AN EFFECTIVE SCHEME OF LOCATION PRIVACY PRESERVING IN MONITORING SYSTEM FOR WSNs

K. Chaithanya Jyothi 1 and V. Srinivas 2
1Student, M.Tech (CSE), Anadapuram, Al-Ameer College of Engineering and Technology, VIZAG, A.P, India
chaithanyakjyothi2012@gmail.com
2Asst.Professor, (CSE), Anadapuram, Al-Ameer College of Engineering and Technology, VIZAG, A.P, India

Abstract - While a wireless sensor network is deployed to monitor certain events and pinpoint their locations, the location in formation is intended only for legitimate users. However, an eavesdropper can monitor the traffic and deduce the approximate location of monitored objects in certain situations. We first describe a successful attack against the flooding-based phantom routing, proposed in the seminal work by Celal Oz-turk, Yanyong Zhang, and Wade Trappe. Then, we propose GROW (Greedy Random Walk), a two-way random walk, i.e., from both source and sink, to reduce the chance an eavesdropper can collect the location information. We improve the delivery rate by using local broadcasting and greedy forwarding. Privacy protection is verified under a backtracking attack model. The message delivery time is a little longer than that of the broadcasting-based approach, but it is still acceptable if we consider the enhanced privacy preserving capability of this new approach. At the same time, the energy consumption is less than half the energy consumption of flooding-base phantom routing, which is preferred in a low duty cycle, environmental monitoring sensor network.

Keywords: Location privacy, wireless sensor networks, location monitoring system, aggregate query processing.

INTRODUCTION

Wireless communication had gained more popularity in recent years. The application driven force behind the popularity is easy deployment and mobility. Besides the wide applications of wireless local network today, emerging applications of wireless communication include wireless sensor networks and Mesh Networks [1]. It can be easily seen that wireless networking will gain more popularity and vast in-formation will be carried on wireless networks in the near future.

However, wireless communication media is a broadcast media, which poses a big challenge of how to protect information running on the network. Despite strong encryption of the data, wireless communication media still exposes some information about the traffic carried on the network. This is an inherent side effect of wireless communication. Mobility means that the communication is expected everywhere in the deployment area, which subsequently exposes the communication to possible attackers. Easy deployment means that there is certain openness in the protocol, which subsequently exposes some protocol information to possible attackers.

Location privacy is an important security issue. Loss of location privacy can enable subsequent exposure of identity information because location information enables binding between cyberspace information and physical world entities. For example, web surfing packets coming out of a home in a Mesh network enable an eavesdropper to analyze the surfing habits of one family if the source location of those packets can be determined.

In a wireless sensor network, location information often means the physical location of the event, which is crucial given some applications of wireless sensor networks. For example, in a battlefield, the location of a soldier should not be exposed if he initiates a broadcast query. In the panda-hunter problem, the location of the panda should not be exposed to hunters [2].

A wireless sensor network can be a low duty cycle net-work. Often, traffic has a strong correlation with a certain event at certain time. This gives big advantages to an eavesdropper since he does not need sophisticated techniques to discriminate traffic among different events. In this paper, we study the source location privacy problem under the assumption of one single source during a specific period. However, we need to point out that such a scenario can happen in a real wireless sensor network.

To preserve location privacy, we propose to use source and sink-based random walk for packet delivery. The sink first sets up a path through random walk which serves as a receptor. Each packet from a source is then randomly forwarded until it reaches the receptor. At that point, the packet is forwarded to the sink through the pre-established path. A random walk greatly reduces the chance of packets being detected. Even if an eavesdropper happens to detect one packet, the next packet is unlikely to follow the same path, thus rendering the previous observation useless.

The reminder of the paper is organized into 5 sections. In Section 2, related work is presented. In Section 3, we show by of an illustrated attack that randomness needs to be introduced carefully into the routing protocol. In Section 4, our implementation is described. In Section 5, simulation results are presented and discussed. In Section 6, we conclude
our paper.

