Privacy preservation for WSN: A Survey

Prakhar Gupta
M.tech Scholar Department Of CSE,
Maulana Azad National Institute of Technology
Bhopal, Madhya Pradesh, India

Meenu Chawla
Associate Professor Department Of CSE,
Maulana Azad National Institute of Technology
Bhopal, Madhya Pradesh, India

ABSTRACT
This paper presents a review of privacy-preserving techniques for wireless sensor networks (WSN). A lot of work has been done to enhance efficient use of power and resources in wireless sensor networks with the help of specific routing algorithms and different layers' protocols. As the popularity and uses of wireless sensor network increase, privacy of individuals is of very high demand. In this paper two main categories of privacy preserving techniques for WSN have been presented, data-oriented and context-oriented. Some interesting open challenges for future research have also been introduced. This paper should provide some fruitful help for further research in privacy preservation for WSN.

Keywords
Wireless sensor network, Privacy, Data mining.

1. INTRODUCTION
In recent years, energy, computational power and resource utilization have been well defined and widely studied [1] for wireless sensor network (WSN). With growth in the popularity and deployment of pervasive computing, privacy of individuals is slowly steaming away. In the early days of WSN privacy issue came up as a second priority but now people are showing more concern about the privacy issue. Currently WSN are widely used in many applications such as [2]:

- Military applications
  - Monitoring, tracking and surveillance of borders
  - Nuclear, biological and chemical attack detection
  - Battle damage assessment
- Environmental applications
  - Flood and oceans detection
  - Forest fire detection
  - Precision agriculture
- Health applications
  - Drug administration
  - Remote monitoring of physiological data
  - Tracking and monitoring doctors and patients inside a hospital
- Home applications
  - Automated meter reading
  - Home automation
  - Instrumented environment
- Commercial applications
  - Monitoring vibration for a building structure
  - Monitoring traffic flow and road condition
  - Vehicle tracking and detection

WSN are capable to automatically collect data through efficient deployment of sensor devices. They offer great benefits to users but also exhibit significant potential for misuse. Particularly relevant concerns are privacy problems [4]. Adversaries can use even seemingly innocuous data to derive sensitive information. For example, in the famous "panda-hunter problem" [3], the hunter can find out the position of pandas by monitoring the traffic or consider a network used for noise and sound monitoring, where people can be tracked based on their voice recognition.

So a responsible design of new technologies should take privacy risks into account. In this paper a review of existing privacy-preserving techniques in WSN has been presented. Focus is on two main categories of privacy-preserving techniques; data-oriented and context-oriented. Data-oriented privacy protections target privacy of data collected by sensor nodes and queries posted in network. In context- oriented privacy, location privacy and temporal privacy has been discussed. At last some open issues in this field have been mentioned with the aim that it will help for future research in privacy preservation in WSN.

In next section an introduction to the related work is given while in Section 3 different privacy-preserving techniques such as data-oriented and context-oriented have been discussed. Some different privacy-preserving techniques have been compared and outline to some open challenges for future research is given in Section 4, followed by concluding remarks in Section 5.

2. RELATED WORK
Privacy issue is widely explored in the field of database, networks, data mining and other field. Lots of techniques are proposed for privacy preservation such as: Cryptographic security [7], K-anonymity [9], Random perturbation technique [8]. These techniques are use to protect data when it flows from one node to other.

A large numbers of attacks are possible in WSN such as Denial of Service attacks, The Sybil attacks, Traffic Analysis attacks, Node Replication attacks, attacks against Privacy, and Physical attacks. Lot of work has done to overcome these attacks [6]. Our area of interest is on attacks against Privacy. Privacy attacks can be further classified into two broad categories data oriented attacks and context oriented attacks. Figure 1 shows the types of privacy-preserving problems in WSN.

3. PRIVACY-PRESERVING TECHNIQUES
This section deals with two main branches of privacy-preserving techniques data-oriented privacy and context-oriented privacy respectively.

