Understanding Source Location Privacy Protocols in Sensor Networks via Perturbation of Time Series

WARWICK

Matthew Bradbury and Arshad Jhumka

InfoCom 2017

Outline

1. Introduction
2. Related work
3. Privacy and Attacker Models
4. How to measure SLP: Information loss and privacy loss
5. How to provide SLP: Competing paths
6. Case study

What is a Wireless Sensor Network?

A wireless sensor network (WSN) is a collection of computing devices called nodes, they have:

- a short range wireless radio
- an array of sensors such as light, heat and humidity
- a simple low powered CPU
- a battery with limited power supply

Applications include:

- Tracking
- **Monitoring** (Environment, **Assets**, . . .)

What is Context Privacy?

- Privacy threats can be classified as either content-based or **context-based**
- Content-based threats have been widely addressed (using cryptography) (Perrig et al. [2])
- Context-based threats are varied
  - Location of event source
  - Location of base station
  - Time at which the event occurred
- We focus on protecting the **location** context of the **event source**

The Problem of Source Location Privacy (SLP)

Given:

- A WSN that detects valuable assets
- A node broadcasting information about an asset

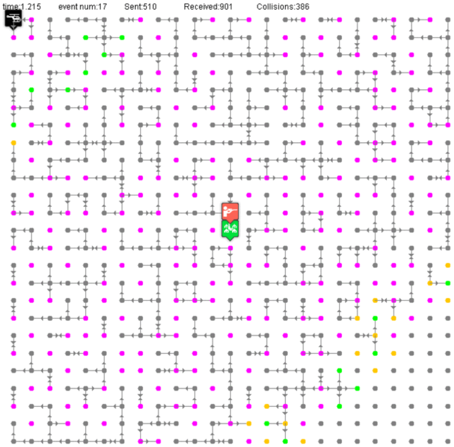
Found:

- An attacker can find the source node by backtracking the messages sent through the network.
- So by deploying a network to monitor a valuable asset, a way has been provided for it to be captured.

The Problem:

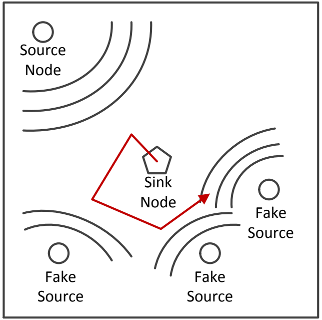
- Panda-Hunter Game
- Difficult

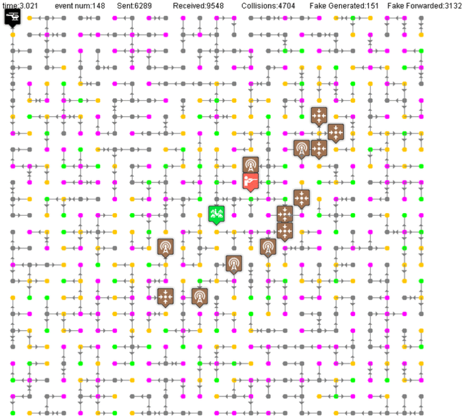
Example: Protectionless Flooding



```
time:1.215     event num:17     Sent:510     Received:901     Collisions:386
```

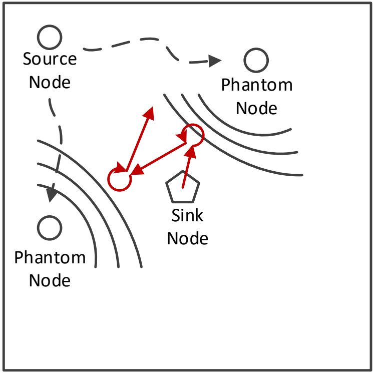
Example: Dynamic Fake Sources

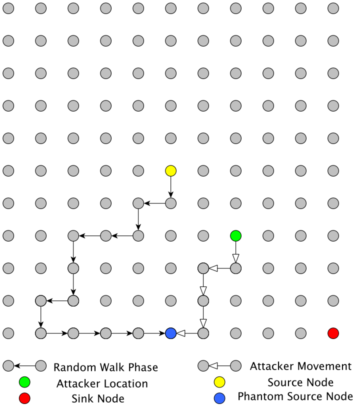




```
time:3.021   event num:148   Sent:6289   Received:9548   Collisions:4704   Fake Generated:151   Fake Forwarded:3132
```

Example: Phantom Routing





Privacy Model

Aim of an SLP protocol: Prevent the attacker from capturing an asset through information the WSN leaks.

