PRESERVING SOURCE-LOCATION PRIVACY IN WIRELESS SENSOR NETWORKS

By

Leron J. Lightfoot

A THESIS

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

MASTER OF SCIENCE

Electrical Engineering

2010

ABSTRACT

PRESERVING SOURCE-LOCATION PRIVACY 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 hinder the successful deployment of wireless sensor networks is source-location privacy. The privacy of the source 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 source-location privacy. For WSNs, source-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 source-location privacy but most of them are based on public-key cryptosystems, while others are either energy inefficient or have certain security flaws. In this thesis, after analyzing the security weakness of the existing scheme, we propose three routing-based source-location privacy schemes. The first scheme routes each message to a randomly selected intermediate node 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 achieve network-level source-location privacy through a technique we call the Sink Toroidal Region (STaR) routing. For each of these routing schemes, both security analysis and simulation results show that the proposed schemes are secure and efficient.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Source-Location Privacy in Wireless Communications

Wireless sensor networks (WSNs) have been envisioned as a technology that has great potential to be widely used in both military and civilian applications. Privacy has been an extensively studied topic in wireless sensor networks. One of the major and unsettled issue in privacy of WSNs is in how to provide adequate routing-based source-location privacy. Sensor networks 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. 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 determine the source location by using RF localization techniques to trace back to the source in a hop-by-hop approach. Therefore, even if a powerful encryption algorithm is used to protect the source identity, the adversary may still be able to determine the location of the source by monitoring the traffic patterns and routing paths.

Privacy in a network consists of not only the privacy of the message content but also the privacy of the source and destination locations. The focus of this thesis is on source-location privacy. The confidentiality of the message content can be protected

by encryption but the source location 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.

In providing adequate source-location privacy, the sensor devices present major limitations. Sensors in the network are meant to be low-cost and energy efficient devices. The sensors 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. These sensors may be deployed in areas where human attending and maintaining the sensors is impractical; 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. Therefore, energy consumption along with source-location privacy are two very vital components for the successful deployment of wireless sensor networks.

In this thesis, we propose a two-phase routing schemes that addresses the source-location privacy issue by using unique routing processes.

1.2 Limitations with Existing Solutions

1.2.1 Limitations with Existing Solutions for Source-Location Privacy

In the past two decades, originated largely from Chaum's mixnet [1] and DC-net [2], a number of source-location private communication protocols have been proposed [3–15]. The mixnet family protocols use a set of "mix" servers that mix the received packets to make the communication source (including the sender and the recipient) ambiguous. The DC-net family protocols [2,5,6] 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 [9,10], 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 [16], which involves location privacy routing and fake message injection. In this thesis, we will address the source-location privacy in wireless sensor networks.

In [11, 12], source-location privacy is provided through broadcasting that mixes valid messages with dummy messages. The main idea is that each node needs to transmit messages consistently. Whenever there is no valid message, 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.

Routing based protocols can also provide source-location privacy through dynamic routing so that it is infeasible for the adversaries to trace back to the source-location through traffic monitoring and analysis. The main idea is to, first, route the message to a node away from the actual message source randomly, then forward the message to the SINK node using single path routing. 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 [13,14] 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 section-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. 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 [15]. 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 "not realistic" for large scale sensor networks.

1.2.2 Summary of Major Limitations

- Public-key based schemes are not suitable for source-location protection in WSNs due to the high computation and communication overhead.
- Broadcasting based source-location protection schemes are energy inefficient.
- Routing based source-location protection schemes leak source-location information to the adversaries.

1.3 Proposed Research Directions

1.3.1 Directions for Source-Location Protection

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

In the first routing scheme, the message source randomly selects an intermediate node 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 far away from the source node 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, we present three more routing schemes.

The second routing scheme provides source-location privacy through a three-phase routing process. In the first phase, each message is transmitted to a randomly selected intermediate node before it is routed to a ring node. This phase provides the local source-location privacy. In the second routing phase, the data packet will be mixed with other packets through 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 certain specific nodes on the mixing ring.

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

1.4 Overview of the Thesis

1.4.1 Design Goals

Our design goals for the routing schemes can be summarized as followed:

- The adversaries should not be able to get the source-location information by analyzing the traffic pattern.
- The adversaries should not be able to get the source-location information even if they are able to monitor certain area of the sensor network and compromise a few network nodes.
- Only the SINK node is able to identify the source-location through 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 [17].

1.4.2 Major Contributions

The major contributions of this thesis are the following:

1. We propose to protect the source-location privacy through a two-phase routing process.

2. We develop source-location privacy through routing to a single randomly selected intermediate node.

3. We develop source-location privacy through a network-level mixing ring.

4. We devise network implementation criteria for source node privacy protection in WSNs.

5. We provide extensive simulation results under ns-2 and Maple for the routing and authentication schemes we propose.

1.4.3 Structure

The thesis is structured as follows.

Chapter II introduces the schemes for source-location privacy protection. There are five sections in this chapter.

The first section presents the network assumptions and adversarial model for our routing-based schemes. The three schemes are provided in the next three sections respectively.

For each of these location privacy routing schemes, there are simulation results to demonstrate that it is efficient and can be used in many practical applications.

Chapter III summarizes the thesis.

CHAPTER 2

Source-Location Privacy Protection

2.1 Models and Assumptions

2.1.1 Network Assumptions

We make the following assumptions about our 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 [18–21].

- The SINK node is the destination location that data messages will be routed to. The information of the SINK node is made public. On detecting an event, a sensor node will generate and send messages to the SINK node through a multi-hop routing.

- Each message will include a unique dynamic ID where the event was generated. Only the SINK node can determine the source node location based on the dynamic ID.

- The sensor nodes are assumed to know their relative locations and the SINK 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 [22–24].

- The key management, including key generation, key distribution and key update, is beyond the scope of this thesis. However, the interested readers are referred to reference [25–27] for more information.

2.1.2 Adversarial Model

We assume that there are some adversaries in the target area. They try to locate the source node through traffic analysis and packet tracking. The adversaries have the following characteristics in this thesis:

- ***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. The adversary also has enough memory to store any information useful to him. If needed, the adversary could compromise some sensor nodes in the network. We also assume that an adversary will never miss the event, such as a panda, when they are close to each other.

- ***Passive:*** To prevent from being detected by the anti-hunting officials, the adversaries should not tamper any contents of the messages transmitted in the sensor network, or do any damage to the equipments, but only carry out some passive attacks, which only involve eavesdropping work.

- ***Traffic-monitoring:*** The adversary is able to monitor the traffic in an area which is important in his opinion, and he could get all of the messages in this area. However, we assume that the adversary is unable to monitor the entire network. If the adversary can monitor all the traffic though the network, he

can just monitor the events directly without relying on monitoring of the sensor network.

2.2 Source-Location Privacy Through Routing to a Single Intermediate Node (RSIN)

In this section, we will describe the proposed schemes on routing through a single intermediate node (RSIN).

2.2.1 Constrained RSIN Scheme

In this scheme, each message will be routed through an intermediate node, which will be selected randomly. The intermediate node is expected to be away from the source node for a minimum distance d_{min} based on the relative location of the sensor node. This design will make it difficult for the adversaries to get the location information of the source node.

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 may not even exist. However, the knowledge of relative location guarantees that the message packet will be forwarded to an intermediate node in an area 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. Then, this node will route the received message to the SINK node.

Suppose the source node is located at the relative location $\left(x_0, y_0\right)$. To transmit a data message, it first determines the minimum distance, d_{min}, that the intermediate node has to be away from the source node. We denote the distance between the source node and the randomly selected intermediate node as d_{rand}. Then we have

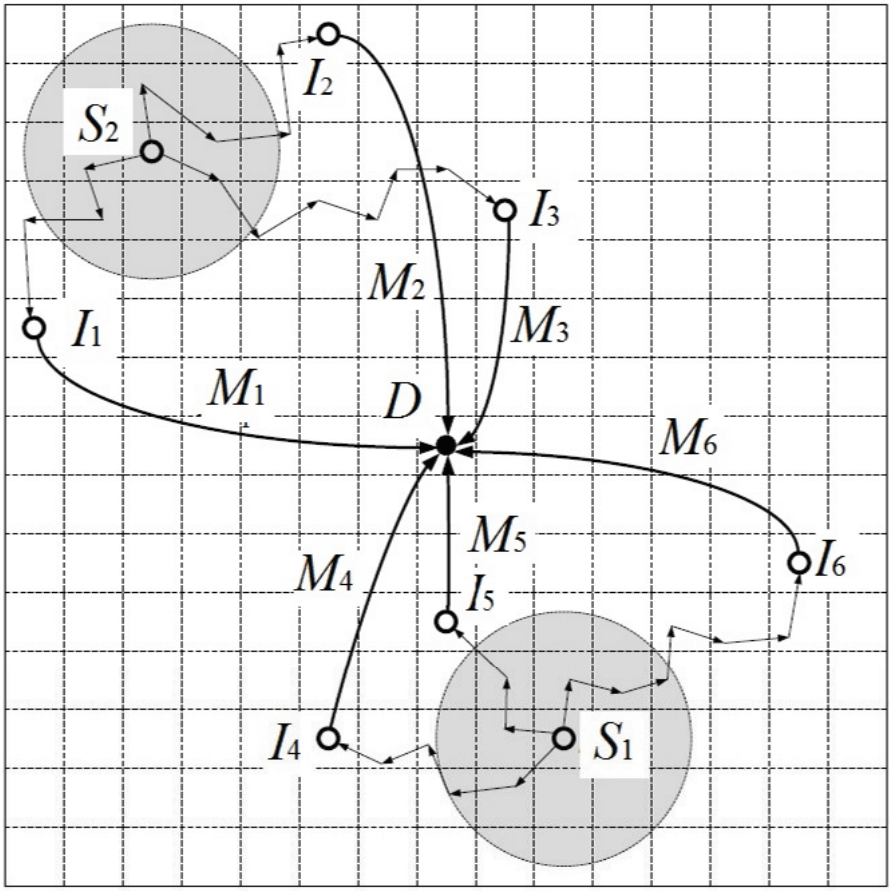


Figure 2.1. Illustration of RSIN

$d_{rand} \geq d_{min}$.

Whenever the source node wants to generate a d_{rand}, it first generates a random number x. This random variable is normally distributed with mean 0 and variance σ^2, i.e., $X \sim N(0, \sigma)$. Then the source node can calculate d_{rand} as follows:

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

Therefore, the probability [28] 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 ρ is a parameter larger than 1,φ_{0,σ^2} is the probability density function of Gaussian distribution [29].

If we choose σ 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:

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

Upon receiving a data message, the intermediate node forwards the message to the SINK node.

In the explanatory example given in Fig. 2.1, S_1, S_2 denote two source nodes in the sensor network, D represents the SINK node and $I_1, \cdots, I_6$ 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 $I_1, \cdots, I_6$ will forward the messages $M_1, \cdots, M_6$ to the SINK node, respectively.

Security Analysis

In the constrained RSIN scheme, the intermediate node is randomly selected by the source node based on relative locations of the sensor nodes. From probability point of view, every node away 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 to the best extend, 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.