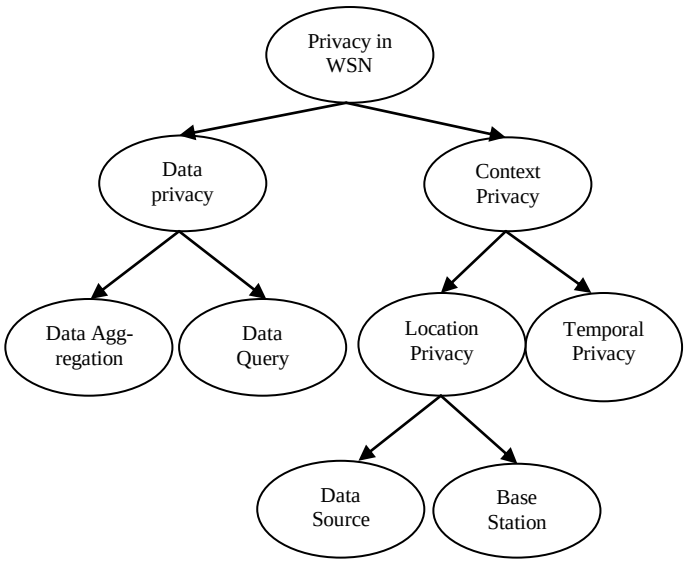


Figure 1 Classification of privacy-preserving problems for WSN [5].

3.1 Data-oriented privacy preserving techniques

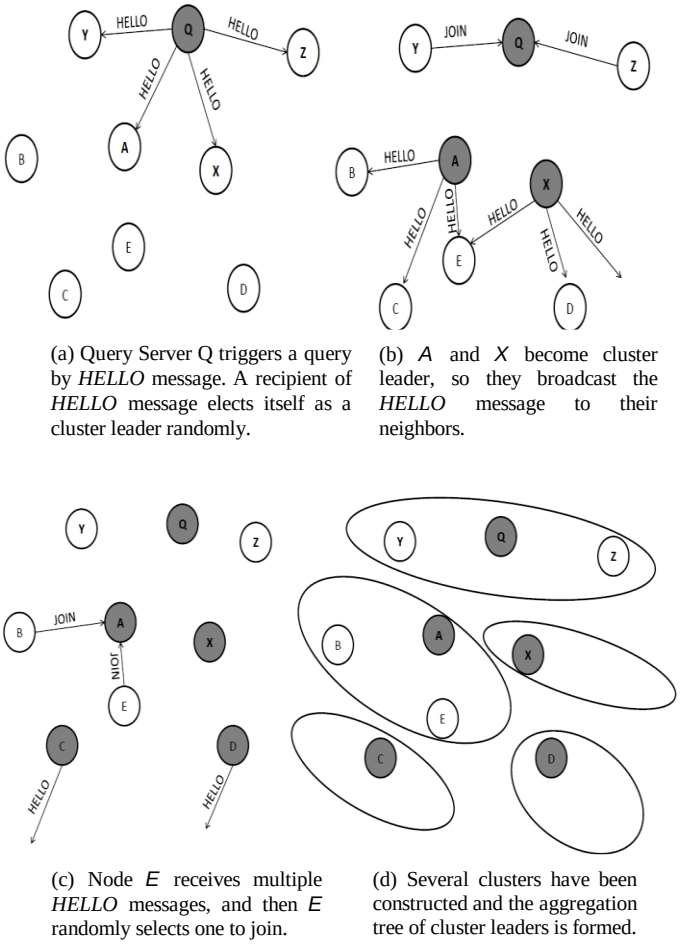
In data-oriented privacy preserving technique main focus is on data which is collected and send to the sink. In the above mentioned example of noise and sound monitoring system; adversary can use sensor data for matching and being able to trace any person.

Two main approaches for data-oriented privacy preserving techniques are data aggregation and private data query.

3.1.1 Data aggregation

Protection to data can be guaranteed by secret keys between sensor nodes [10, 11]. In this approach an intermediate node has to decrypt the received data, then aggregate the data according to the corresponding aggregation function, and finally encrypt the aggregated result before forwarding it. This approach requires lots of computation and due to this the overhead increase. WSNs have limited power so this approach is fairly expensive for WSNs.

