A Source-location Privacy Protection Strategy via Pseudo Normal Distribution-based Phantom Routing in WSNs

Jun Huang
SCIE
Chongqing University of Posts and Telecommunications
Chongqing, China 400065
jhuang@cqupt.edu.cn

Meisong Sun
CS Department
Chongqing University of Posts and Telecommunications
Chongqing, China 400065
mssun1990@126.com

Shitong Zhu, Yi Sun
SCIE
Chongqing University of Posts and Telecommunications
Chongqing, China 400065
stzhu.ac@gmail.com
sunycqupt@gmail.com

Cong-cong Xing
Math. and CS Department
Nicholls State University
Thibodaux, LA 70123, USA
cong-cong.xing@nicholls.edu

Qiang Duan
IST Department
The Pennsylvania State University
Abington, LA 70123, USA
qduan@psu.edu

ABSTRACT

Toward resolving the source-location privacy protection issue in Wireless Sensor Networks (WSNs), a Pseudo Normal Distribution-based Phantom Routing (PNDBPR) protocol is proposed in this paper. The proposed protocol is composed of two critical phases: 1) adjusting the value of minimum-hops between a phantom node and its source, and varying the area of the phantom node distribution region at the network deployment stage; 2) generating a set of Gaussian-distributed random real numbers using the Pseudo Random Generator when the source node is transmitting data packets, and calculating the corresponding random walk hops. Theoretical analyses and simulation results show that PNDBPR can dramatically improve the diversity and the dynamicity of the phantom nodes distribution at the expense of a slight increase in communication overheads when compared with the existing PUSBRF (Source Location Privacy Preservation Protocol in Wireless Sensor Network Using Source Based Restricted Flooding) and HBDRW (A hop-based directed random walk) protocols.

Keywords

Wireless Sensor Network (WSN); source-location privacy; pseudo normal distribution

1. INTRODUCTION

With the technological advancement of microelectronics, Wireless Sensor Networks (WSNs), as the foundational and underlying technology of the Internet of Things (IoT), has become one of the norms in today's technological world [1]. WSNs are widely used in tasks where wired networks are not suitable, for example, remote target tracking, vast environment surveillance, and military information probing [2]. Unfortunately, WSNs are highly vulnerable to attacks due to both the environment in which WSNs are deployed and the nature by which WSNs work. Generally speaking, WSN-related security issues can be characterized into two types [3] data-oriented privacy threats and context-oriented privacy threats (the source-location privacy protection problem discussed in this paper is of the second type). Data-oriented privacy threats refer to the situation when adversaries attempt to acquire the contents of data packets and to obtain further information such as the location and identity of WSN nodes, which can be (well) handled by conventional security techniques such as data packet encryptions [4] and authentications. Context-oriented privacy threats, however, turn out to be a much more challenging issue. This can be seen from the following two aspects. On one hand, since WSN communications are wirelessly broadcasted, adversaries can easily seize important pieces of WSN information such as the emitting time and location of data packets without seeing the actual contents of data packets by using certain kind of data traffic analysis techniques. On the other hand, due to the fact that sensor nodes are typically composed of inexpensive and energy-efficient devices with constrained computing and storage capacities, and usually operate against adversities without battery-replacement support and human interventions, traditional security mechanisms which demand more computing resources (e.g. public key encryption) are no longer reasonable for WSNs. Consequently, acceptable privacy protection techniques for WSNs must be light-weight and resource-aware.

In order to deal with the issues raised in the context-oriented threats, we propose a pseudo normal distribution-based phantom routing (PNDBPR) protocol to protect source locations in WSNs. The major contributions made in this paper are summarized as follows.

SAC'15 April 13-17, 2015, Salamanca, Spain.

http://dx.doi.org/10.1145/2695664.2695843

- Unlike prior protocols, PNDBPR calculates the number of hops from a phantom node to the source based on a randomly-generated and Gaussian-distributed number, which not only renders a uniform distribution of phantom nodes around the source node, but also enhances the diversity and dynamicity of the distribution. Moreover, the stochastically constructed forwarding routing mechanism in PNDBPR reduces the chance of the same path from phantom nodes to the sink node being repeated.
- We theoretically analyze the performance of PNDBPR in terms of protection effectiveness and communication overhead. Our analysis demonstrates that PNDBPR dramatically improves the diversity and the dynamicity of the phantom nodes distribution with a slight compromise in communication overheads when compared with the existing typical peer protocols PUSBRF (Source Location Privacy Preservation Protocol in Wireless Sensor Network Using Source Based Restricted Flooding) and HBDRW (A hop-based directed random walk) protocols.
- We conduct thorough comparisons among PNDBPR, PUSBRF, and HBDRW via extensive simulations. The simulation results show that our proposed work significantly improves the privacy protection of source locations with the trade-off of a slightly higher communication overhead, which further confirms the theoretic analysis.

