: First-To-Die

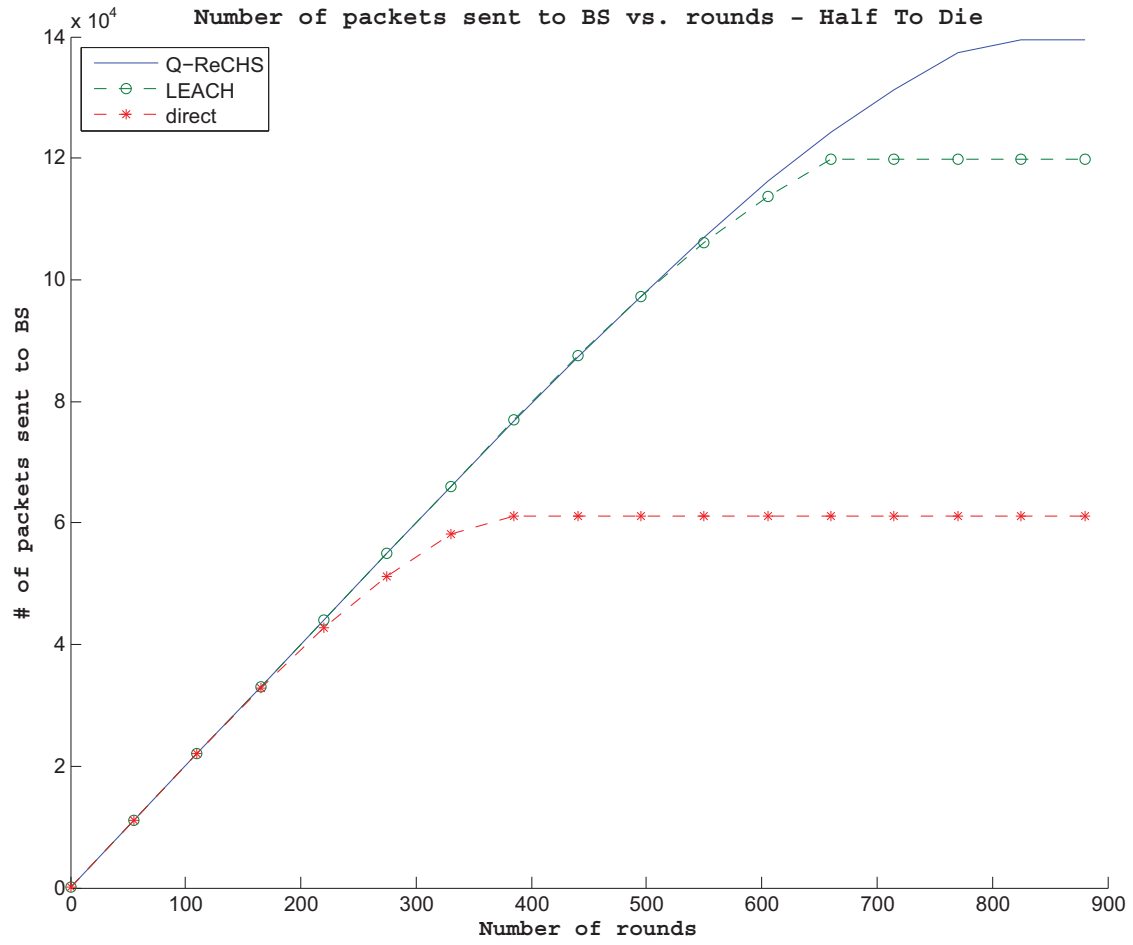


Figure 4.10 200 Nodes in 200m x 200m Environment: Half-To-Die

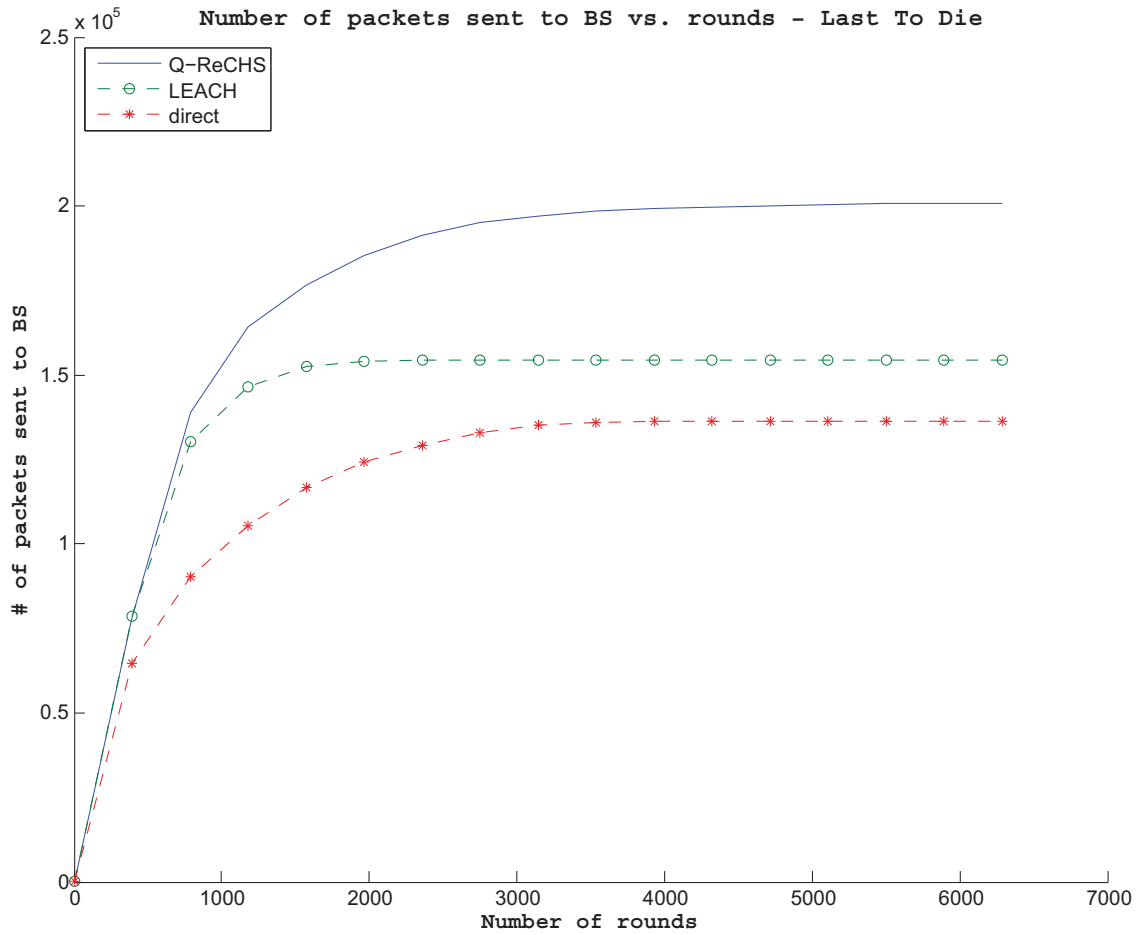


Figure 4.11 200 Nodes in 200m x 200m Environment: Last-To-Die

From the figures, we can see that our Q-ReCHS out performs the LEACH protocol in all threshold simulations. LEACH protocol does not assign CHs based of remaining energy and does not take into account the location of the CHs. While Q-ReCHS selections cluster heads based on region location in the network and the remaining energy to provide much even distribution of energy consumptions among the network nodes.

## CHAPTER 5

### CONCLUSION AND FUTURE RESEARCH

Location privacy is vital to the successful deployment of wireless sensor networks. In this dissertation, we proposed unique energy-efficient schemes to provide location privacy and cluster-based energy-aware routing. The introductory first chapter provides insight on wireless sensor networks and the importance to having an adequate routing-based location privacy technique.

#### 5.1 Conclusion for Source-Location Privacy

In **chapter 2**, we proposed source-location privacy schemes. We provide limitations with existing schemes and proposed 3 unique SLP schemes. The first one is implemented by routing through a single randomly selected intermediate node which provide local level SLP security. The second one achieves the location privacy by routing in a network-level mixing ring which provided global network SLP security. The third scheme achieve network-level source-location privacy through a technique we call the Sink Toroidal Region (STaR) routing which provided both, global and local level SLP security. These schemes were explored with theoretical analysis and simulation results. Our simulation results demonstrate that our proposed SLP schemes can achieve excellent performance in energy consumption, message delivery ratio and delivery latency.

