Abstract: Wireless Sensor Networks (WSNs) are commonly used for animal tracking. Over the years, a significant number of studies have been presented for monitoring moving targets through WSNs. At the same time the location/position information of a target should be available only to authorized entities, e.g., Animal Protection Centers, thus this information should be kept private. The iHIDE is a location privacy mechanism that uses a non-geographical routing scheme for packet delivery over WSN. In this paper we elaborate on that scheme by introducing a routing plan algorithm. We enhance iHIDE by adopting epidemical data dissemination as an enforcing privacy technique. We evaluate through simulations the scheme against other commonly used location privacy overlays in terms of network overhead and safety period and quantify the benefits stemming for its adoption.

Keywords: Wireless sensor network, Location privacy, Panda Hunter game, Epidemical information dissemination.

1. Introduction

Wireless Sensor Networks (WSNs) are collections of autonomous nodes that are commonly used to monitor physical and environment phenomena. Over the last years the use of this technology for animal tracking (hereinafter referred to as data source) becomes significantly important. For instance, scientists/researchers observe animal behavior, identify population exchanges and prevent animal accidents. At the same time, the location information of animals should be kept private from unauthorized use. Several approaches are proposed aiming to address this issue. A distinctive example of a data source protection problem is the so called “Panda Hunter Game” introduced in [1]. In this problem a large number of sensors are deployed in a panda habitat. The sensors are able to track panda movements and report them back to an Animal Protection Center (APC). Hunters located in the same area are trying to detect the data source by tracking the packages exchanged among the sensor nodes in the WSN. The overall goal is to provide a privacy mechanism that constantly informs the APC about the location information of a panda in the habitat. Once detect the presence of a data source, the mechanism should ensure the prompt delivery of this location information packet to the sink node. At the same time, the mechanism should enable countermeasures in order to disorientate the hunter from tracking down the originator of the packet and thus the panda. However these countermeasures should not significantly affect the overall performance of the WSN in terms of communication overhead and energy consumption. An overview of the existing data source protection mechanisms is thoroughly presented in [2]. In the Flooding approach in [1] the idea is to broadcast the data of a sensor node to all neighboring nodes. Since all the nodes participate on flooding information, it is difficult for a hunter to identify the data source. An extension of the standard Flooding is the Probabilistic Flooding [1] aiming to reduce the significant energy consumption of the previous scheme. In this extension only a part of WSN performs data forwarding, while the other part discards the received messages. The retransmission of data packages is based on a forwarding probability. The Phantom Routing [3] combines routing techniques in order to deliver the packet to the sink, e.g., APC. Specifically the packet delivery from the data source to the sink is divided in two routines namely (i) the random walk routine and (ii) single path/flooding routine. In the former routine the packet performs either a pure or directed walk to a fake source according to a random number of hops. In the latter routine a packet is delivered to the sink either through probabilistic flooding or through a single path routing. Another approach is the Greedy Random Walk presented in [3]. The mechanism in this case firstly initializes a random path with a given number of hops from the sink. Then, packets are randomly forwarded from the data source till they reach specific WSN nodes called receptors. Receptors sequentially forward the packets to the sink through a pre-established path initially set by the sink. Alternative approaches introduce the use of fake data sources [3, 20]. The idea is the selection of one or more WSN nodes to impersonate an actual data source and trigger randomly packets. The more fake data sources one uses the better protection this mechanism offers. However, this mechanism aggravates overall network performance in terms of power and communication overhead. Other techniques [1, 4, 5, 19] aim at protecting the location of the data source and introduce fake data packets to confuse the hunter’s traffic patterns techniques. However all of them are facing significant network performance issues as mentioned above.

Finally, the strategy proposed in [21] attempts to prevent the hunter from inference interconnections between WSN nodes by introducing a delay between data send from one node and data received from another node. This method might confuse the hunter but causes significant delays on packet delivery. The iHIDE scheme [7] refers to a location privacy overlay that deploys and applies non-geographic routing plan through the use of multiple rings in WSN connected to a central bus. Packets transmitted from the data source are routed through this overlay towards the sink node or an APC. The contribution of this paper is:

- enhancement of the iHIDE by introducing a routing generation algorithm,
- enforcement of the scheme’s privacy through the
adoption of epidemical information dissemination as a fake data source mechanism and
• evaluation of our scheme against other location privacy techniques in terms of network overhead and privacy metrics.