The rest of the paper is organized as follows: Section 2 reviews existing approaches for providing location privacy in sensor networks. Section 3 presents our network and adversary models. In Section 4, we propose PNDBPR to address source-location privacy issue. Section 5 analyzes the performance of PNDBPR in terms of protection effectiveness and communication overhead theoretically. Section 6 evaluates the proposed techniques via simulation study. Finally, Section 7 concludes this paper.

2. RELATED WORK

The issue of source-location privacy (SLP) in WSNs has been extensively studied in recent years. This issue was first investigated by Oztuk, Zhang, and Trappe [5] in 2004, and was followed by a Panda-Hunter game modeling for it by Kamat et al. [3] in 2005. Chen et al. [6] noticed that phantom nodes and their paths generated by existing SLP protection protocols tend to inhabit only some specific areas, whereby proposed a Source-based Restricted Flooding SLP protection protocol which not only enables the phantom nodes to be as far away as possible from the source but also creates a variety in their paths. Based on [7], a quantitative measurement framework for SLP protections was suggested in [8] and was used to show vulnerabilities of the existing SLP protection mechanisms. Note, however, that due to the full randomness in selecting intermediate nodes in [7] and [8], two such selected intermediate nodes by two adjacent data packets may be too close to be acceptable. Also, the risk of overlapping data transmission paths is increased as a result of this full randomness, causing excessive communication overheads. In [9], Zhou, Wen, and Zhang proposed a Confused Area Scheme (CAS) whereby a designated area is set up in which forwarding routes are purposely changed to confuse the attackers when they attempt to trace the data transmission paths. While CAS is effective in reducing the chances of data packets being traced, it is evident that there is no obvious way to determine the number of needed such areas in a network and the number of needed sensor nodes in each of the areas. Path Extension Method (PEM) was introduced by Tan, Xu, and Wang in [10], which dynamically creates multiple fake source nodes and multiple fake data communication paths as well to confused the adversaries. Although the PEM strategy works effectively, even in the situation where the source is considerably close to the sink, the creation of multiple fake communication paths is the clear communication overhead. All work in [11, 12, 13, 14] centers around the idea of using fabricated data packets to conceal the real ones. Instead of mixing fabricated data packets themselves with the authentic data packets, the transmissions of these two types of data are mixed each node in the network needs to transmit some data packets either at equally-spaced time points or at exponentially-distributed time points. At any such time point, authentic data packets will be transmitted if they are present; otherwise, fabricated data packets will be transmitted. Since all data packets are transmitted in the same fashion, eavesdroppers will have no way of discerning whether a data packet is real or fake, whereby a statistically strong protection of SLP can be achieved. However, mechanism suffers from a serious communication overhead especially when there are not many authentic data packets to be transmitted. In that case, a large amount of network resources will be used to create and transmit fake data packets, causing excessive network burdens and shortening network's lifetime.

Unlike the aforementioned SLP protection techniques, our work (PNDBPR) enables directed random walks of the source by using pseudo Gaussian-distributed processes and also adopts a probabilistic forwarding strategy. This protocol can improve both the diversity of the phantom nodes and the randomness of the forwarding paths. Due to the large-scale nature of WSNs, it would be highly unlikely for an adversary to be able to eavesdrop the traffic of the entire network. We thus in our work focus on the hop-by-hop adversary model only.

3. SYSTEM MODEL

The Panda-Hunter game model presented in [3] is employed in this paper. We assume that a large number of sensor nodes are randomly deployed in a natural resources preservation area to monitor pandas, forming an ad-hoc WSN. The first node that sees pandas will become the source node, and the source node will regularly transmit data packets which contain the information about objects being watched to the sink node.

3.1 Network Model

Throughout the paper, the following assumptions are stipulated about the network:

- There is only one sink in the network; the location information of the sink is public and is available to every node in the network.
- Every node is aware of its neighbors' location information and its own relative location information in the network. Relative location information is broadcasted

to the entire network. For each node, its communication area is ranged within the circle with center n and radius r; that is, each node is allowed to communicate only with its neighborhood nodes. In addition, all nodes have the same communication radius r.

- All data packets being transmitted are encrypted. (The data encryption itself is not a topic discussed in this paper.)

3.2 Adversary Model

For simplicity, we assume that there is only one adversary in the network. This attacker attempts to find the location of the source through traffic analysis and backtracking. Generally, by the way they move and attack, adversaries can be split into two groups: 1) patient adversaries and 2) prudent adversaries. Based on the study in [3], we choose the patient adversary model in our work since the longer an attacker waits, the more information revealing the source may be obtained.

