Boundary Probabilistic Algorithm for Source Location Privacy in Wireless Sensor Networks

T. M. Nishi Kumar Singh
Department of Computer Science and Engg.
North Eastern Regional Institute of Science and Technology
Nirjuli, India

Ningrinla Marchang
Department of Computer Science and Engg.
North Eastern Regional Institute of Science and Technology
Nirjuli, India

H. Saratchandra Sharma
Department of Computer Science and Engg.
North Eastern Regional Institute of Science and Technology
Nirjuli, India

ABSTRACT
With the recent advances in wireless sensor technologies, sensor networks have been used in many applications. One of the applications is subject monitoring and tracking, where the locations of the monitored objects are quite sensitive and need to be protected. So event source anonymity is an attractive and critical security property. For event source anonymity, the concept of source location privacy has been implemented against both local (mote-class) and global (laptop-class) adversaries. In this paper, a scheme called Boundary probabilistic algorithm is proposed to protect against laptop-class attacks. Moreover, a comparison of this proposed scheme with two existing schemes namely- Naive algorithm and Probabilistic algorithm is proposed. In the proposed Boundary Probabilistic algorithm, filtered dummy messages for achieving source location privacy are used. Through simulation results, the proposed scheme is illustrated, the message complexity is reduced. Moreover the delay of the real event message to reach the base station is reduced.

Keywords
Source location privacy (SLP), sensor network.

1. INTRODUCTION
Wireless sensor network (WSN) is an ad-hoc network of wireless sensor nodes. In this network, a large number of sensor nodes are deployed to monitor a vast field, where the operational conditions are most often harsh or even hostile. However, the nodes in WSN have severe resource constraints due to their lack of processing power, limited memory and energy [1][4]. Sensor nodes are capable of sensing an event, collecting information, processing that information and disseminating them. Whatever information is sensed by a node is relayed to the subsequent neighbor nodes until it reaches the base station or the sink node. The moment a node senses the subject, an event is said to have occurred. These capabilities of sensor nodes make them useful in many applications like subject tracking and monitoring, surveillance system, area monitoring and many more. In spite of much usefulness of these networks there are security issues which need to be addressed. The panda hunter game [3][6] as an example is being taken. A node that sensed the panda informs the base station by sending a message that travels via intermediary nodes to the base station. The hunter can kill the panda by tracing the messages from the WSN all the way to the source node that sensed the panda. Hence, the source location privacy (SLP) [2][5] is being aimed. SLP requires more than confidentiality of the messages exchanged between the nodes. SLP requires that the flow of the messages do not give up location of the source node. Since sensor nodes have limited energy supply, it is important to take note of the energy consumption, and also the message delay while devising a method to protect privacy.

2. SOURCE LOCATION PRIVACY USING DUMMY DATA SOURCE
Dummy data sources are nodes that generate and send out dummy packets to other nodes in the network. The dummy packets do not contain information about real events, but are used to obfuscate the real traffic or divert the adversary/attacker by mimicking the presence of a fake subject. Previous work has divided attackers into two types: mote-class and laptop-class [7]. Mote-class attackers are assumed to be limited to small physical devices with capabilities similar to sensor nodes. Thus, at any given time, a mote-class attacker can only monitor communications between a limited numbers of nodes. In contrast, laptop-class attackers are assumed to have much stronger computational capabilities and longer radio range—if they are equipped with hardware powerful enough, the radius of their monitoring area could even cover the entire network. This paper addresses solution against laptop-class attacks in sensor networks using dummy messages and in this work, it is assumed that laptop-class attackers can eavesdrop on all communications in a sensor network.

3. RELATED WORK AGAINST LAPTOP-CLASS ATTACKS
There has been significant past research on source location privacy in wireless networks. Mehta et al. [8] were one of the first to defend against a global adversary with a solution using encryption and dummy traffic called periodic collection (PeCo). Shao et al. [10] proposed a scheme called FitProbRate, which realizes statistically strong source anonymity for sensor networks. Yang et al. [9] introduced a carefully chosen dummy traffic to hide the real source and select some sensors as proxies that filtered dummy messages on their way to the base station. Bicakci et al. [14] proposed another filter-based solution called the optimal filtering scheme (OFS) which allows for the optimal routing and filtering, giving the WSN an optimal lifetime. Lu et al. [11] introduced the timed efficient privacy preservation (TESPˆ2) which also filter out the dummy traffic and uses symmetric keys for encryption and decryption of message. Doomun et al. [13] introduced cloud-based scheme for protecting source

location privacy called source and destination seclusion using clouds (SECLOUD) as a solution that defend against global adversaries. Ouyang et al. [5] introduced schemes, like ConstRate (naive algorithm), globally optimal algorithm (GOA), heuristic greedy algorithm (HGA) and probabilistic algorithm to provide source event unobservability in the network by introducing dummy traffic to hide the real event sources.

