Routing-Based Source-Location Privacy Protection in Wireless Sensor Networks

Yun Li, Leron Lightfoot, Jian Ren
Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824
Email: {liyun1, lightf13, renjian}@egr.msu.edu

***Abstract*—Wireless sensor networks (WSN) have the potential to be widely used in many areas for unattended event monitoring. Mainly due to lack of a protected physical boundary, wireless communications are vulnerable to unauthorized interception and detection. Privacy is becoming one of the major issues that jeopardize the successful deployment of wireless sensor networks. While confidentiality of the message can be ensured through content encryption, it is much more difficult to adequately address the source-location privacy. For WSN, source-location privacy service is further complicated by the fact that the sensor nodes consist of low-cost and low-power radio devices. Therefore, computationally intensive cryptographic algorithms (such as public-key cryptosystems) and large scale broadcasting-based protocols are not suitable for WSN. In this paper, we propose a scheme to provide both content confidentiality and source-location privacy through routing to randomly selected intermediate nodes. While being able to provide source-location privacy for WSN, our simulation results also demonstrate that the proposed scheme is efficient and can be used for practical applications.**

I. INTRODUCTION

Wireless sensor networks have been envisioned as a technology that has great potential to be widely used in both military and civilian applications. Sensor networks rely on wireless communication, which is by nature a broadcast medium that is more vulnerable to security attacks than its wired counterpart due to lack of a physical boundary. In the wireless sensor domain, anybody with an appropriate wireless receiver can monitor and intercept the sensor network communications. The adversaries may use expensive radio transceivers and powerful workstations to interact with the network from a distance since they are not restricted to using sensor network hardware. It is possible for the adversaries to identify the message source or even identify the source location, even if strong data encryption is utilized.

Location privacy is an important security issue. Lack of location privacy can expose significant information about the traffic carried on the network and the physical world entities. While confidentiality of the message can be ensured through content encryption, it is much more difficult to adequately address the source-location privacy. Privacy service in WSN, is further complicated since the sensor nodes consist of only low-cost and low-power radio devices and are designed to operate unattended for a long period of time. Battery recharging or replacement may be infeasible or impossible. Therefore, computationally intensive cryptographic algorithms, such as public-key cryptosystems, and large scale broadcasting-based protocols, are not suitable for WSN. This makes privacy preserving communications in WSN an extremely challenging research task. To optimize the sensor nodes for the limited capabilities and the application specific nature of the networks, traditionally, security requirements were largely ignored. This leaves the WSN vulnerable to network security attacks. In the worst case, an adversary 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 provides source-location privacy through routing. For each message transmission, the message source first randomly selects an intermediate node in the sensor domain, then transmits the data packet to the randomly selected intermediate node. From this intermediate node, the message will be forwarded to the destination node. This scheme could provide local source-location privacy or global location privacy, depending on how the intermediate nodes are selected. We also carried out simulation to demonstrate that the proposed scheme is very efficient and can be used for practical applications, while being able to provide source-location privacy for WSN.

The remainder of this paper is organized as followed: In Section II, the related works are reviewed. The system model and design goals are described in Section III. Section IV details the proposed source-location privacy scheme. Security analysis and performance analysis are provided in Section V and Section VI, respectively. We conclude in Section VII. The acknowledgments are provided in Section VIII.

II. RELATED WORKS

In the past two decades, a number of source-location private communication protocols have been proposed [1]–[13], which originated largely from Chaum's mixnet [14] and DC-net [15]. The mixnet family protocols use a set of "mix" servers that mix the received packets to make the communication source (including the sender and the recipient) ambiguous. The DC-net family protocols [3], [4], [15] utilize secure multiparty

978-1-4244-3355-1/09/$25.00©2009 IEEE

computation techniques. However, both approaches require public-key cryptosystems, which are not suitable for WSN.

In [7], [8], Deng et al. proposed to address the location privacy of base station through multi-path routing and fake message injection. In their scheme, every node in the network has to transmit messages at a constant rate. Another base location privacy scheme is introduced in [16], which involves location privacy routing and fake message injection. This proposed scheme is interested in protecting the destination location privacy, while our scheme cares about source location privacy protection.

