

A Probabilistic Source Location Privacy Protection Scheme in Wireless Sensor Networks

Hao Wang, Guangjie Han, *Senior Member, IEEE*, Wenbo Zhang, Mohsen Guizani, *Fellow, IEEE*, and Sammy Chan, *Member, IEEE*

Abstract—With the recent developments of Wireless Sensor Networks (WSNs), computing and communication have experienced huge advancement. Meanwhile, security has not received the same attention to go along with such developments. In this paper, we focus on the source location privacy problem in WSNs, a hot research topic in security, and propose a probabilistic source location privacy protection scheme (PSLP) for WSNs. A more powerful adversary, which can use Hidden Markov Model (HMM) to estimate the state of the source, is considered in this study. To cope with this type of adversary, phantom nodes and fake sources, which are responsible to mimic the behavior of the source, are utilized to diversify the routing path. Then, the weight of each node is calculated as a criteria to select the next-hop candidate. In addition, two transmission modes are designed to transmit real packets. The simulation results demonstrate that the proposed PSLP scheme improves the safety time without compromising the energy consumption.

Index Terms—Wireless sensor networks, source location privacy, phantom node, fake source.

I. INTRODUCTION

WIRELESS Sensor Networks (WSNs) consist of numerous sensor nodes and protocols, which is the basis of service like information authentication [1], event awareness [2], and node charging [3]. These nodes play the role of microcomputer and are distributed in various environments. There are a lot of data transmissions and communication behaviors between nodes. So, it is essential to preserve the security [4].

Security of WSNs involves many aspects, such as data privacy [5] and location privacy [6]. Data privacy can be protected by encryption algorithms while location privacy cannot be protected to the extreme. Due to the time correlation in data transmission between two nodes, the adversary can infer location information through analysis. From a time correlation perspective, location privacy consists of the source location privacy and the sink location privacy. Given the importance

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Hao Wang and Guangjie Han are with the Department of Internet of Things Engineering, Hohai University, Changzhou, 213022, China. (e-mail: wanghaohu@outlook.com, hanguangjie@gmail.com), Guangjie Han is the corresponding author.

Wenbo Zhang is with the School of Information Science and Engineering, Shenyang Ligong University, Shenyang, 110168, China. (e-mail: zhangwenbo@yeah.net.)

Moshen Guizani is with Department of Computer Science and Engineering, Qatar University, Qatar. (e-mail: mguizani@ieee.org.)

Sammy Chan is now City University of Hong Kong, Hong Kong, China. (e-mail: eeschan@cityu.edu.hk.)

of the source, in this paper, we focus on the source location privacy, which is an emerging research topic in the field of security. There are many techniques, like secure routing [7], fake sources [8], phantom nodes [9], fake cloud [10], and cluster [11], that can be applied to protect the source location privacy. We propose a probabilistic source location privacy protection scheme (PSLP), which adopts phantom nodes and fake sources for the reason that these two techniques can diversify the routing path. The steps of PSLP are as follows:

- 1) Phantom nodes are selected around the source and the visible area is taken into consideration.
- 2) A weight value, which is dynamically updated, is calculated in each node to determine the next-hop candidate.
- 3) Fake sources are generated around the sink to send fake packets, in order to confuse the adversary.

In the above steps, the visible area is a special area. When the adversary backtracks to this area, the source can be recognized immediately. Two types of packets exist in the transmission, which are the real packets and the fake packets. Real packets are generated by the source while fake packets are generated by fake sources. In order to hide the source location, real packets sent by the source are first transmitted to a phantom node through directed random walk. Here, considering the distance between the source and the sink, two transmission modes are taken into consideration and details will be given later. During the transmission of real packets, fake packets are also transmitted to the sink with a fixed period.

The proposed PSLP has exhibited a better performance than two other recent schemes in our simulations with regard to increasing the safety time while balancing the energy consumption. The main contributions of this paper are:

- 1) Both phantom nodes and fake sources are integrated into the proposed PSLP, which enhance the source location privacy.
- 2) A more powerful local adversary, which can use Hidden Markov Model to estimate the state of the source, is taken into consideration.
- 3) Two data transmission modes are designed based on the distance between the source and the sink, which further enhance the source location privacy.

The remainder of this paper is organized as follows. In Section II, we provide a review of studies focused on the source location privacy. In Section III, we introduce the network and the adversary models. In Section IV, we describe the proposed PSLP in detail. In Section V, we analyze the simulation results.

We conclude the paper and describe planned future studies in Section VI.

II. RELATED WORK

Many researchers have paid attention to the location privacy since Ozturk first proposed his concept [12]. Recently, location privacy has been widely researched in industrial wireless sensor networks [13], vehicular ad-hoc networks [14], cloud computing [15], social network [16] and so on.

Location privacy covers the source location privacy and the sink location privacy. In this paper, we focus on the source location privacy protection. Manjula *et al.* used virtual sources to protect the source location privacy [17]. In their scheme, a routing technique was proposed to maximize the safety time. By adding random walk into the routing process, nodes in non-hotspot areas participated in the establishment of multiple routing paths. Hence, the safety time increased without influencing the network lifetime.

Matthew *et al.* proposed two algorithms using fake sources to protect the source location privacy [8]. In the first algorithm, fake sources were dynamically deployed around the sink. Then, the sink used flooding to select fake sources. This algorithm can provide a good source location privacy at the expense of the huge energy consumption. To cope with this, another algorithm called dynamic single path routing algorithm (DynamicSPR) was proposed. By using directed random walk, nodes away from the source were selected as fake sources, which significantly reduced the energy consumption. However, fake sources were related to the relative location of the source and the sink, sensor nodes in a specific area might exhaust energy.

Jing *et al.* considered a more powerful adversary and proposed a privacy enhancing routing algorithm to protect location privacy [18]. In their research, a global adversary using Bayesian maximum-a-posteriori (MAP) estimation strategy tried to monitor the communication between nodes. Then, a decision-making framework was put forward to reduce the adversary's detection probability. Finally, the problem was converted into the adjustment of parameters.

Huang *et al.* focused on the energy utilization rate in WSNs while maintaining the source location privacy [19]. They proposed a redundancy branch-based source location privacy scheme. In their scheme, many redundancy branches were generated from the source to the sink. The number of branches was determined by the energy collected by nodes. In addition, these branches were converged into several routing paths later. However, the number of converged routing paths was not clearly defined and the energy collected by nodes around the sink might be less than the energy costed by transmitting packets.

Chen *et al.* in [20] proposed a constrained random walk mechanism. In their mechanism, a next-hop candidate selection domain was generated based on the offset angle of current node's neighbors and the danger distance, which made the selection domain look like an ellipse. Then, the weight of each node in the domain was calculated by the ratio between a current node's offset angle and the sum of total offset angle.

The smaller the ratio, the higher the probability that this node became the next-hop candidate. However, the offset angle of a node was fixed, and thereby the weight might not change. Thus, nodes which acted as the next-hop candidate might consume too much energy.