#### 5.2 Conclusion for Destination-Location Privacy

In **chapter 3**, we proposed several destination-location privacy schemes. We proposed the bubble routing scheme to protect the destination-location privacy by first routing the message to an bubble region. The other 3 destination-location privacy schemes using a routing technique we call R-STaR routing which is a region located around the source node. We provided theoretical security analysis

for all of the proposed DLP schemes. We also was able to show that the R-STaR routing scheme performance was an improvement compared to the RSIN routing scheme.

### **5.3 Conclusion for Cluster-Based Routing**

In **chapter 4**, we introduced a new cluster-based energy-aware routing protocol for prolonging the network lifetime by evenly distributing the energy load among all the nodes. Our scheme addressed location and selection process for determining the cluster-heads within the network. With simulation, we was able to show that our proposed Q-ReCHS scheme was able to out perform some well known existing cluster-based routing schemes.

### **5.4 Future Work**

For our future work, we will look into testing our Q-ReCHS cluster-based routing scheme with other known schemes, such as, LEACH-C, HEED, Q-LEACH, and CBER. Also, we would design a stronger Q-ReCHS scheme that determines smart cluster-head selections to minimize transmit distances and maximizing network lifetime.

For other privacy protection schemes, we are working on designing privacy-aware information characterization and disclosure protection. With data exposure, we will develop the minimum distribution leakage data disclosure for a given entropy leakage, and the minimum entropy leakage for a given distribution leakage.

## **BIBLIOGRAPHY**

## BIBLIOGRAPHY

- [1] F. Ye, H. Lou, S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," in *IEEE INFOCOM*, March 2004.
- [2] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An interleaved hop-by-hop authentication scheme for filtering false data in sensor networks," in *IEEE Symposium on Security and Privacy*, 2004.
- [3] A. Perrig, R. Canetti, J. Tygar, and D. Song, "Efficient authentication and signing of multicast streams over lossy channels," in *IEEE Symposium on Security and Privacy*, May 2000.
- [4] D. T. D. S. A. Perrig, R. Canetti, "The tesla broadcast authentication protocol," in *Crypto-Bytes*, 2002.
- [5] D. Liu and P. Ning, "Efficient distribution of key chain commitments for broadcast authentication in distributed sensor networks," tech. rep., Raleigh, NC, USA, 2002.
- [6] D. Liu and P. Ning, "Multilevel  $\mu$ tesla: Broadcast authentication for distributed sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 3, no. 4, pp. 800–836, 2004.
- [7] J. Drissi and Q. Gu, "Localized broadcast authentication in large sensor networks," pp. 25–25, july 2006.
- [8] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J. Tygar, "SPINS: Security protocols for sensor networks," in *Seventh Annual International Conference on Mobile Computing and Networks (MobiCOM 2001)*, (Rome, Italy), July 2001.
- [9] C. Blundo, A. De Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, "Perfectly-secure key distribution for dynamic conferences," in *Advances in Cryptology - Crypto'92* (E. F. Brickell, ed.), (Berlin), pp. 471–486, Springer-Verlag, 1992. Lecture Notes in Computer Science Volume 740.
- [10] N. Subramanian, C. Yang, and W. Zhang, "Securing distributed data storage and retrieval in sensor networks," *Pervasive Mob. Comput.*, vol. 3, no. 6, pp. 659–676, 2007.
- [11] W. Zhang, M. Tran, S. Zhu, and G. Cao, "A random perturbation-based scheme for pairwise key establishment in sensor networks," in *MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, (New York, NY, USA), pp. 90–99, ACM, 2007.
- [12] W. Zhang, N. Subramanian, and G. Wang, "Lightweight and compromise-resilient message authentication in sensor networks," in *IEEE INFOCOM*, (Phoenix, AZ.), April 15-17 2008.
- [13] M. Albrecht, C. Gentry, S. Halevi, and J. Katz, "Attacking cryptographic schemes based on "perturbation polynomials"." Cryptology ePrint Archive, Report 2009/098, 2009. <http://eprint.iacr.org/>.

