A sector-based random routing scheme for protecting the source location privacy in WSNs for the Internet of Things

Yu He, Guangjie Han, Hao Wang, James Adu Ansere, Whenbo Zhang

PII: S0167-739X(18)31640-6
DOI: https://doi.org/10.1016/j.future.2019.02.049
Reference: FUTURE 4802

To appear in: *Future Generation Computer Systems*

Received date : 10 July 2018
Revised date : 31 December 2018
Accepted date : 22 February 2019

A Sector-based Random Routing Scheme for Protecting the Source Location Privacy in WSNs for the Internet of Things

Yu He[a], Guangjie Han[a], Hao Wang[a], James Adu Ansere[a], Wenbo Zhang[b]

[a] *College of Internet of Things Engineering, Hohai University, 200 North Jinling Road, Changzhou 213022, China; heyuhhu2018@outlook.com, hanguangjie@gmail.com, wanghaohhu@outlook.com, jaansere@hhu.edu.cn*
[b] *College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China; zhangwenbo@yeah.net*

Abstract

With the development of the Internet of Things (IoT), Smart Data, which effectively support the IoT for planning, operation, monitoring, control, and intelligent decision making, has received extensive interest by researchers. However, the security of the data source has not been entirely resolved. Wireless sensor networks (WSNs) are vital components of the IoT for event monitoring and information gathering. Recently, source location privacy (SLP) protection in WSNs has attracted attentions as an approach to prevent adversaries from performing a backtracking strategy to capture the data sources. In view of the characteristics of the sensor nodes, the restricted computing power and the energy resource, we propose a sector-based random routing (SRR) scheme to address the SLP problem and reduce the energy consumption. In the SRR, the data packets are sent to random phantom sources that are located in different sectors and are distributed in all directions to reach the sink node. In addition, the concept of a hop threshold is presented to control the routing strategies and reduce the energy consumption. The theoretical analysis and experimental results prove that the proposed protocol efficiently reduces backtracking and direction attacks while safekeeping the balance between security and network lifetime.

Keywords: Internet of things, wireless sensor networks, source location privacy, phantom source

1. Introduction

The Internet of Things (IoT), in which numerous physical objects are connected to collect and exchange data, has been applied in various domains, such as home automation, patient and industrial monitoring, smart cities, and smart infrastructures [1-2]. In recent years, Smart Data, which refers to valuable data without noise, has played an important role in supporting the development of the IoT. However, researchers have not entirely solved the security issue related to the location of the data origin. As fundamental components of the IoT for event monitoring and information gathering, wireless sensor networks (WSNs) are comprised of abundant resource-constrained and non-rechargeable sensor nodes that are self organized [3-5]. Unlike wired networks, WSNs are flexible to adapt to complex application scenarios. However, with an appropriate wireless device, a person can monitor the communication signals in a wireless sensor domain [6]. In spite of encryption techniques that protect the communication content exchanged between two sensor nodes, the adversaries mostly use powerful equipment for locating the message source by monitoring the communication patterns between the nodes without accessing the communication content. Therefore,

Preprint submitted to Elsevier *December 31, 2018*

many researchers have focused on source location privacy (SLP) protection in recent years.

SLP is a significant and challenging security issue [6]. In the absence of SLP, vital information on the physical objective may be revealed. Numerous research studies related to SLP have been conducted in the last decade. Ozturk *et al.* [7] proposed the classical Panda-Hunter Game and the phantom routing algorithm. Wang *et al.* [8] first presented the concept of a visible area. Yao *et al.* [9] put forward a scheme based on a multi-ring centered at the sink node. Chen *et al.* [10] used constrained offset angles and probabilistic routing to balance the energy consumption and the security requirements. However, to our best knowledge, most existing schemes do not consider balancing energy consumption and security based on the location of the source node.

In this paper, a sector-based random routing (SRR) scheme is proposed to protect SLP. Our proposed scheme effectively prevents the adversary from utilizing a backtracking mechanism to locate the source node. In SRR, the deployed area is uniformly divided into sectors that have the sink node as a common vertex. Initially, the source node selects phantom sources of the data packets in different sectors. Then the source node sends the data packets to the phantom sources using annular routing. Finally, the phantom sources send the data packets to the sink node via random routing. Hence, the adversary is confused by the varying paths. SRR reduces the energy consumption by assuming that sufficient security exists. A hop threshold (t_{hop}) is set for the sensor nodes. If the hop count of the source node is smaller than t_{hop}, the sectors opposite to the source node are selected to improve the location privacy. Otherwise, all sectors become the candidate domain of the phantom sources but the sectors that are closer to the source node have higher probabilities to be selected to save energy consumption. The main contributions of this paper are as follows:

(1) We put forward a novel SRR scheme that uses multiple dispersed routes to achieve improved SLP.
(2) We introduce the concept of t_{hop} to maintain a balance between the energy consumption and the source node security depending on the location of the source node.
(3) We provide an extensive theoretical analysis and experimental simulations to prove the efficiency of our scheme.

The rest of this paper is organized as follows: Section 2 introduces the related studies on this topic. Section 3 presents the system model and problem statement. Section 4 describes our proposed scheme in detail and Section 5 gives a theoretical analysis of our scheme. Section 6 shows the experimental results based on a simulation. We present the conclusions of the study in Section 7.

2. Related work

The privacy issues of WSNs can be divided into two categories: content privacy and context privacy. Content privacy is usually protected by encryption or authentication but the schemes on content privacy are not discussed here. Context privacy is more intractable because the communication signal is exposed in cyberspace. The SLP is a type of context privacy that protects the location of the source node [11-13].

The SLP problem has gained much attention in the past years since Ozturk *et al.* [7] initially proposed the classical panda-hunter game, as shown in Figure 1. Many sensor nodes are deployed in a habitat at random to monitor the panda's living status. When a panda is detected in the network, the corresponding sensor node becomes the source node and starts sending the collected data to the sink node hop by hop. This is done periodically until the panda disappears from its detection range. Since Kamat *et al.* [14] formalized the SLP problem, the panda-hunter game model has become the fundamental event-driven application scenario for studying SLP.

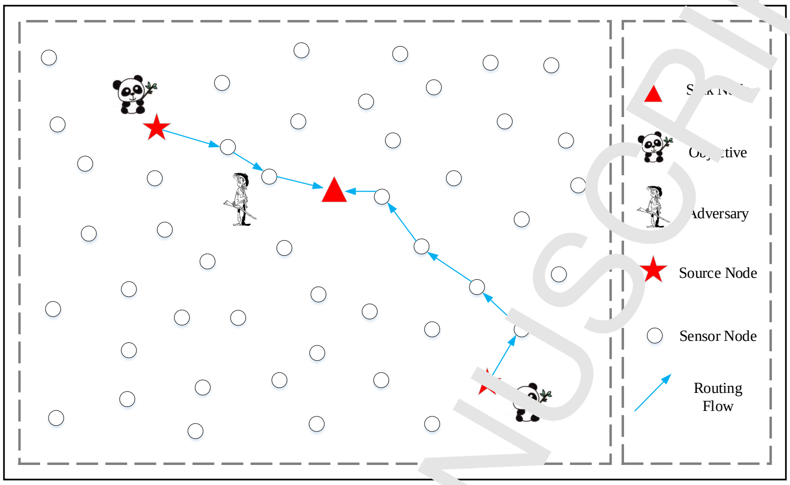


Figure 1: The structure of the panda-hunter game.

