Quantitative Measurement and Design of Source-Location Privacy Schemes for Wireless Sensor Networks

Yun Li, Jian Ren, *Senior Member*, *IEEE*, and Jie Wu, *Fellow*, *IEEE*

Abstract—Wireless sensor networks (WSNs) have been widely used in many areas for critical infrastructure monitoring and information collection. While confidentiality of the message can be ensured through content encryption, it is much more difficult to adequately address source-location privacy (SLP). For WSNs, SLP service is further complicated by the nature that the sensor nodes generally consist of low-cost and low-power radio devices. Computationally intensive cryptographic algorithms (such as public-key cryptosystems), and large scale broadcasting-based protocols may not be suitable. In this paper, we first propose criteria to quantitatively measure source-location information leakage in routing-based SLP protection schemes for WSNs. Through this model, we identify vulnerabilities of some well-known SLP protection schemes. We then propose a scheme to provide SLP through routing to a randomly selected intermediate node (RSIN) and a network mixing ring (NMR). Our security analysis, based on the proposed criteria, shows that the proposed scheme can provide excellent SLP. The comprehensive simulation results demonstrate that the proposed scheme is very efficient and can achieve a high message delivery ratio. We believe it can be used in many practical applications.

Index Terms—Source-location privacy, source-location information leakage, quantitative measurement, wireless sensor networks (WSNs).

1 Introduction

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. Sensor networks rely on wireless communication, which is by nature a broadcast medium and is more vulnerable to security attacks than its wired counterpart due to lack of a physical boundary. In the wireless sensor domain, anybody with an appropriate wireless receiver can monitor and intercept the sensor network communications. The adversaries may use expensive radio transceivers, powerful workstations, and interact with the network from a distance since they are not restricted to using sensor network hardware. It is possible for the adversaries to identify the message source or even identify the source location, even if strong data encryption is utilized.

Source-location privacy (SLP) is an important security issue. Lack of SLP can expose significant information about the traffic carried on the network and the physical world entities. While confidentiality of the message can be ensured through content encryption, it is much more difficult to adequately address the SLP. Preserving SLP is even more challenging in WSNs since the sensor nodes consist of only low-cost and low-power radio devices, and are designed to operate unattended for long periods of time. Battery recharging or replacement may be infeasible or impossible. Computationally intensive cryptographic algorithms, such as public-key cryptosystems, and large scale broadcasting-based protocols, are not suitable for WSNs. To optimize the sensor nodes for the limited node capabilities and the application specific nature of the WSNs, traditionally, security requirements were largely ignored. This leaves WSNs vulnerable to network security attacks. In the worst case, adversaries may be able to undetectably take control of some wireless sensor nodes, compromise the cryptographic keys, and reprogram the wireless sensor nodes.

In this paper, we first propose some criteria to quantitatively measure source-location information leakage for routing-based SLP schemes. Through the proposed measurement criteria, we are able to identify security vulnerabilities of some exiting SLP schemes. We then propose a scheme that can provide both content confidentiality and SLP through a two-phase routing. In the first routing phase, the message source randomly selects an intermediate node in the sensor domain and then transmits the message to the randomly selected intermediate node (RSIN). This phase provides SLP with a high local degree. In the second routing phase, the messages will be routed to a ring node where the messages will be blended through a network mixing ring (NMR). By integrating the NMR, we can dramatically decrease the local degree and increase the SLP. Our simulation results demonstrate that the proposed scheme is very efficient and can achieve a high message delivery ratio. We believe it can be used in many practical applications.

The major contributions of this paper can be summarized as follows:

- *Y. Li is with Microsoft, 15850 NE 40th ST B208, Redmond, WA 98052. E-mail: yunl@microsoft.com.*
- *J. Ren is with the Department of Electrical and Computer Engineering, Michigan State University, 2120 Engineering Building, East Lansing, MI 48824-1226. E-mail: renjian@egr.msu.edu.*
- *J. Wu is with the Department of Computer and Information Sciences, Temple University, 1805 N. Broad St., Room 302, Wachman Hall 302, Philadelphia, PA 19122. E-mail: jiewu@temple.edu.*

Manuscript received 15 Feb. 2011; revised 30 Aug. 2011; accepted 9 Sept. 2011; published online 19 Oct. 2011.
Recommended for acceptance by Y. Liu.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2011-02-0087.
Digital Object Identifier no. 10.1109/TPDS.2011.260.

- We develop a model to quantitatively measure source-location information leakage for routing-based SLP schemes.
- We identify three criteria to measure source-location information leakage for routing-based schemes.
- We propose a two-phase routing scheme to protect routing-based source-location information.
- We provide extensive simulation results using ns-2 to demonstrate the efficiency of our proposed scheme.

The rest of this paper is organized as follows: In Section 2, related work is reviewed. The network models are described in Section 3. The proposed SLP evaluation model is presented in Section 4. Section 5 analyzes the existing schemes. Section 6 details the proposed SLP scheme. Security analysis and performance analysis are provided in Sections 7 and 8, respectively. We conclude in Section 9.

