This is to certify that the
dissertation entitled

PROVIDING SOURCE PRIVACY IN WIRELESS SENSOR NETWORKS

presented by

Yun Li

has been accepted towards fulfillment
of the requirements for the

Ph.D degree in Electrical Engineering

Major Professor's Signature

5/11/2010

Date

MSU is an Affirmative Action/Equal Opportunity Employer

PROVIDING SOURCE PRIVACY IN WIRELESS SENSOR NETWORKS

By

Yun Li

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Electrical Engineering

2010

ABSTRACT

PROVIDING SOURCE PRIVACY IN WIRELESS SENSOR NETWORKS

By

Yun Li

Wireless sensor networks (WSNs) have the potential to be widely used in many areas for unattended event monitoring. Mainly due to the lack of a protected physical boundary, wireless communications are vulnerable to unauthorized interception and detection. Security problem has become one of the major issues that jeopardize the successful deployment of WSNs. While message content confidentiality can be ensured through message encryption, it is much more difficult to adequately protect source privacy, which includes source-location privacy and source message authentication. For WSNs, source privacy protection is further complicated by the fact that sensor nodes consist of low-cost and low-power radio devices, computationally intensive cryptographic algorithms (such as public-key cryptosystems) and large scale broadcasting-based protocols are generally not suitable for WSNs.

While many protocols have been proposed to provide source-location privacy, most of them are based on public-key cryptosystems. Others are either energy inefficient or have certain security flaws. In addition, no model has been proposed to quantitatively evaluate security properties of source-location privacy protection schemes. In this dissertation, we first build a security evaluation model and use this model to analyze some of the existing source-location privacy schemes. Then, using the security model as guidance, we propose a dynamic ID assignment scheme and four routing-based source-location privacy schemes. The first routing-based scheme routes each message to a randomly selected intermediate node before it is transmitted to the SINK node. We introduce three intermediate node selection methods, which are constrained method, totally random method, and ring-band based method respectively.

In the second routing-based scheme, a network mixing ring (NMR) is proposed to provide network-level source-location privacy. The third and the fourth routing-based schemes protect source-location privacy through multiple intermediate nodes, which are selected based on angle and quadrant respectively. For each of these routing-based schemes, we provide detailed security analysis and simulation results.

Message authentication is a crucial issue in source privacy protection. Without authentication, even network administrator cannot get source information. In addition, valid messages cannot be distinguished from fake or corrupted messages. Many symmetric-key based or public-key based schemes have been developed to provide message authentication and source non-repudiation services. Most of them, however, can only provide end-to-end authentication, or have the limitations of high computation and communication overhead. To address these issues, a polynomial-based scheme was introduced recently. However, this scheme and its extensions suffer from a built-in threshold limited by the degree of the polynomial. In this dissertation, we propose a scalable authentication scheme based on elliptic curve cryptography (ECC). While enabling hop-by-hop authentication, the proposed scheme allows any node to transmit an unlimited number of messages without suffering the threshold problem. Both theoretical analysis and simulation results demonstrate that the proposed scheme is secure with light overhead.

Dedicated to my family

ACKNOWLEDGMENTS

I would like to take this opportunity to express my appreciation to my advisor, Dr. Jian Ren, for his constant support, guidance and encouragement throughout the years. He makes his best endeavors to help me in every aspect, from providing advice on research to personal development and growth.

I want to thank Dr. Tongtong Li and Dr. Subir Biswas from Department of Electrical and Computer Engineering, Dr. Richard Enbody from Department of Computer Science and Engineering for serving on my committee. I am deeply indebted to them for their kind support, either in the classroom or in all thoughtful correspondences.

I am grateful to all my friends who have made my life at Michigan State University an enjoyable experience. I would like to extend my heartfelt thanks to Peng Liu, Xiaoyan Lu, Yuanteng Pei, Yun Zhou, Ming Gu, Wei Wang, Lei Zhang, Guokai Zeng, Ting Hong, Chenling Huang, Jinglei Xiang, Chang Liu, Shanwen Ji, Fei Chen for being my great friends and for all the fun we have together. I would also like to send a special thank you to my lab mates Wenbo Qiao, Di Tang, Leron Lightfoot for the valuable discussions on research and for the advice on the daily life in the United States.

Last, but not the least, I would like to thank my parents, my grandparents, my wife, my aunts and uncles for their love and continuous support. A special thanks goes to my old friends Zirui Zhu, Guangjie Zhou, Xing Wang, Qi Huang, Yi Wang, Xia Wang, Jia Hou, Lanlan Li, Peng Wang, Dongwu Li, who always care for me as brothers and sisters and have given me valuable suggestions and ceaseless encouragement.

TABLE OF CONTENTS

List of Tables

List of Figures

"Images in this dissertation are presented in color"

CHAPTER 1

Introduction

1.1 Source Privacy in Wireless Communications

Wireless sensor networks (WSNs) have been envisioned as a technology that has a great potential to be widely used in both military and civilian applications. 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 the lack of a physical boundary. In WSNs, the security problem is more serious because sensor nodes consist of low-cost and low-power radio devices. They are designed to operate unattended for long periods of time. Battery recharging or replacement may be infeasible or impossible. Therefore, computationally intensive cryptographic algorithms, such as public-key cryptosystems, and large scale broadcasting-based protocols, may not be quite suitable for WSNs. In addition, the adversaries, who are not restricted to use sensor networks hardware, may use expensive radio transceivers and powerful workstations to interact with the WSNs or get traffic information in WSNs from a distance. In the worst case, adversaries may be able to take control of some sensor nodes, compromise the cryptographic keys and reprogram some sensor nodes. This makes privacy preserving communication in WSNs an extremely challenging research task. Unfortunately, to optimize the sensor nodes for the limited capabilities and application specific nature of WSNs, traditionally, security requirements were largely ignored. This leaves WSNs vulnerable to security attacks.

Source-location privacy is an important security issue for WSNs. Lack of location privacy can cause exposure of significant information about traffic carried on the network and the physical world entities. While confidentiality of a message can be ensured through content encryption, it is much more difficult to adequately address source-location privacy since there is always certain correlation between traffic pattern and source-location information. Using certain equipments to monitor the transmission direction of any detected message, adversaries can easily trace back to the source node hop by hop or deduce the location of the source node through traffic analysis.

Besides source-location privacy, non-repudiation is another property that cannot be ignored for source privacy in wireless communication. Without the non-repudiation, not only attackers, but also network administrators cannot get any information about the source. This makes managing operations almost impossible for network administrators. Lack of non-repudiation also prevents administrators from distinguishing valid messages from fake and unauthorized messages set by attackers. Therefore, attackers could carry out flooding attack to disable the wireless communications in WSNs.

To summarize, there are two aspects that need to be considered for source privacy: source-location privacy and anonymous source authentication.

1.2 Limitations of Existing Solutions

1.2.1 Limitations of Existing Solutions for Location Privacy

In the past two decades, originated largely from Chaum's mixnet [1] and DC-net [2], a number of anonymous communication protocols have been proposed [3–34]. The mixnet family protocols use a set of "mix" servers that shuffle 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 quite suitable for WSNs. Due to the nature of WSNs, the anonymous protocols [22–26] are not suitable for WSNs.

Multiple schemes have been 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 network has to transmit messages at a constant rate. Another base station location privacy scheme was introduced in [35], which involves location privacy routing and fake message injection. Base station location privacy protection is also discussed in [27,28]. However, base station location privacy and source location privacy are different research issues, because the location of the base station is usually fixed while the source location tends to be dynamic. In this dissertation, we will address the source-location privacy in WSNs.

The authors of [29] proposed to achieve source-location privacy through trusted mechanism built in the WSNs and neighboring nodes categorization. However, this scheme requires a long delay after network distribution to build the trusted reputation infrastructure through WSNs, which is not quite practical for resource constrained WSNs. Two schemes were proposed in [30]: Simple Anonymity Scheme (SAS) and Cryptographic Anonymity Scheme (CAS) for establishing anonymity in clustered WSNs. SAS is implemented through large number of non-contiguous pseudonyms stored in sensor nodes, which is memory inefficient. CAS is implemented through key hash function to generate pseudonyms which is computation inefficient. The most serious problem for these two schemes is the fact that no routing schemes are introduced. So they cannot resist hop-by-hop traceback attacks.

In [11,12,31–34], source-location privacy is provided through broadcasting or dummy messages injection that mixes valid messages with dummy messages. The main idea is that each node needs to transmit messages consistently or probabilistically. Whenever there is no valid message, the node has to transmit dummy messages. The transmission of dummy messages not only increases the networks collisions ratio and decreases the packet delivery ratio, but also consumes significant amount of sensor energy due to the fact that, on average, transmission of one bit consumes about as much power as executing 800-1000 instructions [36]. Therefore, these schemes are not quite suitable for large scale WSNs.

Routing-based protocols can also provide source-location privacy through dynamic

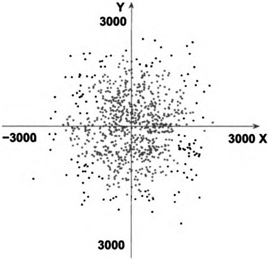


Figure 1.1. Nodes distribution through random routing

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/nodes 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. An example is shown in Fig. 1.1, where the source node is located at $(0, 0)$. This source node generates 1000 packets. Each packet is routed 50 hops until it reaches a randomly chosen node. The transmission range is 250 meters at most for one hop. We can see that most of the randomly selected nodes are located relatively close to the source node. Our statistics shows that the average distance between the source and the randomly selected nodes is only 4.2 hops, and the longest distance is just 12.2 hops. To solve this problem, several approaches have been proposed. In phantom routing protocol [13,14,21], 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. Then 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 contained in the header of the message. The exposure of direction information decreases the complexity for adversaries to trace back to the actual message source in a magnitude of 2^h. Random walks from both the source node and the SINK node were also used 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, in this scheme, the adversaries can recover significant routing information from received messages. In addition, this design is "not realistic" for large scale WSNs.

In addition, although source-location privacy has been discussed in many papers, the research on quantitative measurement and analysis of source-location privacy are largely unfolded. Without a security evaluation model, it will be difficult to quantitatively measure the security properties of different source-location privacy protection schemes

1.2.2 Limitations of Existing Solutions for Message Authentication

In [37,38], symmetric key and hash based authentication schemes were proposed for WSNs. In these schemes, each symmetric authentication key is shared by a group of sensor nodes. An intruder can compromise the key by capturing a single sensor node. Therefore, these schemes are not resilient to a large number of nodes compromising. Another type of symmetric-key scheme requires synchronization among nodes. These schemes, including TESLA and its variants [39–44], can also provide message sender authentication. However, this scheme requires initial time synchronization, which is difficult to be implemented in large scale WSNs. In addition, they also introduce delay in message authentication, and the delay increases as the network scales up.

A secret polynomial based message authentication scheme was introduced in [45]. This scheme offers information theoretic security with ideas similar to a threshold secret sharing, where the threshold is determined by the degree of the polynomial. When the number of messages transmitted is below the threshold, the scheme enables the intermediate node to verify the authenticity of the message through polynomial evaluation. However, when the number of messages transmitted is larger than the threshold, the polynomial can be fully recovered and the system is completely broken. To increase the threshold and the complexity for the intruder to break the secret polynomial, a random noise, also called a perturbation factor, was added to the polynomial in [46–48]. The main idea is to thwart the adversary from computing the coefficient of the polynomial. However, the added perturbation factor can be completely removed using error-correcting code techniques [49].

Many protocols have been proposed to address message authentication through public-key infrastructure [50–56]. Public-key based approaches have simple and clean key management. However, this kinds of approach may cause high overhead in terms of computation and communication cost.

Some key distribution schemes [57–60] have also been proposed to solve these problems. However, in these schemes, the sensor nodes have to store a large number of extra keys which might not be used at all. This design wastes the storage resources of the sensor nodes. In addition, these schemes are vulnerable to a large number of node compromising.

Recently, message sender anonymity based on ring signatures was introduced [61]. This approach enables the message sender to generate a source anonymous message signature with content authenticity assurance. To generate a ring signature, a ring member randomly selects an ambiguity set (AS) and forges a message signature for all other members. Then he uses his trap-door information to glue the ring together. The original scheme has very limited flexibility and very high complexity. Moreover, the original paper only focuses on the cryptographic algorithm, and the relevant network issues were left unaddressed.

For the protocols mentioned above, only [48] can provide hop-by-hop authentication, while others can only achieve end-to-end authentication. For end-to-end authentication, only the receiver can verify the authenticity of messages en-route. This means that no intermediate node can authenticate the message in general. The intermediate nodes may have to forward a manipulated message for many hops before the message can finally be authenticated and dropped by the destination node. This not only consumes extra sensor power, but also increases network collision and decreases message deliver ratio. With hop-by-hop authentication, every forwarder en-route is able to authenticate the forwarded messages and discard faked or corrupted messages immediately.

1.2.3 Summary of Major Limitations

The major limitations of existing schemes can be summarized as follows:

- Public-key based schemes are not suitable for source-location privacy protection in WSNs due to high computation and communication overhead.

- Broadcasting-based source-location privacy protection schemes are not suitable in WSNs due to energy inefficiency.

- Existing Routing-based source-location privacy protection schemes suffer certain security flaw, which leaks source-location information to adversaries.

- No evaluation model has been proposed to quantitatively measure security properties of source-location privacy protection schemes.

- Most of the existing authentication schemes have threshold limitation and are not resilient to a large number of node compromising.

- Most of the existing authentication schemes can only provide end-to-end authentication instead of hop-by-hop authentication.

1.3 Proposed Research Directions

As mentioned before, there are three problems that have not been solved for source privacy: security property evaluation model for source-location privacy protection schemes, source-location privacy protection, and anonymous source authentication.

1.3.1 Evaluation Model For Source-Location Privacy

For source-location privacy protection, the first important task is to build a security evaluation model. With a quantitatively measurable model, we can evaluate the security properties of any proposed source-location privacy protection scheme, and provide guidance for future scheme design. In this dissertation, we will first categorize the methods an adversary may use to locate the source node. Then we will build a security evaluation mode based on the analysis.

After the model is built up, we will use it to evaluate some of the existing source-location privacy protection schemes. We will also describe our own schemes along with theoretical analysis and comprehensive simulation results.

1.3.2 Directions For Source-Location Privacy Protection

As a rule of thumb, in order to provide source-location privacy, no source information should be contained in the message content and the message should not have any identifiable ID information. In this dissertation, we will propose a dynamic ID assignment scheme to prevent the adversaries from locating the source node through ID information contained in the messages.

We will also propose four routing-based schemes to address the source-location privacy in this dissertation.

In the first routing-based scheme, for each message to be transmitted, the message source first randomly selects an intermediate node in the sensor domain, and then transmits this message to the intermediate node before this message is forwarded to the SINK node by the intermediate node. We will propose three different methods to select an intermediate node: constrained method, totally random method and the ring band around the source method.

The second routing-based 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 aims at providing local source-location privacy. In the second routing phase, each message will be mixed with other messages through a network mixing ring (NMR). This phase offers network-level (global) source-location privacy. In the last phase, each message will be forwarded to the SINK node from certain specific nodes on the mixing ring.

The third and the fourth routing-based schemes achieve network-level (global) source-location privacy by routing through multiple intermediate nodes. Selection methods of these intermediate nodes are based on angle and quadrant respectively.

For each of these routing-based schemes, we will provide detailed security analysis and comprehensive simulation results to evaluate its security property and communication performance.

1.3.3 Directions For Anonymous Source Authentication

To address the limitations of existing works for anonymous source authentication, we will develop an energy efficient authentication scheme which can provide hop-by-hop authentication without the threshold problem.

Recent research [56] has shown that elliptic curve cryptography (ECC) can achieve similar security property as public-key infrastructure (PKI) while maintaining small computation and communication overhead. In this dissertation, we propose a scalable authentication scheme based on ECC. While enabling intermediate node authentication, the proposed scheme allows any node to transmit an unlimited number of messages without suffering the threshold problem.

1.4 Overview of the Dissertation

1.4.1 Design Goals

Our design goals for the routing-based schemes can be summarized as follows:

- Adversaries should not be able to get any source-location information by analyzing traffic pattern.

- Adversaries should not be able to get any source-location information even if they are able to monitor certain area of the WSNs and compromise a few network nodes.

- Only the SINK node is able to identify the location of a source node through the received messages. This recovery process should be efficient.

- The length of each message should be as short as possible to save previous sensor node power.

We also want to achieve the following goals for our authentication scheme:

- *Message authentication:* Message receiver should be able to verify whether a received message is sent by the node it claims to be, or by a node in a particular group of nodes. In other words, an adversary cannot pretend to be an innocent node to inject fake messages into the network without being detected.

- *Message integrity:* Message receiver should be able to verify that the message has not been modified en-route by an adversary. In other words, an adversary cannot modify any messages' content without being detected.

- *Hop-by-hop message authentication:* Every forwarder in the routing path should be able to verify the authenticity and integrity of a message upon reception.

- *Identity and location privacy:* An adversary cannot determine the message sender's ID and location by analyzing the message content or the local traffic pattern.

- *Node compromising resilience:* The scheme should be resilient to compromised node. In other words, no matter how many nodes are compromised, remaining nodes should still be secure.

- *Efficiency:* The scheme should be efficient in both computation and communication.

1.4.2 Major Contributions

The major contributions of this dissertation are the following:

1. We build a security evaluation model for source-location privacy protection schemes and analyze security properties of some existing routing-based source-location privacy protection schemes using this model.

2. We propose a dynamic ID assignment scheme to prevent source-location information leaking through correlation-based source identification.

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

4. We develop a source-location privacy protection scheme through a network-level mixing ring.

5. We develop a source-location privacy protection scheme through routing to multiple angle-based randomly selected intermediate nodes.

6. We develop a source-location privacy protection scheme through routing to multiple quadrant-based randomly selected intermediate nodes.

7. We develop a source anonymous message authentication scheme on ECC that can provide unconditional source anonymity.

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

9. We propose an efficient key management framework ensuring the identified compromised node to be isolated.

10. We provide extensive simulation results using ns-2 and Maple for the routing-based schemes and authentication scheme proposed in this dissertation.

1.4.3 Thesis Organization

The dissertation is organized as follows.

Chapter II introduces a quantitative evaluation model for source-location privacy.

The first section serves as an introduction for this chapter. In the second section, we present the network and adversarial assumptions, which will be used throughout this dissertation. In the third section, a security evaluation model for source-location privacy protection schemes is introduced. This model can be used to quantitatively measure source-location privacy and provide guidance for new scheme design. Based on the model, we evaluate some existing source-location privacy protection schemes in the fourth section. This chapter is concluded in the last section.

Chapter III introduces the design for source-location privacy protection.

The dynamic ID assignment scheme aims at preventing adversaries from getting any useful source-location information through correlation-based source identification attacks.

The routing-based schemes aim at protecting source-location information from traffic analysis attacks. The first source-location privacy protection scheme is implemented through routing to a single intermediate node. Three intermediate node selection methods are introduced. Although these methods achieve some improvement comparing to existing schemes, detailed analysis shows that they still have some security flaws. To address these problems, we proposed three source-location privacy protection schemes, which are routing with a network mixing ring (NMR), routing through angle-based multi-intermediate nodes, routing through quadrant-based multi-intermediate nodes, respectively.

For each of these routing-based source-location privacy protection schemes, we provide detailed security analysis and simulation results to demonstrate that it is secure, efficient and can be used in many practical applications.

Chapter IV is mainly about the design for source anonymous message authentication. It consists of five sections.

The first section gives an introduction to the whole chapter. The second section presents some terminologies that will be used in the later sections. The third section describes the authentication scheme in detail. The fourth section discussed the anonymity set selection for source privacy. The fifth section is key management and compromised nodes detection. Performance analysis is provided in the sixth section. We conclude this chapter in the last section.

Chapter V summarizes the contributions and concludes the dissertation. An outline of related future work is also provided.

CHAPTER 2

Quantitative Evaluation Model for Source-Location Privacy

2.1 Introduction

To develop source-location privacy protection schemes, first, we have to specify the network environment that our schemes can be applied to. We also need to define the adversaries these schemes should be against. Second, we have to build a security evaluation model. Without a quantitatively measurable model, it is not accurate to evaluate the security properties of any proposed source-location privacy protection scheme.

In this chapter, we will first present our network assumptions and adversarial model in Section 2.2. Then, in Section 2.3, we will introduce a security evaluation model. Using this evaluation model, we will analyze the security properties of some existing source-location privacy protection schemes in Section 2.4. Section 2.5 concludes this chapter.

2.2 Models and Assumptions

2.2.1 Network Assumptions

We make the following assumptions about our system:

- The networks are evenly divided into small grids. The sensor nodes are randomly distributed in the target area. Sensor nodes in each grid are all fully connected. In each grid, there is one header node that is responsible for communicating with other header nodes nearby. The whole network is fully connected through multi-hop communications. The formation of the grid and the header node selection in each grid have been studied in many literature works [62–73].

