Source Location Privacy Considerations in Wireless Sensor Networks

Ruben Rios  
Computer Science Department  
University of Malaga  
ruben@lcc.uma.es

Javier Lopez  
Computer Science Department  
University of Malaga  
jlm@lcc.uma.es

Abstract

Wireless Sensor Networks is considered to be one of the cornerstones of Ambient Intelligence since they can be used in countless applications, where sensors are unobtrusively embedded into the environment to perform operations like monitoring, tracking and reporting. In such scenarios, privacy issues must be carefully considered since the mere observation of the network operation might reveal great amounts of private information to unauthorised parties. One of the problems that is gaining more attention in the realm of privacy, is the location privacy problem, which aims to prevent attackers from obtaining the location of specific nodes of interest to him. In this paper we provide a general overview of the proposed solutions to counter this threat. Finally, we will also discuss some open challenges and future directions of research for a convenient management of privacy issues in smart environments.

1 Introduction

Wireless Sensor Networks (WSNs) are adhoc networks comprised of a large number of small and costless devices (sensor nodes) which provide traditional computers with the ability to feel and reason about their surroundings, thus providing intelligence to the environment and enabling the Ambient Intelligence (AmI) paradigm.

The reduced cost and size of sensor nodes is one of the main advantages of WSNs but it is also one of its main limitations, since it greatly constrains the capabilities of sensor nodes. These devices must cope with a processor or memory equivalent to that of computers thirty years ago. Moreover, they are mainly battery powered and in most cases these are irreplaceable. Due to the lack of resources, sensor nodes are extremely vulnerable to different types of attacks, from the hardware to the application layer [19].

In general, privacy in AmI environments has traditionally been related to what is known as social privacy, that is, the need to prevent individuals from being tracked without their explicit consent. However, there are also network privacy considerations that must be taken into consideration. An attacker might analyse the network operation in order to retrieve information about the network itself and the data being collected. It is interesting to note that traditional security mechanisms, such as using cryptographic techniques in order to conceal the contents of the packets, cannot provide an appropriate level of privacy protection. The mere presence of messages traversing the network, encrypted or not, allows an observer to infer private information about the network.

In particular, location privacy in WSNs aims to prevent an adversary from being able to estimate the location of special nodes in the network, such as source nodes. Protecting such nodes from being localized is of vital importance since an adversary with that knowledge becomes very powerful. In some cases this information is innocuous (e.g. weather conditions)
though there are many scenarios where the location information of the events being monitored is critical (e.g. homeland security). For example consider a sensor network deployed inside a building (see Figure 1) to improve users’ quality of life. Due to user interaction with the environment, some (source) sensors will immediately generate network traffic in order to inform about the needs of the user and adapt building conditions accordingly. An external attacker might detect traffic variations in the network and thus localize the sources of messages, what finally allows him to approximate the location of the users as well as other private information.

In this work we investigate the location privacy problem and some of the proposed countermeasures to this problem. Firstly, in Section 2 we will introduce the privacy preservation problem in WSNs presenting some potential leaks of sensitive information from the network. Subsequently, in Section 3, we will present and analyse some of the solutions that have been proposed so far to protect the location of the nodes reporting the presence of events in the vicinity of the network. Finally, in Section 4 we will conclude the paper and present future directions of research due to the full integration of WSNs and the Internet.

2 Privacy in WSNs

Privacy and anonymity have been extensively studied since Chaum presented his paper on anonymous mailing systems [3]. This work was the breeding ground for an extensive literature on different aspects of anonymity as well as its application to different scenarios [5]. Every scenario has its own constraints and special features, which requires a specific design of the privacy preserving mechanisms. This is not different in the case of WSNs where a special care has to be taken mainly due to the extreme resource limitations of sensor nodes.