Cluster-based Private Data Aggregation (CPDA) is more effective approach for data aggregation [12]. As Figure 2 shows, the first step of CPDA Query Server Q triggers a query by HELLO message. A recipient of HELLO message elects itself as a cluster leader randomly. Some of the nodes become cluster leader, so they broadcast the HELLO message to their neighbors then it decides to join one of the clusters by broadcasting a JOIN message. As this procedure goes on, multiple clusters are constructed. The second step of CPDA is the intermediate aggregations within clusters with the help of encryption keys. Computation overhead reduces in this approach as compared to previous approach but still high.



(a) Query Server Q triggers a query by *HELLO* message. A recipient of *HELLO* message elects itself as a cluster leader randomly.

(b) *A* and *X* become cluster leader, so they broadcast the *HELLO* message to their neighbors.

(c) Node *E* receives multiple *HELLO* messages, and then *E* randomly selects one to join.

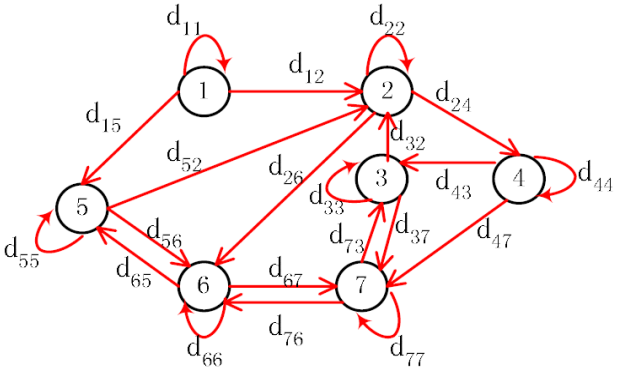
(d) Several clusters have been constructed and the aggregation tree of cluster leaders is formed.

Figure 2 Formation of clusters in CPDA approach [12].

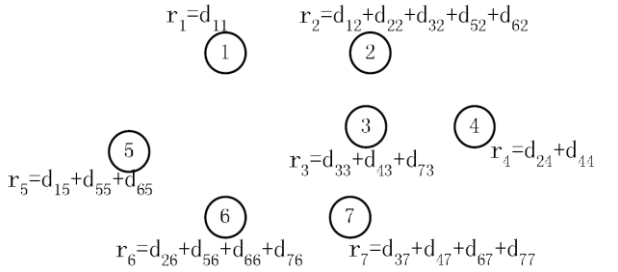
Slice-Mix-Aggregate (SMART) is another approach where computational overhead is reduced at the cost of slightly increase communication overhead [12]. As the name suggest Slice-Mix-Aggregate approach works with three steps. In first step data is sliced into J pieces. One piece is kept by the node itself and rest J-1 pieces are encrypted and sent to neighbor nodes. In second step each node has slices of different node data. Node decrypts all data and sums up all of them. Finally in third step all nodes aggregate the data and the result is send to the query server. All three steps are shown in Figure 3, assuming J=3 and hop count (h) = 1. Computation overhead reduces in this approach in compare to CPDA but computational overhead slightly increased.

3.1.2 Data Query

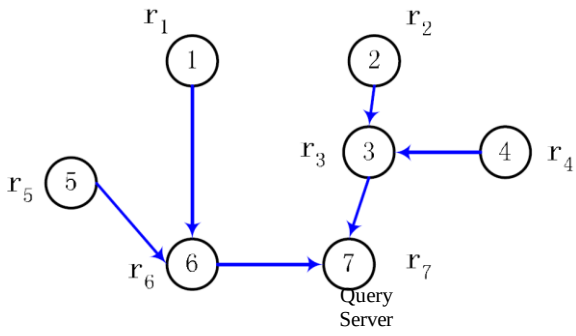
Other privacy problem is due to data query. A query protection technique was proposed in [13]. Instead of enlarging the query regions, this technique intends to disconnect the mapping between a user identity and the query issued by the user. With this technique, a user is able to access data in a WSN after purchasing a certain amount of tokens from the WSN owner. Such a token not only controls access to the sensed data, but also hides the user identity and thereby guarantees query privacy.



