Source Location Privacy using Multiple-Phantom Nodes in WSN

Prabhat Kumar, J.P Singh, Prateek Vishnoi and M.P Singh
Computer Science and Engineering
National Institute of Technology Patna, Bihar, India
Email: {prabhat, jps}@nitp.ac.in, prateek.vishnoi8@gmail.com, mps@nitp.ac.in

***Abstract*—The ever increasing integration of sensor-driven application into our lives has led to sensor privacy becoming an important issue. The locational information of sensor nodes has to be hidden from adversary for the sake of privacy. An adversary may trace traffic and try to figure out the location of the source node. This work attempts to improve the Source Location Privacy by using two phantom nodes, selection of neighbors based on random based approach and random walk upto phantom nodes. Two phantom nodes are selected for each source node in such a way that no two phantom nodes of the same triplet are co-linear with the sink. The proposed protocol can keep the adversary confused within the sensor networks as it generates different paths for different packets for the same source. Here, we are distracting the adversary by creating alternate paths. This results in minimizing the hit-ratio, thereby maximizing the privacy. Analysis of the present work shows that this protocol tends to achieve more privacy and greater safety period as compared to single phantom routing protocol. Flooding techniques and dummy packets have not been used in working phase for the sake of energy efficiency and network congestion.**

Keywords*—*Wireless Sensor Networks, Context Privacy, Source Location Privacy, Phantom Node, Random Walk

I. Introduction

Wireless sensor networks are composed of many small sensor nodes that can sense, collect and spread information for different types of applications. Sensing of data includes sensing physical quantity such as temperature, humidity, pressure, radiation etc. Wireless sensor nodes have limited storage, computing power, and energy supply [1],[2]. After the deployment of sensor nodes, the nodes are left unattended for most of the practical applications [3],[4]. One of the major application is subject tracking and monitoring where not only data but also the location of the sensor node needs to be preserved [5]. Privacy can be defined as "a state in which one is not observed or disturbed by other people". It is the state of being free from public attention. Privacy in wireless sensor networks includes hiding of nodes location, confidentiality, availability and integrity of messages etc. It can be broadly classified into two parts: Content Privacy and Context Privacy. Content privacy deals with the protection of data that is being communicated between sensor nodes while context privacy deals with the context related to the information such as source location, destination (sink) location and time at which the message was created. Context privacy includes hiding the identity and the locality of each node, and hiding the flow of traffic among the nodes. Our focus is on Source Location Privacy(SLP) as shown in Fig.1. Let us take an example of the Panda Hunter game where

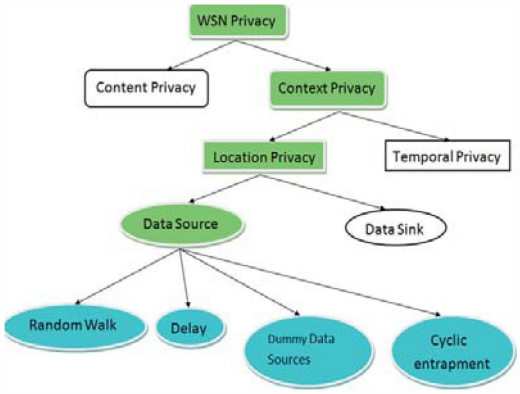

Fig. 1: Privacy in WSN

sensor nodes are deployed to sense the location of panda and inform the base station (sink) with the help of different intermediate nodes [6]. Hunter or adversary tries it's best to capture the Panda. Every time when the hunter or adversary captures the message it moves towards the source location. It is assumed that adversary has high computational power and memory so that it can track the whole path from sink to source. Various solutions have been proposed to attain the SLP but achieving the location privacy is a difficult task as there are different factors that plays an important role in the effectiveness of a solution. For example, mobile and static nodes require different types of SLP solutions [7]. The scope of the adversary, whether local or global, also needs to be taken into consideration. Local adversary can view or analyze some part of the network while the global adversary can view or analyze the whole network at one glance. Another influencing factor that is related to the adversary is whether the adversary can compromise some nodes or not. Some protocols have been proposed in order to gain privacy. This can be broadly classified into four parts as shown in Fig.1. Random Walk are those solutions where we use random walk as a subpart of the protocol. In Delay solutions, messages are delayed at some node for random amount of time before forwarding it. In Dummy data source, we introduce dummy traffic in the network by introducing dummy packets[8][9] and thus, making harder for an adversary to analyze the traffic. Cyclic entrapment is another type of solution where we introduce cycles of messages at some node and thus try to confuse the adversary. The main contribution of our work is: We propose a new routing method for energy constraint WSN

