Preserving Source-Location Privacy in Wireless Sensor Networks

Yun Li and Jian Ren
Department of Electrical and Computer Engineering
Michigan State University
East Lansing, MI 48824
Email: {liyun1, renjian}@egr.msu.edu

***Abstract*—Wireless sensor networks (WSN) have the potential to be widely used in many areas for unattended event monitoring. Mainly due to lack of a protected physical boundary, wireless communications are vulnerable to unauthorized interception and detection. Privacy is becoming one of the major issues that jeopardize the successful deployment of wireless sensor networks. While confidentiality of the message can be ensured through content encryption, it is much more difficult to adequately address the source-location privacy. For WSN, source-location privacy service is further complicated by the fact that the 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 not suitable for WSN. In this paper, we propose a scheme to provide both content confidentiality and source-location privacy through routing to a randomly selected intermediate node (RRIN) and a network mixing ring (NMR), where the RRIN provides local source-location privacy and NMR yields network-level (global) source-location privacy. While being able to provide source-location privacy for WSN, our simulation results also demonstrate that the proposed scheme is very efficient and can be used for practical applications.**

I. Introduction

Wireless sensor networks 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 that 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.

Location privacy is an important security issue. Lack of location privacy 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 source-location privacy. Privacy service in WSN is further complicated 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. Therefore, computationally intensive cryptographic algorithms, such as public-key cryptosystems, and large scale broadcasting-based protocols, are not suitable for WSN. This makes privacy preserving communications in WSN an extremely challenging research task. To optimize the sensor nodes for the limited node capabilities and the application specific nature of the networks, traditionally, security requirements were largely ignored. This leaves the WSN under network security attacks. In the worst case, adversaries may be able to undetectably take control of some sensor nodes, compromise the cryptographic keys and reprogram the sensor nodes.

In this paper, we propose a scheme that provides both content confidentiality and source-location privacy through a two-phase routing process. In the first routing phase, the message source randomly selects an intermediate node in the sensor domain and then transmits the data packet to the randomly selected intermediate node before it is routed to a ring node. This phase provides the local source-location privacy.

In the second routing phase, the data packet will be mixed with other packets through a network mixing ring (NMR). This phase offers network-level (global) source-location privacy. While being able to provide source-location privacy for WSN, our simulation results also demonstrate that the proposed scheme is very efficient and can be used practical applications.

The rest of this paper is organized as follows: In Section II, the related works are reviewed. The system model is described in Section III. In Section IV, the proposed dynamic ID and key management are presented. Section V details the proposed source-location privacy scheme. Security analysis and performance analysis are provided in Section VI and Section VII, respectively. We conclude in Section VIII.

II. Related Works

In the past two decades, originated largely from Chaum's mixnet [1] and DC-net [2], a number of source-location private

978-1-4244-2908-0/09/$25.00 ©2009 IEEE

communication protocols have been proposed [1], [3]–[12]. The mixnet family protocols use a set of "mix" servers that mix the received packets to make the communication source (including the sender and the recipient) ambiguous. They rely on the statistical properties of background traffic that is also referred to as the *cover traffic* to achieve the desired anonymity. The DC-net family protocols [2], [6], [13] utilize secure multiparty computation techniques. However, both approaches require public-key cryptosystems and are not suitable for WSN.

Source-location privacy is also provided through broadcasting that mixes valid messages with dummy messages [14], [15]. The main idea is that each node needs to transmit messages consistently. Whenever there are no valid messages, the node has to transmit dummy messages. The rate of the broadcasting is either fixed or probabilistic. In practical situation, the dummy messages could be several magnitudes higher than the valid messages. The goal of broadcasting based schemes is to make it infeasible for the adversaries who are able to monitor the traffic of the entire network to perform traffic analysis and distinguish valid messages from dummy messages. The transmission of dummy messages not only consumes significant amount of the limited energy in the sensor nodes, but also increases the network collisions and decreases the packet delivery ratio. Therefore, these schemes are not quite suitable for large sensor networks.