2 Related Work

In the past two decades, originated largely from Chaum's mixnet [1], a number of protocols have been proposed to provide SLP [1], [2], [3]. The mixnet family protocols use a set of "mix" servers that blend the received packets so that the communication source (including the sender and the recipient) becomes ambiguous. They rely on the statistical properties of background traffic, also referred to as *cover traffic*, to achieve the desired anonymity. However, these schemes all require public-key cryptosystems and are not suitable for WSNs.

Broadcasting-based schemes provide SLP by mixing the valid messages with the dummy messages so that they become indistinguishable to the adversaries [4]. In a practical situation, the dummy messages can be significantly more than the valid messages, which not only consumes a significant amount of the limited energy, but also increases network collisions and decreases the packet delivery ratio. Therefore, these schemes are not quite suitable for large sensor networks.

Providing SLP through dynamic routing is, in our opinion, one of the most feasible approaches in WSNs [5], [6], [7]. The main idea is to prevent the adversaries from tracing back to the source location through traffic monitoring and analysis. A representative example of a routing-based protocol is the phantom routing protocol, which involves two phases: a random walk phase and a subsequent flooding/single path routing phase. In the random walking phase, the message from the actual source will be routed to a phantom source along a random path or a designed directed path. The phantom source is expected to be far away from the actual source, which will make the actual source's location hard to be traced back by the adversaries. However, theoretical analysis shows that if the message is routed h hops randomly, it is highly possible that the distance between the phantom source and the actual source is within $h/5$. To solve this problem, directed walk, through either a sector-based or a hop-based approach, was proposed. Take the section-based directed walk, for example. The source node first randomly determines a direction that the message will be sent. This direction information is stored in the header of the message. Every forwarder on the random walk path will forward this message to a random neighbor in the same direction as the source node did so that the phantom source can be far away from the actual source. Unfortunately, once the message is captured on the random walk path, the adversaries will be able to get the direction information stored in the header of the message. Therefore, exposure of the direction information decreases the complexity for adversaries to traceback to the actual message source in the order of 2^h.

3 Network Models and Design Goals

SLP is a key security requirement for military and many civilian applications. In the asset monitoring model, WSNs can be used to monitor the activities or presence of animals in a wild animal habitat. However, the information should be kept unavailable to illegal hunters. In military intelligence networks, to protect the message source, both the message source and the routing path have to be protected from adversarial attacks.

Before we describe our proposed SLP scheme in WSNs, we will introduce the system model and adversarial model in this section to capture the relevant features of WSNs and the potential adversaries in SLP applications.

3.1 System Model

Our system is similar to the explanatory Panda-Hunter Game that was introduced in [5], [8]. In this Panda-Hunter Game, a sensor network is deployed to continuously monitor activities and locations of the animals in a wild animal habitat. Once a panda is discovered, the corresponding source node in the nearby area will observe and report data periodically to the SINK node. However, the illegal hunters, who may try to track and locate the panda, should be prevented from acquiring this kind of information. Our goal is to make it infeasible for the adversaries to determine the location of the panda by analyzing the traffic pattern and messages transmitted through the network. We make the following assumptions about our system:

- The SINK node is the only destination for messages to be transmitted to. The information of the SINK node is made public.
- Each message will include a unique *dynamic* ID corresponding to the location where this message is generated. The content of each message will be encrypted using the secret key shared between the node/grid and the SINK node.
- The sensor nodes are assumed to know their relative location. We also assume that each sensor node has the knowledge of its adjacent neighboring nodes.
- The key management, including key generation, key distribution, and key updating, is beyond the scope of this paper.

3.2 Adversarial Model

Because of the high profits related to panda hunting, the adversaries would try their best to equip themselves with advanced equipment, which means they would have some technical advantages over the sensor nodes. In this paper,

the adversaries are assumed to have the following characteristics:

- The adversaries will have a sufficient energy resource, adequate computation capability, and enough memory for data storage. On detecting an event, they could determine and move to the immediate sender by analyzing the strength and direction of the signal they received. The adversaries may also compromise some sensor nodes in the network. We also assume that the adversaries will never miss any event when they are close to the event.
- The adversaries will not interfere with the proper function of the network, such as modifying packets, altering the routing path, or destroying sensor devices, since such activities can be easily identified. However, the adversaries may carry out passive attacks, such as eavesdropping of the communications.
- The adversaries are able to monitor the traffic in an area that is important to them and get all of the transmitted messages. However, we assume that the adversaries are unable to monitor the entire network. In fact, if the adversaries could monitor the entire WSN, then they can monitor the events directly without relying on the sensor network.

3.3 Design Goals

Our design goal can be summarized as follows:

- Build a security evaluation model to facilitate the designing and analyzing of routing-based source-location protection schemes.
- The adversaries should not be able to get the source-location information by analyzing the traffic pattern.
- The adversaries should not be able to get the source-location information even if they are able to monitor a certain area of the sensor network and compromise a few network nodes.
- Only the SINK node is able to identify the source-location through the messages received. The recovery of the source-location from the received message should be very efficient.
- The length of each message should be as short as possible to save the precious sensor node power. This is because on average, transmission of one bit consumes about as much power as executing 800-1,000 instructions [9].