978-1-4799-8641-5/15/$31.00 © 2015 IEEE

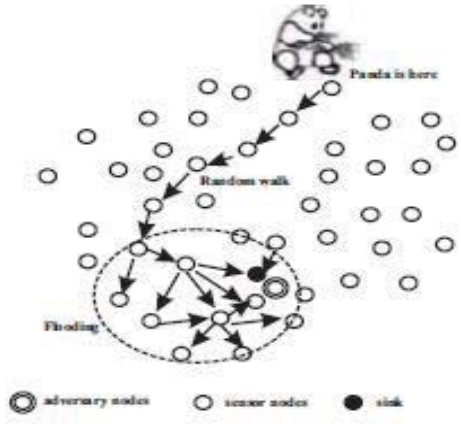

Fig. 2: Phantom Flooding Protocol[8]

deployed for subject monitoring. Using β-angle anonymity combined with phantom source we present a metric used for evaluating the privacy level of the protocol. Lastly, we analyze the effectiveness of the solution with respect to single phantom source and random walk routing solutions.

Remaining portion of the paper is organized as follows. Section II introduces Related Work, Section III discusses the Network model, Section IV explains the Adversary Model, Section V describes Proposed protocol, Section VI analyzes the new protocol and Section VII concludes with a brief discussion of Future Work.