Routing based protocols can also provide source-location privacy [16]–[18]. The main idea is to prevent the adversaries from tracing back to the source-location through traffic monitoring and analysis. A representing example of routing based protocol is the phantom routing protocol [16]. Phantom routing involves two phases: a random walk phase and a subsequent flooding/single path routing phase. In the random walking phase, the message from the real 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 real source, which will make the real source's location hard to be traced back by the adversaries. However, theoretical analysis shows that if the message is routed h hops randomly, then it is highly possible that the distance between the phantom source and the real source is within $h/5$. To solve this problem, directed walk was proposed [16]. Directed walk can be achieved either through sector-based directed random walk or hop-based directed random walk. Take the section-based directed walk for example, the source node first randomly determines a direction that the message will be sent to. This 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 as the source node did. In this way, the phantom source will be far away from the real source. Unfortunately, once the message is captured on the random walk path, the adversaries will be able to get the direction information stored in the header of the message. Therefore, the exposure of direction information decreases the complexity for adversaries to trace back to the true message source in the magnitude of 2^h.

III. Models and Design Goals

Source-location privacy is a key security requirement for many civilian and military applications. In the asset monitoring model, wireless sensor networks can be used to monitor the activities or presence of animals in a wild animal habitat. However, this information should be kept unavailable to the illegal hunters. In military intelligence network, to protect the message source and also the messages delivery process, both the source node and the routing path have to be protected from the adversarial attack.

In order to facilitate the description and analysis of the proposed source-location privacy scheme in wireless sensor networks, we will introduce the system model and adversarial model in this section to capture the relevant features of wireless sensor networks and potential adversaries in source-location privacy applications.

A. The System Model

Our system is similar to the explanatory Panda-Hunter Game that was introduced in [16], [19]. In this Panda-Hunter Game, a sensor network is deployed to continuously monitor activities and locations of the animals in a wild animal habitat.

As soon as 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 knowing 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 that data messages will be transmitted to. The information of the SINK node is made public. On detecting an event, a sensor node will generate and send messages to the SINK node through a multi-hop routing.
- Each message will include a unique node ID corresponding to the location where this message is generated. The content of each message will be encrypted using the shared secret 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 information about the relative location of the sensor domain may also be broadcasted through this network for routing information update [20]–[22].
- The key management, including key generation, key distribution and key update, is beyond the scope of this paper. However, the interested readers are referred to references such as [23]–[26].

B. The Adversaries Model

Because of the high profits related to panda hunting, the adversaries would try their best to equip themselves with advanced equipments, 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 sufficient energy resource, adequate computation capability and enough memory for data storage. On detecting an event, they could determine the immediate sender by analyzing the strength and direction of the signal they received. They can move to this sender's location without too much delay. 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 functioning 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 wireless sensor networks, then they can monitor the events directly without relying on the sensor network.

C. Design Goals

Our design goal can be summarized as follows:

- The adversaries should not be able to get the source-location information by analyzing the traffic pattern.
- The adversaries should not be able to get the source-location information even if they are able to monitor certain area of the sensor network and compromise a few network nodes.
- Only the SINK node is able to identify the source-location through the messages received. The recovery of the source-location from the received message should be very efficient.
- The length of each message should be as short as possible to save the previous sensor node power. This is because that on average, transmission of one bit consumes about as much power as executing 800-1000 instructions [27].

D. Overview of the Proposed Scheme

In our scheme, the network would be 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 [28]–[31]. We assume that the sensor nodes in each grid are all within the direct communication range of each other. In each grid, the header node coordinates the communication with other header nodes nearby. We assume that the whole network is fully connected through the multi-hop communications.