4 SLP Evaluation Model

Although SLP has been discussed in the literature, the research on quantitative measurement and analysis on information leakage of the source-location in routing-based schemes is largely unfolded. In this section, we define some criteria to quantitatively measure the source-location information that can be acquired from routing-based schemes. These criteria can be used to quantitatively analyze the existing routing-based schemes and also provide a theoretic foundation for a new scheme design.

In a network, an adversary may always try to derive the source-location information from a captured message through traffic analysis and/or routing traceback. We divide our analysis into three categories:

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

Traditionally, each transmitted message bears a fixed message ID. If the location of the message ID is already known, then the source location of the message can be easily determined. Otherwise, the adversary can perform a routing traceback attack for all messages with the same or even correlated ID.

To prevent correlation-based source identification, a dynamic ID-based approach [6] can be used to prevent adversaries from relating messages transmitted from each source. This can be done by requiring that each node in the network be preloaded with an *ID-hash-chain* so that a different and uncorrelated ID is attached to each message. The adversaries are no longer able to get any useful information about the source node through correlation-based source identification.

For routing traceback and reducing source node space analysis, we define two criteria to measure the source-location information leakage.

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

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

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

For a routing scheme, if *SSI* is large, it means that the message may be transmitted by many possible source nodes. On the contrary, if *SSI* is small, then the adversary can limit the possible source nodes to a small group.

Therefore, for a SLP scheme, *SSI* should be as large as possible so that the complexity for an adversary to perform an exhaustive search of the message source is maximized.

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

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

5 Analysis of the Existing Schemes

In this section, we will analyze the source privacy of some well-known routing-based schemes using our proposed evaluation criteria.

5.1 Security Evaluation for Fixed Path Routing

We assume the attacker's sensing range is the same as the regular sensor nodes in the wireless sensor domain.

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

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

Proof. This is because, for each message received, the adversary A can move one hop closer to the source node S. Since the source node S and the destination node D are only L hops apart, we only need L messages in order to fully traceback from the destination node D to the source node S. Therefore, we have

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

□

This is the least secure SLP scheme we can imagine. To increase SLP, multiple schemes have been proposed [10], [11], [12], [13], [14] though nonintersected routing paths between the source node and the destination node.

Suppose there are n disjoint routing paths between the source node S and destination node D, and the length of the n paths are: $L_1, L_2, \ldots, L_n$, 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 $SDI_i = \frac{p_i}{L_i}, i = 1, \ldots, n$. Define the overall SDI as

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

We will then have the following result:

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

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

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

Define

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

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

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

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

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

which corresponds to the minimal value of SDI. That is

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

□

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

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

messages to fully determine the location of the source node, i.e., traceback 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 identify the message source with at most $L_1 + L_2 + \cdots + L_n - (n-1)$ received messages. Therefore, to provide SLP in a network, we have to increase the total number of possible routing paths between the destination node and the source node. However, for a practical network configuration, the number of routing paths cannot be increased without limitation. This means that we will always have $SDI > 0$.

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

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

The previous analysis is based on the assumption that the adversary is able to receive and identify the messages transmitted from the actual message source, say S. For each received message, the adversary is able to move one hop closer to the message source in the fixed routing path. In the case that the adversary is unable to correlate the messages received with the message source, and use all the messages to find the actual message source node, the recovery of the source location can be much more difficult. This is because the message received from the other message source may mislead the adversary to move away from the source node S. In this case, the SDI becomes 0 since no information can be linked to the message source S. In particular, Theorem 1 and Corollary 1 are no longer applicable.

5.2 Security Evaluation for Dynamic Routing Path

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

From an adversary's point of view, on corrupting a message in the random walk path, the adversary is able to move one hop closer to the actual source node. However, because the routing path is dynamic, the possibility that the adversary can receive another message sent from the same source may be very small, especially for a large scale network. In other words, the correlation between the source node and the message received in the random path can be viewed as zero, i.e., $SDI \simeq 0$. However, the direction information stored in the message header can facilitate the adversary to narrow the possible area of the source node. Take the section-based random walk as an example. Once a message is corrupted by an adversary on the random path, the adversary can determine to which direction of the current location the actual source node is located.

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

In addition to the $NSSI$-based attacks, in phantom routing, 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 it makes it possible for the adversaries to monitor and link all messages from the same source node together, which may help the adversaries to identify the source location since the IDs correspond to the grids' locations. Whenever the adversaries discover a message sent from a grid with an ID that they already know, they can use this message to move closer to the message source.

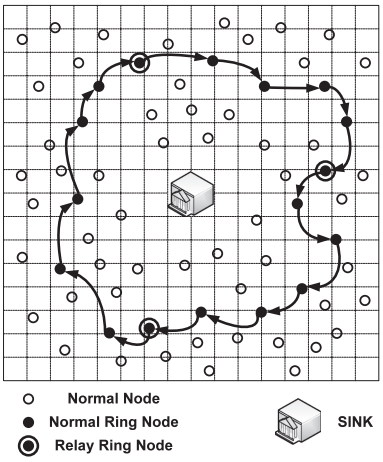

Fig. 1. Grids formation.

