IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2017, accepted August 28, 2017, date of publication September 14, 2017, date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2752179

Constrained Random Routing Mechanism for Source Privacy Protection in WSNs

WENLONG CHEN[1], MINGSHU ZHANG[1], GUANGWU HU[2], XIAOLAN TANG[1], AND ARUN KUMAR SANGAIAH[3]

[1]College of Information Engineering, Capital Normal University, Beijing 100000, China
[2]Shenzhen Institute of Information Technology, Shenzhen 518029, China
[3]School of Computer Science and Engineering, VIT University, Vellore 632014, India

Corresponding author: Guangwu Hu (hugw@sziit.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant 61373161 and Grant 61502320, in part by the Natural Science Foundation of Guangdong Province under Grant 2015A030310492, in part by the Fundamental Research Project of Shenzhen Municipality under Grant JCYJ20160301152145171, and in part by the Beijing Advanced Innovation Center for Imaging Technology.

ABSTRACT In wireless sensor networks, it is a typical threat to source privacy that an attacker performs backtracing strategy to locate source nodes by analyzing transmission paths. With the popularity of the Internet of Things in recent years, source privacy protection has attracted a lot of attentions. In order to mitigate this threat, many proposals show their merits. However, they fail to get the tradeoff between multi-path transmission and transmission cost. In this paper, we propose a constrained random routing mechanism, which can constantly change routing next-hop instead of a relative fixed route so that attackers cannot analyze routing and trace back to source nodes. First, we design a specific selection domain which is located around the sending node according to the dangerous distance and the wireless communication range. Then sending nodes calculate the selected weights of the candidate nodes according to their offset angles in this domain. Finally, the selected weights help to decide which node will become the next hop. In this way, attackers would be confused by the constantly changing paths. The simulation results prove that our proposal can achieve high routing efficiency in multi-path transmission, while only introducing a controllable energy consumption, end-to-end delay and redundant paths.

INDEX TERMS Wireless sensor networks, source privacy protection, offset angle, probability.

I. INTRODUCTION

Wireless sensor networks (WSNs) are composed of a large number of low-cost and non-rechargeable micro sensors through the way of self-organization. They have been widely used in military and civilian applications as a potential technology. Different from traditional wired networks, WSNs are usually deployed in unmanned areas, and message packets are transmitted through wireless electromagnetic signals. Therefore, WSNs are under increasing threat of privacy disclosure, interception or tampering, even if the high density complex data encryption algorithm has already been used. For example, an attacker can locate the positions of rare animals by monitoring the wireless link and capture them [1]. The privacy protection of sensor nodes becomes a significant issue, though lots of methods are presented to provide real-time and efficient routing for data delivery, such as [2]–[4].

The attacker usually uses expensive wireless receivers to determine the position of signal transmitting nodes, and then moves to the node to monitor continuously. Repeating the above procedures, the attacker can perform the backtracing strategy to find the position of the source node. It is extremely urgent to study the source privacy problem exposed in WSNs because it has seriously hindered its development, application, and research on imminent source privacy protection. The existing researches based on multi-path transmission have the disadvantages of high communication delay or high communication energy consumption. Therefore, we focus on improving the efficiency of source privacy protection.

In this paper, a Constrained Random Routing (CRR) mechanism based on the transmitting offset angles and constrained probability is proposed to protect the privacy of source nodes. CRR prevents attackers from performing the backtracing strategy to locate source nodes effectively. Source nodes

do not send message packets to the sink node through any specific optimal transmission strategy (e.g., the shortest-path solution). Sending nodes select next hops randomly under constrained offset angles. Therefore, attackers are confused by the changing paths. CRR fully considers energy consumption in multi-path transmission. The selected relay nodes are relatively close to the sink node to ensure that the lengths of transmission paths are as short as possible compared with the shortest-path solution. System evaluation shows that CRR has great performance to protect source privacy.

Compared to existing studies, the main contributions of our mechanism are as follows:

(1) We propose a random routing mechanism to prevent attackers from tracing back to locate source location hop by hop under the constrained offset angles and constrained probability. With the usage of CRR, it is rather difficult for attackers to obtain the location information of some nodes.

(2) Rectangular coordinate is used in the process of the next-hop selection and the construction of the whole transmission paths.

(3) To minimize the lengths of transmission paths, CRR gives greater priority to the nodes with smaller offset angles to be selected as relay nodes. Meanwhile, the construct of the whole transmission path are under the constraint of the sum of offset angles.

The rest of this paper is organized as follows. Section II summarizes the related work to provide a whole picture in this area. Section III introduces the attack model and our design goals. Section IV and Section V elaborate CRR's detail and its security analysis, respectively. Finally, Section VI concludes the paper.

II. RELATED WORK

The existing studies mainly adopt multi-path transmission. The source node does not choose the shortest path to transmit packets, but choose one or a plurality of camouflage paths as the communication link deliberately to confuse attackers.

In early 2004, researchers began to study the source protection problem. For example, Celal *et al.* [5] proposed a classic model named Panda-Hunter where a large number of sensor nodes are randomly deployed in a large protected area to observe pandas' living habits. As long as pandas are found in the monitoring area, the corresponding sensor nodes will send packets to the base station through multi-hops periodically until the target disappears. Pandurang *et al.* [6] proposed a technique called phantom routing(PR), which PR has been proven flexible and capable of protecting source locations, even the source is mobile in the network. However, it is unable to defend against global attacks. Wang *et al.* [7] proposed a suboptimal but practical privacy-aware routing scheme called WRS. It ensures a longer security period but the location information is leaked. In [8], sink nodes move around along random paths to collect packets from the local nodes, preventing the attackers from predicting their locations and movements.

Two new identities, route and location (IRL and r-IRL) privacy algorithms and data privacy mechanism were proposed by Shaikh *et al.* [9]. They use two notions: direction and trust to provide reliable (non-malicious and non-faulty) secure transmission paths. Chen J et al. proposed source location privacy preservation Protocol Using Source-Based Restricted Flooding(PUSBRF) in [10]. It ensures the first h hops away from the real source node. Enhanced source location privacy preservation Protocol Using Source-based Restricted Flooding (EPUSBRF) was designed under the consideration of an attacker with enhanced visual ability. The protocol marks the nodes in the visible region while in the process of restricted flooding, and it uses a broadcast strategy which makes message packets avoid the visible areas completely in the shortest path routing process. Average safety periods in PUSBRF and EPUSBRF are increased substantially. Kang [11] proposed a Location Privacy Support Scheme (LPSS) which introduced the conception of gradient. In LPSS, the protection strength increases exponentially with the increase in distance between the sink and the source node. A scheme was presented in [12] to hide source information using cryptographic techniques incurring lower overhead. The packet is modified and routed by dynamically selecting nodes to make it harder for an attacker to traceback to a source node.