- Information of the SINK node is public. It is the destination that all data messages will be transmitted to through multi-hop routing paths.

- The content of messages will be encrypted using the secret keys shared between nodes/grids and the SINK node. However, the encryption operation is beyond the scope of this dissertation.

- Sensor nodes are assumed to have knowledge of their relative locations and their adjacent neighboring nodes. Relative location information of the sensor domain may be obtained through network broadcasting [74–82].

- The key generation, key distribution and key update, are beyond the scope of this dissertation. However, the interested readers are referred to references such as [44, 58, 60, 83–90].

2.2.2 Adversarial Model

We assume that the adversaries in the target area will try to locate the source node through traffic analysis and routing traceback. Adversaries have the following characteristics in this dissertation:

- Adversaries have unbounded energy resource, adequate computation capability and sufficient memory for data storage. Adversaries may also compromise some sensor nodes in the networks.

- Adversaries could eavesdrop on messages transmission in the network, and perform traffic analysis or decryption in order to derive some valuable information.

- Adversaries are able to monitor the traffic in an area and get all of the transmitted messages. On detecting an event, they could determine the immediate sender by analyzing the strength and direction of the message they received. However, we assume that the adversaries are unable to monitor the entire WSNs.

- Adversaries can compromise some sensor nodes. Once the sensor nodes are compromised, the adversaries will get all the information contained in the compromised nodes, including the security parameters of the compromised nodes. Adversaries can modify the contents of the messages and inject their own messages.

2.3 Location Privacy Evaluation Model

Although source-location privacy has been discussed in many papers, the research on quantitative measurement and analysis of source-location privacy are largely unfolded. In this section, we define some criteria for routing-based source-location privacy protection schemes. These criteria can be used to quantitatively measure source-location privacy and provide guidance for new scheme design.

In a network, an adversary can always try to derive the source-location information from a captured message through traffic analysis and routing traceback. The analysis that an adversary may try to get location information of the source node can be divided into three categories.

1. *Correlation-based source identification attack*, i.e., source-location recovery based on ID analysis. If a message is received by an adversary, and the adversary already knows the location of the node for this ID contained in the message, then the adversary would be able to locate the source node immediately.

2. *Routing traceback attack*, i.e., the adversary can move to the immediate forwarder on capturing a transmitted message until the source node is reached.

3. *Reducing source space attack*, i.e., on capturing a transmitted message, the adversary can reduce the source node to a subgroup of the nodes in the networks.

With multiple such captured messages, this subgroup may be reduced to contain only a small number of nodes, which means the source-location privacy is almost leaked.

Traditionally, each transmitted message bears a fixed message ID. The adversary can get information of the message source from the fixed ID in one of two ways: the message source, or the messages that transmitted from the same source. For the second approach, the adversary will eventually be able to find the message source based on routing traceback.

To prevent correlation-based source identification, in this dissertation, a dynamic ID based approach is proposed to prevent adversaries from relating messages transmitted from each source. This can be achieved by requiring that each node in the network to be preloaded with an *ID-hash-chain* so that a different 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, we define two criteria to measure the security properties of source-location privacy protection schemes.

Definition 2.1 (Source-location Disclosure Index (SDI)) SDI *measures, from probability point of view, the amount of source-location information that one message can leak to the adversaries.*

For a source-location privacy scheme, if SDI is fixed for one source node S, then it means an adversary needs to detect $\lceil \frac{1}{SDI} \rceil$ messages sent from S in order to successfully locate it. For a good source-location privacy protection scheme, SDI should be as small as possible.

Definition 2.2 (Source-location Space Index (SSI)) SSI *is defined as the set of possible network nodes that a message can be transmitted from.*

For a source-location privacy scheme, if SSI is large for one detected message, it means this message may be sent out by many possible source nodes. Therefore, it will be hard for an adversary to determine which node is the real source node. On the contrary, if *SSI* is small, then the adversary can easily limit the possible source nodes to a small group. For a source-location privacy protection scheme, *SSI* should be as large as possible.

Considering that different networks are composed of different number of nodes, we give a definition for *normalized SSI* to facilitate the comparison of *SSI*s between different networks.

Definition 2.3 (Normalized Source-location Space Index (NSSI)) NSSI *is defined as the ratio of* SSI *over the total number of nodes in the networks domain.*

According to our assumption, sensor nodes are evenly distributed in the target area, which means *NSSI* equals the ratio of *SSI* area over the total area of the networks domain.

Obviously, for an ideal source-location privacy protection scheme, it should have the following properties:

$$SDI = 0,$$

$$NSSI = 1.$$

We define these security properties as *global location privacy* or *network-level location privacy.* Otherwise, the security properties are defined as *local location privacy.* Global location privacy will be our design goal.

2.4 Security Analysis of the Existing Source-Location Privacy Protection Schemes

In this section, we will analyze the security properties of some well-known routing-based source-location privacy protection schemes using our proposed evaluation model.

2.4.1 Security Evaluation for Fixed Path Routing

We assume the WSNs consists of N nodes. We will first assume that an attacker's sensing range is the same as the regular sensor nodes in the sensor domain.

Lemma 2.1 *Suppose there is a fixed routing path between the source node S and the destination node D with a length of L hops. An adversary A, who hides near D, will be able to detect all messages transmitted to D. Then after receiving L messages, A will be able to trace back to S, i.e.,*

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

This is because that for each message received, the adversary A can move one hop closer to the source node S. In other words, for each received message, the SDI equals to $\frac{1}{L}$ of the source location to the attackers. Since the source node S and the destination node D is only L hops away, we only need L messages in order to fully trace back from the destination node D to the source node S.

This is the least secure source-location privacy protection scheme that we can imagine. To increase source-location privacy, multiple schemes have been proposed [91–97] through non-intersected routing paths between the source node and the destination node. As shown in Fig. 2.1, between source node S and destination node D, n routing paths exist. Suppose the lengthes of the n paths are: $L_1, L_2, \cdots, L_n$, 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 $SDI_i = \frac{p_i}{L_i}, i = 1, \cdots, n$. Define the overall SDI as

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

then we have the following result.

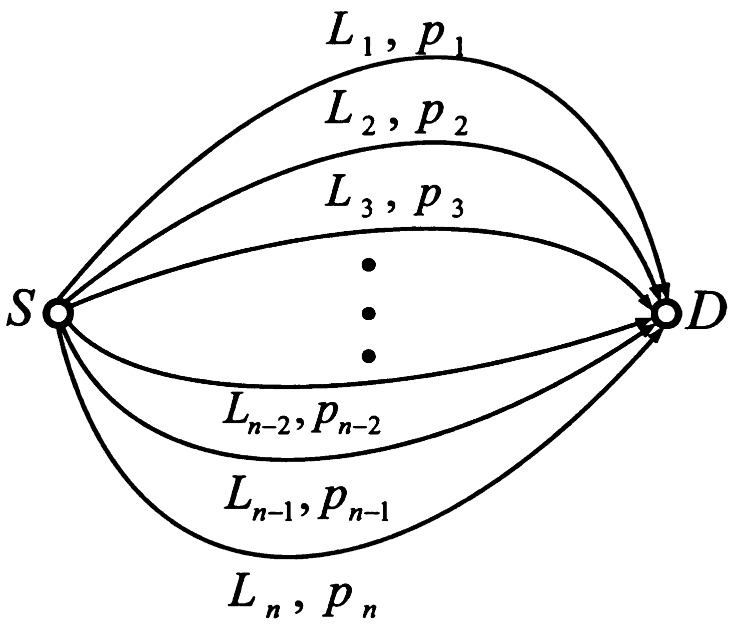


Figure 2.1. Multiple fixed routing paths

Theorem 2.1 *For each received message, when* $p_i = \frac{L_i}{L_1+L_2+\cdots+L_n}$, $i = 1, 2, \cdots, n$, *the SDI is minimized, which is*

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

Proof: To find the minimal of $SDI(p_1, p_2, \cdots, p_n) = \sum_{i=1}^{n} p_i \cdot SDI_i = \frac{p_1^2}{L_1} + \frac{p_2^2}{L_2} + \cdots + \frac{p_n^2}{L_n}$, subject to the constrain $p_1 + p_2 + \cdots + p_n = 1$, we will use Lagrange multipliers to find the minimum SDI.

Define

$$F(p_1, p_2, \cdots, p_n, \lambda) = \frac{p_1^2}{L_1} + \frac{p_2^2}{L_2} + \cdots + \frac{p_n^2}{L_n} + \lambda \cdot (p_1 + p_2 + \cdots + p_{n-1} - 1).$$

Let $\nabla_{p_1,p_2,\cdots,p_n} F(p_1, p_2, \cdots, p_n) = (\frac{\partial F}{\partial p_1}, \frac{\partial F}{\partial p_2}, \cdots, \frac{\partial F}{\partial p_n}, \frac{\partial F}{\partial \lambda}) = 0$, then we have

$$\begin{cases} F'_{p_1} = \frac{2p_1}{L_1} + \lambda = 0 \\ F'_{p_2} = \frac{2p_2}{L_2} + \lambda = 0 \\ \qquad\vdots \\ F'_{p_n} = \frac{2p_n}{L_n} + \lambda = 0 \\ p_1 + p_2 + \cdots + p_n = 1. \end{cases}$$

We can solve that $\lambda = -\frac{2p_1}{L_1} = -\frac{2p_2}{L_2} = \cdots = -\frac{2p_n}{L_n} = -\frac{2}{L_1+L_2+\cdots+L_n}$, and the only stationary point is

$$\left\{ p_1 = \frac{L_1}{L_1 + L_2 + \cdots + L_n}, p_2 = \frac{L_1}{L_1 + L_2 + \cdots + L_n}, \cdots, p_n = \frac{L_n}{L_1 + L_2 + \cdots + L_n} \right\},$$

which corresponds to the minimal value of SDI. □

Corollary 2.1 *Suppose there are n non-intersected routing paths between the source node S and the destination node D. The lengthes of the n paths are* $L_1, L_2, \cdots, L_n$, *respectively, then an adversary needs to receive*

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

messages on average to determine the location of the source node, i.e., trace back to the source node.

Note that for a single adversary, Corollary 1 only gives the average number of packets required to find the message source. If multiple adversaries collaborate and monitor all the routing paths for message transmission, then the adversaries can fully identified the message source with at most $L_1 + L_2 + \cdots + L_n - (n-1)$ received messages. To provide source-location privacy in a network, we have to increase the total number of possible routing paths between the destination node and the source node. However, for any practical network configuration, the number of fixed routing paths cannot be increased without limitation. This means: $SDI > 0$.

We can summarize the two defects of the source-location privacy protection schemes through fixed routing path as follows:

- Positive *SDI*: For fixed path routing, no matter how dedicated the scheme is designed, *SDI* must be positive. In other words, for each message sent out by one source node, from the probability point of view, there is always a fraction of source information to be leaked to adversaries. No matter how small the *SDI* is, given enough number of detected messages, adversaries will be able to locate the source node, i.e., no absolute security can be achieved.

- Limited *SSI*: Because the paths are fixed for each source node, the association between the messages transmitted on a particular path and the source node which is connected to this path is definitely high. In other words, the value of *SSI* cannot be high enough to achieve satisfying security level.

All the analysis above is based on the assumption that an adversary is able to identify messages transmitted from one source. In the case that the adversary is unable to distinguish the messages from different sources and determine which messages are sent by the same source, say S, the situation will be totally different. For instance, the information transmitted from the source S may help an adversary to get closer to the source node. However, information carried in messages generated by other source nodes may mislead the adversary to move away from the source node S. In this case, the SDI may be close to zero since no information can be linked to the message source S. In particular, Theorem 1 and Corollary 1 may no longer be applicable.

As we have analyzed, routing through fixed path cannot achieve our design goals. We propose to achieve source-location privacy through dynamic routing.

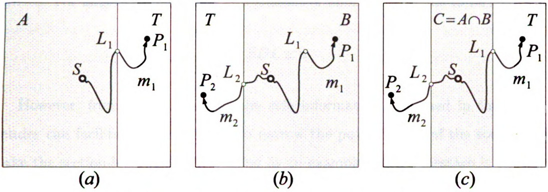


Figure 2.2. NSSI calculation for section-based phantom routing

2.4.2 Security Evaluation for Dynamic Path Routing

As mentioned in the Section 1.2, phantom routing is the most representative protocol for the existing dynamic routing-based source-location protection schemes. In this subsection, we will analyze the security properties of phantom routing scheme in detail.

In phantom routing, the message is first routed to a phantom source through a random path before it is forwarded to the real destination node. We can divide phantom routing into two phases: directed walk phase and routing-to-SINK phase. We will analyze both of them separately.

In the first phase, to make sure that the phantom source is away from the real source node, the side information must be contained in the message's header. In this way, the intermediate nodes on the random path will be able to select the next forwarder on the random path.

From an adversary's point of view, on corrupting a message on the random walking path, the adversary will be able to move one hop closer to the real source node. However, because the routing path is not fixed, the possibility that the adversary can receive another message sent from the same source node is pretty small, especially for a large scale network. In other words, from this aspect, the amount of source-location information that one message on the random path can leak to the adversary is pretty limited. For large scale networks, this amount of information tends to be zero, i.e.:

$$SDI \simeq 0.$$

However, from another aspect, the side information contained in the message header can facilitate the adversary to narrow the possible area of the source node. Take the section-based random walking 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 real source node is located. Fig. 2.2.(a) illustrates this situation, in which T is the target network area, S and P_1 are the real source and the phantom source respectively. Message m_1 is corrupted at L_1, which is located on a random path from S to P_1. On this routing path, the next forwarder is to the right of the current forwarder. Therefore, the adversary will know that S is located to the left of P_1, which is the shaded area A. So we have:

$$NSSI = \frac{area(A)}{area(T)},$$

where $area(X)$ is a function of the area of X. The value of $NSSI$ depends on the location of L_1 and the side information contained in m_1. On average, $NSSI = 50\%$.

If we assume there are multiple adversaries in the target area T and they cooperate to locate the real source node, the value of $NSSI$ can be even smaller. For example, in Fig. 2.2.(b), the real source S sends message m_2 to another phantom source P_2 through a new selected random path, on which the next forwarder is located to the left of the current forwarder. If m_2 is corrupted at L_2, then the possible location area of S determined by m_2 is B. By combining the source-location information got from m_1 and m_2, the possible location area of S will be $C = A \bigcap B$, which is shown in Fig. 2.2.(c), i.e.:

$$NSSI = \frac{area(C)}{area(T)},$$

which is much smaller than both $\frac{area(A)}{area(T)}$ and $\frac{area(B)}{area(T)}$. Suppose there are multiple adversaries to collaborate, the value of $NSSI$ can be further reduced and the source-

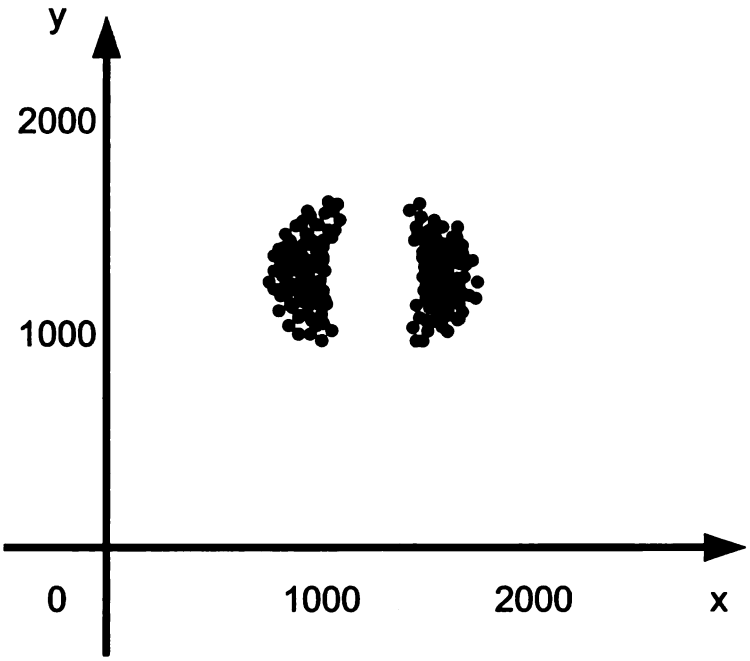


Figure 2.3. Phantom sources distribution for section-based phantom routing scheme

location privacy is no longer well protected.

Now, we will analyze the security property of the routing-to-SINK phase in phantom routing. For large scale WSNs, the possibility that phantom sources are close to the real source node is high. For instance, as shown in Fig. 2.3, the x-coordinate range and the y-coordinate range of the target area are both $[-2500, 2500]$. The source node is located at $(1250, 1250)$. We use section-based directed walk. Each phantom source is 10 hops away from the real source node, while one hop corresponds to 160 at least. We randomly selected 200 intermediate nodes. It can be seen that all these intermediate nodes are located relatively close to the real source node. When an adversary around the SINK node, which is located at $(0, 0)$, collects several messages from this source, he/she will be pretty sure that the source is located in the first quadrant, i.e,:

$$NSSI = \frac{2500 * 2500}{5000 * 5000} = \frac{1}{4}.$$

Obviously, this value indicates that the security level is not high enough. For large scale WSNs, the second phase of phantom routing scheme may not be able to provide network-level source-location privacy.

2.5 Summary

In this chapter, first, we made some assumptions for both the networks and the adversaries. Then, we developed an evaluation model to quantitative measure the security properties of source-location privacy protection schemes. Using this evaluation model, we analyzed some existing schemes which aim at protecting source-location privacy.

CHAPTER 3

Design for Source-Location Privacy Protection

In this chapter, we propose a dynamic ID assignment scheme and four routing-based source-location privacy protection schemes. The dynamic ID assignment scheme aims at preventing source-location information from be leaked in message content. The routing-based schemes protect source location from traffic pattern analysis. The first scheme routes each message to a randomly selected intermediate node before the message is forwarded to the SINK node. There are three different ways to select the intermediate node: constrained method [98], totally random method [99], and ring-band method. Each of them has different security property and communication performance. The second scheme is proposed to achieve network-level source-location privacy through a network mixing ring (NMR) [100, 101], which will mix the messages transmitted in WSNs completely. The third and the fourth schemes achieve source-location privacy through routing to multiple intermediate nodes [102]. The intermediate nodes are selected based on angle and quadrant respectively. For each of these protocols, detailed security analysis and simulation results will be provided to illustrated their security properties and performances.

3.1 Introduction

As mention in Chapter II, there are three major techniques for the adversaries to get the source-location information in WSNs, which are *correlation-based source identification attack, routing traceback attack, reducing source space attack*, respectively.

Through correlation-based source identification attack, the adversaries can get the ID information contained in the messages content, and link the messages from the same source node together to deduce the source location. Although the message content can be encrypted using symmetric-key based algorithm between the source node and the SINK, according to our adversarial assumptions, it is still possible for the adversaries to get the ID information. We need a secure scheme to defend against the correlation-based source identification attack.

Both routing traceback attack and reducing source space attack are carried out through traffic pattern analysis. There are mainly two approaches that can protect the source-location privacy from traffic analysis attacks in WSNs, which are broadcast-based and routing-based, respectively.

For the broadcasting-based schemes [11,12,31–34], source-location privacy is provided through broadcasting or injection of dummy messages. The main idea is to make each node transmit messages consistently or at certain probabilistic pattern so that it is impossible for the adversaries to distinguish the real messages from the dummy messages. If there are no valid messages to transmit, the nodes have to transmit dummy messages to make the transmission pattern consistent all through the WSNs. This kind of schemes can provide source-location privacy protection against an adversary who is able to monitor the traffic information in the whole network area. However, if the adversary is able to monitor the entire network activities, then he/she can just detect any event happened in the networks directly without relying on the traffic pattern of the WSNs. In other words, it makes little sense to design a security protocol against such a 'powerful' adversary. In addition, due to the fact that, on average, transmission of one bit consumes about as much power as executing 800-1000 instructions [36], the transmission of dummy messages consumes significant amount of precious energy resources for WSNs. The networks collisions will also be increased because of the transmission of dummy messages. Therefore, these schemes are not quite suitable for large scale WSNs.