In [9], [10], source location privacy is provided through broadcasting that mixes valid messages with dummy messages. The main idea is that each node needs to transmit messages consistently. Whenever there are no valid messages, the node has to transmit dummy messages. The transmission of dummy messages consumes significant amount of sensor energy while increasing the network collisions and decreasing the packet delivery ratio. Therefore, these schemes are not quite suitable for large sensor networks.

Routing based protocols can also provide source-location privacy. The main idea is to prevent the adversaries from tracing back to the source-location through traffic monitoring and analysis. A phantom routing protocol is introduced in [11], [12]. Phantom routing involves two phases: a random walk phase and a subsequent flooding/single path routing phase. In the random walking phase, the message from the real source will be routed to a phantom source along a random path or a designed directed path. The phantom source is expected to be far away from the real source, which will make the real source's location hard to be traced back by the adversaries. However, theoretical analysis shows that if the message is routed h hops randomly, then it is highly possible that the distance between the phantom source and the real source is within $h/5$. To solve this problem, directed walk was proposed in [11]. Directed walk can be achieved either through section-based directed random walk or hop-based directed random walk. Let's take the section-based directed walk as an example. The source node first randomly determines a direction that the message will be sent. This direction information is stored in the header of the message. Every forwarder on the random walk path will forward this message to a random neighbor in the same direction determined by the source node. In this way, the directed random walk will determine a phantom source that is away from the real source. Unfortunately, once the message is captured on the random walk path, the adversaries will be able to get the direction information stored in the header of the message. Therefore, the exposure of direction information decreases the complexity for adversaries to trace back to the true message source in the magnitude of 2^h. [13] proposed to implement random walk from both the source node and the SINK node. Different from the directed walk, Bloom Filter is proposed to store all the visited nodes in the network for each message to prevent the adversaries from hopping back. However, for large scale sensor networks, this is not realistic.

III. Models and Design Goals

A. The System Model

Our system is similar to the explanatory Panda-Hunter Game that was introduced in [11], [17]. In this Panda-Hunter Game, a sensor network is deployed to continuously monitor activities and locations of the animals in a wild animal habitat.

As soon as a panda is discovered, the corresponding source node in the nearby area will observe and report data periodically to the SINK node. However, the information should be kept unavailable to the illegal hunters who may try to track and locate the panda. Our goal is to make it infeasible for the adversaries to determine the location of the panda by analyzing the traffic pattern and messages transmitted through the network. We made the following assumptions about our system:

- The network is evenly divided into small grids. The sensor nodes in each grid are all fully connected. In each grid, there is one header node responsible for communicating with other header nodes nearby. The whole network is fully connected through multi-hop communications.
- The SINK node is the destination location that data messages will be transmitted to. The information of the SINK node is public. On detecting an event, a sensor node will generate and send messages to the SINK node through a multi-hop routing path.
- The content of each message will be encrypted using the shared secret key between the node/grid and the SINK node. The encryption operation is beyond the scope of this paper.
- The sensor nodes are assumed to know their relative location. We also assume that each sensor node has the knowledge of its adjacent neighboring nodes. The information about the relative location of the sensor domain may also be broadcasted through this network for routing information update [18]–[20].
- The key management, including key generation, key distribution and key update, is beyond the scope of this paper. However, the interested readers are referred to references such as [21]–[24].

B. The Adversaries Model

Due to the high profits related to panda hunting, the adversaries would try their best to equip themselves with advanced equipments. Therefore, they typically 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 resource, 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 also assume that the adversaries will never miss any event when they are close to the event.

- 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 goals can be summarized as followed:

- The adversaries should not be able to get the source-location information by analyzing the traffic pattern.
- The adversaries should not be able to get the source-location information even if they are able to monitor certain area of the sensor network and compromise a few network 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 length of each message should be as short as possible to save the previous sensor node power. This is because that on average, transmission of one bit consumes about as much power as executing 800-1000 instructions [25].

IV. Routing-based Source Location Privacy Scheme

We have analyzed that phantom routing will leak direction information to the adversaries while the messages are forwarded to the phantom sources. To prevent this, we proposed routing through a randomly selected intermediate node (RRIN) [26].