In the realm of AmI environments, privacy has been mostly related to the ability of WSNs of collecting and analysing large amounts of data from the users while the interact or simply move in those environments, namely social privacy. However, the privacy problem can be taken from a different perspective if we consider the privacy of the network. An attacker might capture and analyse network traffic to retrieve private information about the network itself and the data being collected. In fact, there is a link between these two aspects of privacy because the events being monitored by the network might be related to people. The main differences stands in which is the entity who might violate the privacy. In the case of social privacy, the user might not even be aware of being tracked since the devices collecting data are unobtrusively embedded into the environment, which turn the network owner into the privacy perpetrator. However, in the network privacy case, the adversary is an outsider who takes advantage of a sensor network deployed for legitimate purposes in order to obtain private information.

Clearly, the packet payloads might be protected using traditional confidentiality and in-
tegrity mechanisms. In fact, this is a prerequisite for privacy protection. However, an attacker unable to obtain the information contained in the packets can still retrieve sensitive information just by observing and analysing the communications. Pai et al. [15] show that simple observation of network traffic can reveal much information about the context in which the network is deployed. Different sensor nodes platforms communicate within different frequency ranges. Recent sensor platforms (e.g. Imote2) perform in the Gigahertz spectrum while older ones (e.g. cricket) perform at lower frequencies. This apparently innocuous information, can be used by an attacker to launch platform-specific attacks. Also the transmission rate can help an observer to determine the quantity and the nature of the events being monitored. For example, in the case of a body sensor network monitoring the heart rate of an individual with high blood pressure, the transmission of no messages might be an indicator of a heart problem. Moreover, the size of messages can be used to infer the type and precision of the data being collected. The size of packets reported by a sensor node monitoring the state of a light bulb (on/off) is smaller than those sent by sensors collecting data about the luminosity in a room. Also, some data aggregation protocols try to reduce network traffic by forcing nodes to reuse in transit packets to incorporate their own sensed data, thus increasing the size of the packets as they move closer to the base station. Finally, routing protocols might reveal the location of important nodes in the network such as the base station or the sources of messages, since sensor nodes usually send event data to a single or very few base stations in relative stable paths in order to preserve nodes’ batteries.

Another consideration about contextual privacy is made in [6] by Kamat et al. who claim that not only the occurrence of an event is sensitive information but also the time at which this event takes place (temporal privacy). This problem is more serious in the context of mobile asset monitoring, where an adversary can link the time and position of the events being monitored by the network and eventually he will be able to predict future behaviours.

A large amount of contextual information can be gathered by simple observation of network traffic. However, due to space limitations, in this article we focus on the source location problem and its countermeasures to prevent attackers from inferring the location of specific events taking place in the environment.

3 Source Location Privacy in WSNs

The main goal of source location privacy mechanisms is to prevent an attacker capable of performing traffic analysis attacks from determining the location of a node reporting the presence of an event in its vicinity. Indeed, the interest of the attacker is not the node itself but the location of the event. However, he might use that information to get an approximation of the location of the event.

This problem was first described in the well-known “Panda Hunter Game” [7,14]. It proposes a scenario where a large sensor network is deployed to enable biologists to monitor the behaviour of pandas in their environment. Whenever a panda comes into the hearing range of a sensor it starts transmitting messages to the base station. Although the sensor network is deployed for legitimate purposes, an attacker (the panda hunter) takes advantage of the already existing infrastructure to find and hunt pandas.

The attacker might try to gain information about the location of the reported events either from the content of the packets or from the traffic pattern generated due to the operation of the network. Packets contain both information in the payload and the header. Assuming that the packet payload is cryptographically protected, the attacker might still retrieve sensitive information from the headers. Header information is used at every hop for routing purposes and thus contain information about the sender and recipient of the packet (see Figure 2). This
information might be used to determine the location of these nodes. Therefore, it is necessary to protect the real identities of each sensor node taking part in a communication.

