Sink-Location Privacy Protection Based on Differential Privacy in WSNs
Ren-ji HUANG¹,*, Zhi-feng WANG² and Yong-zhen HONG³

¹Department of Information Security, Naval University of Engineering, Wuhan 43033, China
²Unit 61705, PLA, Beijing 100091, China
³Unit 91912, PLA, Ningde 352000, China
*Corresponding author

Keywords: Wireless sensor networks, Location privacy protection, Location-Based routing, Differential privacy.

Abstract. The location privacy of sink nodes is of vital importance for WSN security. Many location privacy protocols need the location of the sink node when routing, while most of them can’t resist an active attack. Hence, this paper proposed a new sink-location privacy protection protocol based on differential privacy (DPSL). The protocol uses anonymity and differential privacy to disturb the real position of the sink node and balances the privacy protection level and network performance by adjusting parameters. Analyzing the security, performance and power consumption of the protocol in Figure 4 to Figure 7, results showed that the protocol can protect location privacy with a little influence to power consumption of the network.

Introduction

Wireless Sensor Network (WSN) is a wireless Ad hoc network which is composed of plenty autonomous sensors[1]. The wide application of WSN also raises concerns about its security problems which includes location privacy leakage [2]. WSN is easier to be attacked because of its wireless and self-organizing. An adversary can destroy the nodes and analyze the structure and purpose of the network once getting their locations. So the location privacy of WSN is of vital importance especially in military applications and environment detection.

The Sink node plays an irreplaceable role in WSN for collecting data from sensors and forwarding them to the users or data processing unit. A network might be collapsed when attacking its sink node, so the location privacy of the sink node is one of the most important thing in WSN security. Many existing location privacy protection protocols need the position of the nodes when finding route from source to destination [3], which makes it inevitable to broadcast locations of nodes in the network. If some nodes or packages are captured, there will be a risk of sink node location privacy leakage. Therefore, we will put forward a new sink location privacy protection protocol to reduce the risk above.

Theoretical Basis

Location-Based Routing

A routing protocol that needs geographical location information to mark the destination or to compute route is called location-based routing [3]. It just needs the location information of nodes to transmit a package in location-based routing, so a route table or a whole path is no longer necessary.

GPSR (Greedy Perimeter Stateless Routing) [4] is one of the most fundamental location-based routing protocol. It requires every node can get the location information of neighbors and the sink. Every time a node transmits the packages, it chooses the neighbor node which is geographically closest to destination to ensure a short and power-saving route. If there isn’t a closer neighbor node since nodes distributed sparsely in the zone which is called a cavity, the node will transmit the packages by the ‘right-hand rule’ which means transmit the packages counterclockwise around the cavity.
Attack Model

There can be divided into general attack model and complex attack model by the method of an adversary in WSN location privacy protection\cite{5}. The adversary in general attack model can eavesdrop the flow, backward tracing\cite{6}, flow analyzing\cite{7} and other passive attack method. On the other hand, complex-attack-model adversaries can capture the node\cite{8}, tamper and forge the data, make denial of service attack, which gives more directly destruction to the network. The complex-attack-model adversary gives a greater threat but only a few protocols can defense it.

Differential Privacy

**Definition 1.** Suppose data set $D$ and $D'$ has the same attribute structure, and the symmetric difference is written as $D \Delta D'$. Then, $|D \Delta D'|$ expresses the data recorded by $D \Delta D'$. If $|D \Delta D'|=1$, we call $D$ and $D'$ as an Adjacent Dataset.

**Definition 2.** Suppose a randomized algorithm $M$ and $P_M$ is the set of all possible output from $M$. For any two adjacent datasets $D$ and $D'$ as well as any subsets $S_M$ of $P_M$, if

$$\Pr[M(D) \in S_M] \leq \exp(\varepsilon) \times \Pr[M(D') \in S_M].$$

(1)

We call the algorithm $M$ provides $\varepsilon$-differential privacy protection\cite{9}. The parameter $\varepsilon$ is called privacy protection budget\cite{10}, which determine the level of privacy protection.

**Definition 3**\cite{11}. Suppose a function $f$. The data set $D$ is the input of $f$ and the $d$-dimension real number vector is the output of $f$. We call $GS_f$ as the sensitivity of function $f$ if for all adjacent datasets $D$ and $D'$,

$$GS_f = \max_{D,D'} \| f(D) - f(D') \|.$$

(2)

The sensitivity of a function shows the influence to the query of data sets by the noise.

In differential privacy, there is little influence to the result that whether a specific record is in the data set or not, hence, the risk of privacy leakage caused by a record addition is controlled in a very small and acceptable range, which make it difficult for an attacker to get accurate individual information by observing and analyzing the calculation results\cite{9}.