**Wireless Sensor Network:**

A wireless sensor network (WSN) is a heterogeneous network composed of a large number of tiny low-cost devices, denoted as nodes (or motes), and one or few general-purpose computing devices referred to as base stations (or sinks). A general purpose of the WSN is to monitor some physical phenomena (e.g., temperature, barometric pressure, light) inside an area of deployment. Nodes are equipped with a proper communication unit (e.g., radio transceiver), processing unit, battery and sensor(s). Nodes are constrained in processing power and energy, whereas the base stations have laptop capabilities and not severely energy resources. The base stations usually act as gateways between the WSN and other networks (e.g., Internet).

There is a wide variety of applications for WSNs, ranging from military applications (e.g., perimeter monitoring through environmental (e.g., animal habitat monitoring) and health applications (e.g., patient health monitoring) to commercial applications (e.g., shopping habits monitoring, bridge structural health monitoring).

WSNs can be classified according to several aspects with impact on the security protocol design. One such aspect is the mobility of nodes and the base station. The nodes can be mobile or placed on static positions. The same holds true for the base station. Another consideration is the way the nodes are placed. The nodes can be deployed manually on specific locations following some predefined network topology or randomly deployed in an area, e.g., by dropping from a plane. The number of nodes is also a very important factor — number of nodes in a network can range from tens to tens of thousands. In our future work, we will focus on WSNs consisting of large number of nodes (hundreds or thousands) deployed without a priori topology design. Both nodes and the base station have static positions. As the reference platform we consider the MICAz node [3]. This node is based on the Atmel ATmega128L microcontroller with 128KB programmable flash memory, 512KB measurement flash memory and 4KB configuration EEPROM. It is equipped with an IEEE 802.15.4 compliant RF transceiver and the energy is supplied by two AA batteries. The node is running the TinyOS operating system.

**Motivation for Our Research:**

WSNs are becoming one of the building blocks of pervasive computing. They provide a simple, and in the near future also quite likely cheap, mechanism for area and entity monitoring. One of the dark sides of the WSN technology is that an inappropriate use can significantly violate privacy of humans. WSNs are frequently deployed to collect sensitive information. Typical example is a WSN monitoring movements in a building or traffic in a city. Such a network can be used to determine mine location of people or vehicles. If this information is available on a wide basis it can easily lead to blackmailing or stalking.

It can be also exploited by terrorists as a targeting tool to impact specific people or buildings. Another example of a WSN application, in which privacy is heavily exposed, is health monitoring. Here, the medical measurements should be available only to the attending physician. Wrong usage of simple commercial WSNs can easily result into serious privacy violations as well. Suppose that the WSN monitors people movements at a supermarket to improve the placement of products within the shelves. If someone is able to find out detailed information related to a particular person, then a seemingly innocent application turns into a privacy violating tracking device.

WSNs are sometimes able to provide a kind of information, which is far away from the purpose the WSN was originally designed for. Consider a network used for noise and sound monitoring, where people can be tracked based on their voice recognition. Such a network then also enables people tracking. This type of information leakage is denoted as a side channel. WSN applications are complex systems that are likely to contain a number of such channels. This example demonstrates that in most cases collected data themselves do not pose a privacy threat. The problem arises when the data can be linked to a specific person. This is why anonymity and proper identity management of the nodes, or their carriers, or the subjects that these nodes monitor, are needed. If an attacker is not able to link measured data with the measuring device or location then this data is of a little value for privacy attacks.

We feel that lot of effort has been put into ensuring traditional network security properties for WSNs, namely availability and confidentiality, and less attention was paid to privacy measures. We also have examined some traditional WSN security issues, especially security of routing algorithms, and proposed a method for automatic attack generation. However, as the examples above show, anonymity and privacy in general are of a great importance too. In [4], we have shown the significance of location privacy of important nodes and identified several open questions in this field. We have thus demonstrated the need for novel privacy preserving mechanisms for WSNs.

**RELATED WORK**

Our work is inspired by [2, 3]. An application scenario of a wireless sensor network for monitoring a panda is presented. Enabling outside monitoring of a panda without exposing the location of the panda to hunters is proposed as the *Panda-Hunter problem*. Phantom routing is used for message delivery from the location of the panda to the sink for preserving its location privacy. The phantom routing algorithm is composed of two phases. In the first phase, the source initiates a random walk. In the second phase, the packet is being delivered through flooding or single path routing. In this paper, we specifically address a possible attack against the flooding-based delivery method.
The idea of using intersecting paths to deliver packets has been proposed in rumor routing [5]. In rumor routing, an event is known by some sensors in the small of at neighborhood of as event location. A query is sent through random walk. A usable delivery ratio is achieved by a large number of query random walks intersecting with each other. This is different from our approach. In our approach, both event and query source use random walk to advertise themselves. Also, our concern is to provide privacy protection; thus a more dynamic structure than rumor routing is needed.