Ozturk *et al.* [7] first developed the phantom routing scheme to solve the SLP problem. As soon as the source node detects an event, the generated data packets are sent to adjacent nodes called neighbor nodes, which continue sending the data packets to their own neighbor nodes in a similar manner. This procedure maintains a predefined hop. Afterwards, each node sends the data packets to all neighbor nodes until the sink node receives the data packets. It becomes difficult for the adversary to capture the source node by backtracking because the routing paths of the data packets are random and unpredictable. However, Kamat *et al.* [14] showed that a routing loop is formed due to a pure random walk. To solve this problem, Tan *et al.* [15] proposed a directed random walk scheme called EDROW. In EDROW, the nodes closer to the sink node are called parent nodes and are responsible for transmitting the data packets. Thus, a sufficient number of optional parent nodes provide better SLP. Luo *et al.* [16] presented a phantom single-path routing scheme, where each phantom source generated a fake path to simulate the behavior of the real source node as a means of inducing the adversary. A multiple phantom scheme was designed by Gupta *et al.* [17]. In this scheme, every three nodes are considered a triplet. When one node of a triplet becomes the source node, the other two nodes play the role of the phantom source. A greedy random walk scheme called GROW was proposed by Xi *et al.* [18]. In GROW, a random walk is initiated by the sink node. In the meantime, the source node also transmits event packets in a random walk. Once the two paths connect, the packets are delivered along the path of the sink node until they reach the sink node. In order to achieve a balance between privacy and energy consumption, Chen *et al.* [10] designed a selection domain that randomly chooses a neighbor node to transmit the data packet. In addition, the sum of offsets is estimated to control the energy consumption of the random routes. Tang *et al.* [19] proposed the CASER protocol. In this protocol, the two adjustable parameters of energy balance control and probabilistic-based random walk are designed to address the conflict between the lifetime optimization and the security.

In addition to random routes, the use of fake message also constitutes a proven measure to provide SLP [20-22]. It is demanding for the adversary to identify the true routes because it is difficult to distinguish the real data packets from the fake data packets. Chen *et al.* [23] proposed three similar approaches that are called bidirectional tree, dynamic bidirectional tree, and zigzag bidirectional tree. By using fake data packets to generate false paths that differ from the real path, the adversary is likely to trace some of the wrong routes. Thus, the source node has more time to safely transmit the data packets. Mahmoud *et al.* [24] first introduced an attack named Hotspot-Locating and a cloud-based scheme was proposed to cope with this attack. A novel tree-based scheme was proposed by Long *et al.* [25]. In this scheme, the energy of the non-hotspot regions is utilized to generate diversionary routings. Fake messages are transmitted in the diversionary routings to tempt the adversary to leave the real path. Proano *et al.* [26] devised a scheme to confront the global adversary. This scheme partitions the WSN into minimum connected dominating sets (MCDSs). When the real data packets pass by the section of an MCDS, fake data packets are generated to deny the adversary access to identify the real routing. Huang *et al.* [27] proposed a redundancy branch convergence-based preserved source location privacy scheme (RBCPSLP), in which every node generates fake messages independently according to the residual energy and the total routes are merged into several backbone routes to reduce the energy consumption of the hotspot area.

In addition, a solution has been proposed to provide SLP by utilizing the cyclic topology [28-30]. A multiring-based scheme was proposed by Yao *et al.* [9]. In this scheme, sensor nodes are divided into different rings dependent on the hop distance to the sink node. Before reaching the sink node, each data packet is transmitted with an angle α along the external ring and with an angle $\pi - \alpha$ along the internal ring. Although the scheme also utilizes fake messages to enhance the security, the network lifetime is reduced due to the additional energy consumption. Zhou *et al.* [31] used a ring to buffer the real data packets. As the fake packets generated by a boundary node pass by the ring, the fake data packets are replaced by the real data packets to be sent to the sink node. However, this method is not secure enough when the adversary appears on the ring where the source node exists.

3. The system model and problem statement

3.1. Network model

The network model in this study is based on the panda-hunter game [32]. The key points of this model are as follows:

(1) There are a large number of sensor nodes uniformly and randomly deployed in the network. Each sensor node is constrained by computing power and energy resource. We assume that the locations of the sensor nodes remain steady after they have been deployed and that any two sensor nodes communicate via multi-hop routing.

(2) A powerful and settled sink node, which is the unique and final destination of all data flows, is located in the center of the deployment area. We assume that the location of the sink node is public and that each sensor node is aware of its own location based on a location algorithm.

(3) The network belongs to an event-monitoring network. When an event is detected, the source node periodically sends the gathered data to the sink node. We assume that events randomly arise in the network and there is only one source node generated at any time.

(4) We assume that a cryptographic technique is used to ensure that the content of each data packet is unknowable to unauthorized users; the details of the cryptographic techniques can be found in [33,34].

3.2. Adversary model

An adversary is a hunter with the ultimate goal to find the source node to capture the pandas. We assume that the characteristics of the adversary are as follows:

(1) Well-equipped. The adversary is equipped with sophisticated radio equipment. The energy resources, computing power, and storage capacity are unlimited.
(2) Passive. The adversary does not perform any active attacks to obstruct the normal operation of the network because such behavior is easily detected by the network administrator. Therefore, the adversary only implements passive attacks such as eavesdropping to determine the traffic pattern of the network.
(3) Local vision. The eavesdropping radius of the adversary is similar to the communication radius of the sensor node. The adversary can estimate the location of the direct sender by calculating the signal strength and direction to quickly move to the estimated site. As this process is repeated, the adversary can perform a backtracking attack.

3.3. Problem statement

In this study, we design a guaranteed and efficient scheme to protect the SLP; the metrics used to characterize the performance of the proposed schemes are as follows.

(1) Safe time: This is the period that begins when the source node transmits the first data packet and ends when the adversary captures the source node. If the transmitting period of the data packets (T) is fixed, the safe time is expressed as follows:

$$\max\left(\textit{safe time}\right) = \max\left(n\right) \times T \tag{1}$$

where n is the number of data packets sent by the source node before the adversary captures the source node.

(2) Capture rate: This is defined as the probability that the source node can be captured by the adversary in a certain amount of time. The security improves as the capture rate declines. We assume that when the maximum hop distance from the source node to the sink node is l and the probability of the transmitting node i being captured is p_i, then the metric is expressed as follows:

$$\min\left(\textit{capture rate}\right) = \min\left(\prod_{i=1}^{l} p_i\right) \tag{2}$$

(3) Lifetime: This is defined as the period that begins when the network starts running and ends when the first dead node occurs. A dead node is a node that runs out of energy. If the energy consumption of node i is e_i and the sum of the sensor nodes equals n, the objective is expressed as follows:

$$\max\left(\textit{lifetime}\right) = \min\left\{\max\left(\sum_{i=1}^{n} e_i\right)\right\} \tag{3}$$

4. SRR scheme

In this section, we introduce the details of the SRR scheme; it consists of three phases. In the first phase, an initialization program is executed by the sink node. Each node obtains the information (e.g., location of the sink node and list of neighbor nodes) at the end of this phase. In the second phase, dispersive phantom sources are generated and data packets are sent from the source node to the different phantom sources. In the last phase, the data packets are sent from the phantom sources to the sink node by random routing.

4.1. Initialization

We assume that the network includes one sink node v_0 and t sensor nodes v_i $(i = 1, 2, \cdots t)$, represent as $\{v_0, v_1, \cdots, v_t\}$. At the time of initialization, the sink node sets its hop count as 0, $v_0.hop = 0$ (the hop count represents the minimum hop distance from a node to the sink node). All sensor nodes set their hop counts to infinity, $v_i.hop = \infty$. Subsequently, the sink node begins to broadcast the beacon message to the nearby nodes using the same transmitting radius as the sensor nodes. The initial hop count recorded in the beacon message is equal to 0 ($beacon.hop = 0$). Once a node v_i receives the beacon message, it checks whether Inequation (4) is satisfied; if that is the case, the node discards the beacon message.

$$v_i.hop \leq beacon.hop \tag{4}$$

Otherwise, the node v_i sets $v_i.hop = beacon.hop + 1$ and increases $beacon.hop$ by one. The node continues to broadcast the beacon message to its neighbor nodes. Other nodes that receive the beacon message repeat the same procedure as node v_i until all hop counts of the nodes are no longer updated or have reached the threshold of the updating time. After the initialization, each node obtains its hop count and those of the neighbor nodes. Furthermore, each node generates three neighbor lists Ls, Ll, and Le to determine the neighbor nodes whose hop counts are smaller, larger, and equal to those of the current node respectively.