6 Proposed Schemes

In this section, we will present our proposed SLP schemes. First, to prevent the adversaries from getting any useful source-location information through *correlation-based source identification*, a dynamic ID proposed in [6] should be used for each message. Then, we introduce a two-phase routing protocol to provide SLP and content confidentiality. In the first phase, the source node routes the messages to a ring node through a single randomly selected intermediate node (RSIN) in the sensor domain before the message is routed to the mixing ring. Although this phase can provide a good SDI, the local degree is still large. In the second phase, the message from the first phase will be forwarded to the network mixing ring (NMR) [6]. The combination of these two phases guarantees the local degree to be small, therefore, providing a high degree of SLP.

In our scheme, the network is evenly divided into small grids, as shown in Fig. 1. The formation of the grid and the header node selection in each grid have been studied in many literature works [6]. We assume that the sensor nodes in each grid are all within the direct communication range of each other. In each grid, the header node coordinates the communication with other header nodes nearby. We assume that the whole network is fully connected through the multihop communications.

After the formation of all the grids, a large ring is generated in the sensor network to provide a network-level traffic mix. This ring is called the *mixing ring*. The mixing

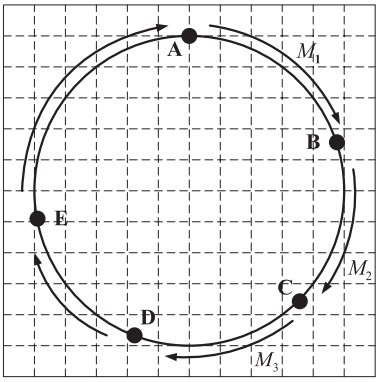

Fig. 2. Message transmission in the ring.

ring is composed of multiple header nodes. We call these header nodes *ring nodes*. The ring nodes are further divided into *relay ring nodes* and *normal ring nodes*. The messages transmitted in the mixing ring are referred to as *vehicle messages*. Vehicle messages will be transmitted in the ring in a clockwise direction, called *ring direction*. Only relay ring nodes can generate vehicle messages. We also define the grids containing ring nodes as *ring grids*. Correspondingly, the grids without ring nodes are called *normal grids*. The sensor nodes in normal grids are defined as *normal nodes*. The messages sent by the normal nodes are referred to as *messages*. When a normal node has a message to transmit, the message will first be sent to the header node in that grid. The header node will then forward this message to a randomly selected intermediate node before it is forwarded to a ring node. The ring transmission provides a network-level traffic mix. The detailed description of the proposed two-phase routing will be described in the subsequent sections.

6.1 Key Management

We require two kinds of keys in our scheme:

- *Grid-key* K_{G_i}*:* the key shared between grid G_i and the SINK node.
- *Ring-key* K_{AB}*:* the key shared between ring grid A and ring grid B.

The grid-keys are used to provide message content confidentiality. When the ith normal grid has a message m to transmit, the message is first encrypted using the grid-key: K_{G_i}, then its dynamic ID $id_j^{(i)}$ is prefixed to the encrypted message. Therefore,

$$Msg = id_j^{(i)} \| E_{K_{G_i}}(m),$$

will be transmitted from the source node to the SINK node, where $E_{K_{G_i}}(m)$ is the ciphertext of m, encrypted using the secret key K_{G_i} shared between the ith grid with dynamic ID $id_j^{(i)}$ and the SINK node.

On receiving a message Msg, the SINK node identifies the source grid and decrypts the message Msg to recover m.

Fig. 2 gives an example of a mixing ring, where A, B, C, D, E are the ring nodes. Each ring node shares a secret key with its predecessor ring node and a secret key with its successor ring node. For instance, ring node B shares a key K_{AB} with node A, and a key K_{BC} with node C.

6.2 Routing to a Single Intermediate Node

As described before, phantom routing has no control over the phantom source without leaking significant side information. To solve this problem, in the proposed protocol, the message source first randomly selects an intermediate node at the sensor domain based on the relative location of the sensor node, defined in the grid shown in Fig. 1. The intermediate node is expected to be far away from the actual source node so that it is difficult for the adversaries to get the information of the actual source node from the intermediate node selected.

Since we assume that each sensor node only has knowledge of its adjacent nodes. The source node has no accurate information of the sensor nodes more than one hop away. In particular, the randomly selected intermediate node may not even exist. However, the relative location can guarantee that the message will be forwarded to the area of the intermediate node. The last node in the routing path adjacent to the intermediate node should 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. The intermediate node then routes the received message to a ring node.