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

2.2.2 Totally Random RSIN Scheme

Although the constrained RSIN scheme works well in some scenarios, its limitation is that the possibility for an intermediate node to be selected is proportional to the distance between this node and the source node. Therefore, for large scale sensor networks, the intermediate nodes tend to be 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} away from the source, as illustrated in Fig. 2.2. In this simulation example, the source node is located at $(-1250, 1250)$. We randomly selected 500 intermediate nodes according to the constrained RSIN scheme with σ equal to 1. It can be seen that all intermediate nodes are distributed around the source node. When all these node are selected to forward messages to the SINK node. The adversaries may very likely be able to locate the source node. Therefore, for large scale sensor networks, the constrained RSIN scheme may not be able to provide adequate global 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 very close to the real source node. Fortunately, the probability for this is very low for large scale sensor networks. Nevertheless, to prevent this from happening, in

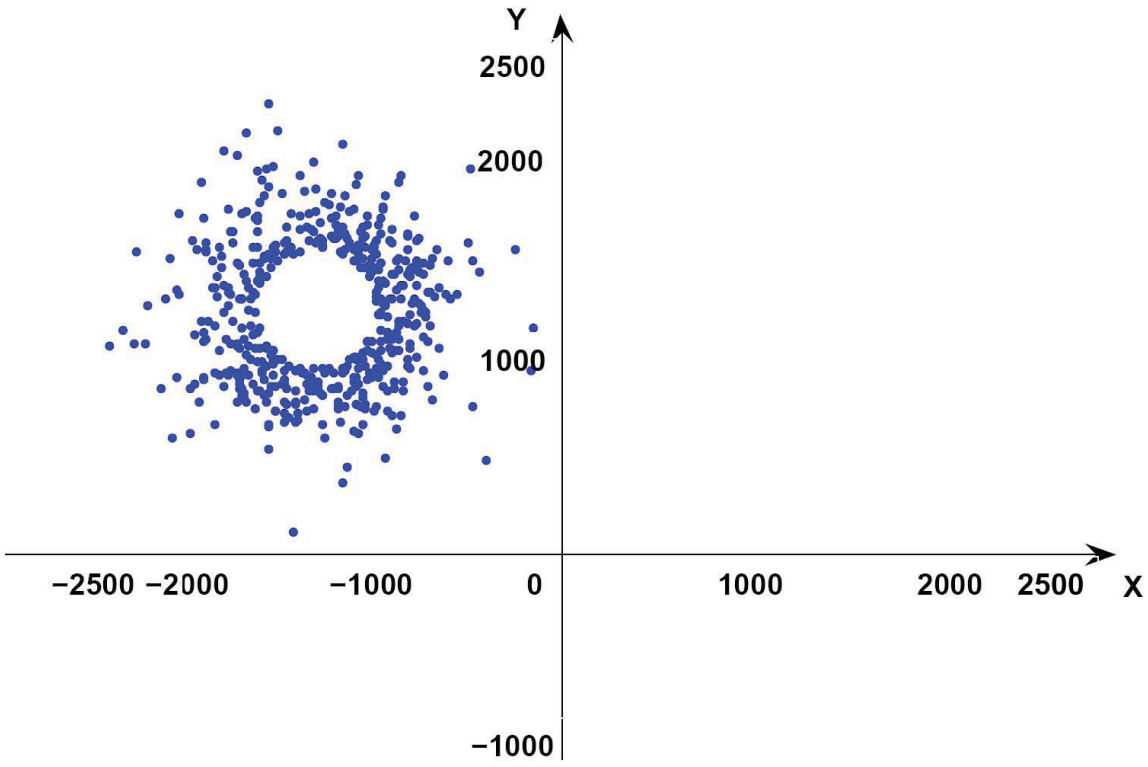


Figure 2.2. Intermediate nodes distribution for constrained RSIN scheme

"For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis."

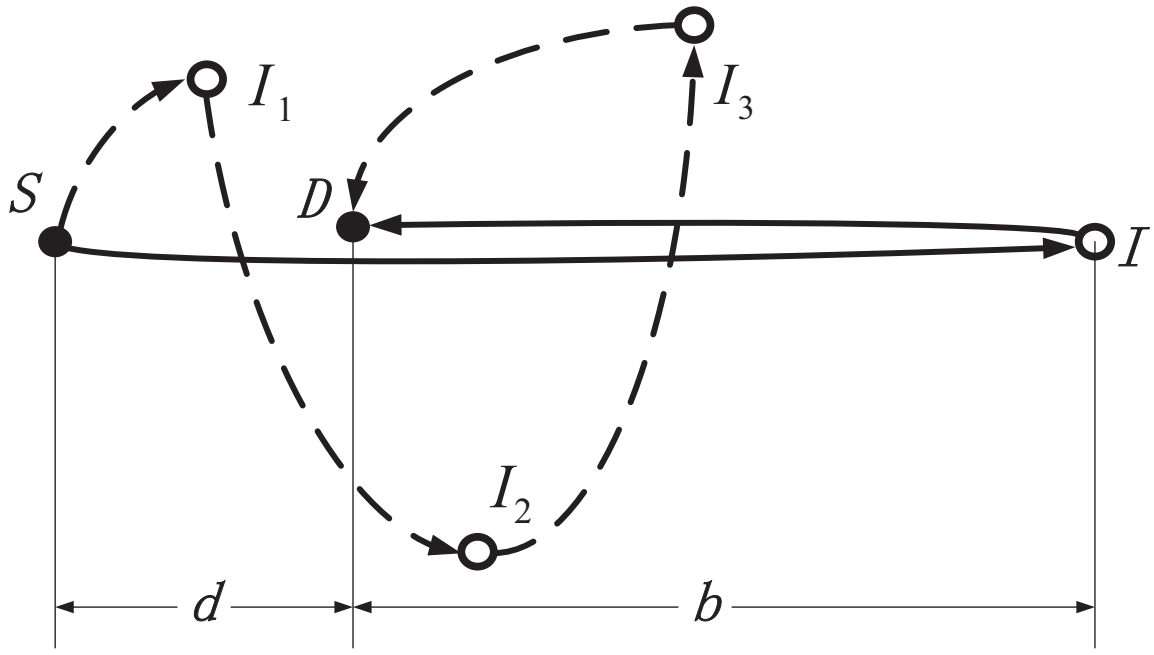


Figure 2.3. Message forwarding through intermediate node(s)

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 Fig. 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 Fig. 2.3 as an example, once a packet is captured by adversaries en-route, the adversaries may get

the direction of the source-location according to transmission direction of the captured packet.

Security Analysis

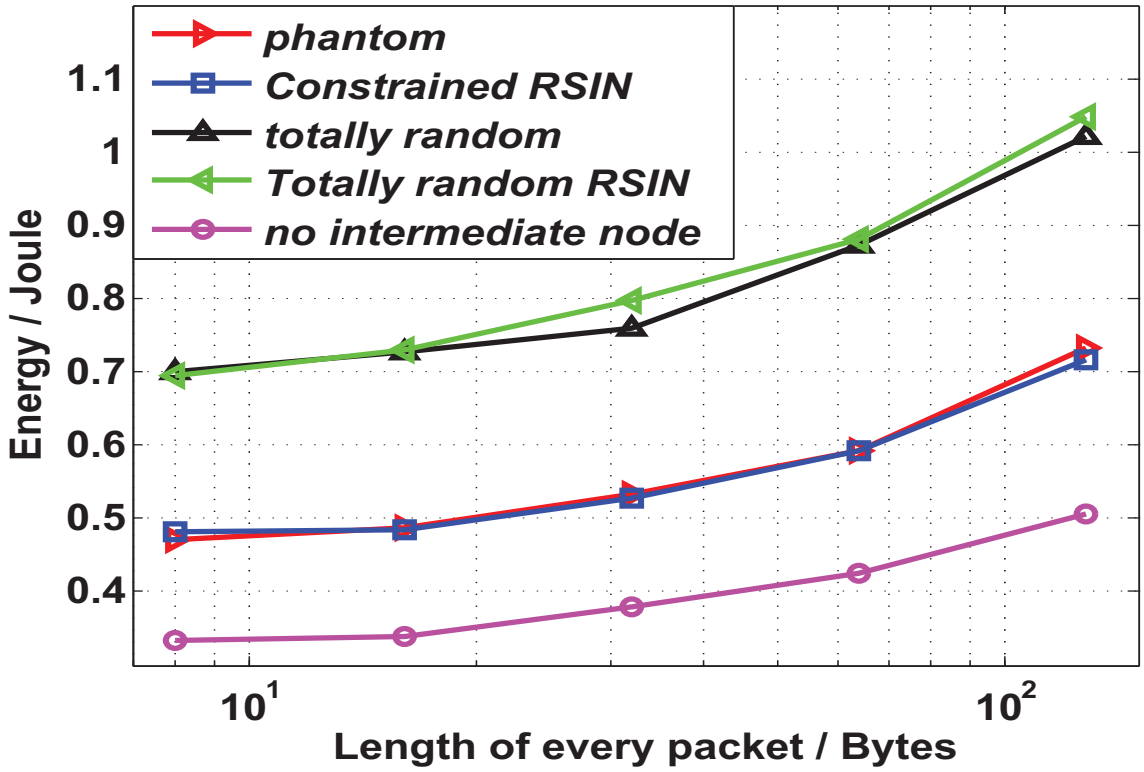
Unlike constrained RSIN scheme, in totally random RSIN scheme, the intermediate nodes randomly selected are evenly distributed in the sensor networks. Every node in the networks has the same possibility of being selected as the intermediate node. The messages can be forwarded to the SINK node from all 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.

2.2.3 Simulation Results and Performance Comparison

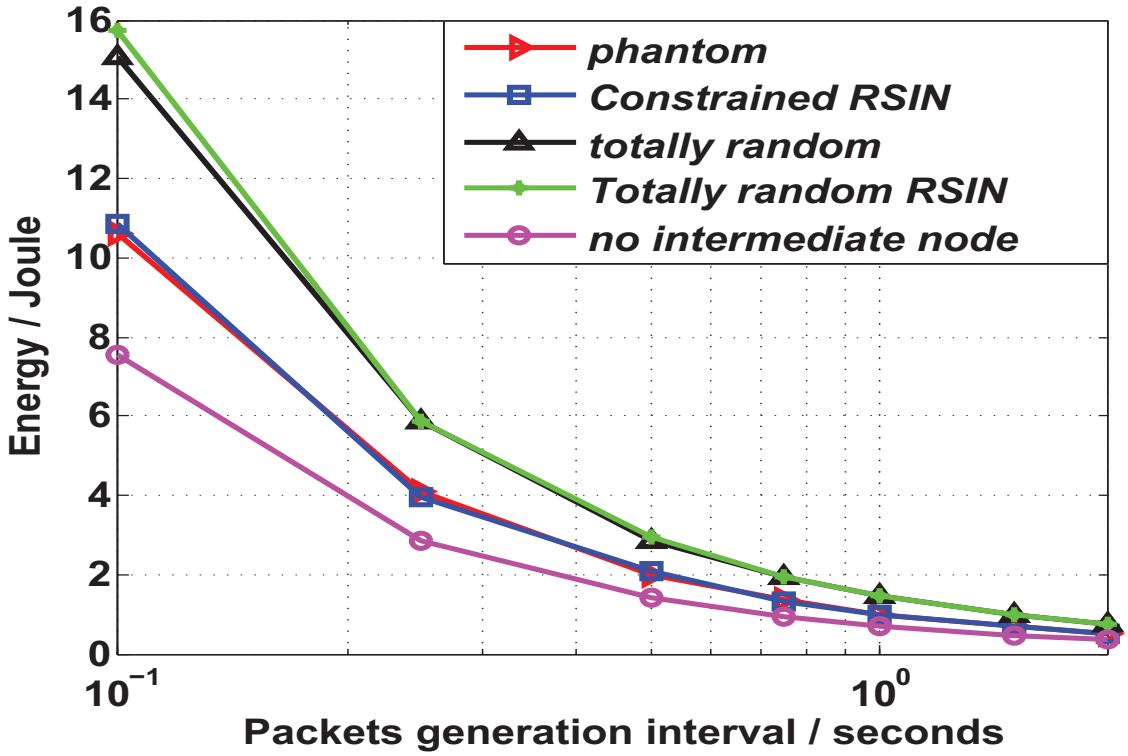
To evaluate the performance of our proposed schemes, and routing through totally randomly selected intermediate nodes without d_{min} restriction, we conduct simulations using ns-2 on RedHat Linux system. In the simulation, 400 nodes are distributed in an area of size 3360×3360 meters. The SINK node is located at the center of the networks.