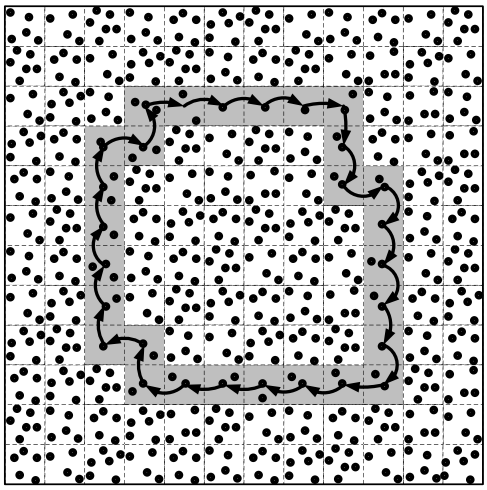


Fig. 1. Grids Formation

After the formation of all the grids, a large ring is generated in the sensor network to provide network-level traffic mix. This ring is called the *mixing ring*. The mixing 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 that will be transmitted in the mixing ring are referred to as *vehicle messages*. Vehicle messages will be transmitted in the ring in the clockwise direction, called *ring direction*. Only relay ring nodes can generate vehicle message. We also define the grids containing ring node as *ring grids*. 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 as referred as *data messages*. When a normal node has an event message to transmit, the message will first be transmitted to the header node in that grid. The header node will then forward this message to a randomly selected intermediate node before it is being 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.

IV. Proposed Dynamic ID and Key Management

A. Dynamic ID Assignment

In [16], 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 of the network 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 *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 is: $\{id_1^{(i)}, id_2^{(i)}, id_3^{(i)}, \cdots, id_{n-1}^{(i)}, id_n^{(i)}\}$, generated as follows:

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

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 computational infeasible to compute id_j from id_{j+1}. As a result, the adversaries can not 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-chains, can correlate the ID of the corresponding source grid with the source-location. In this way, transmitting dynamic IDs with messages will not leak any source-location information to the 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, the length should be as short as possible in practical applications. We propose to use only part of the hash function output as the ID. In addition to efficiency, this design also makes it infeasible for the adversary to link the dynamic IDs of the same node together. 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 the adversaries to perform source-location analysis based on the sensor node ID. 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

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

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

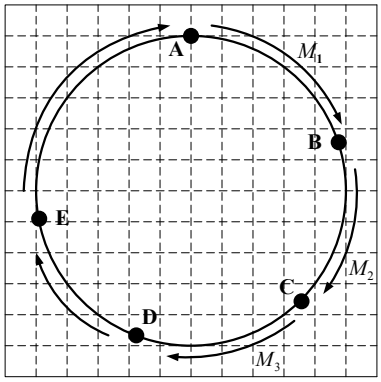


Fig. 2. Message transmission in the ring

B. Key Management and Communication Privacy

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 between ring grid A and ring grid B.

The application of these two keys will be described in the following sections.

1) Grid-Key: The Grid-Keys are used to provide message content confidentiality. When the i^{th} normal grid has a data message m to send, it first encrypts this message using its grid-key: K_{G_i}, then it appends its current ID: $id_j^{(i)}$ to the encrypted message. Finally,

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

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 SINK node and the i^{th} grid with dynamic ID $id_j^{(i)}$.

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

2) Ring-Key: Fig. 2 gives an example of a mixing ring. 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.

V. Proposed Source-Location Privacy Scheme

In this paper, we propose a two-phase routing protocol to provide source-location privacy and content confidentiality. In the first phase, the source node routes the messages to a ring node through a random intermediate node (RRIN) selected from the sensor domain before the message is being routed to the mixing ring. In the second phase, the message packet from the first phase will be forwarded through the network mixing ring (NMR). The RRIN provides local source-location privacy, while NMR offers the network-level source-location privacy.

A. Routing through a Random Intermediate Node (RRIN)

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 goal of this phase is to provide local source-location privacy. The intermediate node is expected to be far away from the real source node so that it is difficult for the adversaries to get the information of the real source 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 packet 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 and the first phase routing is accomplished.

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 will first generate a random number x. The value of this random variable is normally distributed with mean 0 and variance σ^2, i.e., $X \sim N(0, \sigma)$. Then the source node can calculate d_{rand} as follows:

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

Therefore, the probability [32] 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 which is the Gaussian function [33].

The cumulative distribution function (CDF) $\Phi(0, \sigma^2)$ of $N(0, \sigma)$ is defined as follows [34]:

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

If we choose σ to be 1.0, then the probability that d_{rand} falls within the interval $[d_{min}, 2d_{min})$ will be $2\Phi(\frac{1}{1}) - 1 = 0.6827$.