Suppose the source node is located at the relative location (x_0, y_0). To transmit a 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 needs to generate a d_{rand}, it first generates a random number x, which 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

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

The probability 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, and φ_{0,σ^2} is the probability density Gaussian function [15]. The cumulative distribution function (CDF) $\Phi(0, \sigma^2)$ of $N(0, \sigma)$ is defined as follows [16]:

$$\begin{aligned}\Phi_{0,\sigma^2}(x) &= \int_{-\infty}^{x} \varphi_{0,\sigma^2}(u)\, du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\Big(-\frac{u^2}{2\sigma^2}\Big)\, du \\ &= \Phi\Big(\frac{x}{\sigma}\Big).\end{aligned}$$

If we choose $\sigma = 1$, then the probability that d_{rand} falls within the interval $[d_{min}, 2d_{min})$ will be $2\Phi(\frac{1}{1}) - 1 = 0.6827$. If we choose $\sigma = 2$, then we get the probability that d_{rand} is in the interval $[d_{min}, 3d_{min})$ to 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

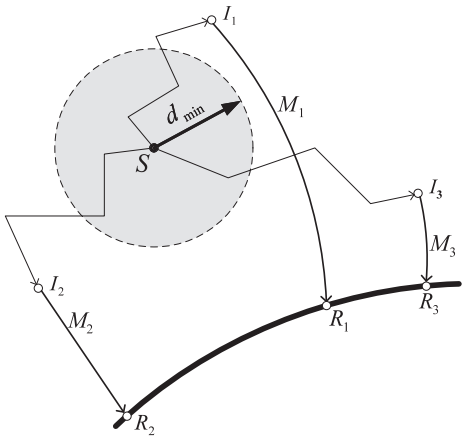

Fig. 3. Illustration of the two-phase routing.

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

Upon receiving the message, the intermediate node forwards the message to the closest ring node.

An example is given in Fig. 3, 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.

Unlike the directed walk proposed in phantom routing, in our proposed RSIN scheme, the selection of the intermediate nodes is entirely random. Therefore, it does not have the security weaknesses of phantom routing, as discussed before. More security analysis will be provided in Section 7.

6.3 Network Mixing Ring (NMR)

In the second routing phase, the messages will be forwarded hop-by-hop in the *mixing ring*. The network mixing ring is a logic ring established by selecting a set of hop-by-hop connected grids that can form a ring. The ring nodes can be either regular nodes or special nodes. In the case that the ring nodes are regular nodes, the ring nodes in the selected grid can take turn to be the ring node to achieve energy balance.

In the mixing ring, each *ring node* can route messages toward its *successor* in the *ring direction*, which is the next hop node in a clockwise direction. The message can hop along the ring direction for a random number of hops before it is transmitted to the SINK node. This routing process provides SLP 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, it would be infeasible for the adversary to identify the source location of the actual message.

In the mixing ring, only the relay ring nodes can initiate the vehicle messages starting with dummy messages, and deliver the vehicle messages to the SINK node. The normal ring nodes can store and forward messages received from the normal node to its successor ring node. The relay ring nodes can be either more powerful than or the same as the normal ring nodes.

Each vehicle message may contain several message units. These units are left unused initially. If a unit in the vehicle message is not used, we name this unit as *dummy unit*, composed of any fixed data structure, such as all 0s. The length of a unit is the same as the message sent by a normal node. Upon receiving a vehicle message, if a normal ring node has an actual message received and there is still a dummy unit in the vehicle message, it can replace this dummy unit with the message. The updated vehicle message will then be forwarded to its successor ring node. If it has not received any messages from the normal nodes, or there is no dummy units left in the vehicle message, it simply forwards this vehicle message.

In our scheme, to thwart the 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.

When an encryption algorithm is used, it is computationally infeasible for the adversary to find the correlation between the input and output of each node. The vehicle message should be sent at a rate which can ensure that all the messages are embedded in vehicle messages and forwarded to the SINK with minimum delay.

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

7 Security Analysis

We will first analyze the SLP from routing to a single RSIN in the first routing phase. We assume the adversary is unable to monitor the entire sensor area of the source node, since otherwise it can monitor the actual event directly.

In our RSIN, the intermediate node is randomly selected by the source node. From a probability point of view, every node away from the source node can be selected as the intermediate node. Moreover, the probability for a node to be selected as the intermediate node multiple times is negligible in large sensor networks.

If an adversary tries to traceback to the source location from the message in the route path through which the packet is being transmitted to the SINK node. To the best extend, the adversaries will be led to the randomly selected intermediate node, instead of the actual message source. Since the intermediate node is randomly selected for each message, the probability that the adversaries will receive the messages from one source node continuously is negligible.

Even if one intermediate node's location is discovered by the adversaries, the source location still cannot be identified because the locations of the intermediate nodes are at least d_{min} away from the actual source node. Therefore, the probability for the same routing path to be selected for multiple events from the same source is negligible for large sensor networks. In other words, we have

$$SSI = 0,$$

with negligible exception.