Simulation results are provided in Fig. 2.4, 2.5, 2.6, 2.7, 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 simulation results from (a) to (f), we set $d_{min} = 480$ meters for totally random RSIN scheme. 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 set to be 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, $R1$, $R2$, $R3$

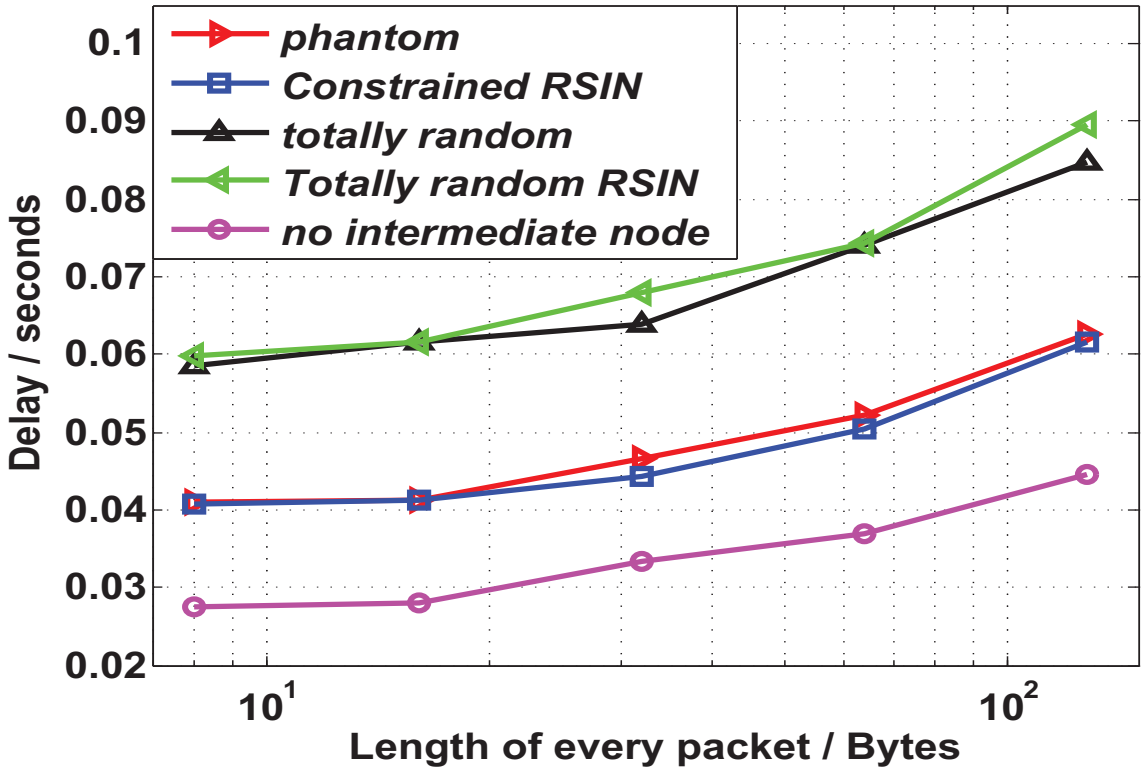


(a) Power consumption for different packet lengths

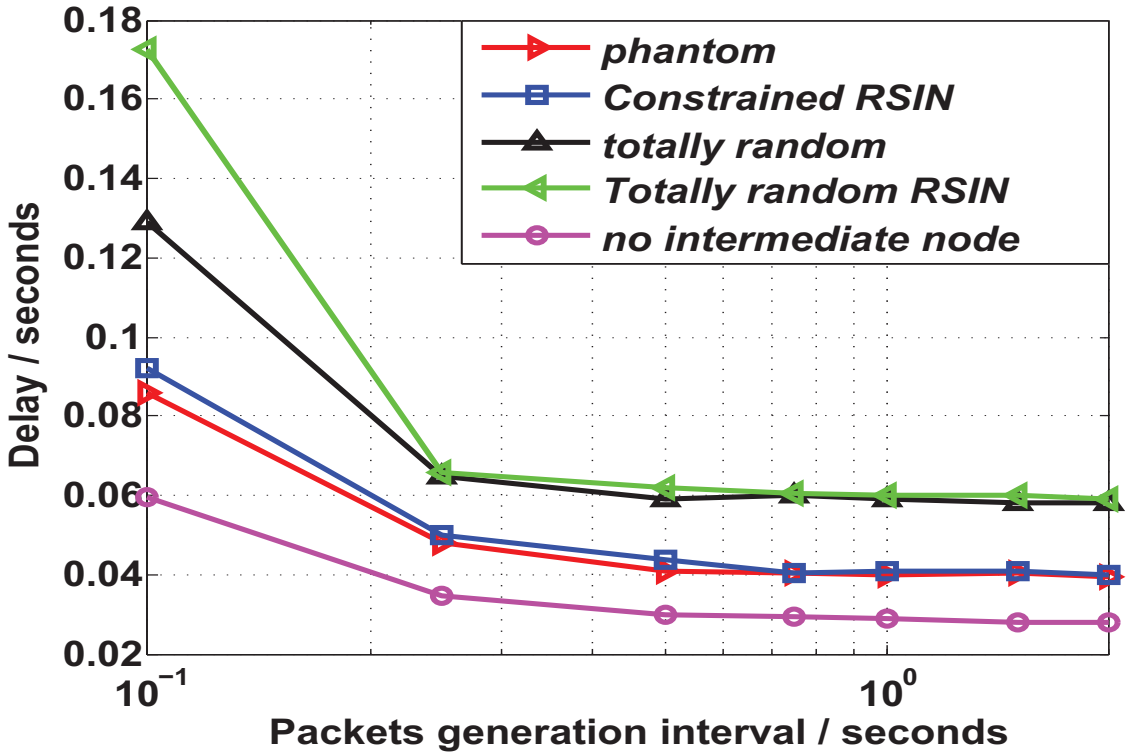


(b) Power consumption for different packet generation intervals

Figure 2.4. Performance of routing by single-intermediate node

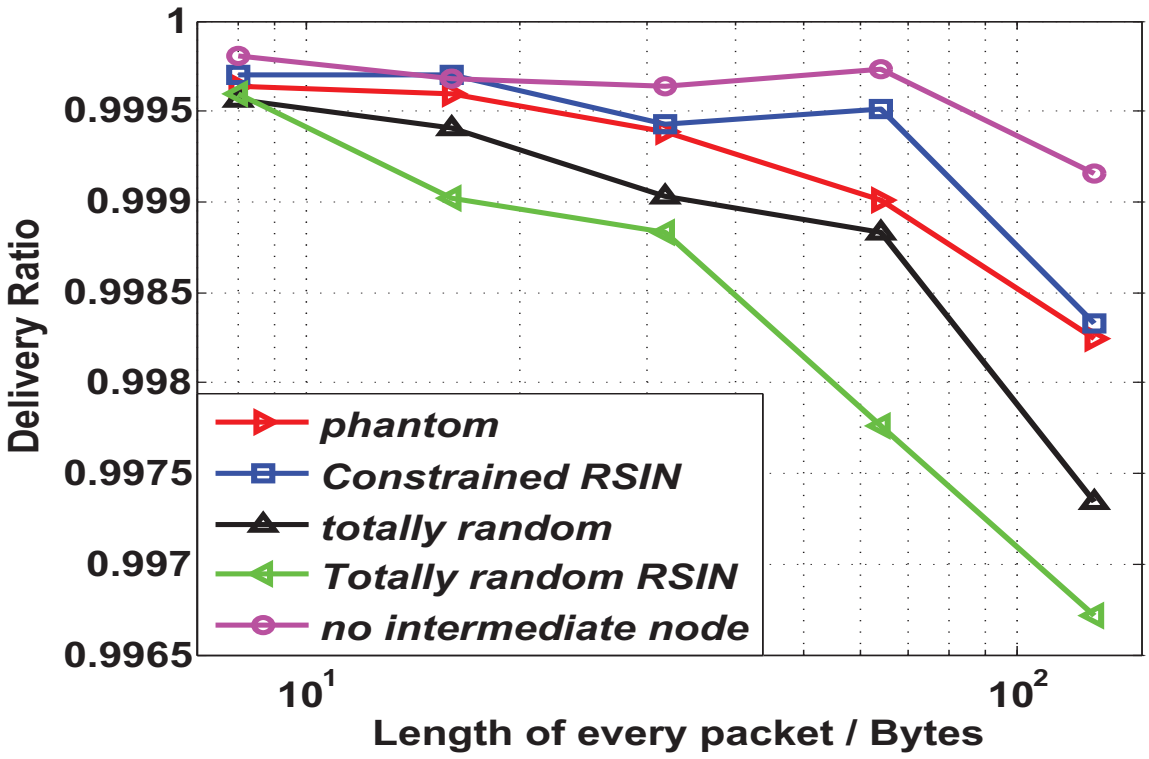


(c) Message latency for different packet lengths

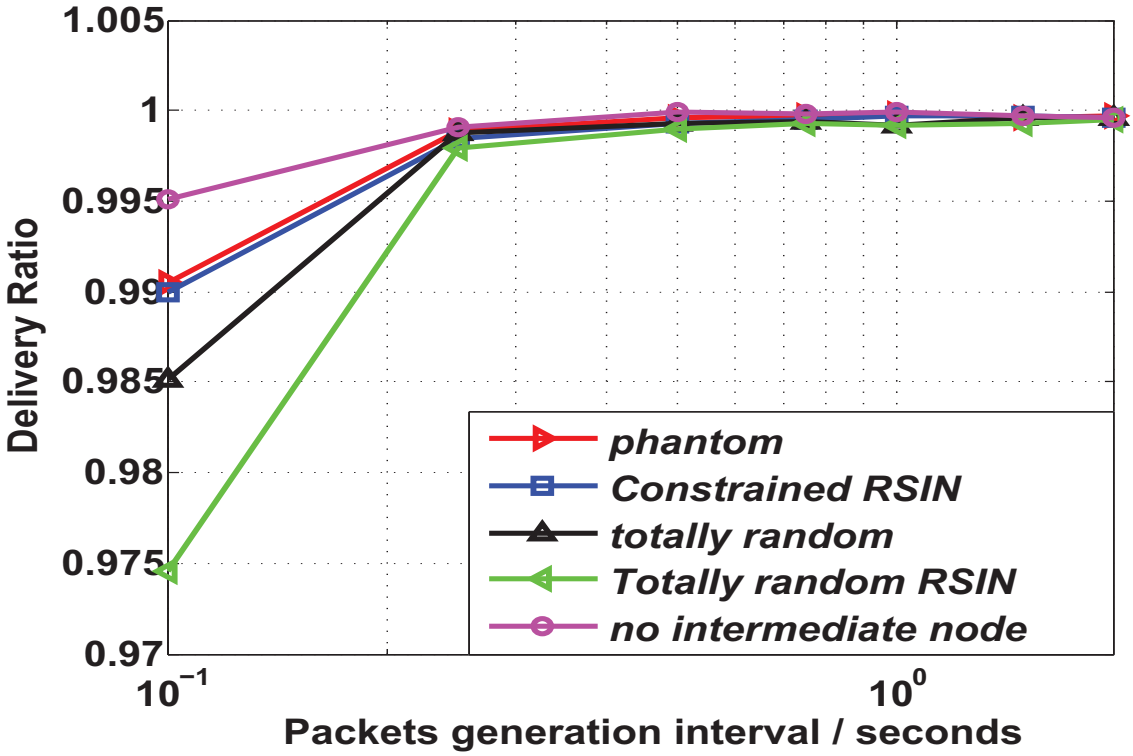


(d) Message latency for different packet generation intervals

Figure 2.5. Performance of routing by single-intermediate node

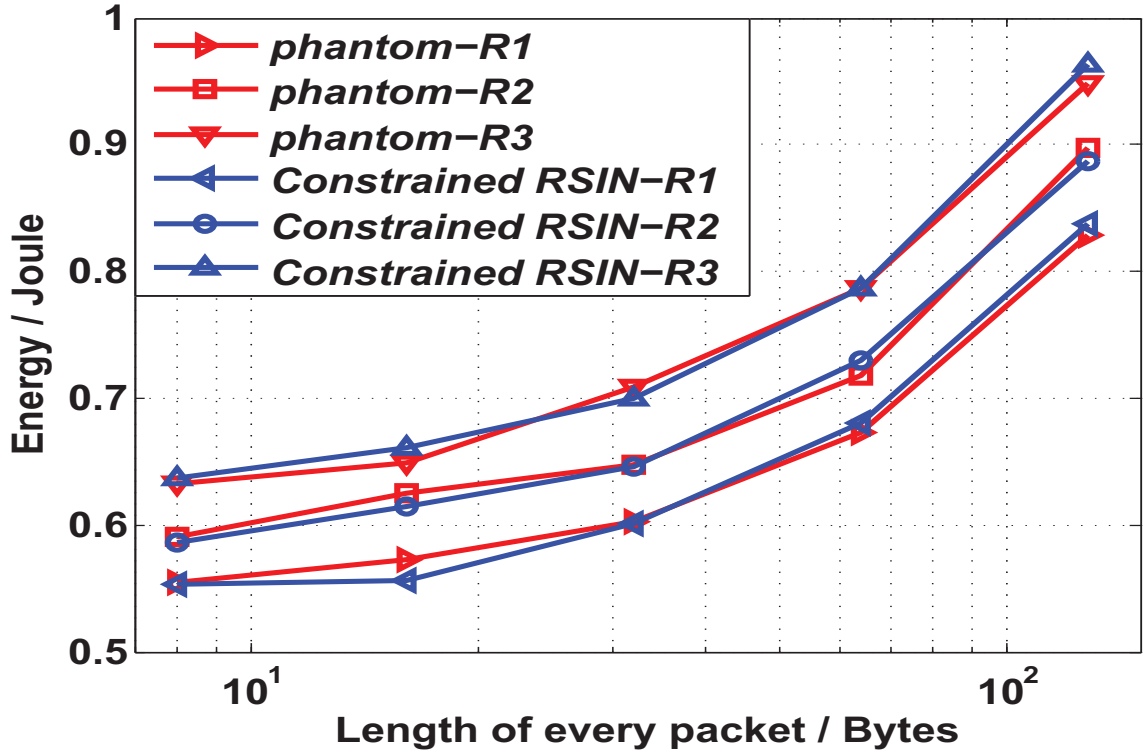