Xi et al. [13] proposed a two-phase privacy protection mechanism Greedy Random Walk (GROW). The source node and the sink node perform random walks at the same time and then the two random-walk paths connect together to form the complete transmission path. GROW can prevent the attacker from tracing back to the source node although with a long latency. Lightfoot L [14] proposed Sink Toroidal Region Routing (STaR) in which sensor nodes change their IDs dynamically and periodically. STaR effectively prevents the attacker from locating the source node through ID correlation analysis. Network Mixing Ring (NMR) was proposed by Li and Ren [15] and Yun *et al.* [16]. They evaluate the strength of source privacy protection qualitatively with three indexes and make attempts to put forward general strength evaluation indexes in privacy protection mechanism. Ren and Tang [17] proposed a scheme where message packets are routed to an intermediate node selected from a hierarchical Connected Dominating Set (CDS) of the network. CDS represents the backbone of the network and the nodes in CDS are located in different regions of the network. The selection of the intermediate node can effectively prevent the adversary from performing routing backtracing attack to identify the message source node. The researchers in [18] proposed an opportunistic mesh networking scheme, where each sensor transmits packet over a dynamic path to the destination, making it difficult for an adversary to trace back hop by hop to the origin of the sensor communication.

The devised solution consisting of two complementary schemes that hinder both traffic analysis and the node capture attack was proposed in [19]. It considers delay, but the intensity of privacy protection needs to be improved. In [20], an energy efficient preserving sensor location

privacy based on the ant colony optimization scheme was proposed to protect the sensor locations. However, it requires the sensor nodes to have strong capability of computing. In [21], Path Extension Method (PEM) was proposed to provide strong protection for source-location privacy. In PEM, fake sources are generated dynamically after the source sends packets to the sink node. It also performs quite well even though an object occurs near the base station. The researchers in [22]–[25] make contributes to increase safety period and energy efficiency by creating alternate paths. With coordinates of source node and sink node, the algorithm proposed in [22] selects a node randomly as the expected phantom source node. By selecting dispersive expected phantom source nodes, transmission paths of adjacent packets are dispersive. [23] forms an Isolated Adversary Zone (IAZ) by sacrificing a certain percentage of sensor nodes and entraps patient attackers in that zone. The researchers in [26]–[30] proposed a variety of multi-path forwarding mechanisms based on angles and the source privacy has been well protected.

III. SYSTEM OVERVIEW

A. ATTACK MODEL

An attacker may be well-equipped with advanced and powerful transceivers to locate the source node through analyzing traffic patterns. Suppose an attacker has the following abilities.

(1) The attacker has sufficient energy, advanced computing ability and enough memory to store data.

(2) The attacker does not interfere with normal network function, such as tampering with data packets or destroy sensor equipment because these behaviors can be found easily. The attacker will implement some negative attacks, such as eavesdropping on communication networks.

(3) The attacker can monitor an area and get all the packets transmitted in this area. Once an event is detected, the attacker analyzes the signal intensity and determines the direction of the sending node then quickly moving to it.

In Fig.1, packets are sent through a fixed path $s \rightarrow N_1 \rightarrow N_2 \rightarrow N_3$, and an attacker stays nearby N_4. Once a message packet is routed to N_3, the attacker will immediately detect the target and quickly move to N_3. Then the attacker continue staying nearby N_3 and waiting for a next packet. The attacker will finally trace back to s hop by hop by repeating the above procedure.

B. SYSTEM GOALS

Our system goals are summarized as follows.

(1) Attackers hardly trace back to locate the source node by analyzing the transmission paths hop by hop. In other words, the implementation cost is extremely high if attackers want to obtain the location information of some nodes.

(2) Energy consumption, end-to-end delay and other transmission performance could be effectively controlled. Sending nodes do not choose specific paths to transmit packets.

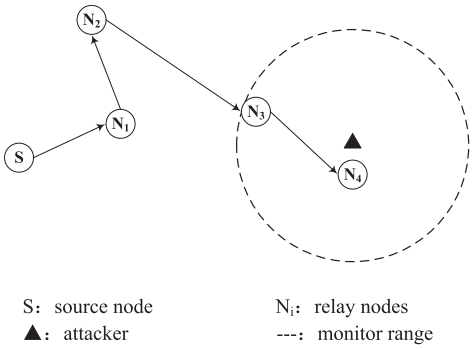


FIGURE 1. Attack model.

But the lengths of transmitting paths are as short as possible in the case of random routing.

C. SYSTEM ARCHITECTURE

The system architecture in this paper is described as follows.

(1) The target network is composed of a number of sensor nodes with limited energy. The sensor nodes are deployed randomly.

(2) There is only one base station called sink node in the area which is the only destination that sensor nodes send message packets to through a multi-hop strategy.

(3) Each node knows the location information of some other nodes. The location of each node corresponds to a point in the plane. Each node has knowledge of the location of its neighbor nodes and the sink node.

Note that data encryption like key management and generation algorithm is beyond the discussion of this paper. Meanwhile, the primary purpose of our mechanism is to monitor and collect packets, thus we only take one-way message transmission into consideration.

IV. CONSTRAINED RANDOM ROUTING MECHANISM

The general idea of CRR is as follows. To prevent attackers from tracing back to locate the source node hop by hop, first, the sending node (source node or each forwarding node) determines a specific selection domain for next-hop transmission according to the dangerous distance and the wireless communication range. Then, it analyzes the offset angles of the candidate nodes based on the direction of the nodes to the sink node. Lastly, the sending node calculates the selected weights of the candidate nodes according to their offset angles, and the selected weights help to decide which node will become the next-hop. Meanwhile, CRR ensures that energy consumption and end-to-end delay are under control of the constrained sum of offset angles. It will only introduce a small amount of redundant hops compared with the shortest-path solution.

A. NEXT-HOP SELECTION STRATEGY

The next-hop selection strategy is designed for the nodes which are going to send a packet towards the sink node. These nodes are called sending nodes, including the source node and the intermediate forwarding nodes.

TABLE 1. Summary of notations.

Symbol	Meaning
N_i	The sensor nodes in the network, $i = 1, 2, ..., n$
N_{src}	The source node
N_{sink}	The sink node
θ_i	The offset angle of N_i, $\|\theta_i\| \in [0°, 180°]$
r_{comm},r_1	The communication range of each sensor node
d_{danger},r_2	The dangerous distance that is easy to trace back to the source node for an attacker
m	The center of ellipse is located in (m,0)
a	The length of elliptical long axis
b	The length of elliptical short axis
e	The eccentricity of ellipse
p_i	The selected weight that the sending node selects N_i as the next relay node
n_{rand}	The random number generated by the system which helps to select the next relay node
α_i	The supplementary angle of θ_i
α_{sum}	The sum of α_i
λ	The sum of offset angles
δ	The the rest of λ
S	The set of candidate nodes of a sending node
k_λ	A coefficient that can affect the value of λ
avg_hop	The number of average hops from a source to the sink node in CRR
cr	The two adjacent path coincidence ratio
$NUM_{coincident}$	The number of coincident hops of two successive transmission paths from the same source node
NUM_{hops}	The number of hops of the last adjacent path
h_{sp}	The number of hops from a source node to the sink node through the shortest-path solution
h_{CRR}	The number of hops from a source node to the sink node in CRR
rpr	The redundant path ratio
$rdtr$	The reserve direction transmission ratio
h_{rdtr}	The number of hops of reverse direction transmission
avg_rdtr	The number of average $rdtr$

Definition 1: Offset Angle. Build a rectangular coordinate and make the sending node as the origin. The sending node becomes the origin of the rectangular coordinate whenever it has a packet to send. Make the connection line from the sending node to the sink node as X-axis. Y-axis is perpendicular to X-axis and is through the sending node. On the plane, the inclination angle formed by X-axis and the connection line of a sensor node to the origin, is called its offset angle. In Fig.2, θ_1, θ_2 and θ_3 are the offset angles of N_1, N_2 and N_3, respectively. Offset angles are carried by the packets during routing.

