R. Rios, J. Cuellar, and J. Lopez, “Robust Probabilistic Fake Packet Injection for Receiver-Location Privacy in WSN”, *17th European Symposium on Research in Computer Security (ESORICS 2012)*, LNCS vol. 7459, pp. 163-180, 2012.
http://doi.org/10.1007/978-3-642-33167-1_10
NICS Lab. Publications: https://www.nics.uma.es/publications

Robust Probabilistic Fake Packet Injection for Receiver-Location Privacy in WSN

Ruben Rios[1], Jorge Cuellar[2], and Javier Lopez[1]

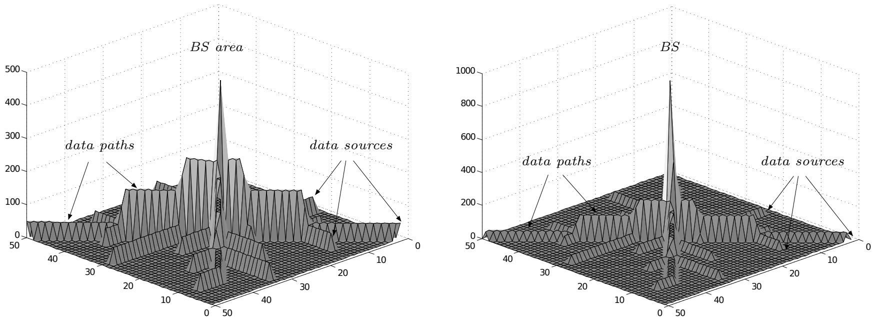
[1] Network, Information and Computer Security (NICS) Lab,
University of Malaga, Spain
[2] Siemens AG, Munich, Germany
{ruben,jlm}@lcc.uma.es
jorge.cuellar@siemens.com

Abstract. The singular communication model in wireless sensor networks (WSNs) originate pronounced traffic patterns that allow a local observer to deduce the location of the base station, which must be kept secret for both strategical and security reasons. In this work we present a new receiver-location privacy solution called HISP (Homogenous Injection for Sink Privacy). Our scheme is based on the idea of hiding the flow of real traffic by carefully injecting fake traffic to homogenize the transmissions from a node to its neighbors. This process is guided by a lightweight probabilistic approach ensuring that the adversary cannot decide with sufficient precision in which direction to move while maintaining a moderate amount of fake traffic. Our system is both validated analytically and experimentally through simulations.

1 Introduction

Wireless Sensor Networks (WSNs) [1] can be seen as an extension of ordinary computers that allow them to sense and react over the environment surrounding them. These networks are composed of battery-powered devices, the sensor nodes, which are capable of measuring the physical phenomena in their vicinity and wirelessly transmit these data to a central node called base station or sink. The base station gathers the packets from different sources and processes them in order to gain insight about the area being monitored.

This technology has raised a tremendous interest in the academia and is finally drawing the attention of companies because of their potential integration into many diverse application scenarios. The criticality of many of these applications together with the hardware limitations of sensor nodes require the development of tailored security mechanisms to guarantee the proper operation of the network in the presence of adversaries [2]. Most of the countermeasures found in the literature have been built on top of cryptographic primitives in order to protect the information traversing the network. However, even when secure encryption algorithms are used to protect message content, traffic analysis reveals sensitive contextual information about the network and the application scenario [3].



(a) Number of packets sent (b) Number of packets received

Fig. 1. Communication pattern in a typical WSN

A noteworthy problem related to contextual privacy is the protection of the location of relevant network nodes. In particular, the location of source nodes is important because it provides the attacker with information about the area where special events occur. Consider, for example, a WSN deployed to control the transportation of hazardous materials into and out of a nuclear or chemical plant. If an attacker obtains the location of source nodes he might be able to approximate the location and movements of the trucks carrying these materials. Moreover, this information might allow the adversary to deduce sensitive information about the distribution of the plant or even the presence of problems in the industrial processes. On the other hand, protecting the location of the base station is tremendously important because if it gets compromised or even destroyed, the whole system is rendered useless. Besides the physical protection of the network, the location of the base station is strategically critical because this key device is most likely housed in a relevant facility within the plant.

The aforementioned privacy problems are extensible to any application scenario because they are caused by the particular way of operation of WSNs. In a typical configuration, packets containing event data are generated at various locations from where they are forwarded in the shortest possible path towards the sink. Fig. 1 represents a WSN consisting of 50 × 50 nodes where 15 nodes are reporting event data using a shortest-path routing protocol. Although this is the most suitable configuration for preserving the limited energy budget of sensor nodes, it produces pronounced traffic patterns that reveal the location of both the source nodes and the base station.

Most of the research so far has focused on the source-location privacy problem while the protection of receiver-location privacy has received much less attention. The main reason is that hiding the base station is a especially difficult task because all the traffic is addressed to this single node with the consequent increase of traffic in its vicinity. These features are exploited by adversaries who may monitor the direction of packet flows or the amount of traffic being transmitted to uncover the location of the sink. To counter these strategies several works have focussed on the use of random routing protocols [4, 5] and the injection of fake traffic [6–8]. Many of these solutions fail to provide a sufficient protection level or they impose prohibitive energy costs and message delivery delays.

The main contribution of this work is the HISP (Homogenous Injection for Sink Privacy) protocol. HISP is based on the idea of locally homogenizing the amount and direction of the packets forwarded from the sensor nodes to their neighbors. Besides, this protection mechanism ensures that event data reach the base station without incurring in significant delays or excessive energy costs. To achieve this, HISP sends real packets using a random walk algorithm and introduces controlled amounts of fake traffic in such a way that the distribution of real packets remains probabilistically hidden.