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

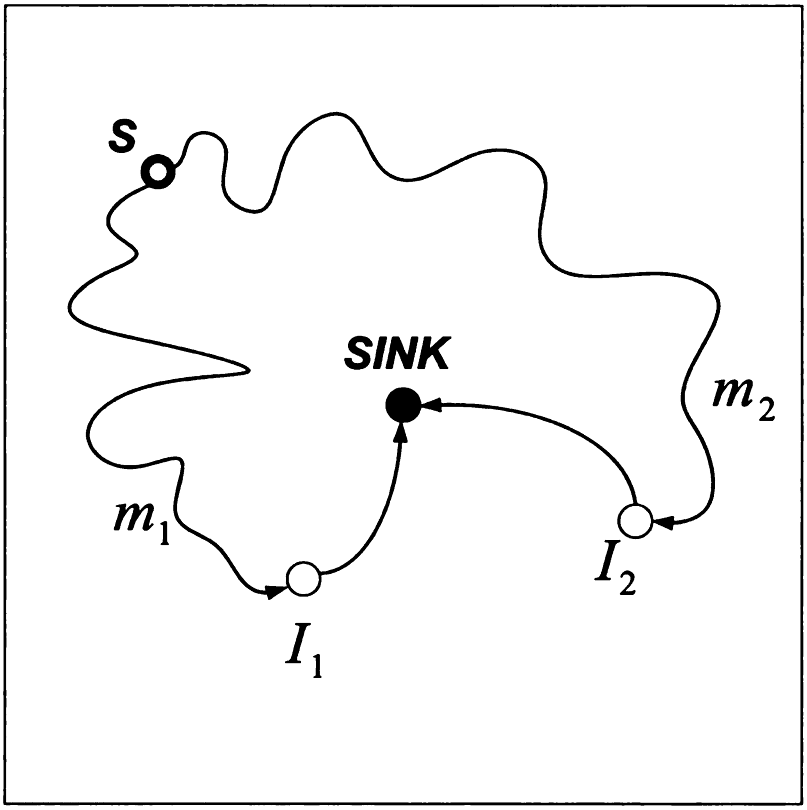


Figure 3.1. Nodes distribution through random routing

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.

As shown in Fig. 3.1, The source S generates two messages: m_1, m_2, which are first forwarded to I_1, I_2 respectively through two different random paths. Then, these two messages are forwarded to the SINK node by I_1 and I_2. In this way, when an adversary near the SINK node captures the messages, he/she will be led to the intermediate node instead of the real source node. If the messages are captured on the random paths, due to the randomness of the path formations, the adversary still cannot get any useful information about the source node.

How to select the random paths becomes the core factor for the success of this kind of schemes. If the next forwarder on the routing path is totally randomly selected, it is highly possible that the end nodes of the paths will be located within $h/5$ hops from the real source node after h hops transmission.

To solve this problem, several approaches have been proposed. In phantom routing protocol [13,14,21], 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. In section-based directed walk, 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. Then 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. In hop-based directed walk, for every forwarder on the random path, the messages are required to be forwarded to nodes which are further away from the real source node than the current forwarders.

Unfortunately, for section-based directed walk, once the message is captured on the random walk path, the adversaries will be able to get the direction information contained in the header of the message. The exposure of direction information decreases the complexity for adversaries to trace back to the actual message source in a magnitude of 2^h. For hop-based directed walk, every node in the network must know whether it is closer to a source node than its neighbors, which is obviously not very practical when the source nodes are dynamically changing.

In this chapter, we will first propose a dynamic ID assignment scheme in Section 3.2 to prevent the source-location information from being leaked in message content. Secondly, there will be four routing-based schemes introduced to prevent attacks through traffic analysis attacks. The first routing-based scheme, introduced in Section 3.3, routes each message to a randomly selected intermediate node before it is transmitted to the SINK node. We introduce three intermediate node selection methods. Section 3.4 describes the second routing-based scheme, in which a network mixing ring (NMR) is proposed to provide network-level source-location privacy. The third and the fourth routing-based schemes will be presented in Section

3.5 and Section 3.6, respectively. They protect source-location privacy through multiple intermediate nodes, which are selected based on angle and quadrant respectively. For each of these routing-based schemes, we provide detailed security analysis and simulation results. We then conclude in Section 3.7.

3.2 Dynamic ID Assignment

To prevent adversaries from getting any useful source-location information through *correlation-based source identification*, we propose a dynamic ID assignment scheme in this section.

In [13], each sensor node is assumed to have a unique ID that corresponds to a physical location. Only the SINK node can tell a node's location from its ID. The source node ID is directly included in the message packet. This ID also serves as the identifier of the encryption key shared between the grid and the SINK node. The problem of this design is that the adversaries could monitor the traffic pattern of the WSNs and link multiple packets from the same sensor node, which may help the adversaries to identify the source location since the IDs correspond to the grids' locations. Whenever the adversaries discover a message sent from a grid with an ID that they already know, they can easily get the source-location information.

To solve this problem, we propose to protect the source-location privacy through the application of dynamic IDs generated using an *ID-hash-chain*. At the beginning phase of the network distribution, each grid in the network is offered one initial ID. The SINK node will build an ID-hash-chain for each grid using this initial ID. As an example, the i^{th} grid is distributed with an initial ID: ID_i. The ID-hash-chain for grid i: $\{id_1^{(i)}, id_2^{(i)}, id_3^{(i)}, \cdots, id_{n-1}^{(i)}, id_n^{(i)}\}$, is generated as follows:

$$\begin{aligned}
id_1^{(i)} &= H(ID_i), \\
id_2^{(i)} &= H(id_1^{(i)}), \\
id_3^{(i)} &= H(id_2^{(i)}), \\
& \ldots\ldots \\
id_n^{(i)} &= H(id_{n-1}^{(i)}),
\end{aligned}$$

where H is a one-way hash function, and n is a preselected system parameter.

After each grid has received its ID-hash-chain, it will use the IDs in reverse order. In other words, the i^{th} grid should send id_n together with its first message, id_{n-1} together with its second message, and id_{n-j} with its $(j+1)^{th}$ message, etc. Because H is a one-way hash function, it is computationally infeasible to compute id_j from id_{j+1}. As a result, the adversaries cannot link multiple packets generated from the same grid even if they can get the current ID of this grid. Only the SINK, which has the full knowledge of the ID-hash-chain, can correlate the ID of the corresponding source grid with the source location. In this way, transmitting dynamic IDs with messages will not leak source-location information to adversaries.

Theoretically, the length of the ID-hash-chain could be the full length of the hash function output. However, because of the limited available resource of sensor nodes, this length should be as short as possible in practical applications. The dynamic ID is implemented as a tradeoff between security and memory resource. As long as the delay for a previously used ID to be reused is long enough, then we can make it very difficult for adversaries to perform source-location analysis based on sensor node IDs. As an example, if the active rate of each sensor node is 10%, each sensor node reports an event every 100 seconds, and the ID-hash-chain recycles on 1000 dynamic IDs, then it takes about

$$\frac{1000 \times 100}{10\%}\ seconds = 277.8\ hours = 11.6\ days,$$

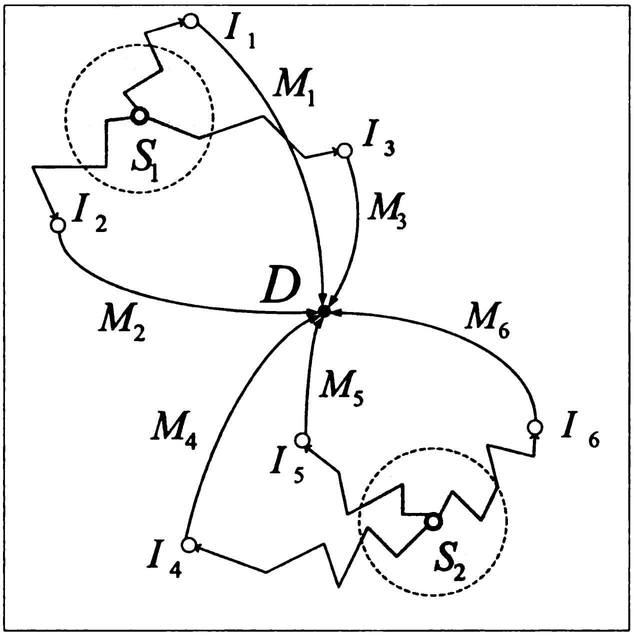


Figure 3.2. Illustration of RSIN

for an ID to be reused. This should be long enough to protect the ID anonymity.

3.3 Source-Location Privacy Protection 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).

3.3.1 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 locations of the sensor nodes. This design will make it difficult for adversaries to get 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, a randomly selected intermediate node may not even exist. However, the knowledge of relative location guarantees that message packets 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 an 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 (x_0, y_0). 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 $d_{rand} \geq d_{min}$.

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:

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

The probability [103] 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 [104].

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 data messages, intermediate nodes forward the messages to the SINK node.

In the explanatory example given in Fig. 3.2, 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

Theorem 3.1 *In the first routing phase, the constrained RSIN scheme can leak direction information to the adversaries. In its second routing phase, local location privacy instead of global location privacy can be achieved.*

Proof: In constrained RSIN, the routing scheme can be divided into two phases. In the first phase, messages are forwarded to randomly selected intermediate nodes. In the second phase, messages are forwarded to the SINK node by intermediate nodes.

(a) For the first phase, if an adversary gets a message on a path from a source node to an intermediate node, he/she will be able to move one hop closer to the source node. However, because the routing path is not fixed, the possibility that the adversary can receive another message sent from the same source node is very small, especially for a large scale network. In other words, the amount of source-location information that one message on the random path can leak to the adversary tends to be zero, i.e.:

$$SDI_1 \simeq 0.$$

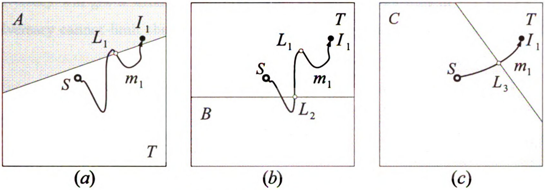


Figure 3.3. NSSI calculation for RSIN

An example is shown in Fig. 3.2, 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 from I_2, the probability for the adversary to receive M_3 is close to zero according to our network assumption. 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, messages in our protocol does not contain any side information. Since intermediate nodes are determined before each data message is sent out by the source, data messages carry no observable side information of the message source's location in its content due to message content encryption. Therefore, our proposed protocol leaks no side information even if the message is corrupted.

We will illustrate our description using an example shown in Fig. 3.3, which is similar to the one used in Section 2.4.2. In this figure, T is the target network area, S and I_1 are the real source and an intermediate node respectively. If message m_1 is captured at L_1 by an adversary as shown in Fig. 3.3.(a), which is located on a random path from S to I_1, then the adversary will assume that S is located in the shaded area A. If the same message m_1 is captured at L_2 in Fig. 3.3.(b), then the adversary will conclude that S is located in the shaded area B. In both cases, the adversary will get a wrong conclusion. This example shows that in our scheme an adversary cannot limit the source node to a specific group of nodes. In other words,

$$SSI_1 = area(T),$$

$$NSSI_1 = 1.$$

However, what we have discussed is just the ideal case for constrained RSIN. In fact, in constrained RSIN, instead of routing through random walking path, messages are routed to intermediate nodes directly. Forwarding paths from source node to intermediate nodes tends to be more straight compared to random walking path in phantom routing scheme. For example, a routing path from S to I_1 is more likely to be the path in Fig. 3.3.(c) rather than the paths in Fig. 3.3.(a) and Fig. 3.3.(b). If message m_1 is captured at L_3, the adversary will assume that it is highly possible that S is located in the shaded area C, which means:

$$NSSI_1 = \frac{area(C)}{area(T)}.$$

We can come to a conclusion that although in some scenarios the first phase routing in constrained RSIN has satisfying security property, in most of the cases, it still suffers the security flaw that is similar to the directed walk phase in phantom routing.

(b) Now, we will analyze the security property of the second phase of constrained RSIN. In constrained RSIN, the possibility for any node to be selected as an intermediate node is proportional to the distance between this node and the source node. If the target area of WSNs is not very large, messages forwarded to the SINK node tend to come from all possible directions. The adversary will not be able to get any useful information about where the source node is located or move closer to the source node. Therefore, we have:

$$NSSI_2 = 1,$$

for small scale networks.

However, for large scale WSNs, the second phase routing has the same security flaw

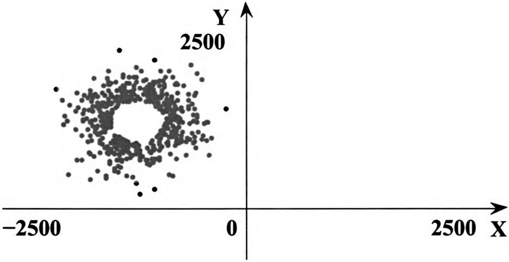


Figure 3.4. Intermediate nodes distribution for constrained RSIN scheme

as the second phase of phantom routing scheme, i.e, the intermediate nodes tend to be located 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. We will illustrate this situation through Fig. 3.4. In this example, the x-coordinate range and the y-coordinate range of the target area are both $[-2500, 2500]$. The source node is located at $(-1250, 1250)$. d_{min} equals to 250. 500 intermediate nodes are selected according to the constrained RSIN scheme with σ equal to 1. The result shows that all intermediate nodes are distributed around the source node. When an adversary around the SINK node, which is located at $(0, 0)$, collects several messages from this source, he/she will assume that the source is located in the second quadrant, i.e,:

$$NSSI_2 = \frac{2500 \times 2500}{5000 \times 5000} = \frac{1}{4}.$$

To summarize, for constrained RSIN, the first phase cannot guarantee provable absolute security level. The second phase can provide good security level for small scale networks, but not for large scale networks.

For WSNs with fixed SINK nodes, an adversary is more likely to stay around the

SINK node waiting to capture transmitted messages, because the area around the SINK has higher possibility to detect the transmitted messages. In other words, in WSNs, the security property of the second routing phase is more important. From this point of view, the phantom routing scheme and constrained RSIN can only guarantee local location privacy instead of global location privacy or network-level location privacy. □

3.3.2 Totally Random RSIN Scheme

As we have analyzed, constrained RSIN can only provide local location privacy because intermediate nodes tend to be located around the source node. In order to provide global location privacy over WSNs, 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 WSNs. Nevertheless, to prevent this from happening, in totally random RSIN scheme, the intermediate nodes is required to be at least d_{min} away from the real source node.

Security Analysis

Theorem 3.2 *The totally random RSIN scheme can achieve global location privacy in its second routing phase.*

Proof: Similar to constrained RSIN, totally random RSIN scheme can also be divided into two phases. The security property of the first phase is almost the same with the first phase of constrained RSIN.

Because the intermediate nodes randomly selected are evenly distributed in the WSNs. Every node in the network has the same possibility to act as an intermediate node. In the second phase of totally random RSIN scheme, messages can be forwarded

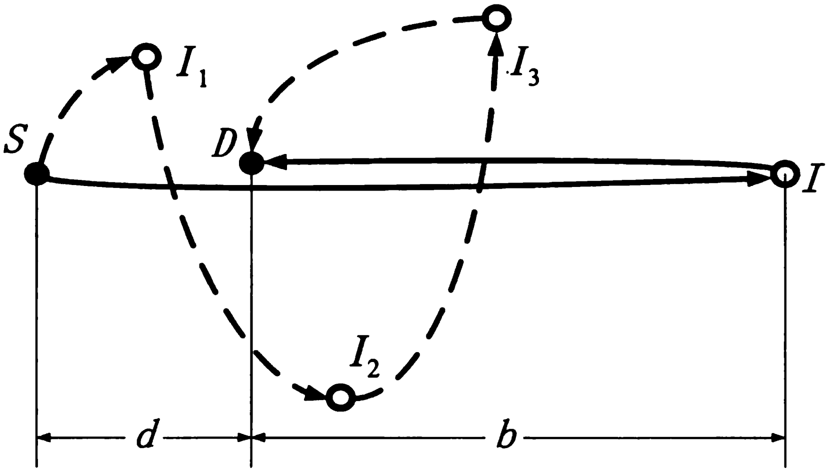


Figure 3.5. Message forwarding through intermediate node(s)

to the SINK node from all possible directions. Even if the location of one intermediate node is successfully identified, the source node is still at least d_{min} distance away. So for the second phase of totally random RSIN, we have:

$$SDI_2 \simeq 0,$$

$$NSSI_2 = 1,$$

□

Obviously, totally random RSIN scheme is more suitable for large scale networks than constrained RSIN from the security point of view. Global or network-level source-location privacy can be achieved in the second routing phase.

3.3.3 Ring-Band RSIN Scheme

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. 3.5, S, D, I are the source node, the SINK node and intermediate node, respectively.

The distance between S, D and I, D are d and b, respectively. 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 the increasing of the routing length.

- The first phase of the scheme might be too long. As we have analyzed in Section 3.3.1, the routing phase from source to an intermediate node may allow adversaries to deduce information of the source location. If the routing path in this phase is too long, the possibility that an adversary will receive this message will increase dramatically. This will weaken the security property of the whole scheme.

In this subsection, we propose another intermediate node selection method: ring-band RSIN. In this scheme, the random intermediate node would be located in a pre-determine region around the SINK node. We call this region a ring-band around the SINK.

The goal of this scheme is to provide global source-location privacy with adequate energy-efficient routing. The ring-band area would be a large area and at least a minimum radius distance, r, from the SINK node to provide global privacy. Also, the ring-band area would guarantee that the intermediate node is not be too far away from the SINK node to limit the energy consumption in the routing phase. This routing scheme would give the illusion that the source node is sending messages to the SINK node from all possible directions. In this way, the ring-band RSIN creates an effect that is similar to the totally random RSIN scheme but with less energy consumption.

We assume that each node has knowledge of the perimeters that is shown in Fig. 3.6. The description of the perimeters is as follows:

- x_0, y_0: The corresponding X and Y coordinates of the SINK node,

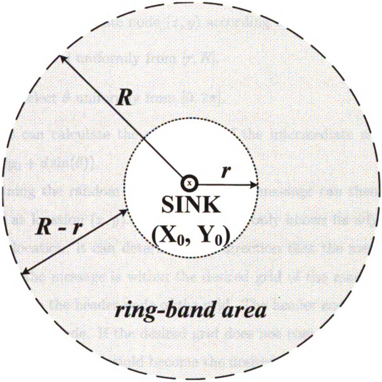


Figure 3.6. Distribution of the ring-band area

- R: The pre-determined radius from the SINK to the outer-edge of the ring-band area,
- r: The pre-determined radius from the SINK node to the inner-edge of the ring-band area.

From these perimeters, $\{x_0, y_0, R, r\}$, the source nodes are able to randomly select intermediate nodes within the ring-band area for each of the messages. Since we assume that the SINK node is located at the relative location (x_0, y_0), the source node selects the random intermediate node (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. An illustration for ring-band RSIN is shown in Fig. 3.7.

Security Analysis

Theorem 3.3 *The ring-band RSIN scheme can achieve global location privacy in its second routing phase.*

Proof: Because the intermediate nodes randomly selected are evenly distributed in the ring-band around the SINK, in the second phase of ring-band RSIN

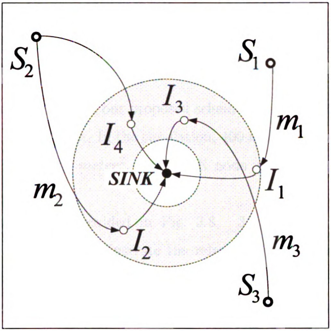


Figure 3.7. Routing illustration of the ring-band RSIN protocol

scheme, messages are forwarded to the SINK node from all possible directions with equal possibilities. For the second phase of ring-band RSIN, we have:

$$SDI_2 \simeq 0,$$

$$NSSI_2 = 1,$$

□

Both totally random RSIN and ring-band RSIN can provide network-level source-location privacy in their second routing phase.

Each of these two schemes has its own advantages. For ring-band RSIN, the ring-band around the SINK is a fixed infrastructure, which facilitates adversaries to carry out attacks and eavesdropping. However, in ring-band RSIN, the path lengths from the source node to intermediate nodes are largely reduced, which can provide better source-location privacy protection in the first routing phase than the totally random

RSIN.

3.3.4 Simulation Results and Performance Comparison

To evaluate the performance of our proposed schemes, we conduct simulations using ns-2 on RedHat Linux system. In the simulation, 400 nodes are evenly distributed in an area of size 3360 × 3360 $meter^2$. The SINK node is located at the center of the networks.

Simulation results are provided in Fig. 3.8, 3.9, 3.10, 3.11, 3.12 where Fig. 3.8.(a), 3.9.(a), 3.10.(a) illustrate the relationship between performance and the packet lengths, Fig. 3.8.(b), Fig. 3.9.(b), Fig. 3.10.(b) show performance with different packet generation intervals, Fig. 3.11, 3.12 show performance with different lengths of random path.

