This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Combining Source-Location Privacy and Routing Efficiency in Wireless Sensor Networks

Jian Ren Di Tang
Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824
Email: {renjian, ditony}@egr.msu.edu

***Abstract*—Wireless sensor networks (WSNs) have been widely used in various applications for continuous event monitoring and detection. The WSNs communication is generally event-driven. While confidentiality of the message content can be ensured through content encryption, it is much more difficult to adequately protect the source-location information of the event. For WSNs, source-location privacy service is further complicated by the fact that the sensor nodes consist of low-cost and low-power radio devices, computationally intensive cryptographic algorithms (such as public-key cryptosystems) and large scale broadcasting-based protocols are not suitable for WSNs. On the other hand, exposure of the source-location can jeopardize the successful deployment of WSNs. In this paper, we propose a scheme to provide both source-location privacy and routing efficiency through routing to an intermediate node selected from a hierarchical connected dominating set (CDS) of the network. The CDS represents the backbone of the network and the nodes in the CDS are located in different regions of the network. As a result, choosing nodes from the CDS can ensure the intermediate node to be away from the actual message source node. The selection of the intermediate node can effectively prevent the adversary from performing routing trace back attack to identify the message source node. In addition, this design guarantees a high message delivery ratio and a high message delivery efficiency.**

***Index Terms*—wireless sensor networks, source-location privacy, connected dominating set, delivery ratio**

I. Introduction

With the maturity of technology in both hardware development and protocol design, wireless sensor networks (WSNs) have the potential to be widely used in many scenarios for unattended event monitoring and target tracking. While confidentiality of the message content can be ensured through content encryption, it is much more difficult to adequately protect the source-location information of the event. For WSNs, source-location privacy service is further complicated by the fact that the sensor nodes consist of low-cost and low-power radio devices, computationally intensive cryptographic algorithms (such as public-key cryptosystems) and large scale broadcasting-based protocols are not suitable for WSNs. On the other hand, the adversaries may use expensive radio transceivers, powerful workstations and interact with the network from a distance since they are not restricted to using sensor network hardware. The WSNs communications is generally event-driven. Whenever an event is detected, the sensor nodes normally needs to send the information back to the sink node where the message could be processed or routed out of the WSNs. Based on the event that an adversary detects, it is possible for the adversaries to identify the message source or even identify the source-location through routing trace back attack, even if strong data encryption is utilized. This makes the event and the source-location privacy one of the major issues that can jeopardize the successful deployment of WSNs.

To optimize the sensor nodes for the limited node capabilities and the application specific nature of the networks, traditionally, security requirements were largely ignored. This leaves the WSNs open to network security attacks. In the worst case, adversaries may be able to undetectably take control of some sensor nodes, compromise the cryptographic keys and reprogram the sensor nodes.

In this paper, we propose a scheme that can balance source-location privacy and routing efficiency by combing dynamic routing with hierarchical connected dominating set (CDS). The basic idea of the algorithm is whenever the source node needs to send a message, it routes the message to an intermediate node. However, unlike the existing schemes [1], [2], in this paper, we propose that the intermediate node be selected from a hierarchical CDS. The CDS is used as a communications backbone, and nodes that are not in this set communicate by passing messages through nodes in the set.

This paper has the following major contributions:

1) It is believed that the minimum CDS problem and the maximum leaf spanning tree problem cannot be solved in polynomial time, which makes it difficult for the adversary to predetermine the CDS, therefore, to predetermine the possible intermediate nodes.
2) When the intermediate node is dynamically selected in the CDS, routing trace back attack can be effectively limited. Therefore, source-location privacy can be provided.
3) By routing through the CDS, it guarantees routing efficiency and a high message delivery ratio.

The rest of the paper is organized as follows: in Section II, the related work of this paper is presented, in Section IV, the algorithm is presented in detail; in Section V, the attacks this algorithm is against are analyzed; in Section VI, the simulation results are given; Section VII concludes the paper.