II. Related Work

Pandurang Kamat and Y.Zhang[10],[11] introduced the phantom routing technique where the random walk is introduced. In phantom routing scheme(PRS), there are two phases-random walk phase and flooding phase. When the source senses any event then the message is forwarded in random fashion for h hops. The node that has received the packet at the end of the random walk termed as the phantom source. After that, the phantom source starts to flood the packet in the network towards the sink as shown in Fig.2. The purpose here is to keep the phantom node away from the source even if the adversary tries for hop to hop trace, it reaches to the phantom source and not the actual source. Phantom Single Path Routing(PSRS) is another protocol which is similar to the PRS but in PSRS shortest path algorithm is used instead of flooding. PSRS is better than PRS, in sense that flooding requires a lot of energy consumption as compared to the shortest path. Kamat et al.[10]shows that pure random walk is unable to keep the phantom source away from the real source. If the message is forwarded to h hops then there is a high probability that the phantom source will be only h hops away. But in some cases, there may be possibility of cycles in the random walk path from source to phantom source. The cycles in a random walk not only ease the tracing for an adversary even they also upsurge the energy consumption of the nodes. So in order to avoid repetition of paths, directed random walk came into existence. There are two types of directed random walk: sector-based directed random walk and hop based directed random walk. In a sector based directed random walk[12], author suggested a technique to divide the neighbors into required sets without the use of the sectional antenna. In the hop-based directed random walk, each sensor node divides its neighbor into two sets P and Q. P contains all the sensor nodes whose hop-count is less than or equal to the node while Q contains all the sensor nodes whose hop-count is greater than the node. This information is kept by each node in its memory. Each node knows its hop-distance to the sink. Information of hop-distance to the sink can be calculated during configuration phase with a simple technique. Sink initiates a flood by sending a message to all its neighbors with a hop-count value zero. Each node forwards the message to other neighbors and increases the hop-count by one. Among all the recorded hop-counts each sensor node chooses the minimum one and in this way each node gets the hop-count to the sink. L. Zhang[12] suggested an enhancement of the sector based directed random walk and introduced the self-adjusting directed random walk(SADRW). According to SADRW, each sensor nodes divides its neighbors into four groups: neighbors to east, neighbors to west, neighbors to north, neighbors to south. When a sensor node senses any event it chooses a random set from the above sets and forwards the message to randomly selected node from that set. Each intermediate node forwards the packet in the same direction for next h hops. If any intermediary node cannot forward it to the chosen direction then it selects a new direction from rest of the set termed as second direction. If the case that when the packet is at the edge of network then the random walk terminates if it travels minimum i hops, $h \geq i$. Otherwise, it selects any random neighbor from remaining two sets. Wang Wei-Ping et al.[13] introduced the Phantom Routing with Location Angle(PRLA). First of all each node calculates the inclination angle between itself and its neighbor nodes. If any source senses any event then it selects neighbor with some probability. After that each intermediate node will forward the message to the next neighbor with the same inclination angle and in this way the random walk is always directed away from the source. After reaching the phantom source, the shortest path algorithm from the phantom source to the sink is applied. Simulation results shows that PRLA improves the safety period as compared to PSRS. Yun Li et al.[14] introduced the RRIN(Randomly selected intermediate node) as an improvement over Phantom Routing Scheme. In RRIN, the source node first selects the intermediate node and sends the message to it and after that the message is sent to the sink. Jun Long et al.[15] introduced the network lifetime maximization through Tree-Based Diversionary Routing. Under his solution they are creating tree branches that are completely homogeneous to the adversary. They are utilizing the energy of those sensor nodes that are away from the sink. In this way, they are minimizing the energy consumption of hotspot regions. Hotspot regions contains those sensor nodes that are situated near the base station. Petro Spachos et al.[16] suggested the opportunistic routing where some sensor nodes act as relay nodes and one node among them will finally forward the packet. In opportunistic routing, next hop relay nodes at each hop has the equal probability to forward the packet. Each time when the message is forwarded from the source to the sink the path changes for every messages as each relay nodes has the equal probability to forward it. Opportunistic routing shows his good performance when the network contains the large number of nodes so that when the

message is forwarded, every time the node choose different nodes. Yun Li et al.[6],[7],[17] introduced the combination of RRIN with the NMR[18]. Firstly, the network is divided into smaller grids. After the formation of grids in the WSN, the network-mixing ring is generated by some sensor nodes called header ring nodes. When the source senses any event then the message is forwarded to any intermediate node, after that the message is passed to the network mixing ring and then the header node would be responsible for deliver it to the sink. Xi et al.[19] introduced the greedy random walk(GROW) in which a sensor node is selected called receptor node. One random walk is started from the source to receptor node and one walk is started from the sink. Then the formed path is used to send the data from source to the sink. Yao et al. [20] provide another enhancement of the random walk by introducing the directed random walk (DROW). DROW is completely dependent on the hop-distance towards the basestation. Each node checks its hop-distance to the sink and then selects the minimum one. It shows the higher privacy when the intermediate node would be more than one node from which the hop distance to the sink is minimum. Younis et al.[21] introduced the cyclic entrapment method(CEM) where some sensor nodes behave as activation node. Activation nodes are those nodes which create a cycle of messages in the network. These cycles confuses the adversary for traffic analysis. But the major drawback in CEM is that if the hearing range of the local adversary is more than the loop then it can bypass the whole cycle after capturing a single message. This method also reduces the network-lifetime as it is introducing the unnecessary messages in the network.

III. Network Model

In our network model, homogeneous sensor nodes are deployed randomly that can behave as source node, intermediate node or phantom node. All these sensor nodes are static in nature which means they cannot move in the network. Sensor node that senses any event and forwards the message to the base station is called as source node. Intermediate nodes are those nodes that forward the received message towards its destination. When a sensor node senses any event then it forwards the message towards the sink with the help of neighbor nodes. These all neighbor nodes are called intermediate nodes. Phantom node is a sensor node that forwards the message of source node with its own identity. All three nodes are similar in nature but performing different task at different time.