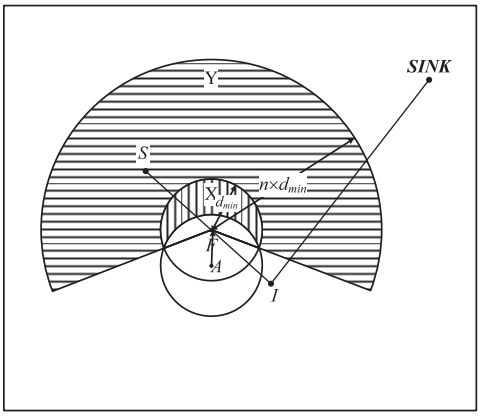

Fig. 4. Local source-location privacy: an adversary receives a message in scenario 1.

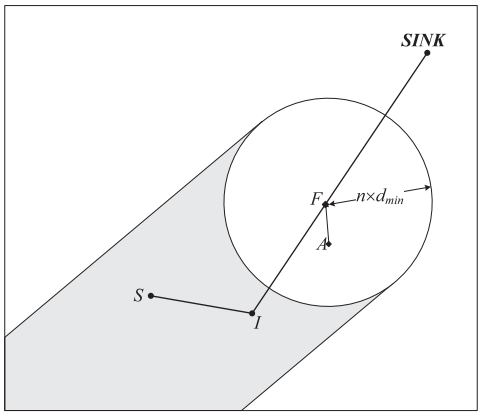

Fig. 5. Local source-location privacy: an adversary receives a message in scenario 2.

However, when an adversary intercepts a message in the first routing phase, that is when the message is routed to a single RSIN, there are two possible scenarios: 1) the adversary receives the message while the message is being transmitted from the source node to the randomly selected intermediate node, and 2) the message is being transmitted from the intermediate node to the SINK node.

For scenario 1) when an adversary at location A receives a message, according to our assumption, he can immediately find the message forwarder node F in Fig. 4. We also know the probability that the source node being in the area with radius $3d_{min}$ to node F is 95.45 percent, and 99.73 percent for radius $4d_{min}$. In addition, if this is the only message that the adversary receives, the next hop node can only be located in area X, which is the areas with vertical lines. In this way, there is a very high probability that the actual source node is located in area X or area Y, which is the area with horizontal lines.

For scenario 2) upon receiving a message, there is a high probability that the adversary is also able to limit the intermediate node and the actual message source node to a ribbon area, shown in Fig. 5. This is because when the adversary at location A receives a message, he can immediately identify the message forward node F. If F is the randomly selected immediate node, then the actual message source will be located in the circle with radius $3d_{min}$; otherwise, based on the direction information and the selection of intermediate nodes, we can derive with high probability that the actual message source is located in the shaded area.

In this way, though our protocol does not directly leak information to the adversaries, the adversary can still narrow the actual message source node to a small area with high probability. Therefore, in both cases, routing through a single RSIN results in a source-location privacy with a high local degree.

However, if we integrate the NMR as the second phase in the routing, the local degree of the SLP becomes negligible. In fact, we have the following theorem:

Theorem 2. *It is computationally infeasible for an adversary to distinguish the message initiator and message forwarder in the mixing ring.*

Proof. (Sketch) As we have described, all message transmissions in the mix ring are encapsulated into a vehicle message and encrypted using the secret key shared between the ring nodes. Particularly, when a ring node B in Fig. 2 receives a message M_1 from its predecessor node A, it transmits message M_2 to its subsequent node C, where M_2 is generated from M_1: $M_2 = E_{K_{BC}}(m)$, and $m = \{D_{K_{AB}}(M_1)\}$ if M_1 is an actual message; otherwise, m can be any message initiated by node B. K_{AB}, K_{BC} are the secret keys shared between A, B and B, C, respectively. The plaintext encapsulated in M_1 and M_2 can be either the same or different, based on whether node B has embedded its own messages. However, it is infeasible for the adversaries to get this information and derive the correlation between these two vehicle messages. This is guaranteed by the diffusion property of the encryption algorithm. Therefore, the adversaries are unable to distinguish whether the ring node has embedded its own message in the updated vehicle message. Consequently, it is computationally infeasible for the adversary to distinguish the message initiator and message forwarder node in the mixing ring. □

It should be point out that without a hop-by-hop message encryption, by comparing the vehicle message that a node receives and transmits, the adversary can determine whether a message has been loaded into the updated vehicle message. In this way, when an adversary receives a message while the message is being transmitted from a ring node to the SINK node, the actual message source node can be anywhere in the entire sensor domain. Therefore, we have $NSSI = 1$ and the local degree is 0.

For large sensor networks, the probability for an adversary to intercept a message while the message is being transmitted to the mixing ring is negligible. Therefore, the local degree is 0 with negligible exception.

The hop-by-hop encryption technique can also be used in the RSIN phase. However, it is not as critical as it is for the mixing ring. The extra energy budget cannot be easily justified due to the limited benefit. It is also possible to have multiple mixing rings. In fact, mixing rings can also be used to provide local SLP. However, this part will not be considered in this paper.