The structure of the paper has as follows: Section 2 reports on the main principles of iHIDE and elaborates on the generation of the routing plan. Section 3 emphasizes on the security threats and further enhances our mechanism by introducing the use of fake sources and epidemic models. In Section 4 we evaluate our scheme against other data source protection mechanisms while Section 5 concludes the paper with future research.

2. The iHIDE Framework

2.1 Rationale

iHIDE considers static WSN nodes and mobile data sources. The core components of iHIDE are:

1. Sensing Nodes (SEN) that track the presence of a pan-da in a given coverage area,
2. Bus Nodes (BUN), and
3. a unique Sink Node (SIN) that is responsible for collection and processing of packets originated from data source.

![Figure 1. The iHIDE Architecture](image)

The SIN is usually an APC (back-end) system with relatively high computation capabilities. Each iHIDE overlay deployment consists of one Bus (BS) and multiple Rings (RNs) as shown in Fig. 1. Each RN consists of SENs and the BS consists of BUNs. The upper end of the BS is always the SIN. The RNs are attached to one BS forming a virtual hub with multiple cycles attached to it. On the connecting points, WSN nodes act both as SEN and BUN. Over this deployment a Routing Scheme (RS) is applied that interconnects the WSN nodes and forwards the data source packets to the SIN. Once a packet has received from the RN, the BUNs forward it through the BS to SIN. On the way up the intermediate BUNs based on a probabilistic model could re-route this packet to their RNs. For more information, the interested reader could refer to [7].

2.2 Routing in iHIDE

2.2.1 Routing Scheme Creation

In this section we introduce the mechanism that generates a RS. Several aspects should be taken into consideration such as WSN topology, coverage area and the ground rules of the mechanism as described in Section 2.1. During the initialization phase we determine the WSN topology and the connectivity between WSN nodes. Several existing protocols can be used for this purpose as well for RS distribution to the WSN nodes, for instance the protocols proposed in [8-12]. As soon as we have the WSN topology information, we can apply the iHIDE overlay. For the RS generation procedure we use the Dijkstra shortest path algorithm. We refer to BusL and RingL, the target Bus and average Ring size, respectively. The steps are depicted on Table 1.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Select a sensor node that is able to communicate directly to the SIN</td>
</tr>
<tr>
<td>2</td>
<td>Set sensor node of step 1 as BUN</td>
</tr>
<tr>
<td>3</td>
<td>Find the neighborhood of BUN(i.e. sensors located within its transmission range). Add these sensor nodes to group called NH0</td>
</tr>
<tr>
<td>4</td>
<td>Randomly select a sensor node from NH0 and set it as BUN</td>
</tr>
</tbody>
</table>
| 5    | While BS size is less than BusL: 
  a. Find the neighborhood NH1 of BUN1 
  b. Select randomly a sensor node from NH1 and set it as BUN 
  c. Set NH = NH0 ∪ NH1, where 1 = 0 |
| 6    | Create a graph containing all sensor nodes of the network |
| 7    | Locate pair of nodes that are able to directly communicate and connected them via an arc |
| 8    | Use a random number generator and assign to each arc a weight |
| 9    | For each node in BS (BUN): 
  a. Apply dijkstra to find the shortest path from BUN to all other nodes 
  b. Select a Random node Ntemp such that the hop distance between BUN and Ntemp is ~RingL/2 
  c. Store the path of step 9b (it will be the half of an iHIDE RN) 
  d. Shuffle the weights of the graph 
  e. Run dijkstra to find the shortest path from Ntemp to all other nodes 
  f. Append to the path of step 9c the path from Ntemp to BUN 
  g. The path corresponds to the RN of BUN |
| 10   | If there are more than one SIN repeat steps 1 to 9 |

The outcome of the RS generation procedure is a Routing Table (RT) per iHIDE component. Each RT record contains the following information:

- InSenNode: the packet originator node,
- OutSenNode: the corresponding destination node,
- pN: probability to forward a packet.
According to the iHIDE scheme the SENs contains one forwarding entry per RN (one SEN might be part of multiple RNs), whist for BUNs it contains exactly one record for the BN. Note that since BUN is also a SEN both RT apply for it. We further describe the use of $pN$ in Section 2.2.2. Upon RS creation, the RTs are distributed through out the network by using a commonly used protocol like directed diffusion [10] or Leach [24], or even better a secure enhancement of them [23].