Chen *et al.* utilized phantom nodes and proposed a limited flooding algorithm to protect the source location privacy [9]. The limited flooding was performed by the source to get the information of nodes in the limited flooding area. Then, nodes on the edge of the limited flooding area were chosen as phantom nodes to simulate the function of the source. If a phantom node stayed behind the source, packets sent by this phantom node first bypassed the visible area and were then transmitted to the sink using the shortest path. However, the limited flooding was repeatedly performed, which might not be suitable for a large scale network.

Li *et al.* in [21] proposed a scheme using random intermediate nodes and ring to protect the source location privacy. First, the authors introduced the criteria to quantitatively measure the source location information leakage. Then, to reduce the leakage probability, random intermediate nodes were added to make the routing path disperse. Packets were first transmitted to an intermediate node and then forwarded to a node in ring around the sink. Packets were routed on the ring for a random hop and then sent to the sink.

Mutalemwa *et al.* divided the whole network into regions and proposed a scheme based on region transmission [22]. In this scheme, the sink was located in the center of the network and regions were generated around the sink. The transmission between regions was implemented by a set of relay nodes which were selected strategically. These strategic relay nodes took up two regions and were responsible for forwarding packets to the sink. However, the distribution of these nodes were close to the sink. Relaying too many packets would consume a lot of energy. Thereby, the average energy efficiency was not high.

Wang *et al.* considered the source location privacy against a new type of adversary in [23]. The adversary model had two properties, global and local. Under normal circumstances, the adversary was a local adversary. When a potential area where the source may stay was located, the adversary became a global adversary in this area. To cope with it, a message mapping sharing method was presented and a cloud containing many dummy packets was created around the source to hide the location. Each message copy was transmitted by random routing, which provided sufficient source location privacy.

Considering the time correlation during the transmission between sensor nodes, Mayank *et al.* used the data mule to protect the source location privacy [24]. Data mule worked as the mobile data collection unit and collected data when the source was in its communication radius. In this condition, the source location privacy was changed into the protection of the mule's moving track. Then, the authors proposed three extended versions of angle-based scheme to protect the source location. However, since the mule moved grid by grid, the protection of the mule was not given enough attention. There was still a lot of research space in reducing the time correlation.

To further reduce the time correlation during the transmission, Proaño *et al.* proposed a traffic decorrelation technique to reduce the threat of a global adversary [25]. The proposed traffic normalization scheme reduced the communication overhead and the transmission delay. In addition, the whole network was partitioned into a set of minimum connected areas with a circular queue, which could reduce the active nodes during transmission and the adversary's eavesdrop probability. The privacy in their work was quantified to the distance between location estimated by the adversary and the location of the source.

From the above works, it can be seen that source location privacy protection has experienced a great improvement, techniques like fake sources, phantom nodes, random walk, and the weight are developed. However, these techniques are only used in a simple way, which gives us an inspiration.

III. SYSTEM MODEL AND ASSUMPTIONS

In this section, the system model contains the network model and the adversary model, and assumptions are interspersed in both two parts. The background application is the protection of wild rare animals. In the wild environment, sensor nodes are randomly deployed. After being deployed, the locations of these sensor nodes keep unchanged. Then, sensor nodes monitor the acts of animals.

A. Network model

The network model in this study is based on the typical Panda-Hunter model [12]. As shown in Fig. 1, a WSN which is composed of many sensor nodes is deployed to monitor the activities of pandas. Once a sensor node detects a panda, it becomes the source and sends packets to the sink through multiple hops. The essence of privacy protection is reducing the probability that the adversary finds the source location. Therefore, we make the following assumptions:

- 1) Sensor nodes are randomly deployed. After being deployed, the location of each sensor node remains unchanged. What's more, all sensor nodes are homogeneous, which means that they have the same initial energy, the same computing ability, and the same cache memory.
- 2) Routing is based on the weight. Each sensor node is assigned a weight that is updated regularly. The weight here represents the probability that this node is selected as the next hop, or it can be understood as the preference in selecting the next hop node, which is related to the residual energy, the communication quality, and the hop count to the sink. Details of this weight will be given later.
- 3) Only one sink exists in the network. As in other schemes or protocols [12], [15], [22], the sink remains in the network center.
- 4) Each sensor node has knowledge of its own adjacent neighbors. Packets sent by each sensor node are encrypted with an encryption algorithm. However, this part is beyond the scope of this study.

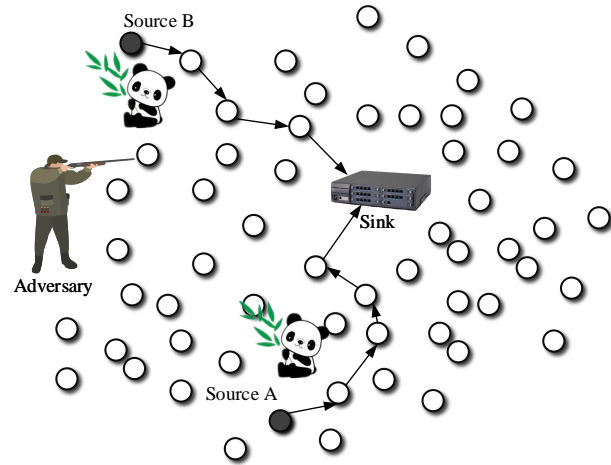


Fig. 1: The Panda-Hunter model.

B. Adversary model

Due to the potential value of the source, the adversary starts from the sink and tries his/her best to find the source location. The monitoring range of the adversary equals to a sensor node's radius, which means that the type of the adversary is a local adversary. The local adversary has a limited monitoring range, which is equal to or a little larger than the communication range of a common node. Thus, the local adversary can only monitor parts of the network. Commonly, the adversary performs passive attacks, such as eavesdropping and backtracking, to avoid being discovered by the network administrator.

We consider a more powerful adversary in this paper. Apart from the passive attack, we assume that the adversary knows the packet type by checking the header of each packet. Then, the adversary can use the Hidden Markov Model (HMM) to infer the possible state of the source for a given time based on its observation. The intention of using HMM to infer the possible state of the source is that, comparing with wandering in the network, it is more effective for the adversary to find the source location from the estimation result of HMM. This is because the estimation of HMM can help the adversary reduce the scope of finding the source.

In HMM, there are three critical elements, which are the initial probability π , the state transition matrix A , $A = P(q_j|q_i) 1 \leq i, j \leq N$, and the confusion matrix B , $B = P(o_i|q_j) 1 \leq i \leq M, 1 \leq j \leq N$ (here we assume M equals to N). In this network, we define three states of the source and thereby π consists of three states, which are the resting state probability P_r , the foraging state probability P_f , and the disporting state probability P_d . After getting the transition probability of each state, the adversary can use the Viterbi algorithm to calculate the state of the source $(i_1^*, \dots, i_{T-1}^*, i_T^*)$ based on the observation $O = (o_1, o_2, \dots, o_T)$ and the model $\lambda = (A, B, \pi)$.