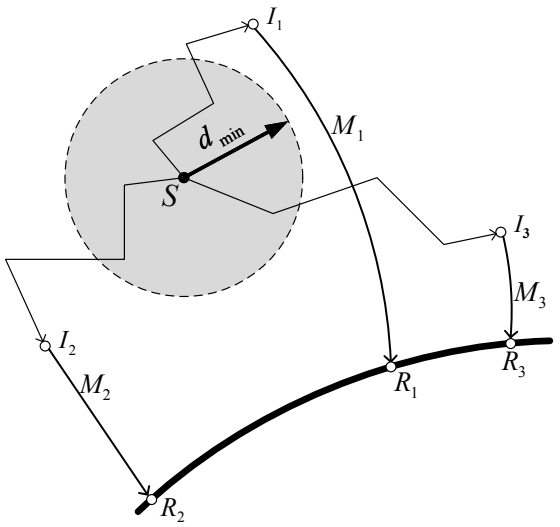


Fig. 3. Illustrate of the two-phase routing

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 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 RRIN scheme, the selection of the intermediate node is entirely random. Therefore, it does not have the security drawbacks of phantom routing discussed before. More security analysis will be provided in Section VI.

B. Network Mixing Ring (NMR)

In the second routing phase, the messages will be forwarded hop-by-hop in the ring. This ring is referred to as the *mixing ring*. Each node in the ring, refer to as a *ring node*, can route messages towards its *successor*, that is the next hop in the clockwise direction of the ring. This direction is referred to as the *ring direction*. 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. As an example, it would be infeasible for the adversaries to tell the real source of a message that reaches node C in Fig. 2 since every node could be the possible source node. Therefore, our goal is to design security mechanisms such that it is infeasible for anyone to distinguish the message source node from the message forwarding node.

In the mixing ring, only a few nodes can initiate the ring transmission starting with dummy messages, called *vehicle messages*, and delivery the ring messages to the SINK node. These nodes are referred to as *relay ring nodes*. All other nodes are referred to as *normal ring nodes*. The normal ring nodes can store and forward data message received from the normal node to its successor ring node. The relay ring nodes could be either more powerful than or the same as the normal ring nodes.

These vehicle messages may contain several data units. These units are left unused initially. If a unit in the vehicle message is not used, we name this unit as *dummy unit*, composed of any fixed data structure, such as all 0s. The length of a unit is the same as the data message sent by a normal node. Upon receiving a vehicle message, if a normal ring node has a real data message received and there is still a dummy unit in the vehicle message, it can replace this dummy unit with the data message. The updated vehicle message will then be forwarded to its successor ring node. If it 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.

In our scheme, to thwart message source analysis, the message transmission in the ring is encrypted. Each ring node shares a secret key with its predecessor ring node and a secret key with its successor ring node. As an example, in Fig. 2, ring node B shares a key K_{AB} with ring node A and a key K_{BC} with ring node C. When node B receives a packet M_1 from node A, it first decrypts M_1 using the share secret key K_{AB}. Let $m_1 = D_{K_{AB}}(M_1)$. Upon decryption, node B will be able to find the dummy unit(s) in m_1 and replace the dummy unit(s) with the data message(s) that it received from the normal nodes. Denote the updated message as $\{D_{K_{AB}}(M_1)\}$. The updated vehicle message will be encrypted using the shared secret K_{BC} before it is transmitted to the node C. Denote the message that generated in node B as M_2, then we have

$$M_2 = E_{K_{BC}}(\{D_{K_{AB}}(M_1)\}). \quad (1)$$

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. 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.

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

VI. Security Analysis

We will first analyze that the proposed routing to a random intermediate node (RRIN) in phase one can provide local source-location privacy. RRIN is not like phantom routing, which has no control over the phantom source without leaking significant side information. We assume the adversaries are unable to monitor the entire sensor area of the source node since it makes little sense for the adversaries to monitor the routing of the sensor node if the adversaries can monitor the real event.