(e) Message delivery ratio for different packet lengths

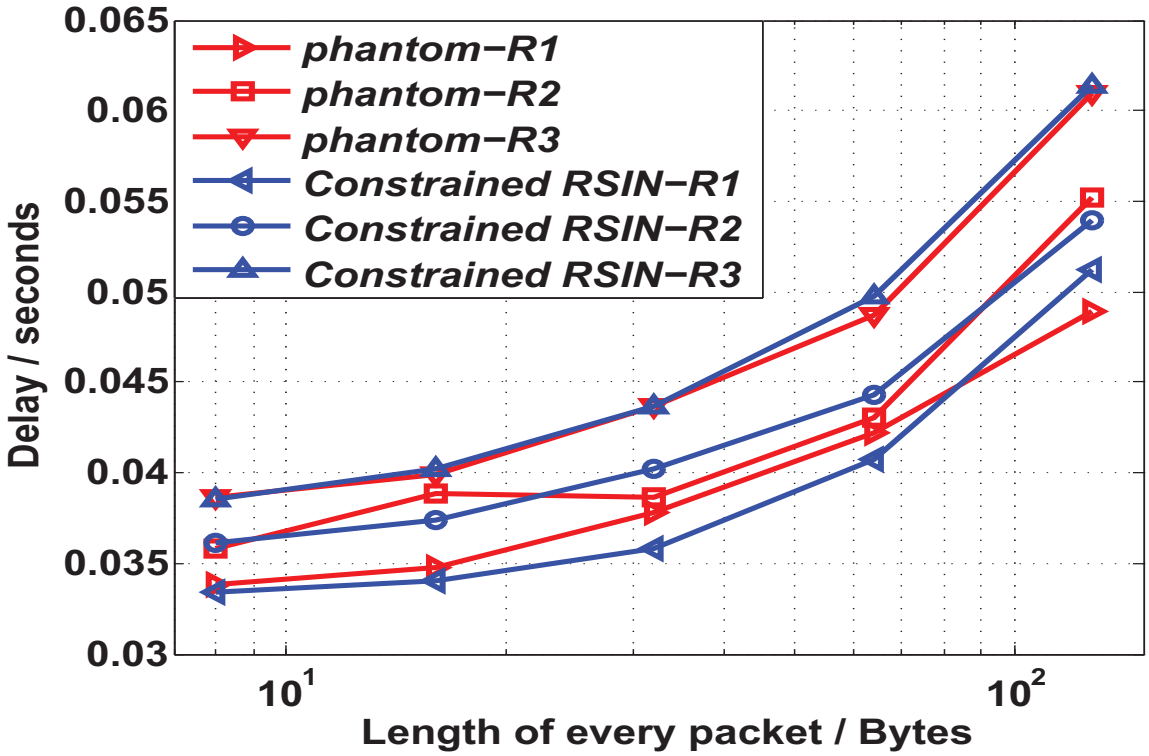


(f) Message delivery ratio for different packet generation intervals

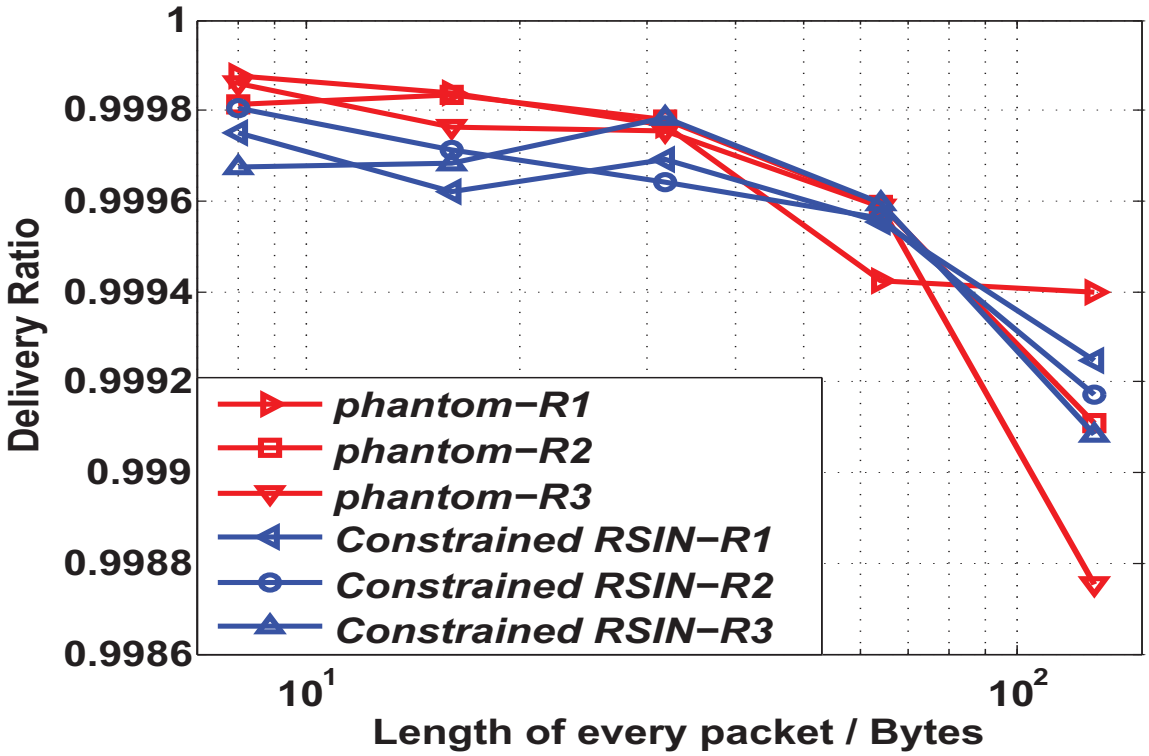
Figure 2.6. Performance of routing by single-intermediate node



(g) Power consumption for different length of random path



(h) Message latency for different length of random path



(i) Message delivery ratio for different length of random path

Figure 2.7. Performance of routing by single-intermediate node

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 4 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; (iii) Totally random RSIN with d_{min} restriction has the worst performance; (iv) Totally random intermediate selection without d_{min} constrain has performance in between these two schemes.

2.3 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*.