IV. Adversary Model

Adversary tries its best to know the source location. It is similar to Panda Hunter game. The basic characteristics of adversary are as follows:

- Adversary is local i.e adversary cannot view the whole network at one glance. It can only view a part of it which means it can eavesdrop a packet only when it is under its hearing range.
- Adversary is mobile i.e it can move from one position to another. It can move towards the immediate sender of the captured message.
- Initially, adversary will be found near the base station. From base station it will start its strategy to capture the source.

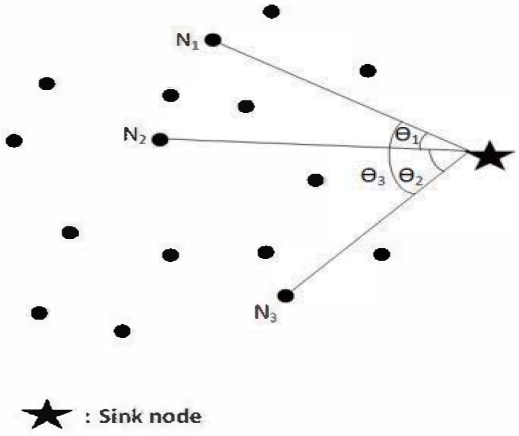

Fig. 3: Triplet Selection

- Adversary is resourceful and has no dearth of resources like storage, computational power. It can save all the captured messages which can be used to identify the routing path.
- Adversary is passive i.e it can only overhear the message, it cannot harm the sensor nodes like destroying sensor nodes, compromising some nodes etc.

V. Multi-Phantom Routing Scheme

Our proposed scheme consists of two phases: (i) configuration phase (involves neighbor discovery, flooding, node reports its hop count from the BS and triplet selection) and (ii) working phase (involves random walk and phantom selection based on given criteria).

A. Configuration Phase

During configuration phase, initially sink starts flooding with a message setting counter zero. Each node stores the counter value with sender ID. After that it forwards the message to its neighbor with incremented value of counter by one. In this way, each sensor node has well knowledge that base station is how much hop distance away. After that each node informs the hop-distance to the sink. Sink maintains a hop distance table from where it creates set of triplets of sensor nodes. In a triplet, each sensor node behaves as phantom node for other two nodes. A triplet is selected in such a way that no two sensor nodes and sink are co-linear and the angle between each two node with the sink should be atleast 30 degree. If sensor nodes are co-linear with the sink then the source node would lie in the path between phantom node and sink. When this condition arises, then the privacy can be easily breached by the adversary. This we can easily understand with the help of Fig. 3

$$m(slope) = \frac{(y_2 - y_1)}{(x_2 - x_1)} \tag{1}$$

$$\theta = \tan^{-1} \frac{(m_2 - m_1)}{(1 + m_1 m_2)} \tag{2}$$

During configuration phase, it is assumed that all sensor nodes have been localized and the sink node has well knowledge of all the sensor nodes. Now base station randomly chooses three sensor nodes that are nearly same hop distance away and then calculates the angle between them. Base station can easily find out the angle between the two nodes with itself as vertex with the help of equation 1 and 2. If all the three angles that are calculated are more than 30 degree then the triplet is selected otherwise choose some different nodes. In our algorithm, we have fixed the position of a node in the triplet and other two can be changed. In the mentioned algorithm 1, we have fixed the position of n_2 while n_1 and n_3 are changing their position to form the triplet. When the angle between n_1 and n_2 is less than 30 degree then n_1 is replaced with some other node. If the angle between n_1 and n_3 or n_2 and n_3 is less than 30 degree then n_3 is replaced with some other node. This has been summarized in algorithm 1 :

Algorithm 1: Triplet Selection Algorithm(At Sink)