Viterbi algorithm is a solution method of solving HMM. In Viterbi algorithm, the first step is the initialization:

$$\delta_1(i) = \pi_i b_i(o_i), \quad i = 1, 2, \dots, N \quad (1)$$

$$\psi_1(i) = 0, \quad i = 1, 2, \dots, N \quad (2)$$

Then, the adversary does the recursion operation on t from 2 to N :

$$\delta_t(i) = \max [\delta_{t-1}(j)a_{ji}|b_i(o_t)], \quad i = 1, 2, \dots, N; 1 \leq j \leq N \quad (3)$$

$$\psi_t(i) = \arg \max [\delta_{t-1}(j)a_{ji}], \quad i = 1, 2, \dots, N; 1 \leq j \leq N \quad (4)$$

Finally, the recursion stops at:

$$P^* = \max \delta_T(i) \quad (5)$$

$$i_T^* = \arg \max [\delta_T(i)] \quad (6)$$

The source state estimated by the adversary is:

$$i_T^* = \psi_{t+1}(i_{t+1}^*) \quad (7)$$

Performing above steps, an adversary can acquire the estimation, which is a most possible state vector for a given observation state. Then the adversary can reduce the searching scope of the source.

However, in this way the adversary only knows the source's state, not the source's location. What we consider here is that if the adversary has enough knowledge about the network, he has a higher probability to find the source from the estimated source state. In our proposed PSLP, the key idea is to make real packets and fake packets to be transmitted from different directions with different states, which attracts the adversary's attention and reduces the accuracy of the estimate.

IV. THE PROBABILISTIC SOURCE LOCATION PRIVACY PROTECTION SCHEME

In this section, a detailed description of PSLP is given. In the initialization process, the beacon message is periodically broadcasted by the sink to sensor nodes. When a node receives the message, it records the hop count stored in it, increases the hop count by one, repackages the packet, and sends to its neighbors. Each node only records the minimum hop count. Subsequently, all nodes know their hop count to the sink and their neighbors. Since the adversary may know the state of the source at a given time while the location of the source is still unknown, we intend to increase the possible locations of the source. PSLP contains three steps: the first step is the determination of phantom nodes; the second step is the determination of fake sources; the third step is the routing from the source to the sink. An overview of PSLP is shown in Fig. 2.

As mentioned in the adversary model, the adversary can use HMM to estimate the state of the source and then perform targeted search. What we need to do is to increase more possible states of the source. Phantom nodes and fake sources perfectly match our needs. Although the function of the phantom node and the fake source is similar, but the definition

of the two is different. The phantom node refers to nodes around or nearby the source, which simulate the function of the source. The fake source also refers to nodes which simulate the function of the source. But the location of fake source is around the sink, which is far from the source. The motivation of combining the phantom node and the fake source together is to create the diversification of the transmission directions. Both phantom nodes and fake sources are selected in non-hotspot area, which has little influence on the network lifetime.

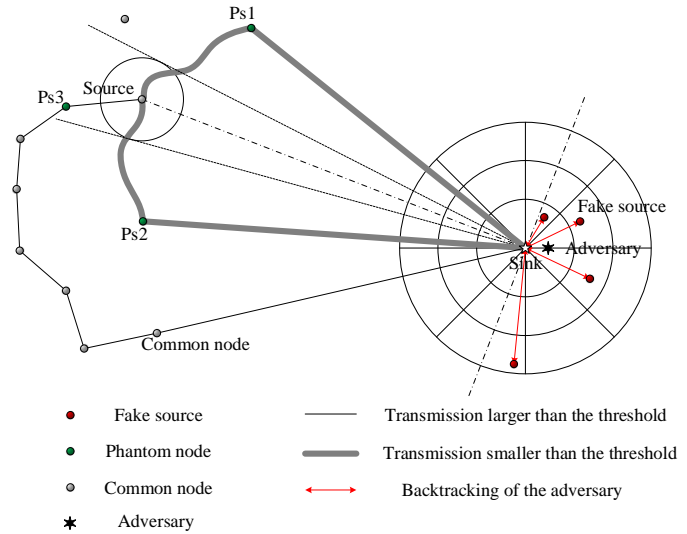


Fig. 2: Overview of PSLP.

A. The determination of phantom nodes

As mentioned before, phantom nodes are nodes deployed around the source to simulate the function of the source. Considering the function of phantom nodes, we can see that the longer the distance between a phantom node and the source, the stronger the privacy protection is. The main purpose of this setup is to direct the adversary away from the real source. However, authors in [17] and [21] have proved the probability that a phantom node stays within 20% of H hops from the source is $1 - e^{-H/25}$. Thus, we decide to use directed random walk to select phantom nodes. In directed random walk, packets are transmitted in a fixed direction. Hence, when directed random walk stops, the selected phantom node stays away from the source.

For more details, when the source appears, it sends packets to one of its neighbors within H hops via directed random walk. Then, the neighbor sends packets to a node in its far neighbor list and decreases H by one. When H becomes zero, the current node changes into a phantom node and forwards packets sent by the source. The phantom node changes during each data transmission. In addition, the phantom node must stay outside the visible area (circle around the source). Because when the adversary backtracks to the visible area, it recognizes the source immediately. Moreover, the source sends packets to the phantom node once during the initialization. So, the transmission between the source and the phantom node is assumed to be safe. Noted that the determination of phantom nodes relates to the distance between the source and the sink, which will be presented later.

B. The determination of fake sources

As described in previous definition, fake sources are generated around the sink to increase directions from where packets come. The deployment range of a fake source is specified by angle θ_2 in Fig. 3. First of all, the sink divides the network into several rings. Then, these rings are divided into n sectors. For the sake of separating fake sources and the source, fake sources are only selected in the right part of the line which is perpendicular to the line linking the source and the sink. The number of fake sources is determined by the actual application. At the initialization, the fake source sequence is generated.

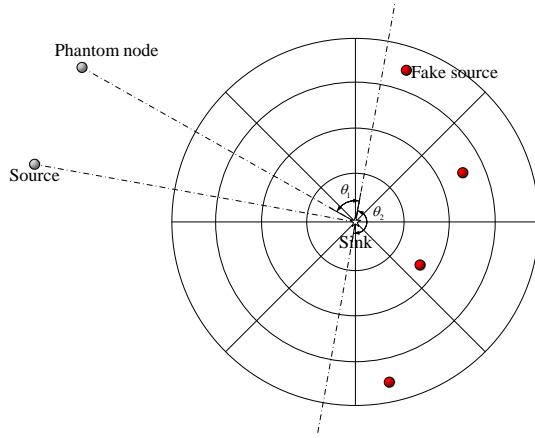


Fig. 3: Ring areas around the sink.

Each fake source is preferably to stay in different sectors, which guarantees that the direction of each fake packet is different. Since the adversary knows the source state in a specific time, it needs to analyze the packet flow to find the source. Therefore, by adopting fake sources to diversify the source location, source location privacy is protected. A node acts as a fake source for a fixed period. When the time period exhausts, another fake source appears. In order to alleviate the energy consumption of fake sources, we assume that there only exists one fake source for a certain period of time.