2.3.1 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

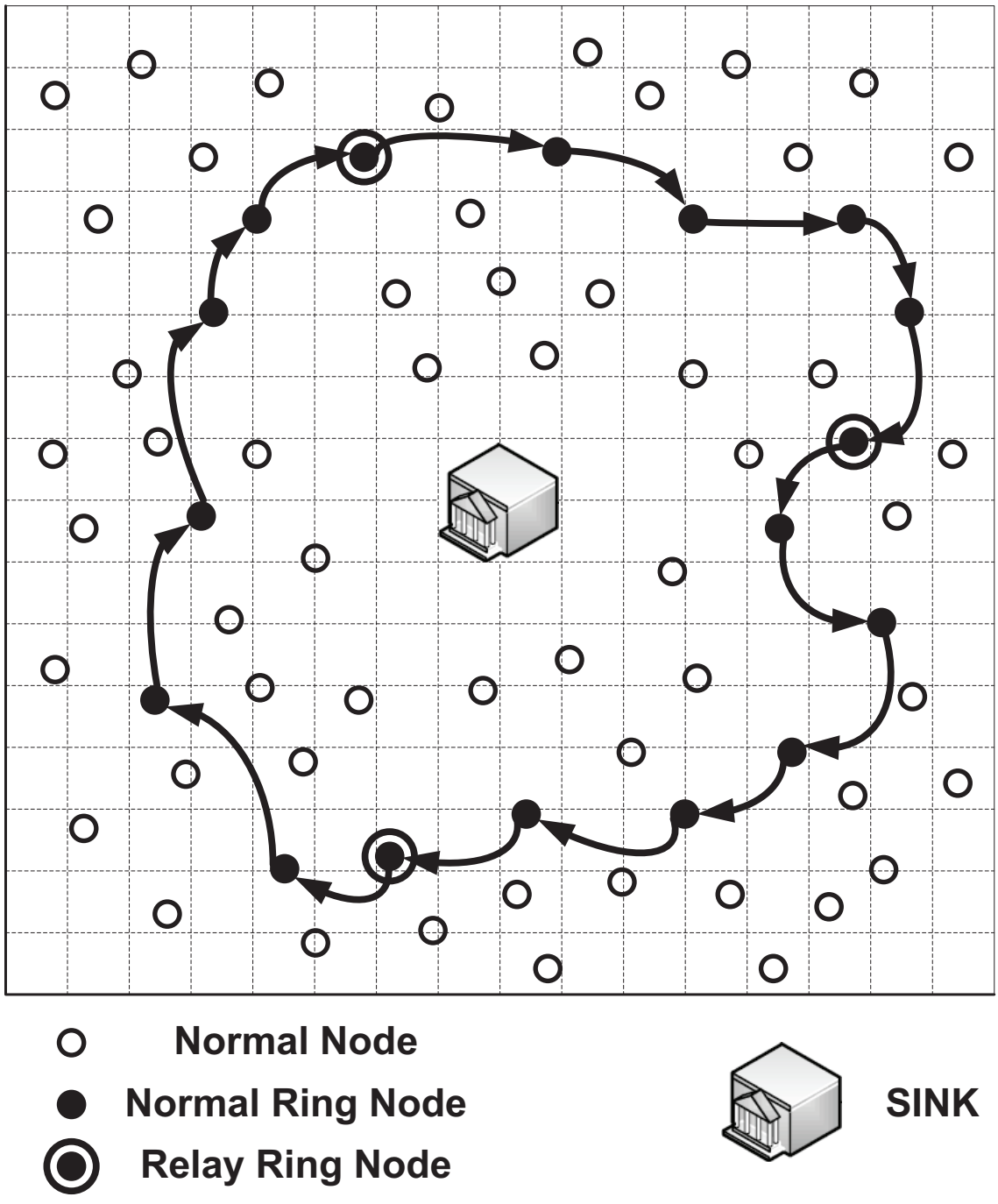


Figure 2.8. Grids Formation

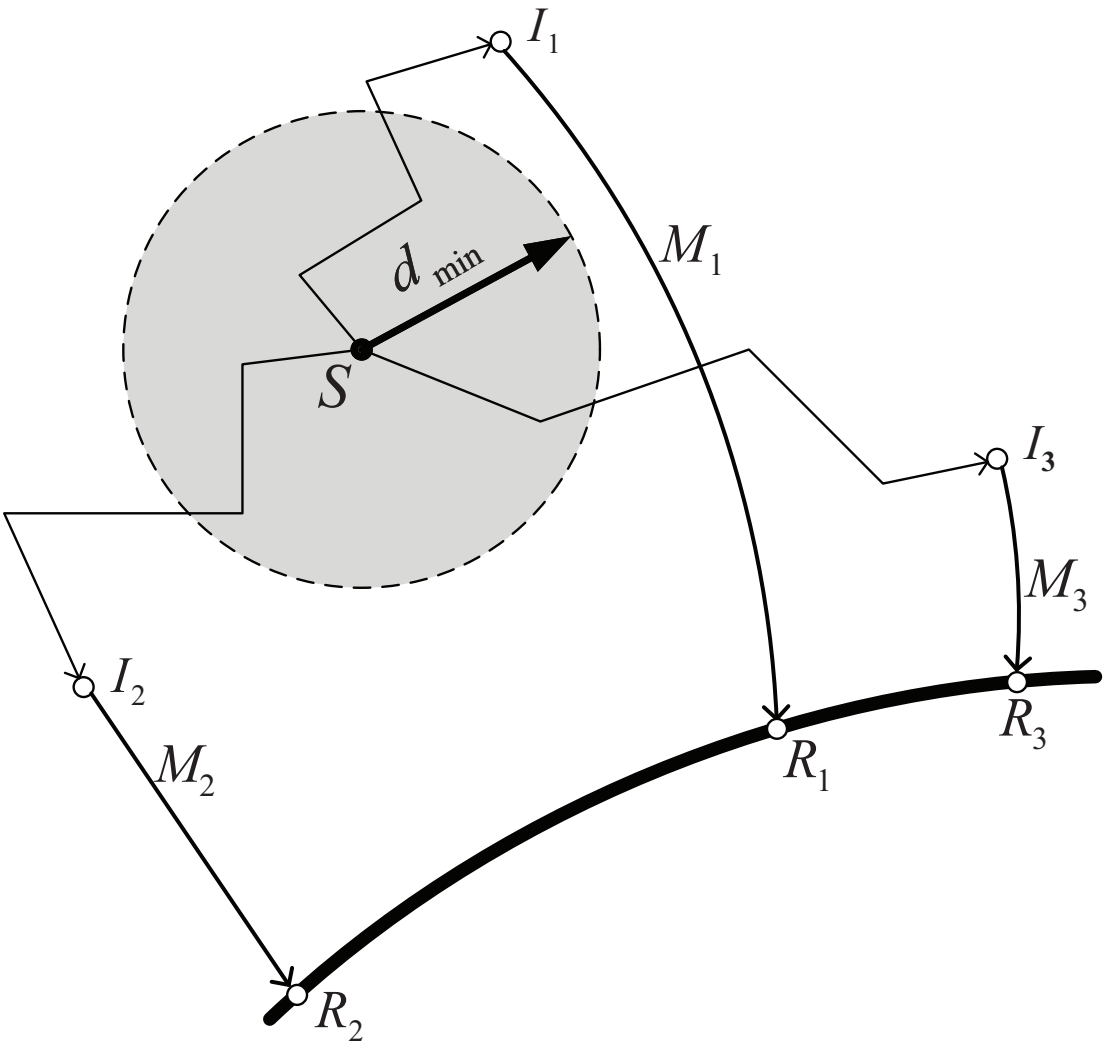


Figure 2.9. Illustrate of the first two phases routing

be forwarded to the nearest ring node by this intermediate node.

An example is given in Fig. 2.9, 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.

2.3.2 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 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 Fig. 2.10, ring node B shares a key K_{AB} with ring node A and a key K_{BC} with ring node

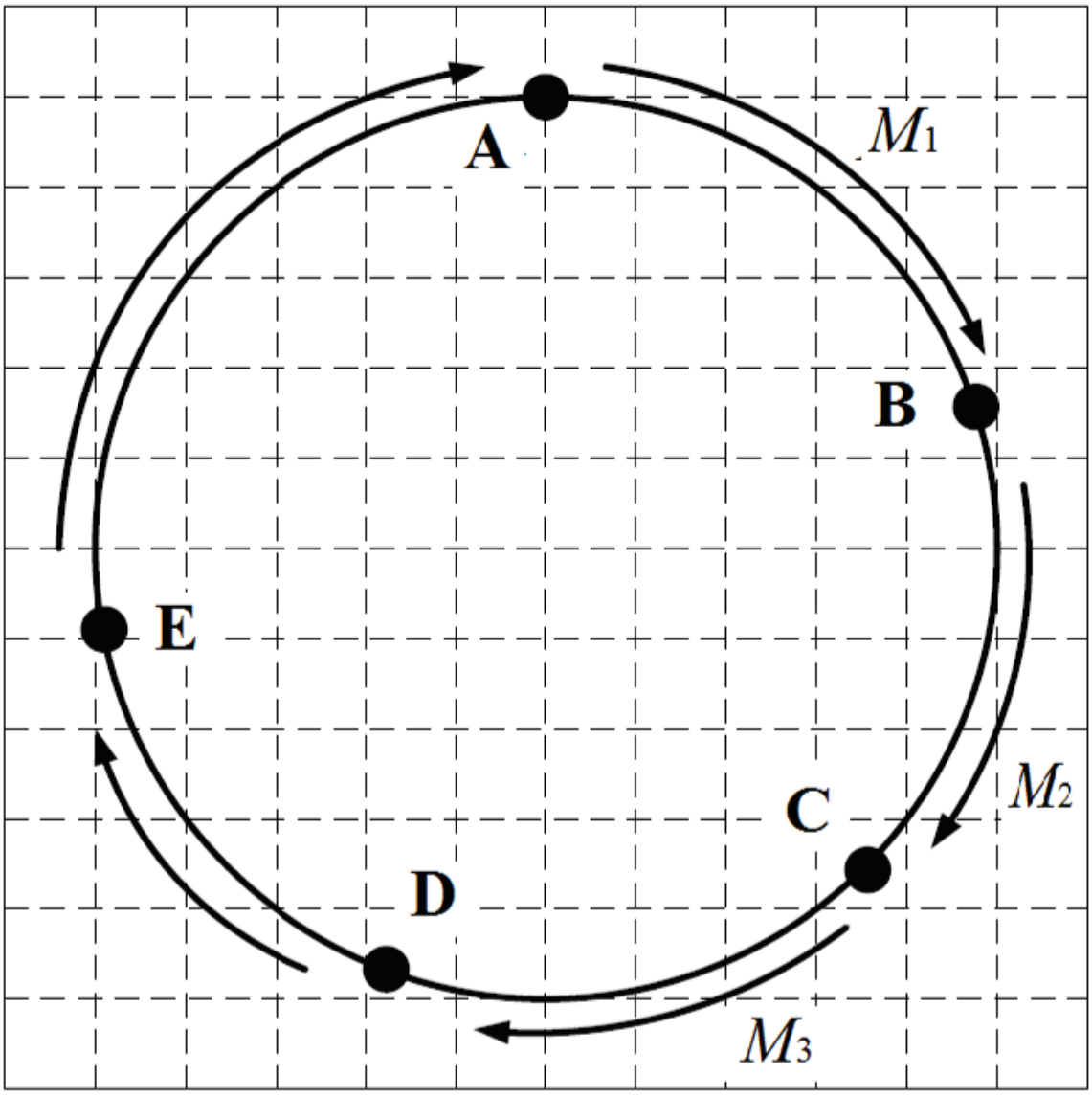


Figure 2.10. Message transmission in the ring

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

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

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.

2.3.3 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.