Table 1: Summary of notations

Symbol	Meaning
v_i	The sensor node i
$v_i.hop$	The hop count of node v_i
$source.hop$	The hop count of the source node
Ls_i	The neighbor list of node v_i with a smaller hop count
Le_i	The neighbor list of node v_i with an equal hop count
Ll_i	The neighbor list of node v_i with a larger hop count
sec_k^i	The sector i selected for the kth data packet
t_{hop}	The hop threshold
P_i	The selected probability of sec_i
θ	The angle between the lines of sink-current and sink-source
R	The communication radius of the nodes
Ts	The safe time
Rc	The capture rate
T	The period of the data packets

4.2. Phantom routing

In general, the following sequence occurs; first, the source node calculates the random expected angles for each data packet. Subsequently, the data packets are sent through annular routes and each node determines whether it should be a phantom source according to the expected angle. Finally, the data packets are sent from the real source node to the different phantom sources successfully.

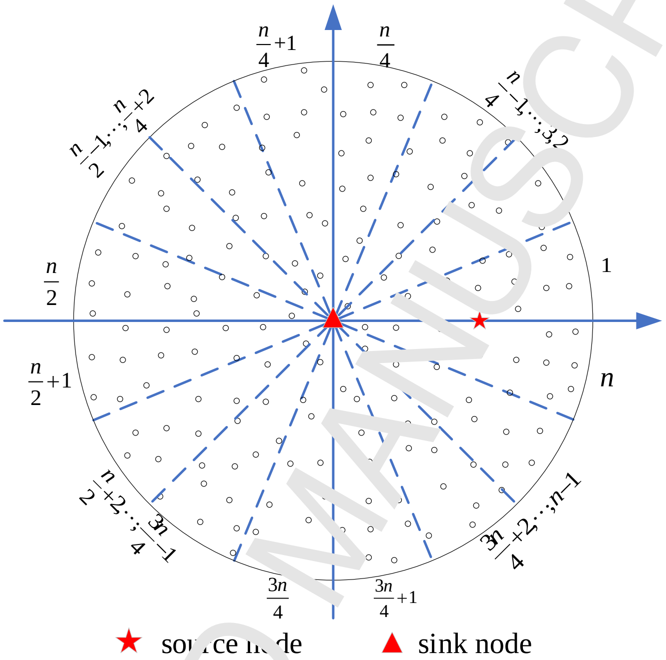


Figure 2: The coordinate system and the separate sectors.

4.2.1. Calculation of the expected angle

Before the source node sends the data packets, it creates a rectangular coordinate system as illustrated in Figure 2. The sink node is the origin and the line from the sink node to the source node is the X-axis. The area of the network is evenly divided into n ($n = 2i, i = 1, 2, 3 \cdots$) sectors, that are designated as $sec_1, sec_2, \cdots, sec_n$ in a counterclockwise direction. The central angle of each sector is $\frac{2\pi}{n}$. Finally, the source node determines the next strategy according to the following comparison:

$$source.hop \leq t_{hop} \tag{5}$$

If Inequation (5) holds, the source node randomly chooses a sector sec_i^1 from the candidate domain $\left[sec_{\frac{n}{4}+1}, sec_{\frac{n}{4}+2}, \cdots, sec_{\frac{n}{2}}\right]$ for the first data packet and selects a random angle from sec_i^1 as the expected angle. For instance, when sec_1 is selected, the corresponding angle ranges from 0 to $\frac{2\pi}{n}$ and the source node assigns a random number in the range of $\left(0, \frac{2\pi}{n}\right)$ to the expected angle. For the kth ($k \geq 2$) data packet, the source node removes sec_i^{k-1} chosen by the prior data packet from the candidate domain $\left[sec_{\frac{n}{4}+1}, sec_{\frac{n}{4}+2}, \cdots, sec_{\frac{n}{2}}\right]$. Afterward, a new sector sec_i^k is selected from the updated candidate domain comprised of the remaining $\left(\frac{n}{4} - 1\right)$ sectors. Likewise, a new expected angle is selected from sec_i^k. Subsequently, except for the first data packet that has a candidate domain of $\frac{n}{4}$ members, the other candidates have the candidate domain of $\left(\frac{n}{4}-1\right)$ members and the adjacent data packets do not select the expected angles from the same sector.

In contrast, if Inequation (5) does not hold, the selection of the expected angle is the same as described above but the selection strategy for the sectors is different due to the tradeoff between the security and the energy consumption. The initial candidate domain becomes $\left[sec_1, sec_2, \cdots, sec_{\frac{n}{2}}\right]$ and creates the corresponding probability set $\left[P_1, P_2, \cdots, P_{\frac{n}{2}}\right]$, where P_i represents the selected probability of sec_i. The principle of setting P_i is as follows: a low probability for selection is set for sec_1 to prevent a direction attack and the values from P_2 to $P_{\frac{n}{2}}$ are decreased gradually to reduce the energy consumption. A small P_1 prevents the concentration of the routing paths in the area surrounding the line between the sink node and the source node to prevent a direction attack. Furthermore, by decreasing the values from P_2 to $P_{\frac{n}{2}}$, the average length of the routing paths is reduced, which reduces the energy consumption. Similarly, the source node selects a sector for the first data packet from the candidate domain based on the initial probability set. For the kth $(k \geq 2)$ data packet, the sector sec_i^{k-1} is removed from the candidate domain. In addition, P_i will become 0 and its value is added proportionally to the other probabilities. Therefore, the updated candidate domain and probability set are as follows:

$$\left[sec_1, \cdots, sec_{i-1}, sec_{i+1}, \cdots, sec_{\frac{n}{2}}\right]$$

$$\left[P_1 + \frac{P_1}{1-P_i} \times P_i, \cdots, P_{i-1} + \frac{P_{i-1}}{1-P_i} \times P_i, P_{i+1} + \frac{P_{i+1}}{1-P_i} \times P_i, \cdots, P_{\frac{n}{2}} + \frac{P_{\frac{n}{2}}}{1-P_i} \times P_i\right]$$

The details are shown in Algorithm 1.

Algorithm 1 Calculation of the expected angle.

```
Establish coordinate system and divide network into $n$ sectors $sec_1, sec_2, \cdots, sec_n$.
**if** $source.hop \leq t_{hop}$ **then**
    Remove $sec_i^{k-1}$ from initial candidate domain.
    Randomly choose $sec_i^k$ from candidate domain.
    Randomly choose an expected angle from $sec_i^k$.
**else**
    Remove $sec_i^{k-1}$ from initial candidate domain.
    Update probability set.
    Choose $sec_i^k$ from candidate domain based on probability set.
    Randomly choose an expected angle from $sec_i^k$.
**end if**
```

4.2.2. Routing to the phantom source

After identifying the expected angle, the source node adds the information on the expected angle to the corresponding data packet. Hence, the process of transmitting the data packet to the phantom source starts.

As illustrated in Figure 3, we assume that the adversary has sufficient memory space to store the visited locations to predict the direction of the data flow. The adversary backtracks to a phantom source and continues to intercept a sufficient number of data packets hop by hop before the source node disappears; an approximate ring path is deduced by the adversary due to the visited locations. Therefore, the adversary continues moving along the deduced ring direction. When the adversary passes by the visible area, there is a high probability that he can find the protected objective.

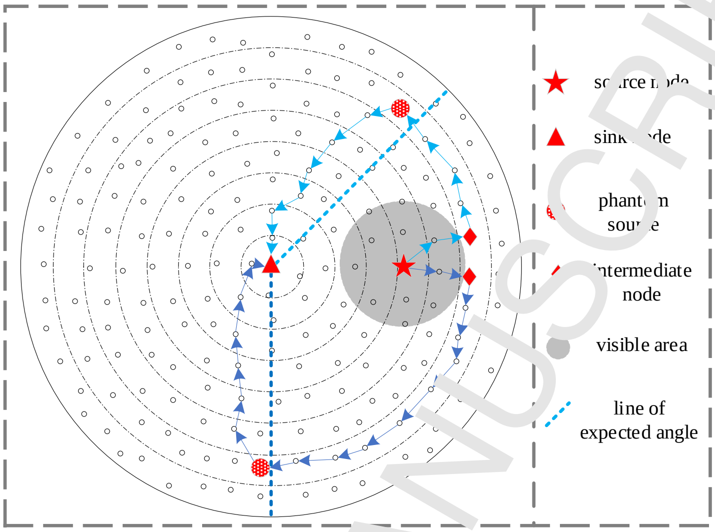


Figure 3: The illustration of the SRR scheme.