There are many methods to transmit fake packets to the sink, such as the shortest path routing. However, in order to defend against the adversary's direction attack, the transmission should be given enough attention. A possible transmission of fake packets is presented in Fig. 4. Several rings are generated according to the fake source sequence. Then, these rings are split into arcs in eight directions. Fake packets are routed to the sink through arcs, just like the orange line in Fig. 4. In this orange line, the transmission direction of fake packets changes each time. With the help of these arcs, the routing path of fake packets varies, and hence the direction attack is resisted and the adversary takes more time to find fake sources. In addition, fake packets will not be cached in the sink memory. On the contrary, fake packets will be discarded after a while.

C. The routing from the source to the sink

After the determination of phantom nodes and fake sources, the next step is the transmission between the real source and the sink. The source transmits a message to inform the

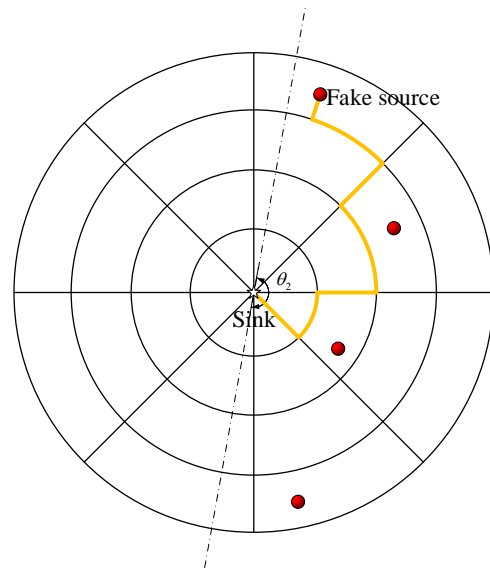


Fig. 4: Possible transmission of fake packets.

sink when it appears. Then, the sink selects a fake source immediately after receiving this message. Considering that the source randomly appears, there exists a possibility that distance between the source and the sink is small. So, in response to this situation, we set a threshold between the source and the sink. Thereby, the routing process from the source to the sink contains two scenarios. The first case is that the hop count between the source and the sink is larger than the threshold. The second case is that the hop count between the source and the sink is smaller than the threshold. In general, as the source first sends packets to a phantom node, the main differences lie in the selection of phantom nodes and the transmission from the phantom node to the sink.

1) *The hop count is larger than the threshold:* In this case, the source first routes packets to a phantom node using directed random walk. Here the selection of phantom node has no restriction due to the hop count between the source and the sink is large enough. After that, the phantom node selects its next-hop candidate according to the weight of each node. This weight is related to the residual energy, the communication quality, and the hop count to the sink. Therefore, the weight is calculated by:

$$w_i = \alpha * \frac{E_{res_i}}{E_0} + \beta * Q_i + (1 - \alpha - \beta) * \frac{H_i}{avg(\sum H_{neighbor})} \quad (8)$$

where E_{res_i} is the residual energy of current node, E_0 is the initial energy of a node, Q_i is the communication quality per hop, which refers to the success rate of the communication transmission, H_i is the number of hop counts from current node to the sink, and $avg(\sum H_{neighbor})$ is the average hop count of neighbors of current node to the sink. Then, the phantom node selects a neighbor with the maximum weight as its next-hop candidate. After a transmission, each node dynamically adjusts its weight and, thereby, the next round of transmission path will change, which balances the energy consumption of nodes.

2) *The hop count is smaller than the threshold:* This case is a little bit complicated. Because the hop count is small,

the potential candidate nodes that can be chosen as phantom nodes are limited. The source first searches nodes whose hop counts to the sink are larger than that of the source itself. These nodes should also stay outside the visible area. If no node satisfies the condition, the source selects another angular direction to select the phantom node. Then, the source selects a phantom node from the searched nodes using directed random walk. After a phantom node is determined, this phantom node transmits packets through the equal hop count routing. The equal hop count routing is constructed by nodes whose hop counts to the sink are the same as the selected phantom node. The equal hop count routing ends when packets are routed to a node whose x-coordinate is near the sink. As shown in Fig. 3, the equal hop count routing lies in angle θ_1 . By building this route, nodes within the visible area can be utilized to transmit packets. This is because the existence of the equal hop count routing makes the transmitted packets bypass the visible area.

When packets are received by the last relay node of the equal hop count routing, this node selects a next-hop candidate in its close neighbor list on the basis of the weight. Since nodes around the sink may consume more energy than other nodes, weight of nodes can significantly balance the energy consumption of these nodes. Since the fake source also transmits fake packets to the sink during the transmission of real packets, there is a possibility that a node relays both real packets and fake packets. However, we are not worried about this situation. This is because transmitting two packets on the same node is a small probability event. Since that the weight is dynamically adjusted, it is unlikely to happen. Even though it happens, it is more likely that the adversary finally finds the fake source instead of the source.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of PSLP. All the results provided in this section are the average values of the experimental data.

A. Overview

In this section, four metrics are evaluated in the simulation, namely, the safety time, the energy consumption, the network lifetime, and the transmission delay. First of all, we give the definition of each metric. The safety time is the difference between the time when the source sends the first packet and when the adversary finds the source's location. To be more specific, we use the hop count of backtracking taken by the adversary to represent the safety time. The energy consumption represents the average energy costed per simulation run. As control packets only take up very little energy, so we ignore this part and mainly focus on the energy consumption during packets transmission. The network lifetime refers to the time difference between the network establishment and the death of the first node. The transmission delay means the average packet transmission and the data processing time per simulation run.

PSLP is compared with two other schemes, which are the dynamic single path routing algorithm (DynamicSPR) [8] and

the enhanced protocol for source location protection (SLP-E) [9]. DynamicSPR uses fake sources to protect the source location, while the SLP-E adopts phantom nodes to implement this. These two methods are integrated in PSLP. Therefore, we choose DynamicSPR and SLP-E for the comparison.

B. Simulation environment and parameter settings

The simulation experiment is performed on MATLAB R2017b and Table 1 is the parameters used in the simulations. Since the network size of DynamicSPR and SLP-E is not the same, we unify the network scale in the simulation.

TABLE I: Parameter settings

Parameter	Symbol	Value
Side length	L	300-900 m
Node density	ρ	0.003 $number/m^2$
Communication radius	R	45 m
Size of a packet	l	1000 $bits$
Threshold	T	5 $hops$
Initial energy	E_0	0.5 J
Distance threshold	d_0	87 m
System parameter in the first mode	ε_{fs}	10 $pJ/bit/m^2$
System parameter in the second mode	ε_{amp}	0.0013 $pJ/bit/m^4$
Transmitting circuit loss	E_{elec}	50 nJ/bit
Data rate	v	1 $kbps$

C. Safety time

The longer the safety time, the safer the network is. As two cases and fake sources are taken into consideration (as mentioned in Section IV), the safety time is different on both sides of the threshold T between the source and the sink, which looks like a split point. As shown in Fig. 5, when the hop count between the source and the sink is five (which is the pre-set threshold), there is an obvious decline. This is because the transmission of packets changes when the hop count is larger than five. When the hop count is larger than five, the next-hop candidate is selected by the weight of nodes, so the safety time is stabilized due to the fact that packets are routed towards the sink. However, when the hop count is less than five, the safety time fluctuates sharply. This is because the selection angle of phantom nodes changes and the hop count of directed random walk varies, which makes the safety time fluctuate. In addition, the communication radius in Fig. 5 is fixed.