In this scheme, the message source first randomly selects an intermediate node at the sensor domain based on the relative location of the sensor node. The intermediate node is expected to be far away from the real source node so that it is difficult for the adversaries to get the information of the real source from the intermediate node selected.

Since we assume that each sensor node only has knowledge of its adjacent nodes. The source node has no accurate information of the sensor nodes more than one hop away. In particular, the randomly selected intermediate node may not even exist. However, the relative location can guarantee that the message packet will be forwarded to the area of the intermediate node. The last node in the routing path adjacent to the intermediate node should be able to tell whether such a randomly selected intermediate node exists or not. In the case that such a node does not exist, this node will become the intermediate node. The intermediate node then routes the received message to the destination node.

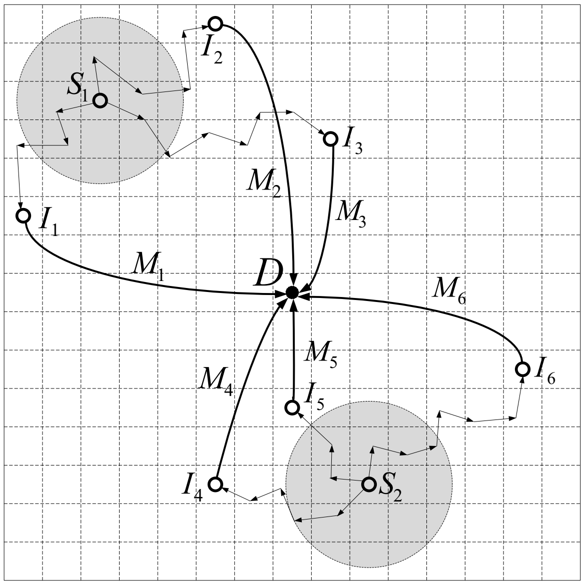


Fig. 1. Illustration of the two-phase routing

Suppose the source node is located at the relative location (x_0, y_0), to transmit a data message, it first determines the minimum distance, d_{min}, that the intermediate node has to be away from the source node. We denote the distance between the source node and the randomly selected intermediate node as d_{rand}. Then we have $d_{rand} \geq d_{min}$.

Whenever the source node wants to generate a d_{rand}, it will first generate a random number x. The value of this random variable is normally distributed with mean 0 and variance σ^2, i.e., $X \sim N(0, \sigma)$. Then the source node can calculate d_{rand} as followed:

$$d_{rand} = d_{min} \times (|x| + 1).$$

Therefore, the probability [27] that d_{rand} is located in the interval $[d_{min}, \rho d_{min})$ is:

$$2\varphi_{0,\sigma^2}(\rho - 1) - 1 = 2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\rho-1)^2}{2\sigma^2}} - 1 = 2\varphi\left(\frac{\rho - 1}{\sigma}\right) - 1,$$

where ρ is a parameter larger than 1, φ_{0,σ^2} is the probability density function which is the Gaussian function [28].

If we choose σ to be 1.0, then the probability that d_{rand} falls within the interval $[d_{min}, 2d_{min})$ will be $2\Phi(\frac{1}{1}) - 1 = 0.6827$. The probability that d_{rand} is in the interval $[d_{min}, 3d_{min})$ will be $2\Phi(\frac{2}{1}) - 1 = 0.9545$.

After d_{rand} is determined, the source node randomly generates an intermediate node located at (x_d, y_d) that satisfies:

$$d_{rand} = \sqrt{(x_d - x_0)^2 + (y_d - y_0)^2} \geq d_{min}.$$

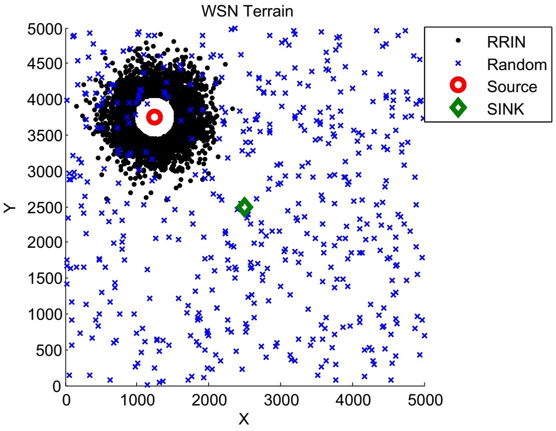


Fig. 2. Distribution of the intermediate nodes

Upon receiving data message, the intermediate node forwards the message to the SINK node.