Differential Privacy Based Sink-Location Privacy Protection Protocol (DPSL)

Most of the existing location privacy protection protocols in WSN defenses the adversary who can implement passive attack but thinks little about the active attacks\cite{12,13,14}. On the other hand, the location-based routing needs to share the location information of nodes in the network and even some location privacy protection protocols are based on the location-based routing protocols. For the sake of security of releasing location information in WSN, especially for preventing the sink-location information leakage, we propose the differential privacy based sink-location privacy protection protocol (DPSL) below. The protocol is to disturb the real location of the sink node before releasing based on the theory of differential privacy, which makes it impossible to compute the real location of the sink node even if they know the disturbed location.

Anonymous Area Generation

An anonymous area is to hide the accurate location information of the sink node in one area. The anonymous area meets the routing needs for source nodes and makes attackers unable to get the precise position as well, which increases the difficulty of finding the sink node. An anonymous area is generated as follows:

**Step1:** The sink node makes a flood of $h$ to the surrounding nodes. The nodes receive the package, record the hop of the package and transmit it to their neighbors with a hop added to the package. Then they forward their hops and location information to the sink node.
Step 2: In a period, the $h$-hops neighbors of the sink node compose an anonymous area and the sink node randomly selects $k$ nodes from them.

Step 3: Computing the center of gravity of the $k$ nodes (average of their location coordinate) as the substitution of the real location of the sink node.

Step 4: Do steps 2-4 again after a period to renew the anonymous area.

Location Disturbed Based on Differential Privacy

We use Laplace Mechanism to realize differential privacy protection to sink-location. Laplace Mechanism is to add noise that is Laplace distribution to the output of data sets and it meets $\varepsilon$-differential privacy. For the noise that is Laplace distribution $\text{Lab}(b)$, the probability density of $\text{Lab}(b)$ is

$$p(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$

Parameter $b$ of $\text{Lab}(b)$ is the scale parameter. For the data set $D$, a query $f$ on $D$, and the sensitivity of the query $\Delta f$, random algorithm $M(D) = f(D) + Y$ provides $\varepsilon$-differential privacy protection$^{[15]}$, and $Y \sim \text{Lap}(\Delta f / \varepsilon)$ is the Laplace distribution noise with the scale parameter $\Delta f / \varepsilon$.

Suppose the anonymous location coordinates of the sink node is $(x_c, y_c)$ and the average distance between adjacent nodes in the direction of coordinate axis is $d$, which means the sensitivity of the node location query is approximately $d$. So we use the Laplace noise with scale parameter $b = d / \varepsilon$ and the final released sink-location coordinates that is disturbed by Laplace Mechanism is

$$(x_i', y_i') = (x_i, y_i) + (Y_i, Y_2), Y_i \sim \text{Lap}(d / \varepsilon).$$

The locations information of other nodes in WSN are also used in location-based routing protocols. Although these nodes are less important than the sink node, the leakage of their location privacy may lead the adversary to find the sink node by tracing their locations. For the sake of security, sensor nodes in WSN need to be disturbed too. A location disturbed node is harder to be discovered and protects other nodes in the network from being captured by adversary. Suppose a random node in the network has the location coordinates $(x_s, y_s)$, then the disturbed released location coordinates in the network is

$$(x_s', y_s') = (x_s, y_s) + (Y_1, Y_2), Y_i \sim \text{Lap}(d / \varepsilon).$$

Routing Based on Location

There are two processes for source nodes to routing their data packages in DPSL. Firstly, packages are transmitted in the almost shortest path. That is a node finding the closest neighbor to the sink node as the next hop just like GPSR. Secondly, when packages arrived at the anonymous area, it will be transmitted according to hop and location information of the adjacent nodes. The routing protocol is showed as follows:

Step 1: Location information update. In a fixed period, the sink node generates an anonymous area, computes the anonymous location, disturbs the anonymous location and broadcasts the disturbed location to the whole network as the final released location coordinates of the sink node. At the same time, every node in the network broadcasts its location which is also disturbed to adjacent nodes.

Step 2: When a source node finishes collecting the data and generates a data package, the source node will use the pre-stored location information of adjacent nodes and the sink node to calculate a node that is closest to the sink node and itself. Then, the source node forwards the data package to the calculated node.
Step3: The relay nodes take the same processing as step 2 to forward the data packages after receiving it from the source nodes or other relay nodes. If there are no adjacent nodes closer to the sink node than itself, use right-hand rule to find the closest node to forward the package. Once there are no more adjacent nodes, backward to the last hop and find another neighbor to forward.

Step4: When the packages arriving at \( h \)-hops away from the sink nodes, they will be transmitted to the closest \((h-1)\)-hops node as the disturbed location of the sink node is no longer useful and hops to the sink node is the only practicable information.

