A Countermeasure against Traffic-Analysis based Base Station Detection in WSN

Shibasis Biswas
Jadavpur University, Kolkata
shibasisbiswas@yahoo.co.in

Sayan Mukherjee
Jadavpur University, Kolkata
ju.sayan@gmail.com

Krishnendu Mukhopadhyaya
Indian Statistical Institute, Kolkata
krishnendu@isical.ac.in

Abstract: With the increasing popularity of civilian and military based applications of Wireless sensor networks (WSN), the WSN are being subjected to numerous security threats. One of the major security threats is the detection of the base station, the centre of data processing in WSN, by an adversary through continuous vigilant analysis of traffic flow. The objective of this paper is to suggest an algorithm to prevent base station detection by an adversary by creating a situation where a number of ordinary nodes appear to functioning just like the real base station, so that the adversary is tricked into detecting one of these pseudo base stations as the real one, while at the same time not affecting the normal lifetime of the WSN, i.e. the secure WSN will have the same lifetime as the normal, non-secure one.

1. Introduction

Wireless sensor networks (WSN) are becoming popular because of their relative low cost, intricate sensing technology, processing power and wireless communication abilities. They are finding a lot of applications for both domestic as well as large scale purposes, in civilian as well as military fields. But at the same time the security threats that WSN are being subjected to are considerable. But the problem with the WSN is that their resource constraints. These nodes are typically resource-constrained, with limited energy lifetime, low-power micro sensors and actuators, slow embedded processors, limited memory, and low-bandwidth radios. [10] These resource constraints render the traditional network security techniques to be unsuitable to be applied for securing WSN. A sensor network has a base station and a number of nodes. Each node processes data that it received from its group of neighbouring sensor nodes and sends that processed data to the base station through multiple hops. The base station is the most critical part of a sensor network as all the relevant data collected by the sensor nodes are directed towards the base station where the data is aggregated and processed. So if an adversary can detect the base station and compromise it, the entire WSN will be rendered useless.

An adversary can be guided by any one of the following two motives in attacking the WSN [1]

- **Benefit from data**: The adversary may be keen to gain access to the sensitive data being transmitted or monitored.
- **Mission interference**: The adversary may intend to damage the WSN rather than gain access to the data being relayed.

If an adversary is guided by any one of these motives, it is obvious that detecting and compromising the base station will be of most benefit to it. Since the base station is the centre of data aggregation, an adversary by compromising the base station will be able to access majority of data flowing through the network. Also if it makes the base station non-functional, the entire WSN will collapse.

In this paper we discuss a method by which an adversary may attempt to detect the base station and then we proceed to give an energy-efficient solution for that intrusion strategy.

2. Related works

With the increasing concern for security issues in sensor networks, there had been several works addressing this issue. Lee et al [11] and Walters et al [8] have addressed the general security issues in WSN and the proposed security mechanisms and their effectiveness. Abu-Ghazaleh et al [1] propose an application-specific security context as the combination of a potential attacker's motivation and the WSN vulnerability. Gu et al [9] have given a solution for securing a WSN against search-based physical attacks.

A very relevant discussion on the traffic-analysis issue has been carried out by Deng et al in [6] and [7]. They outline three techniques to reduce the uniform directionality of traffic flow towards the base station. **Multi-parent routing scheme** [7] suggests that the messages are not necessarily transmitted through the shortest path but multiple paths are selected randomly between the same source and destination. A node randomly selects one of its parent nodes to forward the packets. In **Random walk scheme** [7] the network is visualized as a tree. Here the messages are forwarded up the tree to the parent and ultimately reach the base station, which is the root of the tree. Here in order to protect against the rate monitoring attack, the message is forwarded not only to its parent but also to its neighbours with some probability. In **Fractal propagation method** [7] when a node senses that its neighbour is sending a message to the base station, it generates fake messages with some probability and forwards to one of its neighbour nodes. The generation of fake messages creates more randomness in the network.

3. A common traffic-analysis based base station detection mechanism and a corresponding counteracting strategy

Sensor data is typically routed along relatively fixed paths from sensor nodes towards the base station. This produces quite pronounced traffic patterns that reveal the direction towards and hence the location of the base station. [7] One common strategy for the adversary is to check the number of messages transmitted and the number of messages received by a particular node. Now since the base station acts as the centre of data accumulation, it receives large volumes of messages. Compared to the no. of received messages, the no. of transmitted messages for the base station will be really less. But for an ordinary node, the no. of transmitted and received messages will be equal, as whenever a node receives a message from any of its neighbours in its radio range, it transmits the same. So if an adversary keeps track of the difference between the number of messages received and the number of messages transmitted for each node, it can track the base station as for the base station this difference will be relatively large, whereas for other nodes this difference will be almost zero. In doing so the adversary may not need to know the pattern of the message that flows in the network, it will just keep track of the no. of messages.