An example is given in Fig. 1, where $S1, S2$ indicate two source nodes in the network, D indicates the Destination node and I_1 to I_6 are six intermediate nodes. The selection of d_{rand} guarantees that none of the intermediate nodes will be in the shaded areas. Then I_1 to I_6 will forward these messages M_1 to M_6 to the destination node, respectively.

Unlike the directed walk proposed in phantom routing, in our proposed RRIN scheme, the selection of the intermediate node is entirely random. Therefore, it does not have the security drawbacks of phantom routing as discussed before.

However, in this scheme, the possibility of being selected as an intermediate node for a sensor in the WSN is proportional to the distance between this sensor and the source node. Therefore, for large scale sensor networks, the intermediate nodes tend not to be too far away from the source node. In other words, the intermediate nodes are highly likely to concentrate in a circle area centered at the source node. We carry out a small simulation to illustrate this. As shown in Fig.2, we generate 500 intermediate nodes in the target WSN area of size 5000×5000 meters according to the RRIN algorithm above with σ equals to 1. We could see that all the intermediate nodes are located not too far away from the source node and the distribution looks like a circle. Therefore, nearly all the messages generated by this source node would be forwarded to the SINK node from the intermediate nodes in this circle. The adversaries can be pretty sure that the source node is located at the left top corner of this target WSN terrain. So for large scale sensor networks, RRIN could only provide local location privacy.

In order to provide global location privacy over the sensor network, the selection of intermediate nodes have to be totally random, i.e., every sensor node in the networks has the same probability of being selected as the intermediate node for any source nodes. We also generate 500 intermediate nodes totally random, and the distribution for them is shown in Fig.2. In this situation, the intermediate nodes are evenly distributed in the WSN terrain and the messages could be forwarded to the SINK from all possible directions.

However, if the selection is totally random, then some intermediate nodes' location could be very close to the real source node. Once such an intermediate node is located by the adversaries, the source's location will also be exposed. To prevent this from happening, the location of the intermediate nodes should be at least d_{min} away from the real source node.

V. Security Analysis

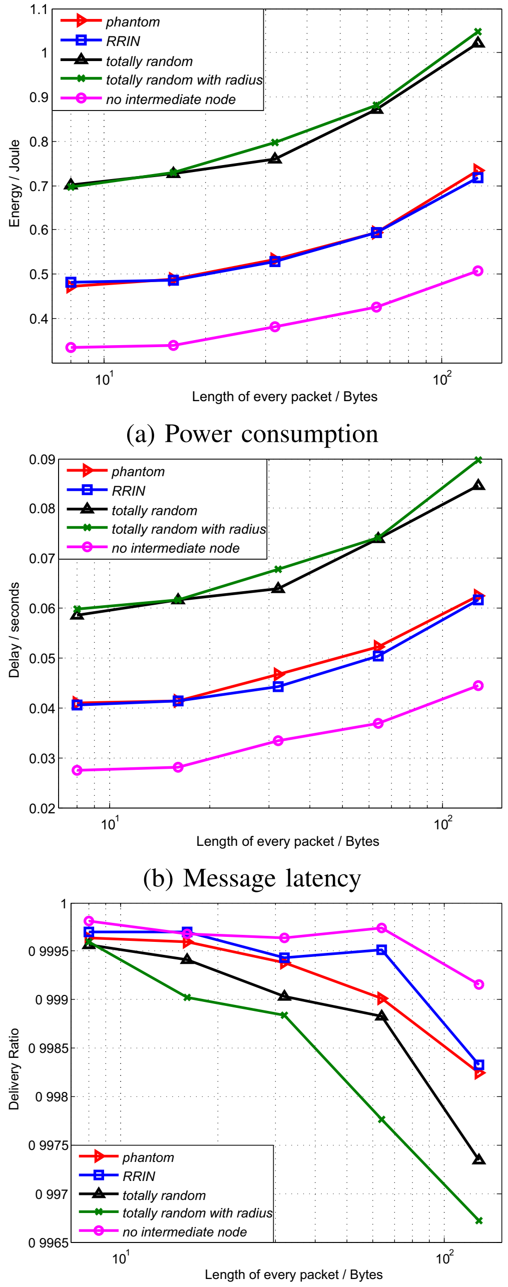
In this section, we will analyze that the proposed scheme can provide source-location privacy. We assume the adversaries are unable to monitor the entire sensor area of the source node. If the adversaries can monitor all real events, it makes little sense for the adversaries to monitor the routing of the sensor node.