To tackle this issue, at first, an integer is predetermined based on the range of the visible area and is recorded in the data packet. In this study, we assume that the radius of the visible area is $2R$ and the integer is randomly chosen and is either 2 or 3. Therefore, the source node randomly chooses a neighbor node from its Ll and sends the data packet to the neighbor node. When the data packet is received by the node v_i, v_i checks the data packet and decreases the integer by one. If the new integer equals 0, v_i becomes the intermediate node. Otherwise, v_i sends the data packet to a random neighbor node in Ll_i. The details are shown in Algorithm 2.

Algorithm 2 Generation of the intermediate node.

```
Check the integer in the data packet.
Decrease the integer by one.
if the integer equals 0 then
    Current node becomes intermediate node.
else
    Send the data packet to a random neighbor in $Ll$.
end if
```

As the intermediate node appears, it randomly determines to transmit the data packets clockwise or counterclockwise with equal probability. The data packet is immediately sent to a random neighbor node of the intermediate node and the neighbor node should be both on the determined direction and in the Le of the intermediate node. When the data packet is received by the node v_j, v_j calculates an angle θ according to Equation (6):

$$\theta = \cos^{-1}\frac{a^2+b^2-c^2}{2ab} \tag{6}$$

where a represents the distance between the sink node and the current node; b represents the distance between the sink node and the source node; c represents the distance between the source node and the current node. The result θ is the angle between the sink-current line and the line sink-source line.

If $\theta \geq$ *expected angle*, the node v_j becomes the phantom source. Otherwise, v_j sends the data packet in the determined direction to a random neighbor node that is in the Le_j. By repeating this process, the data packet is transmitted from the intermediate node to the phantom source. The details are shown in Algorithm 3.

Algorithm 3 From the intermediate node to the phantom source.

```
Determine clockwise direction or counterclockwise direction.
if $\theta \geq$ expected angle then
    Current node becomes the phantom source.
else
    Send the data packet in the determined direction to a random neighbor node that is in the Le.
end if
```

4.3. Random routing

After the phantom source appears, it sends the data packet to the sink node via random routing. From the phantom source to the sink node, when a node v_k receives the data packet, v_k transmits the data packet to a random neighbor node that is in the Ls_k. This process continues until the data packet reaches the sink node.

However, a common problem needs to be considered. How do the boundary nodes find the paths to the sink node? As shown in Figure 5, when a source node arises on the border of the network, it is hard to transmit data packets to the sink node by directly applying the SRR scheme because there are not enough nodes to form the ring on the border. Depending on the network size, a threshold is set by the sink node and is sent to all nodes. When the distance from the source node to the sink node is larger than the threshold, the data packet is sent for some hops through the neighbor nodes in the Ls at first so that the data packet is received by a node whose distance to the sink node is smaller than the threshold. Then the SRR scheme can be launched successfully by the node. The threshold and the hops can be adjusted depending on the actual conditions of the network.

5. Security analysis

In this section, we introduce the theoretical analysis of the security of our SRR scheme in terms of likely attack patterns of the adversary.

5.1. Direction attack

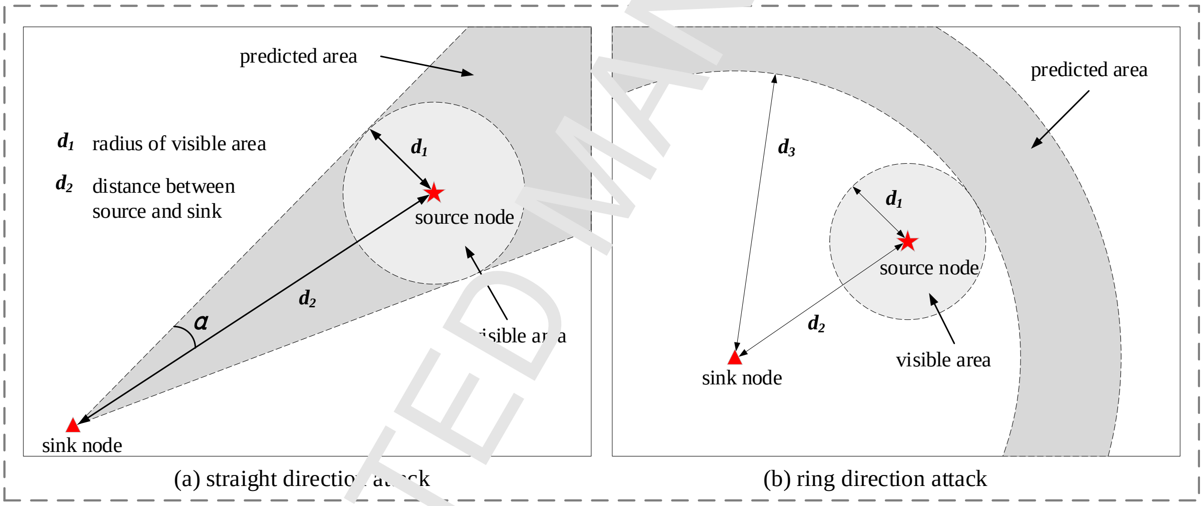
In this study, we focus on the straight direction attack and the ring direction attack. As demonstrated in Figure 4(a), when a sufficient number of data packets, whose routing paths are located in the shaded area, are intercepted, the adversary can launch a straight direction attack. When the adversary passes by the visible area in a straight direction along the shaded area, there is a high probability that the source node will be captured with a high probability. In the SRR, if the hop count of the source node is not larger than t_{hop}, the routing path from the phantom source to the sink node are located in $sec_{\frac{n}{4}+1} \sim sec_{\frac{n}{2}}$; therefore, the probability that the source node is captured by a straight direction attack equals 0. If the hop count of the source node is larger than t_{hop}, namely $d_2 > t_{hop} \times R$, the angle α is given by Equation (7)

$$\alpha = \sin^{-1}\frac{d_1}{d_2} \tag{7}$$

To improve the security of the source node, we can adjust t_{hop} as $t_{hop} > \frac{d_1}{R\sin\frac{2\pi}{n}}$ to guarantee $\alpha < \frac{2\pi}{n}$. In this case, the attack probability is expressed as:

$$\underbrace{P_1 \times \cdots \times P_1}_{k} \times \frac{d_1}{d_2 \sin\frac{2\pi}{n}} = \frac{d_1(P_1)^k}{d_2 \sin\frac{2\pi}{n}} \tag{8}$$

where P_1 denotes the probability that the routing path is located in sec_1 or sec_n and k denotes the number of locations used to predict the attack direction by the adversary.



(a) straight direction attack (b) ring direction attack

Figure 4: The illustration of the direction attack.

As shown in Figure 4(b), similar to the previous case, when the routing paths of the intercepted data packets are annular, a specific ring direction can be inferred by the adversary. It is also possible to capture the source node when the adversary passes by the visible area and moves in the direction of the ring. In the SRR, we assume that the radius of the visible area is $2R$ and we choose the integer 2 or 3 to ensure that the distance between the intermediate node and the source node is at least $2R$. Therefore, the probability of being captured by the ring direction attack equals 0. In theory, if the shortest distance d_3 from the ring routing to the sink node satisfies $d_3 \geq d_2 + d_1$, the security of the source node is guaranteed with regard to a ring direction attack.