We assume that an adversary has the following features:

- **Fully equipped:** The adversary is device- and resource-rich, equipped with advanced devices such as antenna and spectrum analyzers. These devices allow an adversary to easily detect the signal strength of received data packets, determine the sender of the data packets, and decide whether to take actions or not. We also assume that an adversary will not miss any packets within its detection range.

- **Non-malicious:** The adversary only monitors the network traffic and attempts to locate the source. It is not interested in malicious actions such as altering the contents of data packets, changing routing information, and damaging any nodes, since none of these actions will be of any help in expediting the finding of the source location for the adversary. Rather, they might risk the adversary to be detected by the intrusion defense mechanism installed in the network.

- **Initially close to sink & knowledgeable:** The adversary initiates its attempt nearby the sink and can listen to all data traffic going into the sink. Only if a message originated from the source is detected, the adversary will start backtracking, hop-by-hop, to the message originator. By the Kerckhoffs' principle, we assume that the adversary is aware of everything about the protection of the panda.

- **Range-constrained:** The listening range and the viewing area of the adversary are both equal to sensors' transmission radius.

4. PROPOSED PNDBPR PROTOCOL

We now present the PNDBPR protocol proposed in this paper. All parameters used in his paper and their meanings are given in Table 1.

4.1 Network Initialization

The purpose of network initialization is to configure the state of node neighborhood and to set up the minimum number of hops from any node to the sink. Each node in the network is assigned a unique identification label and the sink broadcasts a global message $Sink\_Inf$ to the entire network by flooding. The information contained in $Sink\_Inf$ includes $\{ID, Coor\_Pos, Sink\_Hop\}$ where ID is the identification label of the node, $Coor\_Pos$ is the coordinates of the node, and $Sink\_Hop$ is the minimum number of hops from the node to the sink. The initial value of $Sink\_Hop$ is set to be zero. When a node receives the $Sink\_Inf$ for the first time, it will increment $Sink\_Hop$ by 1, record the sender's ID, $Coor\_Pos$ and $Sink\_Hop$, then forward this message to its neighbors. Otherwise, the $Sink\_Inf$ will be just discarded. This procedure will be repeated until $Sink\_Inf$ is received by all nodes. After the initialization, each node in the network will be aware of its neighbors and know its own $Sink\_Hop$. For any node u, its neighborhood nodes can be split into two sets: $u.parent$ and $u.child$ where the $Sink\_Hop$ value of any node in $u.parent$ is less than that of u, i.e., $H_{(v,B)} < H_{(u,B)}$, $v \in u.parent$; and the $Sink\_Hop$ value of any node in $u.child$ is greater or equal to that of u, i.e., $H_{(v,B)} \geq H_{(u,B)}$, $v \in u.child$.

Table 1: Parameters List

Parameter	Note
$Sink\_Hop$	Minimum number of hops between the current node and the sink
$Source\_Hop$	Minimum number of hops between the current node and the source
d_{min}	Minimum number of hops between phantom nodes and the source
d_{ps}	Random number of hops between phantom nodes and the source
$drand$	Average of d_{ps} (with equal probabilities)
$d''rand$	Average of d_{ps} (with unequal probabilities)
n_p	Phantom node that has d_{ps} hops from the source
$H_{(v,S)}$	Hop count from source S to node v
$H_{(v,B)}$	Hop count from sink B to node v
H	Euclidean distance from source to sink
a	Angle intersecting by line from source to a phentom node and line from source to sink

4.2 Phantom Nodes Generation

A regular sensor node becomes the source node upon detection of any target object. A source node broadcasts the message $Source\_Inf$ in the same fashion as the broadcasting of $Sink\_Inf$ in the network initialization phase known as the source-based restricted flooding [6]. A $Source\_Inf$ message is similar to $Sink\_Inf$, and is composed of $\{ID, Coor\_Pos, Source\_Hop\}$, where ID and $Coor\_Pos$ denote the same thing as before and $Source\_Hop$ denotes the minimum number of hops between the node and the source. Again, in a similar way, each node has two separated neighborhood node sets $u.parent$ and $u.child$, where the value of $Source\_Hop$ of any node in $u.parent$ is less than or equal to that of u, i.e., $H_{(v,S)} \leq H_{(u,S)}$, $v \in u.parent$; and the value of $Source\_Hop$ of any node in $u.child$ is greater than that of u, i.e., $H_{(v,S)} > H_{(u,S)}$, $v \in u.child$.