4. PROBLEM DEFINITION

In implementing the solution, the following is being defined:

4.1 Network Model

In this scheme a WSN consisting of a large number of randomly deployed sensor nodes and a base station is being considered. Each node, relatively stationary, has the same radio range of say *r*. It is assumed that every sensor node has a routing table set up to route the packets before the proposed scheme is employed. And also assumed that the messages are protected using an encryption algorithm.

4.2 Communication Model

There are mainly two operational phases in WSNs to set up the communication: topology discovery and data transmission phase [12]. In the topology discovery phase the base station floods the WSN by sending a message to each and every node, so that each node knows the *hopcount* (which is the number of hops) up to the base station and thus the shortest path to the base station. In the data transmission phase, source node transmits sensed data to the base station through multiple hops and the time period to reach the base station is called the *delivery time*.

4.3 Attack Model

The following assumptions have been made about the attacker. The attacker can know the location of every sensor node because of his powerful detection equipment and has a global view of the sensor network. The attacker can always hear all on-going communications in the sensor networks. The attacker is able to sense, store, and analyze all transmitted data, but is not able to understand them (based on the assumption that data can be safely encrypted for transmission). The attacker only eavesdrops on communications between sensor nodes and will not physically compromise sensor nodes. The attacker's goal is to find the location of the source node which is originating an event message.

4.4 Metrics

The following metrics have been used to evaluate the methods for source location protection methods.

4.4.1 Security

The probability that an attacker successfully identifies the source node originating a real event message.

4.4.2 Delivery time

The time taken by an event message to travel from the source to the base station.

4.4.3 Energy cost

The energy cost of sensor nodes consists of computational cost and communication cost. Communication is much more expensive than computation in terms of energy. To simplify the analysis, the communication cost of the network is considered. In this paper, the number of messages transmitted in the network is calculated to measure the energy cost.

5. NAIVE ALGORITHM

The Naive Algorithm [5] is a simple solution that provides protection against laptop-class attackers. In this idea, each node transmits a dummy message to its neighbors at the end of each fixed time period. These dummy messages are called maintenance messages and the periods as maintenance periods. The maintenance period is usually made long enough to make sure that the sensor nodes do not run out of battery quickly due to dummy messages (maintenance messages). When a source wants to send an event message, the event message can masquerade as a dummy message, i.e. by delaying this event message and replacing the next dummy message with it. The receiver (sensor node) of this event message needs to wait until the end of its current maintenance period and forward the message to the next node along the routing path to the base station.

6. PROBABILISTIC ALGORITHM

In probabilistic algorithm [5], the hopcount of each node up to the base station is calculated first as done in the naive algorithm. If a node originates or receives an event message, it will select the immediate next node and forward the message to this node at the end of the current maintenance period. If a node is not in the delivery path of any real event message, it will send a dummy message at the end of every maintenance period with a probability of *Th*.

7. BOUNDARY PROBABILISTIC ALGORITHM

In this section, the proposed algorithm is presented. The Probabilistic algorithm is modified to generate the proposed Boundary Probabilistic algorithm for achieving better performance. In Probabilistic algorithm every node generates dummy messages whereas in Boundary Probabilistic algorithm, the nodes at the boundary of the sensor network are only allowed to generate the dummy messages. The idea is that the messages from the boundary nodes go through the interior nodes to reach the base station. Thus, interior nodes need not generate dummy messages. The proposed algorithm mainly focuses on reducing the overall energy cost of WSNs by filtering the generation of dummy messages.

7.1 Algorithm for finding the boundary nodes

Algorithm 1 determines the boundary nodes of the deployed WSN. Here, G is an undirected graph representing the sensor network. Array A is the array that stores index numbers of the boundary nodes. The following notations are:

node.no	Index of this node
node.prev	Previous node in the routing table of this node
node.hp	Hopcount of this node upto the base station
N	Number of nodes in WSN
Th	Preset threshold probability of WSN
t_m	Maintenance time

Algorithm 1: Determining the boundary nodes

```
Input: Graph G
Output: Array A containing boundary nodes

 for j=1 to n
       I[j] :=  node[j].no
       P[j]:=  node[j].prev
 end
 for i:=1 to n
    for j:=1 to n
        if I[i].no != P[j].prev
           A[i] := I[i].no
        end
    end
end
```

Figure 1 depicts a communication scenario in WSN. When a node senses an event, the real message is forwarded to next node in the delivery path masqueraded as dummy message by delaying the real event message and replacing next dummy message with it. Boundary nodes generate the dummy messages and forward the messages to interior nodes to reach the base station.