II. Related Work

The related work of this paper can be divided into two categories: the efficient construction of CDS and designing

978-1-4244-9268-8/11/$26.00 ©2011 IEEE

routing protocols based on the CDS, and the existing source-location privacy schemes.

For the CDS construction, the CDS is constructed as the backbone of the whole network to broadcast the global topology of the network [3]. After every node in the ad-hoc network knows the topology of the entire network, they will run an algorithm similar to a link-state routing protocol to compute the shortest paths to other nodes within themselves. In [4], two generic algorithms for computing the CDS of the network were proposed. These algorithms work well in static environments. The idea of the algorithms is similar to calculating the spanning tree originated from some node in the network. After the spanning tree is built, the nodes in the spanning tree, excluding the nodes on the leaves, are nodes that will form the CDS. These algorithms do not work well in ad-hoc networks in the sense that it converges very slowly, because the construction of the spanning tree expands at most two steps at a time. In addition, when the topology of the network changes, the algorithm has to start from the very beginning, which makes it hard to be used in dynamic environment. For WSNs, distributed algorithms are more advantageous and more prominent in calculating the CDSs. The algorithms proposed by Wu and Dai in [5]–[8] are among the most efficient algorithms. Their algorithms are totally distributed. It is up to the nodes to decide whether they are a member in a CDS based only on the information from their neighbors within two hops. Another types of distributed algorithm includes two steps [9]. In the first step, the cluster headers are selected to form a dominating set of the network. Then gateway nodes are chosen to connect these cluster headers. In this paper, our proposed privacy preserving algorithm is based on constructing the CDS [8].

The research on source-location privacy communication was initiated by Chaum [10]. The importance of source-location privacy preservation for WSNs has been explicitly clarified in the panda-hunter game in [11], [12]. In the panda-hunter model, the WSNs are deployed to monitor the activities of pandas. When a panda is discovered, the sensor nodes will report the event to the sink node. The contents of messages can be protected through symmetric encryption. However, if the source node routes the messages to the sink node through the same path for all messages, then by monitoring all incoming messages and performing routing trace back attack, the adversary is able to find the location of the message source node. To solve this security problem, several routing based schemes have been proposed [1], [2], [11]. One of the advantages of routing based schemes is that they do not require extra dummy or carrier traffic while providing the source-location privacy. However, one of the problems of these schemes is that they generally detached routing efficiency with source-location privacy.

III. Models and Assumptions

A. The System Model

We assume that the WSNs are composed of a large number of sensor nodes and a sink node. The sensor nodes are randomly deployed throughout the sensor domain. Each sensor node will have a very limited and non-replenishable energy resource. The sink node is the only destination that every sensor node will send message packets to through a multi-hop routing strategy. The information of the sink node is made public. For security management purpose, each sensor node may also be assigned a node ID corresponding to the location where this message is generated. To prevent adversaries from recovering the source location from the node ID, dynamic ID can be used. In addition, the content of each message can also be encrypted using the shared secret key between the node/grid and the sink node.

We also assume that each sensor node knows its relative location in the sensor domain and has knowledge of its adjacent neighboring nodes in the hierarchical CDS. The key management, including key generation, key distribution and key update, are beyond the scope of this paper.

B. The Adversarial Model

In WSNs, the adversary may try to recover the message source node or jam the packet from being delivered to the sink node. The adversaries would try their best to equip themselves with advanced equipments, which means they would have some technical advantages over the sensor nodes. In this paper, the adversaries are assumed to have the following characteristics:

- The adversaries will have sufficient energy resources, adequate computation capability and enough memory for data storage. On detecting an event, they could determine the immediate sender by analyzing the strength and direction of the signal they received. They can move to this sender's location without too much delay. The adversaries may also compromise some sensor nodes in the network. We assume that the adversaries will never miss any event close to them.
- The adversaries will not interfere with the proper functioning of the network, such as modifying packets, altering the routing path, or destroying sensor devices, since such activities can be easily identified. However, the adversaries may carry out passive attacks, such as eavesdropping of the communications.
- The adversaries are able to monitor the traffic in an area that is important to them and get all of the transmitted messages. However, we assume that the adversaries are unable to monitor the entire network. In fact, if the adversaries could monitor the entire wireless sensor networks, then they can monitor the events directly without relying on the sensor network.