In [6], asymptotic of three query strategies over a sensor network are discussed. Proofs are given that the probability of unsuccessful delivery using source and receiver driven ‘sticky’ Brownian motion decays much faster than using a single Brownian motion with increasing random walk length. \((T-5/8)\) vs. \((\log (t)) - 1\) where \(t\) is how long the Brownian motion has lasted This result gives us a lower bound on the performance for our approach. In a real sensor network, the performance can be improved due to a limited size net- work. Also, in our approach, pure Brownian motion is not required for providing enough privacy protection. In [7], the problem of hiding the location of the base station in sensor networks is discussed. An attack model of determining the base station location through traffic analysis is used. To hide the traffic pattern, randomly delaying the sending time is proposed to hide the parent-child relationship given a traffic rate model. Our work instead addresses the spatial pattern of the traffic.

In [8], the problem of sharing the location information without revealing the identity privacy in the mobile data collection applications, such as a cell phone periodically reporting its location, is discussed. Multi target tracking algorithms can be used to identify each trajectory even when there is no identity information. A perturbation algorithm over multiple user paths is proposed to confuse the attacker. The algorithm takes advantage of the possible intersections of different paths and modifies location samples according to a nonlinear optimization solution. The artificially generated errors because wrong trajectories being calculated by the attacker. This is different from our problem. In our model, the location information is not explicitly included in the packets.

This section provides an overview of the current state of research in the area of privacy in WSNs. We first discuss privacy in the context of WSNs. Then we describe taxonomy of privacy protections for WSNs. With respect to this taxonomy, we also present a survey of state-of-the-art privacy protections. Related issues that do not fit to the taxonomy are discussed separately at the end of the section.

We base our work on a common perception of privacy. Privacy is the right to autonomy, and it includes the right to be let alone. Privacy encompasses the right to control information about ourselves, including the right to limit access to that information. Privacy in the context of WSNs involves both privacy of monitored subjects and privacy of nodes and base stations. Privacy of these parties is usually bound together to some extent. Breach of node privacy can lead to violation of the monitored subject privacy and vice versa. Privacy in WSNs can be classified into two categories – content-oriented privacy and context-oriented privacy.

Content-oriented privacy is threatened by an adversary who aims to manipulate and/or read the content of messages sent over a WSN. In contrast, context-oriented privacy is concerned about a protection of contextual information surrounding the content. Typical contextual information is location where the data has been sensed or time of the measurement.

Taxonomy of privacy protections for WSNs follows the previous privacy classification and provides further refinement. Protections are first classified into data-oriented (content-oriented) and context-oriented. Data-oriented protections are then categorized into privacy protections during data aggregation and private data query techniques. Context-oriented privacy protections can be split into location privacy preserving techniques, that cover data source location protections and sink location protections, and temporal privacy preserving techniques. An overview of the taxonomy is depicted in figure 1.

For completeness, we provide some references to most critical topics related to privacy in WSNs. Survey on wireless sensor network security is presented in below. Taxonomy of attacks on WSNs is given in below and summary of attacks on routing protocols is provided in below.

**Data Privacy:**

Data privacy protections target privacy of data collected by a network and queries posted to a network. There are two types of adversaries threatening the data privacy – external adversary and internal adversary. The external adversary only eavesdrops communication in a network. This kind of adversary can be easily defeated by encryption techniques such as SPIPS or pDCS. On the other hand, the internal adversary controls one or more nodes and usually has an access to encryption keys of these nodes. In such a case, the easiest way to protect privacy of data sent from nodes to the base station is to use end-to-end encryption based on keys shared between the sending node and the base station.
However, such encryption makes data aggregation within the network impossible. Therefore, one of the challenges is to provide secure and privacy preserving data aggregation in the presence of an internal adversary.

Multiple schemes were proposed to solve this problem. Query privacy in a similar setting was also investigated in survey. Since our future research includes the data-oriented privacy only partially, we do not explore this idea further.

