Analysis of Location Privacy/Energy Efficiency Tradeoffs in Wireless Sensor Networks

Sergio Armenia[1], Giacomo Morabito[2], and Sergio Palazzo[2]

[1] CNIT - Research Unit of Catania, Italy
sarmenia@diit.unict.it
[2] Universita' di Catania, DIIT, V.le A.Doria, Catania, Italy
gmorabi@diit.unict.it, spalazzo@diit.unict.it

Abstract. In this paper an analytical framework is proposed for the evaluation of the tradeoffs between location privacy and energy efficiency in wireless sensor networks. We assume that random routing is utilized to improve privacy. However, this involves an increase in the average path length and thus an increase in energy consumption. The privacy loss is measured using information theory concepts; indeed, it is calculated as the difference between the uncertainties on the target location before and after the attack. To evaluate both privacy loss and average energy consumption the behavior of the routing protocol is modeled through a Markov chain in which states represent the nodes traversed by a packet in its way to the sink. The analytical framework can be used by designers to evaluate the most appropriate setting of the random routing parameters depending on the privacy and/or energy efficiency requirements.

1 Introduction

It is well known that wireless networks have serious privacy problems. This is mainly because of the broadcast nature of the radio channel that allows all stations in proximity of the sender to overhear the frames sent. Even if network devices make use of encryption algorithms, confidentiality is usually provided for the data field only, whereas the header/tail fields remain in plain text. Therefore, given that during the normal activity there are frequently packets sent using broadcast address as destination, an eavesdropper can receive and process them without any effort and thus can obtain information about the sender. This, joined to the fact that wireless devices usually have a fixed address, gives the attackers the possibility to link device address to user identity or to device position as well to the type of application utilized.

In wireless sensor networks (WSNs) the above problems are amplified and new issues arise. In fact, WSNs are based on the wireless multihop communication paradigm and therefore, eavesdropping attacks can be accomplished more easily. Furthermore, WSN applications are pervasive by nature and as a consequence, a lot of user sensible information can be stolen by attackers.

In the recent past a lot of attention has been devoted to the key distribution in the WSN cryptography domain. Accordingly, several solutions have been proposed for pre-distributing keys or for reducing their size.

I.F. Akyildiz et al. (Eds.): NETWORKING 2007, LNCS 4479, pp. 215–226, 2007.

However, secure cryptography does not guarantee privacy, as we have already said. Indeed, some research work has recently appeared that deals with the relationship between routing and location privacy in WSN. In fact, radio activity at an intermediate node can be used to obtain information about the position of the information source.

In [3] a formal model of the source-location privacy problem is provided, and two popular classes of routing protocols, namely, flooding protocols and single path protocol, are analysed from the privacy and energy consumption standpoints. Based on such analysis a new technique called *phantom routing* is proposed that combines the advantages of both the above mentioned classes of routing protocols and provides suitable protection of the source location while not causing a noticeable increase in energy consumption. In [5] the authors propose GROW (Greedy Random Walk): a two way random walk to reduce the chance an eavesdropper can collect the source-location information. Note that both the above research contributions are simulations-based.

Differently, in this paper we introduce an analytical framework for the evaluation of the tradeoff between location privacy and energy efficiency in wireless sensor networks. To this end we extend the definition of privacy loss based on information theory concepts, proposed in [1] for data mining systems, to the case of location privacy in sensor networks. More specifically, we focus on the relationship between random routing design choices and privacy loss as well as energy efficiency. Accordingly, we will derive a Markov-based model of the random routing behavior that allows to calculate the privacy loss as well as the average energy consumption. Numerical results confirm that, as expected, energy efficiency and privacy are competing requirements. The framework can be used by protocol designers to set appropriate tradeoffs between the two above requirements.

The remainder of this paper is organised as follows: in Section 2 we present the system model along with a statement of the problem. In Section 3 we evaluate the privacy loss and the energy consumption. Some numerical results are provided in Section 4 and, finally, conclusions are drawn in Section 5.

2 Problem Statement and System Model

In this section we first state the problem of location-privacy in wireless sensor networks (WSNs). More in detail, in Section 2.1 we will define the problem using the *panda-hunter game* scenario, then, in Section 2.2 we introduce the system model that will be utilized in the following of the paper.

2.1 Statement of the Problem: The Panda-Hunter Game

The panda-hunter game is a well known reference scenario utilized for the study of source location privacy in WSNs [4,3].

Suppose that a set of sensor nodes has been deployed by the *Save The Panda Organisation*, in a random way within a large area in order to study and monitor panda habit. Sensor nodes are able to detect panda's presence. At any time, while the panda freely moves, there is always a sensor node, called *source node*, that detects panda's position. Such an observation must be periodically reported to a sink node, via multihop routing techniques. In this way the current position of the panda is approximately the position of the current source node. Thus, when the sink node receives a message from the source node, it will know the panda position.

We suppose that transmissions are encrypted, so the source node ID field cannot be read by attackers. Moreover we assume that relationship between node ID and node location is known only by the sink node.