C. Design Goals

Our design goal can be summarized as follows:

- To achieve high packet delivery ratio, our routing protocol should try to avoid packet dropping when alternative paths exist.
- The adversaries should not be able to get the source location information by analyzing the traffic pattern.

- The adversary should not be able to get the source location information if he is only able to monitor certain area of the WSNs and compromise a few sensor nodes.
- Only the sink node is able to identify the source location through the messages received. The recovery of the source location from the received message should be very efficient.
- The routing protocol should maximize the probability that the packet is being delivered to the sink node when adversaries are only able to jam a few sensor nodes.

IV. Proposed Efficient Source-Location Privacy Scheme

In this paper, we propose a scheme that combines routing efficiency and source-location privacy. The basic idea is as follows: when the source node needs to send a message to the sink node, it first transmits the message to an intermediate node, which is chosen from the hierarchical CDS of the network. Then the intermediate node routes the message to the sink node. Each time the source node needs to transmit a message, it will select a random level of CDS. After that, it selects a random node in that CDS and then transmits the message to this intermediate node where the message is forwarded to the sink node. The random selection of the CDS levels guarantees that the routing path is dynamic for each message. Since it is believed that the minimum CDS problem cannot be solved in polynomial time, no adversary is able to predetermine the CDSs. Therefore, it is infeasible for any adversary to perform static routing monitoring and routing trace back attack.

A. Construct the CDS of the original network

The CDS forms the backbone of the network. A CDS has two properties: (i) Any two nodes in the CDS can reach each other by a path that stays entirely within the CDS, and (ii) Every vertex in the graph either belongs to the CDS or is adjacent to a vertex in the CDS.

In this paper, we have adopted the algorithm proposed in [6] in constructing the CDS. The reasons for choosing this algorithm are because the algorithm is completely distributed and each node can run the algorithm simultaneously so that the delay is relative short. In addition, it also adapts to the dynamic change of the network topology.

The general idea of this algorithm is as follows. First, each node in the network will run an algorithm to decide whether it belongs to CDS. If the node has two neighbors that are not neighbors of each other, the node can mark itself as a node in the CDS. The next step is to unmark the marked nodes in the CDS. If the neighbors of one node are covered by other nodes, the node can unmark itself.

In our algorithm, we have introduced a new term, the *hierarchy of CDS*. Previously, the CDS we have discussed is the Level-1 CDS, which is constructed from the original network. So Level-1 CDS is the backbone of the original network. Likewise, Level-2 CDS is constructed from Level-1 CDS and Level-2 CDS is the backbone of Level-1 CDS. The reason why we have introduced CDS of different levels is because nodes in a higher level CDS are much more difficult to predict than nodes in a lower level CDS. In large-scale network, the size of the CDSs reduces rapidly when the level of the CDSs increases, as shown in Table 1, which makes the nodes in the higher-level CDSs even more difficult to predict.

TABLE I
Number of sensor nodes in the CDSs of different levels. The results were derived as an average of 10 random deployment, where the sensor transmission range is assumed to be 50 meters

Sensor Nodes	Field Size	Level-1 CDS	Level-2 CDS	Level-3 CDS	Level-4 CDS	Level-5 CDS
50	100×100	6.2	2.5	0.6	0	0
100	200×200	23.0	13.2	9.2	6.6	4.5
100	300×300	38.1	26.3	19.7	15.4	12.2
200	400×400	75.8	52.1	41.8	34.7	29.4
300	400×400	80.4	53.1	40.3	31.9	26.0
400	500×500	119.7	79.2	61.7	50.9	43.2
500	500×500	122.9	80.6	64.0	53.0	44.2
600	600×600	170.9	114.6	91.5	77.2	66.8
700	700×700	224.9	153.2	122.9	102.6	89.8
800	800×800	288.0	199.9	159.2	135.4	117.4
900	900×900	344.8	243.1	194.9	165.3	145.2
1000	1000×1000	412.8	295.2	241.4	206.2	182.0