In order to counteract this traffic analysis strategy, we propose a solution which will not allow the adversary to track the base station just by analyzing the difference between the number of messages received and the number of messages transmitted for each node, whereas at the same time there will be no extra overhead for the network as a whole, i.e. the lifetime of the network will remain the same. The neighbouring nodes of the base station are the nodes that have to transmit maximum messages in the WSN. So these nodes are the ones that are expected to run out of power ahead of the other nodes. We are making the following assumptions:

- If a single node dies, the lifetime of the entire network is over.
- The receipt of messages means sending of an acknowledgement message which costs negligible energy in comparison to the transmission of message.

The first assumption is valid as that single node will be one of the neighbours of the base station, and as the average no. of messages transmitted by each of the neighbours of the base station will not be radically different, if a single neighbour dies, it is most likely that the other neighbours of the base station are also expected to run out of power shortly.

Now when the network lifetime is over, there will be several nodes, mostly not in the vicinity of the base station, which will have retained most of their power, as these remote nodes have to send very few numbers of messages. This power will be unnecessarily wasted. So our strategy is to utilize this excess energy in some fake message generation and transmission, which will guarantee that the base station is not detected. Firstly we select several nodes to act as **pseudo base stations**. Our aim will be to make the difference between the number of messages received and the number of messages transmitted for these **pseudo base stations** greater than or equal to that for the real base station.

➢ Method to find out the level and number of neighbours of each node:

During the initialization phase, the base station will send a beacon message to each of its neighbouring nodes. The beacon message will have a counter value which will be initialized to one by the base station. As a node receives a beacon message for the first time, it will save the counter value and will transmit the message by incrementing the counter value. Also it will send an acknowledge message to the sending node. If a node receives a beacon message for the second time or third time, it will not save the counter value, but will only send the acknowledge message. When all the nodes have received the beacon message the counter value saved in each node will be an indication of its hop count from the base station (we call it its level) and the total number of acknowledgement messages received by a node will be the measure of the number of its neighbours.

- **Criteria for selection of pseudo base stations:** (1) The level of a pseudo-base station will not be close to 1. (2) Two or more pseudo base stations should not be neighbours. (3) A pseudo base station should have maximum no. of neighbours. (4) Two pseudo base stations should not share the same neighbour.

Our proposed algorithm:

$\mathbf{x_1}$ = No. of maximum messages (fake + real) that a node can transmit in the 1st time slice

m = Average number of neighbours of a node

n = Number of neighbours of the real base station

D = Desired difference between the number of messages received and the number of messages transmitted for the real base station as well as the pseudo base stations

c = No. of intervals in a single time slice

$\mathbf{X_i}$ = Total no. of messages transmitted by an *ordinary* node up to the *i-th* interval

$\mathbf{X_1 = x_1}$

Let $\mathbf{y_1, y_2, y_3 \ldots y_n}$ are the total number of messages transmitted by the neighbours of the real base station upto the (i -1) th time slice

```
for the i-th time slice {
        $x_i$ = No. of maximum messages (fake + real) that a node can transmit in the i-th time slice
if (average ( $y_1, y_2, y_3, \ldots\ldots\ldots, y_n$) >= $X_{i-1}$ )
                    $x_i$ = {average ( $y_1, y_2, y_3, \ldots\ldots\ldots, y_n$)}/(i-1) ;
else  $x_i$ = {{average ( $y_1, y_2, y_3, \ldots\ldots\ldots, y_n$)}/(i-1)}- { $X_{i-1}$ - average ( $y_1, y_2, y_3, \ldots\ldots\ldots, y_n$)} ;

  D = $x_i$ * m - $x_i$ ;
  $X_i$ = $X_{i-1}$ + $x_i$ ;

 if ( a node is an ordinary node ) {
 if (at the j-th interval the total number of messages (real or fake) transmitted upto that point in the time slice < ($x_i$ /c)* j )
      transmit / generate {($x_i$ /c)* j – total number of messages transmitted in that time slice} fake messages;
  else    do not accept any fake message;
  }
if ( a node is a pseudo base station ) {
      accept all the messages ( real + fake) transmitted by its neighbours;
      transmit all the real messages that are intended for it ;
      drop all other real and fake messages ;
}
if ( a node is the REAL base station) {
   if (at the j-th interval the total number of messages (real or fake) transmitted upto that point in the time slice < (total number of messages received in that time slice – (D/c)* j ) {
    generate {(total number of messages received in that time slice –(D/c)* j ) – total number of messages transmitted in that time slice} fake messages;
    }
    }
}
```