(a) Slicing.



(b) Mixing.



(c) Aggregation

Figure 3 Three steps in SMART [12].

3.2 Context-oriented privacy protection techniques

Although privacy of data can be achieved through effective protection techniques, sensor events are so sensitive that it is needed to protect all information surrounding these sensor events. The context-oriented information includes information on source location, sink location and timing of events. Adversary can obtain this type of information by using traffic analysis techniques [14]. Context-oriented privacy is summarized in the following two subsections.

3.2.1 Location privacy

Location privacy plays an important role in WSN such as, location of special sensor node, data source and the base station. The famous Panda-Hunter Problem [3] in WSN sensor nodes are used to locate pandas in their local habitat. Adversary can find out location of sensor node that monitors the panda, successfully localize and capture the panda. Similarly, the location of base station is needed by adversary to mount different network attacks on the base station. Location privacy is further classified into two categories; privacy for location of data source and location of base station.

3.2.1.1 Location privacy for data source

Flooding [17] has been used to preserve data source location. In this technique each node broadcasts data packets to their neighbor node to forward that data packet towards the sink. This technique required lots of energy due to broadcasting of data packet. Probabilistic flooding [17] is proposed to overcome this problem. This technique is more energy efficient. A node forwards a message with some probability so that only a subset of nodes is involved in forwarding. Phantom flooding [17] technique is also been proposed to reduce energy consumption. This technique is carried out with two steps. In the first step a message takes random or directed walk to a random node in the network. Then in the second step, the message is flooded by the phantom node into the network to reach the base station. In comparison to previous two techniques higher level of source location privacy can be achieved by Phantom flooding.

In order to improve the phantom flooding arguments are proposed in [18, 19]. They argue that both pure random walk and directed random walk have major drawbacks. Pure random walk may stick around the source node, whereas in directed random walk, an adversary can obtain sensitive information from source node. Instead of random walk, they propose to route a message through single or multiple randomly selected intermediate nodes from the network. However, the adversary is able to obtain approximate location of the source node; similar approaches still provide local source location privacy only.

In order to achieve global network-wide source location privacy Network Mixing Ring (NMR) [20] is proposed that extends the routing through a random intermediate node. Logical ring containing large number of nodes is formed around the base station to mix messages globally. The actual routing has three steps. In the first step message is routed through a random intermediate node to some node in the ring, which provides local source location privacy. In the second step, the message circulates in the ring to achieve the global source location privacy. In the last step, the message is simply routed from a random node in the ring to the base station. Source location privacy problem was also taken up in [21] to trap an adversary in a false cyclic path. Several cyclic paths are pre-generated in the network which results in improved safety period. Once a real message crosses one of the cyclic paths, fake messages start to circulate along this path. Then, adversary which backtracks on real messages can be stuck onto the cyclic path. The safety period is improved by this technique while maintaining low latency.

The idea of fake source and dummy traffic [22] are proposed to hide real messages and sources within a dummy traffic and fake sources. Two techniques are proposed for above idea namely Short-lived and Persistent Fake Source techniques. Again, these techniques cannot protect the source location against a global adversary, because the fake sources are activated by real messages. In this way first source can be recognized as the real one by global adversary. To achieve protection against global adversary; techniques are proposed

such as Periodic Collection [23] and the Source Simulation [23]. Periodic Collection provides optimal source location privacy. In this technique every sensor node periodically sends a dummy or real message to the sink at pre-defined intervals thus the real messages are covered within the dummy ones and the network traffic is totally independent from the sensed events. In second technique several fake event sources are added in the network. These fake sources' behavior is identical with the real sources to confuse the adversary. The biggest challenge of the Source Simulation technique is to precisely simulate the source behavior. A statistical framework for evaluation of source anonymity in WSNs was proposed in [24]. All the above techniques are based on the fake sources and the dummy traffic that ensures source location privacy in the presence of global adversary on the cost of huge communication overhead or higher latency. Instead of gaining great source location privacy these techniques cannot be practically implemented over a long period of time.