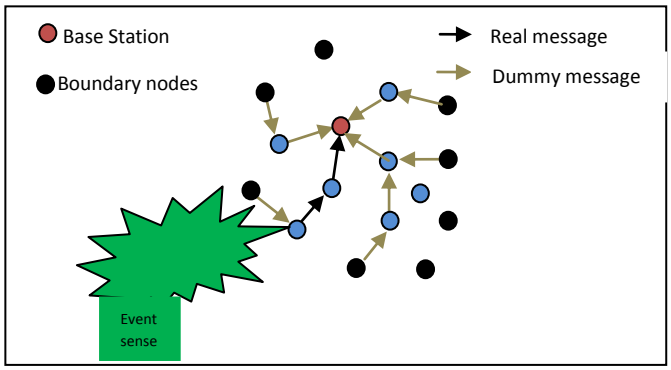

Figure 1: Boundary Probabilistic Scheme

In this scheme, all the boundary nodes are assigned a random number, say 0 < p < 1. If a node senses an event message, it will select the next node and forward the message at the end of its maintenance period.

Algorithm 2: Boundary Probabilistic algorithm

```
while true do
      if it receives a message from a neighbor
            wait for the next sending time;
            forward the message;
      else
          if it is a boundary node
             get  a random number 0 < p < 1;
             if  p > Th then wait for the next sending time;
             else
                    wait for the next sending time;
                    send out a maintenance message;
             end
          end
      end
end
```

Otherwise the boundary nodes generate the dummy message at the end of every maintenance period if its randomly assigned number has a value less than predefined threshold, *Th*.

7.1.1 Security

In the boundary probabilistic algorithm, every node in the delivery path from the boundary node to the base station sends a message after every random period, which is in the range *a*–*b*. As long as the attacker can hear one message in any period duration *b*, the attacker cannot determine if that message is an event message or a dummy message. The probability of the event messages being discovered by an attacker is very less and hence the security is high.

7.1.2 Delivery time

In the boundary probabilistic algorithm the messages are routed in the shortest path with nodes having varying random maintenance period. As compared to naive and probabilistic algorithm, the number of nodes transmitting dummy message is less since only the boundary nodes generate the dummy messages. Hence, there is not much delay in the event message reaching the base station Thus, the delivery time for the event messages is less.

7.1.3 Energy cost

Compared to the probabilistic algorithm, only boundary nodes generate a dummy message and nodes in the delivery path relay those messages after every random period in the boundary probabilistic algorithm. Thus, the number of messages transmitted is less and hence energy cost expended is less than in the Probabilistic algorithm.

8. EXPERIMENTS

This section describes the comparison of the simulation results of the algorithms: *Naive, Probabilistic and Boundary Probabilistic.*

In this simulation, 200 sensor nodes are uniformly distributed in a rectangular area of 400m×400m, the radio range of each sensor node is 40m, the probabilistic threshold (*Th*) is set to 0.31 and the assumed transmission time between two sensors is 10msec. Performance in terms of delivery time and energy cost is shown by simulation graph in Figs. 2(a) and (b).

The plot with the label *Probabilistic (t_m = 1-5sec, Th= 0.31)* denotes that in Probabilistic algorithm, the maintenance period of every node in WSN is varying and is in the range of 1-5 seconds and the threshold probability (*Th*) is set to 0.31. The label Naive (t_m=3sec) denotes that in naive algorithm the maintenance period of every node in WSN is fixed and is set to 3 sec and the label *Boundary Probabilistic (t_m = 1-5sec, Th= 0.31)* represents the same specifications as probabilistic algorithm mentioned above.

Fig. 2(a) is a graph of the delivery time (in seconds) vs. the number of hops. The number of hops is the *hopcount* of a node that sensed the event up to the base station. In the naive algorithm, the maintenance period, t_m, of every node is fixed which is set to 3 seconds. Each and every node within the WSN transmits dummy messages at the end of every maintenance period. Because event message is always forwarded by the nodes along the routing path to the base station masqueraded as dummy messages, the event message needs to stay at every intermediate node until the end of the current fixed maintenance period.