2.2.2 Probabilistic Information Dissemination

Once a SEN$_{origin}$ detects the presence of a panda, it creates a packet containing the location information of the panda. In turn, it encrypts this content using the SIN’s public key and signs the encrypted content with its private key. Sequentially, it sends the packet to the forwarding RN node, which is a SEN in its RN. Note that, if SEN$_{origin}$ participates in multiple RNs, it randomly selects one RN to forward the packet. The packet circulates among the SENs of the RN until it reaches the BUN$_{origin}$. The BUN$_{origin}$ stores the packet to its local repository and places it into the BS after a randomly selected time period. Once the following BUN in the BS receives the packet, it forwards the packet to the next BUN node and reroutes a copy of that packet to its RN with some probability $pN$. The copied packet is disposed from the RN through the use of a Time To Live (TTL) flag. Note that the BUN$_{origin}$ reroutes the packet into its RN with $pN$ to ensure a packet circulation around the RN and disorientate the hunter.

3. Enhancement of the iHIDE

3.1. Injection of fake sources

In order to enhance the privacy mechanism we adopt a fake packet dissemination policy. The main idea is to introduce fake data sources inside the WSN that randomly trigger a data source packet. By adopting such a policy we can further disorientate the hunter since the latter is not able to distinguish a fake from a real packet due to the applied content’s encryption. Nonetheless, the extra transferred packets on the WSN will significantly increase the energy consumption and the communication overhead. Therefore, our goal is to find a way to use fake sources but at the same time to limit the packet retransmissions. To this end, we adopt an epidemic-like packet dissemination protocol that ensures the minimum number of retransmissions for a finite time period. Note that the tradeoff between the privacy offered and the system performance is an issue that comes up when trying to solve privacy issues [22].

3.2. Epidemical dissemination of fake sources

A WSN node can store a piece of information in order to disseminate it to nearby nodes under certain time-space constraints. The epidemic-based spreading model [17] adopts a simplistic routing scheme where dissemination proceeds on a local basis and does not require central coordination or complex routing schemes. In addition all pieces of information are transmitted in the form of local (1-hop) broadcast and a node opportunistically forwards information to neighbors. According to [18], an individual (corresponding to a WSN node) can be in three states:

- Infected: an individual is infected with epidemic – corresponding to a valid piece of information,
- Susceptible: an individual is prone to be infected and
- Removed: an individual is immune as it has recovered from the disease.

This kind of model is usually referred to as Susceptible-Infected-Recovered (SIR). A simplified version of that model is the Susceptible-Infected-Susceptible (SIS) model in which an individual can exist in only susceptible and infected states. This means that an individual never gets immune after its contact with the epidemic. In the Susceptible-Infected (SI) model an individual never turns susceptible if infected once. We can abstract the epidemiological analogies if the SIS model for setting up of a probabilistic approach for disseminating pieces of information in a WSN. We adopt the SIS model in which WSN nodes that carry a piece of information (infected nodes) can disseminate it to the neighboring nodes that do not have any information (susceptible nodes). After a period of time, infected nodes may recover from the epidemic (data) and then transit to the susceptible state with a recovery rate $\delta > 0$. In that state they can get infected again with an infection rate $0 < p_{inf} < 1$ thus, in the limit, any individual perpetually moves between the two states: Susceptible – Infected.

One might argue that the epidemic model can be used as a location privacy mechanism itself. However even though the epidemic model disorientates the hunter, once triggered from fake sources it has the opposite results in case it is initiated from the real data source. In the latter case, the transmitted packets follows a directional root starting from the data source thus the hunter could back-track this path (similar to shortest path) and reach the panda much faster. Hence, in our approach the epidemical model is used as a disorientation mechanism triggering from the fake data sources. For the rest on the paper we call the application of Epidemic models to iHIDE as iHIDE$^E_{\text{epidemic}}$.

Every sensor node can initiate a fake packet. We denote as $P_r$ the probability that a sensor node triggers a fake packet at any some time instance. We can regulate the number of the fake packets that travel in the network. $P_r$ varies based on the WSN establishment. For WSN establishments, where the number of tracked entities is small and thus the corresponding iHIDE traffic is low, $P_r$ can be higher. On the other hand, in relatively dense WSNs with many data sources, $P_r$ can be lower since there would be no need for frequent fake packet dissemination.