The rest of this paper is organized as follows. Sec. 2 describes the network and threat model. A detailed description of the HISP protocol is presented in Sec. 3. Subsequently, in Sec. 4, we evaluate and analyze the main features and potential limitations of our approach. Moreover, Sec. 5 presents a discussion about the privacy protection level provided by the proposed solution. Sec. 6 compares this work with previous solutions in the area of location privacy in WSNs. Finally, Sec. 7 concludes the paper.

2 Problem Statement

This section presents the main features of WSNs as well as the adversarial model under consideration. Moreover, it introduces the main assumptions applicable to the rest of this work.

2.1 Network Model

We consider WSNs used for monitoring purposes. Usually, this type of networks follow an event-driven model, which means that the decision of transmitting data to the base station is made by individual sensor nodes upon the occurrence of special events. Consequently, this implies a many-to-one communication model where all the information flows from source nodes to a single base station.

Also, we assume that the deployed WSN is comprised of numerous sensor nodes which are deployed in a vast area. This prevents the adversary from controlling the communications in a large portion of the network as well as having all sensors within easy reach. Moreover, sensor nodes could be hidden or placed out of the visual field of the adversary. Sometimes this is not a strong assumption, for instance if we consider application scenarios such as under-water or under-ground sensor networks.

We focus on highly-connected sensor networks composed of n sensor nodes, where every node is aware of its adjacent neighboring nodes and the direction towards the sink. We require sensor nodes to have relatively high connectivity, that is, every node has several neighbors with which they share keys in order to be able to transmit to or receive packets from various locations. Note that, in sparse WSNs, an adversary can identify the route followed by messages more easily because the number of potential senders or receivers is rather limited.

Finally, we assume that sensor nodes make use of secure encryption algorithms that prevent an adversary from obtaining any identifiable information from packet payloads. In other words, the encryption mechanism under consideration must be robust to cryptanalysis attacks and also provide indistinguishability between real and fake transmissions. The key management scheme is beyond the scope of this paper. A survey can be found in [9].

2.2 Adversarial Model

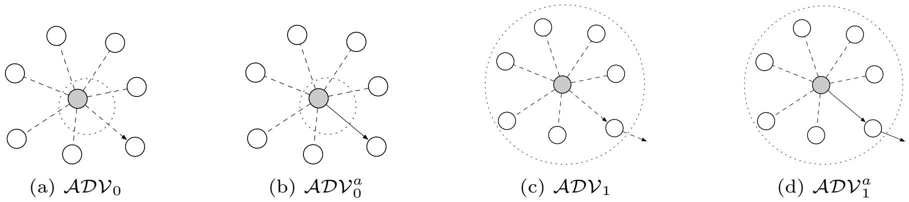
The adversarial model considered is external, passive and mobile. An *external* adversary does not control sensor nodes and thus has no access to the key material. A *passive* attacker does not interfere with the communications or the normal operation of the network. In general, passive adversaries limit their actions to performing traffic analysis attacks. These attacks depend on the hearing range of the adversary, which is typically equivalent to that of an ordinary sensor node[3]. Moreover, a *mobile* adversary is capable of moving in the field based on his observations according to a particular strategy.

First, we define adversaries based on their eavesdropping capabilities. In particular, we take into consideration both the hearing range and the ability to retrieve packet header information. With respect to the hearing range, we might find adversaries capable of observing the transmissions of a single and adversary capable of monitoring all the communications in the network. On the other side, we distinguish between adversaries who, by observing a message, are capable to recognize the addressee of the next hop and those unable to retrieve this information. This information is contained in the header of the packets but it might be protected by means of some pseudonyms mechanism[11]. Next, we provide a formal definition of the adversarial model:

Definition 1 **($\mathcal{ADV}$)** *Let $X = \{x_1, x_2, \cdots, x_m\}$ be the set of sensor nodes comprising the network and let x_i be an ordinary sensor node in the proximity of the adversary. We define the following adversaries:*

- *$\mathcal{ADV}_n$ chooses first a node x_i, and then observes the transmissions of node x_i and all its neighbors within distance n. On the next round he may choose a different node $x_{i'}$. The choice of the next $x_{i'}$ depends on the movement strategy, see for instance time-correlation and rate monitoring, below.*
- *$\mathcal{ADV}_n^a$ is similar to the previous one: he observes the transmissions of node x_i and all its neighbors within distance n, but this observation includes also the addressees of all those transmissions.*

[3] The hearing range of current sensor nodes operating outdoors is around 100 meters for low power configurations [10]. However, these values might be altered by many factors such as the signal frequency or the presence of obstacles.



(a) $\mathcal{ADV}_0$ (b) $\mathcal{ADV}_0^a$ (c) $\mathcal{ADV}_1$ (d) $\mathcal{ADV}_1^a$

Fig. 2. Adversarial Model Examples

Fig. 2 provides a visual representation of the different adversarial models at distances no larger than 1. The central node, x_i, broadcasts a message that is received by all its immediate neighbors. Transmissions are depicted by means of lines and arrows. An arrow represents that the packet is addressed to that particular node while dashed lines represent that these nodes are passive observers. When the arrow is dashed we mean that the node identifier cannot be retrieved by the attacker while the ordinary arrow represents that the identifier is accessible. Finally, the dotted circles represent the hearing range of the adversary.