- A *stationary asset* cannot be protected as an attacker can perform an exhaustive search.
- A *mobile asset* will only stay in detection range of a WSN node for a certain amount of time.
- The SLP problem can only be considered when it is time-bounded.
- The *safety period* is how long the asset will be protected for.

Attacker Model

Aim of an Attacker ($\mathcal{A}$): to reach the source (s) within the safety period (λ).

The attacker:

- is present in the network
- is mobile
- has a limited range
- starts at the sink
- follows the first new packet it receives

Transforming Routing from Protectionless into SLP

Transform the protectionless routing protocol $\mathcal{R}_{\mathcal{N}}$ into a SLP routing protocol $\mathcal{R}_{\mathcal{S}}$, via an SLP transformation $\mathcal{P}$.

$$\mathcal{R}_{\mathcal{N}} \xrightarrow{\mathcal{P}} \mathcal{R}_{\mathcal{S}}$$

Want to ensure that when using $\mathcal{R}_{\mathcal{S}}$:

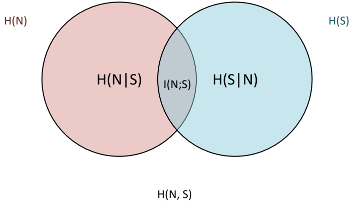
- There exists a path from source s to the sink
- The attacker $\mathcal{A}$ reaches s with probability δ within the safety period λ
- The attacker experiences greater information loss as it should lead to reduced privacy loss

$\mathcal{R}$ is a routing matrix where $\mathcal{R}[i, j]$ represents the probability j receives a message from i.

Measuring Privacy Loss - Mutual Information

- $\mathcal{N}$ is a random variable of attacker transitions under a protectionless routing protocol $\mathcal{R}_{\mathcal{N}}$
- $\mathcal{S}$ is a random variable of attacker transitions under a SLP routing protocol $\mathcal{R}_{\mathcal{S}}$
- A transition is a move the attacker makes from one node to another.
- Mutual information (I) between protectionless ($\mathcal{N}$) and SLP ($\mathcal{S}$) random variables:

$$I(\mathcal{N};\mathcal{S}) = H(\mathcal{N}) - H(\mathcal{N} \mid \mathcal{S}) \quad (1)$$



- If the entropy (H) is the same, then the presence of any SLP routing protocol has no effect of the way the attacker responds to the transitions in $\mathcal{N}$.

Measuring Privacy Loss - How To Calculate It

The probability the attacker takes transition n within λ' steps if transition f is the next transition, where $\lambda' \propto \lambda$, is given by $\Pr(\mathcal{N} = n, \mathcal{S} = f)$.

$$\Pr(\mathcal{N} = n, \mathcal{S} = f) = \sum_{\tau=0}^{\lambda'} \Pr(\mathcal{N} = n, \mathcal{T} = \tau \mid \mathcal{S} = f) \Pr(\mathcal{S} = f) \tag{2}$$

$$\Pr(\mathcal{N} = n \mid \mathcal{S} = f) = \sum_{n \in \mathcal{N}} \left(\omega^f \cdot \sum_{\tau=0}^{\lambda'} (\mathcal{R}'_{\mathcal{S}})^\tau \cdot {\omega^n}^\top \right) \tag{3}$$

$$\omega^x = \begin{cases} 1 & \text{if } x\text{th entry} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\mathcal{R}'_{\mathcal{S}}[i,j] = \begin{cases} \mathcal{R}_{\mathcal{S}}[i,j] & \text{if } (i,j) \neq n \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Privacy Preserving Data Mining