D. Energy consumption

The energy consumption has an impact on the network lifetime, the less energy the node costs, the larger the network lifetime is. However, given that increasing the safety time adds extra energy consumption, a balance should be kept between the safety time and the energy consumption. In this study, the energy consumption is calculated by:

$$\begin{cases} E_t = lE_{elec} + l\varepsilon_{fs}d^2 & d \leq d_0 \\ E_t = lE_{elec} + l\varepsilon_{amp}d^4 & d \geq d_0 \end{cases} \quad (9)$$

$$E_r = lE_{elec} \quad (10)$$

As shown in Fig. 6, the energy consumption fluctuates with the hop count between the source and the sink in a fixed

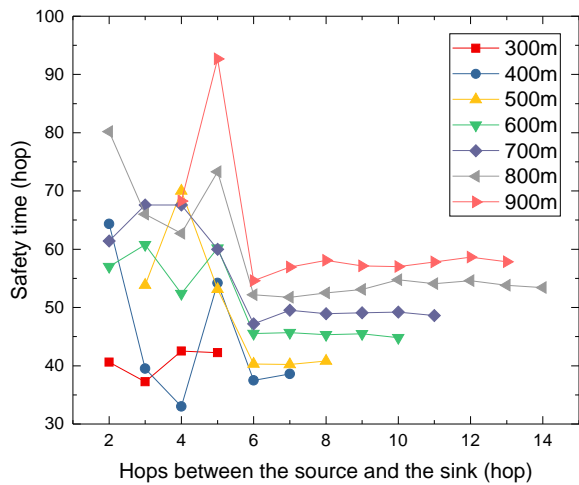


Fig. 5: Safety time versus various hops between the source and the sink.

communication radius. When the hop count is larger than five, nodes on the routing path are selected by the weight and, therefore, the energy consumption is stabilized. When the hop count is less than five, the randomness of the selection of phantom nodes makes the energy consumption fluctuate. In addition, the weight is considered in the selection of the next-hop candidates, and as a result each node's next-hop candidate may change. So the energy consumption is balanced in each node.

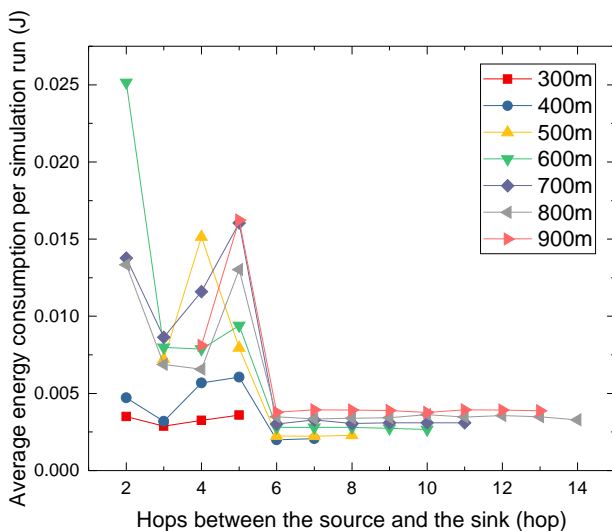


Fig. 6: Average energy consumption versus various hops between the source and the sink.

The 3-D residual energy distribution of two transmission modes is presented in Fig. 7. Figure 7 shows the energy consumption per simulation run. The X and Y axes are the network side length in meter, while the Z axis is the residual energy of each node per transmission in Joule. The white area indicates that the node has consumed energy, which are responsible for the transmission from the source to the sink. Considering there are two transmission modes in PSLP. When the hop count between the source and the sink is smaller than the threshold T , routing path is constructed based on the weight of nodes and the residual energy consumption of

each node is shown in Fig. 7(a). When the hop count between the source and the sink is larger than the threshold T , the transmission of packets is changed into the equal hop count routing. Hence, the white area in Fig. 7(b) is larger than that in Fig. 7(a). Even though there are some peaks in both Fig. 7(a) and Fig. 7(b), the energy cost per simulation run is not large and the magnitude is 10^{-3} , which is acceptable.

E. Network lifetime

The network lifetime is influenced by many factors and the energy consumption of nodes occupies a large proportion. As there are two transmission modes in PSLP and the threshold T plays a vital role in this scheme, we intend to explore the relationship between the network lifetime and the distance between the source and the sink. The result is presented in Fig. 8. We can observe that with the increase of the hop count from the source to the sink, the average network lifetime increases with side length ranging from 300 m to 900 m for a fixed communication radius. Due to the existence of the threshold T , the network life on both sides of the threshold T is completely different. When distance between the source and the sink is smaller than the threshold T , the transmission in this condition consumes more energy, and hence the network lifetime is not long. In addition, the network lifetime decreases with the increase of the network's side length. When the side length increases, a packet is delivered using more hops to the sink, which in turn influences the network lifetime.

F. Influence of the number of phantom nodes and fake sources

The influence of the number of phantom nodes and fake sources on the safety time and the average energy consumption is shown in Fig. 9. As we can see in Fig. 9(a), the number of phantom nodes has a little influence on the safety time. This is because only one phantom node works per data transmission. Hence, the difference of the safety time in each transmission is not obvious, which is only related to the relative position between the phantom node and the source. There is a decrease when the number of phantom nodes is 4. This is due to the randomness of location between the source and the sink, and data transmission mode changes during each simulation run. The safety time increases with the number of fake sources. This is because fake source works in a relay. Thus, the more fake sources, the longer the safety time.

In Fig. 9(b), the number of phantom nodes has a little influence on the average energy consumption except the 4 phantom nodes case. This is because only one phantom node works in each simulation run. When there are 4 phantom nodes, the distance between the source and the sink tends to be larger than the threshold T . Thus, the length of routing path is not long and the average energy consumption decreases. The average energy consumption increases with the number of fake sources. The reason for the 4 fake sources case is the same as that of the 4 phantom nodes case.