### 3.1 Node Identity Protection

Pseudonyms are commonly used to protect the identity of a node. A pseudonym is a name or identifier that can be used instead of a real name. However, fixed pseudonyms eventually provides no protection because the attacker is able to relate a pseudonym to a given node. For that reason, several mechanisms have been proposed to periodically renew pseudonyms.

The Simple Anonymity Scheme (SAS) [11] provides each node with a randomly distributed set of pseudonyms ranges from a network-wide pool of pseudonyms and, for each transaction, the node will select one of the assigned pseudonyms for a specific neighbour. The main drawback of this approach is the amount of memory needed to store the complete pseudonyms space, which must be long enough to avoid repetitions. In order to reduce the large memory waste in SAS, the Cryptographic Anonymity Scheme (CAS) is proposed [11]. In CAS neighbouring nodes share a random number and a sequence number, which are used as an input to a keyed hash function. For every new transaction, the sequence number is incremented resulting in a fresh pseudonym.

In these schemes sensor nodes internally store all the material to generate fresh pseudonyms. However, if an attacker is able to compromise the nodes he might be able to easily obtain past and future identifiers. To counter this problem, Ouyang et al. [13] propose two methods based on keyed hash chains. In Hashing-based ID Randomization (HIR) every node shares a pairwise secret key with its neighbours that are used to generate a new pseudonym for each message that is obtained by hashing the previous one, thus creating a keyed hash chain making more difficult for an adversary to obtain old pseudonyms. To further reduce the risk of node compromise attacks, Reverse HIR firstly creates the hash chain and uses it in reverse order.

Node identity concealment does not provide complete location privacy protection since an attacker can still determine which node sent a particular message when in its vicinity. However, by simply using this information the attacker is not able to determine the real originator of the data. Nonetheless, the attacker has other abilities that might help him to determine the source.

### 3.2 Traffic Pattern Protection

The traffic generated by the network can lead an attacker to the location of an event even if he is only capable of monitoring a small portion of the network. This type of attacker is usually known as local adversary and the strategy he follows consists of placing himself somewhere in the network (usually close to the base station) and waits for the arrival of messages. Upon the reception of the first message he is able to determine its origin and subsequently he moves in that direction until he reaches the sender node. This process is repeated until the attacker arrives at the real source of the data. Thus, the attacker is able to find the event source because the packets tend to follow relatively static paths to reach the base station. The aim of source location privacy mechanisms is to mislead the adversary in order to increase the safety period,
which is the number of packets sent by the source node before the source is localized. Several approaches were devised to counter this problem and we summarize them in this section.

3.2.1 Multiple Random Paths

The most commonly used strategy is based on the randomization of routes. In Phantom Flooding [14], every newly generated packet experiences two phases: a walking phase, in which the packet travels $h$ hops in a random walk to reach a phantom source, and a (baseline or probabilistic) flooding phase, which is intended to finally deliver the packet to the base station (see Figure 3). Since every packet follows a different random route, the attacker needs much more effort to find the source. The ideas proposed in this work were extended to further reduce the likelihood of successfully reaching the source nodes.

A two-way random walk named Greedy Random Walk (GROW) is presented in [22]. The source node sends every packet on a greedy random walk that will eventually intersect with a static random walk originated from the base station, which is called path of receptors. Once the message reaches a receptor node, it forwards the packet to the base station following the established path. It is a greedy algorithm because it attempts to expand itself as far as possible by avoiding revisiting nodes. Thus reducing the concern with random walks staying close to the source and the creation random paths with non intermediate nodes in common.

This concern is also discussed in [20], where Wang et. al propose the use of a Random Parallel routing. In such scheme, every sensor is pre-assigned $n$ parallel paths to the base station. Whenever the node sends a packet it randomly chooses one of the paths to convey the data. Since the paths are parallel and well separated the attacker is forced to stay in one of the paths, thus significantly increasing the safety period. In practice, deriving the paths is a complex task in large sensor networks, where the topology is quite unstable due to the nature of wireless links.