2.3.4 Security Analysis

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 Fig. 2.9, 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 Fig. 2.9, 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

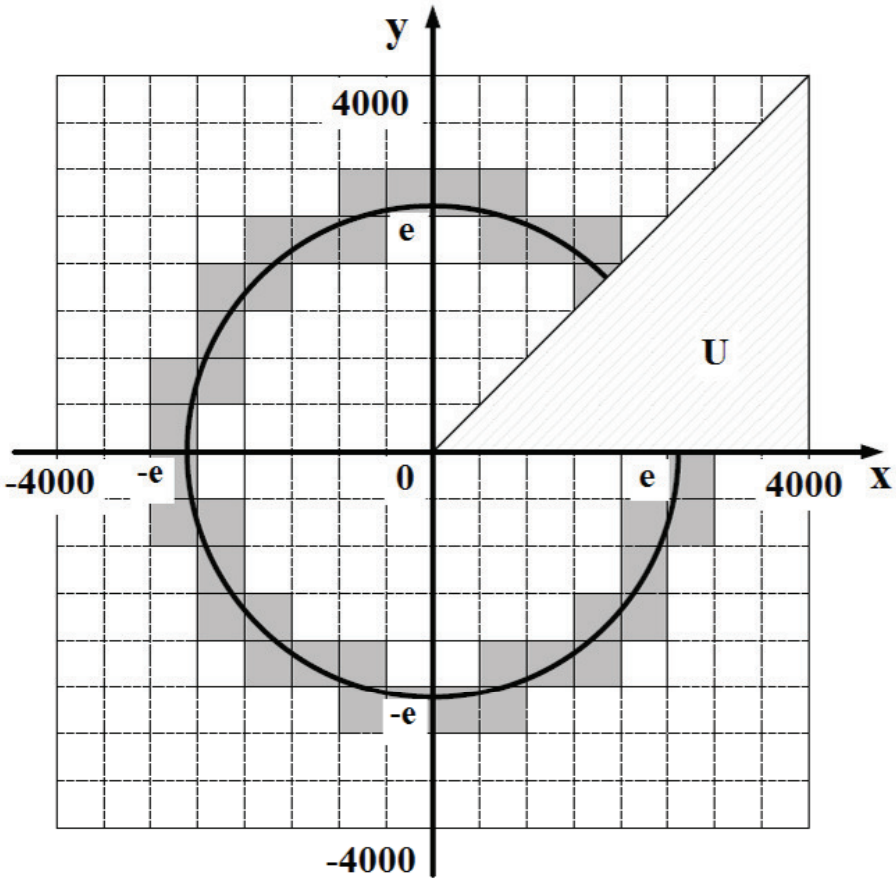


Figure 2.11. Ring selection in simulation setup

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 Fig.2.9, 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.

2.3.5 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.

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

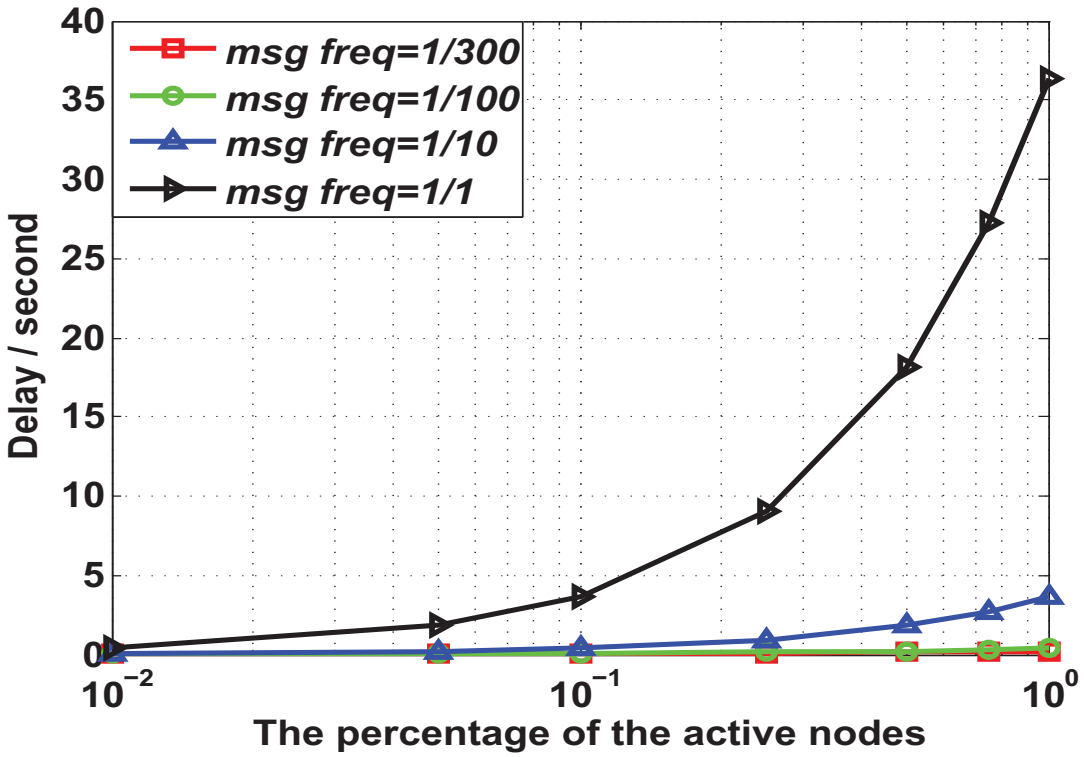
where $\mathcal{E}$ denotes the energy consumption, α is a constant parameter and d is the distance of the transmission. Fig. 2.11 gives an example of a target area of size 8000×8000 meters. The shaded grids are selected as the ring grids. The line in the middle of the shaded area is indicated by the solid line. If the density of the sensor nodes in the sensor network is λ, then the total energy consumption for each sensor in this area to transmit one message to a ring node can be calculated as follows:

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

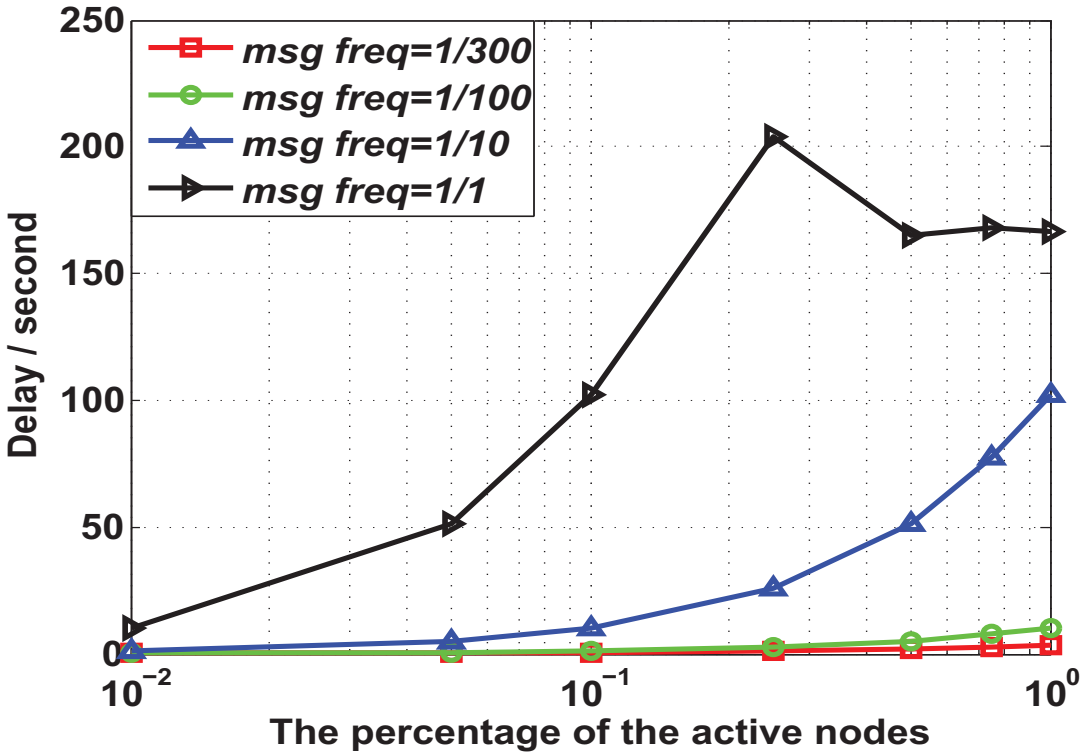
where $\mathcal{E}_U$ is the energy consumption for area U as demonstrated in Fig. 2.11. It can be calculated that when $e = 3061$, the overall power consumption $\mathcal{E}_{total}$ achieves the minimum. In this way, we get the optimal ring location.

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: τ, 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