It has been proved that the task of finding a minimal CDS with global network information is NP-complete. The unpredictability of CDS, and the intermediate node, can provide high routing path security. This will be analyzed in more detail in later part of the paper.

B. Choosing Intermediate Node from a hierarchical CDS

In the previous section, we have mentioned the hierarchy of CDS in the original network and known that when the level of CDS in the network increases, the node number in the CDS decreases dramatically. Our proposed algorithm is based on choosing intermediate node from a random level of the CDSs.

As mentioned before, in order to provide routing security, the general idea is to distract the adversary from the real source. In our algorithm, we aim to provide random routing path each time the source node sends a message back to the sink node. It is realized by routing the message first to an intermediate node in a CDS of level i, where i is a random number selected from a predetermined range. The intermediate node will serve as a relay node. This intermediate node will further transmit the message back to the sink node.

The intermediate node in this algorithm should satisfy several requirements. First, the chosen intermediate node should not be in the vicinity of the source node. If the algorithm always chooses the intermediate node in the vicinity of the source node, the adversary may be able to trace back to the source node and discover the event. The second requirement is that each time the source node chooses a random intermediate node. In this way, it is possible to provide different routing paths each time the source node transmits a message.

In our scheme, the network will first go through an initialization phase. In this phase, the algorithm for generating the

CDS hierarchy will first be executed in the network. Through this phase, each node in the network will know its CDS level, and its neighbors in each CDS levels. According to the network size, we choose the intermediate node from the CDS of a certain level. If we have decided to choose an intermediate node from Level-3 CDS, for instance, then the nodes in Level-3 CDS will broadcast its information only in Level-3 CDS for message transmission from the intermediate relay node to the sink node.

In fact, we have the following result.

Lemma 1. *In the WSNs, every node is at most i hops away from the i-th level CDS.*

As a result, if the sink node is also located in the sensor networks, then it is also at most i hops away from the i-th level CDS. Every node can choose the intermediate node based on the message it has received. The message will only be broadcast in the corresponding level of CDS. When a node receives a message, the node in the corresponding CDS does not know the hop distance between the actual message and the CDS. So to every node that has received the message, it is blind about the detailed topology structure of the network or the structure of CDS of the network.

C. A Detailed Description of the Proposed Scheme

The proposed scheme includes two major phases: node initialization and CDS generation phase, and message routing through CDS phase.

Node Initialization and CDS Generation:

1) The nodes in the WSNs run the CDS construction algorithm to determine whether it belongs to any levels of the CDS. The algorithm may run several times to construct the hierarchy of CDS in the WSNs.
2) The nodes in the chosen level of the CDS broadcast their information to the other nodes in the same CDS level. After this step, each node in the CDS of this level will know the topology.
3) The nodes will broadcast the structure of the CDS to its neighbor nodes. The neighbor nodes receiving such information will further forward the message to other nodes that haven not received the message. After this step, all nodes in the network will store the information of the nodes from which they might choose the intermediate nodes from. When the nodes broadcast the information, a node may receive several different messages. The node adopts the rule "First Come, First Accepted." It accepts the first received message about the CDS topology and abandons the later messages of the same kind.

Intermediate Node Selection and Message Routing:

1) Choose a random level i of the hierarchical CDS for the intermediate node to be selected.
2) Select an intermediate node from the i-th CDS that is close to the message source node.
3) When the source node needs to route a message to the sink node, it first sets the chosen intermediate node as the destination. After the message arrives at the intermediate node, the intermediate node further forwards the message to the sink node.