Instead of using a traditional random walk, the authors in [9] propose to randomly select a single intermediate node that will eventually send the information packets to the base station. The scheme is named RRIN. In order to avoid privacy problems, the intermediaries must be well away from the source node. Also, since the source node might not be aware of all existing nodes in the network it actually sends the data to a relative position in the network.
3.2.2 Directed Random Paths

In order to reduce the delivery time and to increase the protection level, some authors proposed the use of directed walk phases instead of traditional random walks. Both Phantom Flooding [14] and Phantom Routing [7] were the first to include a directed walk in order to avoid having phantom sources located close to the source. The directed walk separates neighbouring nodes into two groups, the former point to the base station and the latter point to the opposite direction.

Phantom Routing with Location Angle (PRLA) [21] introduces inclination angles to direct random walks. In many situations, increasing the length of the random walks does not increase the protection level since the phantom source is not necessarily placed in a privileged location to initiate the second phase. An adversary placed on the direct line from the base station to the source will be more likely to find the source when the phantom source is close to the shortest path. Therefore, phantom sources with a larger inclination angle are prioritized.

In order to overcome the limitations of the Random Parallel routing (Section 3.2.1), Wang et. al [20] also proposed the Weighted Random Stride routing (WRS). In WRS, every sensor makes its own routing decisions according to a forwarding angle and a stride. The node selects an angle (or sector) and choose a neighbour that matches that angle. The message is forwarded to the chosen node, which subsequently forwards the message to one of its neighbours that matches the predefined angle. This process is repeated for a number of hops defined by the stride (see Figure 4). Once the stride is finished, the receiving node starts a new stride. Larger forwarding angles are also prioritized in this scheme.

A similar approach is devised in the Identity, Route and Location privacy (IRL) scheme [16], which introduces the notion of trust and reputation [4] in the routing process. Every node classifies its neighbours in four different groups depending on their position with respect to the base station: forward (F), right backward (Br), left backward (Bl) and middle backward (Bm). Also, every nodes classify its neighbours into trustworthy or untrustworthy based on the number of successfully forwarded packets. The trust values are updated after each interaction between neighbours. When a node sends a packet it checks for trustworthy nodes in the direction of the base station. If all the nodes are untrustworthy, the same process is repeated with another group (Br or Bl). Finally, Bm is checked and if no trustworthy nodes are found, the packet is dropped.

Li and Ren [9] extend the RRIN scheme by selecting multiple random intermediaries. An angle-based and a quadrant-based approach are proposed for the selection of intermediaries. In
the angle-based approach, the source selects a maximum angle, $\beta$, to limit the location of the last intermediate node, within the range $(-\beta, \beta)$. Then the source selects the other intermediate nodes to be between the last intermediate and itself. The process in the quadrant-based approach is similar but the source divides the network into four quadrants according to a random angle $\alpha$ and chooses the last intermediate node to be in some of them. The authors claim to obtain global source location privacy, which must not be confused with source location privacy against a global attacker.

3.2.3 Network Loop Methods

The Cyclic Entrapment Method (CEM) [12] is based on the creation of traps, in the form of network loops, to keep the adversary away from the real path to the source. A network loop consists of several nodes transmitting decoy messages in a circular fashion. Local adversaries tracing back the path to the source node will at some time reach the loop, where the path forks in several possible directions. At this point, the adversary must decide in which direction to move. In case of making the wrong choice, he will be trapped until he discovers.

The Network Mixing Ring (NMR) approach [8] is based on the idea of mix zones proposed in [2], where the identities of the users moving in a smart environment are made indistinguishable (mixed) between each other. NMR creates a virtual ring of nodes around the base station, where messages are clockwise relayed in order to mix them (see Figure 5). For every new message, the source selects a random intermediate node to forward the message to the ring. Messages within the ring change their appearance at every hop to thwart message analysis. Also, messages hop along the ring a random number of times before being transmitted to the base station.