Obviously, the offset angle of each ordinary node is always a fixed value due to the unmovable position of the sink node in the network.

The method to select the next relay node according to CRR is as follows.

1) DETERMINE THE SELECTION DOMAIN

The sending node builds a rectangular coordinate as Definition 1. First, it determines a dangerous distance d_{danger} according to the specific network requirement and selection strategy. It is easy to trace back to the sending node in this distance range. Then, the sending node determines the next-hop selection range according to the offset angles of the candidate nodes. That is, the smaller the offset angle is, the longer the next-hop selection domain boundary. Therefore, the next-hop selection range is represented by an ellipse. It is assumed that the communication range of each sensor node is r_{comm}. Shown as the shadow area in Fig.2, the selection domain for the next relay node is an ellipse which is calculated according to r_{comm}, d_{danger}. The calculation formula for the selection domain is,

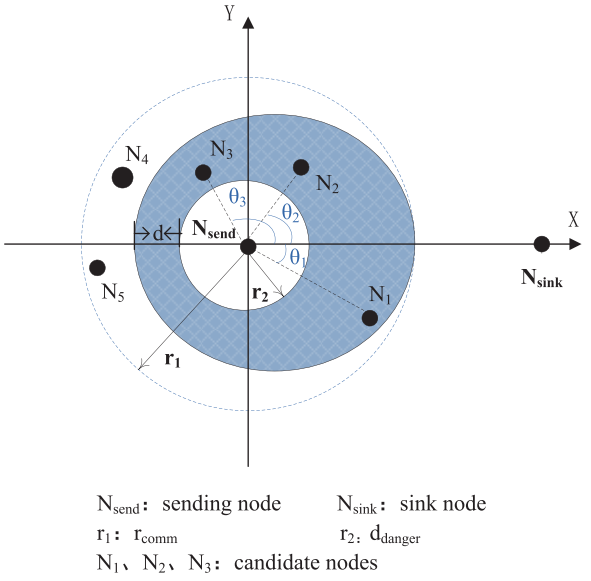


FIGURE 2. Next-hop selection domain.

$$\frac{(x-m)^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{1}$$

The related parameters in this formula are explained as follows.

(1) The sending node is located at the origin of the coordinate plane;

(2) r_1 stands for r_{comm}, and r_2 for d_{danger};

(3)The center of the ellipse is located in (m,0), and $m = \frac{1}{2}(-d - r_2 + r_1)$;

(4) a is the length of elliptical long axis, and $a = r_1$-m;

(5) b is the length of elliptical short axis which is determined by eccentricity e. That is, $b = a\sqrt{1-e^2}$;

(6) d can be set dynamically but it satisfies $0 < d \leq r_1 - r_2$.

The sending node selects a next relay node in the selection domain. The size of the selection domain varies with different value of r_1, r_2 and d. That is, the number of nodes in the domain changes as well, which affects the randomness of the next-hop selection. Moreover, a node with smaller offset angle has more probability to be selected as the next relay node even if it is further to the sink node, such as N_1 in Fig.2. In order to minimize the lengths of transmission paths, nodes with larger offset angles are left out of the candidate nodes, such as N_4 and N_5. Thus, the transmission paths are as short

as possible compared with the paths of the shortest-path solution.

2) HOW TO SELECT THE NEXT RELAY NODE

It is assumed that there are n candidate nodes in the selection domain. The offset angle of N_i is $|\theta_i|$, $i = 1, 2, \ldots, n$. For the convenience of expression, the offset angle is expressed in absolute value. Maintaining the generality, the nodes are arranged with their offset angles:

$$|\theta_1| < |\theta_2| < \ldots < |\theta_n|, 1 \le i \le n, |\theta_i| \in [0°, 180°] \quad (2)$$

We try to minimize the transmission paths in consideration of reducing the energy consumption while selecting the next hop in the selection domain. So the smaller the offset angle a node has, the more possibility it has to be selected as the next hop. The details of the method are as follows.

Let offset angle of each node correspond to a supplementary angle α_i, $\alpha_i = \pi - |\theta_i|$, and $\alpha_{sum} = \sum_{i=1}^{n} \alpha_i$. The selected weight that the sending node selects N_i as the next relay node is:

$$p_i = \frac{\alpha_i}{\alpha_{sum}} \quad (3)$$

It is clear that $\sum_{i=1}^{n} p_i = 1$ and $p_i > 0$. Then the next relay node is selected according to the random number n_{rand} which is generated by the system, $0 < n_{rand} \le 1$. The method is as follows with the usage of n_{rand}.

(1) If $0 < n_{rand} \le p_1$, N_1 is selected to be the next relay node;

(2) If $\sum_{i=1}^{j-1} p_i < n_{rand} \le \sum_{i=1}^{j} p_i$, N_i is selected to be the next relay node.

After the sending node selects the next relay node, the selected node begins to send packets to the sink node as the new message source. It still adopts the above method to select its next relay node.

Now an example is used to illustrate our mechanism. Shown in Fig.2, there are three nodes in the selection domain, N_1, N_2 and N_3, whose offset angles are $|\theta_1| = 30°$, $|\theta_2| = 50°$, $|\theta_3| = 135°$ and the weighs to be selected as the next relay node are $p_1 = 0.46$, $p_2 = 0.4$, $p_3 = 0.14$, respectively. n_{rand} is generated to decide which one should be selected.

(1) If $0 < n_{rand} \le 0.46$, N_1 will be selected;

(2) If $0.46 < n_{rand} \le 0.86$, N_2 will be selected;

(3) If $0.86 < n_{rand} \le 1$, N_3 will be selected.

B. THE WHOLE TRANSMISSION PATH STRATEGY

After the sending node selects the next relay node successfully, the selected node becomes the new sending node. The whole transmission path selection strategy in CRR is described as follows.

1) TRANSMISSION PATH CONSTRUCTION

Definition 2: Sum of Offset Angles. It refers to the sum of the offset angles of a series of selected next-hops that relay message packets to the sink node, expressed by λ.