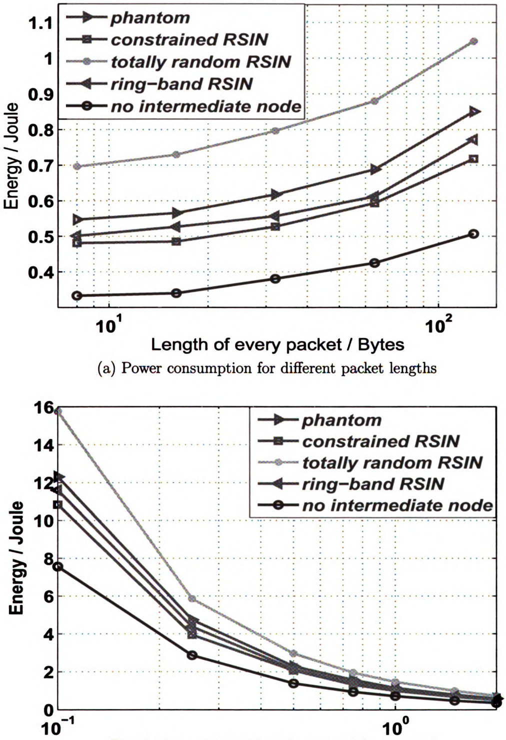
For simulation results in Fig. 3.8, 3.9, 3.10, we set $d_{min} = 480$ meters for totally random RSIN scheme. The simulation shows that after four hops, the average distance between the phantom source node and the real source node for phantom routing is 526.12 meters. While for constrained RSIN scheme, the average distance between the intermediate node and the source is set to be 529.14 meters. For ring-band RSIN, the inner-radius of the band is 480 meters, while the outer-radius is 640 meters.

For simulation results in Fig. 3.11, 3.12, $R1$, $R2$, $R3$ 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 the following findings:

- While direct routing without intermediate node provides the least source-location privacy, it has the best performance;

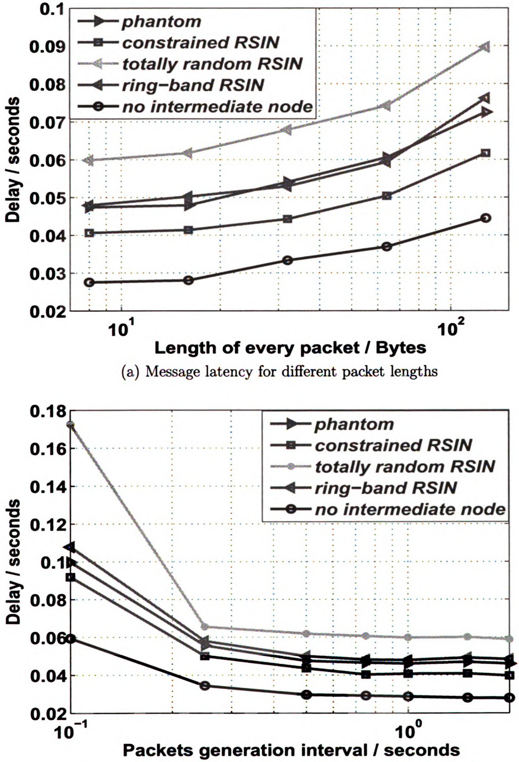
- Constrained RSIN scheme provides comparable source-location privacy protection than phantom routing, while constrained RSIN has better communication performance. This is because the path from the real source node to phantom



(a) Power consumption for different packet lengths

(b) Power consumption for different packet generation intervals

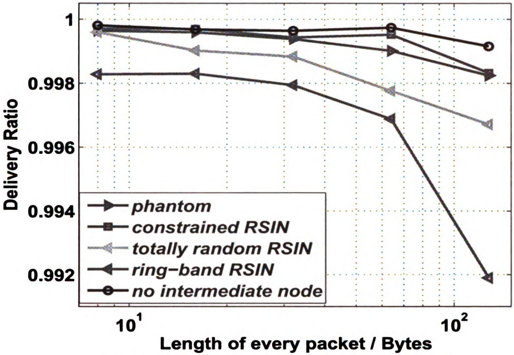
Figure 3.8. Performance of routing by single-intermediate node



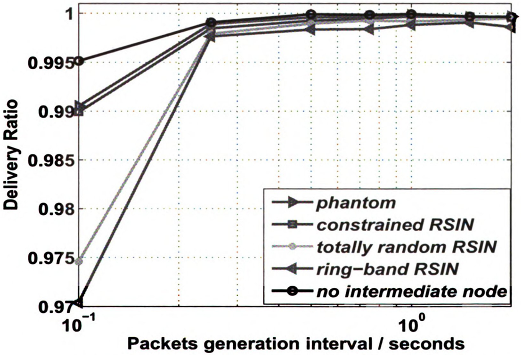
(a) Message latency for different packet lengths

(b) Message latency for different packet generation intervals

Figure 3.9. Performance of routing by single-intermediate node

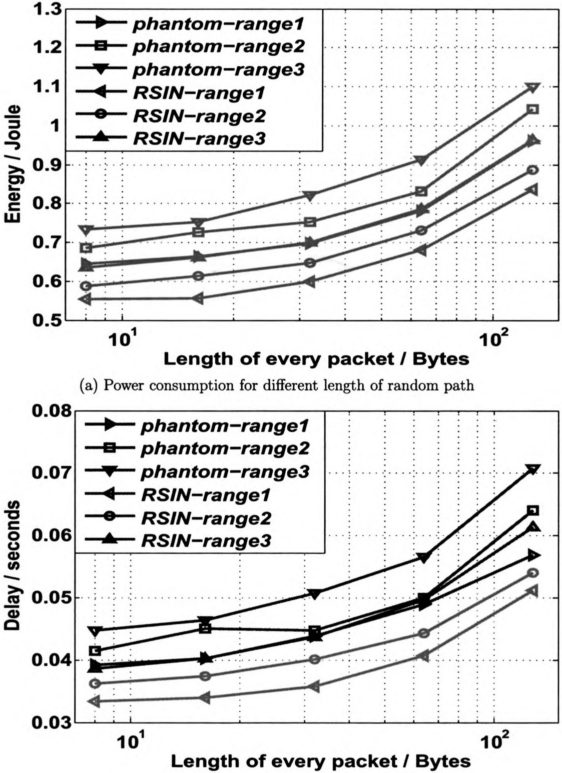


(a) Message delivery ratio for different packet lengths



(b) Message delivery ratio for different packet generation intervals

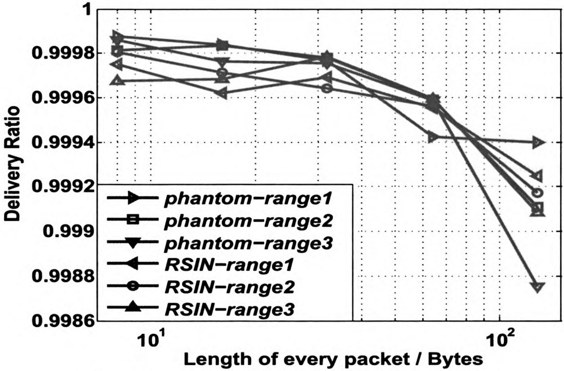
Figure 3.10. Performance of routing by single-intermediate node



(a) Power consumption for different length of random path

(b) Message latency for different length of random path

Figure 3.11. Performance of routing by single-intermediate node



Message delivery ratio for different length of random path

Figure 3.12. Performance of routing by single-intermediate node

sources in phantom routing is more curved than the path from the source to intermediate nodes in constrained RSIN. The average length of routing paths in phantom routing is longer than constrained RSIN when similar security properties are guaranteed;

- The performance of the totally random RSIN is worse than the constrained RSIN;

- For energy consumption and transmission delay, the ring-band RSIN is better than the totally random RSIN, but worse than the constrained RSIN. However, ring-band RSIN has the lowest delivery ratio than the other two methods. The reason for this lies in the fact that for ring-band RSIN, all the messages transmitted in the network have to be delivered to the ring-band area, while in the constrained RSIN and the totally random RSIN, the traffic pattern are evenly distributed in the networks. Therefore, there will be more message collisions in the ring-band area, which will decrease the delivery ratio.

- The performance of constrained RSIN and phantom routing is inversely proportional to the average distance between the source node and intermediate nodes or phantom sources.

- The performance of ring-band RSIN is inversely proportional to the average distance between the SINK and intermediates nodes.

- When the average distance between the SINK and the intermediate nodes is fixed, the performance of ring-band RSIN is proportional to the width of the ring-band.

3.4 Source-Location Privacy Protection with Network Mixing Ring

In the previous section, we have proposed three source-location schemes through routing to a single intermediate node. However, each of these schemes suffers either

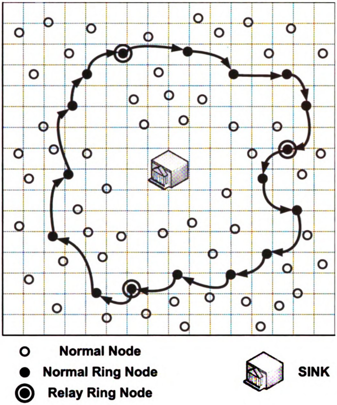


Figure 3.13. Grids Formation

security flaws or performance downgrading. To address these problems, we propose three advanced schemes, which will be presented in Section 3.4, Section 3.5 and Section 3.6, respectively.

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 network-level source-location privacy. The last phase forwards the message to the SINK node.

The networks formation for this scheme is illustrated in Fig. 3.13. 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*. Messages that are transmitted in the mixing

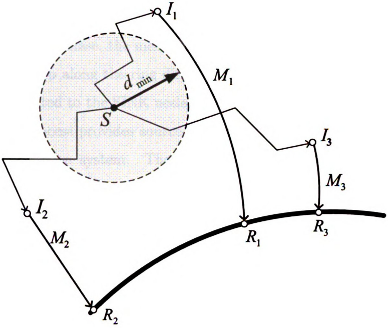


Figure 3.14. Illustrate of the first two phases routing

ring are referred to as *vehicle messages*. Vehicle messages will be transmitted in the mixing ring in 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 to as *data messages*.

3.4.1 Constrained RSIN

In this phase, messages will be forwarded to an intermediate node in the same way as the constrained RSIN introduced in Section 3.3. Then messages will be forwarded to the nearest ring node by this intermediate node.

An example is given in Fig. 3.14, 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.

3.4.2 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. 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. Normal ring nodes can store data messages received from normal nodes. 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 such a unit as *dummy unit*, composed of any fixed data structure, such as all 0s. The length of a unit is the same as a data message sent by any 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. 3.15, ring node B shares a key K_{AB} with ring node A and a key K_{BC} with ring node C. When

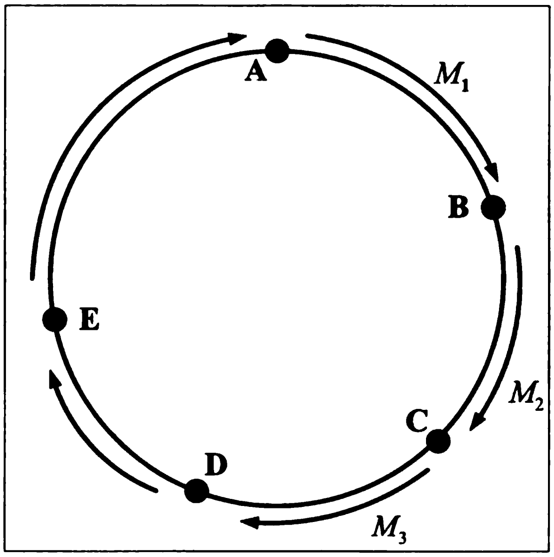


Figure 3.15. Message transmission in the ring

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

$$M_2 = E_{K_{BC}}(\{D_{K_{AB}}(M_1)\}). \tag{3.1}$$

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

3.4.3 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 the next relay ring node is reached.

3.4.4 Security Analysis

Theorem 3.4 *The NMR based source-location privacy protection scheme can achieve security level:*

$$SDI = 0,$$

$$NSSI = 1.$$

Proof: The mixing ring based source-location privacy scheme is composed of three phases: (a) Constrained RSIN; (b) Routing on NMR; (c) Routing to the SINK node. We will analyze the security property of this scheme through analyzing each of these three routing phases.

(a) *Constrained RSIN*: The security property of constrained RSIN has been discussed in Section 3.3.1. As we have analyzed, constrained RSIN can provide local location privacy, i,e., for a network which is not too large, constrained RSIN can achieve good security level, i.e,:

$$SDI_1 \simeq 0,$$

$$NSSI_1 = 1.$$

For large scale WSNs, the area for constrained RSIN to take place is greatly reduced because each source node just needs to forward its message to an intermediate node nearby. In other words, constrained RSIN can provide satisfying security property in the first phase.

As shown in Fig. 3.14, 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, data messages received from one ring node could also be transmitted from many different source nodes in the network.

(b) *Routing on NMR*: 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 messages a ring node received and transmitted, adversaries can determine whether a data message has been loaded into the vehicle message by this ring node or not. 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. In this way, messages across the network are totally mixed up. As shown in Fig. 3.14, 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, on receiving a message on the NMR, an adversary cannot get any useful information about the source location, i.e,:

$$SDI_2 = 0,$$

$$NSSI_2 = 1.$$

(c) *Routing to SINK node*: In this phase, relay ring nodes forward messages to the SINK node. After the first phase and the second phase, locations of relay ring nodes have already no correlation with locations of source nodes. Capturing a message in the third phase would leak no information to adversaries, which means:

$$SDI_3 = 0,$$

$$NSSI_3 = 1.$$

To summarize, the mixing ring based source-location privacy protection scheme could provide provable satisfying security level in each of its three phases. When these three phases are combined to protect source-location privacy, the following security

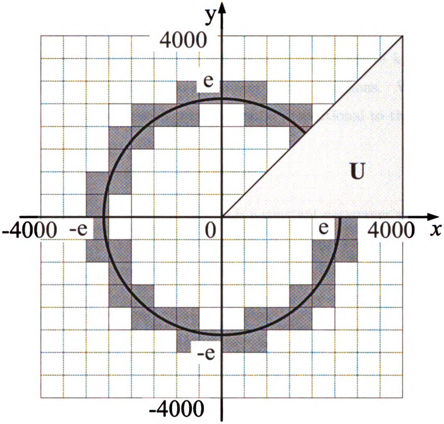


Figure 3.16. Ring selection in simulation setup

properties can be achieved:

$$SDI = SDI_1 \times SDI_2 \times SDI_3 = 0,$$

$$NSSI = NSSI_1 \times NSSI_2 \times NSSI_3 = 1.$$

□

3.4.5 Performance Analysis and Simulation Results

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

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

where $\mathcal{E}$ denotes the energy consumption, α is a constant parameter and d is distance of the transmission. Fig. 3.16 gives an example of a target area of size 8000 × 8000 $meter^2$. 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:

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

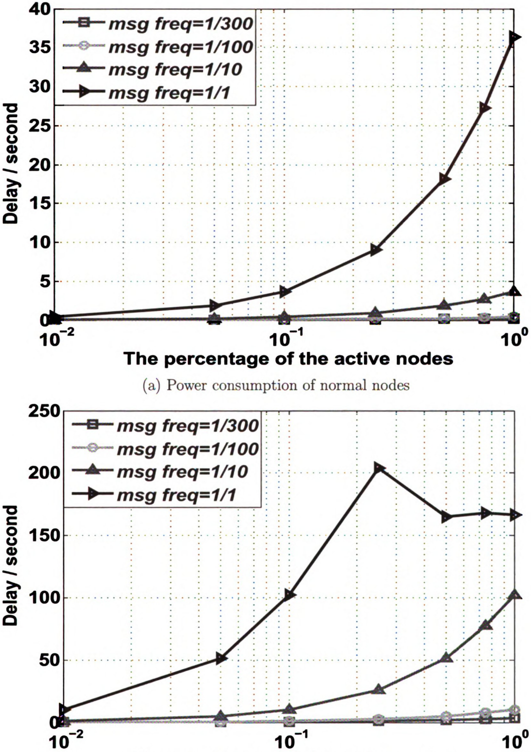
where $\mathcal{E}_U$ is the energy consumption for area U as demonstrated in Fig. 3.16. It can be calculated that when $e = 3061$, the overall power consumption $\mathcal{E}_{total}$ achieves the minimum. In this way, we get the optimal ring location.

Lemma 3.1 *Suppose the target area is of dimension $A \times A$ $meters^2$, the radius of the NMR, which is located in the center of the target area, is R meters. Then when:*

$$R : A = 0.76525 : 1,$$

the overall power consumption for each sensor in this area to transmit one message to a ring node is minimized.

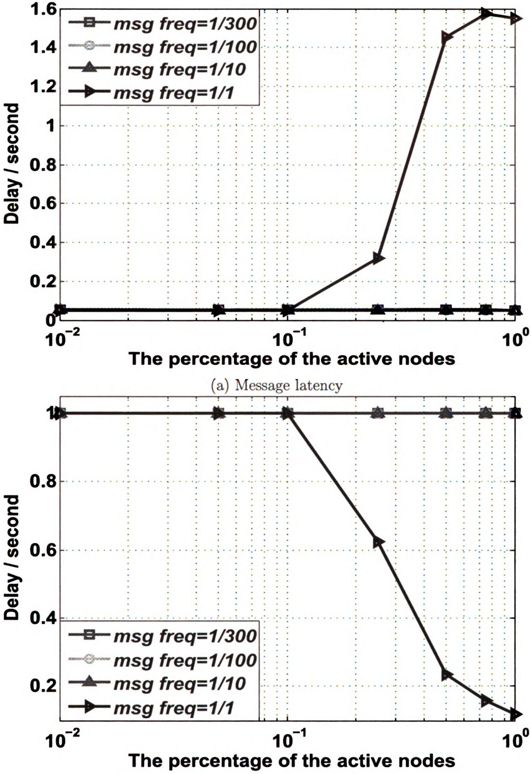
In practical application, for large sensor network, usually only a small fraction of the sensor nodes in the network has events to report. We name these nodes as



(a) Power consumption of normal nodes

(b) Power consumption of ring nodes

Figure 3.17. Performance of the proposed routing and encryption scheme



(a) Message latency

(b) Message delivery ratio

Figure 3.18. Performance of the proposed routing and encryption scheme

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 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 relay ring nodes on the mixing ring can generate vehicle messages. If there are n relay ring nodes on the mixing 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. 3.17, 3.18 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 ns-2 on Linux system. In the simulation, the target area is a square field of size 8000 × 8000 $meter^2$. 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 each 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. 3.17.(a) and (b), we can see that ring nodes consume more energy than normal nodes. To solve this problem, nodes in ring grids can take turns to be ring nodes. It is also noticed that delivery ratio drops exponentially when traffic volume increases. It is primarily because of traffic collisions and packet losses caused by 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 to be relatively low. In other words, traffic volume may be low. In this scenario, we can ensure almost 100% delivery ratio, as shown in Fig. 3.18.(b). The simulation results demonstrate that the proposed scheme is very efficient and can be used for many practical applications.

3.5 Source-Location Privacy Protection Through Angle-Based Multi-Intermediate Nodes

In this section and the next section, we propose routing through multiple randomly selected intermediate nodes for large scale WSNs.

The intermediate nodes are preselected before a message is sent out from the source node. If information of the intermediate nodes is contained in the header of messages, adversaries can get information of all the intermediate nodes from a captured packet, and the routing path formed by these intermediate nodes. To solve this problem, we assume that information of the previous intermediate node(s) will be deleted from the message header before a message is forwarded. In this way, no information of previous intermediate node(s) can be obtained from a message header.

3.5.1 Angle-based Multi-Intermediate Nodes Selection

In *angle-based* intermediate nodes selection, prior to each message transmission, the source node needs to determine a maximum angle β between the last intermediate node and the source node with the SINK node as the vertex, where $\beta \in [0°, 180°]$. After β is determined, the source node chooses an actual angle θ between the last intermediate node and itself with the SINK node D as the vertex, where θ is a random

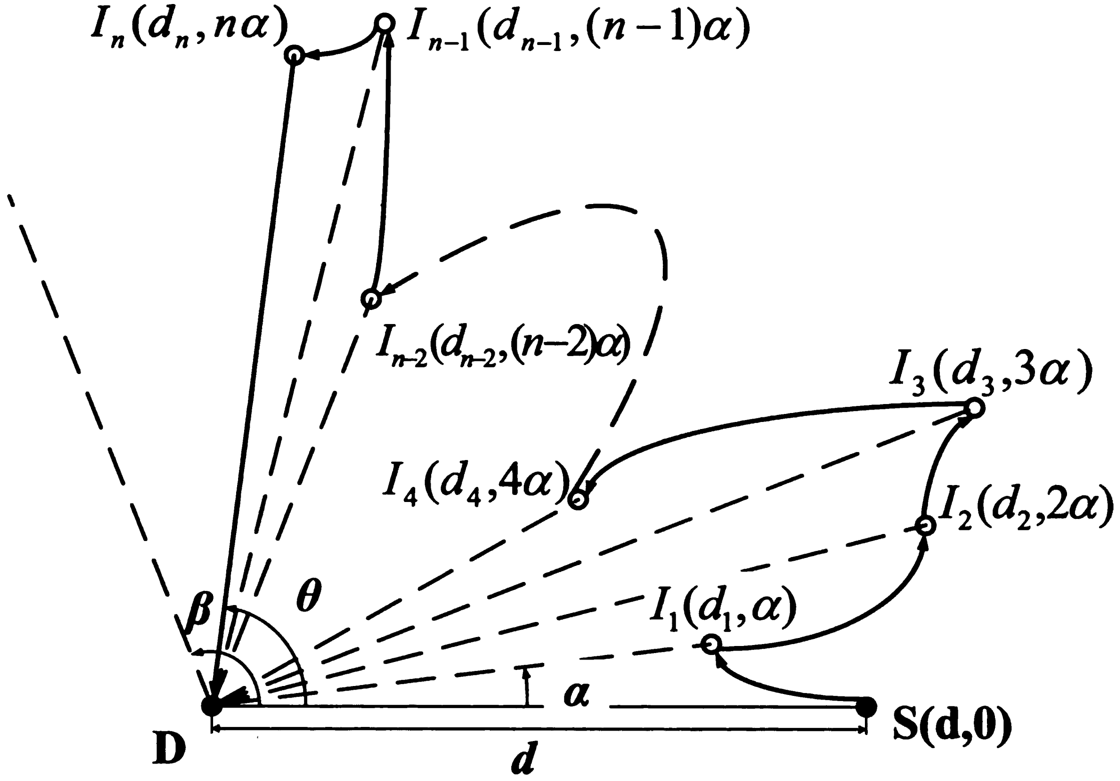


Figure 3.19. Angle-based intermediate nodes selection

variable evenly distributed in range $(-\beta, \beta)$. Then the source node needs to determine the number of intermediate nodes, say n.