We can define other types of attackers that are not able to see all the neighbors within a certain distance but a partial set of them. These type of attackers and their analysis will be left for future work. The attacker model considered in this work has a limited hearing range, similar to those depicted in Fig. 2. This is the typical hearing range considered in the literature, which focusses on adversaries with eavesdropping capabilities equivalent to an ordinary sensor node. Based on his observations and the peculiarities of the communication model, the adversary decides in which direction to move in order to reach the sink. Also, we are consistent with the two potential strategies proposed in the literature.

The adversary might perform two types of attacks to decide on the next move. In the *time-correlation* attack, the adversary observes the transmission times of a node and its neighbors. Based on the assumption that a node forwards a received packet shortly after receiving it, the adversary is able to deduce the direction to the sink and move accordingly. In the *rate-monitoring* attack, the adversary moves in the direction of the nodes transmitting a higher number of packets. This attack is based on the fact that nodes in the vicinity of the base station must transmit their own data as well as forward the traffic from remote sources. This strategy is less efficient because it requires the adversary to capture a sufficient number of packets before moving. Additionally, this attack is not effective when there are very few data sources or the adversary is not close to the sink.

3 Homogenous Injection for Sink Privacy

This section provides a detailed description of the HISP protocol. We present an overview of its main features as well as some fundamental properties that must be hold to ensure a robust privacy-preserving transmission protocol and the arrival of packets to the sink. Also, the neighbor discovery process is described since it is crucial for the subsequent data transmission stage.

3.1 System Overview

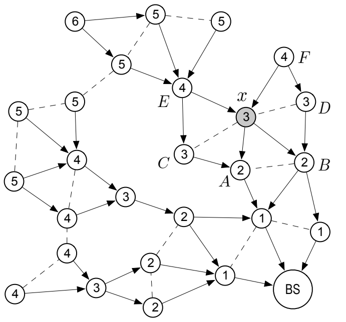
The HISP protocol is basically a biased random routing reinforced with the injection of controlled amounts of fake traffic. Upon the reception of a real message, the sensor node decides the next hop in the route based on some probability, which is dependent on the connectivity of the node. Fake packets are incorporated to prevent the adversary from being able to determine the direction to the sink when observing the number of packets being forwarded in his vicinity. In this way, messages are evenly distributed among all the neighbors of a node without introducing significant delays in the delivery of packets.

We devised a computationally inexpensive approach to determine the recipients of fake and real messages. Whenever a node has to transmit event data it picks a pair of neighbors. This pair is obtained from the combination of two elements without repetitions from all neighbors in its routing table. The routing table of the sensor is sorted incrementally (see Fig. 3), such that neighbors closer to the base station are placed first, then neighbors at the same distance, and finally neighbors in the opposite direction. This arrangement give rise to combinations of neighbors where nodes closer to the sink are more likely to appear in the first position of the pair while the second position contains equally distant or further neighbors. Thus, the random selection of these pairs leads to an homogeneous distribution of messages among all the neighbors of the node. HISP takes advantage of these features to send real packets to the first element and fake packets to the second.

3.2 Neighbor Discovery Process

Shortly after the deployment of the network, a network discovery protocol is launched to allow every sensor node to be aware of a routing path to the base station. This information is usually obtained by means of a discovery message broadcast by the base station. This message contains a hop count that is initially set to zero and is incremented at every hop by its recipients. On reception, every node stores the minimum distance value received from all of its neighbors. In this way, every node generates a routing table that contains its neighbors at distance $n-1$, n, and $n+1$, where n is the number of hops from the node to the base station. The result of this process is depicted in Fig. 3. The numbers represent the minimum distance of the node to the base station, the arrows indicate the direction towards the sink and the dashed lines indicate links to neighbor at the same distance. In the following, we will refer to nodes closer, equal, and further or $neigh_{n-1}$, $neigh_n$, and $neigh_{n+1}$ to mention these groups of neighbors.

The neighbor discovery process is essential to the rest of protocol. The reason is that the number and distribution of neighbors affects to both the privacy-protection level and the delivery of event messages to the base station as we will show in the following sections.



neighs(x)	distance
A	$n-1$
B	$n-1$
C	n
D	n
E	$n+1$
F	$n+1$

Fig. 3. Routing table of shaded node x

3.3 Data Transmission Properties

This section presents several key properties which are intended to limit the information gain of a local adversary during the transmission of data. Moreover, the fulfillment of these properties ensures the timely delivery of event data.

The protocol we are aiming at uses both real and fake messages. The source node, as well as any node that receives a real message, sends a real and a fake message, which should be indistinguishable to the intruder but not to the addressees. Property 2 aims to balance the amount of traffic being delivered by a node among its neighbors. By doing this, a local adversary cannot make a decision on which direction to follow based on the number of packets forwarded to neighboring nodes. While the paths of fake messages have relatively short length (this is a parameter of the solution), the path of real messages is intended to converge to the sink. This is established by Property 1: real messages must be transmitted to nodes closer to the base station with a high probability. These two properties together ensure that both real packets reach the base station and also that the flow of real messages is hidden by fake messages since they are indistinguishable. An additional technical property ensures that the transmission of every pair of messages is sent to *two* different nodes.

Property 1 (Convergence) *Let x be an arbitrary sensor node and BS be the base station. Also, let $neigh(n)$ be the set of immediate neighbors of a particular node n. Then we say that a path is convergent if x chooses the next node $x' \in neigh(x)$ such that:*

$$E(dist(x', BS)) < E(dist(x, BS))$$

where E is the mathematical expectation and $dist$ is the distance between two particular nodes.