- [14] C. K. Wong and S. S. Lam, "Digital signatures for flows and multicasts," *IEEE/ACM Trans. Netw.*, vol. 7, no. 4, pp. 502–513, 1999.
- [15] R. Watro, D. Kong, S. fen Cuti, C. Gardiner, C. Lynn, and P. Kruus, "Tinypk: securing sensor networks with public key technology," in *In SASN q̄04: Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 59–64, ACM Press, 2004.
- [16] N. Gura, A. Patel, A. W. H. Eberle, and S. C. Shantz, "Comparing elliptic curve cryptography and rsa on 8-bit cpus," pp. 119–132, 2004.
- [17] A. S. Wander, N. Gura, H. Eberle, V. Gupta, and S. C. Shantz, "Energy analysis of public-key cryptography for wireless sensor networks," in *PERCOM '05: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications*, (Washington, DC, USA), pp. 324–328, IEEE Computer Society, 2005.
- [18] G. Gaubatz, J.-P. Kaps, E. Ozturk, and B. Sunar, "State of the art in ultra-low power public key cryptography for wireless sensor networks," in *PERCOMW '05: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications Workshops*, (Washington, DC, USA), pp. 146–150, IEEE Computer Society, 2005.
- [19] A. Liu and P. Ning. [Online] <http://discovery.csc.ncsu.edu/software/TinyECC/>, 2005.
- [20] H. Wang, S. Sheng, C. Tan, and Q. Li, "Comparing symmetric-key and public-key based security schemes in sensor networks: A case study of user access control," in *IEEE ICDCS*, (Beijing, China), pp. 11–18, 2008.
- [21] H. Chan and A. Perrig, "Security and privacy in sensor networks," *IEEE Computer Magazine*, pp. 103–105, Oct. 2003.
- [22] W. Du, J. Deng, Y. S. Han, P. K. Varshney, J. Katz, and A. Khalili, "A pairwise key predistribution scheme for wireless sensor networks," *ACM Trans. Inf. Syst. Secur.*, vol. 8, no. 2, pp. 228–258, 2005.
- [23] D. Liu and P. Ning, "Establishing pairwise keys in distributed sensor networks," in *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, (New York, NY, USA), pp. 52–61, ACM, 2003.
- [24] H. Chan and A. Perrig, "Pike: peer intermediaries for key establishment in sensor networks," *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1, pp. 524–535 vol. 1, March 2005.
- [25] R. L. Rivest, A. Shamir, and Y. Tauman, "How to leak a secret," in *Advances in Cryptology—ASIACRYPT*, Lecture Notes in Computer Science, vol 2248/2001, Springer Berlin / Heidelberg, 2001.
- [26] <http://www.panda.org/>.



- [27] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source-location privacy in sensor network routing," *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pp. 599–608, June 2005.
- [28] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, February 1981.
- [29] D. Chaum, "The dining cryptographer problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, vol. 1, no. 1, pp. 65–75, 1988.
- [30] L. von Ahn, A. Bortz, and N. Hopper, " $k$ -anonymous message transmission," in *Proceedings of CCS*, (Washington D.C., USA.), pp. 122–130, 2003.
- [31] A. Beimel and S. Dolev, "Buses for anonymous message delivery," *J. Cryptology*, vol. 16, pp. 25–39, 2003.
- [32] P. Golle and A. Juels, "Dining cryptographers revisited," in *Advances in Cryptology - Eurocrypt 2004*, LNCS 3027, pp. 456–473, 2004.
- [33] S. Goel, M. Robson, M. Polte, and E. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Tech. Rep. 2003-1890, Cornell University, Ithaca, NY, February 2003.
- [34] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous connections and onion routing," *IEEE J. on Selected Areas in Communications*, vol. 16, no. 4, pp. 482–494, 1998.
- [35] M. Reiter and A. Rubin, "Crowds: anonymity for web transaction," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66–92, 1998.
- [36] J. Deng, R. Han, and S. Mishra, "Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks," in *DSN '04: Proceedings of the 2004 International Conference on Dependable Systems and Networks*, (Washington, DC, USA), p. 637, IEEE Computer Society, 2004.
- [37] J. Deng, R. Han, and S. Mishra, "Countermeasures against traffic analysis attacks in wireless sensor networks," *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, pp. 113–126, Sept. 2005.
- [38] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in *WiSec '08: Proceedings of the first ACM conference on Wireless network security*, (New York, NY, USA), pp. 77–88, ACM, 2008.
- [39] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards statistically strong source anonymity for sensor networks," *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 51–55, April 2008.