Step5: The packages won’t be transmitted to the sink node directly until arriving at the 1-hop nodes.

The routing protocol of DPSL is shown in the Figure 1:

![Routing Diagram](diagram.png)

Figure 1. The Routing of DPSL.

**DPSL Analysis**

DPSL provides sink-location privacy protection in WSN by the combination of anonymous area and Laplace Mechanism and it protects the sink node from active attacks like nodes captured. We will analyze the protocol from the security, routing performance and power consumption of the network.

**Security**

The released node location information is disturbed by Laplace noise, so an adversary can’t get the precise location even if he captured some nodes. The disturbed location information will mislead the adversary to put more effort to find another node. Hence, it takes more time for an attack and makes it even more difficult to close the sink node. On the other hand, because of the differential privacy protection of Laplace Mechanism, the adversary will not calculate the real location of nodes even by means of multi-node query and crossing attack\(^{[16]}\). At last, the anonymous area hides the location information of the sink node since it just need an approximately location of the sink node in the routing protocol. Hence the ambiguous location enlarges the exploring range to the adversary.

Suppose the communication distance between nodes in the network is \( r \), and then the maximum anonymous area that might be generated is \( \pi (hr)^2 \). So the number of nodes in anonymous area are

\[
N = \sum_{x=1}^{h} \left( \frac{2\pi x r}{r} \right) + 1 = \sum_{x=1}^{h} 2\pi x + 1. 
\]  

As a result, the probability of discovering the sink node in the anonymous area for the adversary is

\[
P = \frac{1}{N} = \frac{1}{\sum_{x=1}^{h} 2\pi x + 1}. 
\]  

Under the protection of differential privacy, the probability of clarifying the adjacent data sets query for the adversary is no more than \( e^\epsilon \), so we could ensure the security of sink-location privacy if selecting a sufficient privacy protection budget \( \epsilon \) and suitable \( h \) of the anonymous area.
Routing Performance

A good route from source node to the sink node has the least number of forwarding (hops) and the shortest length of the path for it cost less resource of the network. Comparing to GPSR which is a typical location-based routing protocol, there is a deviation in shortest path calculation because of the anonymous location information. On the other hand, it doesn’t use the location information routing in the h-hop anonymous area, which makes the path a little longer. Figure 2 shows the difference in routing of two protocols.

![Figure 2. Compare the Routing of two protocols.](image)

The imaginary line in the Figure 2 is the route of GPRS and the full line is DPSL. The additional path caused by the location disturbed of the sink node is no longer than the distance between \( h \) -hop nodes on their paths \( X_h \). Let \( X_s \) be the distance between the real location and the released location of the sink node, so \( X_h < X_s \). The additional path caused by the last \( h \) -hop routing is no longer than \( \sum_{i=1}^{h-1} X_i \), and \( X_i \) is the distance between nodes in \( i \)-hop. To sum up, the DPSL has an additional path

\[
l_{\text{exa}} < X_s + \sum_{i=1}^{h-1} X_i < h \cdot X_s.
\]  

(8)

Power Consumption of Network

Power consumption of the whole network is mainly from three processes including network initialization, data collection by source nodes and data package forwarding along with the routing. Data collection is the basic function of the sensor node and the power consumption from data collection is independent of routing protocol, so we just leave it out. As for the other two processes, power consumption is mainly caused by data package forwarding and the consumption of power for a node in a package forwarding is almost the same. Hence, we can use total package forwarding times in the whole network to evaluate power consumption of the network\(^{[17]}\).

Suppose \( E_0 \) is the power needed in a package forwarding. In the network initializing stage, the sink node makes a flood of \( h \) hops and receives location information from surrounding nodes. The power consumption in this stage is

\[
E_c = E_0 \cdot \left( \sum_{i=0}^{k-1} n_i + \sum_{i=1}^{h} n_i \right).
\]  

(9)

\( n_i \) is the number of nodes in the \( i \) hop, \( n_0 \) is the sink node so \( n_0 = 1 \). All nodes in the network broadcast their location information to their neighbors. Let \( N \) be the number of nodes in WSN and they cost power \( N \cdot E_0 \).

Let \( H \) be hops from source nodes to the sink node so the power for package routing is \( H \cdot E_0 \). Suppose every source node generates a package in a period, then the whole power consumption is about

\[
E = E_c + N \cdot \left( E_0 + H \cdot E_0 \right).
\]  

(10)
For a sensor node in the network, it forwarding more times if closed to the sink node, so it cost more power. On the other hand, if the node is far away from the sink node, there will be relatively less power consumption.