We divide the entire lifetime of the network into a number of time slices. We consider the difference between the number of messages received and the number of messages transmitted for each node in each time slice. For each time slice we set a limit ($\mathbf{x_i}$) on the number of messages to be transmitted by each node which is NOT a pseudo base station as well as NOT a neighbour of the real base station (We call such a node as an *"ordinary"* node). If within that time slice a node does not transmit that many number of real messages, it is made to generate/ transmit required number of **fake messages** so that at the end of the time slice the no. of messages transmitted by each node is equal to the set limit. The generated fake messages will contain an address field which will refer to a pseudo base station and the choice of the pseudo base station will be in the round robin basis for nodes which are not neighbours of any pseudo base station. A pseudo base station will accept all the messages (real + fake) transmitted by its neighbours, transmit all the real messages that are intended for it and drop all other real and fake messages. But since the pseudo base stations do not transmit any fake messages, the difference between the number of messages received and the number of messages transmitted increases considerably for each of the pseudo base station. Since each ordinary node is guaranteed to transmit x_i messages in a time slice and a pseudo base station is expected to transmit no more than x_i messages, we can get the difference between the no. of received messages and transmitted messages considering average number of neighbours for the pseudo base stations. This same difference will be maintained for the REAL base station by making the REAL base station to transmit the required no. of fake messages.

Arguments in favor of the algorithm

I. Lifetime of the network is not affected:

For the *i-th* slice, the expected average number of messages transmitted by a neighbour of the real base station is the average number of messages transmitted per slice by a neighbour of the real base station for the (i-1) slices. Now if the average of the total number of messages transmitted by all the neighbours of the real base station up to the *(i-1)-th* slice {i.e. average (y1, y_2, y_3,……….., y_n) } is greater than or equal to the total no. of messages transmitted by an ordinary node up to the *(i-1)-th* slice (i.e. X_{i-1}) , then we set the limit of the no. of messages to be transmitted by an ordinary node in the *i-th* slice equal to expected average number of messages transmitted by a neighbour of the real base station in the *i-th* slice. Otherwise we decrease this limit from expected average number of messages transmitted by a neighbour of the real base station in the *i-th* slice by the same margin as the difference between average (y1, y_2, y_3,……….., y_n) and X_{i-1}. This will ensure that at the end of the *i-th* slice the average of the total number of messages transmitted by all the neighbours of the real base station will remain greater than or equal to the total no. of messages transmitted by an ordinary node. There will be some neighbours of the real base station which will transmit more messages than the average value {i.e. average (y1, y_2, y_3, ………..,y_n) } and none of the ordinary nodes will die before these nodes, so that the normal lifetime of the network is not affected.

II. The difference between the no. of received messages and transmitted messages remains the same for both the real base station and the pseudo base stations:

The pseudo base stations are so selected that they have maximum neighbours. So the no. of messages received by the pseudo base stations in any slice = (x_i * No. of neighbours of the pseudo base station). This value >= (x_i * m) since m is the average no. of neighbours of any node. So for any pseudo base station,
The difference between the no. of received messages and transmitted messages
= (x_i * No. of neighbours of the pseudo base station) **-** x_i
>= D where D is the set value of difference between the no. of received messages and transmitted messages for the real base station.

4. Performance Evaluation

In this section we show the performance evaluation of our algorithm in counteracting the detection of the real base station by an adversary. In the simulation process we create an environment consisting of 100-1000 sensor nodes, randomly distributed. We specify the maximum number of levels in the network and the upper bound of the no. of neighbours of a node. Based on the above mentioned criteria a suitable number of pseudo base stations is selected. As seen from our simulation, the no. of pseudo base stations that can be selected satisfying the given criteria is around 10-15% of the total no. of nodes in the network. An important measure of efficiency of the algorithm is the ratio of the average no. of pseudo base stations detected per time slice and the no. of pseudo base stations selected during the initialization phase (We call this ratio as $\mathbf{R}_{\textbf{pbs detected / pbs selected}}$). From our studies it is revealed that this ratio is almost constant and has a value of around 0.5 to 0.6 (**Fig 1**). Thus on an average around 7-8% of the total no. of nodes is detected by an adversary as base stations in each time slice. This result is satisfactory, as from the adversary's point of view it will now have to compromise 7-8% of the total no. of nodes in to damage the entire WSN. Also that the lifetime of the WSN remains unaffected is shown in **Fig 2** by the comparison of the lifetime of a secure WSN with that of the corresponding non-secure WSN in terms of the no. of time slices that the network will work without failure.