4. Performance & Comparative Assessment

4.1. Simulation environment

During our first set of experiments presented in [7] we evaluated the basic principles of iHIDE using a fixed WSN with predefined iHIDE RS applied. However, in an actual iHIDE overlay the WSN and the corresponding RS varies as we discuss in Section 2.2.1. Hence the simulation code was rewritten and, for the second set of experiments, the simulation environment was developed in Java®.

For the simulations we used a 200x200 unit wide area for WSN. In each experiment, 500 WSN sensors are randomly placed on that area. We consider a single panda moving according to the Random Waypoint (RWP) [13] model which is a commonly used synthetic model for mobility...
traces. Each experiment’s duration is 24 hours (or 86400 ticks) and can be interrupted if the hunter locates the panda.

We also consider a hunter that is aware of the physical location of the WSN nodes and employs data traffic analyzer equipment which is able to intercept packets transmitted between WSN nodes. Her strategy is to maintain a position close to a WSN node until she intercepts a packet and checks over its content. Then, she needs 15 ticks to move towards the direction of the packet originator. Furthermore, she maintains her position for 60 ticks and randomly moves to another in case she doesn’t overhear a packet. On our experiments, the hunter starts from a random point inside the WSN.

4.2. Models

We implemented five routing schemes namely

(i) Phantom Routing [3].
(ii) Epidemic, [17].
(iii) Flooding, [1].
(iv) iHIDE
(v) iHIDE-Epidemic

Table 2 depicts special parameters that we set in the simulations.

<table>
<thead>
<tr>
<th>Overlay</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phantom Routing</td>
<td>The radius of the Flooding area was 100 units around the sink.</td>
</tr>
<tr>
<td>Epidemic</td>
<td>The cure probability was 3% and the infection probability $p_{inf}$ was variable depending of the experiment.</td>
</tr>
<tr>
<td>Flooding</td>
<td>No specific parameter. Each node sends each new received packet to all neighbors.</td>
</tr>
<tr>
<td>iHIDE</td>
<td>The length of the bus was 100, the length of each ring was 50 and the probability for rerouting packets thru the RN was $pN = 10%$.</td>
</tr>
<tr>
<td>iHIDE-Epidemic</td>
<td>See iHIDE and Epidemic. $P_e$ is set to 10%</td>
</tr>
</tbody>
</table>

In the Phantom Routing scheme the privacy level is related to the size of the area where the flooding phase is applied. In case that the flooding area is small then the hunter can infer more information during the random walk phase, since the packet follows a specific path for a longer period of time. On the contrary, wider Flooding areas can perform better in terms of privacy and disorientate the hunter. In our simulations we assume that the Flooding area covers the 37.5% of the total WSN area. Hence the packet performed a random walk till it reached the flooding area and, sequentially, it was flooded to the sink. Note that we selected this area size because it was proven being an equivalent scenario in terms of network overhead with the other overlays we wanted to evaluate.

Regarding iHIDE our goal is to apply an iHIDE establishment taking into account, among others, the WSN topology. The ideal iHIDE establishment would situate nodes in such a way that, with a relatively small number of hops, comparing to the overall WSN size a packet would reach the sink fast. Imagine an iHIDE establishment that contains one BN and two extremely large RNs spreading over the WSN (note that the iHIDE overlay is non-geographical). In this case a packet will probably circulate around the network before it reaches the BN and forwarded to the sink. Hence our goal is to focus on small rings. Another objective is to minimize the number of sensor nodes that participate in multiple rings and thus reducing the packets that each WSN node exchanges. During our simulations, upon testing multiple multiple establishments using several combinations for the BS and RN sizes, we concluded that the following values can establish a iHIDE overlay closer to specifications: SB size = 100 and average RN size = 50.

Finally, regarding the epidemic routing scheme we have to select a curing probability that it will be small to cure nodes thus preventing rerouting packets but, in parallel, should not dramatically slow down the infection process. In the simulations the curing probability was set to 0.01.

4.3. Performance metric: Network overhead

Our first objective is to measure the network overhead when location privacy overlays are applied to the WSN. Hence, we measured the number of transmitted packets when iHIDE, Phantom Routing, Epidemic and iHIDE-Epidemic were applied to the network. Especially for the latter case we measured the system performance when the $p_{inf}$ of the epidemic SIS model was 1, 0.8, 0.5, and 0.1, respectively; $\delta = 0.01$ to ensure epidemic dissemination [17]. Fig. 2 shows our findings on the average number of packages; we executed 1000 experiments per case.