The angle generated by one intermediate node should be: $\alpha = \theta/n$. The angles between all intermediate nodes and the source node with the SINK node as the vertex are: $\alpha, 2\alpha, 3\alpha, \cdots, n\alpha$, respectively, where $n\alpha = \theta$ is the angle between the last intermediate node and the source node.

After all the angles are determined, the source node generates distances between the destination node and the n intermediate nodes: $d_1, d_2, d_3, \cdots, d_n$, where d_i $(i = 1, 2, \cdots, n)$ is a random variable evenly distributed in range $(0, R)$, and R is the radius of the network.

Assume a polar coordinate system is built on the network, the SINK node and the source node are located at the origin and $(d, 0°)$, respectively, and d is the distance between the source node and the SINK, then the locations for all the intermediate nodes will be: $(d_1, \alpha), (d_2, 2\alpha), (d_3, 3\alpha), \cdots, (d_n, n\alpha)$.

Fig. 3.19 illustrates this intermediate nodes selection, in which S is the source node, $I_1, \cdots, I_n$ are the intermediate nodes and D is the SINK node.

After all the intermediate nodes' locations are determined, the source node will forward the message to the first intermediate node, I_1, with all the intermediate nodes' information contained in the message header.

When an intermediate node, say I_i, receives a message, it will first replace the information about itself in the message header with some dummy information. If I_i is not the last intermediate node, then it will forward this message to the next forwarder I_{i+1} as indicated in the message header. If I_i is the last intermediate node, then it will forward this message to the SINK node.

3.5.2 Security Analysis

Theorem 3.5 *The source-location privacy protection scheme through angle-based multi-intermediate nodes selection can achieve security level:*

$$SDI = 0,$$

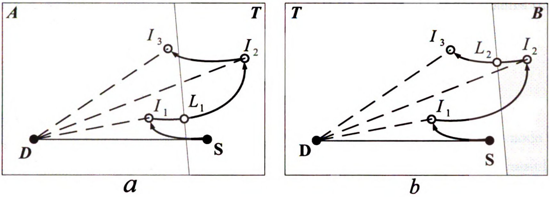


Figure 3.20. NSSI calculation for angle-based intermediate nodes selection

$$NSSI = \frac{\beta}{180}.$$

Proof: The source-location privacy protection scheme through angle-based multi-intermediate nodes selection consists of two routing phases. We will analyze the security property of these two phases separately.

(a) In the first phase, messages are forwarded by multiple intermediate nodes before sent to the SINK node. The routing paths are different for different messages generated by one source node. Capturing one message will not help the adversary to move closer to the source node, because the possibility that a same routing path is used repeatedly is virtually zero for large scale WSNs. That is,

$$SDI_1 \simeq 0.$$

In addition, the routing direction in the first phase will be changed at each intermediate node. Message transmission direction carries no information about the source-location information.

As shown in Fig. 3.20, S, D are the source node and the SINK node respectively. I_1, I_2, I_3 are three intermediate nodes and I_3 is the last intermediate node. We can see that the transmission direction changes completely when the message is forwarded by each of these three intermediate nodes. If a message is captured at L_1, based on the transmission direction, the adversary will assume that S is located in the shaded area A, which is wrong. The random selection of the intermediate nodes ensures that no source-location information can be leaked to the adversaries, i.e.,

$$NSSI_1 = 1.$$

(b) In the second phase, messages are forwarded from the last intermediate node to the SINK node. We will analyze that even if the adversaries are able to successfully identify the location of the last intermediate node I_n, determination of the source location S is still very difficult according to our assumption.

After the first phase, an adversary cannot move closer to the source node based on a captured message, which means:

$$SDI_2 = 0$$

in the second phase. In the case that the location of I_n is known, a polar coordinate system is built on the networks with D located at the origin and I_n at $(d, 0^\circ)$, where d is the distance from D to I_n. The possible location of S is in the shaded area shown in Fig. 3.21, i.e, the radian measure range of $(-\beta, \beta)$, where β is a configurable parameter ranging from 0° to 180°. Therefore:

$$NSSI_2 = \frac{\beta}{180}.$$

To summarize, the security properties of the source-location privacy protection scheme through angle-based multi-intermediate nodes selection are:

$$SDI = SDI_1 \times SDI_2 = 0,$$

$$NSSI = NSSI_1 \times NSSI_2 = \frac{\beta}{180}.$$

□

The larger β is, the better source-location privacy can be achieved. When β is

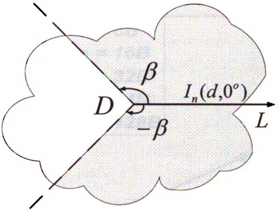


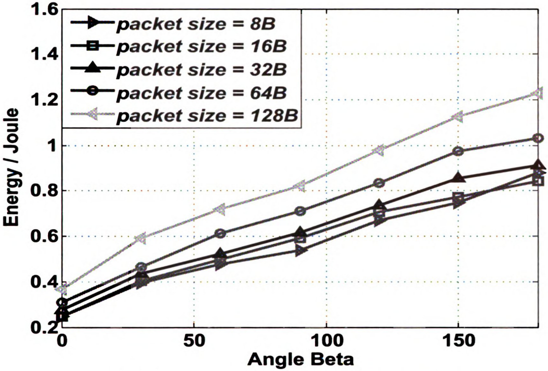
Figure 3.21. Possible source-location for angle-based scheme

dynamic and cannot be determined by adversaries, the source node S can possibly be located anywhere in the whole sensor domain. In other words, global source-location privacy is achieved.

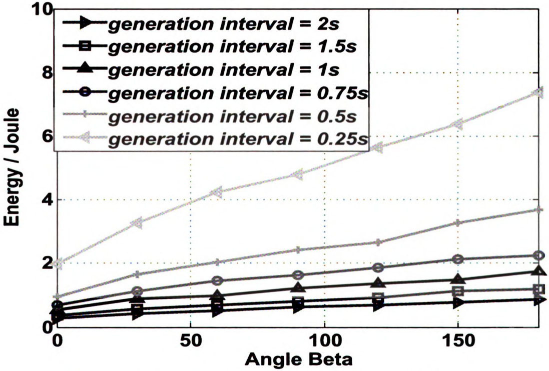
3.5.3 Simulation Results

We carry out simulations to evaluate performance of the angle-based multi-intermediate nodes selection scheme using ns-2 on RedHat Linux system. In the simulation, the target area is still 3360 × 3360 $meter^2$. The SINK node is located at the center of the networks. In this simulation, $\beta = 0$ means the messages are transmitted to the SINK node directly without relying on any intermediate nodes.

Simulation results are provided in Fig. 3.22, 3.23, 3.24 to demonstrate the tradeoff between the angle β and the performance, where Fig. 3.22.(a), 3.23.(a), 3.24.(a) show the performance of different packet sizes and β; and Fig. 3.22.(b), 3.23.(b), 3.24.(b) demonstrate the performance of different message transmission intervals and β. From these figures, we can see that as β increases, the performance decreases, however, the security level improves.

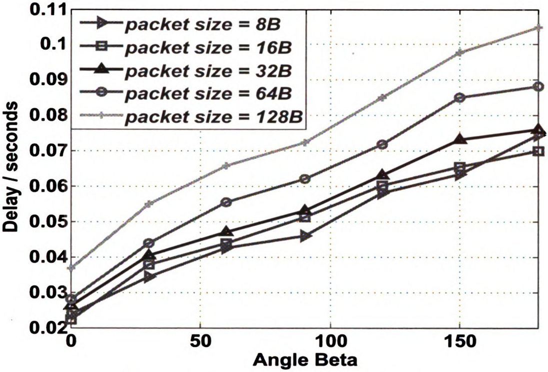


(a) Power consumption for different packet lengths

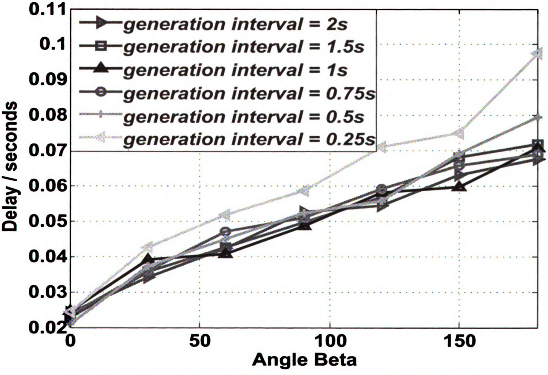


(b) Power consumption for different packet generation intervals

Figure 3.22. Performance of angle-based multi-intermediate nodes

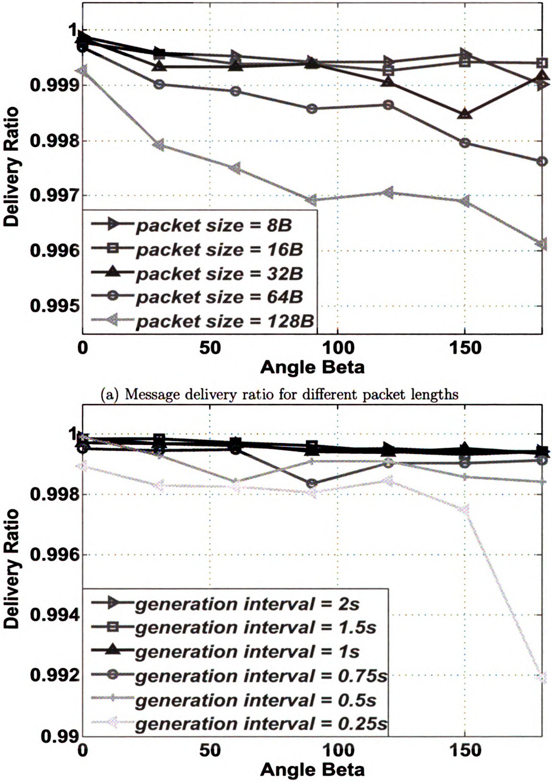


(a) Message latency for different packet lengths



(b) Message latency for different packet generation intervals

Figure 3.23. Performance of angle-based multi-intermediate nodes



(a) Message delivery ratio for different packet lengths

(b) Message delivery ratio for different packet generation intervals

Figure 3.24. Performance of angle-based multi-intermediate nodes

3.6 Source-Location Privacy Protection Through Quadrant-Based Multi-Intermediate Nodes

In this section, we present *quadrant-based* intermediate nodes selection scheme.

3.6.1 Quadrant-Based Multi-Intermediate Nodes Selection

In quadrant-based approach, the whole network is divided into four quadrants according to locations of the source node and the SINK node.

Prior to each message transmission, the source node has to form the quadrants. As shown in Fig. 3.25.(a), S, I, D are the source node, the last intermediate node and the SINK node, respectively. The distance between S and D is d. A reference frame is built on this network for source node S. The SINK node D is located at the origin with coordinate $(0,0)$. The source node S has coordinate (x_S, y_S), its location in *quadrant*1 is: $x_S = d \times cos(\alpha)$, $y_S = d \times sin(\alpha)$, where α is an evenly distributed random variable located in range of $(0°, 90°)$.

After the reference frame is built up, the source node needs to select an intermediate node in *quadrant*2 or *quadrant*4 in the reference frame as the last intermediate node. It is the shaded area in Fig. 3.25.(a). After the last intermediate node is determined, other intermediate nodes can be selected in a similar way as the angle-based multi-intermediate nodes selection scheme.

Once all the intermediate nodes are selected, the following procedure is similar to angle-based multi-intermediate nodes selection scheme.

3.6.2 Security Analysis

Theorem 3.6 *The source-location privacy protection scheme through quadrant-based multi-intermediate nodes selection can achieve security level:*

$$SDI = 0,$$

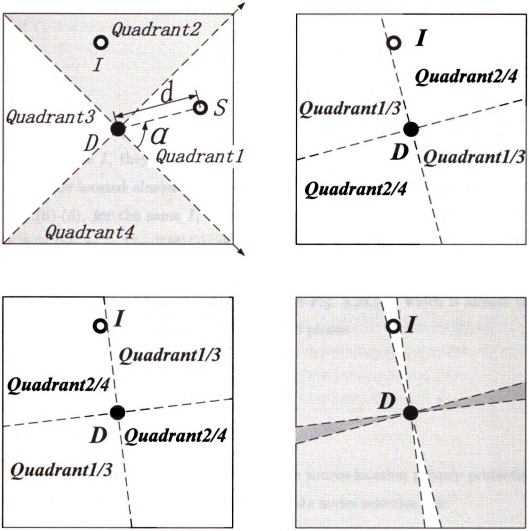


Figure 3.25. Quadrant-based intermediate nodes selection

$$NSSI = 1.$$

Proof: This scheme also consists of two routing phases.

(a) The first phase is similar to the first phase of angle-based intermediate nodes selection scheme. We have analyzed in Section 3.5.2 that in this phase:

$$SDI_1 \simeq 0,$$

$$NSSI_1 = 1.$$

(b) In the second phase, even if adversaries can determine location of the last intermediate node I, they still cannot get information of the source node location since S can be located almost anywhere in the sensor domain. As an example, in Fig. 3.25.(b)-(d), for the same I, based on the formation of the quadrants shown in Fig. 3.25.(b), or in Fig. 3.25.(c), the source node S can be located in either the shaded area in Fig. 3.25.(b), or the shaded area in Fig. 3.25.(c). Therefore, the possible location of the source node is the shaded area in Fig. 3.25.(d), which is almost the whole network area. In other word, in the second phase:

$$SDI_2 = 0,$$

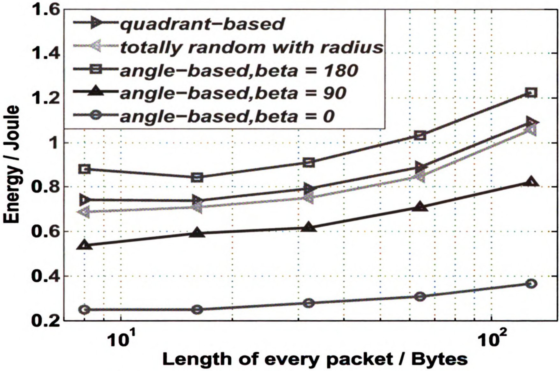
$$NSSI_2 = 1.$$

To summarize, the security properties of the source-location privacy protection scheme through quadrant-based multi-intermediate nodes selection are:

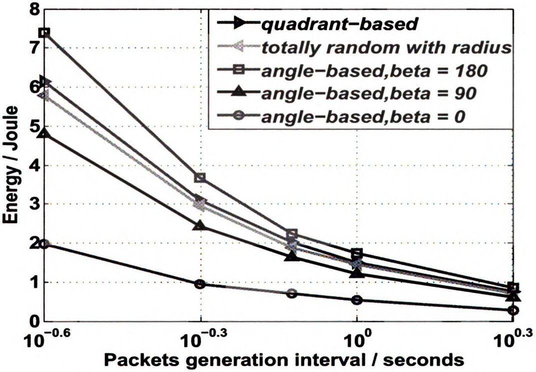
$$SDI = SDI_1 \times SDI_2 = 0,$$

$$NSSI = NSSI_1 \times NSSI_2 = 1.$$

□

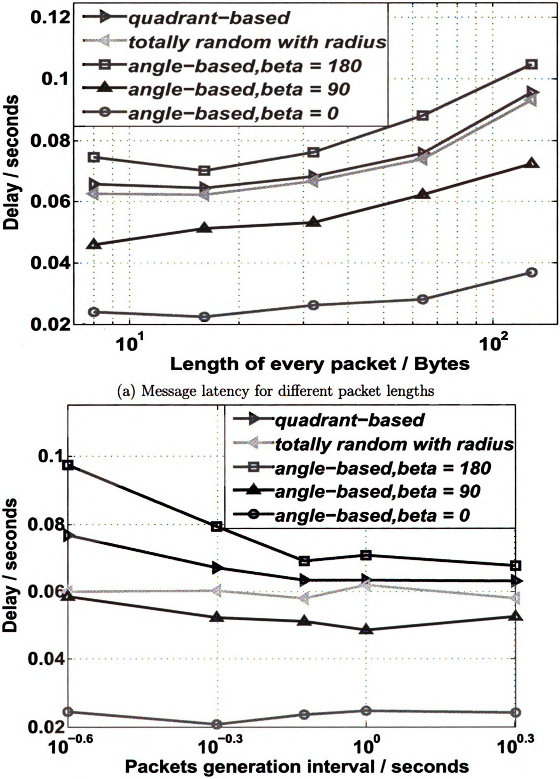


(a) Power consumption for different packet lengths



(b) Power consumption for different packet generation intervals

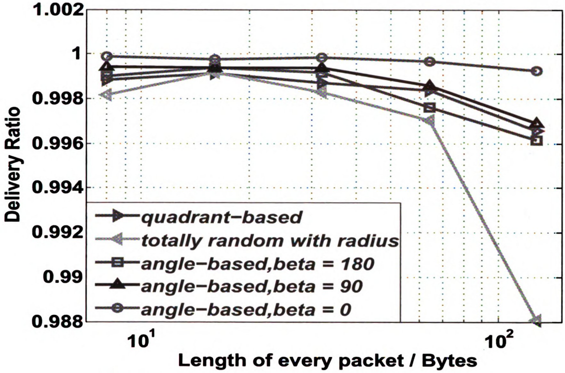
Figure 3.26. Performance of quadrant-based multi-intermediate nodes



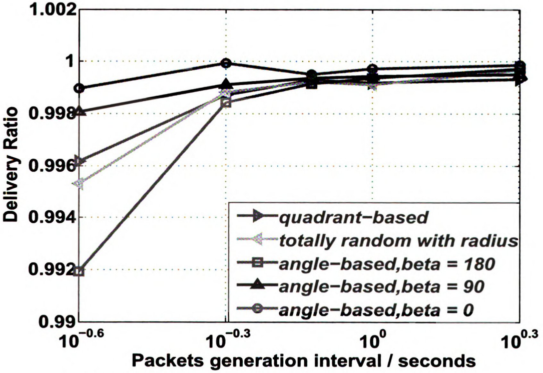
(a) Message latency for different packet lengths

(b) Message latency for different packet generation intervals

Figure 3.27. Performance of quadrant-based multi-intermediate nodes



(a) Message delivery ratio for different packet lengths



(b) Message delivery ratio for different packet generation intervals

Figure 3.28. Performance of quadrant-based multi-intermediate nodes

3.6.3 Simulation Results

We conduct simulations to compare performance of the quadrant-based intermediate nodes selection scheme and the angle-based intermediate nodes selection scheme. The setup for this simulation is same as the angle-based approach. The simulation results are shown in Fig. 3.26, 3.27, 3.28, where Fig. 3.26.(a), 3.27.(a), 3.28.(a) show the performance of multiple schemes with different packet lengths; Fig. 3.26.(b), 3.27.(b), 3.28.(b) demonstrate the performance of different message transmission intervals.

The simulation results illustrate that the quadrant-based approach is more efficient than the angle-based approach with β equal to 180°, while both of these two schemes can achieve excellent global source-location privacy.