1: Sort the table in ascending order of hop count value.
2: Choose three different nodes n_1, n_2, n_3 randomly.
3: Cal. angle between them θ_1(n_1 and n_2), θ_2(n_2 and n_3), θ_3(n_3 and n_1) in degree.
4: **if** ($\theta_1 \geq 30$)
5: **if**($\theta_2 \geq 30$)
6: **if**($\theta_3 \geq 30$)
7: Triplet Selected
8: Inform(n_1, n_2, n_3)
9: **goto** 2
10. **else**
11. replace(n_3 with some other node)
12: **goto** 5
12. **else**
13: replace(n_3 with some other node)
14: **goto** 5
15: **else**
16: replace(n_1 with some other node)
17: **goto** 4

After the selection of triplet, base-station informs to all the sensor nodes about its triplet by sending a message including the ID of other two sensor nodes. Each sensor node stores the ID of the two sensor nodes that will behave as phantom node for it. Each sensor node will behave as phantom for the two different nodes.

B. Working Phase

After the completion of configuration phase, the working phase starts during which the communication between the source node and the sink node is performed. Triplet selection has been done during configuration phase. Each node has ID of two different sensor nodes that are in triplet. These two sensor nodes will behave as phantom node one at a time. When a source node senses any event then it randomly generates a number within 1 to 10. If the generated number is greater than 5 then the first node is selected otherwise second node is selected as phantom. After selecting the phantom node, source node forwards the message to randomly selected neighbor with phantom node as destination. Selected neighbor forwards the message towards the destination phantom node with shortest path algorithm. It also includes its ID in message content.

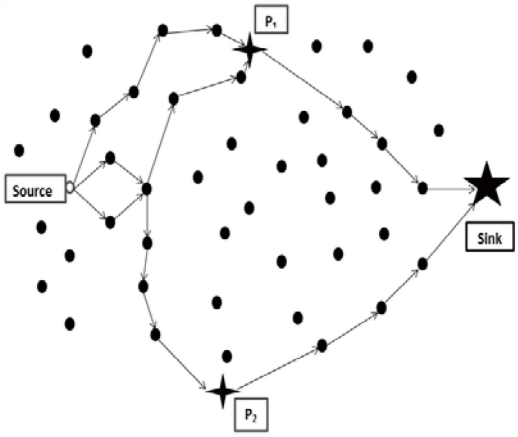

Fig. 4: Working Phase

After receiving the message at phantom node, it checks the sender of the message. If source is its phantom node then it forwards the message towards the sink with its own ID by using shortest path algorithm. This we can explain with the help of fig.4. Here, source first randomly generates a number a between 1 and 10. After that it is checked whether a is greater than 5 or less than equal to 5. If the a is less than or equal to 5 then the P_1 is selected as phantom node otherwise P_2 is selected as phantom node. After the selection of phantom node, source randomly chooses x from set of neighbors N. Now, source passes message M to x with P as destination. Then, after some intermediate nodes message reaches to the phantom P with the help of shortest path algorithm. Now phantom node checks whether source is its phantom. If the condition satisfies then the P forwards the message to the sink with the help of shortest path algorithm. This has been summarized in algorithm 2:

Algorithm 2: Working Phase algorithm

1: Source S generates number a between [1,10]
2: **if** ($a \leq 5$)
3: $P = P_1$
4: **else**
5: $P = P_2$
6: S randomly chooses node $x \in$ N (Set of neighbors)
7: S passes message M to x with destination P
8: x forwards the M to P
9: **if** (S==phantom(P))
10: change($S = P$)
11: P forwards M to sink node

Each time before sending the message, source first generates a number and checks the condition and then the phantom node is selected. Thus, we are trying to create alternate paths from source to sink with the help of random neighbor and phantom node. These alternate paths from source to sink makes job harder for an adversary to trace the source location.