5.2. Backtrack attack

This refers to the case when an adversary locates a directed transmission of the message based on the signal intensity and angle and then moves to the location to continue eavesdropping on the new messages. By repeating this process, the adversary may capture the source node. In this section, we describe the theoretical analysis of the security of our scheme for the most vulnerable situation, i.e., when the adversary captures the source node successfully in the least amount of time. The detailed analysis steps are as follows:

Part 1: From the sink node to the phantom source. We assume the hop count of the source node is h; the safe time (Ts) and the capture rate (Rc) of this phase are expressed as:

$$Ts_1 = 2\,(h+2)\,T \tag{9}$$

$$Rc_1 = \begin{cases} \left(\frac{2}{n-4}\right)^{h+2}, h \le t_{hop} \\ (P_i)^{h+2}, \quad h > t_{hop} \end{cases} \left(i = 1, 2, \cdots, \frac{n}{2}\right) \tag{10}$$

Proof: According to Algorithm 2, the hop count of the phantom source is $h+2$ or $h+3$ and least secure situation is $h+2$. Because the adversary starts backtracking from the sink node to the phantom source, the fewest number of intercepted data packets is equal to the hop distance between them, namely $h+2$. In addition, because the SRR mechanism ensures that the adjacent data packets do not select the expected angles from the same sector, Ts is at least twice as much as $(h+2)\,T$. Therefore, Equation (9) is satisfied. When $h \le t_{hop}$, the probability that the routing path just passes by the sector where the adversary remains is $\frac{1}{2} \times \frac{1}{\frac{n}{4}-1} = \frac{2}{n-4}$ for each data packet. As a result, the adversary has a probability of $\underbrace{\frac{2}{n-4} \times \cdots \times \frac{2}{n-4}}_{h+2} = \left(\frac{2}{n-4}\right)^{h+2}$ to capture the phantom source. When $h > t_{hop}$, if the adversary is located in sec_i, the probability of intercepting the data flow is P_i. Therefore, similar to the previous case, the probability of capturing the phantom source equals $(P_i)^{h+2}$.

Part 2: From the phantom source to the source node. When the adversary gets to the phantom source whose hop count is $h+2$ and the angle at the source node is θ ($0 \le \theta \le \pi$), then Ts and Rc of this phase are expressed as

$$Ts_2 = (\theta\,(h+2) + 2)\,T \tag{11}$$

$$Rc_2 = \left(\frac{1}{4}\right)^{\theta(h+2)} \tag{12}$$

Proof: In this phase, the routing path between the phantom source and the intermediate node is similar to an arc; therefore, the hop distance is equal to $\theta\,(h+2)$ and the least hop distance from the intermediate node to the source node is equal to 2. The sum $\theta\,(h+2)+2$ is also the minimum number of data packets needed by the adversary. Therefore, Equation (11) is satisfied. According to Algorithm 2 and Algorithm 3, a data flow has the probability of $\frac{1}{2}$ to occur in the same arc where the adversary is located. In addition, the probability that the direction of annular data flow is precisely towards the adversary is also $\frac{1}{2}$. Therefore, for each data packet, the probability of being intercepted by the adversary equals to $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. If we assume that the probability of being intercepted by the adversary from the source node to the intermediate node is 1, the probability that the adversary captures the source node by intercepting $\theta\,(h+2)+2$ data packets is equal to

$\left(\frac{1}{4}\right)^{\theta(h+2)}$. Therefore, Equation (12) is satisfied. In summary, the minimum Ts and corresponding Rc of the entire process are expressed as:

$$Ts = Ts_1 + Ts_2 = \left(\left(\theta + 2\right)\left(h + 2\right) + 2\right)T \tag{13}$$

$$Rc = Rc_1 \times Rc_2 = \begin{cases} \left(\frac{2}{n-4}\right)^{h+2}\left(\frac{1}{4}\right)^{\theta(h+2)}, h \leq t_{hop} \\ \left(P_i\right)^{h+2}\left(\frac{1}{4}\right)^{\theta(h+2)}, \quad h > t_{hop} \end{cases} \left(i = 1, 2, \cdots \frac{n}{2}\right) \tag{14}$$

6. Simulation results and performance analysis

6.1. Simulation environment and parameter configuration

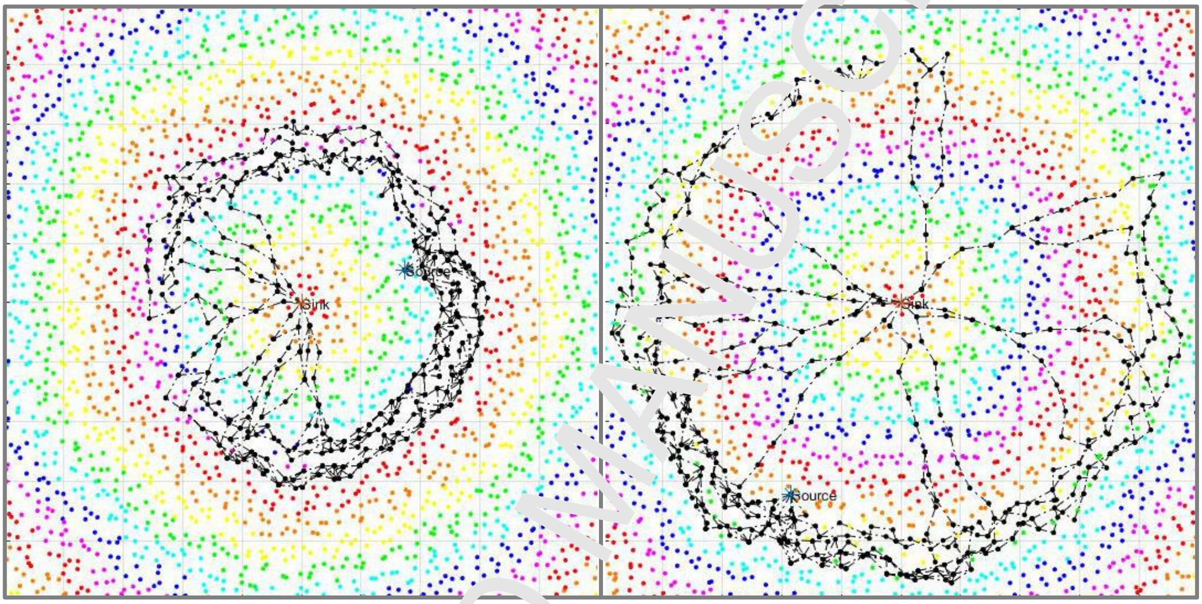
In this study, we compare the proposed scheme with other schemes introduced by Yao *et al.* [9] and Chen *et al.* [10] as the Constrained Random Routing (CRR) scheme. The two schemes are similar to our proposed scheme. In addition, all schemes have the same goal, which is to provide SLP and to minimize the energy consumption. MATLAB R2016a is used to simulate the performance of the schemes. The simulation parameters are shown in Table 2.

Table 2: Simulation parameters

Parameter	Value
Network size	$1000m \times 1000m$
Number of sensor nodes	2500
Coordinate of sink node	$(500, 500)$
Initial energy of sensor nodes	$50J$
Data packet size	$1000bit$
Period of data packet (T)	$5s$
Hop threshold (t_{hop})	5
Radius of sensor nodes (R)	$50m$

The simulation environment and parameter configuration are as follows. There are 2500 nodes distributed over an area of $1000m \times 1000m$. To simulate the realistic deployment of the sensor nodes, the monitored area is divided into 2500 square grids whose size is $20m \times 20m$ and only one sensor node is randomly deployed in each grid. The communication radius of each node is $50m$ and is equal to the eavesdropping radius of the adversary. In the experiment, the total number of sectors is 12 and the probability set for the source node whose hop count is larger than t_{hop} is $[0.05, 0.36, 0.27, 0.18, 0.09, 0.05]$ according to the principle described in Section 4.2.1. In this setting, the phantom sources have a probability of 0.81 $(0.36 + 0.27 + 0.18)$ to occur in the area ranging from sec_2 to sec_4; in other words, the routing paths are also centralized in this area. Therefore, the straight direction attack can be prevented and the extra energy consumption caused by long distance routing can be reduced. Each group of experimental results is obtained by more than 100 simulations to ensure good accuracy.

Figure 5(a) and Figure 5(b) illustrate the results of 20 simulations of the SRR scheme when the hop count of the source node is smaller and larger than t_{hop} respectively. In the figure, the small dots represent the sensor nodes. These nodes are divided into several layers indicated by different colors to represent the different hop counts. From inside to outside, the hop counts of the nodes increase layer by layer. In addition, the black dots and lines denote the routing nodes and paths respectively. As shown in Figure 5(a), the routing paths from the phantom sources to the sink node are centralized in the region opposite to the source node because the angles of the selected sectors range from $\frac{\pi}{2}$ to $\frac{3\pi}{2}$. Similarly, if the hop count of the source node is larger than t_{hop}, each sector has different probabilities for being chosen but the main probabilities are deployed in the sectors ranging from $\frac{\pi}{6}$ to $\frac{2\pi}{3}$ and from $\frac{4\pi}{3}$ to $\frac{11\pi}{6}$. Therefore, most routings from the phantom sources to the sink node are centralized in the scopes and the simulation result in Figure 5(b) are in agreement with the theoretical analysis.