3.7 Summary

In this chapter, we proposed four routing-based source-location privacy protection schemes. The first source-location privacy protection scheme routes each message to a randomly selected intermediate node before the message is transmitted to the SINK node. There are three intermediate node selection methods introduced: constrained RSIN method, totally random RSIN method, and ring-band RSIN method. Security analysis and simulation results demonstrate that each of them has its own advantages and drawbacks, which make it hard to say which method of these three is better than the other two. The second source-location privacy protection scheme is implemented through routing in a network-level mixing ring. This scheme consists of three routing phases. The first phase is constrained RSIN, which can provide local source-location privacy. The second phase route messages in a network mixing ring (NMR), which aims at network-level source-location privacy. In the last phase, messages are forwarded to the SINK node. Security analysis shows that this scheme can guarantee provable network-level source-location privacy protection. The third and the fourth schemes are based on routing through multiple intermediate nodes. The third scheme selects the intermediate nodes in a angle-based method, while the fourth scheme in a quadrant-based method.

CHAPTER 4

Design for Source Anonymous Message Authentication

This chapter is mainly about the development of an efficient anonymous source authentication protocol for WSNs. This protocol aims at providing hop-by-hop authentication without suffering the threshold problem. First, the terminologies *SAMA* and *MES* are introduced. Secondly, based on ECC, the authentication protocol is introduced in detail. Thirdly, in order to prevent source-information leaking, anonymity set (AS) selection method is presented. Fourthly, we discuss the key management scheme for the authentication protocol and how to detect the compromised nodes. In the last section, through security analysis and simulation results, we demonstrate that our design goals are achieved. Compared to traditional schemes, our protocol can provide comparable or even better computation and communication performance.

4.1 Introduction

Message authentication plays a key role in protecting source privacy, carrying out network-level administration work, and thwarting unauthorized and corrupted packets from being circulated in networks to save precious sensor energy. For this reason, many schemes have been proposed in literature to provide message authenticity and integrity in network communications [37–39, 45, 48]. These schemes can largely be divided into public-key based and symmetric-key based approaches.

For the public-key based approach [50–56], each message is transmitted along with the digital signature of the message generated using the sender's private key. Then, every intermediate forwarder and the final receiver can authenticate the message using the sender's public key [105, 106]. One of the limitations of the public-key based scheme is the high computational overhead. Some of the key distribution schemes [57–60] may solve some of these problems. However, in these schemes, the sensor nodes have to store a large number of extra keys which might not be used at all. This is a waste of the storage resources of the sensor nodes. What is more, compromising one sensor node might expose confidential information of other nodes to the attackers. They also cannot resist to a large number of node compromising. However, the recent progress [56] on elliptic curve cryptography (ECC) shows that the public-key schemes can be more advantageous in terms of memory usage, message complexity, and security resilience, since public-key based approaches have simple and clean key management.

In the symmetric-key based approaches, the shared key is used by the sender to generate a message authentication code (MAC) for each transmitted message. The receiver can verify the authenticity and integrity of the message using its shared secret key. However, for this kind of methods, the authenticity can only be verified by the node with the shared secret key. The secret key is generally shared by a group of sensor nodes. An intruder can compromise the key by capturing a single sensor node. In addition, this method cannot authenticate messages that are multicast. In [37, 38], symmetric key and hash based authentication schemes were proposed for WSNs. In these schemes, each symmetric authentication key is shared by a group of sensor nodes. An intruder can compromise the key by capturing a single sensor node. Therefore, these schemes are not resilient to a large number of nodes compromising. Another type of symmetric-key scheme requires synchronization among nodes. These schemes, including TESLA and its variants [39–44], can also provide message sender authentication. However, this scheme requires initial time synchronization, which is not easy to be implemented in large scale WSNs. In addition, they also introduce delay in message authentication, and the delay increases as the network scales up.

To solve the scalability problem, a secret polynomial based message authentication scheme was introduced in [45]. The idea of this scheme is similar to threshold secret sharing, where the threshold is determined by the degree of the polynomial. It also offers information theoretic security of the shared secret key when the number of messages transmitted is less than the threshold. The scheme enables the intermediate nodes to verify the authenticity of the message through a polynomial evaluation. However, when the number of messages transmitted is larger than the threshold, the polynomial can be fully recovered and the system is completely broken. The threshold is determined based on the degree of the polynomial. To increase the threshold and the complexity for the intruder to break the secret polynomial, a random noise, also called a perturbation factor, was added to the polynomial in [46–48]. The idea is to add a random noise, also called a perturbation factor, to the polynomial so that the coefficients of the polynomial cannot be easily solved. However, a recently study shows that the random noise can be completely removed from the polynomial using error-correcting code techniques [49].

Recently, message sender anonymity based on ring signatures was introduced [61]. This approach enables the message sender to generate a source anonymous message signature with content authenticity assurance. To generate a ring signature, a ring member randomly selects an ambiguity set (AS) and forges a message signature for all other members. Then he uses his trap-door information to glue the ring together. The original scheme has very limited flexibility and very high complexity. Moreover, the original paper only focuses on the cryptographic algorithm, and the relevant network issues were left unaddressed.

In this chapter, we propose an unconditionally secure and efficient source anonymous message authentication (SAMA) scheme on elliptic curves, based on the optimal modified ElGamal signature (MES) scheme on elliptic curves. This is because the MES scheme is secure against no-message attacks and adaptive chosen-message attacks in the random oracle model [107]. To the best of our knowledge, this is the first scheme that provides hop-by-hop node authentication without the threshold limitation, while with performance better than the symmetric-key based schemes. The distributed nature of our algorithms enables these schemes to be utilized in decentralized networks to provide message authentication.

The rest of this chapter is organized as follows: Section 4.2 defines the problem and presents terminology and the preliminary that will be used in this dissertation. Section 4.3 describes the proposed source anonymous message authentication on elliptic curves. Section 4.4 devises that anonymity set selection and source privacy. Section 4.5 offers key management and possible compromised node detection. Performance analysis and simulation results are provided in Section 4.6. We then conclude in Section 4.7.

4.2 Terminology

Privacy is sometimes referred to as anonymity. Communication anonymity in information management has been discussed in a number of previous works [2,8,108–111]. It generally refers to the state of being unidentifiable within a set of subjects. This set is called the ambiguity set (AS). *Sender anonymity* means that a particular message is not linkable to any sender and no message is linkable to a particular sender.

4.2.1 Source Anonymous Message Authentication Scheme (SAMA)

We will start with the definition of the unconditionally secure source anonymous message authentication scheme.

Definition 4.1 (SAMA) *A* SAMA *consists of the following two algorithms:*

- Generate $(m, Q_1, Q_2, \cdots, Q_n)$*: Given a message* m *and the public keys* $Q_1, Q_2, \cdots, Q_n$ *of the ambiguity set (AS)* $\mathcal{S} = \{A_1, A_2, \cdots, A_n\}$*, the actual message sender* $A_t, 1 \leq t \leq n$*, produces an anonymous message* $\mathcal{S}(m)$ *using its own private key* d_t*.*

- Verify $\mathcal{S}(m)$*: Given a message* m *and an anonymous message* $\mathcal{S}(m)$*, which includes the public keys of all members in the AS, a verifier can determine whether* $\mathcal{S}(m)$ *is generated by a member in the AS.*

The security requirements for SAMA include:

- Sender ambiguity*: The probability that a verifier successfully determines the real sender of the anonymous message is exactly* $1/n$*, where* n *is the total number of members in* AS*.*
- Unforgeability*: An anonymous message scheme is unforgeable if no adversary, given the public keys of all members of the AS and the anonymous messages* $m_1, m_2, \cdots, m_n$ *adaptively chosen by the adversary, can produce in polynomial time a new valid anonymous message with non-negligible probability.*

In this dissertation, the user ID and user public key will be used interchangeably without making any distinctions.

4.2.2 Modified ElGamal Signature Scheme (MES)

Definition 4.2 (MES) *The modified ElGamal signature scheme [106, 112] consists of the following three algorithms:*

Key generation algorithm *Let* p *be a large prime and* g *be a generator of* $\mathbb{Z}_p^*$*. Both* p *and* g *are made public. For a random private key* $x \in \mathbb{Z}_p$*, the public key* y *is computed from* $y = g^x \bmod p$*.*

Signature algorithm *The MES can also have many variants [113, 114]. For the purpose of efficiency, we will describe the variant, called* optimal scheme*. To sign a message* m*, one chooses a random* $k \in \mathbb{Z}_{p-1}^*$*, then computes the exponentiation* $r = g^k \bmod p$ *and solves* s *from*

$$s = rxh(m, r) + k \bmod (p - 1), \tag{4.1}$$

where h *is a one-way hash function. The signature of message* m *is defined as the pair* (r, s)*.*

Verification algorithm *The verifier checks whether the signature equation* $g^s = ry^{rh(m,r)} \bmod p$. *If the equality holds true, then the verifier* Accepts *the signature and* Rejects *otherwise.*

4.3 Proposed Source Anonymous Message Authentication (SAMA) on Elliptic Curve

In this section, we propose an unconditionally secure and efficient source anonymous message authentication scheme (SAMA). The main idea is that for each message m to be released, the message sender, or the sending node, generates a source anonymous message authentication for the message m. The generation is based on the MES scheme on elliptic curve. For a ring signature, each ring member is required to compute a forgery signature for all other members in the AS individually. In our scheme, the entire SAMA generation requires only three steps, which link all non-senders and the message sender to the SAMA alike. In addition, the design enables the SAMA to be verified through a single equation without individually verifying the signatures.

4.3.1 Proposed MES Scheme on Elliptic Curve

Let $p > 3$ be an odd prime. An elliptic curve E is defined by an equation of the form:

$$E: \; y^2 = x^3 + ax + b \bmod p,$$

where $a, b \in \mathbb{F}_p$, and $4a^3 + 27b^2 \not\equiv 0 \bmod p$. The set $E(\mathbb{F}_p)$ consists of all points $(x, y) \in \mathbb{F}_p$ on the curve, together with a special point $\mathcal{O}$ called the point at infinity.

Let $G = (x_G, y_G)$ be a base point on $E(\mathbb{F}_p)$ whose order is a very large value N. User A selects a random integer d_A in the $[1, N-1]$ as his private key. Then, he can compute his public key Q_A from $Q_A = d_A \times G$.

Signature generation algorithm For Alice to sign a message m, she follows these steps:

1. Select a random integer k_A, $1 \leq k_A \leq N-1$.
2. Calculate $r = x_A \bmod N$, where $(x_A, y_A) = k_A G$. If $r = 0$, go back to step 1.
3. Calculate $h_A \xleftarrow{l} h(m,r)$, where h is a cryptographic hash function, such as SHA-1, and $\xleftarrow{l}$ denotes the l leftmost bits of the hash.
4. Calculate $s = r d_A h_A + k_A \bmod N$. If $s = 0$, go back to step 2.
5. The signature is the pair (r, s).

When computing s, the string h_A resulting from $h(m,r)$ shall be converted into an integer. Note that h_A can be greater than N but not longer.

Signature verification algorithm For Bob to authenticate Alice's signature, he must have a copy of her public key Q_A.

1. Check that $Q_A \neq \mathcal{O}$, otherwise invalid
2. Check that Q_A lies on the curve
3. Check that $nQ_A = \mathcal{O}$

After that, Bob follows these steps to verify the signature:

1. Verify that r and s are integers in $[1, N-1]$. If not, the signature is invalid.
2. Calculate $h_A \xleftarrow{l} h(m,r)$, where h is the same function used in the signature generation.
3. Calculate $(x_1, x_2) = sG - r h_A Q_A \bmod N$.
4. The signature is valid if $r = x_1 \bmod N$, invalid otherwise.

In fact, if the signature is correctly generated, then

$$\begin{aligned}(x_1, x_2) &= sG - rh_AQ_A \\ &= (rd_Ah_A + k_A)G - rh_AQ_A \\ &= k_AG + rh_AQ_A - rh_AQ_A \\ &= k_AG.\end{aligned}$$

Therefore, we have $x_1 = r$ and the verifier should Accept the signature.

4.3.2 Proposed SAMA on Elliptic Curve

Suppose that the message sender (say Alice) wishes to transmit a message m anonymously from her network node to any other nodes. The AS includes n members, $A_1, A_2, \cdots, A_n$, e.g., $\mathcal{S} = \{A_1, A_2, \cdots, A_n\}$, where the actual message sender Alice is A_t, for some value $t, 1 \le t \le n$. In this dissertation, we will not distinguish between the node A_i and its public key Q_i. Therefore, we also have $\mathcal{S} = \{Q_1, Q_2, \cdots, Q_n\}$.

Authentication generation algorithm Suppose m is a message to be transmitted. The private key of the message sender Alice is $d_t, 1 \le t \le N$. To generate an efficient SAMA for message m, Alice performs the following three steps:

1. Select a random and pairwise different k_i for each $1 \le i \le n-1, i \neq t$ and compute r_i from $(r_i, y_i) = k_iG$.
2. Choose a random $k_i \in \mathbb{Z}_p$ and compute r_t from $(r_t, y_t) = k_tG - \sum\limits_{i \neq t} r_ih_iQ_i$ such that $r_t \neq 0$ and $r_t \neq r_i$ for any $i \neq t$, where $h_i \xleftarrow{l} h(m, r_i)$.
3. Compute $s = k_t + \sum\limits_{i \neq t} k_i + r_td_th_t \bmod N$.

The SAMA of the message m is defined as

$$\mathcal{S}(m) = (m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, s).$$

4.3.3 Verification of SAMA

Verification algorithm For Bob to verify an alleged SAMA $(m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, s)$, he must have a copy of the public keys $Q_1, \cdots, Q_n$. Then he checks:

1. Check that $Q_i \neq \mathcal{O}, i = 1, \cdots, n$, otherwise invalid

2. Check that $Q_i, i = 1, \cdots, n$ lies on the curve

3. Check that $nQ_i = \mathcal{O}, i = 1, \cdots, n$

After that, Bob follows these steps:

1. Verify that $r_i, y_i, i = 1, \cdots, n$ and s are integers in $[1, N-1]$. If not, the signature is invalid.

2. Calculate $h_i \overset{l}{\longleftarrow} h(m, r_i)$, where h is the same function used in the signature generation.

3. Calculate $(x_0, y_0) = sG - \sum\limits_{i=1}^{n} r_i h_i Q_i$

4. The signature is valid if the first coordinate of $\sum\limits_{i}(r_i, y_i)$ equals x_0, invalid otherwise.

In fact, if the SAMA has been correctly generated without being modified, then we compute

$$\begin{aligned}
(x_0, y_0) &= sG - \sum_{i=1}^{n} r_i h_i Q_i \\
&= (k_t + \sum_{i \neq t} k_i + r_t d_t h_t)G - \sum_{i} r_i h_i Q_i \\
&= \sum_{i \neq t} k_i G + (k_t G - \sum_{i \neq t} r_i h_i Q_i) \\
&= \sum_{i \neq t} (r_i, y_i) + (r_t, y_t) \\
&= \sum_{i} (r_i, y_i).
\end{aligned}$$

Therefore, the verifier should always Accept the SAMA.

4.3.4 Security Analysis

In this subsection, we will prove that the proposed SAMA scheme can provide unconditional source anonymity and provable unforgeability against adaptive chosen-message attacks.

Anonymity

In order to prove that the proposed SAMA can ensure unconditional source anonymity, we have to prove that (i) for anybody other than the members of $\mathcal{S}$, the probability to successfully identify the real sender is $1/n$, and (ii) anybody from $\mathcal{S}$ can generate SAMAs.

Theorem 4.1 *The proposed source anonymous message authentication scheme (SAMA) can provide unconditional message sender anonymity.*

Proof: The identity of the message sender is unconditionally protected with the proposed SAMA scheme. This is because, regardless of the sender's identity, there are exactly $(N-1)(N-2)\cdots(N-n)$ different options to generate the SAMA. All of them can be chosen by any members in the AS during the SAMA generation procedure with equal probability without depending on any complexity-theoretic assumptions. The proof for the second part, that anybody from $\mathcal{S}$ can generate the SAMA, is straightforward. This finishes the proof of this theorem. □

Unforgeability

The design of the proposed SAMA relies on the ElGamal signature schemes. Signature schemes can achieve different levels of security. Security against existential forgery under adaptive-chosen message attacks is the maximum level of security.

In this section, we will prove that the proposed SAMA is secure against existential forgery under adaptive-chosen message attacks in the random oracle model [115]. The security of our result is based on elliptic curve cryptography (ECC), which assumes that the computation of discrete logarithms in elliptic curve is computationally infeasible. In other words, no efficient algorithms are known for non-quantum computers.

We will introduce two lemmas first. Lemma 1 is the Splitting Lemma, which is a well-known probabilistic lemma from reference [107]. The basic idea of the Splitting Lemma is that when a subset Z is "large" in a product space $X \times Y$, it will have many "large" sections. Lemma 2 is a slight modification of the Forking Lemma presented in [107]. The proofs of the two lemmas are mainly probability theory related. We will skip the proofs of these two lemmas here.

Lemma 4.1 (The Splitting Lemma) *Let* $Z \subset X \times Y$ *such that* $\Pr[(x,y) \in Z] \geq \varepsilon$. *For any* $\alpha < \varepsilon$, *define* $W = \{(x,y) \in X \times Y | \Pr_{y' \in Y}[(x,y') \in Z] \geq \varepsilon - \alpha\}$, *and* $\bar{W} = (X \times Y) \backslash W$, *then the following statements hold:*

1. $\Pr[W] \geq \alpha$.

2. $\forall (x,y) \in W, \Pr_{y' \in Y}[(x,y') \in Z] \geq \varepsilon - \alpha$.

3. $\Pr[W|Z] \geq \alpha/\varepsilon$.

Lemma 4.2 (The Forking Lemma) *Let $\mathcal{A}$ be a Probabilistic Polynomial Time (PPT) Turing machine. Given only the public data as input, if $\mathcal{A}$ can find, with non-negligible probability, a valid SAMA $(m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, h_1, \cdots, h_n, s)$ within a bounded polynomial time T, then with non-negligible probability, a replay of this machine, which has control over $\mathcal{A}$ and a different oracle, outputs another valid SAMA $(m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, h'_1, \cdots, h'_n, s)$, such that $h_i = h'_i$, for all $1 \leq i \leq v, i \neq j$ for some fixed j.*

Theorem 4.2 *The proposed SAMA is secure against adaptive chosen-message attacks in the random oracle model.*

Proof: (Sketch) If an adversary can forge a valid SAMA with non-negligible probability, then according to the Forking Lemma, the adversary can get two valid SAMAs $\mathcal{S}(m) = (m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, h_1, \cdots, h_n, s)$, and $\mathcal{S}(m) = (m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, h'_1, \cdots, h'_n, s)$, such that for $1 \leq i \leq n, i \neq j, h_i = h'_i, h_j \neq h'_j$ and $sG - \sum_{i=1}^{n} r_i h_i Q_i = \sum_i (r_i, y_i)$, $s'G - \sum_{i=1}^{n} r_i h'_i Q_i = \sum_i (r_i, y_i)$.

Subtracting these two equations, we obtain $(s - s')G = r_j(h_j - h'_j)Q_j$. Equivalently, we have

$$Q_j = \frac{s - s'}{r_j(h_j - h'_j)} G.$$

We can compute the elliptic curve discrete logarithm of Q_j in base G with non-negligible probability, which contradicts the assumption that it is computationally infeasible to compute the elliptic discrete logarithm of Q_j in base G. Therefore, it is computationally infeasible for any adversary to forge a valid SAMA. □

4.4 Anonymity Set Selection and Source Privacy

The appropriate selection of AS plays a key role in message source privacy, since the actual message source node will be hidden in the AS. In this section, we will discuss techniques that can prevent the adversaries from tracking the message source through AS analysis in combination with local traffic analysis.

Before a message is transmitted, the message source node selects an AS from the public-key list in the SS of its choice. This set should include itself together with some other nodes. When an adversary receives a message, he can possibly find the direction of the previous hop, or even the real node of the previous hop. However, the adversary is unable to distinguish whether the previous node is the actual source node or simply a forwarder node if the adversary is unable to monitor the traffic of the previous hop. The selection of the AS should create sufficient diversity so that it is infeasible for the adversary to find the message source based on the selection of the AS itself.

Some basic criteria for the selection of the AS can be described as follows:

- To provide message source privacy, the message source needs to select the AS

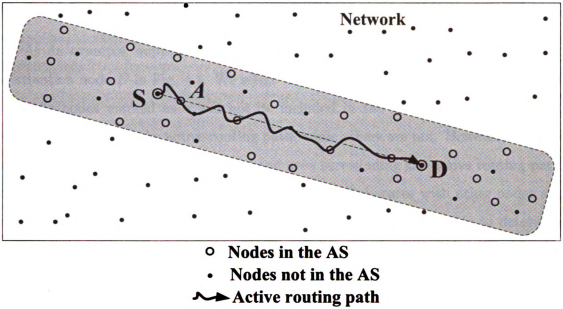


Figure 4.1. Anonymous set selection in active routing

to include nodes from all directions of the source node. In particular, the AS should include nodes from the opposite direction of the successor node. In this way, even the immediate successor node will not be able to distinguish the message source node from the forwarder based on the message that it receives.