In a WSN where nodes are uniformly distributed, the number of hops between two nodes can be measured by the Euclidean distance between these two nodes [6]. Let $\{x_0, y_0\}$

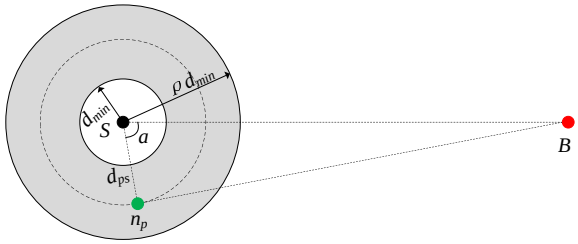

Figure 1: An illustrative example of phantom nodes generation.

be the location of the source and d_{ps} be the number of hops between a phantom node and the source. Every time the source produces a packet, the pseudo random number generator generates a random number x where the set X follows Normal Distribution ($X \sim N(0, \sigma)$), and x is used to calculate d_{ps}:

$$d_{ps} = d_{min}(|x| + 1). \tag{1}$$

If we let $\{x_1, y_1\}$ to be the location of a phantom node, then

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} = d_{ps} \geq d_{min}. \tag{2}$$

Thus d_{ps} produced by formula (1) is guaranteed to be greater than or equal to d_{min}, which puts phantom nodes at locations far away from the source.

Moreover, it can be seen from equation (1) that $d_{min} \leq d_{ps} \leq \rho d_{min}$ holds with probability:

$$2\varphi_{0,\sigma^2}(\rho - 1) - 1 = 2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\rho-1)^2}{2\sigma^2}} - 1 = 2\varphi(\frac{\rho - 1}{\sigma}) - 1, \tag{3}$$

where $\rho > 1$ and $\varphi_{(0,\sigma^2)}$ are the normal probability density functions. For example, when σ= 1, the probability of $d_{ps} \in [d_{min}, 2d_{min}]$ is $2\phi(\frac{1}{1}) - 1 = 0.6827$; the probability of $d_{ps} \in [d_{min}, 3d_{min}]$ is $2\phi(\frac{2}{1}) - 1 = 0.9545$.

The attribute $Source\_Hop$ is assigned to be d_{ps} when the source starts transmitting packets. If $d_{ps} > \rho d_{min}$, then we set $d_{ps} = \rho d_{min}$. d_{ps} will be decremented by 1 ($d_{ps} = d_{ps} - 1$) whenever the packets arrive at a forwarding node u, and the next forwarding node will be selected randomly from $u.child$. This message-forwarding process continues until d_{ps}=0, and at that time the current node will become a phantom node. The mechanism that each next-forwarding node is always selected from $u.child$ guarantees that all hops are directed toward the opposite direction of the source, and the randomness of this selection creates a more dynamic and diverse phantom path, resulting in an improved SLP protection.

d_{min} and ρd_{min} and are respectively the minimum and maximum allowed values for d_{ps}. By properly setting up these two parameters, the distribution of phantom nodes can be well controlled in such a way that their distances from the source may remain sufficiently large.

This is conceptually illustrated in Figure 1, where S and B are the source and sink respectively, the gray area is the phantom region, and n_p is the phantom node which is d_{ps} hops away from S. Clearly, a larger $[d_{min}, \rho d_{min}]$ would give rise to a wider phantom node distribution area with ensuing richer path diversity and less path overlapping.

Algorithm 1 PNDBPR

1: **if** (Node u receives the message for the first time) **then**
2: Record the information included in the message and broadcast it;
3: **else**
4: Record the information included in the message then drop it.
5: **end if**
6: Every node u in the network divides its neighbours into two sets: $u.parent$ and $u.child$;
7: **if** ($\rho d_{min} > 0$) **then**
8: **if** (Node v receives the message for the first time) **then**
9: Record the ρd_{min} in the message;
10: Update the ρd_{min} in the message: $\rho d_{min} = \rho d_{min} - 1$
11: Broadcast the modified message;
12: **else**
13: Drop message;
14: **end if**
15: **else**
16: Stop flooding;
17: **end if**
18: Every node v in the random walk area divides its neighbors into two sets: $v.parent$ and $v.child$;
19: $RandomNum = d_{ps}$
20: **if** ($d_{ps} > \rho d_{min}$) **then**
21: $d_{ps} = \rho d_{min}$
22: **else**
23: Select a node from as the next node and transmit the message;
24: $d_{ps} = d_{ps} - 1$
25: **end if**
26: **while** ($d_{ps} = 0$) **do**
27: Send the message to Sink along the probabilistic forwarding path;
28: **end while**

4.3 Probabilistic Forwarding Routing

In order to transmit packets to the sink as quickly as possible and minimize the path overlaps, we add a new attribute $Flag$ to nodes. A node set its $Flag$ to be after transmitting some packets. When a node u is ready to select its next-hop forwarding destination, it will check the $Flag$ of all of its neighborhood nodes v. If $v.Flag = False$, then the packets will be forwarded to v and $v.Flag$ will be flipped to $True$; otherwise, $v.Flag$ will be flipped to $False$ and u will randomly choose another node to which to forward the packet.