G. Comparison

The comparison of the transmission delay is shown in Fig. 10 and the communication radius of each node is fixed in this

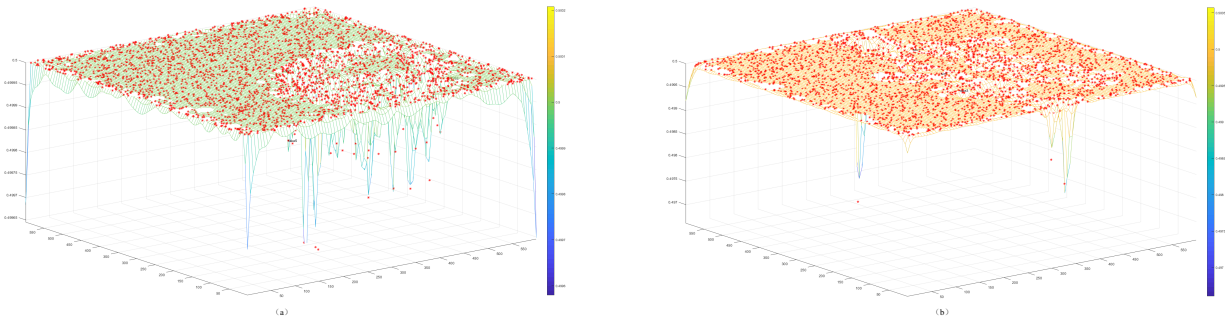


Fig. 7: The 3-D residual energy distribution of PSLP, and the x, y, and z axes are network side length, the network side length, and the residual energy. (a) The hop count between the source and the sink is larger than the threshold T ; (b) The hop count between the source and the sink is smaller than the threshold T .

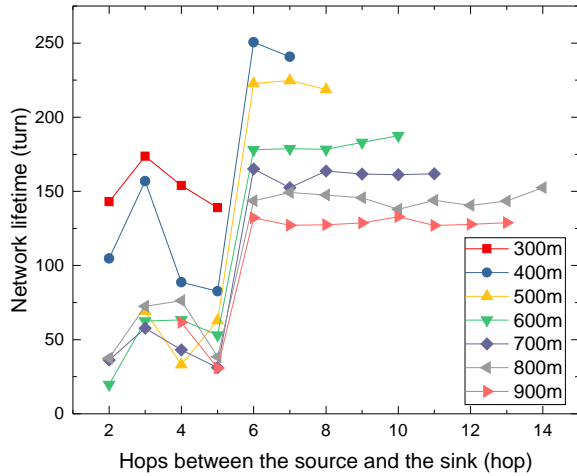


Fig. 8: Network lifetime.

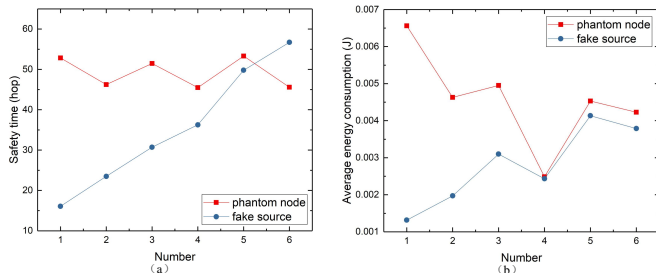


Fig. 9: (a) Influence of the safety time; (b) Influence of the average energy consumption.

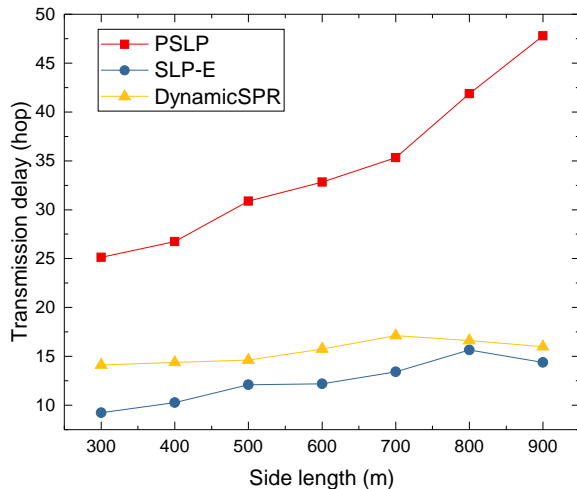


Fig. 10: Transmission delay.

figure. We use hop as the unit of delay. The reason is that the transmission delay is generated mainly due to the data transmission time and the data processing time, and the data transmission time is related to the length of routing path and the data rate. The data rate is a fixed value and thus we use the unit of routing path as the unit of transmission delay. In SLP-E and DynamicSPR, real packets are transmitted using the shortest path, and thereby the transmission delay in these two schemes is small. Although the side length of the network is growing, the impact is not very obvious. For a further elaboration, with the increase of network scale, the limited flooding area in SLP-E also increases. So the transmission delay in SLP-E gradually increases. Since two transmission modes exist in PSLP and the hop count between the source and the sink varies in each simulation run, this causes a little fluctuation in PSLP. In addition, with the increase of the side length, the routing path also increases, so does the transmission delay.

The comparison of the safety time and the energy consumption are presented in Fig. 11. The safety time relates to the routing path. We mainly compare these three algorithms from two aspects: the communication radius and the network size. The Safety time relates to the routing path. The longer the routing path, the more time an adversary will spend on tracking. In Fig. 11(a), the safety time of PSLP fluctuates a lot. However, the safety time in SLP-E and DynamicSPR increases at a lower rate with a given side length. This is because packets in SLP-E and DynamicSPR are transmitted using the shortest path. When a node's communication radius increases, the average length of routing path reduces. Thereby, the safety time decreases with the communication radius. On the other hand, two transmission modes are considered in PSLP, with the increase of the communication radius, the case in which the hop count is smaller than the threshold occurs more often. So it is the threshold that makes the safety time of PSLP fluctuate with the communication radius.

In Fig. 11(b), the safety time increases with the side length. As the side length increases, the average length of routing path increases. Thus, the safety time increases correspondingly. Since the equal hop count routing is adopted in PSLP, the safety time of PSLP is the largest. Also, when the network scale increases, it is more probable that the distance between the source and the sink is larger than the threshold. Thus, the

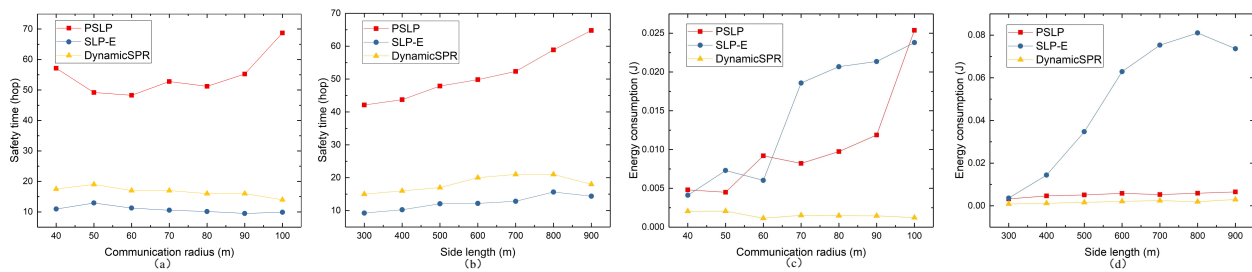


Fig. 11: (a) Comparison of the safety time with various communication radii; (b) Comparison of the safety time with various side lengths; (c) Comparison of the average energy consumption with various communication radii; (d) Comparison of the average energy consumption with various side lengths.

safety time in this condition grows steadily. Many fake sources are deployed around the sink in DynamicSPR, so the average safety time is larger than that of SLP-E.