If the path from the initial source node to the sink node has k relay nodes and their offset angles are $|\theta_{1\sim k}|$, then Formula (4) is established.

$$\sum_{i=1}^{k} |\theta_i| \le \lambda \quad (4)$$

When a packet is going to be forwarded to the j^{th} relay node N_j ($j < k$), the rest of λ is $\delta = \lambda - \sum_{i=1}^{j-1} |\theta_i|$. Then build a set S of the candidate nodes. Put all nodes whose offset angles satisfying Formula (5) into S.

$$S = \{N_i | |\theta_i| \le \delta\} \quad (5)$$

(1) If S is not $\emptyset$, the sending node selects a node randomly from S as the j^{th} forwarding node N_j according to the previous description;

(2) If S is $\emptyset$, the sending node selects the next hop based on the traditional shortest-path solution.

The next-hop selection method for node N_j repeats the procedure above until the message is successfully delivered to the sink node. In particular, if the sink node is in the communication range of the sending nodes, the message is forwarded directly to the sink node.

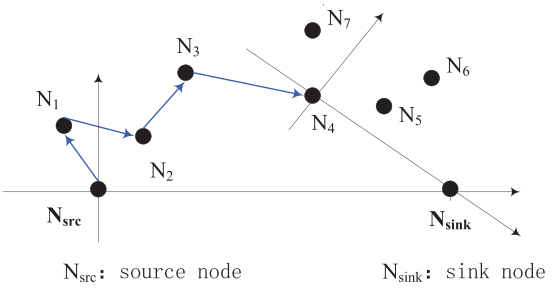


FIGURE 3. The whole transmission path strategy.

Now an example is given to illustrate the above strategy. As shown in Fig.3, N_{src} sends a message packet to N_4 through N_1, N_2, N_3. Their offset angles are $|\theta_1| = 140°$, $|\theta_2| = 5°$, $|\theta_3| = 50°$, $|\theta_4| = 15°$. Let $\lambda = 270°$. Now N_4 is the sending node and there is $\delta = \lambda - \sum\theta = 60°$. N_5, N_6, N_7 are three candidate relay nodes for N_4 and their offset angles are $|40°|$, $|50°|$ and $|130°|$, respectively. That is, $S = \{N_5, N_6\}$. N_4 will select N_5 or N_6 as the next relay node. It is assumed that N_4 selects N_6 as the next relay node. Then N_6 becomes the new sending node and $\delta = 10°$. At this time, the offset angles of N_6's candidate relay nodes are all larger than $10°$. Thus, $S = \emptyset$. Then N_6 will select its relay node based on the traditional shortest-path solution.

2) DETERMINE λ

In different network environments, the sum of offset angles λ should be set dynamically according to the random degree and transmission path redundancy. If an initial source node needs h_{sp} hops to send a message packet to the sink node through the shortest-path solution, then λ will be set as (6).

k_λ is a coefficient that can be set arbitrarily.

$$\lambda = k_\lambda * h_{sp} * 90^\circ \tag{6}$$

V. SECURITY ANALYSIS

In CRR, the source node calculates the selected weights according to the offset angles of the candidate nodes in the selection domain. Then the selected next hop becomes a new sending node and continues to select the next hop randomly through the constrained strategy. Meanwhile, the whole transmission path selection is limited by the sum of offset angles λ so as to ensure that the whole transmission paths are as short as possible. First, we will analyze that the distance from the next relay node to the sending node should be at least d_{danger}, which makes it difficult for the attacker to obtain the location of the real source node. Then the next relay node is selected according to the selected weights randomly, so the attacker does not know which node is selected exactly.

In fact, the attacker must receive message packets continuously from one source node if he wants to perform the backtrace analysis. That is, source location is easily traced if packets are transmitted on the fixed path undoubtedly. Assume that the probability of being captured is 1. An example is shown in Fig.1 that a message packet is sent to N_3 through N_1 and N_2 from the source node. Suppose that,

(1) Because the source node has m_0 nodes to select and the selected weight of each node is p_{0i} $(1 \leq i \leq m_o)$, $\sum_1^{m_0} p_{0\,i} = 1$;

(2) N_1 has m_1 nodes to select and the selected weight of each node is p_{1j} $(1 \leq j \leq m_1)$, so $\sum_1^{m_1} p_{1j} = 1$;

(3) N_2 has m_2 nodes to select and the selected weight of each node is p_{2k} $(1 \leq k \leq m_2)$, so $\sum_1^{m_2} p_{2k} = 1$.

If the source node selects N_1 as the next relay node with the selected weight p_{02}, N_1 selects N_2 as the next hop with the selected weight p_{13} and N_2 selects N_3 as the next hop with the selected weight p_{21}, we can get:

(1) The probability the attacker finds the previous relay node N_2 is deduced to p_{21}.

(2) The probability the attacker performs the backtrace analysis and finds the whole transmission paths correctly is deduced to $p_{02} * p_{13} * p_{21}$. Instead of determining a fix path beforehand, our mechanism chooses a path during the packets are transmitted through the network. Therefore, CRR can improve the complexity for the attacker of tracking significantly and provide effective source privacy protection.

VI. SYSTEM EVALUATION

In order to evaluate our proposal's performance, we set up a network model and evaluate from three aspects: convergence of transmission paths, coincidence ratio of transmission paths (cr) and reverse direction transmission ratio ($rdtr$). The simulation analysis is also presented.

The network area is divided into grids with the same side-length 1. Sensor nodes are deployed in the grid intersection points. A node is selected as the source node N_{src}. Then, build a rectangular coordinate and make N_{src} as the origin which is located in (0,0). There are 100 packets sent from N_{src} but the sink node varies in each experiment, shown in Table 2. The parameters of the elliptic curves are set as follows. $r_{comm} = 3$, $d_{danger} = 1$, $d = 1$, and λ is set dynamically according to the number of hops through the shortest-path solution. The elliptic equation is

$$\frac{(x-0.5)^2}{2.5^2} + \frac{y^2}{2.2^2} = 1 \tag{7}$$

The selection domain in experiments is shown as Fig.4. The red node is the sending node. The blue points are the candidate nodes that have different probabilities to be selected as the next relay node. It is clear that some nodes with larger offset angles are left out of the candidate nodes, such as some of the black ones.

TABLE 2. Experiment results Of *cr*.