**Context Privacy:**

Even though data privacy might be sufficiently protected, a sensor network may still leak valuable context-oriented information. Typical context-oriented information is information on source location, sink location and timing of events. This kind of information can be usually obtained by an external adversary using traffic analysis techniques. We summarize state-of-the-art protections in the following subsections.

**Location Privacy:**

Location privacy is extremely important in WSNs. Information on location of events or on location of base stations can be of a primary concern of an adversary. Suppose the Panda-Hunter Game where a WSN is employed to monitor endangered pandas in their habitat. It is sufficient for the adversary to find out location of sensors currently monitoring the panda to successfully localize and capture the panda. Similarly, the adversary only needs to find out location of the base station to be able to mount a physical or other DoS attack on the base station and thus inactivate the whole network. There are two basic types of adversaries considered when evaluating the location privacy – local adversary and global adversary. The local adversary has limited radio range and is able to monitor traffic only in a small part of the network at a time. On the contrary, the global adversary is capable of monitoring the whole network at a time and is able to immediately localize all transmitting nodes.

**Temporal Privacy:**

In addition to location, sensitive contextual information that can be inferred by an external adversary is timing of monitored events or a message rate. A clever adversary may abuse such information for example for victim tracking. Knowing the time and place of the message creation, she can estimate the victim movements. The problem of temporal information protection is referred to as the temporal privacy. The concept of the temporal privacy in WSNs. They have formalized the problem and proposed the Rate-Controlled Adaptive Delaying (RCAD) to protect the temporal privacy. In the RCAD, every node buffers an incoming message and randomly delays its retransmission according to the exponential distribution. Buffer preemption strategy is included to cope with the problem of overloaded buffers. When the node buffer is full, this strategy chooses a message to be transmitted immediately without further delay. Several such strategies are proposed and evaluated in. The RCAD is suitable for WSN applications where a reasonable delay can be tolerated. Note that also schemes protecting location privacy have the potential to protect the temporal privacy.

Many of them are based on the random walks and introduce time delays as well. Furthermore, the Periodic Collection scheme, where the sending rate is constant among the whole network, seems to provide the optimal temporal privacy, a because the traffic is independent of the event occurrence. Similar holds true for the Fit Probe Rate scheme, for example.

**WHAT IS REQUIRED FOR PRESERVING SOURCE LOCATION PRIVACY?**

We consider an extreme case for preserving privacy in which there is traffic only from a single source in a network. This enables the eavesdropper to use just the spatial traffic pattern to compromise the source location privacy. This is a reasonable assumption. First, sensor networks are low duty cycle networks. The time spent for delivering a packet from the source to the sink can be much shorter than the source packet interval. Second, if the eavesdropper has access to the packet source information, he can isolate the source traffic from the rest of the traffic.

**An Example Attack against the Flooding based Phantom Routing:**

In this section, we illustrate a simulated attack against the flooding-based phantom routing. We assume that the eavesdropper has minimum physical capability, which is the ability to detect the presence of a radio transmission. Also, to get a good estimate of the source location, the eaves dropper must be a part of a group of devices distributed in the network. Each device at a different location is considered an observation point. However, as we argued before, the number of observations is limited. The purpose of the attack is to show that by using only a limited number of observation points the source location can be approximated without much effort.

At each observation point, the eavesdropper can record the time of a radio packet. The propagation speed can be modeled as a Gaussian distribution and is unknown. Also, the time when the algorithm begins to flood a packet is un-known. So, the parameters to be estimated comprise the following tuple: \((x, y, v, t)\), where \((x, y)\) are the coordinates of the location where flooding begins, \(v\) is the propagation speed, and \(t\) is the time when flooding begins. Suppose that the coordinates of each observation point are \((x_i, y_i)\) and the packet is observed at time \(t_i\). The true distance between an observation point and the flooding source is:

\[
D_i = \sqrt{(x_i-x)^2 + (y_i-y)^2} \tag{1}
\]

The distance can also be written as:

\[
V D_i = v (t_i - t) \tag{2}
\]