(a) Hop count of source is smaller than hop threshold (b) Hop count of source is larger than hop threshold

Figure 5: The experimental results of the SRR scheme.

6.2. Performance analysis

(1) Safe time.

Figure 6 depicts the simulation results of the safe time. It is observed that the SRR scheme has the longest safe time of the three schemes. Unlike the scheme described in Ref. [9], our scheme sets sectors for selecting the phantom sources. Because the SRR ensures that two contiguous phantom sources do not exist in the same sector, the adversary has a smaller probability to intercept new messages in succession. Therefore, the adversary spends more time trying to capture the source node. In the CRR scheme, the nodes with smaller offset angles are assumed to have a larger probability to be chosen as the next hop. Therefore, the routing paths of the CRR scheme are identical to the directed random routing from the source node to the sink node. This results in the shortest safe time for the CRR scheme among the three schemes. In addition, as the distance to the sink node increases, the routing paths between the source node and the sink node become longer on average. Correspondingly, the adversary spends more time on the backtracking process due to the longer routing path and thus, the safe time increases.

In Figure 7, we show the changes in the safe time for different side lengths of the grid. It is observed that the safe time exhibits a decreasing trend with increasing side length. We already mentioned that only one sensor node is randomly deployed in each grid; therefore, the side length of the grid is proportional to the distribution density of the sensor nodes. By expanding the grid area, the densities of the sensor nodes and the optional neighbor nodes decrease. Therefore, the source node can be captured more easily by the adversary as the randomness of the routing paths decreases.

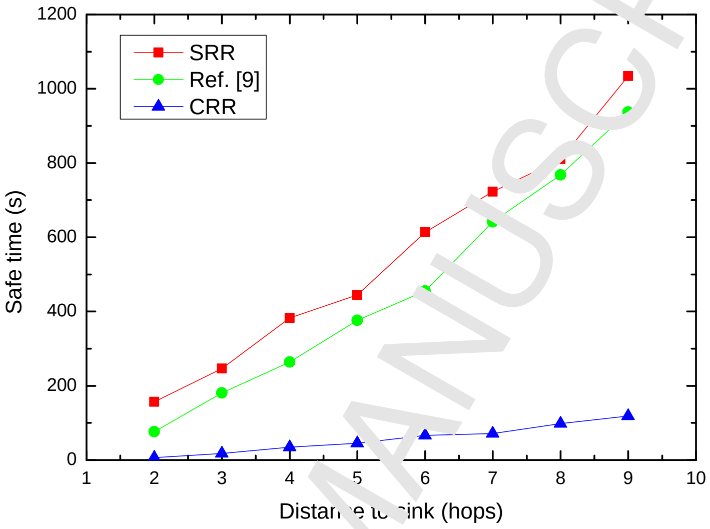


Figure 6: The safe time with different hop distance from source to sink.

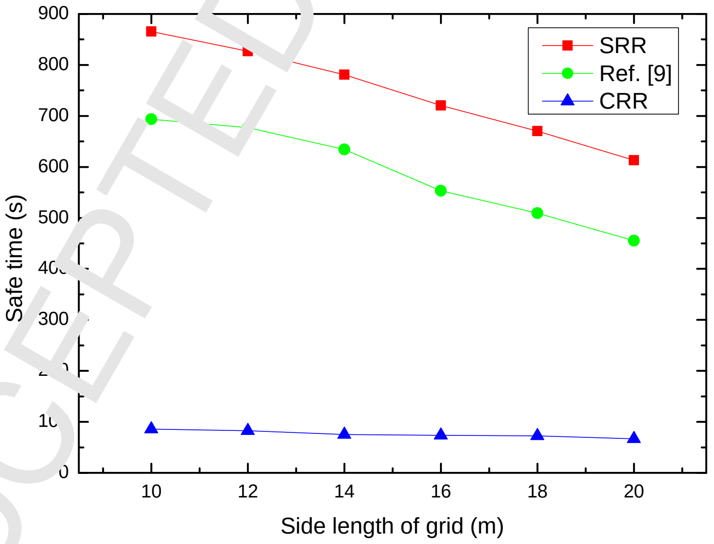


Figure 7: The safe time with different side length of deployment grid.

(2) Intercept rate.

In Figure 8 and Figure 9, the intercept rate is used as a benchmark for determining the security. The intercept rate is defined as the ratio between the number of data packets that is eavesdropped on by the adversary and the number of data packets sent by the source node. Figure 8 shows that the SRR scheme has the lowest intercept rate among the three methods and the intercept rate exhibits a declining trend as the distance between the source node and sink node increase. This occurs because the farther the distance is from the source node to the sink node, the more difficult it is for the adversary to intercept data packets sent on random routing paths. In the SRR scheme, there is a rapid decline in the intercept rate when the distance to the sink node changes from five to six. When the hop count of the source node is larger than five and exceeds t_{hop}, the number of candidate sectors for phantom sources increases. The diversity of routing paths is also increased and eventually results in a decline in the intercept rate. In the CRR scheme, the routing paths are concentrated in an area between the source node and the sink node because of the constrained offset angles. Therefore, there is a high probability that the adversary can intercept new data packets, which explains the high intercept rate.

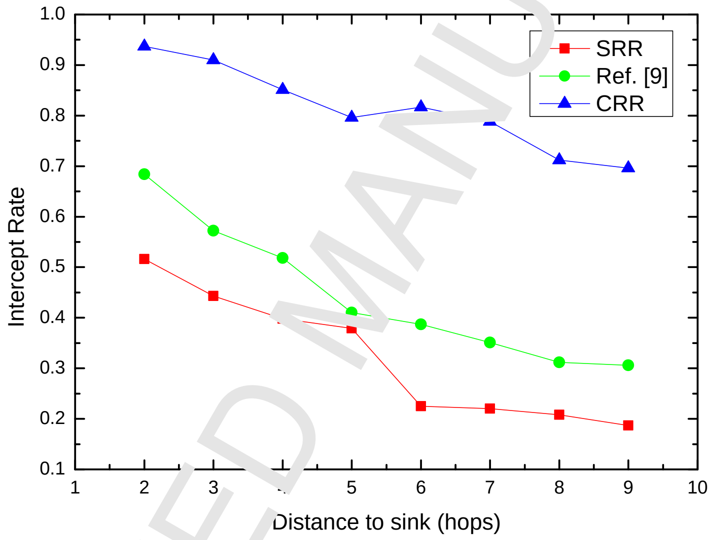


Figure 8: The intercept rate with different hop distance from source to sink.

Figure 9 shows an increasing trend in the intercept rate with increasing side length of the grid. Similar to the result shown in Figure 7, an increase in the side length means that fewer neighbor nodes can be selected as the next hop when transmitting the data packets. Therefore, it is more likely that the adversary can intercept new signals, which is reflected in the increasing trend of the intercept rate.

(3) Energy consumption.

In this experiment, the energy consumption refers to the average energy consumed by transmitting one data packet from the source node to the sink node. As shown in Figure 10, we can analyze the simulation results based on two aspects. First, when the hop count of the source node is smaller than t_{hop}, the scheme of Ref. [9] performs slightly better than our scheme. This happens because our scheme centralizes the routes in the sectors away from the source node as shown in Figure 5(a). The average routing length is longer than in the scheme of Ref. [9]; therefore, our scheme consumes more energy on average. Second, if the hop count of the source node is larger than t_{hop}, our scheme decreases the average routing length by reducing the probability of selecting the sectors that are far away from the source node. Therefore, our scheme performs clearly better than the scheme of Ref. [9] in terms of average energy consumption. In the CRR scheme, the offset angles reduce the randomness of the routings, which shortens the average length of the routing paths. Therefore, the average energy consumption of the CRR scheme is the lowest.