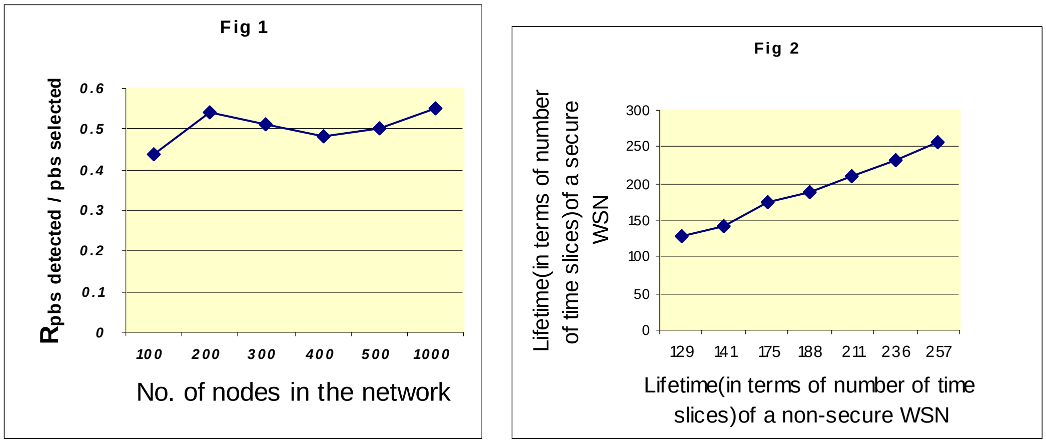

Fig 1. The value of $\mathbf{R}_{\text{pbs detected / pbs selected}}$ for different no. of nodes. **Fig 2.** The lifetime of a secure WSN (in terms of no. of time slices) vs. the lifetime of the same non-secure WSN under different simulating conditions. (WSN consists of 300 nodes)

5. Conclusion

In this paper we have addressed a particular technique adopted by an external adversary in detecting the base station of a WSN and have proposed a scheme to counteract the method while not affecting the normal lifetime of the WSN. We have presented detailed simulation results which reveal that our aim of creating a situation of the illusionary presence of a number of pseudo base stations to trick the adversary is satisfyingly achieved, and at the same time the lifetime of the network is unaffected. We however believe that there is scope of further development in this aspect. We have not considered the overhead involved in message receiving and also the possibility of an internal adversary which can detect the difference between a real and a fake message. We are currently focusing our attention on improving the algorithm by addressing these issues.

6. References

1. An Application Driven Perspective on Wireless Sensor Network Security (Eric Sabbah, Adnan Majeed, Kyoung Don, et.al), 2006.
2. Security for Sensor Networks (Jeffery Undercoffer, Sasikanth Avancha, Anupam Joshi and John Pinkston), 2004.
3. Secure Wireless Sensor Networks: Problems and Solutions (Fei Hu, Jim Ziobro, Jason Tillett , Neeraj K. Sharma) CADIP Research Symposium, 2002.
4. Honeybees: Combining Replication and Evasion for Mitigating Base-station jamming in Sensor Networks (Sherif Khattab, Daniel Mosse, and Rami Melhem) Parallel and Distributed Processing Symposium, 2006
5. Secure Hierarchical In-Network Aggregation in Sensor Networks (Haowen Chan, Perrig, et.al) 13th ACM conference, 2006.
6. Defending against Traffic Analysis Attack in Wireless Sensor Networks (Jing Deng, Richard Han and Shivakant Mishra), USENIX Security Symposium 2004.
7. Countermeasures against Traffic Analysis Attacks in Wireless Sensor Networks (Jing Deng, Richard Han, Shivakant Mishra), SECURECOMM 2005.
8. Wireless Sensor Network Security: A Survey (John Paul Walters, Zhengqiang Liang, Weisong Shi,and Vipin Chaudhary), 2005.
9. Defending Against Physical Attacks in Sensor Networks (Wenjun Gu, Xun Wang, Sriram Chellappan and Dong Xuan), IEEE MASS 2005
10. Intrusion Tolerance and Anti-Traffic Analysis Strategies For Wireless Sensor Networks (Jing Deng, Richard Han and Shivakant Mishra), DCN 2004.
11. Security in Wireless Sensor Networks: Issues and Challenges (Al-Sakib Khan Pathan, Hyung-Woo Lee and Choong Seon Hong), ICACT 2006.