Property 2 (Homogeneity) *Let x be an arbitrary sensor node and $neigh(n)$ be the set of immediate neighbors of a particular node n. We say that the transmissions of a node x hold the homogeneity property if:*

$$\forall y, z \in neigh(x) \quad Frec_m(x, y) \simeq Frec_m(x, z)$$

where $Frec_m(x, y)$ represents the number of messages (real and *fake) transmitted by node x to node y.*

Property 3 (Exclusion) *Let m and m' be a pair of messages and t be a particular transmission time. Let $send(m, x, y, t)$ denote that x sends to y the message m at time t. The exclusion property states that:*

$$\forall m, m', x, y, t \quad send(m, x, y, t) \wedge m \neq m' \Rightarrow \neg send(m', x, y, t)$$

3.4 Transmission Protocol

We devised a message transmission protocol that is consistent with the properties defined in Sec. 3.3. This protocol introduces insignificant computational and memory overhead because it is based on straightforward operations. More precisely, it requires a simple sorting operation and a pseudo-random number generator [12].

Since we send two messages, the combinations of two elements without repetitions from all neighbors in the routing table is an elegant and lightweight mechanism for the selection of neighbors that is consistent with the provisions of Property 3. Moreover, if the routing table is incrementally ordered in terms of the distance of its neighbors to the base station (i.e., $[neigh_{n-1}, neigh_n, neigh_{n+1}]$) we achieve that most of the resulting combinations have a closer or equally distant neighbor in the first position of the tuple. Therefore, Property 1 is satisfied because the real packet is transmitted always to the first neighbor. Also Property 2 holds provided that we randomly select any pair from all possible combinations.

In Algorithm 1 we describe the behavior of a node upon the reception of a packet. The algorithm uses as input the received packet, a data structure which contains the combinations of two neighbors once sorted, and a network parameter that controls the durability of fake packets in the network. Initially, the algorithm decides the random pair of neighbors to whom packets will be addressed (line 1). Subsequently, if the received packet is real then it is be forwarded to $neigh1$ while $neigh2$ receives a fake packet whose time-to-live is set to MAX_TTL (line 3). This parameter is dependent on the hearing range of the adversary and provides a trade-off between energy consumption and privacy. Also, note that the packets are sent in random order to prevent the adversary from trivially learning which is the real message. The described behavior is identical in case that the node, rather than being an intermediary, is a source node which signals the occurrence of an event in the field.

On the contrary, if the received packet is fake, the node first obtains the time-to-live (TTL) of the packet and decrements its value by one (line 5). This prevents fake messages from flooding the network. In case the new TTL is greater than zero, the node sends two fake messages with the current TTL value (line 7).

Algorithm 1 Transmission strategy

```
Input: $packet \leftarrow receive()$
Input: $combs \leftarrow combinations(sort(neighs), 2)$
Input: $MAX\_TTL$
$\{neigh1, neigh2\} \leftarrow select\_random(combs)$
if $isreal(packet)$ then
    $send\_random(neigh1, packet, neigh2, fake(MAX\_TTL))$
else
    $TTL \leftarrow get\_time\_to\_live(packet) - 1$
    if $TTL > 0$ then
        $send\_random(neigh1, fake(TTL), neigh2, fake(TTL))$
    end if
end if
```

Since we consider adversaries with a hearing range similar to an ordinary sensor nodes (i.e., the family $\mathcal{ADV}_1$), fake messages might be forwarded only once but still exceed the reach of the adversary.

4 Protocol Analysis

This section presents a detailed analysis on the potential limitations that might hinder the successful deployment of the HISP scheme in WSNs. First, we explore the impact of the network topology and the expected number of hops for real messages to reach the base station. Finally, we analyze the overhead introduced by our solution in terms of fake packet transmissions.

4.1 Bounding the Number of Neighbors

The distribution of real and fake messages is clearly impacted by the number of the neighbors in each of the groups of the routing table. In other words, Property 1 could be unsatisfied in case the number of neighbors in $neigh_{n-1}$ is significantly lower than the number of neighbors in $neigh_{n+1}$.

This problem is dependent on the topology of the network and the hearing range of the nodes. To have a clearer picture of how much this poses a real limitation to our protocol, we provide a numerical analysis on the number of $neigh_{n+1}$ that any sensor node can withstand without sacrificing any of the properties defined in Sec. 3.3.

Definition 2 *A real message converges to the base station if for any node in the route it traverses $\mathbb{P}_c > \mathbb{P}_f$, where $\mathbb{P}_c$ is the probability of transmitting the message to a node closer to the base station, and $\mathbb{P}_f$ is the probability of sending the message to a further node.*

In order to yield this property, several conditions must be met. In particular, let S be the total number of neighbors of an arbitrary node such that $S =$

$C+E+F$, where C, E, and F are the number of neighbors in $neigh_{n-1}$, $neigh_n$, and $neigh_{n+1}$, respectively. The theorem below gives a sufficient condition on C, F and S to ensure the desired property.

Theorem 1 *Real messages reach the base station if $F < \sqrt{2C(S-C)}$ for any sensor sensor in the route.*

Proof. We want to show that if $F < \sqrt{2C(S-C)}$ then $\mathbb{P}_c > \mathbb{P}_f$, such that $\mathbb{P}_c$ and $\mathbb{P}_f$ are the probabilities of sending a correct message to a node in $neigh_{n-1}$ and $neigh_{n+1}$, respectively.