In our RRIN, the intermediate node is randomly selected by the source node. From probability point of view, every node away from the source node could be selected as the intermediate node. In fact, the source node selected intermediate node may not even exist since according to our assumption, the source node does not have full knowledge of the sensor node more than one hop away. Therefore, based on our assumption, it is impossible for the adversaries to trace back and identify the real message source based on an individual traffic monitoring. This is because the probability for multiple events from the same source to use the same routing path and intermediate node is very low for large sensor networks.

If an adversary tries to trace back the source-location from the message packet in the route through which the packet is being transmitted to the mixing ring, then the adversaries will be led to the randomly selected intermediate node to the best extend, instead of the real message source. Since the intermediate node is randomly selected for each data message, the probability that the adversaries will receive the messages from one source node continuously is virtually impossible. As shown in Fig. 3, if the adversaries receive M_2 forwarded by I_2, it would be led to I_2. However, the next intermediate node I_3 is far from I_2, so the adversaries could not receive M_3.

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

Unlike the directed walk used in random walk, our protocol does not leak side information to the adversaries. Since the intermediate node is determined before each data message is transmitted by the source-location, the data message carries no observable side information of the message source-location in its content due to message content encryption and the application of dynamic ID. Therefore, our proposed protocol can provide the local source-location privacy.

As shown in Fig. 3, the intermediate nodes I_1, I_2, I_3 forward messages to ring nodes R_1, R_2, R_3, respectively. This means

that messages generated from one source node will not be forwarded to a specific ring node. Conversely, the data messages received from one ring node could also be transmitted from many different source nodes in the network.

The routing in the mixing ring is the second phase routing. The phase provides network-level source-location privacy. In fact, we have the following theorem.

Theorem 1. *It is computationally infeasible for an adversary to distinguish the message initiator and message forwarder in the mixing ring.*

Proof: (Sketch) As shown in Fig. 3 and described in equation (1), the data messages transmitted in the mixing ring is encapsulated as a vehicle message and encrypted using the shared secret key between the ring nodes. Particularly, for ring node B in Fig. 3, the received vehicle message is M_1, while the transmitted vehicle message is M_2. 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. ■

Without hop-by-hop message encryption, by comparing the vehicle message that a node received and transmitted, the adversaries can determine whether a data message has been loaded into the updated vehicle message.

The hop-by-hop encryption technique can also be used in RRIN phase. However, it is not as critical as it is for the mixing ring since the location of the ring and the routing path can be public information for the adversaries. Hop-by-hop encryption is necessary to prevent the adversaries from determining a particular data message comes from which part of the ring. While the local source-location privacy is protected by RRIN since the intermediate node is randomly selected. The entire domain traffic monitoring is not justifiable. The extra energy budget cannot be justified due to the limited benefit. In this way, while being able to provide source-location privacy for WSN, our proposed scheme is also quite energy efficient. The performance of the proposed scheme will be further analyzed in the next section.

It is also possible to have multiple mixing rings. In fact, mixing rings can also be used to provide local source-location privacy. However, this part is will not be considered in this paper.

VII. Performance Analysis and Simulation Results

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

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

where $\mathcal{E}$ denotes the energy consumption, α is a constant parameter and d is the distance of the transmission. Fig. 4 gives an example of a target area of size 8000×8000 meters. The shaded grids are selected as the ring grids. The line in the middle of the shaded area is indicated by the solid line. If the density of the sensor nodes in the sensor network is λ, then the total energy consumption for each sensor in this area to transmit one message to a ring node can be calculated as follows:

$$\begin{aligned}\mathcal{E}_{total} &= 8\mathcal{E}_U + 8\mathcal{E}_V + 4\mathcal{E}_W \\ &= 8\alpha\lambda \int_0^e x(e-x)^2 dx + 8\alpha\lambda \int_e^{4000} e(x-e)^2 dx \\ &\quad +4\alpha\lambda \int_e^{4000} \int_e^{4000} (x-e)^2 + (y-e)^2 dxdy,\end{aligned}$$

where $\mathcal{E}_U, \mathcal{E}_V, \mathcal{E}_W$ are the energy consumption for area U, V, W as demonstrated in Fig. 4. This is because that the entire sensor domain includes 8 pieces of U, 8 pieces of V and 4 pieces of W. It can be calculated that when $e = 2711$, the overall power consumption $\mathcal{E}_{total}$ achieves the minimum. In this way, we get the optimal ring location. It is interesting to point out that the total energy consumption is irrelevant to the density of the sensor nodes in the sensor network. In general, the distance between the optimal mixing ring with respect the center and the outside edge should be about $0.6778 : 0.3222 \approx 2.1 : 1$.