The energy consumption also relates to the routing path. In Fig. 11(c), the average energy consumption cost per simulation run in DynamicSPR is smooth. Because only one fake source lasts for a fixed period and packets are transmitted using the shortest path. The average energy consumption of PSLP and SLP-E increases with the communication radius for a given side length. Energy in SLP-E is mainly consumed in the limited flooding area. With the randomness of random walk, the energy consumption fluctuates. Again, it is the two types of packets transmission that makes the energy consumption fluctuate. In addition, with the increase of the communication radius while the network scale keeps unchanged, the hop count between the source and the sink in PSLP may become small. Thereby it consumes more energy in this condition.

In Fig. 11(d), the communication radius is fixed. With the increase of the network side length, the number of random walk hop counts increases with the side length. Therefore, the energy consumption in a limited flooding area increases sharply, and this is the main cause of the high energy consumption. The difference of the energy consumption between PSLP and DynamicSPR lies in the use of phantom nodes and the equal hop count routing. As a result, the energy consumption in PSLP is a little larger than that in DynamicSPR. In addition, parameters of the communication radius in PSLP is important. It is worth to mention that even though the energy consumption in Fig. 11(d) is very smooth, if we change the communication radius into a different value, the result varies. In Fig. 11(d), it seems that the communication radius used in the simulation makes the transmission mode more biased towards the case in which the hop count between the source and the sink is larger than the threshold.

The comparison of the network lifetime is presented in Fig. 12. Among the three schemes, the network lifetime of DynamicSPR is the largest. This is because the fake source only appears in the furthest distance to the sink and only one fake source exists in the network for a fixed time. Moreover, real packets are routed to the sink using the shortest path, which consumes less energy and reduces the transmission delay. However, the distance between the sink and the fake source is different in each simulation, which makes the network lifetime fluctuate. In SLP-E, the limited flooding is repeated during each simulation run and, therefore, most of energy is consumed in the limited flooding step, which adds an

extra burden on nodes in limited flooding area. In PSLP, even though there are two transmission modes in the simulation, the network lifetime of PSLP declines slowly and steadily. Given that we adopt the average value as final results, the difference between the two transmission modes is counteracted. Again, due to the use of fake packets, the network lifetime of PSLP is larger than that of DynamicSPR. However, it is relatively small compared with SLP-E.

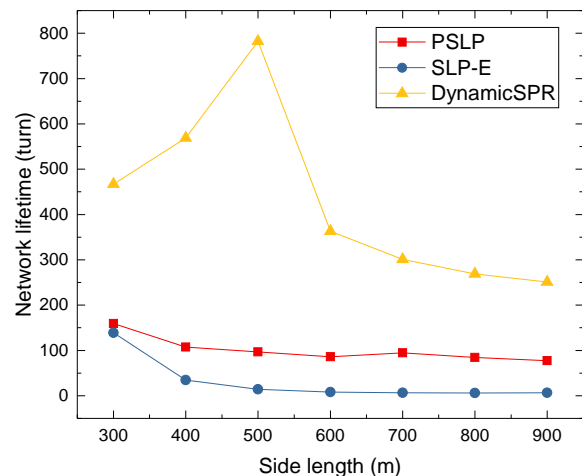


Fig. 12: Network lifetime of three schemes.

VI. CONCLUSIONS AND FUTURE STUDIES

Studying security in WSNs became increasingly important during the last decade. In this paper, we focused on the source location privacy, a research hotspot in security, and proposed a probabilistic source location privacy protection scheme (PSLP) based on WSNs. A powerful adversary which utilizes Hidden Markov Model (HMM) is considered in this study. To cope with it, phantom nodes, fake sources, and weight are adopted to change the packets' transmission directions. Considering the distance between the source and the sink, two types of routing modes are designed. Compared with DynamicSPR and SLP-E, the simulation results demonstrate that the proposed PSLP achieves a high safety time and balances the energy consumption of each node. Future studies will concentrate on protecting the source location by reducing the adversary's monitoring probability and secure communication among nodes.

ACKNOWLEDGMENT

The work is supported by the National Key Research and Development Program, No.YS2017YFGH001945 and the Na-

tional Natural Science Foundation of China-Guangdong Joint Fund under Grant No.U1801264 and supported by Six talent peaks project in Jiangsu Province, No.XYDXXJS-007.