The number of combinations of two neighbors where at least the first element belongs to $neigh_{n+1}$ is:

$$\binom{F}{2} = \frac{F(F-1)}{2}$$

while the number of combinations of two neighbors where the first element of the duple is a node in $neigh_{n-1}$ is:

$$\binom{C}{2} + C(E+F)$$

Consequently, the probability of selecting a closer neighbor is higher than the probability of selecting a further neighbor iff the number of combinations with a closer neighbor in the first position of the duple is larger than those with the first element being a further neighbor. Formally:

$$\mathbb{P}_c > \mathbb{P}_f \Leftrightarrow C(C-1) + 2C(E+F) > F(F-1)$$

In order to simplify the analysis we make some generalizations which are less restrictive but still provide a sufficient condition for the proof.

$$2C(E+F) > F^2 \Rightarrow C(C-1) + 2C(E+F) > F(F-1)$$

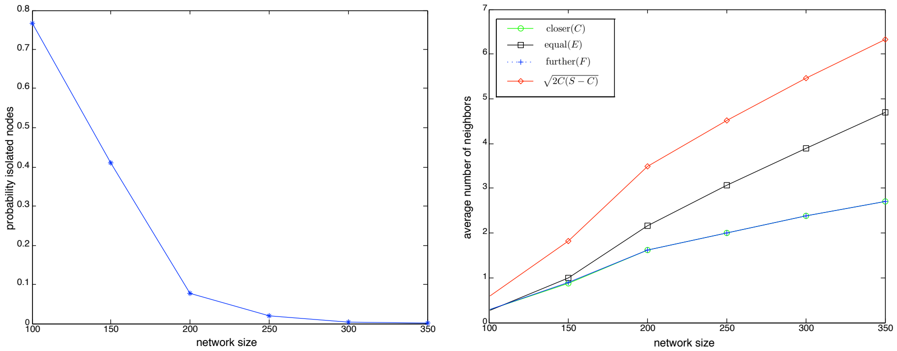
Provided that $C+E+F = S$, the previous equation can be expressed as:

$$F < \sqrt{2C(S-C)} \tag{1}$$

Therefore, we might say that if equation 1 is satisfied, then the following implication holds:

$$F < \sqrt{2C(S-C)} \Rightarrow \mathbb{P}_c > \mathbb{P}_f$$

Intuitively, the imposed restriction can be satisfied in networks deployed by hand following a particular topology (e.g., grid or mesh). Still, we deem necessary to validate the feasibility of our restriction in randomly deployed networks by means of experimental simulations. In particular, Fig. 4 depicts the average results over 50 repetitions of our network discovery protocol for various network sizes. We considered the following network parameters: (i) a square field area of



(a) Probability of isolated nodes

(b) Neighbors restriction

Fig. 4. Node connectivity in randomly deployed networks

side 1, (ii) the transmission radius of the nodes is set to 0.1, and (iii) networks ranging in size from 100 to 700 nodes randomly located. In Fig. 4a we show that the probability of isolated nodes drops significantly when the network size is over 200 nodes. Moreover, Fig. 4b presents the average number of neighbors closer, equal and further for any node in the network. In this figure we also show that the restriction imposed by Equation (1) on the maximum number of further neighbors is satisfied at all times.

Note that the results shown in Fig. 4b are average values and there might be some particular nodes not satisfying the restriction. However, this would only pose some additional delay unless there are network regions with a high concentration of nodes unable to fulfill the imposed condition. This issue might cause network packets to continuously move back and forth impeding their progress towards the base station. This is not the case when the node density is sufficient.

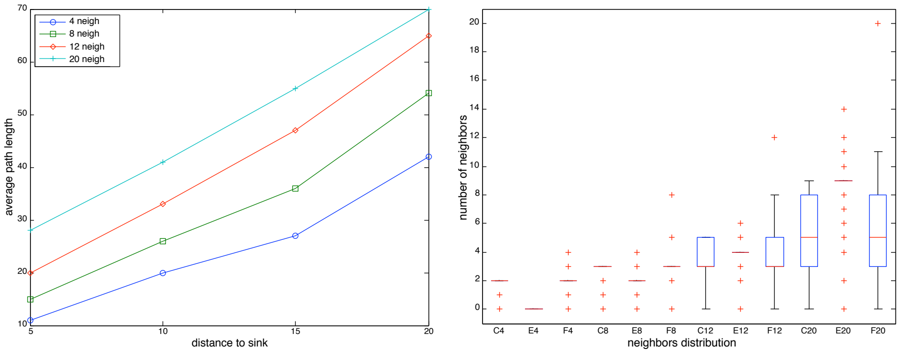
In general, we can state that when the number of nodes in a randomly deployed network is over 350 per square kilometer there is a high probability of full connectivity considering a transmission range of 100 meters. Also, in this case, the restriction on the number of neighbors is always satisfied.

4.2 Message Delivery Time

The probabilistic nature of our protocol introduces some uncertainty on the delivery of messages to the sink. This issue has some implications both on the reaction time of the network and the energy consumption of the nodes. Therefore, we provide some insights on the expected number of hops to reach the base station for a packet originated n hops away.