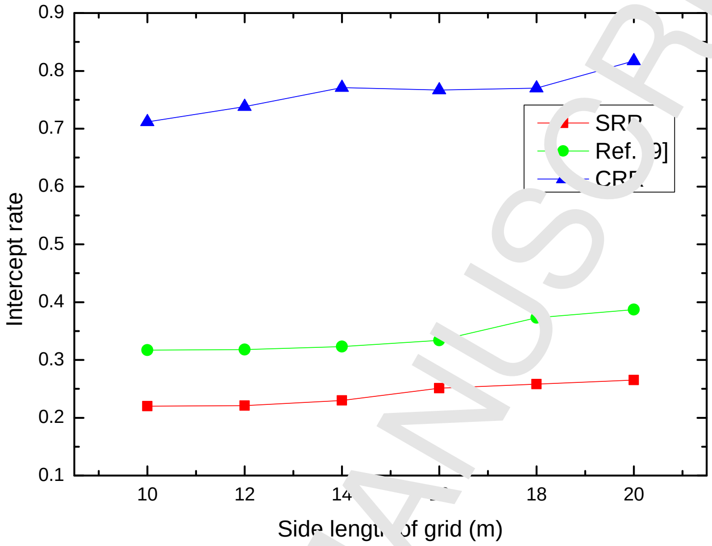


Figure 9: The intercept rate with different side length of deployment grid.

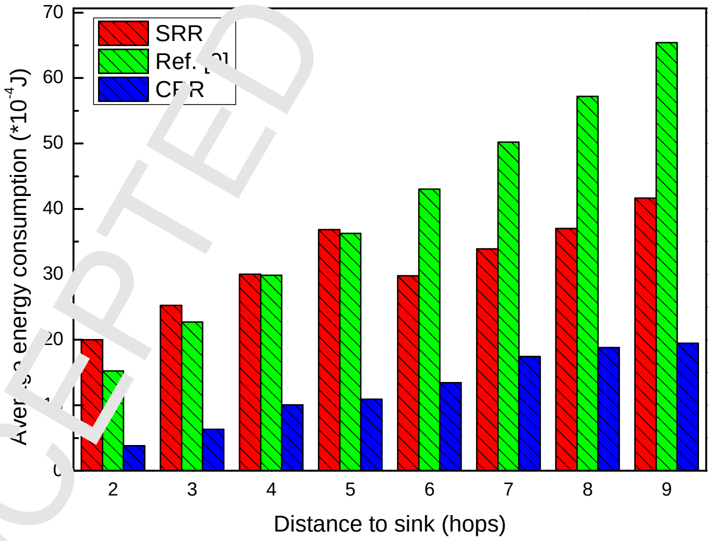


Figure 10: The average energy consumption with different hop distance from source to sink.

(4) Transmission delay.

Figure 11 shows the transmission delay versus the hop count of the source node. A low transmission delay ensures the timeliness of the collected information. It is observed in the figure that the transmission delay is lowest for the CRR scheme. This occurs because the sum of the offset angles is preset by the source node and the lengths of the random routes are constrained in the CRR scheme. The data packets are transmitted by the shortest route instead of using random routing in the final stage of transmission. In addition, when the hop count of the source is less than t_{hop}, although the average routing length is theoretically larger for the SRR scheme than the scheme in Ref. [9], in reality, the difference is not very large due to the random routing. That is the reason why the performance of our scheme is similar to that of Ref. [9]. However, our scheme performs better than that of Ref. [9] when the hop count of the source node is larger than t_{hop} on account of controlling the length of the routing paths by the dynamic candidate domain and the probability set.

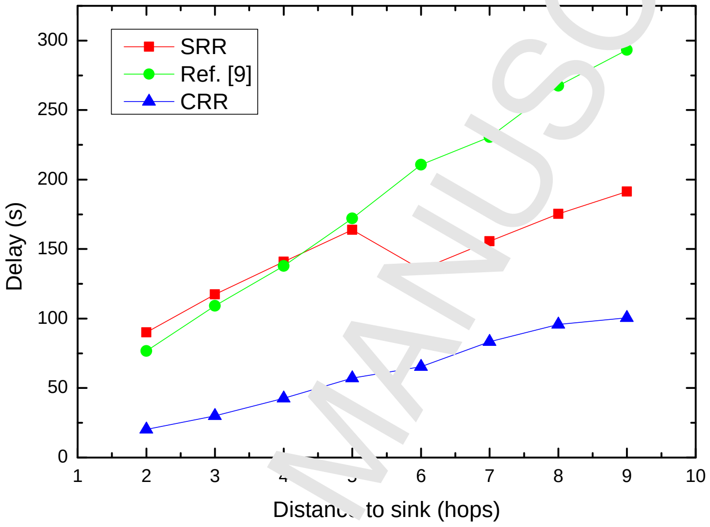


Figure 11: The transmission delay with different hop distance from source to sink.

7. Conclusions

In this paper, we proposed an SRR protocol for WSNs to protect SLP and balance the energy consumption. In the SRR, the source node divides the network into several virtual sectors. The data packets are transmitted through the different sectors and the dispersed routing paths to improve the security. In addition, we reduce the selection probability of the sectors close to the source node and keep the annular routing away from the visible area to prevent a direction attack. In addition, we use a hop threshold to adjust the routing strategy based on the relationship between the hop threshold and the hop count of the source node to ensure a balance between the security and the network lifetime. The analysis and simulation results demonstrate that our scheme provides efficient protection for SLP. In a future study, we plan to explore more energy-efficient methods to address the SLP issue based on multiple source nodes or mobile sink nodes.

Acknowledgements

The work is supported by the National Key Research and Development Program, No.2018YFC0407900 and the National Natural Science Foundation of China under Grant No.61572172 and No.61872124, and supported by the Fundamental Research Funds for the Central Universities, No.2016B10714 and supported by Six talent peaks project in Jiangsu Province, No.XYDXXJS-007.

References

[1] Jianbing Ni, Kuan Zhang, Xiaodong Lin, Xuemin Shen, “Securing Fog Computing for Internet of Things Applications: Challenges and Solutions,” *IEEE Communications Surveys and Tutorials,* 2017, vol. 20, no. 1, pp. 601 - 628.

[2] Ke Zhang, Supeng Leng, Yejun He, Sabita Maharjan, Yan Zhang, “Mobile Edge Computing and Networking for Green and Low-Latency Internet of Things,” *IEEE Communications Magazine,* 2018, vol. 56, no. 5, pp. 39 - 45.

[3] Guangjie Han, Li Liu, Wenbo Zhang, Sammy Chan, “A Hierarchical Jammed-Area Mapping Service for Ubiquitous Communication in Smart Communities,” *IEEE Communications Magazine,* 2018, vol. 56, no. 1, pp. 92-98.

[4] Guangjie Han, Xuan Yang, Li Liu, Wenbo Zhang, “A Joint Energy Replenishing and Data Collection Algorithm in Wireless Rechargeable Sensor Networks,” *IEEE Internet of Thing Journal,* 2018, vol. 5, no. 4, pp. 2596-2604.

[5] Guangjie Han, Xuan Yang, Li Liu, Wenbo Zhang, Mohsen Guizani, “A Disaster Management-Oriented Path Planning for Mobile Anchor-Based Localization in Wireless Sensor Networks,” *IEEE Transactions on Emerging Topics in Computing,* 2017, DOI: 10.1109/TETC.2017.2687319.

[6] Yun Li, Jian Ren, Jie Wu, “Quantitative measurement and design of source-location privacy schemes for wireless sensor networks,” *IEEE Transactions on Parallel and Distributed Systems,* 2012, vol. 23, no. 7, pp: 1302-1311.

[7] Celal Ozturk, Yanyong Zhang, Wade Trappe, “Source-location privacy in energy-constrained sensor network routing,” *ACM workshop on Security of ad hoc and sensor networks,* 2004, vol. 255, pp. 88-93.

[8] W-P Wang, Liang Chen, J-X Wang, “A source-location privacy protocol in WSN based on locational angle,” *IEEE International Conference on Communications,* 2008, pp. 1630-1634.

[9] Lin Yao, Lin Kang, Fangyu Deng, “Protecting source-location privacy based on multirings in wireless sensor networks,” *Concurrency and Computation: Practice and Experience,* 2015, vol. 27, no. 15, pp. 3863-3876.

[10] Wenlong Chen, Mingshu Zhang, Guangwu Hu, “Constrained Random Routing Mechanism for Source Privacy Protection in WSNs,” *IEEE Access,* 2017, vol. 5, pp. 23171-23181.