- Data mining can occur over a series of events in chronological order $\langle e_1 \cdot e_2 \cdot \dots \cdot e_n \rangle$
- To preserve privacy during data mining events can be inserted, removed or reordered while maintaining enough information about the sequence of events.
- We can calculate the information loss between a clear series $(D_{\mathcal{N}})$ and a noisy series $(D_{\mathcal{S}})$ by:

$$IL(D_{\mathcal{N}}, D_{\mathcal{S}}) = \frac{\sum_{i=1}^{n} |f_{D_{\mathcal{N}}}(i) - f_{D_{\mathcal{S}}}(i)|}{\sum_{i=1}^{n} f_{D_{\mathcal{N}}}(i)} \tag{6}$$

- $f_D(i)$ represents the frequency of the data item i in domain D

Applying Information Loss to Source Location Privacy

Applying this technique to SLP we get:

$$IL(D_\mathcal{N}, D_\mathcal{S}) = \frac{\sum_{i=1}^{n} \left| \mathcal{F}_{D_\mathcal{N}}(i) - \mathcal{F}_{D_\mathcal{S}}(i^\lambda) \right|}{\sum_{i=1}^{n} \mathcal{F}_{D_\mathcal{N}}(i)} \tag{7}$$

Where $\mathcal{F}_{\mathcal{D}_\mathcal{N}}(i)$ and $\mathcal{F}_{\mathcal{D}_\mathcal{S}}(i^{\lambda'})$ are defined as:

$$\mathcal{F}_{\mathcal{D}_\mathcal{N}}(i) = \begin{cases} 1 & \text{if transition } i \text{ is used in } \mathcal{N} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$\mathcal{F}_{\mathcal{D}_\mathcal{S}}(i^\lambda) = \begin{cases} 1 & \text{if } i \text{ is not taken within } \lambda' \text{ steps in } \mathcal{S} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Equation 7 states that the more dissimilar the set of transitions taken within λ' time units are, the greater is the information loss, hence the lesser the privacy loss.

Implications of Information Loss

- To maximise information loss $(IL(D_{\mathcal{N}}, D_{\mathcal{S}}) = 1)$ then $D_{\mathcal{N}} \cap D_{\mathcal{S}} = \emptyset$
- To minimise privacy loss $\mathcal{R}_{\mathcal{N}}$ and $\mathcal{R}_{\mathcal{S}}$ cannot share any transitions
- We allows transitions to be shared as long as they occur beyond λ' steps
- Ideally, an attacker should take a transition in $\mathcal{R}_{\mathcal{S}}$ rather than in $\mathcal{R}_{\mathcal{N}}$ before λ'

Competing Path

Definition (Competing Paths)

Given a network $G = (V, E)$ and a protectionless routing protocol $\mathcal{R}_{\mathcal{N}}$, two distinct paths p_1 and p_2 under $\mathcal{R}_{\mathcal{N}}$ compete at a node $n \in V$ iff the following are satisfied:

- p_1 and p_2 are source-converging paths
- $\exists (i, j), (i, j') \in E : (i, j) \in p_1 \wedge (i, j') \in p_2 : i = n$
- $\mathcal{R}_{\mathcal{N}}[j, n] > 0 \wedge \mathcal{R}_{\mathcal{N}}[j', n] \geq 0,\ j \neq j'$

- If p_1 is used in $\mathcal{R}_{\mathcal{N}}$ then the attacker can be made to follow p_2 in $\mathcal{R}_{\mathcal{S}}$
- Competing paths increase entropy at the node they compete at
- Not all competing paths can lead the attacker away from the source

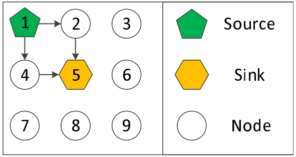
Proper Competing Path

Definition (Proper Competing Paths)

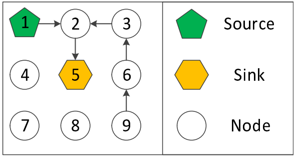
Given a network $G = (V, E)$ and a protectionless routing protocol $\mathcal{R}_{\mathcal{N}}$, two distinct paths p_1 and p_2 under $\mathcal{R}_{\mathcal{N}}$ compete properly at a node $n \in V$ iff the following are satisfied:

- p_1 and p_2 are source-converging paths
- $\exists (i,j),(i,j') \in E : (i,j) \in p_1 \wedge (i,j') \in p_2 : i = n$
- $\mathcal{R}_{\mathcal{N}}[j,n] > 0 \wedge \mathcal{R}_{\mathcal{N}}[j',n] = 0$