V. Security Analysis

Our proposed routing algorithm includes two routing phases. First, the source node sends the message to the intermediate destination. Then the intermediate destination forwards the message to the sink node.

We will first analyze that the proposed routing to a random intermediate node in the hierarchical CDS can provide source-location privacy when routing message to the sink node. The source node needs to first randomly select a level i of the CDS for message to be forwarded, it has no control over the intermediate node. Since finding a minimal hierarchical CDS with global network information is NP-complete. The adversary is able to predict the level of the CDS, the selection of the CDS and the intermediate node a priori.

If an adversary tries to trace back the source-location from the message in the route through which the message is being transmitted, then the adversary will be led to the randomly selected intermediate node to the best extend, instead of the real message source. Since the intermediate node is randomly selected in the hierarchical CDS for each data message, the source node is most likely to select a different routing path for each message transmission. In fact, the probability that the adversaries will receive the messages from one source node continuously is virtually impossible. In this case, if the adversary performs trace back analysis, it will not get constant message flow. In fact, it is also very difficult for the adversary to distinguish messages from different sources and collect multiple messages from one message source that can be used to perform routing trace back attack and to recover the message source node. On the other hand, even if one intermediate node's location is discovered by the adversaries, the source-location is still i hops away from the real source node.

Unlike the directed walk used in random walk, our protocol does not leak side information to the adversaries, since the intermediate node is determined before each data message is transmitted by the source-location. Therefore, our proposed protocol can provide the local source-location privacy.

VI. Performance Analysis and Simulation Results

In this section, to evaluate the performance of the schemes proposed, extensive simulations have been conducted using numerical analysis and ns-2 on RedHat Linux system.

A. Statistical Results of CDS construction

We provide statistical results of the hierarchical CDS sizes in the network. We have chosen different network scales and then run the CDS construction algorithm to see how the CDS sizes change when the hierarchical levels increase.

In this analysis, the sensing range of each node is assumed to be 50 meters. The number of nodes spread in the sensor networks varies from 50 to 1000. The area of the network size varies from 100×100 meters to 1000×1000 meters. For each

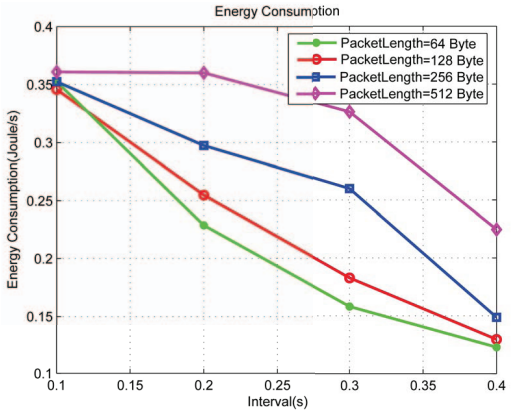

Fig. 1. Energy efficiency of the proposed 2-step routing

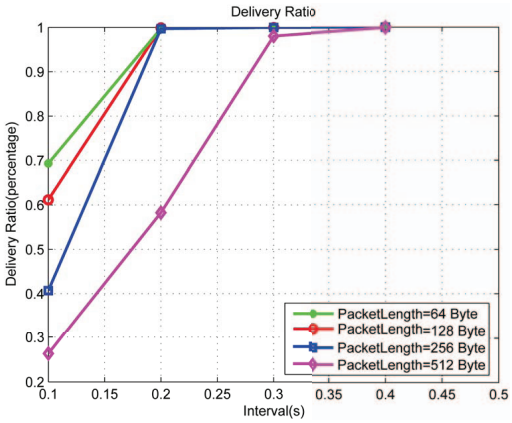

Fig. 2. Message delivery ratio of the proposed 2-step routing

scenario, the simulation has been run for ten times. This result given in Table 1 is the average of these ten simulations.