- Though the message source node can select any node in the AS, some nodes in the AS may not be able to add any ambiguity to the message source node. These nodes are not appropriate candidates for the AS. They should be excluded from the AS for performance advantages.

- To balance the source privacy and efficiency, we should try to limit the nodes to be within a predefined distance range from the routing path. We recommend selecting AS from the nodes in a band that covers the active routing path. However, the AS does not have to include all the nodes in the routing path.

- The AS does not have to include all nodes in that range, nor does it have to include all the nodes in the active routing path. In fact, if all nodes are included in the AS, then this may help the adversary to identity the possible routing path and find the source node.

As an example, suppose we want to transmit a packet from source node S to destination node D in Fig. 4.1. We select the AS to include only nodes marked with $\circ$, while nodes marked as $\bullet$ will not be included in the AS. Of all these $\circ$ nodes, some of them are on the active routing path, while others are not. However, all these nodes are located within the shaded band area surrounding the active routing path. Suppose node A is compromised; unless node A collaborates with other nodes and can fully monitor the traffic of the source node S, it will not be able to determine whether S is the source node, or simply a forwarder. Similar analysis is also true for other nodes.

Any node in the active routing path can verify the contents' authenticity and integrity. However, anybody who receives a packet in the transmission can possibly exclude some of the nodes in the WSNs as the possible source node. Inclusion of these nodes in the AS does not increase the source privacy. Nevertheless, the more nodes included in the AS, the higher the energy cost will be. The selection of the AS has to be done with care so that the energy cost and the source privacy can both be optimized.

In addition, to balance the power consumption between authenticity and integrity verification, and the possibility that corrupted messages are being forwarded, the verification service may not have to take place in every hop; instead, it may be configured to take place in every other hop, for instance.

4.5 Key Management and Compromised Node Detection

In our scheme, we assume that there is a security server (SS) whose responsibilities include public-key storage and distribution in the WSNs. We assume that the SS will never be compromised. However, after deployment, the sensor node may be captured and compromised by the attackers. Once compromised, all information stored in the sensor node will be accessible to the attackers. We further assume that the compromised node will not be able to create new public keys that can be accepted by the SS.

For efficiency, each public key will have a short identity. The length of the identity is based on the scale of the WSNs.

Compromised Node Detection

As a special case scenario, we assume that all sensor information will be delivered to the SINK node, which can be co-located with the SS. As described in Section 4.4, when a message is received by the SINK node, the message source is hidden in an AS. Since the SAMA scheme guarantees that the message integrity is untampered, when a bad or meaningless message is received by the SINK node, the source node is viewed as compromised. If the compromised source node only transmits one message, it would be very difficult for the node to be identified without additional network traffic information. However, when a compromised node transmits more than one message, the SINK node can narrow the possible compromised nodes down to a very small set.

As shown in Fig. 4.2, we use the circle to represent an AS. When only one message is transmitted, the SINK node can only get the information that the source node will be in a set, say AS_1. When the compromised source node transmits two messages, the SINK node will be able to narrow the source node down to the set with both vertical lines and horizontal lines. When the compromised source node transmits three messages, the source node will be further narrowed down to the shaded area. Therefore, if the SINK node keeps tracking the compromised message, there is a high probability that the compromised node can be isolated.

When a node has been identified as compromised, the SS can remove its public key from its public key list. It can also broadcast the node's short identity to the entire sensor domain so that any sensor node that uses the stored public key for AS selection can update its key list. Once the public key of a node has been removed from the public-key list, and/or broadcasted, any message with the AS containing the compromised node should be dropped without any process to save the precious

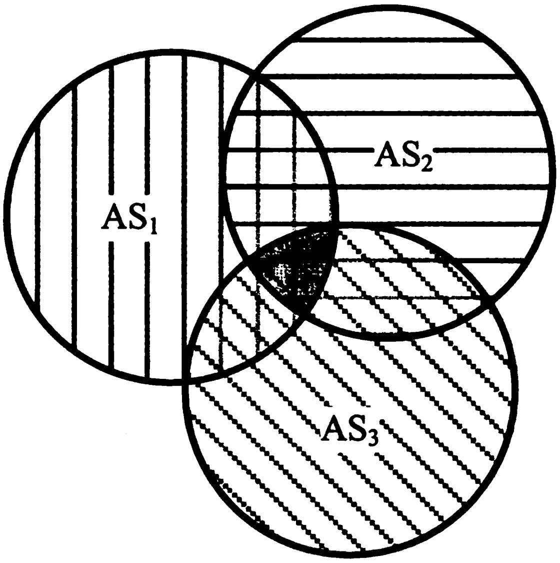


Figure 4.2. Compromised node detection

sensor power.

4.6 Performance Analysis for SAMA

In this section, we will evaluate our proposed authentication scheme through both theoretical analysis and simulation demonstrations. We will compare the proposed scheme with the bivariate polynomial-based symmetric-key scheme described in [45, 48]. We will provide multiple simulation results to demonstrate our analysis.

4.6.1 Theoretical Analysis

Key management is one of the major issues for secret-key based authentication schemes. This is especially true for large scale WSNs. While many of these schemes are designed to provide node authentication, they can only provide end-to-end node authentication using the secret key shared between the two nodes, which implies that only the receiver can verify the authenticity of the messages en-route. This means that no intermediate node can authenticate the message in general. The intermediate nodes may have to forward a manipulated message for many hops before the message can finally be authenticated and dropped by the receiving node. This not only consumes extra sensor power, but also increases the network collision and decreases the message delivery ratio. In addition to performance improvement, enabling intermediate node authentication will thwart adversaries from performing denial-of-service attacks through message manipulation to deplete the energy and communication resources of the wireless network. Developing a protocol that can provide hop-by-hop intermediate node authentication is an important research task.

While hop-by-hop authentication can be achieved through a public-key encryption system, the public-key based schemes are not preferred mainly due to their high computational overhead. Most of the authentication schemes are based on symmetric-key schemes, including the polynomial evaluation based threshold authentication scheme [48]. The secret bivariate polynomial is defined as [45]:

$$f(x,y) = \sum_{i=0}^{d_x} \sum_{j=0}^{d_y} A_{i,j} x^i y^j,$$

where each coefficient $A_{x,y}$ is an element of a finite field $\mathbb{F}_p$, and d_x and d_y are the degrees of this polynomial. d_x and d_y are also related to the message length and the computational complexity of this scheme. From the performance aspect, d_x and d_y should be as short as possible.

On the other hand, it is easy to see that when either more than $d_y + 1$ messages transmitted from the base station are received and recorded by the intruders, or more than $d_x + 1$ sensor nodes have been compromised, the intruders can recover the polynomial $f(x, y)$ via Lagrange interpolation. In this case, the whole security system is totally broken and cannot be used anymore. This property requires that both d_x and d_y must be very large for the scheme to be resilient to compromised node.

An alternative approach based on perturbation of the polynomial was also explored. The main idea is to add a small amount of random noise to the polynomial in the original scheme so that the adversaries will no longer be able to solve the coefficients using Lagrange interpolation. However, this technique is proved to be vulnerable to security attacks [49], since the random noise can be removed from the

Table 4.1. Comparable Key Sizes in Terms of Computational Effort for Cryptanalysis

Symmetric Scheme (key size in bits)	RSA/DSA (modulus size in bits)	ECC-Based Scheme (size of n in bits)
56	512	112
80	1024	160
112	2048	224
128	3072	256
192	7680	384
256	15360	512

polynomial using error-correcting techniques.

For our scheme, each SAMA contains an AS of n randomly selected nodes that dynamically changes for each message. Even if one message is corrupted, other messages transmitted in the network can still be secure. Therefore, n can be much smaller than the parameters d_x and d_y. In fact, even a small n may provide adequate source privacy while ensuring high system performance. The performance analysis and simulation results will be provided in the following sections.

In addition, in the bivariate polynomial-based scheme, there is only one base station that can send messages. All the other nodes can only act as intermediate nodes or receivers. This property makes the base station easy to attack and severely narrows the applicability of this scheme. In fact, the major traffic in WSNs is packet delivery from the sensor nodes to the SINK node. In this case, our scheme enables every node to transmit the message to the SINK node as a message initiator.

The recent progress on elliptic curve cryptography (ECC) has demonstrated that the public-key based schemes have more advantages in terms of memory usage, message complexity, and security resilience, since public-key based approaches have simple and clean key management [52, 55, 56]. According to Table 4.1 [116], to achieve the same level of security, the key size of elliptic curve-based schemes is much shorter than that of the traditional public-key cryptosystem. This progress facilitates the implementation of authentication schemes using ECC.

Table 4.2. Process Time for the Two Schemes (ms)

	Polynomial based approach						Proposed approach							
	$d_x, d_y = 80$		$d_x, d_y = 100$		$d_x, d_y = 150$		$n = 10$		$n = 15$		$n = 20$		$n = 30$	
	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify
$l = 24$	33.75	94.37	52.10	172.40	138.85	589.33	19.38	12.65	30.47	21.09	39.84	29.85	61.32	45.06
$l = 32$	37.66	123.59	60.30	243.40	178.80	927.70	29.84	22.34	44.85	33.12	59.54	44.37	89.57	66.30
$l = 40$	41.10	158.28	69.53	320.70	214.60	1266.20	46.41	34.68	70.31	51.87	92.97	69.22	139.63	103.23
$l = 64$	43.91	264.06	75.80	539.60	245.30	2289.40	115.47	86.25	173.13	128.75	231.25	172.97	346.01	258.30

4.6.2 Experimental Results

In this section, we implement the bivariate polynomial-based scheme and our proposed scheme in a real world comparison. The comparison is based on comparable security levels.

Simulation parameter setup

The bivariate polynomial-based scheme is a symmetric-key based implementation, while our scheme is based on ECC. This requires us to determine the comparable key sizes. According to [116], also summarized in Table 4.1, if we choose the key size to be l for the symmetric-key cryptosystem, then the key size for our proposed ECC will be $2l$, which is much shorter than the traditional public-key cryptosystem. This progress facilitates the implementation of the authentication scheme using ECC.

In the simulation setting, we choose four security levels, which are indicated by the symmetric-key sizes l: 24-bit, 32-bit, 40-bit and 64-bit, respectively. The comparable key sizes of our scheme are 48-bit, 64-bit, 80-bit and 128-bit, respectively.

We also need to determine d_x and d_y for the bivariate polynomial-based scheme and the n for our scheme. In our simulation, we select d_x equal d_y and choose three values for them: 80, 100 and 150. We assume that WSNs does not contain more than 2^{16} nodes in the simulation, which is reasonably large. For size n of the AS, we choose four values in the simulation: 10, 15, 20, 30, although we believe that 30 is a little bit larger than necessary.

We will compare the *computational overhead*, *communication overhead*, *delivery ratio*, *energy consumption*, *transmission delay*, and *memory consumption* of our proposed scheme with those of the bivariate polynomial-based scheme.

Computational overhead

For a public-key based authentication scheme, computational overhead is one of the most important performance measurements. Table 4.2 shows the process time of our scheme and the bivariate polynomial-based scheme for both authentication generation and verification. In the simulations, we assume that the key length of our scheme is $2l$.

The simulations were carried out using Maple 12, which runs on an Intel Core Duo CPU 6420 2.13GHz machine with 1.99GB memory. From Table 4.2, we have the following findings:

- For the bivariate polynomial-based scheme, the authentication generation time is much shorter than the verifying time; meanwhile, for our proposed scheme, the verification time is much shorter than the authentication generation time.
- Our scheme is more efficient for hop-by-hop authentication under comparable security levels. More importantly, the verification time for our scheme is much shorter than the bivariate polynomial-based scheme's since verification will be conducted in multiple hops.

Communication overhead and message transmission delay

The communication overhead is determined by the message length. For the bivariate polynomial-based scheme, each message is transmitted in the form of $< m, MAF_m(y) >$, where $MAF_m(y)$ is defined as: $MAF_m(y) = f(h(m), y) = \sum_{j=0}^{d_y} M_j y^j$. $MAF_m(y)$ is represented by its $d_y + 1$ coefficients $M_i, \in \mathbb{Z}_p, 0 \leq i \leq d_y$, where $p \in (2^{l-1}, 2^l)$ is a large prime number. The total length of $< m, MAF_m(y) >$ is $l(d_y + 1)$.

For our scheme, the message format is: $(m, \mathcal{S}, r_1, y_1, \cdots, r_n, y_n, s)$, where m, s, r_i, y_i are all numbers with length L. $\mathcal{S}$ is the ID list for all the nodes included in the AS. Assuming the network is composed of λ nodes in total, each ID will be of the length: $\lceil \log_2 \lambda \rceil$. When n nodes are included in the AS, the length of $\mathcal{S}$ is $n \lceil \log_2 \lambda \rceil$. Therefore, the total length of one message for our scheme is: $2L(n + 1) + \lceil \log_2 \lambda \rceil$.

Table 4.3 enumerates the message length for multiple scenarios with $\lambda = 2^{16}$. From this table, we can see that for each security level, the length of our scheme is also slightly shorter than the bivariate polynomial-based scheme.

Table 4.3. The Message Length for the Two Schemes (Byte)

	Polynomial approach			**Proposed approach**			
	d_x, d_y			n			
	80	100	150	10	15	20	30
$l = 24$	246	306	456	152	222	292	432
$l = 32$	328	408	608	196	286	376	556
$l = 40$	410	510	760	240	350	460	680
$l = 64$	656	816	1216	372	542	712	1052

The large communication overhead of the polynomial-based scheme will increase the energy consumption and message delay. The simulation results in Fig. 4.6.2 and 4.6.2 demonstrate that our proposed scheme has a much lower energy consumption and message transmission delay. These simulations were carried out in ns-2 on RedHat Linux system. The security levels 1, 2, 3, 4 correspond to symmetric key sizes 24-bit, 32-bit, 40-bit, 64-bit, and elliptic curve key size 48-bit, 64-bit, 80-bit, 128-bit, respectively.

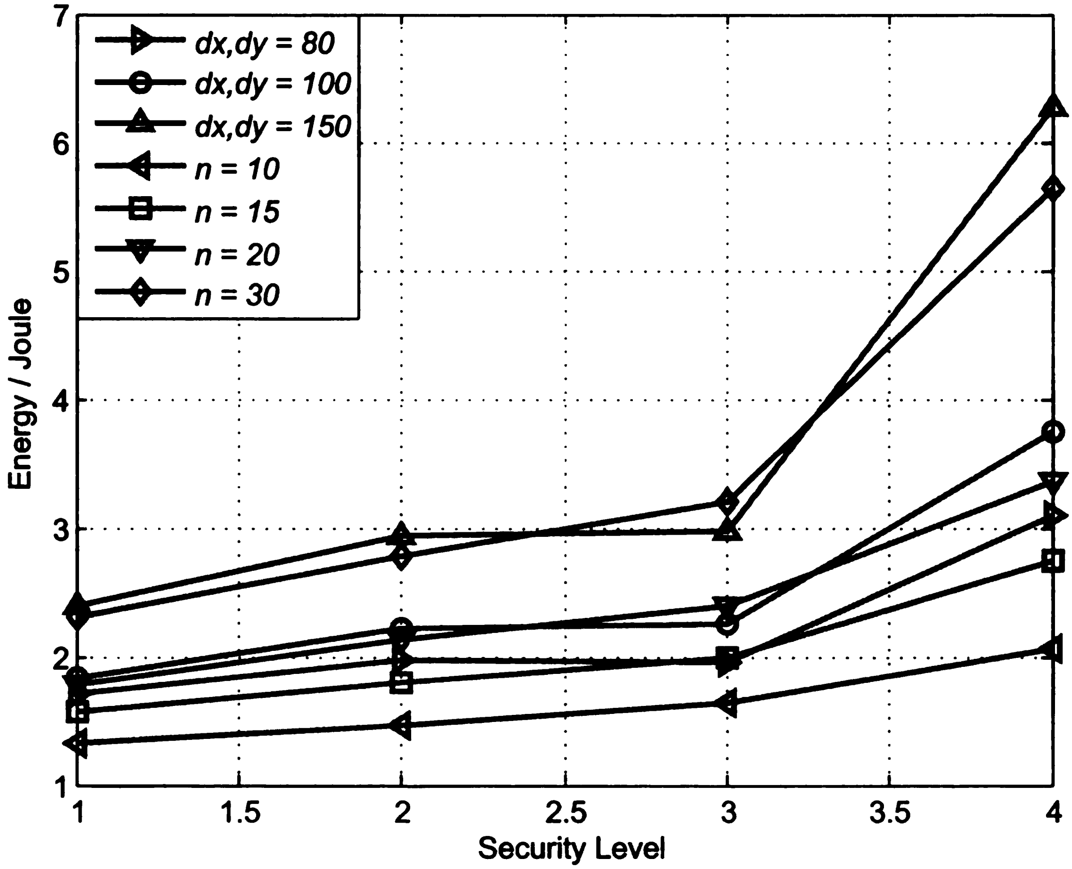


Figure 4.3. Energy consumption comparison

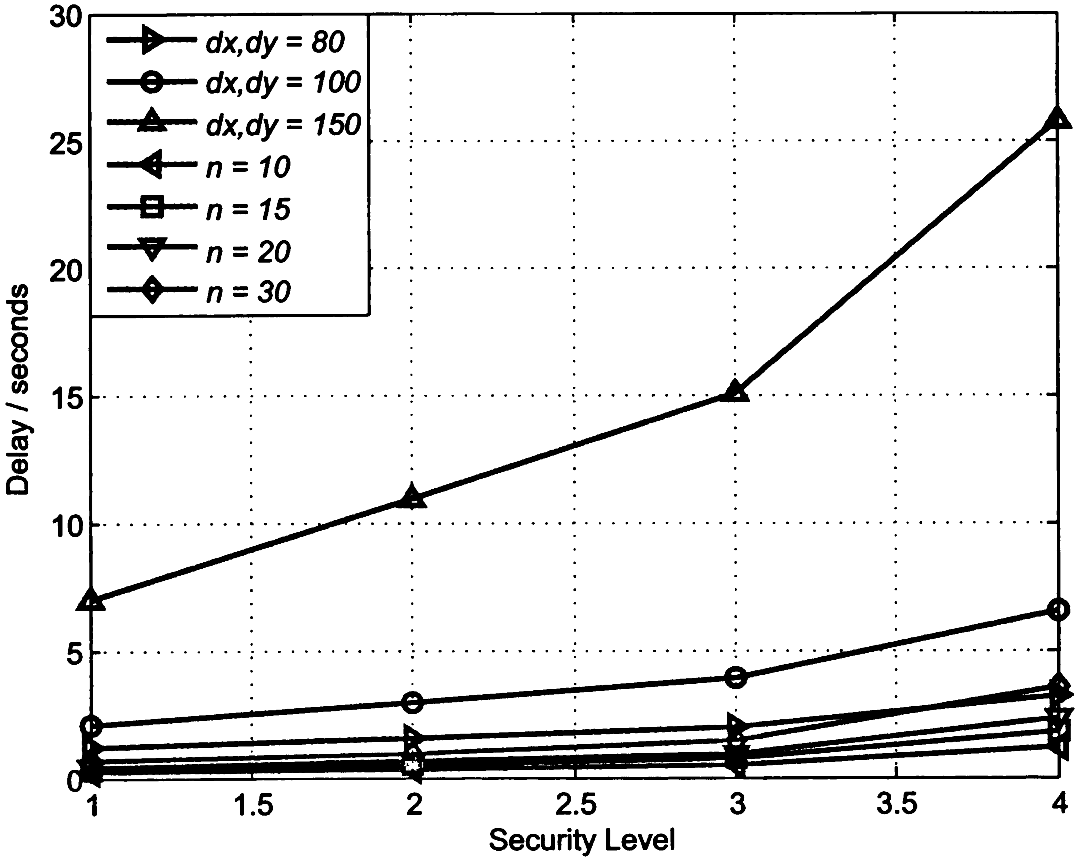


Figure 4.4. Message delay comparison

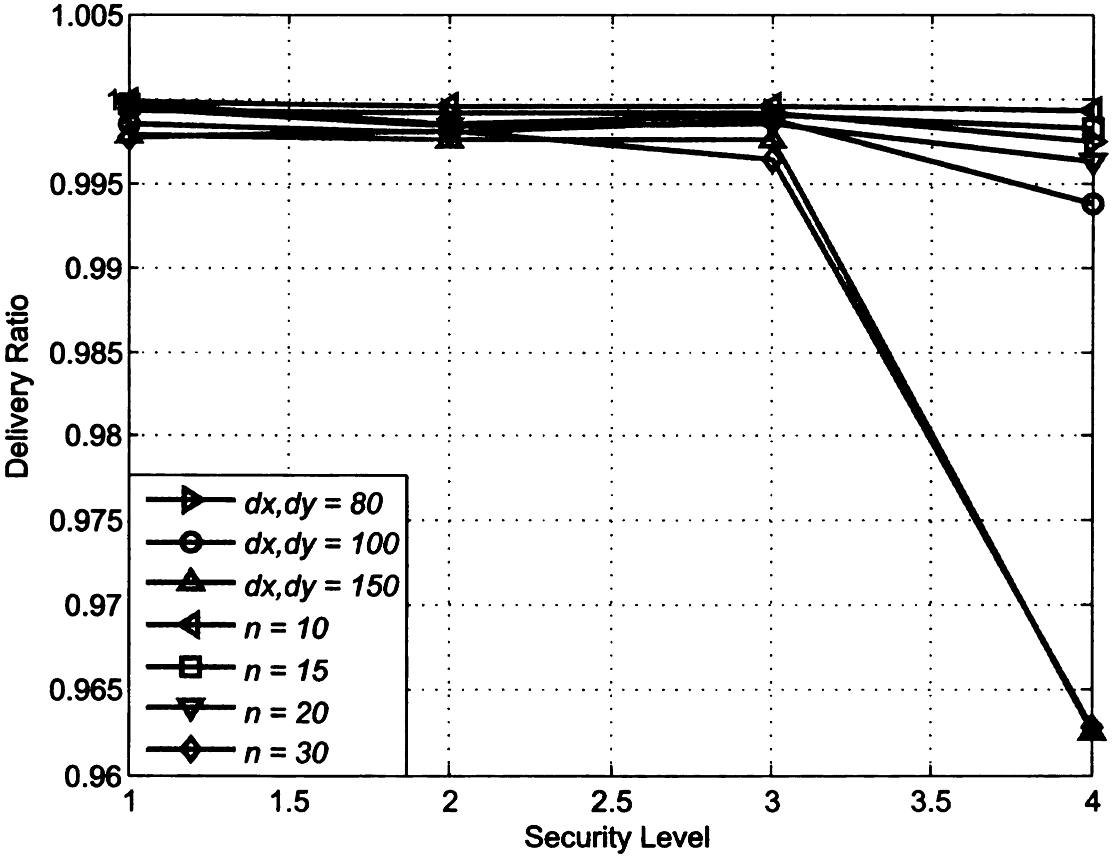


Figure 4.5. Message delivery ratio comparison

Table 4.4. Memory Consumption for the Two Schemes (KB) (Maple 12)