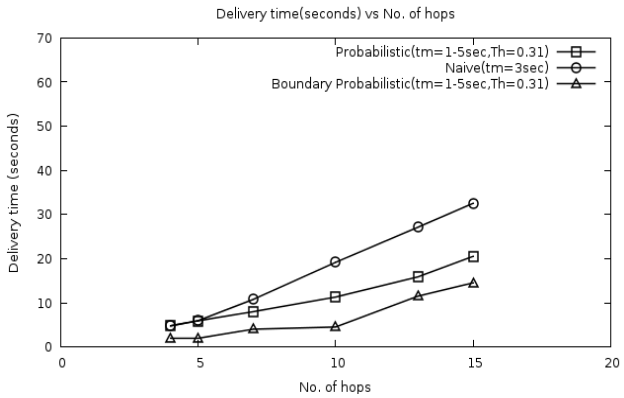

Figure 2(a): Comparison on delivery time of naive, probabilistic and boundary probabilistic algorithm.

Thus the delivery time is very high in the naive algorithm. In the probabilistic algorithm the number of messages generated is lesser since nodes generate dummy messages with some probability, thus the event message need not have to wait longer in the intermediate node. Thus, the delivery time in the case of probabilistic algorithm is better than the naive algorithm. The delivery time of event messages in boundary probabilistic algorithm is the least because the number of dummy messages generated is very less since only the boundary nodes generate the dummy messages and thus the event when masqueraded as dummy message need not have to wait longer in the intermediate nodes.

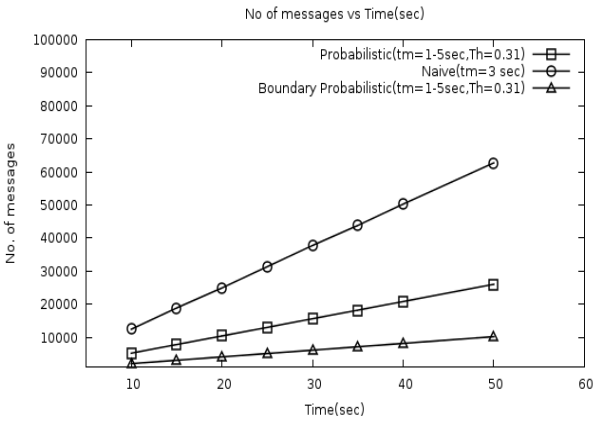

Figure 2(b): Comparison on energy cost of naive, probabilistic and boundary probabilistic algorithm.

The energy cost of WSN is directly proportional to the number of messages transmitted. Figure 2(b) is a graph plot between the number of messages transmitted and time. Here time implies the simulation time. In the naive algorithm, since every node within the WSN transmits dummy messages at the end of every maintenance period, the number of messages generated and transmitted is very high. In the probabilistic algorithm the number of messages transmitted is lesser since nodes generate dummy messages with some probability. The number of messages transmitted is the least in boundary probabilistic algorithm compared to the naive and probabilistic algorithm because the dummy messages are generated only by the boundary nodes with some probability and the interior nodes only forward those messages at the end of their maintenance period. Thus the energy cost of the proposed scheme is better without compromising privacy of the network.

The performance of the boundary probabilistic algorithm for different threshold values Th, i.e. 0.31,0.35 and 0.41 in the same simulation environment is illustrated as above. The performance graph is shown in figure 3(a) and (b).

In Figure 3(a), the delivery time of event messages remains almost the same because event messages are only forwarded in the delivery path to the base station and since only boundary nodes generate the dummy messages, the event message need not have to wait for long in the intermediate node. In Figure 3(b), as the preset threshold probability increases the probability of boundary nodes generating dummy messages is higher and hence the number of messages transmitted in the network increases.

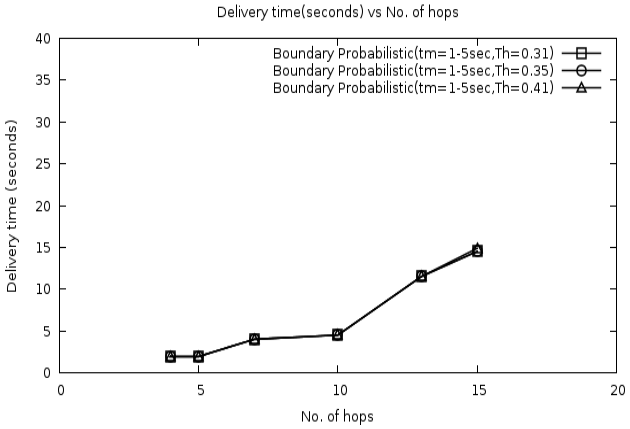

Figure 3(a): Comparison on delivery time of boundary probabilistic algorithm with different threshold values.