Ideally, at each observation point we have \(D_i = V D_i\). However, to estimate those four parameters, multiple observations at different locations are needed to solve the equation. Due to noise, the estimates at each observation will not be consistent. To find the optimal solution, we use the mean square error approach. We minimize the following
Ideally, four observation points should be enough for this purpose. However, in the simulation, we found that using six observation points yields much better estimates. Using six observation points compared with using four observation points is still acceptable. So, we present the simulation results with six observation points only. To illustrate this attack, we have implemented the flooding-based phantom routing algorithm with TOSSIM [9]. We vary the number of hops during the random walk phase to check how this parameter affects the attack. The attack is being run over a network of 5000 nodes. We chose a large network size to show that even a large network can be susceptible to this attack. It’s hard to preserve source location privacy in a small network under the assumption of only one single traffic existing in the network during a specific period. We define the estimation error as the distance between the estimated location and true location. To measure the effectiveness of the attack, we fixed attackers at six locations in the network and varied the location of the source. The simulated network spans a rectangular area of size 100×100. The communication range of every sensor is 2.25.

The six locations of attackers are (10, 90), (10, 10), (90, 10), (90, 90), (40, 60), and (60, 40). The choice of the six locations is rather arbitrary provided that they are relatively far from each other and have good coverage of the network. Note that the chosen locations are not necessarily close to the real source. Figure 2 shows the estimation errors for different scenarios within a period in which 50 source packets were sent out. Table 1 shows the estimation errors and the summations of mean square errors. We deliberately return very large cost values for unreasonable solutions so that the optimization can converge faster. For example, scenario 2 in Table 1 has a large cost value. The reason is that the real location is outside the convex set of the observation points while the optimization is trying to find some point within this convex set. We adopt the following strategy to overcome this limitation. An inaccurate estimate has a very large cost value, which can be used by the eavesdropper to trigger the movement of the observation points. To illustrate this strategy, we moved the center of the original observation points toward the estimated location and re-estimate the location. However, during the moving process, if some observation points would move outside the network, we keep them at the boundary of the network. The whole process can be repeated. We use this strategy for the above example and the result is shown in Figure 3. To investigate the effectiveness of the attack given different random walk steps, we vary the length of the random walk. During the simulation, we found that there are many local minimum in the topology we used above, where a node inside the network does not have any neighbor in one direction. This causes many packets to be dropped before reaching the flooding phase and deteriorates the estimate quickly. However, there is no suggestion on dealing with this problem in the phantom routing algorithm. To avoid the local minimum problem, we run the simulation on a network with 5000 sensors. The sensors are uniformly distributed in a 100×100 rectangle area. The increased density makes the local minimum a rare case.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Estimation Error</th>
<th>Mean-square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.0</td>
<td>2345.6</td>
</tr>
<tr>
<td>2</td>
<td>31.3</td>
<td>2.9 × 1034</td>
</tr>
<tr>
<td>3</td>
<td>4.9</td>
<td>1588.2</td>
</tr>
<tr>
<td>4</td>
<td>6.7</td>
<td>2319.7</td>
</tr>
</tbody>
</table>

Figure 2: Estimation Results for Four Sources

Figure 3: Strategy to Close in on Source 2
Without loss of generality, we chose a source at (25.0, 70.4). The random forward hop count is chosen for the values 5, 10, 15, 20, 25, and 30. The simulation results are shown in Figure 3. The estimation errors are larger for higher hop count values. However, even for a large hop count of 25, the estimate is still usable. Part of the reason that the estimation error gets worse is the way the phantom routing is designed. In our implementation, the forward directions are categorized according to sensors’ $x$ coordinate. For random forward hop counts of 25 and 30, some of the packets are forwarded to the boundary of the network and dropped since there is no recovery mechanism defined. In Figure 4, this is shown as an increased packet loss rate. Since the source is located closer to one side of the network, only packets being forwarded to the closer side are lost. This causes the estimate to move toward the other side of the network. For those hop count values without packet loss, the increase in estimation error grows only linearly with the hop count and the growing speed is much slower than that of the hop count value. It shows that varying only the random forward hop count is not effective for providing better source location privacy.

**Drawbacks of Flooding:**

Privacy is lost when the adversary is able to predict the source location within a reasonable period of time. In the above illustrated attack, the adversary can predict the approximate position of the source when a single packet is flooded. Although randomness is introduced through the random walk phase, the adversary can improve the prediction through statistical estimation.