VI. Analysis

Achieving the source location privacy is a difficult task. Let us suppose if the messages are forwarded continuously

from source to base-station with shortest path algorithm then in worst case after capturing *h* messages the local adversary will be able to capture the source location where *h* is the hop-distance from source to sink. Now, suppose if we choose single phantom node to forward the message from source to sink then again the source location can be traced after short amount of time. Again, the phantom node with the source and the sink must not be co-linear. Thus, in our proposed approach we are trying to overcome these problems with multi-phantom source location privacy. In the configuration phase, we are selecting triplets of sensor nodes that are not co-linear. As the above algorithm 1 shows that each time we are checking whether the angle between each pair of triplet of sensor nodes is greater than or equal to the 30 degree. If the angle between each pair of nodes is greater than 30 degree then only the triplet is selected otherwise we would select different sensor nodes. In the working phase, source is selecting the phantom node based on some probability. Here we are assuming that source generates a random number between 1 and 10 and checks whether it is lesser than or equal to 5. If number is lesser than or equal to 5 then first node is selected otherwise second node is selected as shown in algorithm 2. After that source randomly chooses the neighbor to forward the message towards the phantom node. In our proposed approach, there is no restriction on the position of the base-station. Base-station may be present at the corner or in the middle of the network.

Now, lets assume source *S* senses an event with neighbors *n* then probability of selecting a particular neighbor will be $\frac{1}{n}$. If there are *p* phantom nodes for each node then the probability of selecting a phantom node will be $\frac{1}{p}$. Thus, total number of different routing paths would be *np*. If we assume that adversary can only reach to the immediate neighbor after capturing a single message then it would need to capture the *(x+y+1)* number of messages to locate source, where *x* and *y* are the hop distance from source to phantom node and phantom node to the sink. Probability that the message will be transmitted from a particular path would be $\frac{1}{np}$. Thus, it shows that adversary would move one hop distance towards the source after capturing atleast *np* messages from current node. Now hit ratio can be defined as:

$$Hit\ Ratio = \frac{No.\ of\ messages\ captured}{No.\ of\ messages\ sent} \tag{3}$$

$$H.R = \frac{1}{np} \tag{4}$$

Hit ratio is the parameter to calculate the privacy. Privacy and hit ratio are inversely proportional to each other. If the hit ratio is minimum then privacy will be maximum and the vice-versa. Hit ratio would be reduced if the number of messages captured by the local adversary can be minimized. This can be done if each time we use different path to send the message from source to sink. Here we are creating alternate paths by selecting random neighbor and multi-phantom nodes. Thus, from the above formula, it can be concluded that greater the number of neighbors and the phantom nodes higher would be the privacy achieved.

VII. Conclusion and Future Works

Source location privacy is a serious issue for many monitoring and remote sensing applications. In many scenarios, an adversary may be able to trace back to the source location if not handled properly. In this paper, we have proposed a multi-phantom routing protocol to confuse the adversary by creating alternate paths from source to sink. This protocol also keep in mind the energy issues of WSN and avoids the use of dummy packets and flooding in working phase. The proposed protocol works better than single phantom based approach. Future work may be done on analyzing the performance with respect to increase of phantom nodes. The proposed protocol may be further analyzed by including dummy packets and flooding for privacy improvements.

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, 7 2012.

[2] Z. Eu, H. Tan, and W. Seah, "Design and performance analysis of mac schemes for wireless sensor networks powered by ambient energy harvesting", *Ad Hoc Networks,*, vol. 9, no. 3, pp. 300-323, 5 2011.

[3] A. Cardenas, T. Roosta, "Rethinking security properties, threat models, and the design space in sensor networks: A case study in scada systems", *Ad Hoc Networks,*, vol. 7, no. 8, pp. 1434-1447, 11 2009.

[4] X. Mingjun, H. Liusheng, X. Hongli, W. Yang, and P. Zegen, "Privacy preserving hop-distance computation in wireless sensor networks", *Chinese J. Electronics,*, vol. 19, no. 1, pp. 191194, 1 2010.

[5] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey", *Computer networks,*, vol. 38, no. 4, pp. 393422, 3 2002.

[6] 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. 15011514, 11 2009.