B. Network Simulation

We provide simulations to measure the routing efficiency and energy consumption, the packet delivery ratio, and message delay using ns-2 on RedHat Linus system. In our simulations, we have scattered 200 sensor nodes in an area of 450 meters by 450 meters. The total number of message source nodes is set to 15.

For these simulations, we set multiple message lengths: 16 bytes, 32 bytes, 64 bytes, 128 bytes and 256 bytes. We also allow the message transmission to vary at multiple time intervals: 0.1s, 0.2s, 0.3s, 0.4s. The simulation results are given in Fig. 1 to Fig. 3.

VII. Conclusions

In this paper, we have proposed an efficient 2-step source-location privacy routing scheme. That is the message is first routed to a node in a hierarchical CDS before it is forwarded to the sink node. This scheme takes advantages of the hierarchical CDS for routing efficiency. It also guarantees a high message delivery ratio and a short message delay. We also provide simulation results to demonstrate the proposed theoretical results.

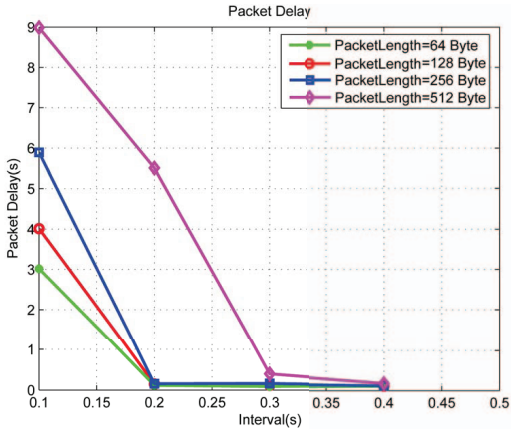

Fig. 3. Message delay of the proposed 2-step routing

The proposed scheme can be widely used in many scenarios to provide message source-location privacy without introducing a high communication overhead.

Acknowledgement

This research is was supported in part by the NSF under grants CNS-0845812, CNS-0848569 and CNS-1050326.

References

[1] Y. Li and J. Ren, "Preserving source-location privacy in wireless sensor networks," in *Proceedings of IEEE SECON 2009*, (Rome, Italy.), June 22-26, 2009.

[2] Y. Li and J. Ren, "Source-location privacy through dynamic routing in wireless sensor networks," in *Proceedings of IEEE INFOCOM 2010*, (San Diego, USA.), March 15-19, 2010.

[3] B. Das and V. Bharghavan, "Routing in ad-hoc networks using minimum connected dominating sets," in *IEEE International Conference on Communications (ICC)*, pp. 376–380, 1997.

[4] S. Guha and S. Khuller, "Approximation algorithms for connected dominating sets," *Algorithmica*, vol. 20, pp. 374–387, April 1998.

[5] J. Wu and F. Dai, "A generic distributed broadcast scheme in ad hoc wireless networks," *IEEE Transactions on Computers*, vol. 53, pp. 1343–1354, October 2004.

[6] J. Wu, S. Yang, and F. Dai, "Iterative local solutions for connected dominating set in ad hoc wireless networks," *IEEE Transactions on Transactions on Computers*, vol. 57, May 2008.

[7] J. Wu and H. Li, "On calculating connected dominating sets for efficient routing in ad hoc wireless networks," in *Proceedings of the 3rd international workshop on Discrete algorithms and methods for mobile computing and communications*, pp. 7–14, 1999.

[8] F. Dai and J. Wu, "An extended localized algorithm for connected dominating set formation in ad hoc wireless networks," vol. 15, pp. 908–920, October 2004.

[9] P.-J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed construction of connected dominating set in wireless ad hoc networks," in *Mobile Networks and Applications*, pp. 141–149, 2002.

[10] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, February 1981.

[11] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pp. 88–93, 2004.

[12] J. Ren and J. Wu, "Survey on anonymous communications in computer networks," *Computer Communications*, vol. 33, no. 4, pp. 420–431, 2010.