Let x_n be the expected number of hops for a packet originated at distance n. The proposed transmission protocol can be modeled by the following recurrence equation:

$$x_n = 1 + px_{n-1} + qx_n + rx_{n+1} \qquad (2)$$



(a) Expected number of hops

(b) Distribution of neighbors

Fig. 5. Protocol performace for various network configurations

This equation represents a biased random walk where, after sending the packet and increasing the number of hops by one, the packet will be forwarded to a neighbor. At each hop, we have a probability p of delivering the packet to a node closer to the base station, a probability q of staying at the same distance, and a probability r of moving in the opposite direction. Therefore, the average speed towards the base station is $p - r$.

In general, the above result is true for constant values of p and r but this is not always the case in sensor networks. The reason is that not all sensor nodes present the same distribution of neighbors. This is dependent on the hearing range of the nodes, the network topology and their location in the network. In Fig. 5 we present the performance of our protocol for WSNs deployed in a grid with equal transmission power for all nodes. We consider various configurations by increasing the transmission power, which in turn changes the connectivity of the network. On average, every node has 4, 8, 12 or 20 neighbors. Also, for every configuration we place the source at various distances from the base station: 5, 10, 15 and 20 hops. Several source nodes are selected for each distance and every single source node generates 500 data packets to be received by the base station.

The results show that the expected number of hops increases with the distance to the sink as well as with the connectivity of the nodes. As the number of neighbors available to a node increases, the more difficult it is for the adversary to make a decision on which of the recipients is actually closer to the base station. However, a significant increase in the number of neighbors has also implications on the delivery time because as the transmission range grows, more nodes will be in the equal list of the node. This issue is shown in Fig. 5b, where we provide a box-plot representation of the number of neighbors closer (C), equal (E), and further (F) for the simulated network configurations. For example, C_4 indicates closer neighbors in the $4neigh$ network configuration.

Additionally, note from Fig. 5a that, for all the configurations, the average speed of the packets decreases when they are close to the sink. Consider, for example, the 4*neigh* configuration. When the distance to the sink is 5, the expected delivery time is 11, while a packet at distance 20 will be delivered after 42 hops. This means that the time difference from distance 20 to 5 is 31 and thus, the average speed is $15/31 = 0.484$. However, in the proximities of the base station (from distance 5 to 0) the speed drops to $5/11 = 0.454$. The reason is that the distribution of neighbors for nodes around the base station is different from distant nodes. More precisely, the nodes in close vicinity of the base station have very few nodes in the closer list but the number of nodes at the same distance or further away is high. The imbalance between the lists of neighbors grows with the transmission range of the nodes, being more significant for the 20*neigh* configuration. In this case, the speed drops from 0.358 to 0.179 in the vicinity of the sink.

4.3 Fake Traffic Overhead

The injection of fake traffic is a fundamental feature of the HISP protocol since it covers the flow of real messages. However, the amount of fake traffic must be kept as low as possible in order to extend the lifetime of the nodes. To control the propagation of fake messages, HISP defines a system parameter, MAX_TTL, which depends on the hearing range of the adversary.

Instead of transmitting fake messages at regular intervals, which would provide the best privacy protection but would deplete sensors' batteries rapidly, the devised protocol injects fake traffic triggered by the presence of real messages. The scope of fake messages is conditioned by the eavesdropping capabilities of the adversary. Thus, if the adversary under consideration belongs to the $\mathcal{ADV}_0$ family, the value of the system parameter can be set to zero, while if the adversary is a global observer, this value is to be as large as the diameter of the network. In the latter case, the energy cost would be similar to transmitting at regular intervals with the difference that fake messages will remain in the network only in the presence of events.

In Fig. 6 we illustrate the fake traffic overhead imposed by HISP for different values of the MAX_TTL parameter in the various network configurations considered. More precisely, we show the ratio of fake over real messages that is introduced to balance the transmissions in a band around the real path. When MAX_TTL is set to zero the ratio is 1 because every real packet is transmitted in conjunction with a fake packet, which is no longer propagated. As the time-to-live grows, the ratio increase is on the order of $\mathcal{O}(2^{n+1})$ where n is the hearing range of the adversary. In any case, given the adversarial model considered in this work the overhead imposed by this approach is moderate.

The overhead imposed by fake messages might be reduced by half if we introduce a slight modification. Instead of sending two packets upon the reception of traffic, we might send a single packet with two identifiers. In this way, and assuming that the identifiers are hidden to potential observers, the two recipients receive the packet and continue with the forwarding process. The first identifier indicates the real recipient and the second indicates the fake recipient. This

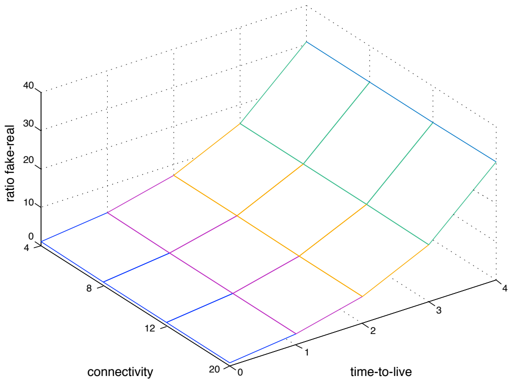


Fig. 6. Overhead of fake messages

improvement is possible due to the broadcast nature of wireless transmissions, which allows all the neighbors from a node to overhear its messages.

Finally, as shown in Fig. 6, the ratio is not affected by different network topologies. This is not surprising since the number of transmissions performed by the protocol is independent of the connectivity of the sensor nodes.

5 Discussion

The devised receiver-location privacy mechanism is aimed to protect from local adversaries capable of performing various traffic analysis attacks. The strategy of the adversary is to repeatedly move closer to the base station by observing the transmissions along the communication path. Starting at any point of the network he eventually finds a data sender. From this location, the adversary attempts to determine the direction to the base station by observing the communications of the data sender and its neighbors.