3.2.4 Fake Message Transmissions

Previous approaches turn out to be ineffective against more powerful adversaries with global eavesdropping capabilities. This type of attacker, known as global adversary, is able to monitor the transmission rate of every node in the network. Thus, a global adversary can easily spot the source of messages among mere intermediaries because upon the observation of an event, a message is immediately transmitted to the base station, thus revealing the location of the source. The underlying idea of this type of protection mechanism is to hide the presence of
real events within the transmission of fake messages. Consequently, both types fake and real 
messages must appear indistinguishable from the attacker’s point of view.

The idea of introducing fake messages was first used in [14] with the Short-lived and Persistent Fake Source strategies. In both strategies some nodes simulate the presence of real events in their vicinity. In the former, when a node receives a real message it decides based on some probability to produce a fake packet and flood the network with it. In the latter, every node decides whether to become a fake source based on a certain probability. These strategies are still inefficient against a global adversary because they are activated by real messages or the real sources are dynamic while the fake ones are persistent, respectively.

Mehta et al. [10] are the first to consider the threat of global adversaries. The authors suggest the Periodic Collection scheme, where every sensor node sends messages at a fixed rate $\Delta$. In case there is no real data to report, the sensor node transmits a bogus packet in that slot. When a real event occurs, the message is stored temporarily until the next transmission period. This method provides the best level of protection (perfect event source unobservability) because network traffic is independent of the occurrence of real events (see Figure 6). However, depending on the value of $\Delta$ this technique might either impose a large message delay or a high energy consumption.

3.2.5 Energy-Aware Approaches

Although the Periodic Collection scheme (Section 3.2.4) provides an optimal level of protection, it is too resource consuming for a constraint scenario like WSNs. Thus, some solutions attempt to solve the problem of providing source location privacy without introducing an excessive delay in nodes transmissions, while preserving nodes’ batteries. Most of these solutions are also based on the transmission of fake messages but in a more sophisticated way.

The Source Simulation scheme [10], reduces the energy waste by decreasing the number of fake sources. The idea is to simulate the presence of the object being monitored in the field. This strategy requires in-depth knowledge about the behaviour of the assets being monitored in order to be able to appear as real to the adversary. In the case of monitoring a moving object, having a static subset of sensors constantly sending messages can be easily detected by the adversary as a decoy mechanism. Therefore, sensor nodes must be carefully programmed to transmit fake messages following a coherent pattern with the events being monitored.

A Proxy-based Filtering Scheme (PFS) is proposed in [23] to reduce network traffic and thus save energy. In this approach sensor nodes produce real or fake message and transmit it to some proxy nodes which are strategically placed in the network. The proxy nodes, upon the reception of a real message it re-encrypts it and stores it temporarily for later forwarding. However, upon the reception of fake traffic, it simply drops it unless there are no real messages to transmit. To further reduce dummy traffic more proxy layers might be placed in the direction of the base station. This approach is called the Tree-based Filtering Scheme (TFS).

Shao et al. [17] propose a scheme which aims to reduce the additional overhead produced
Figure 7: Statistically Strong Unobservability

by dummy traffic by taking advantage of beacon messages, which are periodically broadcast in WSNs for configuration purposes. Event data is cryptographically hidden into beacon messages, thus introducing no extra traffic in the network. After several beaoning intervals the beacon frame containing event information reaches the base station. The main drawback of this approach is a large delivery time, which is highly dependent on the number of hops between source and destination. Beacon frames are broadcast at intervals ranging from milliseconds to several hundreds of seconds. A cross-layer approach is also proposed by the authors to reduce the latency but this is no longer resilient to global adversaries.