- [40] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, (New York, NY, USA), pp. 88–93, ACM, 2004.
- [41] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks.," in *IPDPS*, IEEE, 2006.
- [42] O. Berthold, H. Federrath, and S. Köpsell, "Web MIXes: A system for anonymous and unobservable Internet access," *Lecture Notes in Computer Science*, pp. 115–129, 2001.
- [43] B. Möller, "Provably secure public-key encryption for length-preserving chaumian mixes," in *Proceedings of CT-RSA 2003*, LNCS 2612, pp. 244–262, April 2003.
- [44] C. Gülcü and G. Tsudik, "Mixing email with babel," in *Proceedings of the Symposium on Network and Distributed System Security*, (San Diego, CA), 1996.
- [45] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster protocol," July 2003. Version 2.
- [46] J. Kong and X. Hong, "Anodr: Anonymous on demand routing with untraceable routes for mobile ad-hoc networks," 2003.
- [47] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Enhancing wireless location privacy using silent period," in *IEEE Wireless Communications and Networking Conference (WCNC 2005)*, (New Orleans, NL, U.S.A), 2005.
- [48] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *SECURECOMM '05: Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks*, (Washington, DC, USA), pp. 194–205, IEEE Computer Society, 2005.
- [49] B. Zhu, Z. Wan, M. S. Kankanhalli, F. Bao, and R. H. Deng, "Anonymous secure routing in mobile ad-hoc networks," in *LCN '04: Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, (Washington, DC, USA), pp. 102–108, IEEE Computer Society, 2004.
- [50] W. Y.Zhang, W.Liu, "Anonymous communications in mobile adhoc networks," in *IEEE Infocom*, 2005.
- [51] R. A. Shaikh, H. Jameel, B. J. d'Auriol, S. Lee, Y.-J. Song, and H. Lee, "Network level privacy for wireless sensor networks," in *IAS '08: Proceedings of the 2008 The Fourth International Conference on Information Assurance and Security*, (Washington, DC, USA), pp. 261–266, IEEE Computer Society, 2008.
- [52] S. Misra and G. Xue, "Efficient anonymity schemes for clustered wireless sensor networks," *Int. J. Sen. Netw.*, vol. 1, no. 1/2, pp. 50–63, 2006.

- [53] Y. Ouyang, Z. Le, D. Liu, J. Ford, and F. Makedon, "Source location privacy against laptop-class attacks in sensor networks," in *SecureComm '08: Proceedings of the 4th international conference on Security and privacy in communication networks*, (New York, NY, USA), pp. 1–10, ACM, 2008.
- [54] J. Yao and G. Wen, "Preserving source-location privacy in energy-constrained wireless sensor networks," pp. 412–416, june 2008.
- [55] X. Li, X. Wang, N. Zheng, Z. Wan, and M. Gu, "Enhanced location privacy protection of base station in wireless sensor networks," pp. 457–464, dec. 2009.
- [56] Y. Jian, S. Chen, Z. Zhang, and L. Zhang, "Protecting receiver-location privacy in wireless sensor networks," pp. 1955–1963, may 2007.
- [57] L. Kang, "Protecting location privacy in large-scale wireless sensor networks," pp. 1–6, june 2009.
- [58] D. K.Mehta and M.Wright, "Location privacy in sensor networks against a global eavesdropper," 2007.
- [59] M. Shao, W. Hu, S. Zhu, G. Cao, S. Krishnamurth, and T. La Porta, "Cross-layer enhanced source location privacy in sensor networks," pp. 1–9, june 2009.
- [60] Y. Jian, S. Chen, Z. Zhang, and L. Zhang, "A novel scheme for protecting receiver's location privacy in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 7, pp. 3769–3779, October 2008.
- [61] M. Ye, C. Li, G. Chen, and J. Wu, "Eecs: an energy efficient clustering scheme in wireless sensor networks," *Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International*, pp. 535–540, April 2005.
- [62] W. B. Heinzelman, *Application-specific protocol architectures for wireless networks*. PhD thesis, 2000. Supervisor-Anantha P. Chandrakasan and Supervisor-Hari Balakrishnan.
- [63] J. Neander, E. Hansen, M. Nolin, and M. Bjorkman, "Asymmetric multihop communication in large sensor networks," *Wireless Pervasive Computing, 2006 1st International Symposium on*, pp. 7 pp.–, Jan. 2006.
- [64] O. Younis and S. Fahmy, "Heed: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transaction on Mobile Computing*, vol. 3, vol. 4, 2004.
- [65] Y. Zhang, W. Liu, Y. Fang, and D. Wu, "Secure localization and authentication in ultra-wideband sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, pp. 829–835, April 2006.
- [66] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "Tps: a time-based positioning scheme for outdoor wireless sensor networks," *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, pp. 2685–2696 vol.4, March 2004.