3.2.1.2 Location privacy for base station

Since base station collects data from entire network security for a base station is highly essential. If an adversary is able to locate the base station, it can apply several attacks such as Denial of Service attacks [6]. Hence protecting the location of the base station is extremely important.

There are three basic traffic analysis techniques [25] for discovering a base station location: the rate monitoring attack, the time correlation attack and the content analysis attack. In the rate monitoring attack, an adversary observes node message sending rate and moves closer to a node with the highest rate. In the time correlation attack, an adversary monitors a correlation in sending times between a node and its neighbors. The adversary tries to detect which node forwards the current message and traces the path directly to a base station. In the content analysis attack, an adversary tries to acquire valuable information from message headers and payloads.

To get protection against the rate monitoring attack, every sensor node sends messages at a constant rate. A child node continuously transmits data packet until it is accepted by a parent node. If the child node has no packet to send, it injects a dummy packet. This technique guarantees constant sending rate over the network at the same time it decreases the lifetime of the WSN. Different encryption techniques can be used to protect the base station against content analysis but they are not adequate. These techniques do not solve the problem of message identity, which remains the same along the path and easy to track. Hop-by-hop re-encryption technique [25] can be use to overcome this problem.

Different techniques [26, 27] are proposed to counter the traffic analysis attacks. The rate monitoring attack can be partially prevented by the multiple parents routing technique since traffic spreads along multiple paths. In this technique, every sensor node has multiple parent nodes, which route messages to the base station. Packet is forwarded through any parent node which is randomly selected by its child nodes. This technique can be extended by the controlled random walk. A node forwards a message to one of its parent nodes with probability p. With probability 1 - p the node forwards the message randomly to one of its neighbors including the parent nodes. This technique introduces delivery time delays, which are proportional to extra hops taken by the messages. This technique is still error-prone to the time correlation attack. Therefore, Multi-parent routing technique with fractal propagation is used to avoid errors. When a node observes that a neighbor forwards a message to the base station, the node generates a fake message with some probability and forwards it to one of its neighbors. The main problem with this technique is that it generates a large amount of traffic near the base station, because nodes near the base station usually forward more messages.

All the above discussed techniques provide sink location privacy against the local adversary. A completely randomized [28] technique is used to protect against global adversary. In this technique, a data packet randomly travels through a network for a predefined number of hops discarding the fact that it reaches the base station or not. To improve the probability of data packet delivery, multiple copies of the same data packet can be sent out by the source or a history of already visited nodes can be appended to the message to prevent loops in the routing path. In addition, this technique provides protection against an internal adversary, because no information on the sink is included in a data packet. This technique may seem to be impractical for larger networks, because the probability of delivery decreases rapidly with an increasing number of nodes.

3.2.2 Temporal privacy

Temporal privacy concerns the time when event data is created at source, data collected by a sensor node and transmitted to the base station. Temporal privacy is of primary importance, especially in the mobile target tracking application of WSN, for example consider a forest scenario where presence of animal is sensed by the sensor node and reported to the network sink. Now if an adversary is able to associate the origin time of the packet, then the adversary will be able to track the animal's behavior and use that information for hunting. Or consider a military field where assets are sensed by a sensor node an adversary with knowledge of such timing information may be able to predict the moving path of the mobile target in the future, thus violating the privacy of the target.

In order to protect the temporal privacy some random delay can be added in data packet transmission. This technique is useful for delay tolerant application where timely delivery of data packet is not important. Different techniques are proposed to protect the temporal privacy such as Rate-Controlled Adaptive Delaying [15]. In Rate-Controlled Adaptive Delaying, every node buffers an incoming data packet and randomly delays its retransmission according to the exponential distribution. Buffer preemption technique is included to handle with the problem of overloaded buffers. When the node buffer is full, this technique chooses a message to be transmitted immediately without further delay. Several such techniques are proposed and evaluated in [16].