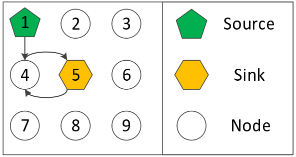
n is a *junction node* where $\mathcal{R}_{\mathcal{S}}$ adds a path that the attacker would not usually take in $\mathcal{R}_{\mathcal{N}}$



Junction Point at 5



Junction Point at 2



Junction Point at 4

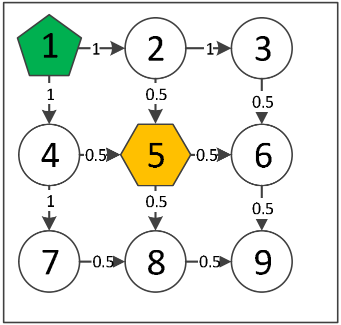
Case Study — $\mathcal{R}_{\mathcal{N}}$ (Flooding)

$\mathcal{R}_{\mathcal{N}}$:

Receiving Nodes

Sending Node

$$\begin{array}{cc} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$



Set of paths $= \{\langle(5,2)\cdot(2,1)\rangle, \langle(5,4)\cdot(4,1)\rangle\}$. Safety period $\lambda = 4$.

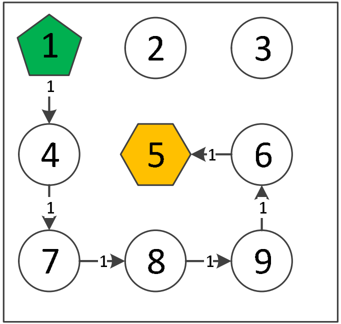
Case Study — $\mathcal{R}_{\mathcal{S}}$

$\mathcal{R}_{\mathcal{S}}$:

Receiving Nodes

Sending Node

$$\begin{array}{c|ccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 1 & 0 & \cancel{0.5}\ \mathbf{0} & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{0} & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{0} & 0 & 1 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{0} & 0 & \cancel{0.5}\ \mathbf{0} & 0 \\
6 & 0 & 0 & 0 & 0 & \cancel{0}\ \mathbf{1} & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{0} \\
7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{1} & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cancel{0.5}\ \mathbf{1} \\
9 & 0 & 0 & 0 & 0 & 0 & \cancel{0}\ \mathbf{1} & 0 & 0 & 0
\end{array}$$



Set of paths $= \{\langle(5,6)\cdot(6,9)\cdot(9,8)\cdot(8,7)\cdot(7,4)\cdot(4,1)\rangle\}$. Minimum path length $= 6$.
Which is greater than the λ of 4.

Discussion

- We do not expect to maximise information loss (minimise privacy loss) as the SLP routing $\mathcal{R}_{\mathcal{S}}$ will contain aspects of the protectionless routing $\mathcal{R}_{\mathcal{N}}$

Exclusions:

- This work on applies to routing-based SLP techniques that need to be transformed to obtain SLP. Techniques such as using data mules are out of its scope (Li et al. [1]).
- This work assumes an attacker present and mobile in the network. It does not apply to global attackers.

Assumptions:

- Links are bidirectional and reliable

Summary

- Formalised creating an SLP-aware routing protocol as a transformation problem
- A way to evaluate the difference in *information loss* between a routing protocol and a SLP version of it by using ideas from privacy preserving data mining
- Introduced the idea of *competing paths* as a way to model SLP techniques

Thank You for Listening

Any Questions?

References

[1] Na Li, Mayank Raj, Donggang Liu, Matthew Wright, and Sajal K. Das. Using data mules to preserve source location privacy in wireless sensor networks. In *Proceedings of the 13th International Conference on Distributed Computing and Networking*, ICDCN'12, pages 309–324, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-25958-6. doi: 10.1007/978-3-642-25959-3_23.

[2] Adrian Perrig, John Stankovic, and David Wagner. Security in wireless sensor networks. *Commun. ACM*, 47(6):53–57, June 2004. ISSN 0001-0782. doi: 10.1145/990680.990707.