[7] Mauro Conti, Jeroen Willemsen, Bruno Crispo "Providing Source Location Privacy in Wireless Sensor Networks: A Survey" *IEEE Communications Surveys and Tutorial,* VOL. 15, NO. 3, THIRD QUARTER 2013

[8] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Towards a statistical framework for source anonymity in sensor networks", *IEEE Trans. Mobile Computing,* vol. 10, no. 12,2011

[9] A. Jhumka, M. Leeke, and S. Shrestha, "On the use of fake sources for source location privacy: trade-offs between energy and privacy", *The Computer Journal* vol. 54, no. 6, pp. 860-874, 2 2011.

[10] P. Kamat, Y. Zhang and W. Trappe "Enhancing source location privacy in sensor network routing", *Proc. 25th IEEE International Conference on Distributed Computing Systems*, IEEE. Los Alamitos, CA, USA: IEEE Computer Society, 06 2005, pp. 599-608

[11] C. Ozturk, Y. Zhang, and W. Trappe "Source-location privacy in energy-constrained sensor network routing", *Proc. 2nd ACM workshop on Security of Ad hoc and Sensor Networks*, SASN 04, ACM. New York, NY, USA: ACM, 10 2004, pp. 88-93.

[12] L. Zhang, "A self-adjusting directed random walk approach for enhancing source-location privacy in sensor network routing", *Proc. 2006 international conference on Wireless communications and mobile computing*, SASN 04, ACM. New York, NY, USA: ACM, 10 2004, pp. 88-93.

[13] W. Wei-ping, C. Liang, and W. Jian-xin, "A source-location privacy protocol in wsn based on locational angle", *Communications, 2008. IEEE International Conference on*, ser. ICC08, IEEE. Piscataway, USA: IEEE, 5 2008, pp. 1630-1634

[14] Yun Li, Leron Lightfoot, Jian Ren, "Routing-Based Source-Location Privacy Protection in Wireless Sensor Networks", *Electro/Information Technology, 2009. IEEE International Conference on*, EIT 09, IEEE. Piscataway, NJ, USA: IEEE, 6 2009, pp. 29-34

[15] Jun Long, Mianxiong Dong "Achieving Source Location Privacy and Network Lifetime Maximisation Through Tree Based Diversionary Routing in Wireless Sensor Networks", *IEEE Translation* Japan 2014

[16] P. Spachos, L. Song, and D. Hatzinakos "Opportunistic routing for enhanced source-location privacy in wireless sensor networks", *25th Biennial Symposium on Communications,* , QBSC 2010, IEEE. Stoughton, WI, USA: The Printing House, Inc., 5 2010, pp. 315-318.

[17] Yun Li; Jian Ren "Preserving source-location privacy in wireless sensor networks,", *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. 6th Annual IEEE Communications Society Conference on* , SECON09, IEEE. Piscataway, NJ, USA: IEEE Press, 6 2009, pp. 1-9.

[18] Y. Li and J. Ren, "Mixing ring-based source-location privacy in wireless sensor networks", *Computer Communications and Networks, 2009. Proceedings of 18th Internatonal Conference on* , ICCCN 2009, IEEE. Washington, DC, USA: IEEE Computer Society, 8 2009, pp. 1-6

[19] Y. Xi, L. Schwiebert, and W. Shi "Preserving source location privacy in monitoring-based wireless sensor networks", *20th International Parallel and Distributed Processing Symposium,* IPDPS 2006, IEEE. Piscataway, USA: IEEE, 4 2006, p. 8 pp.

[20] J. Yao and G. Wen, Preserving source-location privacy in energyconstrained wireless sensor networks, in Distributed Computing Systems Workshops, 2008. 28th International Conference on, ser. ICDCS08, IEEE. Los Alamitos, CA, USA: IEEE Computer Society, 6 2008, pp. 412416.

[21] O. Younis and S. Fahmy "Entrapping Adversaries for Source Protection in Sensor Networks", *Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'06)* , vol3 pp 366-379,2006