4.4 PNDBPR SLP Protection Algorithm

The Pseudo-code of the PNDBPR algorithm is shown as Algorithm 1, which consists of three phases: $SinkFlood$ (from Line 1 to 6), $SourceFlood$ (from Line 7 to 18), and $RandomWalk$ (from Line 19 to 28). If we assume that the number of participating nodes in $SinkFlood$ phase, $SourceFlood$ phase, and probabilistic forwarding routing phase are n, n_1, and n_2 respectively, and that there are m neighbours surrounding each node in average, then the time complexity of the algorithm can be written as: $T_{PNDBPR} = \max\{T_1(n*m) + T_2(n_1*m) + T_3(n_2)\}$. Since m is a constant, $T_{PNDBPR} = O(n)$.

5. PERFORMANCE ANALYSIS

Protection effectiveness and communication overhead are the most important two indicators in evaluating the performance of a SLP protection protocol. In this section, we accordingly compare our proposed PNDBPR strategy with the existing a hop-based directed random walk (HBDRW) and Source Location Privacy Preservation Protocol in WSN Using Source-Based Restricted Flooding (PUSBRF) protocols in regard to these two aspects.

5.1 Protection Effectiveness

Phantom routing is one of the common techniques used in WSN SLP protections. As the topological component of phantom routing, the distribution of phantom nodes determines the effectiveness of the phantom routing technique. Therefore we will assess our proposed SLP protection strategy, in comparison with HBDRW and PUSBRF, using the number of hops d_{ps} from phantom nodes to the source.

In large-scale and uniformly distributed WSNs, directed random walk routing protocols that select a node's next-hop destination by the value of its neighbors' hops from the sink (or the source), such as HBDRW (or PUSBRF), all generate phantom nodes with h random walk hops from the source. However, in PNDBPR, the distance $d_{ps} \in [d_{min}, \rho d_{min}]$ (measured by hops) from phantom nodes to the source is a random value that varies in accordance with another (Gaussian distributed) random value x. When d_{ps} is randomly chosen from $[d_{min}, \rho d_{min}]$ with equal chances and with a sufficiently large number of repetitions, we can calculate the average of d_{ps} as follows:

$$\begin{aligned} drand =& \frac{d_{min} + (d_{min}+1) + \ldots + \rho d_{min}}{\rho d_{min} - d_{min} + 1} \\ =& \frac{\rho d_{min} + d_{min}}{2}. \end{aligned} \tag{4}$$

Note that the probability for d_{ps} to be in the interval $[d_{min}, \rho d_{min}]$ is $2\varphi(\frac{\rho-1}{\sigma}) - 1$ due to the fact that $x \sim N(0, \sigma)$, and that values out of the bounds will be regarded as ρd_{min}, so the actual average of d_{ps} should be recalculated as :

$$\begin{aligned} d''rand =& drand \times [2\varphi(\frac{\rho-1}{\sigma}) - 1] \\ &+ \rho d_{min} \times [2 - 2\varphi(\frac{\rho-1}{\sigma})]. \end{aligned} \tag{5}$$

When $\sigma = 1$, the probability that d_{ps} falls within interval $[d_{min}, 2d_{min}]$ is equal to $2\varphi(\frac{1}{1}) - 1 = 0.6827$. Accordingly,

$$\begin{aligned} d''rand =& \frac{d_{min} + (d_{min}+1) + \ldots + 2d_{min}}{2d_{min} - d_{min} + 1} \times 0.6827 \\ &+ 2d_{min} \times (1 - 0.6827) = 1.66 d_{min}, \end{aligned} \tag{6}$$

the probability that d_{ps} falls within interval $[d_{min}, 3d_{min}]$ is equal to $2\varphi(\frac{2}{1}) - 1 = 0.9545$. Accordingly,

$$\begin{aligned} d''rand =& \frac{d_{min} + (d_{min}+1) + \ldots + 3d_{min}}{3d_{min} - d_{min} + 1} \times 0.9545 \\ &+ 3d_{min} \times (1 - 0.9545) = 2.05 d_{min}. \end{aligned} \tag{7}$$

Equations (6) and (7) show that the d_{ps} produced by PNDBPR yields a 16% and 5% increase over that produced by PUSBRF and HBDRW respectively, in the case of $\rho = 2$ and $\rho = 3$. In order to have a reasonable comparison with peer work, we set $h = \frac{(d_{min} + \rho d_{min})}{2}$ which is used in all calculations in the next section.