**Simulation Experiment**

We use computer simulation to test the effect of DPSL in the security of location privacy protection and the length of routing path. The simulation environment is Windows 7 64 bit operation system with Inter Core i7, 8 GB RAM and Matlab 2017a. We simulate 1,200 sensor nodes uniformly distributed in an area of 1,000m×1,000m, and observe the data forwarding in the network using DPSL. Let the communication distance of nodes be 50m and the parameters of DPSL are as table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Variables</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy protection budget of sink node</td>
<td>$\epsilon_c$</td>
<td>1</td>
</tr>
<tr>
<td>Privacy protection budget of source node</td>
<td>$\epsilon_s$</td>
<td>5</td>
</tr>
<tr>
<td>Hop of sink-flooding</td>
<td>$h$</td>
<td>1-5</td>
</tr>
<tr>
<td>Number of nodes in anonymous area</td>
<td>$k$</td>
<td>5</td>
</tr>
<tr>
<td>Scale parameter of Laplace distribution</td>
<td>$b$</td>
<td>50</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>$N$</td>
<td>1600</td>
</tr>
<tr>
<td>Number of collected-data in a period</td>
<td>$m$</td>
<td>1000</td>
</tr>
<tr>
<td>Number of periods in a simulation</td>
<td>$t$</td>
<td>10</td>
</tr>
<tr>
<td>Location of sink node</td>
<td>$(x,y)$</td>
<td>(460,670)</td>
</tr>
</tbody>
</table>

Firstly, use rand to generate 1,200 nodes range from 1 to 1,000 (1600 nodes in range of 1,000m×1,000m) uniformly and a sink node. Secondly generate the anonymous area and releasing the disturbed location information by adding noise. Thirdly, randomly select $m$ nodes to generate a package and records the path of every hops of DPSL; Finally, generate a new anonymous area and change the hops of the sink node. These are shown in the Figure 3.

![Flow chat of simulation](image)

Then, plot the anonymous area and disturbed location of the sink node in different $h$ according to the simulation, which is shown by Figure 4.
With the parameter $h$ growing, the range of the anonymous area becomes larger and the released location of the sink node is varied. At the same time, the average distance between real and disturbed sink-location is longer with a higher $h$, which shown in Table 2, hence protects the sink node better.

Table 2. Average Distance Between Real and Disturbed Sink with Different $h$.

<table>
<thead>
<tr>
<th>$h$</th>
<th>Distance [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.74</td>
</tr>
<tr>
<td>2</td>
<td>26.70</td>
</tr>
<tr>
<td>3</td>
<td>44.82</td>
</tr>
<tr>
<td>4</td>
<td>54.31</td>
</tr>
<tr>
<td>5</td>
<td>70.78</td>
</tr>
</tbody>
</table>

Next, we compare paths of DPSL and GPSR. Randomly select a sensor as the source node and set the parameter $h = 3$, Figure 5 shows differences between the two protocol.

DPSL’s path is a little longer than GPSR for the disturbed location. Record all paths in the simulation and Figure 6 shows the path length changing with the distance between source and destination.
As a whole, the paths of DPSL is a little longer in the percentage of 12.04%, 12.33%, 10.68% and 8.34% with $h$ grows from 1 to 4. With the distance between source and the sink growing, the length of paths grows faster. Of course, the additional length of paths is acceptable as it provides sink-location privacy protection. Different $h$ influence only a little to the paths of DPSL and paths may become shorter in a higher $h$ within a certain distance.

At last, we analyze the power consumption of DPSL. Comparing of the average hops with the distance of source nodes changing into two protocols are shown in Figure 7.

In the anonymous area, average hops of two protocols are almost the same, but that of the DPSL is relevantly less. This is because DPSL is the shortest hop routing in the anonymous. When source nodes out of the anonymous area, average hops of DPSL becomes a little larger comparing to GPSR because of the disturbed location. By the way, DPSL costs a little more power in generating the anonymous area.

**Summary and Prospect**

For problems of sink-location privacy leakage in location-based routing, we proposed the differential privacy based sink-location privacy protection protocol and analyzed the performance. Through the simulation, the security of location privacy especially protecting the sink node from active attack is verified and the additional costs are acceptable. Characteristics of DPSL are as follows:

- Providing differential privacy protection with anonymous area, so the released sink-locations vary.
- Routing paths are not much influenced and close to the shortest.
- Locations of source nodes are also disturbed to provide more privacy protection.
- Calculation in the protocol for a sensor node is easy so it costs a little power and computing resources.

Since DPSL uses the basic routing, the flows in the network are unbalanced and paths of packages forwarding are easy to be tracked, which may cause location privacy problems. As a result,
we will study a multi-method location privacy protection protocol to unite DPSL and other location privacy methods which will resist more attackers.

References