$No.$	N_{sink}	$packets$	h_{sp}	λ	avg_hop	cr
1	(14,0)	100	5	450	8.25	0.08
2	(18,0)	100	6	540	10.20	0.05
3	(20,0)	100	7	630	11.60	0.04
4	(24,0)	100	8	720	13.37	0.03
5	(26,0)	100	9	810	14.55	0.01

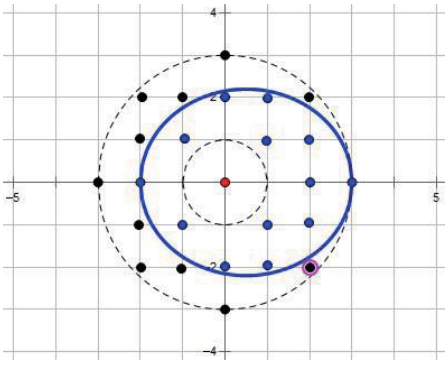
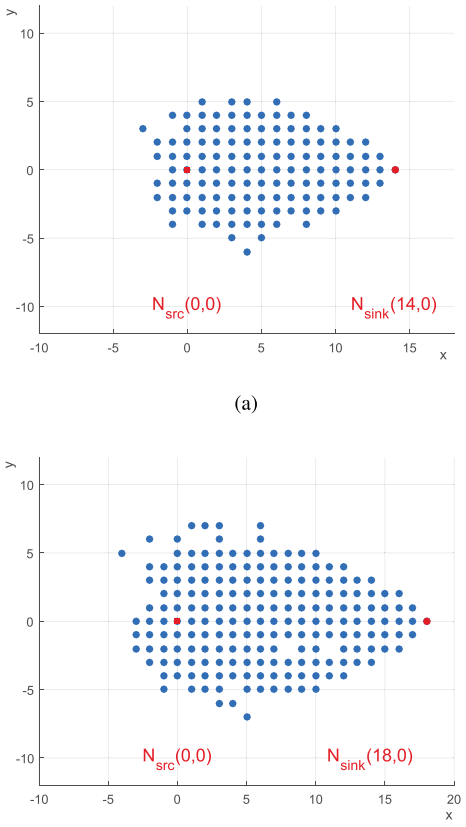


FIGURE 4. Selection domain.

A. THE CONVERGENCE OF TRANSMISSION PATHS

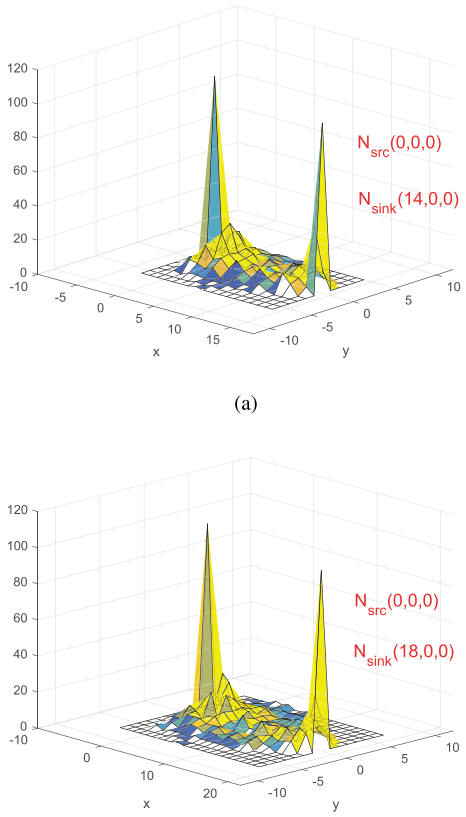
λ makes the transmission paths as short as possible. Therefore, the transmission paths are convergent. We would like to show the convergence of transmission paths more intuitively in CRR through five groups of experiments above. N_{sink} stands for the location of the sink node and λ is set dynamically according to h_{sp} which satisfies Formula (6), shown in Table 2.

The experiment results of No.1 and No.2 in Table 2 are used to illustrate the convergence of transmission paths which is shown in Fig.5. In each experiment, red points denote the source node and the sink node. Blue points are sensor nodes that participate in forwarding packets. It is clear that the transmission paths in CRR are random but convergent. The nodes that participate in forwarding message packets are distributed between the sending node and the sink node. Obviously, the distribution trend of them are convergent and around the shortest path formed through the shortest-path solution. More importantly, most of packets are sent to the direction of the sink node, and only a few are sent to the



(a) (b)

FIGURE 5. Convergence of transmission paths in CRR. (a) Convergence of experiment No.1. (b) Convergence of experiment No.2.



(a) (b)

FIGURE 6. Three-dimensional convergence of transmission paths in CRR. (a) Three-dimensional convergence of experiment No.1. (b) Three-dimensional convergence of experiment No.2.

reverse direction. Three-dimensional convergence of path is shown in Fig.6. The value of Z-axis stands for the number of packets a node forwards. The lighter the grid color is, the more packets it forwards. Namely, the smaller the node offset angle is, the more possible it is to be selected as the next hop.

B. THE COINCIDENCE RATIO OF TRANSMISSION PATHS

The attacker must receive message packets continuously from one source node if he wants to perform the backtrace analysis. That is, source location is easily traced if packets are transmitted on the fixed path. Therefore, we study the two adjacent paths coincidence ratio cr,

$$cr = \frac{NUM_{coincident}}{NUM_{hops}} \tag{8}$$

(1) $NUM_{coincident}$: the number of coincident hops of two successive transmission paths from the same source node;

(2) NUM_{hops} : the number of hops of last adjacent path.

Two adjacent paths are shown in Fig 7. The number of hops of Path 1 is 7. The number of coincident hops in Path 2 compared with Path 1 is 2. The coincident hops are $A \rightarrow B$ and $F \rightarrow G$. Thus, $cr = 0.29$.

Further study on coincidence ratio of two adjacent paths are made through five groups of experiments above. h_{sp} stands

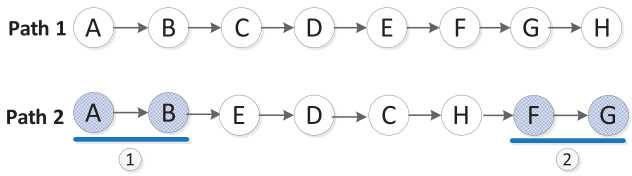


FIGURE 7. Coincidence ratio in 2 adjacent paths.

for the number of hops from a source node to the sink node through the shortest-path solution, λ for the sum of offset angles in CRR and *avg_hop* for the number of average hops from the source to the sink node in CRR. Then the cr of the two adjacent paths is shown in TABLE 2. It is clear that,

(1) CRR will only introduce a few redundant paths shown by *avg_hop*;

(2) cr is extremely low, and it decreases when *avg_hop* increases.

The values of cr prove that the number of coincident hops of two successive transmission paths from the same source node is little. That is, the probability that continuous message packets forwarded along the same path is extremely low. In other words, it is proved that message packets are sent on different paths in CRR. Hence CRR prevents the

attacker from finding the source node by trace-back analysis effectively.

C. REVERSE DIRECTION TRANSMISSION ANALYSIS

In CRR, a sending node selects its next relay node randomly in the selection domain with constrained probability. CRR gives greater priority to the nodes with smaller offset angles to be selected as relay nodes. Meanwhile, λ is used to guarantee that the transmission paths in CRR will not deviate too much from the paths in the shortest-path solution. However, the nodes with larger offset angles still can be selected as relay nodes. Now, we analyze the reverse direction transmission.

Definition 3: Reverse Direction Transmission. If a sending node selects N_i as its relay node with $|\theta_i| > 90°$, this situation is called reverse direction transmission. For example, N_{src} selects N_1 as its next relay node in Fig.3. This situation is a reverse direction transmission because $|\theta_1| = 140°$.