4. COMPARISONS AND CHALLENGES

Table 1: Comparison between different techniques

	Privacy preservation efficacy	***Over-head***	***Delay***	***Power consumption***
CPDA	Excellent (For data privacy)	Fair	Delay due to computation	High due to communication data and Calculation
SMART	Excellent (For J ≥ 3)	Small	Delay due to Slicing and recombine data	Slightly high due to computation (slices of data)
Flooding	Excellent (For local adversary)	No extra over-head	No extra delay	High due to flooding of data in network
Probabilistic flooding	Good (For local adversary)	Small	Delay can increase	Low as compare to flooding
Phantom flooding	Excellent (For local adversary)	Small	Delay can increase	Low as compare to flooding
Network Mixing Ring (NMR)	Excellent (For both local and global adversary)	Fair	Delay increase while improving the safety period	High as it circulates in the ring
Fake sources and the dummy traffic	Excellent (For global adversary)	Huge	No extra delay	No extra power consume
Injects a dummy packet for uniform sending rate	Excellent (For base station privacy)	No extra over-head	No extra delay	Fair but decreases the lifetime of WSN
Hop by hop re-encryption	Excellent	Fair	Delay due to encrypting and decrypting data	Increase due to computation.
Multiple-parent routing	Excellent (For local adversary)	No extra over-head	No extra delay	No extra power consumption
completely randomized scheme	Excellent (For both local and global adversary)	No extra over-head	Delay due to travels for a predefined number of hops	Fairly increase due to extra communication between hops
RCAD	Excellent (For hiding creation time of packet)	No extra over-head	Add random delay	No extra power consumption

While some work has been proposed in privacy protection in WSN, there are still many open challenges. As network behavior or weather changes, properties for sensor nodes change accordingly so in future researches one can propose privacy preserving techniques for WSN which can handle all these possibilities. Apart of this a lot of improvement can be possible in every technique with respect to power consumption or delay.

5. CONCLUDING REMARKS

This paper presents a review of privacy-preserving techniques for wireless sensor networks (WSN). Two main categories of privacy preserving techniques have been presented; data-oriented and context-oriented respectively. The existing techniques have been compared in terms of privacy preservation efficiency, overhead, delay and power consumption. Through comprehensive analysis the best suitable privacy technique for a WSN scenario can be found along with its limitation. This paper draws a complete picture for privacy preservation in WSN also introduces some open challenges for future research.

6. REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci. Wireless sensor networks: a Survey: Computer Networks 38 (4) (2002) 393–422.

[2] K. Sohraby, D. Minoli and T. Znati. Wireless Sensor Network: Technology, Protocols and Applications: John Wiley & Sons, 2007,pg10-11.

[3] 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, 2004.

[4] M. Gruteser, G. Schelle, A. Jain, R. Han, and D. Grunwald. Privacy-aware location sensor networks: In 9th USENIX Workshop on Hot Topics in Operating Systems (HotOS IX), 2003.

[5] Na Li, Nan Zhang, Sajal K. Das, and Bhavani Thuraisingham. Privacy preservation in wireless sensor networks: A state-of-the-art survey. Ad Hoc Networks 7 (2009) 1501–1514.

[6] John Paul Walters, Zhengqiang Liang, Weisong Shi, and Vipin Chaudhary. Wireless Sensor Network Security: A Survey:Security in Distributed, Grid, and Pervasive Computing Yang Xiao,(Eds.) Auerbach Publications, CRC Press.

[7] R. Agrawal, A. Evfimievski, R. Srikant, Information sharing across private databases, in: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, 2003, pp. 86–97.

[8] R. Agrawal, R. Srikan, Privacy-preserving data mining, in: Proceedings of the 2000 ACM SIGMOD on Management of Data, Dallas, TX USA, May 15–18, 2000, pp. 439–450.

[9] L. Sweeney, K-anonymity: a model for protecting privacy, International Journal on Uncertainty, Fuzziness and Knowledgebased Systems 2 (2) (2002) 557–570. pp. 86–97.