5.2 Communication Overhead

Communication overhead refers to the total number of times of sending and receiving packets of all nodes. The entire communication overhead of PNDBPR includes the costs in following areas: broadcasting originated at the sink, restricted flooding originated at the source, random walks from the source, and probabilistic forwarding routing from phantom nodes to the source. Since the cost of the sink's broadcasting is the same as that in [3] and [6] and the associated analysis has been detailed in these two papers, here, we omit the discussion of it due to the space limitation of the paper. Also, because the source floods only once with $\rho d_{min} \ll n$ the associated cost can be ignored. Thus communication overheads in our work will be the sum of the cost of random walks from the source and the cost of probabilistic routing from phantom nodes to the sink.

Chen et al. [6] have showed that the average communication overhead of the PUSBRF protocol is:

$$\overline{E}_1 = h + \frac{1}{\pi} \int_0^{\pi} \sqrt{h^2 + H^2 - 2hHcosa}\mathrm{d}a, \tag{8}$$

and the average communication overhead of the HBDRW protocol is:

$$\begin{aligned} \overline{E}_2 =& h + (\int_0^{\theta} \frac{\sqrt{h^2 + H^2 - 2hHcosa}}{2\theta}\mathrm{d}a \\ &+ \int_{\pi}^{\pi+\theta} \frac{\sqrt{h^2 + H^2 - 2hHcosa}}{2\theta}\mathrm{d}a), \end{aligned} \tag{9}$$

where phantom nodes generated by the HBDRW protocol are distributed over a circle with radius 4θ.

For PNDBPR, as illustrated in Figure 1, packets randomly walk d_{ps} hops to arrive at a phantom node n_p ,and from there is transmitted to sink B with its forwarding routing. The average number of hops from source S to n_p is equal to:

$$\frac{d_{min} + (d_{min}+1) + \ldots + \rho d_{min}}{\rho d_{min} - d_{min} + 1} = \frac{d_{min} + \rho d_{min}}{2}. \tag{10}$$

The average number of hops from n_p to B can be calculated as:

$$\frac{1}{\rho d_{min} - d_{min} + 1} \sum_{d_{min}}^{\rho d_{min}} \int_0^{\pi} \frac{\sqrt{h^2 + H^2 - 2hHcosa}}{\pi}\mathrm{d}a. \tag{11}$$

Hence, using equations (10) and (11), the average communication overheads of PNDBPR is:

$$\begin{aligned} \overline{E}_3 =& (d_{ps} \\ &+ \lambda \sum_{d_{min}}^{\rho d_{min}} \int_0^{\pi} \frac{\sqrt{(d_{ps})^2 + H^2 - 2(d_{ps})Hcosa}}{\pi}\mathrm{d}a) \\ &\cdot p_1 + (\rho d_{min} \\ &+ \int_0^{\pi} \frac{\sqrt{(\rho d_{min})^2 + H^2 - 2(\rho d_{min})Hcosa}}{\pi}\mathrm{d}a) \cdot p_2, \end{aligned} \tag{12}$$

where $\lambda = \frac{1}{\rho d_{min} - d_{min} + 1}$, $p_1 = 2\varphi(\frac{\rho-1}{\sigma}) - 1$, $p_1 = 2 - 2\varphi(\frac{\rho-1}{\sigma})$, and $d_{ps} \in [d_{min}, \rho d_{min}]$.

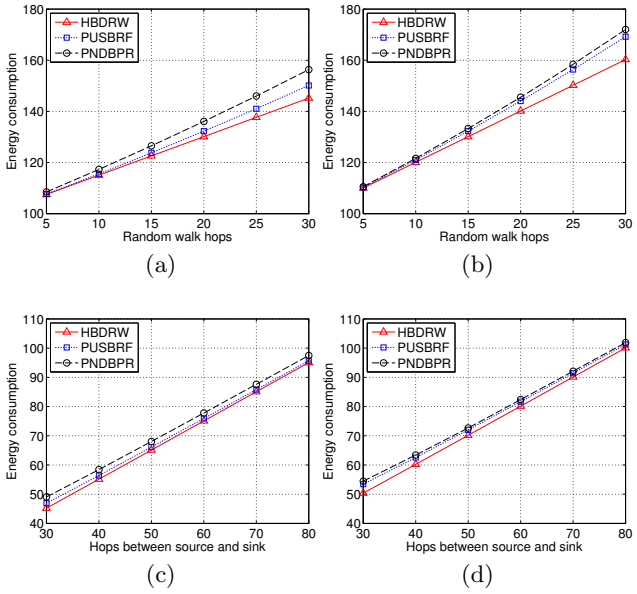

Figure 2: Comparisons of the communication overheads among HBDRW, PUSBRF, and PNDBPR.