We use *rdtr* (reverse direction transmission ratio) to make quantitative analysis. It is assumed that a source needs h_{CRR} hops to send a packet to the sink node and h_{rdtr} hops are reverse direction transmission, obviously $h_{rdtr} \leq h_{CRR}$. Then *rdtr* is expressed as (9).

$$rdtr = \frac{h_{rdtr}}{h_{CRR}} \tag{9}$$

If there are n paths from the source node to the sink node, $rdtr_i$ is the reverse direction transmission ratio of each path. Then the number of average reverse direction transmission ratios of all paths, *avg_rdtr*, is expressed as (10).

$$avg\_rdtr = \frac{\sum_{i=1}^{n} rdtr_i}{n} \tag{10}$$

TABLE 3. Experiment results of *avg_rdtr* with different parameters.

	N_{sink}	(14,0)	(18,0)	(20,0)	(24,0)	(26,0)
	h_{sp}	5	6	7	8	9
$k_\lambda = 1$	λ_1	450	540	630	720	810
	avg_hop_1	9.66	12.26	13.82	16.42	17.94
	h_{rdtr1}	2.66	3.26	3.82	4.42	4.94
	avg_rdtr_1	0.27	0.26	0.27	0.27	0.27
$k_\lambda = 2$	λ_2	900	1080	1260	1440	1620
	avg_hop_2	14.64	18.48	20.86	24.63	26.95
	h_{rdtr2}	5.37	6.48	7.54	8.65	9.69
	avg_rdtr_2	0.37	0.35	0.36	0.35	0.36
$k_\lambda = 3$	λ_3	1350	1620	1890	2160	2430
	avg_hop_3	19.87	24.54	28.03	32.36	36.23
	h_{rdtr3}	7.87	9.54	11.03	12.69	14.25
	avg_rdtr_3	0.39	0.39	0.39	0.39	0.39

We conduct five experiments in the same network experiment. TABLE 3 is the experiment results of *avg_rdtr* with different parameters. The location of N_{sink} and h_{sp} are the same with the values in TABLE 2. First, the value of λ is proportional to the value of k_λ from Formula (6). Then, *avg_hop* increases when the value of λ raises. Last, there are more h_{rdtr} when λ is set lager. Correspondingly, the value of *avg_rdtr* adds. Figure 8 shows that there exists the situation as Definition 3, but the value of *avg_rdtr* is quite low and steady.

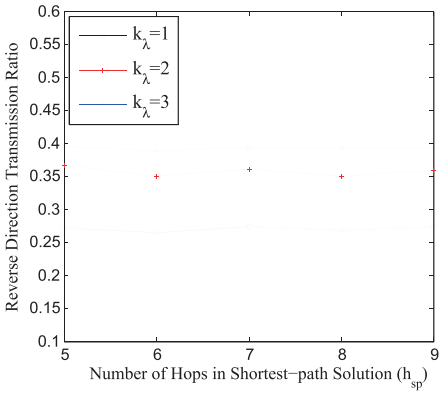


FIGURE 8. Reverse direction transmission ratio with different value of k_λ.

TABLE 4. Simulation environment setting.

Parameter	Value
Simulation area	$1500 \times 1500cm$
r_1	50m
r_2	250m
d	100m
Scenario_1	number of nodes: 100, 150, 200 interval of packet sending: 5s
Scenario_2	number of nodes: 100 interval of packet sending: 5s, 15s, 20s

D. SIMULATION ANALYSIS

We implemented CRR, ADRS [22] and the shortest-path solution in OPNET. For comparison, three mechanisms are run over the same topology. The sink node is set in the center of the area. Other sensor nodes are deployed randomly around the sink node. The simulation environment setting in CRR is presented in TABLE 4.

Suppose that each sensor node has sufficient energy without considering energy exhaustion. In this paper, the energy consumption of CRR is low. When a node is idle, the energy consumption is E_i and the idle time is t_i. When a node sends a message, the energy consumption is E_t and the transmission time is t_t. When a node receives a message, the energy consumption is E_a and the receiving time is t_a. Therefore, the energy consumption of each node is $E = E_i * t_i + E_t * t_t + E_a * t_a$. The total energy of all the nodes in the network is $E_{sum} = \sum E$. If n packets are received in the monitor area, then the average energy consumption of routing one packet is $E_{avg} = E_{sum}/n$. In the simulation, $E_i = 0.01J$, $E_t = 0.18J$ and $E_a = 0.15J$. As to end-to-end delay, we assume that m sources send packets to the sink node and the end-to-end delay are $t_{1\sim m}$, respectively. Then the average end-to-end delay in CRR is calculated as Formula (11).

$$t_{delay} = \frac{\sum_{i=1}^{m} t_i}{m} \tag{11}$$

Figure 9 shows the simulation performance of three mechanisms. Figures 9(a)(b) show average energy consumption of sending a packet and end-to-end delay in different network scales: The number of nodes are 100, 150, 200 and the interval