REFERENCES

- [1] H. Lu and J. Li, "Privacy-preserving authentication schemes for vehicular ad hoc networks: a survey," *Wireless Communications and Mobile Computing*, vol. 16, no. 6, pp. 643-655, Apr. 2016.
- [2] G. Han, X. Yang, L. Liu, S. Chan, and W. Zhang, "A Coverage-Aware Hierarchical Charging Algorithm in Wireless Rechargeable Sensor Networks," *IEEE Network Magazine*, pp. 1-7, Nov. 2018, DOI: 10.1109/MNET.2018.1800197
- [3] G. Han, H. Guan, J. Wu, S. Chan, L. Shu, and W. Zhang, "An Uneven Cluster-Based Mobile Charging Algorithm for Wireless Rechargeable Sensor Networks," *IEEE Systems Journal*, pp. 1-12, Nov. 2018, DOI: 10.1109/JSYST.2018.2879084
- [4] G. Han, H. Wang, J. Jiang, W. Zhang, and S. Chan, "CASLP: A Confused Arc-Based Source Location Privacy Protection Scheme in WSNs for IoT," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 42-47, Sept. 2018.
- [5] H. Lu, J. Li, and M. Guizani, "Secure and Efficient Data Transmission for Cluster-based Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 750-761, Mar. 2014.
- [6] G. Han, L. Zhou, H. Wang, W. Zhang, and S. Chan, "A Source Location Protection Protocol Based on Dynamic Routing in WSNs for Social Internet of Things," *Future Generation Computer Systems*, vol. 82, no. 5, pp. 689-697, Aug. 2018.
- [7] H. Lu, J. Li, and H. Kameda, "A Secure Routing Protocol for Cluster-based Wireless Sensor Networks Using ID-based Digital Signature," *Proceedings of IEEE Global Communications Conference*, Dec. 2010.
- [8] M. Bradbury, A. Jhumka, and M. Leeke, "Hybrid online protocols for source location privacy in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 115, pp. 67-81, May 2018.
- [9] J. Chen, Z. Lin, Y. Hu, and B. Wang, "Hiding the Source Based on Limited Flooding for Sensor Networks," *Sensors*, vol. 15, no. 11, pp. 29129-29148, Nov. 2015.
- [10] G. Han, X. Miao, H. Wang, M. Guizani, and W. Zhang, "CPSLP: A Cloud-Based Scheme for Protecting Source-Location Privacy in Wireless Sensor Networks Using Multi-Sinks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2739-2750, Jan. 2019.
- [11] G. Han, H. Wang, M. Guizani, S. Chan, and W. Zhang, "KCLP: A k-means Cluster-Based Location Privacy Protection Scheme in WSNs for IoT," *IEEE Wireless Communications Magazine*, vol. 25, no. 6, pp. 84-90, Dec. 2018.
- [12] C. Ozturk, Y. Zhang, and W. Trappe, "Source-Location privacy in energy-constrained sensor network routing," *ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 88-93, Jan. 2004.
- [13] J. Wang, R. Zhu, S. Liu, and Z. Cai, "Node Location Privacy Protection Based on Differentially Private Grids in Industrial Wireless Sensor Networks," *Sensors*, vol. 18, no. 2, pp. 410-425, Jan. 2018.
- [14] A. Boualouache, S. Senouci, and S. Moussaoui, "A Survey on Pseudonym Changing Strategies for Vehicular Ad-Hoc Networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 770-790, First quarter. 2018.
- [15] Y. Gong, C. Zhang, Y. Fang, and J. Sun, "Protecting Location Privacy for Task Allocation in Ad Hoc Mobile Cloud Computing," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 1, pp. 110-121, Mar. 2018.
- [16] J. Du, C. Jiang, K. Chen, Y. Ren, and H.V. Poor, "Community-Structured Evolutionary Game for Privacy Protection in Social Networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 574-589, Mar. 2018.
- [17] R. Manjula and D. Raja, "A novel source location privacy preservation technique to achieve enhanced privacy and network lifetime in WSNs," *Pervasive and Mobile Computing*, vol. 44, pp. 58-73, Feb. 2018.
- [18] J. Koh, D. Leong, G. Peters, I. Nevat, and W. Wong, "Optimal Privacy-Preserving Probabilistic Routing for Wireless Network," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2105-2114, Sept. 2017.
- [19] C. Huang, M. Ma, Y. Liu, and A. Liu, "Preserving Source Location Privacy for Energy Harvesting WSNs," *Sensors*, vol. 17, no. 4, pp. 724-755, Mar. 2017.
- [20] W. Chen, M. Zhang, G. Hu, X. Tang, and A. Sangaiah, "Constrained Random Routing Mechanism for Source Privacy Protection in WSNs," *IEEE Access*, vol. 5, pp. 23171-23181, Sept. 2017.
- [21] Y. Li, J. Ren, and J. Wu, "Quantitative measurement and design of source-location privacy schemes for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 7, pp. 1302-1311, July 2012.
- [22] L. Mutalemwa and S. Shin, "Strategic Location-Based Random Routing for Source Location Privacy in Wireless Sensor Networks," *Sensors*, vol. 18, no. 7, July 2018, DOI:10.3390/s18072291.
- [23] N. Wang, J. Fu, J. Zeng, and B. Bhargava, "Source-location privacy full protection in wireless sensor networks," *Information Sciences*, vol. 444, pp. 105-121, May 2018.
- [24] R. Mayank, N. Li, D. Liu, W. Matthew, and K. Sajal Das, "Using data mules to preserve source location privacy in Wireless Sensor Networks," *Pervasive and Mobile Computing*, vol. 11, pp. 244-260, Apr. 2014.
- [25] A. Proaño, L. Lazos, and M. Krunz, "Traffic Decorrelation Techniques for Countering a Global Eavesdropper in WSNs," *IEEE Transactions on Mobile Computing*, vol. 16, no. 3, pp. 857-871, Mar. 2017.



Hao Wang is currently pursuing his Ph.D. degree from the Department of Computer Science and Technology at Hohai University. He received his B.S. degree in Electronic Information Science and Technology from Nanjing Agricultural University, China, in 2015. His current research interests include location privacy protection of wireless sensor networks.



Guangjie Han [S'03, M'05, SM'18] is currently a Professor with the Department of Information and Communication System, Hohai University, Changzhou, China and a Distinguished Professor of Dalian University of Technology, Dalian, China. He received the Ph.D. degree from Northeastern University, Shenyang, China, in 2004. From 2004 to 2005, he was a Product Manager for the ZTE Company. From 2005 to 2006, he was a Key Account Manager for the Huawei 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 October 2011, he was a Visiting Research Scholar with Osaka University, Suita, Japan. From January 2017 to February 2017, he was a Visiting Professor with City University of Hong Kong, China. He is the author of over 330 papers published in related international conference proceedings and journals, including the IEEE COMST, IEEE TII, IEEE TMC, IEEE TVT, IEEE TIE, IEEE TPDS, IEEE TETC, IEEE IoT Journal, IEEE TETCI, IEEE TCC, IEEE Systems, IEEE Sensors, IEEE Wireless Communications, IEEE Communications, IEEE Network, etc, and is the holder of 125 patents. Currently, his H-index is 32 and i10-index is 88 in Google Citation (Google Scholar). Total citation of his papers by other people is more than 4426 times. His current research interests include Internet of Things, Industrial Internet, Machine Learning and Artificial Intelligence, Mobile Computing, Security and Privacy. 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 16 international journals, including the IEEE JSAC, IEEE Network, IEEE Systems, IEEE ACCESS, IEEE/CCA JAS, Telecommunication Systems, etc. He has guest edited a number of special issues in IEEE Journals and Magazines, including the IEEE Communications, IEEE Wireless Communications, IEEE Transactions on Industrial Informatics, Computer Networks, etc. He has served as a Reviewer of more than 60 journals. He had been awarded the ComManTel 2014, ComComAP 2014, Chinacom 2014 and Qshine 2016 Best Paper Awards. He is a Senior Member of IEEE.



Wenbo Zhang is currently a professor of School of Information Science & Engineering, Shenyang Ligong University, China. He received his Ph.D. in Computer science at Northeastern University, China, in March 2006. He had been awarded the ICINIS 2011 Best Paper Awards, the National Science and Technology Progress Award and Youth Science and Technology Awards from China Ordnance Society. His current research interests are Ad hoc networks, Sensor Networks, Satellite networks, Embedded system.



Mohsen Guizani [S'85-M'89-SM'99-F'09] received the B.S. (with distinction) and M.S. degrees in electrical engineering, the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor at the CSE Department in Qatar University, Qatar. Previously, he served in different academic and administrative positions at the University of Idaho, Western Michigan University, University of West Florida, University of Missouri-Kansas City, University of

Colorado-Boulder, and Syracuse University. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is currently the Editor-in-Chief of the IEEE Network Magazine, serves on the editorial boards of several international technical journals and the Founder and Editor-in-Chief of Wireless Communications and Mobile Computing journal (Wiley). He is the author of nine books and more than 500 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He also served as a member, Chair, and General Chair of a number of international conferences. Throughout his career, he received three teaching awards and four research awards. He also received the 2017 IEEE Communications Society WTC Recognition Award as well as the 2018 AdHoc Technical Committee Recognition Award for his contribution to outstanding research in wireless communications and Ad-Hoc Sensor networks. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer. He is a Fellow of IEEE and a Senior Member of ACM.



Sammy Chan [S'87-M'89] received the BE and MEngSc degrees in electrical engineering from the University of Melbourne, Australia, in 1988 and 1990, respectively, and the Ph.D. degree in communication engineering from the Royal Melbourne Institute of Technology, Australia, in 1995. From 1989 to 1994, he was a senior research engineer and project leader in Telecom Australia Research Laboratories. He is currently an associate professor in the Department of Electronic Engineering, City University of Hong Kong.