While being able to provide SLP for WSNs, our proposed scheme is also quite energy efficient. The

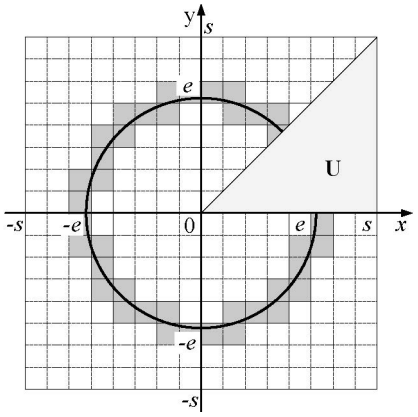

Fig. 6. Optimal NMR determination.

performance of the proposed scheme will be further analyzed in the next section.

8 Performance Analysis and Simulation Results

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

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

where $\mathcal{E}$ denotes the energy consumption, α is a constant parameter, and d is the distance of the transmission. Fig. 6 gives an example of a target area of size $2\text{s} \times 2\text{s}$. The shaded grids are selected as the ring grids. The line in the middle of the shaded area is indicated by the solid line. Assume the density of the sensor nodes in the sensor network is λ, then the total energy consumption for each sensor node in this area to transmit one message to its ring node can be calculated as follows:

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

where $\mathcal{E}_U$ is the energy consumption for area U, as demonstrated in Fig. 6. To find the e for minimum energy consumption, we let

$$\frac{d\mathcal{E}_{total}}{de} = 8\alpha\lambda \int_0^{\pi/4} \int_0^{s/\cos\theta} (-2) \times (r-e) r dr d\theta = 0.$$

We can derive

$$e = \frac{s}{3}\sqrt{2} - \frac{s}{6}\ln(2) + \frac{s}{3}\ln(2+\sqrt{2}) \approx 0.765\ \text{s}.$$

In this way, we derive the optimal ring location.

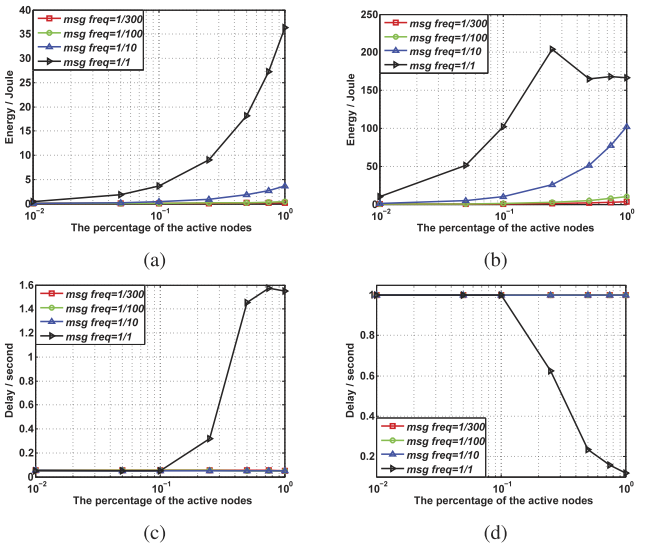

Fig. 7. Performance of the proposed routing and encryption scheme: (a) Power consumption of normal nodes; (b) Power consumption of ring nodes; (c) Message latency; (d) Message delivery ratio.

In a practical application, for large sensor networks, it is possible that only a small fraction of the sensor nodes in the sensor network has events to report at any given time. We name these nodes *active nodes*. We also define two parameters in our simulation: τ, the number of messages a normal node generates in each second, and a, the active node ratio.

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

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

messages.

If vehicle messages are L bits long, the number of vehicle messages generated by a ring node in one second is

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

Since only the relay ring nodes on the ring can generate vehicle messages, if there are n relay ring nodes in the ring, then each relay ring node needs to generate

$$\frac{gla\tau}{\gamma L} \times \frac{\gamma}{n} = \frac{gla\tau}{nL},$$

vehicle messages each second.

We conduct simulations using ns-2 on a Linux system to measure the performance of our proposed scheme. The results are provided in Fig. 7 to demonstrate the power consumption (for both normal nodes and ring nodes), message latency, and message delivery ratio of the proposed scheme.

In the simulation, the target area is a square field of size $8{,}000 \times 8{,}000$ meters. We partition this field into 2,400 normal grids/nodes. The parameters are selected as follows:

1. $\gamma = 80$, i.e., the mixing ring is composed of 80 grids;
2. $d_{min} \geq 600$, i.e., the randomly selected intermediate node is at least 600 meters away from the actual message source;
3. $n = 4$, i.e., there are four relay ring nodes in the mixing ring;
4. $l = 8$, i.e., the messages are 8 bits long; and
5. $L = 16$, i.e., the vehicle messages are 16 bits long.

From Figs. 7a and 7b, we can see that ring nodes consume more energy than normal nodes. To solve this problem, the nodes in ring grids can take turns being ring nodes. It is also noticed that the delivery ratio drops exponentially when the traffic volume increases. This is primarily due to increased traffic collisions and packet losses. For large sensor networks, the percentage of concurrent active nodes may be very low. The transmission frequency also tends to be low. In other words, the traffic volume may be low. In this scenario, we can ensure almost 100 percent delivery ratio, as shown in Fig. 7d.

A comparison of the performance of our proposed and other existing schemes can be found in reference [17]. These simulation results demonstrate that the proposed scheme is very efficient and is suitable for practical applications.