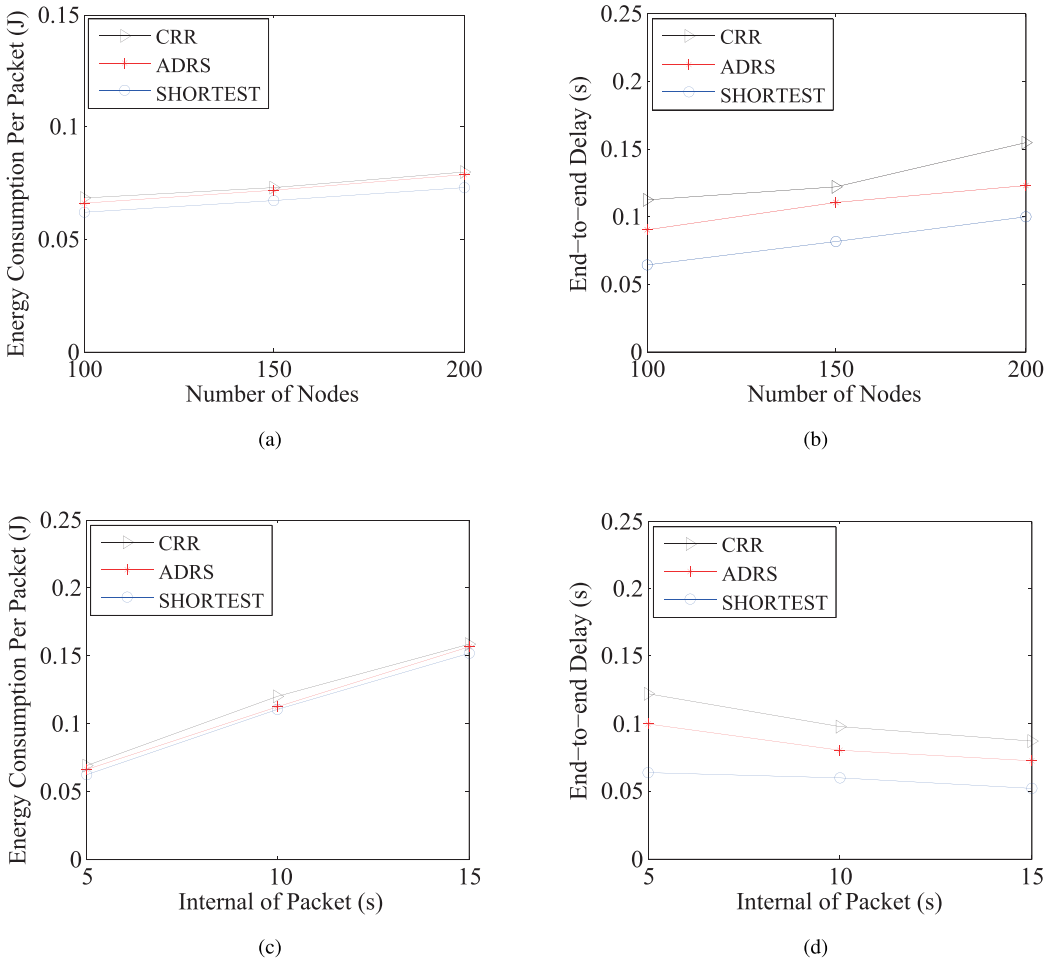


FIGURE 9. Simulation performance of CRR. (a) Energy consumption per packet in scenario 1. (b) End-to-end delay in scenario 1. (c) Energy consumption per packet in scenario 2. (d) End-to-end delay in scenario 2.

of packet sending is 5s. Figures 9(c)(d) show the energy consumption and end-to-end delay in different internals of packet sending: The number of nodes is 100 and intervals of packet sending are 5s, 15s and 20s. In terms of energy consumption, packets transmitted through the shortest-path solution consumes the least energy. Energy consumption in CRR is a little higher than the other two mechanisms but the difference is little. That is, the multi-path transmission formed by CRR will not consume too much energy and the paths are more random than ADRS. Also, the energy consumption increases with the increase of the node scale and the interval of packet sending. As to the end-to-end delay, CRR is longer than that of other two mechanisms, but it is under control. It increases when the number of sensor nodes gets larger. While it decreases when the interval of packet sending gets longer. Finally, no packet is lost in CRR in the simulation. That is, the delivery ratio is 100%. In terms of storagecost, each node has to store the locations of its neighbor nodes and the sink node. It is assumed that a node N_i has i neighbor nodes and each location information (longitude and latitude) occupies 4 bytes. Then, N_i has to store $4(i+1)$ bytes of location information. As to bandwidth cost, each package adds two bytes of storage to store the rest of sum of offset angles.

In CRR, the source node does not choose the shortest path to transmit packets. Therefore, it will inevitably produce redundant paths. If a packet needs h_{sp} hops along the shortest path from the source to the sink and it needs h_{CRR} hops in CRR, then the redundant path ratio (rpr) is as follows.

$$rpr = \frac{h_{CRR} - h_{sp}}{h_{CRR}} \tag{12}$$

The average redundant path ratio in CRR as shown in Fig.10(a) is the average rpr in different network scales where the number of sensor nodes are 100, 150, 200 and the interval of packet sending is 5s. Figure 10(b) is the average rpr under different intervals of packet sending that are 5s, 15s, 20s and the number of nodes is 100, as in the scale where the number of nodes is 150, the average rpr is 0.316 in CRR, and is 0.297 in ADRS. We can see that both CRR and ADRS will introduce some redundant paths. Although the average rpr of CRR is a little higher than that in ADRS, the transmission paths in CRR are more random than ADRS.

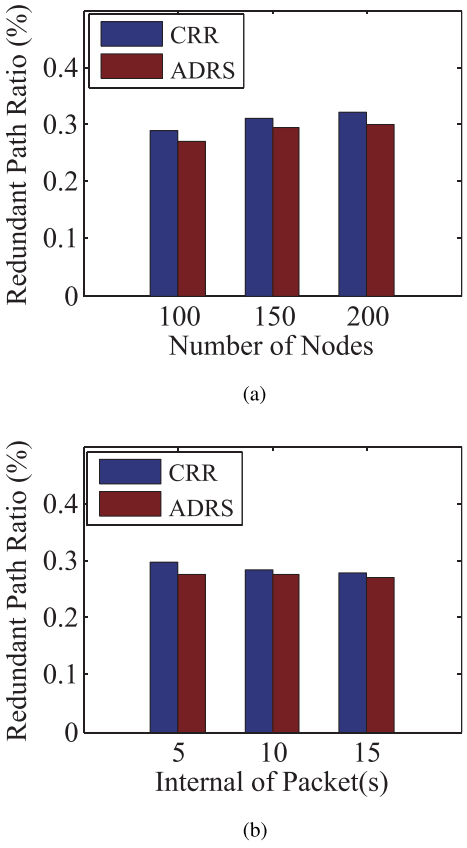


FIGURE 10. Average redundant path ratio. (a) *rpr* in scenario 1. (b) *rpr* in scenario 2.

VII. CONCLUSION

In order to prevent attackers from tracing back to the source nodes through analyzing the transmission paths, a source privacy protection mechanism based on offset angles and constrained probability is proposed in this paper. In CRR, sending nodes select their next relay nodes in a specific selection domain, an ellipse formed by multiple parameters such as the dangerous distance and the communication range. Then calculate the selected weights of the candidate nodes according to their offset angles, and decide which node will become the next hop based on the selected weights. Finally, system evaluation shows that CRR has great performance to protect source privacy though it introduces a small amount of redundant paths. Energy consumption and end-to-end delay are a little higher but under control and no packet is lost. In this paper, rectangular coordinate is used in the process of the next-hop selection and the construction of the whole transmission path. To minimize the lengths of transmission paths, CRR gives greater priority to the nodes with smaller offset angles to be selected as relay nodes. Meanwhile, the construction of the whole transmission path is under the constraint of the sum of offset angles. However, there is still room for us to optimize this mechanism. Our future work is to find more efficient ways to achieve the balance of the energy consumption control, security intensity and communication quality. We will try our best to decrease the energy consumption and end-to-end delay under the premise of ensuring high source privacy protection.