The comparison of communication overheads of these three protocols is depicted in Figure 2. While Figure 2(a) and (b) illustrates the energy consumptions with respect to d_{min} when $H = 100$ and $\rho = 2$ and $\rho = 3$ respectively. Figure 2(c) and (d) illustrates the energy consumptions with respect to H when $d_{min} = 10$ and $\rho = 2$ and $\rho = 3$ respectively.

5.3 Discussion

For all three protocols PUSBR, HBDRW, and PNDBPR, their protection effectiveness will be improved and communication overhead exacerbated as the value of d_{ps} increases. This can be explained by the following observations: with a larger value of d_{ps}, (1) all protocols produce a wider source-based restricted flooding region leading to a higher communication overhead; and (2) all protocols produce more directed random paths, causing the adversary to take a longer time to trace back to the source from phantom nodes. On the other hand, with a smaller value of d_{ps}, all three protocols will have less communication overheads. For instance, all three protocols' communication overheads will be reduced to minimum when $d_{ps} = d_{min}$. However, in this case, the region of the source-based restricted flooding of these protocols is also reduced to the minimum with only a few generated random paths, which allows the adversary to easily trace and find the source from phantom nodes and comprise the security of the system. Hence, a balance between the protection effectiveness and the communication overhead reached by choosing proper values of d_{min} and σ is needed.

6. SIMULATION RESULT

For HBSRW, PUSBRF, and PNDBRP, simulation experiments were conducted on the MATLAB platform with regard to their respective average number of hops and distance from phantom nodes to the source. We consider two randomly deployed WSNs with 300 (network 1) and 600 (network 2) nodes, both of which are situated within a $1000m \times$

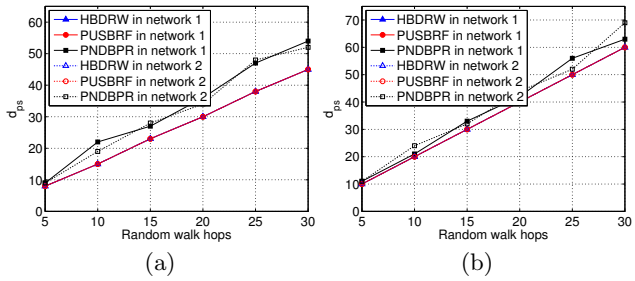

Figure 3: Average d_{ps} with respect to three protocols.

$1000m$ square with the source located at $(500, 500)$.

For each of the two WSNs and in regard to the three protocols, our simulation experiments processed 100 packet transmissions with different values of d_{min} and $\sigma = 1$ respectively. Generated results of the average number of hops and the average Euclidian distance between phantom nodes and the source are shown in Figures 3 and 4 respectively.

Figure 3 shows the d_{ps} comparisons among HBDRW, PUSBRF, and PNDBPR in two different topologies. The results in Figure 4(a) indicate that PNDBPR has 35% (in network 1) and 30% (in network 2) higher d_{ps} than HBDRW and PUSBRF. This is exactly aligned with previous theoretic analysis results. Similarly, Figure 4(b) presents the same insight: PNDBPR has 17% (in network 1) and 18% (in network 2) higher d_{ps} than HBDRW and PUSBRF, which also matches the theoretic analysis.

The comparisons of average Eculidean distance between the phantom node and the source node for three protocols are shown in Figure 4. Figure 4(a) and (b) are the results for $\rho = 2$ and $\rho = 3$, respectively. Since Eculidean distance between any two nodes can be interpreted as the hop counts between them, we can derive the same observation with Figure 3. On the other hand, it is noticed that the average Eculidean distance increases with the network scale expands. For example, in Fig 4(a) the distance of HBDRW and PUSBRF in network 2 is 66% higher than those in network 1, while the distance of PNDBPR in network 2 is 71% higher than that in network 1.

Figures 3 and 4 demonstrate that the deployment of a WSN only impacts the average Euclidian distance from phantom nodes to the source. Specifically, in a fixed area, a larger

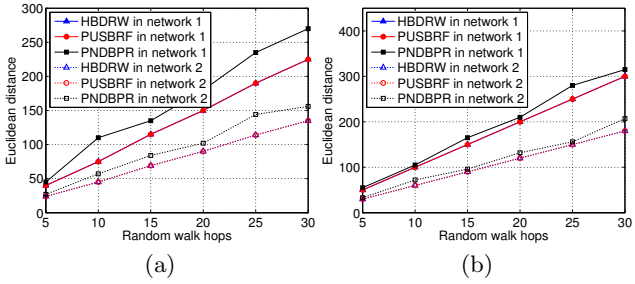

Figure 4: Average Euclidian distance between the phantom node and the source with respect to three protocols.

number of nodes in this area results in a smaller Euclidian distance with respect to the same number of hops; and vice versa. Also, note that the changes of d_{ps} in PNDBPR is caused by the random number x, not by the deployment of the WSN.