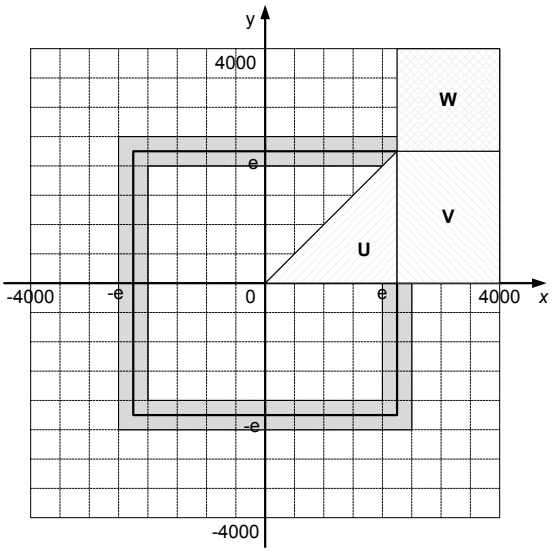
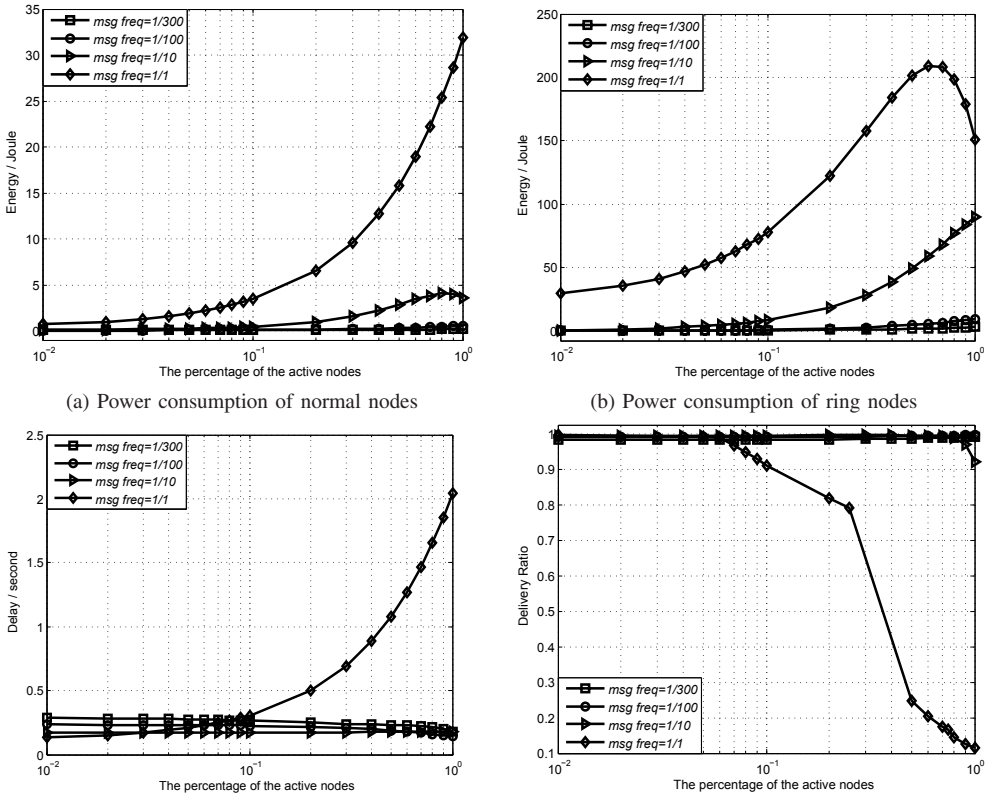


Fig. 4. Ring selection in simulation setup



(a) Power consumption of normal nodes

(b) Power consumption of ring nodes

(c) Message latency

(d) Message delivery ratio

Fig. 5. Performance of the proposed routing and encryption scheme

In practical application, for large sensor network, usually only a small fraction of the sensor nodes in the network has events to report. We name these nodes as *active nodes*. We also define two parameters in our 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 the relay ring nodes on the ring can generate vehicle messages. If there are n relay ring nodes on the ring, then each relay ring node needs to generate at least

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

vehicle messages each second.