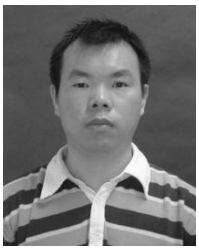
REFERENCES

[1] P. Hui, C. Hong, Z. XiaoYing, F. YongJian, L. CuiPing, and L. DeYing, "Location privacy preservation in wireless sensor networks," *J. Softw.*, vol. 26, no. 3, pp. 617–639, 2015.
[2] H. GuangJie, J. Jinfang, M. Guizani, and J. J. Rodrigues, "Green routing protocols for wireless multimedia sensor networks," *IEEE Wireless Commun. Mag.*, vol. 23, no. 6, pp. 140–146, Dec. 2016.
[3] G. Han, L. Liu, S. Chan, R. Yu, and Y. Yang, "HySense: A hybrid mobile crowdsensing framework for sensing opportunities compensation under dynamic coverage constraint," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 93–99, Mar. 2017.
[4] H. GuangJie, Y. Xuan, L. Li, Z. WenBo, and M. Guizani, "A disaster management-oriented path planning for mobile anchor node-based localization in wireless sensor networks," *IEEE Trans. Emerg. Topics Comput.*, to be published.
[5] O. Celal, Z. Y. Yong, and T. Wade, "Source-location privacy in energy-constrained sensor network routing," in *Proc. 2nd ACMWorkshop Secur. Ad Hoc Sensor Netw. (SASN)*, 2004, pp. 88–93.
[6] K. Pandurang, Z. Yanyong, T. Wade, and O. Celal, "Enhancing source-location privacy in sensor network routing," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, Jun. 2005, pp. 599–608.
[7] H. Wang, B. Sheng, and Q. Li, "Privacy-aware routing in sensor networks," *Comput. Netw.*, vol. 53, no. 9, pp. 1512–1529, 2009.
[8] E. C.-H. Ngai and R. Ioana, "On providing location privacy for mobile sinks in wireless sensor networks," *Wireless Netw.*, vol. 19, no. 1, pp. 115–130, 2013.
[9] R. A. Shaikh, H. Jameel, B. J. D'Auriol, H. Lee, S. Lee, and Y.-J. Song, "Achieving network level privacy in wireless sensor networks," *Sensors*, vol. 10, no. 3, pp. 1447–1472, 2010.
[10] C. Juan, F. Bin-Xing, Y. Li-Hua, and S. U. Shen, "A source-location privacy preservation protocol in wireless sensor networks using source-based restricted flooding," *Chin. J. Comput.*, vol. 33, no. 9, pp. 1736–1747, 2010.
[11] L. Kang, "Protecting location privacy in large-scale wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 603–608.
[12] P. Kanthakumar and L. Xiao, "Maintaining source privacy under eavesdropping and node compromise attacks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1656–1664.
[13] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks," in *Proc. 20th IEEE Int. Conf. Parallel Distrib. Process. Symp. (IPDPS)*, Apr. 2006, p. 8.
[14] L. Lightfoot, Y. Li, and J. Ren, "Preserving source-location privacy in wireless sensor network using STaR routing," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.
[15] Y. Li and J. Ren, "Mixing ring-based source-location privacy in wireless sensor networks," in *Proc. 18th IEEE Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2009, pp. 1–6.
[16] L. Yun, R. Jian, and W. Jie, "Quantitative measurement and design of source-location privacy schemes for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 7, pp. 1302–1311, Jul. 2012.
[17] J. Ren and D. Tang, "Combining source-location privacy and routing efficiency in wireless sensor networks," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–5.
[18] P. Spachos, L. Song, F. M. Bui, and D. Hatzinakos, "Improving source-location privacy through opportunistic routing in wireless sensor networks," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2011, pp. 815–820.
[19] L. Javier, R. Ruben, and C. Jorge, "Preserving receiver-location privacy in wireless sensor networks," in *Proc. Int. Conf. Inf. Secur. Pract. Exper.*, 2014, pp. 15–27.
[20] L. Zhou and Q. Wen, "Energy efficient source location privacy protecting scheme in wireless sensor networks using ant colony optimization," *Int. J. Distrib. Sensor Netw.*, vol. 2014, no. 1, pp. 1–14, 2014.
[21] W. Tan, K. Xu, and D. Wang, "An anti-tracking source-location privacy protection protocol in WSNs based on path extension," *IEEE Internet Things J.*, vol. 1, no. 5, pp. 461–471, Oct. 2014.

[22] L. Bai, L. Li, S. Qian, and S. Zhang, "Privacy protection algorithm based on expected phantom source node in wireless sensor network," in *Proc. IEEE Int. Conf. Softw. Eng. Service Sci.*, Aug. 2016, pp. 985–988.
[23] A. Nassiri, M. A. Razzaque, and A. H. Abdullah, "Isolated adversary zone for source location privacy in wireless sensor networks," in *Proc. IEEE Wireless Commun. Mobile Comput. Conf.*, Sep. 2016, pp. 108–113.
[24] L. Sangho, K. Jong, and K. Yoonho, "Preserving source- and sink-location privacy in sensor networks," *Comput. Sci. Inf. Syst.*, vol. 13, no. 1, pp. 115–130, 2016.
[25] P. Kumar, J. P. Singh, P. Vishnoi, and M. P. Singh, "Source location privacy using multiple-phantom nodes in WSN," in *Proc. IEEE Region 10th Conf. TENCON*, Nov. 2015, pp. 1–6.
[26] K. Shin, K. Kim, and S. Kim, "ADSR: Angle-based multi-hop routing strategy for mobile wireless sensor networks," in *Proc. IEEE Asia–Pacific Services Comput. Conf.*, Dec. 2011, pp. 373–376.
[27] P. Spachos, D. Toumpakaris, and D. Hatzinakos, "Angle-based dynamic routing scheme for source location privacy in wireless sensor networks," in *Proc. Veh. Technol. Conf.*, May 2015, pp. 1–5.
[28] Z. ZeMao, L. Yang, Z. Fan, Z. JianQin, and Z. Pin, "Research on source location privacy routing based on angle and probability in wireless sensor networks," *J. Shandong Univ.*, vol. 48, no. 9, pp. 1–9, 2013.
[29] W. Choi, S. K. Das, and K. Basu, "Angle-based dynamic path construction for route load balancing in wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 4. Mar. 2004, pp. 2474–2479.
[30] Z. YaNan, Y. Xong, and X. Wu, "Enhanced source-location privacy preservation protocol using random angle," *Computer Engineering and Applications*. 2016.

WENLONG CHEN received the Ph.D. degree in communication and information system from University Science and Technology, Beijing in 2011. He is currently an Associate Professor with the College of Information Engineering, Capital Normal University. His research interests include network protocol, Internet architecture, high performance router and wireless sensor networks.

MINGSHU ZHANG is currently pursuing the master's degree with the College of Information Engineer, Capital Normal University, Beijing. Her research interests include Internet of Things and wireless sensor networks.



GUANGWU HU received the Ph.D. degree in computer science and technology from Tsinghua University in 2014. He held a postdoctoral position with the Graduate School at Shenzhen, Tsinghua University. He is currently an Assistant Professor with the Shenzhen Institute of Information Technology. His research interests include software-defined networking, next-generation Internet, and Internet security.

XIAOLAN TANG received the Ph.D. degree with the School of Computer Science and Engineering, Beihang University, in 2014. She is currently a Lecturer with the College of Information Engineering, Capital Normal University. Her research interests include vehicular networks, wireless sensor networks, and microsoft research.

ARUN KUMAR SANGAIAH received the Ph.D. degree in computer science and engineering from VIT University, Vellore, India. He is currently an Associate Professor with the School of Computer Science and Engineering, VIT University, India. His area of interest includes software engineering, computational intelligence, wireless networks, bio-informatics, and embedded systems.