	Polynomial based approach						**Proposed approach**							
	$d_x, d_y = 80$		$d_x, d_y = 100$		$d_x, d_y = 150$		$n = 10$		$n = 15$		$n = 20$		$n = 30$	
	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify	Gen	Verify
$l = 24$	352.12	326.35	424.04	392.61	795.09	786.01	96.97	95.78	96.19	96.12	96.76	96.08	96.60	96.84
$l = 32$	487.33	456.46	645.80	611.72	1193.33	1149.64	99.81	99.21	99.65	98.98	99.65	98.74	100.04	99.32
$l = 40$	600.78	568.15	799.62	770.54	1459.04	1399.26	104.39	102.77	104.15	103.35	104.45	104.25	104.90	104.61
$l = 64$	601.36	569.69	794.81	772.61	1441.66	1383.91	121.28	119.43	121.84	121.22	122.27	121.33	122.15	121.39

We also conduct simulations to compare the delivery ratios and memory usage of the two schemes. The results are given in Fig. 4.6.2 and Table 4.4. The first simulation was carried out using ns-2 on RedHat Linux system, while the last one was derived using Maple 12. The security levels are configured the same way as in the previous simulations.

The results show that our scheme is slightly better than the bivariate polynomial-based scheme in delivery ratio. However, our scheme has a much lower memory consumption.

4.7 Summary

In this chapter, we proposed an unconditionally secure and efficient SAMA scheme on ECC, based on the optimal MES. Our scheme could provide hop-by-hop authentication. Our scheme also allows any node in WSNs to transmit an unlimited number of messages without suffering the threshold problem. Both theoretical analysis and simulation results demonstrate that the proposed scheme is secure, efficient in computation and communication. In addition, we proposed network implementation criteria for source node privacy protection and an efficient key management framework to isolate identified compromised nodes.

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions

Source privacy, which is composed of two conflicting requirements: source-location privacy and anonymous source authentication, is critical to the applications for WSNs. In this dissertation, we propose schemes to protect source privacy.

5.1.1 Source-Location Privacy Protection Schemes

For source-location privacy, we first build a security evaluation model to quantitatively measure the security properties of different source-location privacy protection schemes. Then using this model, we analyze some of the existing schemes. Under the guidance of these analysis results, we propose a dynamic ID assignment scheme and four routing-based source-location privacy protection schemes.

The dynamic ID assignment scheme aims at preventing the source-location information from being leaked in message content. Through this scheme, the adversaries cannot link messages generated by the same source node together. In this way, correlation-based source identification attack is prevented.

The routing-based schemes are proposed to defend the traffic pattern analysis attacks. In the first source-location privacy protection scheme, each message is being routed to a randomly selected intermediate node before the message is transmitted to the SINK node. There are three intermediate node selection methods introduced: constrained RSIN method, totally random RSIN method, and ring-band RSIN method. Security analysis shows that constrained RSIN method can only provide limited source-location privacy protection over large scale WSNs, while totally random RSIN and ring-band RSIN can achieve network-level source-location privacy. Simulation results illustrate that the constrained RSIN can provide the optimal communication performance among these three methods. The performance of totally random RSIN is much worse than the constrained RSIN. For energy consumption and transmission delay, the performance of ring-band RSIN is between constrained RSIN and totally random RSIN. However, the ring-band RSIN approach offers the lowest delivery ratio. It is hard to say which method of these three is better than the other two, users can make choices based on their specific situations.

The second source-location privacy protection scheme is implemented through routing in a network-level mixing ring. This scheme consists of three routing phases. The first phase is constrained RSIN, which can provide local source-location privacy. The second phase routes messages in a network mixing ring (NMR), which aims at network-level source-location privacy. In the last phase, messages are forwarded to the SINK node. Security analysis shows that this scheme can guarantee provable network-level source-location privacy.

The third and the fourth schemes are based on routing through multiple intermediate nodes. The intermediate nodes are selected in an angle-based method and a quadrant-based method respectively. For the third scheme, there is a tradeoff relationship between the communication performance and the security property. The users can determine whether this scheme should be more efficient or more secure based on their requirements, which means the third scheme is tunable. The fourth scheme can only provide network-level source-location privacy. However, when both the third scheme and the fourth scheme are providing global source-location privacy, the fourth scheme is more efficient.

5.1.2 Anonymous Source Authentication Schemes

The most secure authentication protocols for computer networks are usually implemented through public-key cryptosystems. However, computationally intensive cryptographic algorithms and large scale broadcasting-based protocols may not be quite suitable for WSNs because of the limited resources of sensor nodes. This makes source message authentication a challenging task.

After analyzing some of the existing message authentication protocols for WSNs, most of which suffer threshold limitation problem or could only provide end-to-end authentication, we propose to address these problems through ECC technology in this dissertation. Our scheme cannot only provide hop-by-hop authentication, but also allows any node in WSNs to transmit an unlimited number of messages without suffering the threshold problem. Both theoretical analysis and simulation results demonstrate that the proposed scheme is secure with light overhead.

In addition, we propose network implementation criteria for source node privacy protection and an efficient key management framework which ensures the identified compromised node to be isolated.

5.2 Related Future Work

In this section, we discuss the related research topics as a proposal for future work.

5.2.1 Further Research on Theoretical Security Analysis of Source-Location Privacy Protection Schemes

In Chapter 2, an evaluation model is proposed to quantitatively measure the security properties of the source-location privacy protection schemes. Using this model, we have analyzed the proposed schemes in Chapter 3. Our further research direction is to formally prove that our proposed security evaluation framework is equivalent to information-theoretical security.

5.2.2 Further Research on Multiple Network Mixing Ring(MNMR)

In Section 3.4, we introduce a routing-based scheme, in which a network mixing ring (NMR) is proposed to provide network-level source-location privacy. As a future research task, we will study protocol design with multiple NMR (MNMR) in WSNs.

We expect that the average transmission distance from the source node to the ring node may be decreased while the security level of the constrained RSIN phase may be increased. In addition, the multiple level NMR may be more secure than single NMR.

5.2.3 Further Research on Secure and Energy Aware Routing

Energy and security are both important design issues for WSNs. An interesting research topic that we will be investigate is to develop novel secure and energy aware routing protocols that can address these two issues concurrently through balanced energy consumption and probabilistic random walking. Based on the tradeoff relationship between security and energy, this protocol should provide tunable security level and energy consumption pattern.

5.2.4 Further Research on Compromised Node Identification of ECC-based SAMA

In Chapter 4, we introduce an ECC-based SAMA scheme. We also make a discussion about the anonymity set (AS) selection, key management, and compromised node detection methods to support the authentication scheme. More detailed further researches in these supportive topics need to be carried out. Further theoretical analysis and simulation results are necessary.

APPENDICES

APPENDIX A

List of Abbreviations and Acronyms

AES	Advanced Encryption Standard
AS	ambiguity set
DES	Data Encryption Standard
DSA	Digital Signature Algorithm
ECC	elliptic curve cryptography
MAF	Message Authentication Function
MES	Modified ElGamal Signature Scheme
NMR	network mixing ring
NSSI	Normalized Source-location Space Index
RSIN	Routing to a Single Intermediate Node
SAMA	Source Anonymous Message Authentication Scheme
SDI	Source-location Disclosure Index
SSI	Source-location Space Index
WSNs	Wireless Sensor Networks

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 Communications, 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] O. Berthold, H. Federrath, and S. K□ opsell, "Web MIXes: A system for anonymous and unobservable Internet access," Lecture Notes in Computer Science, pp. 115-129, 2001.

[17] B. Moller, "Provably secure public-key encryption for length-preserving chaumian mixes," in Proceedings of CT-RSA 2003, LNCS 2612, pp. 244-262, April 2003.

[18] R. D. G. Danezis and N. Mathewson, "Mixminion: design of a type iii anonymous remailer protocol," IEEE Symposium on Security and Privacy, pp. 2-15, 2003.

[19] C. Gulcu and G. Tsudik, "Mixing email with babel," in Proceedings of the Symposium on Network and Distributed System Security, (San Diego, CA), 1996.

[20] U. M□ oller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster protocol," July 2003. Version 2.

[21] J. Yao and G. Wen, "Preserving source-location privacy in energy-constrained wireless sensor networks," pp. 412 -416, june 2008.

[22] J. Kong and X. Hong, "Anodr: Anonymous on demand routing with untraceable routes for mobile ad-hoc networks," 2003.

[23] 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.

[24] 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.

[25] 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.

[26] W. Y.Zhang, W.Liu, "Anonymous communications in mobile adhoc networks," in IEEE Infocom, 2005.

[27] 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.

[28] Y. Jian, S. Chen, Z. Zhang, and L. Zhang, "Protecting receiver-location privacy in wireless sensor networks," pp. 1955 -1963, may 2007.

[29] 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.

[30] 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.

[31] L. Kang, "Protecting location privacy in large-scale wireless sensor networks," pp. 1 -6, june 2009.

[32] D. K.Mehta and M.Wright, "Location privacy in sensor networks against a global eavesdropper," 2007.

[33] 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.

[34] M. Shao, W. Hu, S. Zhu, G. Cao, S. Krishnamurth, and T. La Porta, "Crosslayer enhanced source location privacy in sensor networks," pp. 1 -9, June 2009.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] D. T. D. S. A. Perrig, R. Canetti, "The tesla broadcast authentication protocol," in CryptoBytes, 2002.

[41] D. Liu and P. Ning, "Efficient distribution of key chain commitments for broadcast authentication in distributed sensor networks," tech. rep., Raleigh, NC, USA, 2002.

[42] D. Liu and P. Ning, "Multilevel tesla: Broadcast authentication for distributed sensor networks," ACM Trans. Embed. Comput. Syst., vol. 3, no. 4, pp. 800-836, 2004.

[43] J. Drissi and Q. Gu, "Localized broadcast authentication in large sensor networks," pp. 25 -25, july 2006.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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/.

[50] 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.

[51] Watro, D. Kong, S. fen Cuti, C. Gardiner, C. Lynn, and P. Kruus, "Tinypk: securing sensor networks with public key technology," in SASN 04: Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks, pp. 59-64, ACM Press, 2004.

[52] 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.

[53] 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.

[54] 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.

[55] A. Liu and P. Ning. [Online] http://discovery.csc.ncsu.edu/software/TinyECC/, 2005.

[56] 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.

[57] H. Chan and A. Perrig, "Security and privacy in sensor networks," IEEE Computer Magazine, pp. 103-105, Oct. 2003.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

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

[64] J. Neander, E. Hansen, M. Nolin, and M. Bjorkman, "Asymmetric multihop communication in large sensor networks," Wireless Pervasive Computing, 2006 1st International Symposium on Jan. 2006.

[65] 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.

[66] M. Zamini and L. Zare, "A reward based method to wireless sensor network clustering," pp. 1 -7, oct. 2009.

[67] D. Xia and N. Vlajic, "Near-optimal node clustering in wireless sensor networks for environment monitoring," pp. 632 -641, may 2007.

[68] H. Alipour, M. Abbaspour, M. Esmaeili, H. Mousavi, and H. Shahhoseini, "Daca: Dynamic advanced clustering algorithm for sensor networks," pp. 518-525, dec. 2007.

[69] M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," Parallel and Distributed Systems, IEEE Transactions on, vol. 20, pp. 1844 -1856, dec. 2009.

[70] F. Ishmanov and S. W. Kim, "Distributed clustering algorithm with load balancing in wireless sensor network," vol. 1, pp. 19 -23, 31 2009-april 2 2009.

[71] F. Tashtarian, A. Haghighat, M. Honary, and H. Shokrzadeh, "A new energy efficient clustering algorithm for wireless sensor networks," pp. 1 -6, sept. 2007

[72] A. Taherkordi, R. Mohammadi, and F. Eliassen, "A communication-efficient distributed clustering algorithm for sensor networks," pp. 634 -638, march 2008.

[73] M.-W. Park, J.-Y. Choi, Y.-J. Han, and T.-M. Chung, "An energy efficient concentric clustering scheme in wireless sensor networks," pp. 58 -61, aug. 2009.

[74] 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.

[75] "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.

[76] 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.

[77] P. Zhang and M. Martonosi, "Locale: Collaborative localization estimation for sparse mobile sensor networks," pp. 195 -206, april 2008.

[78] T. Srinath, "Localization in resource constrained sensor networks using a mobile beacon with in-ranging," pp. 5 pp. -5, 0-0 2006.

[79] D. min Chen and Y. Zhang, "Research of wsn localization algorithm based on entropy function," vol. 1, pp. 229 -233, march 2009.

[80] Z. Chaczko, R. Klempous, J. Nikodem, and M. Nikodem, "Methods of sensors localization in wireless sensor networks," pp. 145 -152, march 2007.

[81] X. Ji and H. Zha, "Robust sensor localization algorithm in wireless ad-hoc sensor networks," pp. 527 - 532, oct. 2003.

[82] M. Rudafshani and S. Datta, "Localization in wireless sensor networks," pp. 51-60, april 2007.

[83] 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.

[84] S. Zhu, S. Setia, and S. Jajodia, "Leap: efficient security mechanisms for largescale 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.

[85] E. Henze, "A solution of the general equation for public key distribution systems," in Advances in Cryptography (A. Gersho, ed.), (Santa Barbara, California, USA), pp. 140-141, University of California, Santa Barbara, 1982.

[86] R. Blom, "Non-public key distribution," in Advances in Cryptology: Proceedings of Crypto'82 (D. Chaum, R. L. Rivest, , and A. T. Sherman, eds.), (New York, USA), pp. 231-236, Plenum Publishing, 1982.

[87] T. Matsumoto and H. Imai, "On the key predistribution system: a practical solution to the key distribution problem," in Advances in Cryptology - Crypto'87 (C. Pomerance, ed.), (Berlin), pp. 185-193, Springer-Verlag, 1987. Lecture Notes in Computer Science Volume 293.

[88] E. Okamoto, "Key distribution systems based on identification information," in Advances in Cryptology - Crypto'87 (C. Pomerance, ed.), (Berlin), pp. 194-202, Springer-Verlag, 1987. Lecture Notes in Computer Science Volume 293.

[89] M. Tatebayashi, N. Matsuzaki, and D. B. J. Newman, "Key distribution protocol for digital mobile communication systems," in Advances in Cryptology- Crypto'89 (G. Brassard, ed.), (Berlin), pp. 324-334, Springer-Verlag, 1989. Lecture Notes in Computer Science Volume 435.

[90] Y. Yacobi and Z. Shmuely, "On key distribution systems," in Advances in Cryptology - Crypto'89 (G. Brassard, ed.), (Berlin), pp. 344-355, Springer-Verlag, 1989. Lecture Notes in Computer Science Volume 435.

[91] J. W. Suurballe, "Disjoint paths in a network," Wiley Periodicals, vol. 4, pp. 125-145, 1974.

[92] R. E. T. J. W. Suurballe, "A quick method for finding shortest pairs of disjoint paths," Wiley Periodicals, vol. 14, pp. 325-336, 1984.

[93] R. Bhandari, "Optimal physical diversity algorithms and survivable networks," Computers and Communications, IEEE Symposium on, vol. 0, p. 433, 1997.

[94] A. Srinivasl and E. Modianol, "Finding minimum energy disjoint paths in wireless ad-hoc networks," Wireless Networks, vol. 11, pp. 401-417, July 2005.

[95] R. Andersen, F. Chung, A. Sen, and G. Xue, "On disjoint path pairs with wavelength continuity constraint in wdm networks," in INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, vol. 1, pp. -535, March 2004.

[96] J. Tang, G. Xue, and W. Zhang, "Interference-aware topology control and qos routing in multi-channel wireless mesh networks," in MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, (New York, NY, USA), pp. 68-77, ACM, 2005.

[97] H. Wang, B. Sheng, and Q. Li, "Privacy-aware routing in sensor networks," Comput. Netw., vol. 53, no. 9, pp. 1512-1529, 2009.

[98] J. Ren, Y. Li, and T. Li, "Routing-based source-location privacy in wireless sensor networks," in Communications, 2009. ICC '09. IEEE International Conference on, pp. 1 -5, june 2009.

[99] 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.

[100] 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.

[101] Y. Li and J. Ren, "Mixing ring-based source-location privacy in wireless sensor networks," in Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on, pp. 1 -6, aug. 2009.

[102] Y. Li and J. Ren, "Source-location privacy through dynamic routing in wireless sensor networks," in Infocom '10. IEEE International Conference on, March 2010.

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

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

[105] R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," Communications. of the Assoc. of Comp. Mach., vol. 21, no. 2, pp. 120-126, 1978.

[106] T. A. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," IEEE Transactions on Information Theory, vol. 31, no. 4, pp. 469-472, 1985.

[107] D. Pointcheval and J. Stern, "Security proofs for signature schemes," in Advances in Cryptology - EuroCrypt'96 (U. Maurer, ed.), (Berlin), pp. 387-398, Springer-Verlag, 1996. Lecture Notes in Computer Science Volume 1070.

[108] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Communications of the ACM, vol. 24, pp. 84-88, February 1981.

[109] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, unobservability, pseudonymity, and identity management a proposal for terminology." http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.31.pdf, Feb. 15 2008.

[110] A. Pfitzmann and M. Waidner, "Networks without user observability-design options.," in Advances in Cryptology - EuroCrypt'85 (F. Pichler, ed.), (Berlin), pp. 245-253, Springer-Verlag, 1985. Lecture Notes in Computer Science Volume 219.

[111] M. Waidner, "Unconditional sender and recipient untraceability in spite of active attacks," in Advances in Cryptology - EuroCrypt'89 (J.-J. Quisquater and J. Vandewalle, eds.), (Berlin), pp. 302-319, Springer-Verlag, 1989. Lecture Notes in Computer Science Volume 434.

[112] D. Pointcheval and J. Stern, "Security arguments for digital signatures and blind signatures," Journal of Cryptology, vol. 13, no. 3, pp. 361-396, 2000.

[113] L. Harn and Y. Xu, "Design of generalized ElGamal type digital signature schemes based on discret logarithm," Electronics Letters, vol. 30, no. 24, pp. 2025-2026, 1994.

[114] K. Nyberg and R. A. Rueppel, "Message recovery for signature schemes based on the discrete logarithm problem," in Advances in Cryptology - EuroCrypt'94 (A. D. Santis, ed.), (Berlin), pp. 182-193, Springer-Verlag, 1995. Lecture Notes in Computer Science Volume 950.

[115] M. Bellare and P. Rogaway, "Random oracles are practical: A paradigm for designing efficient protocols," in CCS'93, pp. 62-73, 1993. [116] BlueKrypt, "Cryptographic key length recommendation." http://www.keylength.com/en/3/.