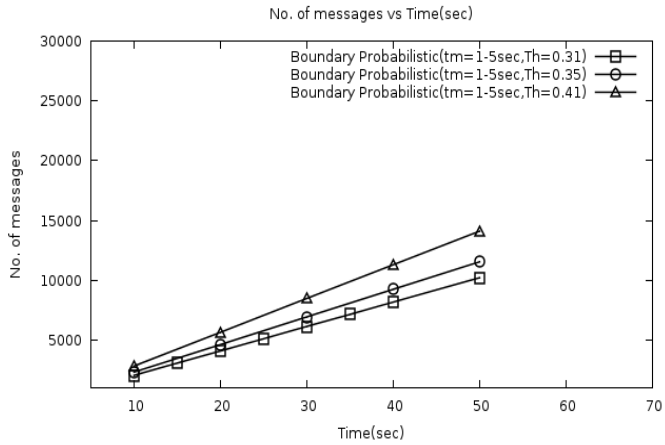

Figure 3(b): Comparison on energy cost of boundary probabilistic algorithm with different threshold values

From the above comparison, the delivery time of boundary probabilistic algorithm is almost the same with the increase in threshold and the number of messages increases with the increase in threshold.

9. CONCLUSION

A source location privacy scheme called Boundary Probabilistic algorithm which is a modification of Probabilistic algorithm is proposed, so that a laptop-class attacker cannot identify the source location even though he/she can monitor the traffic of the entire network. In this new algorithm, the generation of maintenance messages is further reduced compared to Probabilistic algorithm without compromising the privacy of the source node and the delay of

the event messages is further reduced. From the simulation results, it can be concluded that the Boundary Probabilistic algorithm is an efficient algorithm and is also practical solution for providing source location privacy. In the coming future it is aimed to upgrade the boundary nodes responsible for the generation of dummy messages in terms of variable maintenance time and generation of dummy messages by random boundary nodes to increase the security and lifespan of the entire network.

10. REFERENCES

[1] D. Noh and J. Hur, "Using a dynamic backbone for efficient data delivery in solar-powered WSNs," J. Network and Computer Applications, vol. 35, no. 4, pp. 1277 – 1284, July 2012.

[2] M. Conti, J. Willemsen, and B. Crispo, "Providing Source Location Privacy in Wireless Sensor Networks: A Survey," Communications Surveys & Tutorials, IEEE, vol. 15, no. 3, pp. 1238-1280, Third Quarter 2013.

[3] N. Li, N. Zhang, S. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," Ad Hoc Networks, vol. 7, no. 8, pp. 1501–1514, November 2009.

[4] E. Ekici, S. Vural, J. McNair, and D. Al-Abri, "Secure probabilistic location verification in randomly deployed wireless sensor networks," Ad Hoc Networks, vol. 6, no. 2, pp. 195–209, April 2008.

[5] Y. Ouyang, Z. Le, D. Liu, J. Ford, and F. Makedon, "Source location privacy against laptop-class attacks in sensor networks," in SECURECOM:Proc. 4th international conference on Security and privacy in communication networks, ACM, September 2008.

[6] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source location privacy in sensor network routing," in Proceedings of the 25th IEEE International Conference on Distributed Computing Systems(ICDCS), pages 599–608, June 2005.

[7] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications 2003, pp.113-127, 11 May 2003.

[8] K. Mehta, D. Liu, and M. Wright, "Location privacy in sensor networks against a global eavesdropper," in IEEE International Conference on Network Protocols, 2007, pp. 314-323, 16-19 Oct. 2007.

[9] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in Proc. first ACM conference on Wireless network security, pp. 77–88, April 2008.

[10] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards statistically strong source anonymity for sensor networks," in INFOCOM 2008: Proceedings of the 27th IEEE Conference on Computer Communications, 13-18 April 2008.

[11] R. Lu, X. Lin, H. Zhu, and X. Shen, "Tesp2: Timed efficient source privacy preservation scheme for wireless sensor networks," Communications (ICC), 2010 IEEE International Conference on Communication, pp. 1-6, 23-27 May 2010.

[12] A. A. Nezhad, D. Makrakis, A. Miri, "Anonymous Topology Discovery for Multihop Wireless Sensor Networks," in Proc. of the 3rd ACM workshop on QoS and security for wireless and mobile networks, pp. 78-85, 2007.

[13] R. Doomun, T. Hayajneh, P. Krishnamurthy, and D. Tipper, "Secloud:Source and destination seclusion using clouds for wireless ad hoc networks," Computers and Communications, 2009. ISCC 2009. pp. 361-367, 5-8 July 2009.

[14] K. Bicakci, H. Gultekin, B. Tavli, and I. Bagci, "Maximizing lifetime of event-unobservable wireless sensor networks," Computer Standards & Interfaces, vol. 33, no. 4, pp. 401–410, June 2011.