Modeling the routing as a random process, the effectiveness of the adversary’s strategy depends on how randomness is introduced and on how the adversary can sample this process. Given a known random process, every sample contributes to the adversary’s estimation of the invariant parameters. In our case, the parameters are the $x$ and $y$ coordinates of the source. To deter the adversary from predicting the exact location of the source, we would like to slow down the speed at which the adversary can sample this process.

Assume that the source sends multiple packets to the sink over a period of time and uses consecutive sequence numbers to label those packets. The interval of packets received by the eavesdropper is defined as:

$$T = S_i - S_{i-1}$$

(4)

Where $S_i$ is the sequence number of the $i^{th}$ packet from the source arriving at the same physical location. $T$ is a random variable. The larger $T$'s mean, the longer it takes for the adversary to get a good enough estimate of the source location. Note that the sequence number is used only for analyzing. The packet does not have to have a sequence number. Flooding is the worst method for protecting source location privacy in terms of $T$, which will take a fixed minimum value of 1 for all the locations in the network. Flooding enables the eavesdropper to accumulate information about the source location very quickly.

**GROW ALGORITHM**

Previous analysis of random walk is based on a planar graph. However, this is not the actual communication graph in a wireless sensor network. If we treat the communication graph as a non-planar graph during the implementation of the random walk, the probability of the source path and the sink path intersecting is much less than the previous asymptotic result. The scenario is shown in Figure 5(a). We use local broadcasting to solve this problem. Whenever a sensor forwards a packet, all its neighbors overhear this packet and create a route entry for the source pointing to the forwarding sensor. This does not require additional transmissions. Essentially the random walk is sticky not only for the sensors on the forwarding path but also for the neighboring sensors of this path. In effect, we build a pipe along the forwarding path.

The scenario not only exists between two paths, but also exists on a single random path itself. A random path might backtrack to itself after some time. However, we would like the path to extend as far as possible and as quickly as possible. In Figure 5(b), the sensor might forward the packet to one of its previous hop’s neighbors. Such a forwarding decision is not good since the random walk does not make much progress. To prevent this case, we use a Bloom filter [10] to store all current neighbors in the forwarding packet. When the next hop randomly picks up one of its neighbors, it checks whether that neighbor is already in the filter. Given a limited number of neighbors, the probability of false positives can be made very small by using a reasonable size filter within a packet. In other words, the packet will be forwarded to a sensor that has not seen the packet before with high probability. However, the potential for backtracking still exists. The only possible way to prevent backtracking is to remember all the sensors which have already seen this packet. This is not realistic for a large scale network. Currently, we did not address this issue in this paper. Instead, we rely on increasing the random path length to increase the coverage of the path. We are working on an improved method to address this issue.
To decrease the chance of backtracking, each sensor keeps a Bloom filter to store those neighbors that have already participated in the forwarding. Each time a sensor is forwarding a packet, it will store the last hop from which the packet came and the next hop which it forwards the packet to. When the random walk backtracks to a sensor, it will choose one neighbor that has never forwarded the packet before. In this way, we hope to maximize the coverage given a fixed path length. If the source and the sink are close to each other, the two random paths have a greater chance to intersect, thus the intersection points are closer to the source and the sink. This enables the eavesdropper to possibly trace the path. To prevent this from happening, we require a minimum path length of the source random walk. Note that we do not assume any routing infrastructure in GROW for generality. If extra information is available, we can certainly use the information to improve the performance. For example, if geographical locations of sensors are known, it is easy to identify which part of the network has not been visited. Thus a more effective greedy forwarding based on this information can be used.

![Figure 5: Non-Planarity in Communication Graph](image)

CONCLUSION AND FUTURE WORK

In this paper, we describe a possible attack against the flooding-based phantom routing. We propose GROW, a source and sink-based random walk as the alternative against this kind of attack. We improve the basic random walk by using local broadcasting and a Bloom filter. Simulation results show that it is practical to use our approach in a large scale wireless sensor network to protect source location privacy. Energy consumption is greatly reduced compared to the flooding-based phantom routing while there is only slight additional delay for message delivery. However, the delay is still acceptable. We believe that random walk is a basic approach for protecting source location privacy. However, there is still room for us to optimize the performance of this approach. Our future work is to find more efficient ways to build random paths.

REFERENCES