Firstly, the adversary might perform a time-correlation attack and move in the direction of the neighbor forwarding the first message transmitted by the data sender. Given the features of our solution several cases may occur depending on whether the packet is real or fake. If the packet is real, the adversary is highly likely to reduce by one his distance to the base station. However, this is not necessarily the case because real traffic might be also forwarded in other directions. Moreover, the probability of following a real packet is lower than the probability of following a fake packet. The reason is that, as real messages move, they generate pairs of messages, one real and one fake, while fake messages trigger the transmission of pairs of fake messages. Also, note that the adversary can only be certain of whether he made the right choice when he follows a fake packet that is no longer propagated. In fact, this issue provides the adversary with no information about the direction to the base station because fake messages are forwarded in any direction.

Alternatively, the adversary might choose to perform a sufficient number of observations before making a decision on the next move. In that case, the adversary will move towards the neighbor with the higher transmission rate. To reduce the success of this strategy, the HISP protocol makes nodes to evenly distribute messages among their neighbors, thus locally homogenizing the number of packets being observed by a potential adversary. Again, the adversary cannot determine which packets are real and which are fake unless he observes a node that after receiving a packet does not forward it. This implies that he is at the edge of the band of fake messages surrounding the path of real data. Being able to precisely determine the limits of the band of fake messages could provide the adversary with information on how to reach the base station. However, the number and behavior of events being reported by the sensor nodes may be extremely dynamic, which hinders the process of bounding the aforementioned band. Moreover, real packets are sent following a random walk which causes the band to be rather arbitrary. Consequently, even if the adversary was capable of delimiting the edges of the band at some point, this information does not necessarily lead him to the base station.

Defining a sound strategy is rather difficult even when the adversary is fully aware of the protection mechanism in place. The highly dynamic nature of events in the field results in irregular communication flows which greatly complicates the definition of the most effective strategy to reduce the distance to the target.

6 Related Work

This section compares the HISP scheme with previous solutions developed to protect both source- and receiver-location privacy in WSNs.

6.1 Source-Location Privacy

The source-location privacy problem was introduced in [13]. This work proposes the Phantom Routing protocol to counter adversaries tracing back packets to the source node. This protocol sends every message on a random or directed walk to a phantom source, which finally forwards the packet to the sink using a flooding-based or a single-path routing. In this way, every packet appears to be originated from a different source. This protocol presents several drawbacks specially in the walking phase, which tends to stay close to the original source. New solutions [14] concentrated on guiding this walking phase, while in other solutions [15] the phantom sources are placed in a ring where the messages are mixed with fake traffic.

To hide the presence of events to adversaries with a global hearing range, [16] makes all sensors to transmit messages at a fixed rate regardless of the existence of real events. This provides perfect privacy but the cost is unacceptable for battery-powered devices. Several authors concentrated on reducing the energy implications of this approach. In [17] a filtering scheme is proposed to reduce the amount of fake traffic at various network locations. Also, some statistical approaches [18, 19] were devised to modify the real and fake transmission frequency without arousing suspicion on the attacker.

In general, the presented solutions are based the randomization of the routes and the injection of fake traffic, which misleads the adversary or hides the presence of real packets.

6.2 Receiver-Location Privacy

Receiver-location privacy was originally investigated in [4, 6] where various load balancing techniques were designed. They proposed a multi-parent routing technique that randomly selects the next hop in the path from all available nodes closer to the sink instead of sending packets always to the same node. To further complicate traffic analysis, this technique is complemented with random walks in any direction and the injection of fake packets with a given probability distribution.

Other approaches [6, 8] concentrated on the creation of hot-spots, which are areas with high volumes of fake traffic that aim to attract adversaries performing rate-monitoring attacks. The authors in [7] propose to make all nodes transmit the same number of packets so that the traffic rate is homogenized regardless of the proximity to the base station. This strategy provides the best protection but it also imposes the highest energy requirements. Besides, in [20] the base station mimics the behavior of ordinary nodes (i.e. forward some of the packets it receives) to enhance its privacy. Additionally, the authors propose to move the base station to a safer location based on its own measured privacy level.

The work closest to ours is [5] because it makes use of a path diversification and fake packet injection. In this work, the authors propose to forward packets to nodes closer to the base station with some probability $1 - p_f > 0.5$ and to nodes further with probability $p_f < 0.5$. These probabilities ensure that packets eventually reach the sink but, after a sufficient number of observations, the adversary is able to deduce the direction to the sink. To reduce this problem, fake packets are injected in the opposite direction based on a certain probability p_{fake} only after the reception of a real packet. In general, the adversary cannot distinguish real from fake traffic, however, if he observes that a node that receives a packet does not forward it, he can be certain that this is fake packet whose time-to-live has expired. Since fake packets are only sent to further nodes, the must move in the opposite direction to find the sink.

We propose a packet transmission protocol also based on random route generation and fake packet transmissions that is capable of circumventing the problems presented by the previous works.

7 Conclusions

This work presents a new receiver-location privacy scheme for WSNs called HISP. The proposed solution is based on the injection of fake traffic to hide the flow of real traffic which is sent to the base station using a random walk. The goal is to probabilistically homogenize the overall number of packets that a node distributes among its neighbors. More precisely, the devised protocol preserves three critical properties (i.e., convergence, homogeneity, and exclusion), which ensure the delivery of event data to the base station as well as the robustness of receiver-location privacy against local adversaries.