Simulation results are provided in Fig. 5 to demonstrate the power consumption for both normal nodes and ring nodes, message latency and message delivery ratio of the proposed scheme. Our simulation was performed using NS2 on Linux system. In the simulation, the target area is a square field of size 8000×8000 meters. We partition this field into 2400 normal grids/nodes. The mixing ring is composed of 80 grids, i.e., $r = 80$. There are four relay ring nodes in the mixing ring, i.e., $n = 4$. We assume that the randomly selected intermediate node is at least 600 meters away from the real message source. The data messages are 8-bit long, i.e., $l = 8$. The vehicle messages are 16-bit long, i.e., $L = 16$. Our simulation results demonstrate that the proposed scheme is very efficient and can be used for practical applications.

From the Fig. 5.(a) and (b), we can see that ring nodes consume more energy than normal nodes. To solve this problem, the nodes in ring grids can take turns to be the ring nodes. It is also noticed that the delivery ratio drops exponentially when the traffic volume increases. It is primarily because of the traffic collisions and packet losses caused by the increased traffic volume. For a large sensor network, it is usually not necessary for all the sensor nodes to be active at the same time. In practice, the percentage of active nodes might be very low. The transmission frequency also tends not to be very high. In other words, the traffic volume may be low. In this scenario, we

can ensure almost 100% delivery ratio, as shown in Fig. 5.(d).

VIII. Conclusions

Source-location privacy is critical to the successful deployment of wireless sensor networks. In this paper, we have proposed a scheme that can achieve source-location privacy in the wireless sensor networks through a two-phase routing: the routing to a randomly selected intermediate node (RRIN) and routing through the network mixing ring (NMR). The optimal location for the mixing ring is also derived. Our proposed scheme provides excellent local source privacy and global source-location privacy. Simulation results demonstrate that the proposed scheme can achieve very good performance in energy consumption, message delivery latency while assuring high message delivery ratio.

IX. Acknowledgments

This research was partially supported by the US National Science Foundation under grants CNS-0716039, CNS-0848569, and CNS-0845812.

References

[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. Ahn, A. Bortz, and N. Hopper, "*k*-anonymous message transmission," in *Proceedings of the 10th ACM conference on Computer and Communications Security*, (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] O. Berthold, H. Federrath, and S. Köpsell, "Web mixes: A system for anonymous and unobservable Internet access," *Lecture Notes in Computer Science*, pp. 115–129, 2001.

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

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

[8] 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.

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

[10] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster protocol," July 2003. Version 2.

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

[12] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for web transaction," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66–92, 1998.

[13] S. Goel, M. Robson, M. Polte, and E. G. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Tech. Rep. 2003-1890, Cornell University, Ithaca, NY, February 2003.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] http://www.panda.org/.

[20] 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.

[21] "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.

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

[23] 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.

[24] 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.

[25] 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.

[26] S. Zhu, S. Setia, and S. Jajodia, "Leap: efficient security mechanisms for large-scale distributed sensor networks," in *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, (New York, NY, USA), pp. 62–72, ACM, 2003.

[27] 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.

[28] 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.

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

[30] J. Neander, E. Hansen, M. Nolin, and M. Bjorkman, "Asymmetric multihop communication in large sensor networks," *Wireless Pervasive Computing, 2006 1st International Symposium on*, pp. 7 pp.–, Jan. 2006.

[31] 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.

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

[33] S. M. Stigler, *Statistics on the Table*. Harvard University Press. chapter 22 (History of the term "normal distribution".

[34] N. SEMATECH, "Engineering statistics handbook." http://www.itl.nist.gov/div898/handbook/eda/section3/eda362.htm#CDF.