- [67] P. Traynor, R. Kumar, H. Choi, G. Cao, S. Zhu, and T. La Porta, “Efficient hybrid security mechanisms for heterogeneous sensor networks,” *Mobile Computing, IEEE Transactions on*, vol. 6, pp. 663–677, June 2007.
- [68] S. Zhu, S. Setia, and S. Jajodia, “Leap: efficient security mechanisms for large-scale distributed sensor networks,” in *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, (New York, NY, USA), pp. 62–72, ACM, 2003.
- [69] J. Hill, R. Szewczyk, S. H. A. Woo, D. Culler, and K. Pister, “System architecture directions for networked sensors,” in *Proceedings of ACM ASPLOS IX*, November 2000.
- [70] Wikipedia, “Normal distribution.” [http://en.wikipedia.org/wiki/Normal\\_distribution](http://en.wikipedia.org/wiki/Normal_distribution).
- [71] S. M. Stigler, *Statistics on the Table*. Harvard University Press. chapter 22.
- [72] L. Lightfoot, Y. Li, and J. Ren, “Star: Design and quantitative measurement of source-location privacy for wireless sensor networks,” *Wiley: Security and Communication Networks 2012*, 2012.
- [73] Y. Li and J. Ren, “Preserving source-location privacy in wireless sensor networks,” in *SECON'09: Proceedings of the 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks*, (Piscataway, NJ, USA), pp. 493–501, IEEE Press, 2009.
- [74] L. Lightfoot, Y. Li, and J. Ren, “Preserving source-location privacy in wireless sensor networks using star routing,” *IEEE GLOBECOM 2010*, Dec. 2010.
- [75] Y. Li, L. Lightfoot, and J. Ren, “Routing-based source-location privacy protection in wireless sensor networks,” *IEEE EIT 2009*, June 2009.
- [76] Y. Jian, S. Chen, Z. Zhang, and L. Zhang, “Protecting receiver-location privacy in wireless sensor networks,” *INFOCOM 2007. Twenty-six Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 1955–1963, May 2007.
- [77] Y. Li, L. Lightfoot, and J. Ren, “Routing-based source-location privacy protection in wireless sensor networks,” in *Electro/Information Technology, 2009. eit '09. IEEE International Conference on*, pp. 29–34, June 2009.
- [78] L. Lightfoot and J. Ren, “Providing destination-location privacy in wireless sensor using bubble routing,” *International Conference on Communications Signal Processing and Systems 2012*, 2012.
- [79] D. Tang, T. Li, J. Ren, and J. Wu, “Cost-aware routing (car) protocol design for wireless sensor networks,” 2013.
- [80] L. Lightfoot and J. Ren, “R-star destination-location privacy schemes in wireless sensor networks,” *IEEE International Communication Conference 2015 (ICC '15)*, 2015.

- [81] A. Mammu, A. Sharma, U. Hernandez-Jayo, and N. Sainz, "A novel cluster-based energy efficient routing in wireless sensor networks," *IEEE International Conference on Advance Information Networking and Applications '13*, 2013.
- [82] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An applicaton-specific protocol architecture for wireless microsensor networks," *in proceedings of HICSS '00*, 2000.
- [83] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *IEEE Transactions on Wireless Communications '02, Vol. 1, No. 4*, 2002.