The feasibility of the HISP protocol has been validated both analytically and experimentally. In particular, we have analyzed the impact of the connectivity of the network on the convergence of the packets to the base station and the privacy protection level. Also, we have investigated on the expected convergence time of packets in order to gain insights on the potential applicability of our solution to time-critical applications. Finally, we have explored the overhead imposed in terms of fake traffic injection for adversaries with different eavesdropping capabilities.

As future work we consider investigating new ways of reducing the fake traffic overhead required to protect against adversaries with a large hearing range. Also, we will explore the robustness of our scheme against more skilled adversaries. To that end, we first need to define a set of strategies based on the knowledge of the adversary about the network and the privacy protection protocol in use. The adversary may change his strategy depending on the context of the network. Countering such powerful adversaries may also require the development of new and more sophisticated protection mechanisms not considered so far.

Acknowledgements

This work has been partially funded by the European Commission through the FP7 project NESSoS (FP7 256890) and the Spanish Ministry of Science and Innovation through the ARES (CSD2007-00004) and SPRINT (TIN2009-09237) projects. SPRINT is co-financed by FEDER (European Regional Development Fund). The first author is supported by the Spanish Ministry of Education through the F.P.U. Program. Also, special thanks to Martín Ochoa for his valuable comments in preliminary versions of this work.

References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Computer Networks **52**(12) (2008) 2292 – 2330
2. Walters, J., Liang, Z., Shi, W., Chaudhary, V.: Wireless Sensor Network Security: A Survey. In: Security in Distributed, Grid, and Pervasive Computing. Auerbach Pub (2007) 367–409
3. Pai, S., Bermudez, S., Wicker, S., Meingast, M., Roosta, T., Sastry, S., Mulligan, D.: Transactional Confidentiality in Sensor Networks. IEEE Security & Privacy **6**(4) (July-Aug. 2008) 28–35
4. Deng, J., Han, R., Mishra, S.: Countermeasures Against Traffic Analysis Attacks in Wireless Sensor Networks. In: 1st International Conference on Security and

Privacy for Emerging Areas in Communications Networks (SECURECOMM '05). (2005) 113–126

5. Jian, Y., Chen, S., Zhang, Z., Zhang, L.: Protecting receiver-location privacy in wireless sensor networks. In: 26th IEEE International Conference on Computer Communications (INFOCOM 2007). (2007) 1955–1963
6. Deng, J., Han, R., Mishra, S.: Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks. Pervasive and Mobile Computing **2**(2) (2006) 159–186
7. Ying, B., Gallardo, J.R., Makrakis, D., Mouftah, H.T.: Concealing of the Sink Location in WSNs by Artificially Homogenizing Traffic Intensity. In: 1st International Workshop on Security in Computers, Networking and Communications. (2011) 1005–1010
8. Chang, S., Qi, Y., Zhu, H., Dong, M., Ota, K.: Maelstrom: Receiver-Location Preserving in Wireless Sensor Networks. In: Wireless Algorithms, Systems, and Applications. Volume 6843 of LNCS. Springer (2011) 190–201
9. Zhang, J., Varadharajan, V.: Wireless sensor network key management survey and taxonomy. J. Netw. Comput. Appl. **33**(2) (2010) 63 – 75
10. Gómez, C., Paradells, J., Caballero, J.E.: Sensors Everywhere: Wireless Network Technologies and Solutions. Fundación Vodafone España (2010) ISBN 978-84-934740-5-8.
11. Misra, S., Xue, G.: Efficient anonymity schemes for clustered wireless sensor networks. Int. J. Sen. Netw. **1**(1/2) (2006) 50–63
12. Latif, R., Hussain, M.: Hardware-Based Random Number Generation in Wireless Sensor Networks. In: Advances in Information Security and Assurance. Volume 5576 of LNCS. Springer (2009) 732–740
13. Ozturk, C., Zhang, Y., Trappe, W.: Source-Location Privacy in Energy-Constrained Sensor Network Routing. In: 2nd ACM workshop on Security of ad hoc and sensor networks (SASN '04). (2004) 88–93
14. Wang, H., Sheng, B., Li, Q.: Privacy-aware routing in sensor networks. Computer Networks **53**(9) (2009) 1512–1529
15. Li, Y., Ren, J.: Preserving Source-Location Privacy in Wireless Sensor Networks. In: 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '09). (2009) 493–501
16. Mehta, K., Liu, D., Wright, M.: Location Privacy in Sensor Networks Against a Global Eavesdropper. In: IEEE International Conference on Network Protocols (ICNP '07). (2007) 314–323
17. Yang, Y., Shao, M., Zhu, S., Urgaonkar, B., Cao, G.: Towards Event Source Unobservability with Minimum Network Traffic in Sensor Networks. In: 1st ACM conference on Wireless network security (WiSec '08). (2008) 77–88
18. Shao, M., Yang, Y., Zhu, S., Cao, G.: Towards Statistically Strong Source Anonymity for Sensor Networks. In: 27th IEEE Conference on Computer Communications (INFOCOM 2008). (2008) 466–474
19. Alomair, B., Clark, A., Cuellar, J., Poovendran, R.: Statistical Framework for Source Anonymity in Sensor Networks. In: IEEE Global Telecommunications Conference (GLOBECOM 2010). (2010) 1 –6
20. Acharya, U., Younis, M.: Increasing base-station anonymity in wireless sensor networks. Ad Hoc Networks **8**(8) (2010) 791–809