7. CONCLUSION

SLP protection is a notable issue and an on-going research topic in WSNs. Towards improving WSN SLP protections, we have proposed a pseudo normal distribution-based phantom routing scheme against patient adversaries. On the basis of the theoretical analyses and simulation experiments conducted in comparison with representative SLP protection protocols HBDRW and PUSBRF, we argue that the major work of this paper includes the following three aspects:

- The calculation of the number of hops from a phantom node to the source is based on a randomly-generated and Gaussian-distributed number. This machinery not only renders a uniform distribution of phantom nodes around the source node, but also enhances the diversity and dynamicity of the distribution.
- Stochastically constructed forwarding routing reduces the chance of the same path from phantom nodes to the sink node being repeated.
- Our proposed work significantly improves the privacy protection of source locations with the trade-off of a slightly higher communication overhead.

As for the future work, we plan to conduct comprehensive, multidimensional, and in-depth simulation experiments with regard to the comparison of three protocols PNDBPR, HBDRW, and PUSBRF.

Acknowledgement

The research of Jun Huang is supported by NSFC under grants 61309031 and 61272400, Postdoctoral Science Foundation of China (Grant No. 2014M551740), Program for Innovation Team Building at Institutions of Higher Education in Chongqing under grant KJTD201310, NSF of Chongqing under grant cstc2013jcyjA40026, Science and Technology Research Program of Chongqing Municipal Education Commission under grant KJ130523, and CQUPT Research Fund for Young Scholars under grant A2012-79.

8. REFERENCES

[1] Jun Huang, Yu Meng, Xuehong Gong, Yanbing Liu, and Qiang Duan. A novel deployment scheme for green internet of things. *IEEE Internet of Things Journal*, 1(2):196–205, 2014.

[2] Mauro Conti, Jeroen Willemsen, and Bruno Crispo. Providing source location privacy in wireless sensor networks: A survey. *IEEE Communications Surveys and Tutorials*, 15(3):1238–1280, 2013.

[3] Pandurang Kamat, Yanyong Zhang, Wade Trappe, and Celal Ozturk. Enhancing source-location privacy in sensor network routing. In *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pages 599–608. IEEE, 2005.

[4] Kashif Saleem, Mohammed Sayim Khalil, Norsheila Fisal, Adel Ali Ahmed, and Mehmet Ali Orgun. Efficient random key based encryption system for data packet confidentiality in wsns. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on*, pages 1662–1668. IEEE, 2013.

[5] Celal Ozturk, Yanyong Zhang, and Wade Trappe. Source-location privacy in energy-constrained sensor network routing. In *SASN*, volume 4, pages 88–93, 2004.

[6] Yin Li-Hua Chen Juan, Fang Bing-Xing. A source-location privacy preservation protocol in wireless sensor network s using source based restricted flooding. *Journal of Computers*, pages 1736–1747, 2010.

[7] Yun Li and Jian Ren. Source-location privacy through dynamic routing in wireless sensor networks. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[8] Yun Li, Jian Ren, and Jie Wu. Quantitative measurement and design of source-location privacy schemes for wireless sensor networks. *Parallel and Distributed Systems, IEEE Transactions on*, 23(7):1302–1311, 2012.

[9] Liming Zhou, Qiaoyan Wen, and Hua Zhang. Protecting sensor location privacy against adversaries in wireless sensor networks. In *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, pages 1384–1387. IEEE, 2013.

[10] Wei Tan, Ke Xu, and Dan Wang. An anti-tracking source-location privacy protection protocol in wsns based on path extension. 2012.

[11] Kiran Mehta, Donggang Liu, and Matthew Wright. Protecting location privacy in sensor networks against a global eavesdropper. *Mobile Computing, IEEE Transactions on*, 11(2):320–336, 2012.

[12] Wuchen Xiao, Hua Zhang, Qiaoyan Wen, and Wenmin Li. Passive rfid-supported source location privacy preservation against global eavesdroppers in wsn. In *Broadband Network & Multimedia Technology (IC-BNMT), 2013 5th IEEE International Conference on*, pages 289–293. IEEE, 2013.

[13] Hyungbae Park, Sejun Song, Baek-Young Choi, and Chin-Tser Huang. Passages: Preserving anonymity of sources and sinks against global eavesdroppers. In *INFOCOM, 2013 Proceedings IEEE*, pages 210–214. IEEE, 2013.

[14] Jorge Cuellar and Radha Poovendran. Toward a statistical framework for source anonymity in sensor networks. *IEEE Transactions on Mobile Computing*, 12(2), 2013.