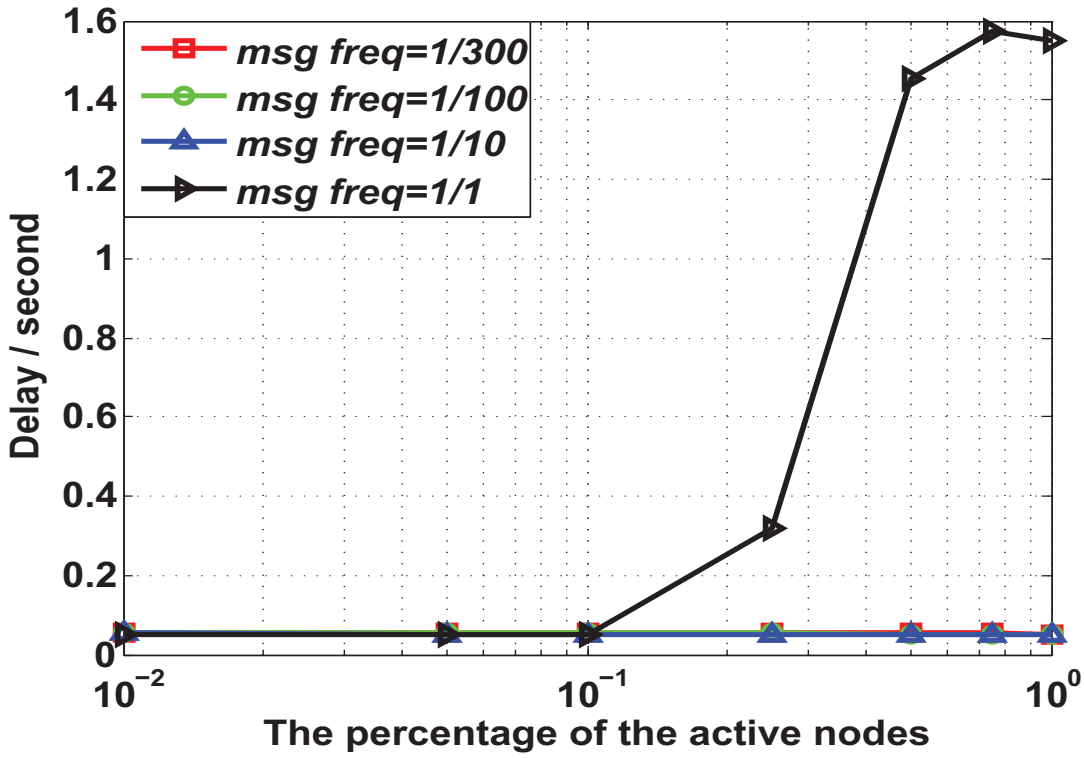


(a) Power consumption of normal nodes

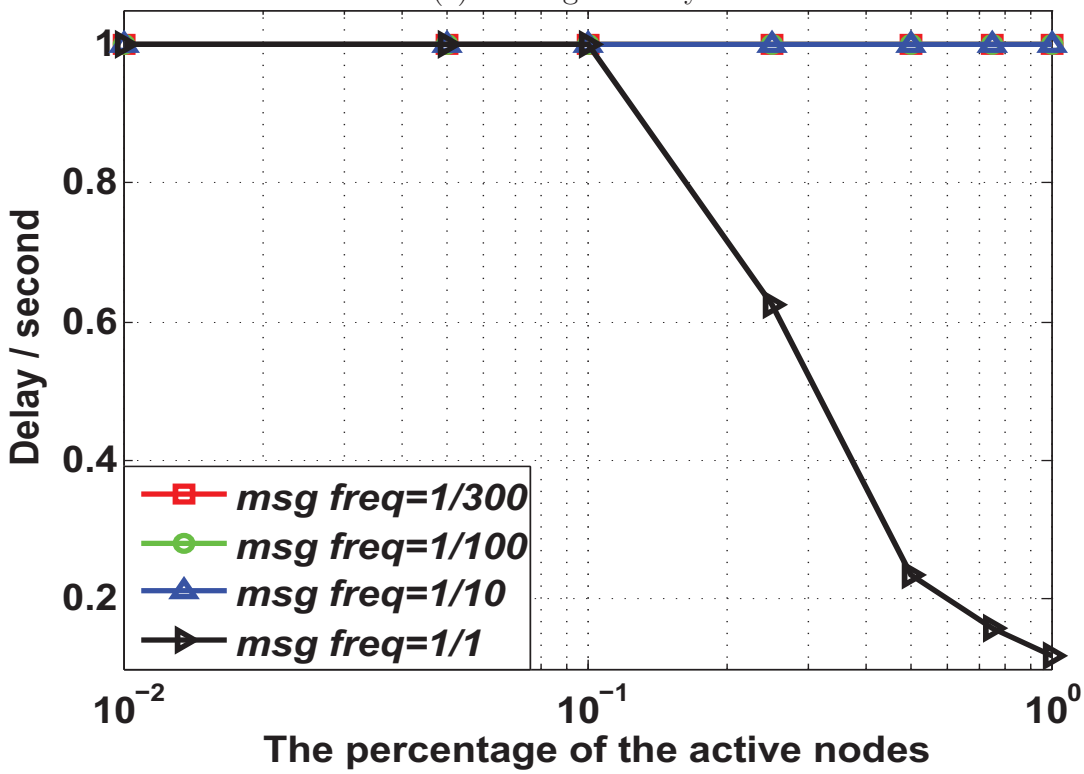


(b) Power consumption of ring nodes

Figure 2.12. Performance of the proposed routing and encryption scheme



(c) Message latency



(d) Message delivery ratio

Figure 2.13. Performance of the proposed routing and encryption scheme

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:

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

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

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

vehicle messages each second.

Simulation results are provided in Fig. 2.12, 2.13 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 8000×8000 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 Fig. 2.12.(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 Fig. 2.13.(d). The simulation results demonstrate that the proposed scheme is very efficient and can be used for practical applications.

2.4 Source-Location Privacy using STaR Routing

In this section, we propose a two-phase routing protocol to provide source-location privacy. In the first phase, the source node randomly selects an intermediate node at the sensor domain and routes the message to the random intermediate node. The random intermediate node services as a fake source when the message is forwarded to the SINK node. In this scheme, the random intermediate node would be located in a pre-determine region around the SINK node. We call this region the Sink Toroidal Region (STaR) [30]. In the second phase, the intermediate node then forwards the message to the SINK node by single-path routing.

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 away 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 distance R away from the SINK node to limit the energy consumption in the routing paths.

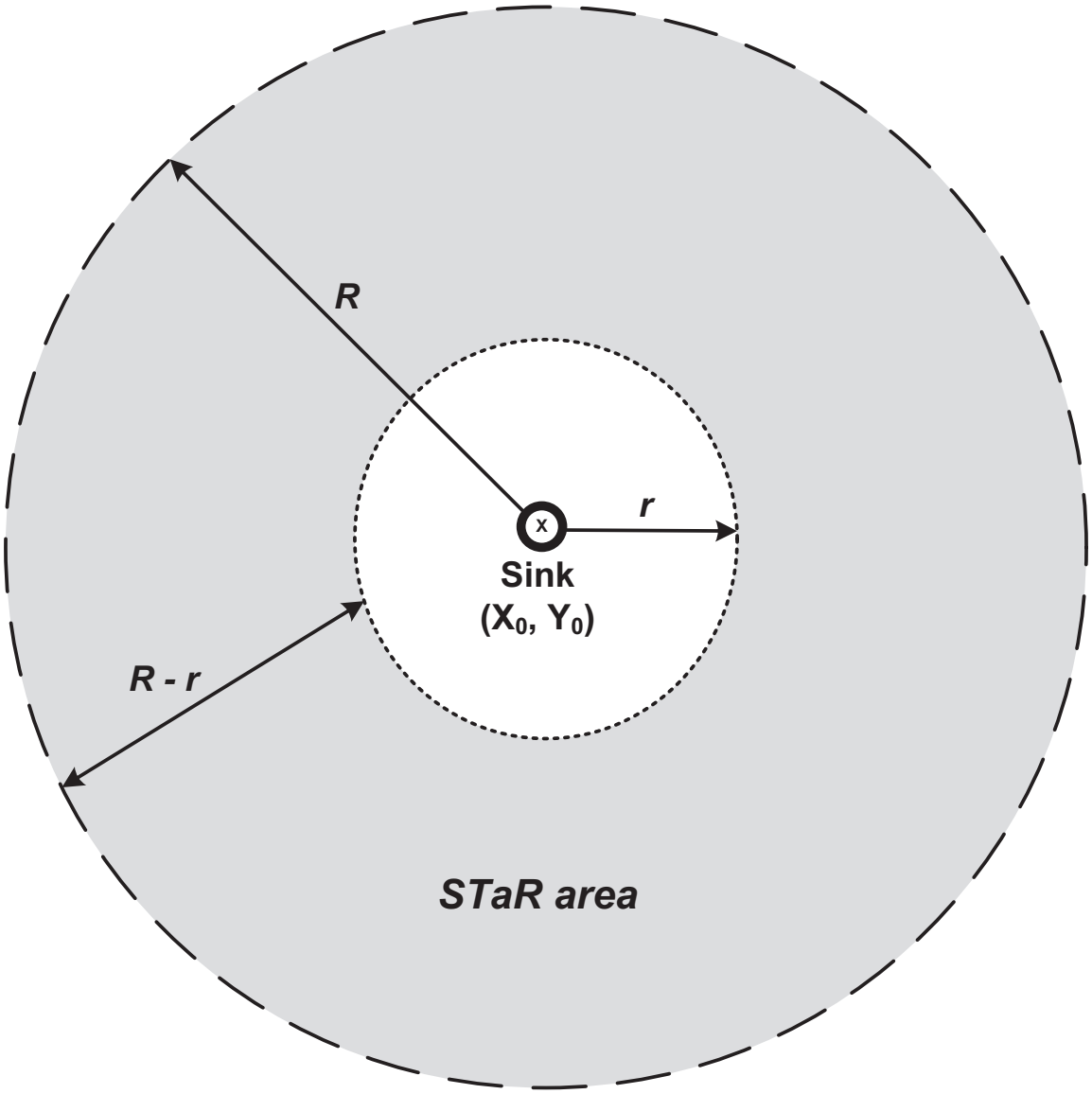


Figure 2.14. Distribution of the STaR area

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 [31] but with less energy consumption.

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 perimeters that is shown in Fig. 2.14. The description of the perimeters 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 perimeters, $\{x_0, y_0, R, r\}$, 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 θ 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 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. The intermediate node then routes the received message to the SINK node using single-path routing.

The proposed scheme will provide adequate source-location privacy since it will repeat this procedure for every message sent out. In general, the source node will send out messages periodically. For every message, the source node will choose a new intermediate node within the STaR area using the procedure described above.

2.4.1 Security Analysis

In this section, we will analyze that the proposed STaR routing scheme can provide source-location privacy.

In our scheme, a random intermediate node is selected from the STaR area. We assume that the STaR area 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. Unlike the directed walk of the phantom routing scheme, our protocol does not leak direction information to the adversaries.

The security of the proposed STaR routing scheme can be analyzed based on the location of the adversarial attacks: i) the adversary monitors traffic between the source node and the randomly selected intermediate node, and ii) the adversary monitors traffic between the randomly selected intermediate node in STaR and the SINK node. For case i), the message source may be located anywhere 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. It is virtually impossible for an adversary intercept multiple messages from the same source, as shown in Fig. 2.15. Therefore, STaR provides local source location privacy.

For case ii), the STaR area is at least a minimum radius distance, r, from the SINK