9 Conclusions

SLP is critical to the successful deployment of WSNs for many applications. In this paper, we have proposed some criteria to quantitatively measure SLP for routing-based schemes. Based on these criteria, we have proposed a scheme that can achieve SLP in WSNs through a two-phase routing: routing to a single RSIN and routing through the NMR. The optimal location for the mixing ring is also derived. Our proposed scheme provides provable local SLP and global SLP. Simulation results demonstrate that the proposed scheme can achieve very good performance in energy consumption and message delivery latency, while assuring a high message delivery ratio.

Acknowledgments

This research was supported in part by US National Science Foundation (NSF) grants CNS-0845812, CNS-0848569, CNS-1050326, CNS-1117831, CCF-1028167, CNS-0948184.

References

[1] D. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Comm. the ACM,* vol. 24, pp. 84-90, Feb. 1981.
[2] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous Connections and Onion Routing," *IEEE J. Selected Areas in Comm.*, vol. 16, no. 4, pp. 482-494, May 1998.
[3] M. Reiter and A. Rubin, "Crowds: Anonymity for Web Transaction," *ACM Trans. Information and System Security,* vol. 1, no. 1, pp. 66-92, 1998.
[4] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks," *Proc. IEEE INFOCOM '08,* pp. 51-55, Apr. 2008.
[5] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source-Location Privacy in Sensor Network Routing," *Proc. IEEE 25th Int'l Conf. Distributed Computing Systems (ICDCS '05),* pp. 599-608, June 2005.
[6] Y. Li and J. Ren, "Preserving Source-Location Privacy in Wireless Sensor Networks," *Proc. IEEE SECON '09,* June 2009.
[7] Y. Li and J. Ren, "Source-Location Privacy through Dynamic Routing in Wireless Sensor Networks," *Proc. IEEE INFOCOM '10,* Mar. 2010.
[8] http://www.panda.org/, 2011.
[9] J. Hill, R. Szewczyk, S.H.A. Woo, D. Culler, and K. Pister, "System Architecture Directions for Networked Sensors," *Proc. ACM Ninth Int'l Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS IX),* Nov. 2000.
[10] J.W. Suurballe, "Disjoint Paths in a Network," *Wiley Periodicals,* vol. 4, pp. 125-145, 1974.
[11] J.W. Suurballe and R.E. Tarjan, "A Quick Method for Finding Shortest Pairs of Disjoint Paths," *Wiley Periodicals,* vol. 14, pp. 325-336, 1984.
[12] R. Bhandari, "Optimal Physical Diversity Algorithms and Survivable Networks," *Proc. IEEE Second Symp. Computers and Comm.,* pp. 433-441, 1997.
[13] A. Srinivas1 and E. Modiano, "Finding Minimum Energy Disjoint Paths in Wireless Ad-Hoc Networks," *Wireless Networks,* vol. 11, pp. 401-417, July 2005.
[14] H. Wang, B. Sheng, and Q. Li, "Privacy-Aware Routing in Sensor Networks," *Computer Networks,* vol. 53, no. 9, pp. 1512-1529, 2009.
[15] S.M. Stigler, *Statistics on the Table.* Chapter 22. Harvard Univ. Press, 2002.
[16] N. Sematech, "Engineering Statistics Handbook," http://www.itl.nist.gov/div898/handbook/eda/section3/eda362.htm#CDF, 2011.
[17] Y. Li, L. Lightfoot, and J. Ren, "Routing-Based Source-Location Privacy Protection in Wireless Sensor Networks," *Proc. IEEE Int'l Conf. Electro/Information Technology (EIT '09),* June 2009.

Yun Li received the BE degree from Xidian University in 2005, and the PhD degree in electrical and computer engineering from Michigan State University in May 2010. He joined the Network Operating System Technology Group (NOSTG) of Cisco System in 2010. His current research interests include wireless sensor networks and network security.

Jian Ren received the BS and MS degrees both in mathematics from Shaanxi Normal University, and the PhD degree in EE from Xidian University, China. He is an associate professor in the Department of ECE at Michigan State University. His current research interests include cryptography, network security, energy efficient sensor network security protocol design, privacy-preserving communications, and cognitive networks. He is a recipient of the US National Science Foundation (NSF) Faculty Early Career Development (CAREER) award in 2009. He is a senior member of the IEEE.

Jie Wu is the chair and a professor in the Department of Computer and Information Sciences, Temple University. Prior to joining Temple University, he was a program director at US National Science Foundation (NSF). His research interests include wireless networks and mobile computing, routing protocols, fault-tolerant computing, and interconnection networks. He serves in the editorial board of the *IEEE Transactions on Computers* and *Journal of Parallel and Distributed Computing*. He is a program cochair for IEEE INFOCOM 2011. He was also general cochair for IEEE MASS 2006, IEEE IPDPS 2008, and DCOSS 2009. He is serving as an ACM distinguished speaker and is the chairman of the IEEE Technical Committee on Distributed Processing (TCDP). He is a fellow of the IEEE.