[10] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in Proceedings of the 9th ACM Conference on Computer and Communications Security, November 2002, pp. 41–47.

[11] D. Liu and P. Ning, "Establishing pairwise keys in distributed sensor networks," in Proceedings of 10th ACM Conference on Computer and Communications Security (CCS03), October 2003, pp. 52–61.

[12] W.B. He, X. Liu, H. Nguyen, K. Nahrstedt, T. Abdelzaher, PDA: privacy-preserving data aggregation in wireless sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007, pp. 2045–2053.

[13] R. Zhang, Y. Zhang, K. Ren, DP^2AC: distributed privacy-preserving access control in sensor networks, to appear in: Proceedings of the 28th IEEE International Conference on Computer Communications (INFOCOM 2009), pp.1298–1306.

[14] Jean-Franois Raymond. Tra_c analysis: Protocols, attacks, design issues and open problems. In Proceedings of International Workshop on Design Issues in Anonymity and Unobservability, pages 10{29. Springer- Verlag New York, Inc., 2001.

[15] Pandurang Kamat, Wenyuan Xu, Wade Trappe, and Yanyong Zhang. Temporal privacy in wireless sensor networks. In ICDCS '07: Proceedings of the 27th International Conference on Distributed computing Systems, pages 23-30, Washington, DC, USA, 2007. IEEE Computer Society.

[16] Pandurang Kamat, Wenyuan Xu, Wade Trappe, and Yanyong Zhang. Temporal privacy in wireless sensor networks: Theory and practice. ACM Transactions on Sensor Networks, 5(4):1-24, 2009.

[17] Celal Ozturk, Yanyong Zhang, and Wade 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, pages 88-93, New York, NY, USA, 2004. ACM.

[18] Yun Li and Jian Ren. Providing source-location privacy in wireless sensor networks. In WASA '09: Proceedings of the 4th International Conference on Wireless Algorithms, Systems, and Applications, pages 338-347, Berlin, Heidelberg, 2009. Springer-Verlag.

[19] Yun Li, Leron Lightfoot, and Jian Ren. Routing-based source-location privacy protection in wireless sensor networks. In IEEE International Conference on Electro/Information Technology, 2009, pages 29-34, 2009.

[20] Yun Li and Jian Ren. Mixing ring-based source-location privacy in wireless sensor networks. In International Conference on Computer Communications and Networks, pages 1-6, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[21] Yi Ouyang, Zhengyi Le, Guanling Chen, James Ford, and Fillia Makedon. Entrapping adversaries for source protection in sensor networks. In WOWMOM '06: Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks, pages 23-34,Washington, DC, USA, 2006. IEEE Computer Society.

[22] Celal Ozturk, Yanyong Zhang, and Wade 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, pages 88-93, New York, NY, USA, 2004. ACM.

[23] K. Mehta, Donggang Liu, and M. Wright. Location privacy in sensor networks against a global eavesdropper. In IEEE International Conference on Network Protocols, 2007. ICNP 2007, pages 31-323, October 2007.

[24] Basel Alomair, Andrew Clark, Jorge Cuellar, and Radha Poovendran. Statistical framework for source anonymity in sensor networks. Technical Report 3, Network Security Lab (NSL), College of Engineering, University of Washington, 2009.

[25] Jing Deng, Richard Han, and Shivakant 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, pages 637-646, Washington, DC, USA, 2004. IEEE Computer Society.

[26] Jing Deng, Richard Han, and Shivakant Mishra. Countermeasures against traffic analysis attacks in wireless sensor networks. In SECURECOMM '05: Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks, pages 113-126, Washington, DC, USA, 2005. IEEE Computer Society.

[27] Jing Deng, Richard Han, and Shivakant Mishra. Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks. Pervasive and Mobile Computing, 2(2):159-186, April 2006.

[28] Edith C.-H. Ngai. On providing sink anonymity for sensor networks. In IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing, pages 269-273, New York, NY, USA, 2009. ACM.