[11] Xiaoyan Wang, Tao Yang, Michel Wortmann, “Analysis of multi-dimensional hydrological alterations under climate change for four major river basins in different climate zones,” *Climatic Change,* 2017, vol. 141, no. 3, pp. 483-498.

[12] Tao Yang, Tong Cui, Chong-Yu Xu, “Development of a new IHA method for impact assessment of climate change on flow regime,” *Global and Planetary Change,* 2017, vol. 156, no. 9, pp. 68-79.

[13] Ching-Sheng Huang, Tao Yang, Hund-Der Yeh, "Review of analytical models to stream depletion induced by pumping: Guide to model selection," *Journal of Hydrology,* 2018, vol. 561, no. 6, pp. 277-285.

[14] Celal Ozturk, Yanyong Zhang, Wade Trappe, "Enhancing source-location privacy in sensor network routing," *Distributed Computing Systems,* 2005, pp. 599-608.

[15] Guangbao Tan, Wei Li, Jie Song, "Enhancing source location privacy in energy-constrained wireless sensor networks," *International Conference on Computer Science and Information Technology,* 2014, vol. 255, pp. 279-289.

[16] Xi Luo, Xu Ji, Myong-Soon Park, "Location privacy against traffic analysis attacks in wireless sensor networks," *International Conference on Information Science and Applications,* 2010, vol. 154, no. 5, pp. 1-6.

[17] Shruti Gupta, Prabhat Kumar, J. P. Singh, M. P. Singh, "Privacy Preservation of Source Location Using Phantom Nodes," *Information Technology: New Generations,* 2016, vol. 448, pp. 247-256.

[18] Yong Xi, Loren Schwiebert, Weisong Shi, "Privacy preserving shortest path routing with an application to navigation," *Pervasive and Mobile Computing,* 2014, vol. 13, pp. 142-149.

[19] Di Tang, Tongtong Li, Jian Ren, "Cost-aware secure routing (CASER) protocol design for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems,* 2015, vol. 26, no. 4, pp. 960-973.

[20] Yi Yang, Min Shao, Sencun Zhu, "Towards statistically strong source anonymity for sensor networks," *ACM Transactions on Sensor Networks,* 2013, vol. 9, No. 3.

[21] Wenbo Yang, Wen Tao Zhu "Protecting source location privacy in wireless sensor networks with data aggregation," *International Conference on Ubiquitous Intelligence and Computing,* 2010, vol. 6406, pp. 252-266.

[22] Arshad Jhumka, Matthew Leeke, Sambid Shrestha, "On the use of fake sources for source location privacy: Trade-offs between energy and privacy," *The Computer Journal,* 2011, vol. 54, no. 6, pp. 860-874.

[23] Honglong Chen, Wei Lou, "From nowhere to somewhere: protecting end-to-end location privacy in wireless sensor networks," *Performance Computing and Communications Conference,* 2010, pp: 1-8.

[24] Mohamed M. E. A Mahmoud, Xuemin Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems,* 2012, vol. 23, no. 10, pp. 1805-1818.

[25] Jun Long, Mianxiong Dong, Kaoru Ota, "Achieving source location privacy and network lifetime maximization through tree-based diversionary routing in wireless sensor networks," *IEEE Access,* 2014, vol. 2, pp. 633-651.

[26] Proano Alejandro, Loukas Lazos, Marwan Krunz, “Traffic decorrelation techniques for countering a global eavesdropper in WSNs,” *IEEE Transactions on Mobile Computing,* 2017, vol. 16, no. 3, pp. 857-871.

[27] Changqin Huang, Ming Ma, Yuxin Liu, “Preserving source location privacy for energy harvesting WSNs,” *Sensors,* 2017, vol. 17, no. 4.

[28] Ouyang Y, Le Z, Chen G, “Entrapping adversaries for source protection in sensor networks, International Symposium on World of Wireless,” *Mobile and Multimedia Networks,* 2006, pp. 23-24.

[29] L. Kazatzopoulos, C. Delakouridis, G. F. Marias, “iHIDE: Hiding sources of information in WSNs,” *International Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing,* 2006, pp: 41-48.

[30] Kiran Mehta, Donggang Liu, Matthew Wright, “ Protecting location privacy in sensor networks against a global eavesdropper,” *IEEE Transactions on Mobile Computing,* 2012, vol. 11, no. 2, pp. 320-336.

[31] Guangjie Han, Lina Zhou, Hao Wang, “A source location protection protocol based on dynamic routing in WSNs for the Social Internet of Things,” *Future Generation Computer Systems,* 2018, vol. 82, pp. 689-697.

[32] Mauro Conti, Jeroen Willemsen, Bruno Crispo, “Providing source location privacy in wireless sensor networks: A survey,” *IEEE Communications Surveys and Tutorials,* 2013, vol. 15, no. 3, pp. 1238-1280.

[33] Hongwei Li, Dongxiao Liu, Yuanshun Dai, “Engineering searchable encryption of mobile cloud networks: When QoE meets QoP,” *IEEE Wireless Communications,* 2015, vol. 22, no. 4, pp. 74-80.

[34] Hongwei Li, Yi Yang, Tom H Luan, “Enabling Fine-Grained Multi-Keyword Search Supporting Classified SubDictionaries over Encrypted Cloud Data,” *IEEE Transactions on Dependable and Secure Computing,* 2016, vol. 13, no. 3, pp. 312-325.

Guangjie Han is currently a Professor with the Department of Information and Communication System, Hohai University, Changzhou, China. He received the Ph.D. degree from Northeastern University, Shenyang, China, in 2004. From 2004 to 2006, he was a Product Manager for the ZTE Company. In February 2008, he finished his work as a Postdoctoral Researcher with the Department of Computer Science, Chonnam National University, Gwangju, Korea. From October 2010 to 2011, he was a Visiting Research Scholar with Osaka University, Suita, Japan. He is the author of over 220 papers published in related international conference proceedings and journals, and is the holder of 90 patents. His current research interests include sensor networks, computer communications, mobile cloud computing, and multimedia communication and security. Dr. Han has served as a Co-chair for more than 50 international conferences/workshops and as a Technical Program Committee member of more than 150 conferences. He has served on the Editorial Boards of up to 14 international journals, including the IEEE ACCESS, Telecommunication Systems, International Journal of Ad Hoc and Ubiquitous Computing, Journal of Internet Technology and KSII Transactions on Internet and Information Systems. He guest edited a number of special issues in IEEE Journals and Magazines. He has served as a Reviewer of more than 50 journals. He had been awarded the ComManTel 2014, ComComAP 2014, Chinacom 2014 and Qshine 2016 Best Paper Awards. He is a member of IEEE and ACM.

Yu He received the B.S. degree in College of Internet of Things from Hohai University, China, in 2017, where he is currently pursuing the M.S. degree. His current research interests are location privacy protection for wireless sensor networks and internet of things.

Hao Wang received the B.S. degree from Nanjing Agriculture University, Nanjing, China, in 2015. He is currently pursuing the M.S. degree with College of Internet of Things Engineering. His research interests include Protection of location privacy in wireless sensor networks.

James Adu Ansere received his BSc in Physics in 2007, from University of Cape Coast, Ghana and MSc in Telecommunication Engineering from Blekinge Institute of Technology, Sweden, 2012. He was a lecturer at Sunyani Technical University. He is a peer reviewer for International Telecommunication System and has published twelve (12) papers in both international and national papers. Currently, he is PhD Student at the Department of College of Internet of Things

Engineering, Hohai University, China. His research interests are in 5G wireless networks, internet of things and wireless sensor networks. He is a recipient of the *Sparbanksstiftelsen Kronan* award in Sweden for Master Thesis, 2012 and Fellowship Award (*FCIDA*) from Civilian Institute of Democratic Administration, West Africa in 2017.

Highlights for this paper are listed as follows:

Location privacy is a significant and challenging security issue in WSNs.
In this paper, we propose a SRR algorithm for WSNs to protect source location privacy.
Dividing deployment area into sectors to enhance randomness of routing paths.
Concept of hop threshold is designed for balancing security and energy consumption