In our scheme, the intermediate node is selected by the source node in a totally random way. From probability point of view, every node away from the source node, except nodes in d_{min} range, has the same possibility of being selected as the intermediate node. In fact, the source node selected intermediate node may not even exist since according to our assumption, the source node does not have full knowledge of the sensor node more than one hop away. Therefore, based on our assumption, it is impossible for the adversaries to trace back and identify the real message source based on an individual traffic monitoring. This is because the probability for multiple events from the same source to use the same routing path and intermediate node is very unlikely for large sensor networks.

If an adversary could carry out traffic analysis and could trace back to the message sender by analyzing the message packet in the route through which the packet is being transmitted to the destination, he/she would 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 for each data message, the probability that the adversaries will receive the messages from one source node continuously is highly unlikely. As shown in Fig. 1, the intermediate nodes I_1, I_2, I_3 forward messages to the destination node D using different routing paths, respectively. If the adversaries receive M_2 forwarded by I_2, it would be led to I_2. However, the next intermediate node I_3 is far from I_2, so the adversaries could not receive M_3.

Even if one intermediate node's location is discovered by the adversaries, the source-location is still well protected because the locations of the intermediate nodes are at least d_{min} away from the real source node.

Unlike the directed walk used in random walk, our protocol does not leak direction information to the adversaries. Since the intermediate node is determined before each data message is transmitted by the source-location, the data message carries no observable direction information of the message source-location in its content. Therefore, our proposed protocol can provide the source-location privacy.



(a) Power consumption

(b) Message latency

(c) Message delivery ratio

Fig. 3. Performances of routing by single-intermediate node

VI. Performance Analysis and Simulation Results

To evaluate the performances of the schemes proposed, we have done some simulations using NS2 on Linux system. In the simulation, 400 nodes are distributed in a square target area of size 3360 × 3360 meters, while the SINK node is located in the center of the network. We set hop count of directed walking of phantom routing to be four, which on average the phantom source was found to be 506.12 meters away from the real source. For RRIN scheme, the average distance from the source node to the intermediate nodes is 529.14 meters. We also illustrate the performances of the totally randomly selected intermediate nodes. In the figures, 'total random' means the selection of the intermediate nodes are completely random without considering d_{min}. 'total random with radius' means the intermediate nodes are at least d_{min} from the source node, while d_{min} equals to 480 meters in this simulation.

Simulation results are provided in Fig.3. From the figures, it it reasonable to observe that sending the massages to the destination directly without intermediate node gives the best performances. The performances of RRIN and phantom are of the same level, while RRIN is better in location privacy protection. If the intermediate nodes selection is made totally random, the performances are not as good as the others. The performances of the totally random intermediate nodes selection with the d_{min} constraint are the worst. This is reasonable, because usually you have to make tradeoff between the level of security and the performance.

VII. Conclusions

Source location privacy is critical to the successful deployment of wireless sensor networks. In this paper, we first introduced a randomly selected intermediate node (RRIN) scheme for local source location privacy protection. Then based on this scheme, we proposed global source location privacy protection scheme. For each of these schemes, we carried out simulations to evaluate the performances and compared their performances with others. Simulation results demonstrate that the proposed schemes can achieve satisfiable performances in energy consumption, message delivery latency while assuring high message delivery ratio.