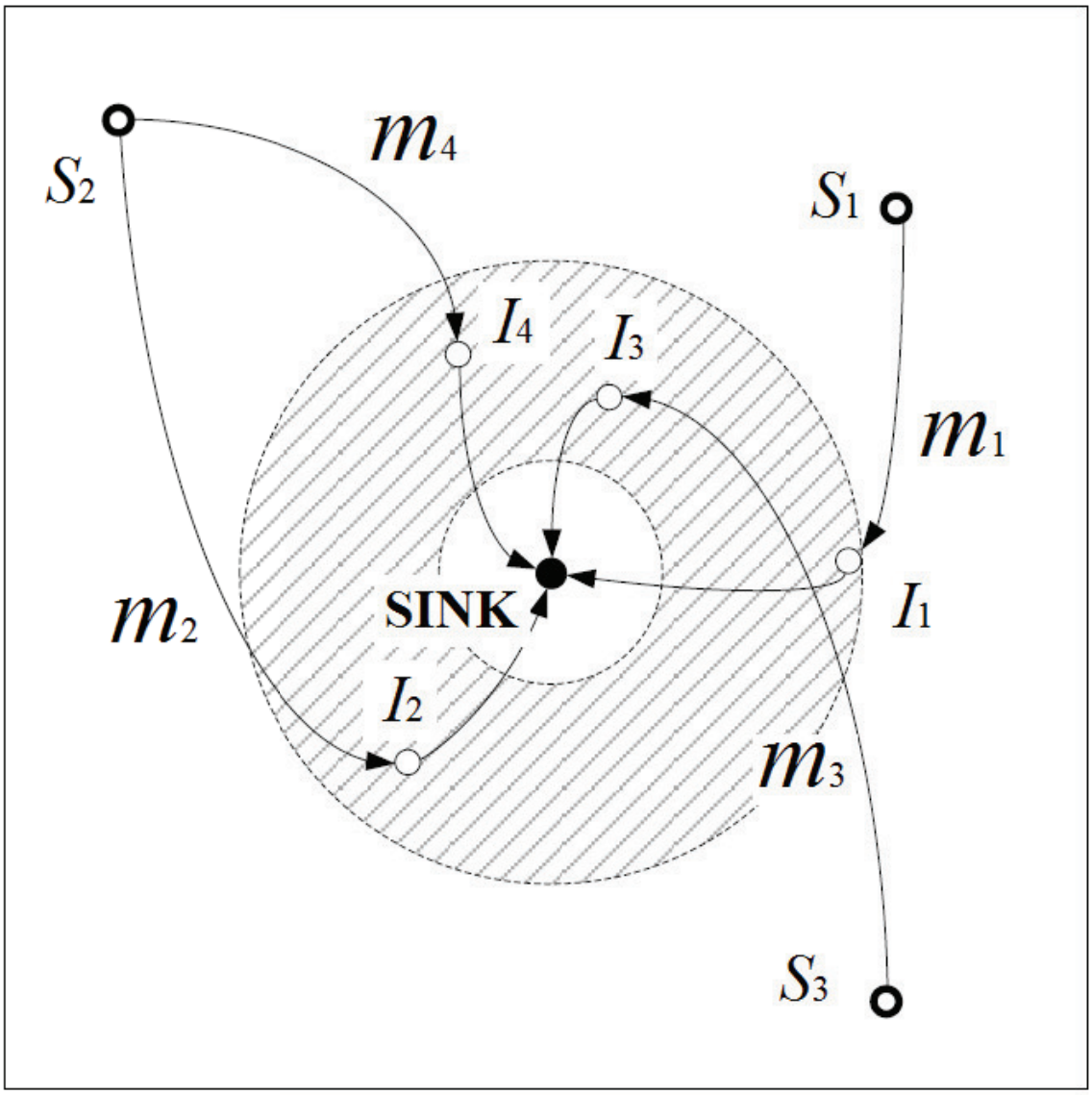


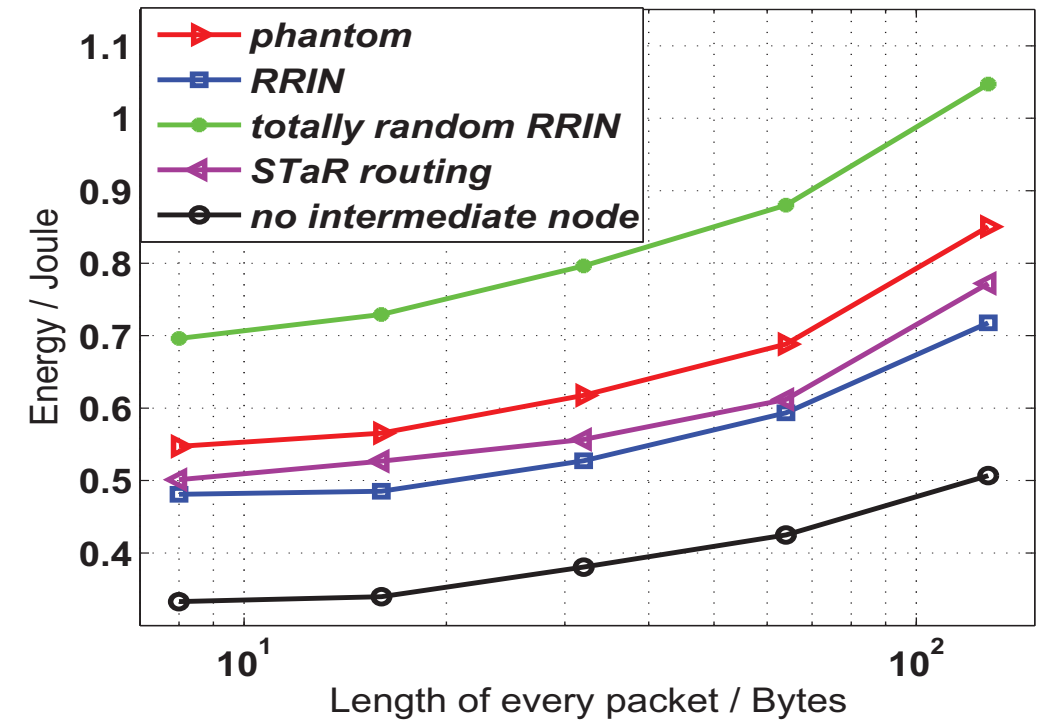
Figure 2.15. Routing illustration of the STaR protocol

node and traffic will be transmitted to the SINK node from all possible directions with equal probability. Therefore, it is quite impossible for an adversary to predict the direction of the source node. 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 way, global source location privacy can be assured.

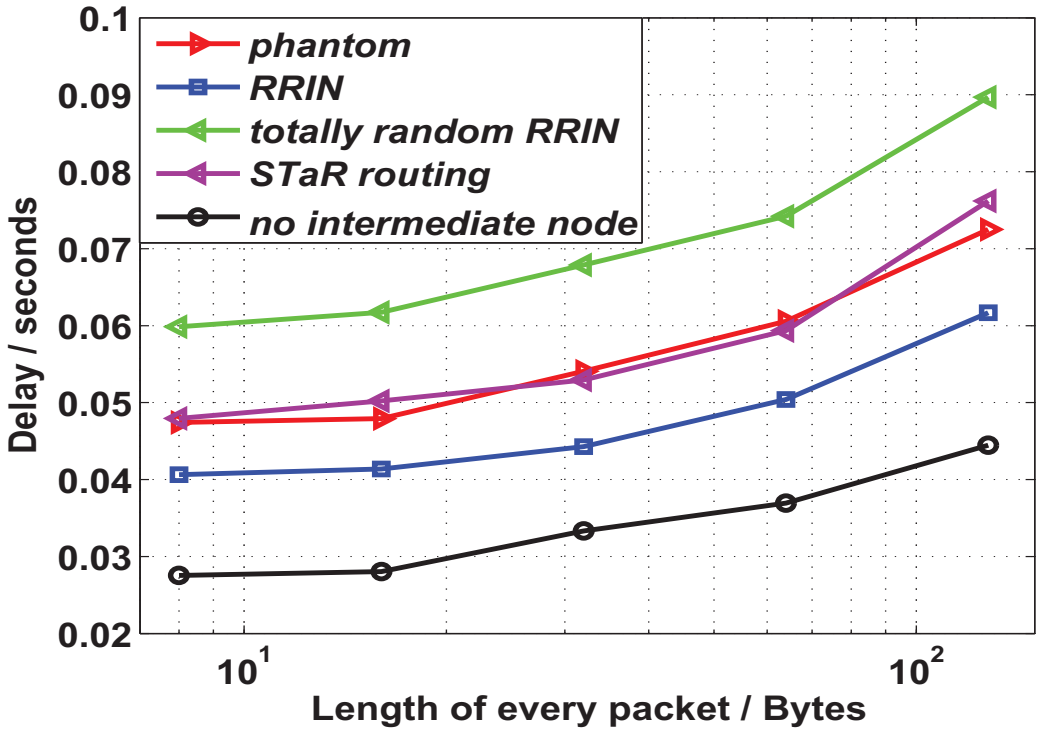
2.4.2 Performance Analysis and Simulation Results

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. 2.16, 2.17. In the simulation, 400 nodes are randomly distributed in a square target area of size 3360×3360 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.

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.

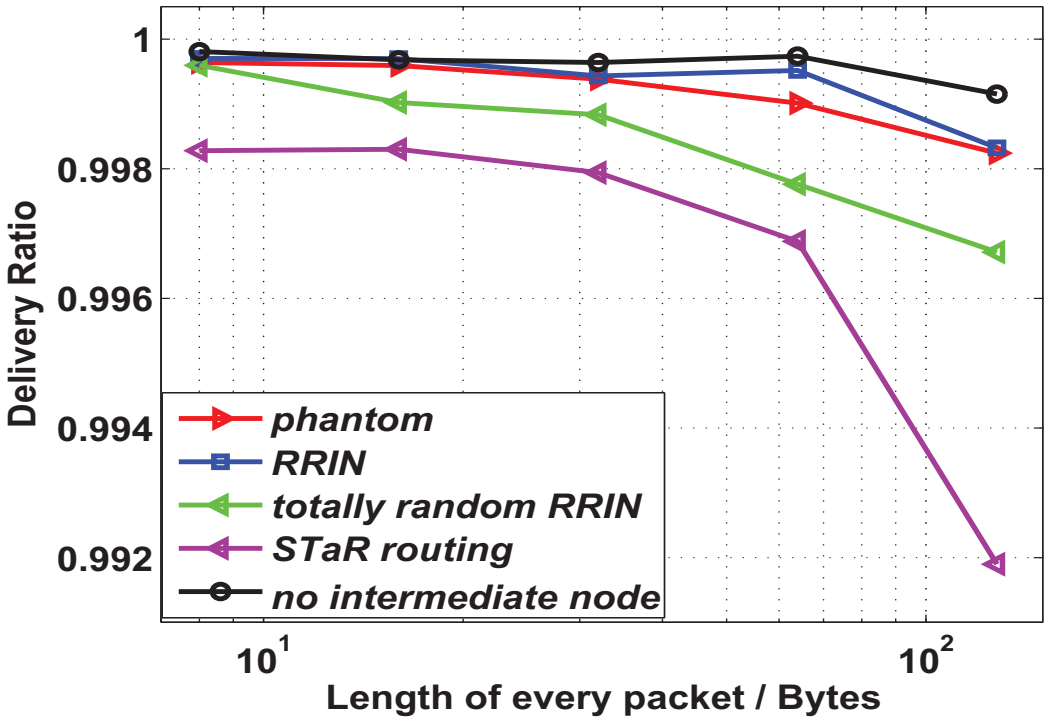


(a) Power consumption



(b) Message latency

Figure 2.16. Performance of routing by single-intermediate node



(c) Message delivery ratio

Figure 2.17. Performance of routing by single-intermediate node

CHAPTER 3

Conclusions

Source-location privacy is vital to the successful deployment of wireless sensor networks. In this report, we propose three schemes to protect the source-location privacy. The first one is implemented by routing through a single randomly selected intermediate node. The second one achieves the location privacy by routing in a network-level mixing ring. The third scheme achieve network-level source-location privacy through a technique we call the Sink Toroidal Region (STaR) routing. We carried out theoretical analysis to evaluate the security and the performance of the proposed schemes and compared it with other existing schemes. Our simulation results demonstrate that the proposed schemes can achieve excellent performance in energy consumption, message delivery ratio and delivery latency.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, February 1981.

[2] D. Chaum, "The dinning cryptographer problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, vol. 1, no. 1, pp. 65–75, 1988.

[3] L. von Ahn, A. Bortz, and N. Hopper, "k-anonymous message transmission," in *Proceedings of CCS*, (Washington D.C., USA.), pp. 122–130, 2003.

[4] A. Beimel and S. Dolev, "Buses for anonymous message delivery," *J. Cryptology*, vol. 16, pp. 25–39, 2003.

[5] P. Golle and A. Juels, "Dining cryptographers revisited," in *Advances in Cryptology - Eurocrypt 2004*, LNCS 3027, pp. 456–473, 2004.

[6] 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.

[7] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous connections and onion routing," *IEEE J. on Selected Areas in Coomunications*, vol. 16, no. 4, pp. 482–494, 1998.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks.," in *IPDPS*, IEEE, 2006.

[16] 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.

[17] 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.

[18] 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.

[19] W. B. Heinzelman, *Application-specific protocol architectures for wireless networks*. PhD thesis, 2000. Supervisor-Anantha P. Chandrakasan and Supervisor-Hari Balakrishnan.

[20] 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.

[21] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 3, pp. 366–379, Oct.-Dec. 2004.

[22] 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.

[23] L. Hu and D. Evans, "Localization for mobile sensor networks," in *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, (New York, NY, USA), pp. 45–57, ACM, 2004.

[24] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "Tps: a time-based positioning scheme for outdoor wireless sensor networks," *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 4, pp. 2685–2696 vol.4, March 2004.

[25] 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.

[26] 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.

[27] 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.

[28] Wikipedia, "Normal distribution." `http://en.wikipedia.org/wiki/Normal_distribution`.

[29] S. M. Stigler, *Statistics on the Table*. Harvard University Press. chapter 22.

[30] L. Lightfoot, Y. Li, and J. Ren, "Preserving source-location privacy in wireless sensor networks using star routing," in *Proc. IEEE Global Telecommunications Conference*, December 2010.

[31] Y. Li, L. Lightfoot, and J. Ren, "Routing-based source-location privacy protection in wireless sensor networks," in *Proc. IEEE International Conference on Electro/Information Technology*, June 2009.