We observe that the Flooding overlay creates significantly higher network overhead than the other schemes. The only exception is iHIDE-Epidemic when $p_{inf} = 1$ is applied. Should we compare the Phantom Routing, iHide and iHIDE-Epidemic overlays, we can conclude that both iHide and iHIDE-Epidemic perform better than Phantom Routing in all scenarios. However, we have to mark that when $p_{inf}$ is greater than 0.5, the number of transmitted packets increases significant and thus the communication cost is very high. From an energy consumption perspective we can relay the number of exchanged packets with power consumption models like [14-16] and estimate the required energy.

4.4. Performance metric: Safety period

Our next goal is to calculate the Safety Period accomplished by the privacy overlays. This period refers to the time till the panda was captured by the hunter.

Fig. 3 depicts the Safety Period achieved by the Flooding, Phantom Routing, iHIDE and iHIDE-Epidemic overlays. As we can observe the Safety Period achieved by iHide and iHIDE-Epidemic is higher than all the other overlays. The only
exception is the Flooding algorithm but this is achieved with a significant network overhead cost as discussed in Section 4.2. Hence, iHIDE and iHIDE\textsubscript{Epidemic} receive the higher rank should we consider the privacy the corresponding network and power overheads.

Since we have concluded that iHIDE\textsubscript{Epidemic} performs relatively better we decided to evaluate its performance in terms of privacy for different values of the $p_{inf}$. Fig. 4 depicts that despite the deviation of $p_{inf}$ the Safety Period is not significantly affected. As we mentioned in Section 4.2, the use of the epidemic SIS model with $p_{inf} = 0.5, 0.8$ and 1 is not efficient in terms of network overhead. Hence, we can conclude that both in terms of privacy and performance, the use of high $p_{inf}$ is not efficient.

By taking a closer look on Fig. 3 and Fig. 4, we will note that the privacy level of iHIDE is enhanced by 25% when the $p_{inf}$ is 0.1, while, by setting $p_{inf} = 1.0$, the privacy level is enhanced by an additional factor of 30%. In other words, although the $p_{inf}$ is ten times greater, the privacy level is increased just by 30%. This happens because the epidemic routing scheme can guarantee the fast information propagation (taking in account the cure probability). Hence even though $p_{inf}$ can be low (for instance, $p_{inf} = 0.1$), the entire WSN can exchange fake packets fast enough comparing to the hunters’ movements.

Hitherto, in our experiments we considered the presence of a single hunter. In order to enforce hunter’s capabilities we can consider the existence of multiple hunters on the same area. Our objective is to evaluate the effectiveness of the iHide, Epidemic and iHIDE\textsubscript{Epidemic} overlays against them. We conducted 1000 experiments per overlay while one, two and three hunters overheard the network. In Fig. 5 we can observe that the use of an Epidemic model for real sources packet transmission performs much worse than iHIDE, whereas the use of iHIDE\textsubscript{Epidemic} increases the privacy level of iHide on average for 23%. As discussed in Section 3.2, this proves our initial intention to use Epidemic model as hunter’s disorientation strategy rather than a location privacy policy. Even though, from a network overhead point of view, as depicted in Fig. 2, the iHIDE\textsubscript{Epidemic} requires 3 times more packets than the iHIDE, it is still less than Phantom Routing overhead.

5. Conclusions and Future work

In this paper we elaborated on the iHIDE privacy mechanism by defining a routing plan generation algorithm and further enhanced iHIDE by introducing the use of fake sources and epidemic models. The evaluation results showed that iHIDE\textsubscript{Epidemic} enforces the privacy level of our original technique while keeping communication overhead lower than other proposed overlays, even when more than one hunters are trying to locate the panda. Another interesting conclusion emerging from our experiments is that even though we manage to increase the Safety Period, the hunter eventually manages to locate the panda by observing the packets exchanged through the WSN. Hence, we have to update our RS before she inferences panda’s location. At the same time we do not want to change the RS frequently since it causes additional communication overhead. Hence, a decision model should be defined that will evaluate our risk and decide when to take action.

References