VIII. Acknowledgments

This research was partially supported by the US National Science Foundation under grants CNS-0716039, CNS-0848569, and CNS-0845812.

References

[1] L. Ahn, A. Bortz, and N. Hopper, "k-anonymous message transmission," in *Proceedings of the 10th ACM conference on Computer and Communications Security*, (Washington D.C., USA.), pp. 122–130, 2003.

[2] A. Beimel and S. Dolev, "Buses for anonymous message delivery," *J. Cryptology*, vol. 16, pp. 25–39, 2003.

[3] P. Golle and A. Juels, "Dining cryptographers revisited," in *Advances in Cryptology - Eurocrypt 2004*, LNCS 3027, pp. 456–473, 2004.

[4] S. Goel, M. Robson, M. Polte, and E. G. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Tech. Rep. 2003-1890, Cornell University, Ithaca, NY, February 2003.

[5] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous connections and onion routing," *IEEE J. on Selected Areas in Coomunications, Special Issue on Copyrigh and Privacy Protection*, vol. 16, no. 4, pp. 482–494, 1998.

[6] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for web transaction," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66–92, 1998.

[7] J. Deng, R. Han, and S. Mishra, "Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks," in *DSN '04: Proceedings of the 2004 International Conference on Dependable Systems and Networks*, (Washington, DC, USA), p. 637, IEEE Computer Society, 2004.

[8] J. Deng, R. Han, and S. Mishra, "Countermeasures against traffic analysis attacks in wireless sensor networks," *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, pp. 113–126, Sept. 2005.

[9] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in *WiSec '08: Proceedings of the first ACM conference on Wireless network security*, (New York, NY, USA), pp. 77–88, ACM, 2008.

[10] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards statistically strong source anonymity for sensor networks," *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 51–55, April 2008.

[11] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source-location privacy in sensor network routing," *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pp. 599–608, June 2005.

[12] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, (New York, NY, USA), pp. 88–93, ACM, 2004.

[13] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks.," in *IPDPS*, IEEE, 2006.

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

[15] D. Chaum, "The dinning cryptographer problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, vol. 1, no. 1, pp. 65–75, 1988.

[16] Y. Jian, S. Chen, Z. Zhang, and L. Zhang, "A novel scheme for protecting receiver's location privacy in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 7, pp. 3769–3779, October 2008.

[17] http://www.panda.org/.

[18] Y. Zhang, W. Liu, Y. Fang, and D. Wu, "Secure localization and authentication in ultra-wideband sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, pp. 829–835, April 2006.

[19] "Localization for mobile sensor networks," in *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, (New York, NY, USA), pp. 45–57, ACM, 2004.

[20] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "Tps: a time-based positioning scheme for outdoor wireless sensor networks," *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 4, pp. 2685–2696 vol.4, March 2004.

[21] H. Chan and A. Perrig, "Pike: peer intermediaries for key establishment in sensor networks," *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1, pp. 524–535 vol. 1, March 2005.

[22] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J. Tygar, "SPINS: Security protocols for sensor networks," in *Seventh Annual International Conference on Mobile Computing and Networks (MobiCOM 2001)*, (Rome, Italy), July 2001.

[23] P. Traynor, R. Kumar, H. Choi, G. Cao, S. Zhu, and T. La Porta, "Efficient hybrid security mechanisms for heterogeneous sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 6, pp. 663–677, June 2007.

[24] S. Zhu, S. Setia, and S. Jajodia, "Leap: efficient security mechanisms for large-scale distributed sensor networks," in *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, (New York, NY, USA), pp. 62–72, ACM, 2003.

[25] J. Hill, R. Szewczyk, S. H. A. Woo, D. Culler, and K. Pister, "System architecture directions for networked sensors," in *Proceedings of ACM ASPLOS IX*, November 2000.

[26] J. Ren and Y. Li, "Routing-based source-location privacy in wireless sensor networks," ICC 2009 Communication and Information Systems Security Symposium, 2009.

[27] Wikipedia, "Normal distribution." http://en.wikipedia.org/wiki/Normal_distribution.

[28] S. M. Stigler, *Statistics on the Table*. Harvard University Press. chapter 22 (History of the term "normal distribution".