Also some statistical approaches have been devised to minimize the latency. In [18], the authors propose to transmit fake messages according to a probability distribution ($F_i$) in such a way that upon the occurrence of a real event ($E_1$) it is transmitted in the shortest time possible ($F_{42}$), before the next scheduled fake transmission ($F_{41}$), without altering the parameters of the distribution (see Figure 7). Thus, an adversary is not able to detect real messages by performing statistical tests on inter-transmission times within a transmission window. As real messages are re-scheduled, the presence of bursts of events might skew the mean of the distribution. To overcome this problem, a mean recovery mechanism, which delays subsequent transmissions. However, Alomair et al. [1] discussed that an adversary might attempt to spot differences between any two transmission windows to detect the presence of real events. Since in the presence of real events, short inter-transmission times followed by long inter-transmission times are more likely to happen due to mean recovery mechanisms, an adversary could count the number of occurrences of short-long inter-delays, and thereby distinguish the interval containing real events.

4 Conclusions

The extensive benefits that the Ambient Intelligence paradigm will bring to our society clearly deserves due care from the research community. Several challenges are still to be met until AmI environments are widely deployed. Among these challenges, privacy is of uttermost importance. In this work, we pay special attention to privacy issues resulting from the deployment of WSNs as we believe that it will be a predominant technology in these environments. In particular, we focused on the source location privacy problem and the most relevant solutions proposed to date.

Source location privacy in sensor networks poses new challenges due to the extreme limitations of sensor nodes. Thus, these solutions must carefully trade-off between an adequate protection level and the overhead incurred in the application of these countermeasures. We have categorized the proposed solutions based on the nature of their approach (see a summary in Table 1). It is clear that a more powerful attacker requires more resources from the network in order to protect the location of the source nodes. In general, the solutions are based on modifying the traffic pattern of the network.
Table 1: Summary of Source Location Privacy Solutions in WSNs

<table>
<thead>
<tr>
<th>Adversary</th>
<th>Approach</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protection</td>
<td>Cryptographic schemes</td>
<td>[11, 13]</td>
</tr>
<tr>
<td>Traffic Pattern</td>
<td>Local Multiple random paths</td>
<td>[9, 14, 20, 22]</td>
</tr>
<tr>
<td>Protection</td>
<td>Directed random paths</td>
<td>[7, 9, 14, 16, 20, 21]</td>
</tr>
<tr>
<td></td>
<td>Network loop methods</td>
<td>[8, 12]</td>
</tr>
<tr>
<td></td>
<td>Global Fake message transmissions</td>
<td>[1, 10, 18, 23]</td>
</tr>
<tr>
<td></td>
<td>Energy-aware approaches:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Source simulation</td>
<td>[10]</td>
</tr>
<tr>
<td></td>
<td>- Cross-layer scheme</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>- Message filtering</td>
<td>[23]</td>
</tr>
<tr>
<td></td>
<td>- Statistical approaches</td>
<td>[1, 18]</td>
</tr>
</tbody>
</table>

It is interesting to note that privacy problems exist at different levels. In this paper, we investigated the privacy problem at the communications level. However, in the realm of AmI scenarios, where the final aim is the provision customized services to the users, it is also necessary to incorporate privacy preservation mechanisms at higher levels. Consider the case where a user moves in an intelligent environment, the services being provided will probably adapt or move with him. In such case, an observer could manage to infer the type of service being accessed and relate it to a user regardless of using privacy techniques at the communication level. Therefore, in order to tackle the privacy problem it must be considered in a holistic way, from the communication to the application layer.

Finally, the interconnection of various smart environments through the Internet will result in a very promising area of research. This will allow the exchange of information between domains giving raise to more intelligent environments capable of providing enhanced services to the users. In such scenarios, where sensor data might be remotely accessed or communicated, new threats to privacy will appear as well as new and more skilled adversaries. Similarly, new countermeasures will need to be devised.

Acknowledgements

This work has been partially supported by the Ministry of Science and Innovation through the ARES (CSD2007-00004) and SPRINT (TIN2009-09237) projects. The latter is co-financed by FEDER (European Regional Development Fund). The first author has been funded by the Spanish Ministry of Education through the National F.P.U. Program.

References