In the area there is a hunter as well, with the role of adversary. He aims to catch the panda, thus he is an enemy from the Save The Panda Organisation standpoint. The hunter is not able to decrypt messages therefore, he cannot learn, at least not directly, location of the source node, but in order to get the worst case we considered the hunter, as in [3], *non malicius*, i.e. does not interfere with proper function of the network, *device rich*, i.e. he is equipped in such a way he can measure signal strenght and angle of arrival of any message, *resource rich*, i.e. he has unlimited amount of power, and *informed*, i.e. he knows location of the sink, and the network structure and protocols.

Using his devices and resources the hunter can analyse messages at RF level, so he can try to capture panda by back-tracing the routing path used by messages until the source.

As an example, consider the sensor network represented in Figure 1. There are $N = 11$ sensor nodes n_0, n_1, ..., n_{10}, with n_0 representing the sink.

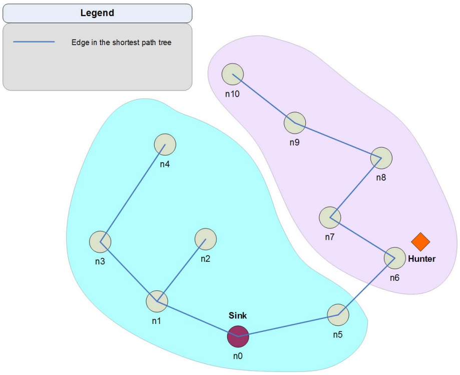

Fig. 1. Example of hunting activity

In Figure 1 we show the shortest path routing tree connecting each sensor node n_i to the sink n_0, i.e., node n_0 is the root of the tree. If the hunter is located near node n_6 and detects radio activity, then a node in the set $\{n_7, n_8, n_9, n_{10}\}$ is the source node. Instead, if no activity is detected, then the panda is near one of the remaining nodes, i.e., a node in the set $\{n_1, n_2, n_3, n_4, n_5\}$ is the source node.

Observe, that in any case the hunter splits the network and obtains information about the panda location.

This leads to a strict connection between location privacy and routing protocol in a WSN. Routing protocols must be privacy-aware in order to save, or at least prolong, panda's life.

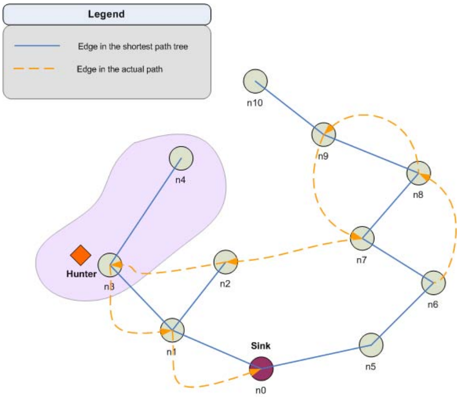

Fig. 2. Example of random routing

A simple way to improve privacy is to introduce some randomness in the routing behavior. Indeed, in *random routing* the next relay is chosen randomly between all the neighbors of the current relay. As an example, in Figure 2 we show a path obtained applying random routing in the same WSN shown in Figure 1. In this case, the fact that node n_3 is forwarding a packet does not mean that the source node is in the set $\{n_3, n_4\}$.

However, the length of path followed by packets in random routing can increase significantly, which involves large energy consumption. In other words, the increase in privacy is achieved at the expenses of higher energy consumption. It follows that appropriate tradeoffs are needed.

2.2 System Model

Let us consider a WSN composed of M nodes denoted as n_0, n_1, ..., n_{M-1}. For any node n_i, with $i < M$, we define $\Phi(n_i)$ the set of *neighbors*[1] of n_i and $\phi(n_i)$ the number of its neighbors, i.e., $\phi(n_i) = |\Phi(n_i)|$. Now suppose that n_0 is the sink and let us call $d(n_i)$ the distance between node n_i and the sink n_0. Obviously, $d(n_0) = 0$ and $d(n_i) = \min_{n \in \Phi(n_i)}\{d(n) + 1\}$.

Observe that routing of packets towards the sink in a sensor networks can be modeled by means of a matrix $Q \in \Re^{(M-1)\times(M-1)}$, the generic element of which, $[Q]_{i,j}$, represents the probability that the next relay of a packet transmitted by n_i is n_j, with i and $j \in [1, (M-1)]$. We define Q as the *routing matrix*.

[1] We say that two nodes are neighbors if they are in the radio coverage of each other.

In order to model random routing we define the *best next relay* $\Psi(n_i)$ as the neighbour of n_i which is closest to the sink, i.e., it is a node that satisfies the following relationship

$$d\left(\psi(n_i)\right) \leq d(m), \forall m \in \Phi(n_i) \tag{1}$$

Let us stress that even if several nodes may satisfy the relationship in eq. (1), for each n_i only one node $\psi(n_i)$ is selected. Accordingly, if shortest path routing is utilized $[Q]_{[i,j]}$ is equal to 1 if n_j is the best next relay, i.e., if $n_j = \psi(n_i)$, and is equal to 0, otherwise.

We define as *p-random routing* a routing algorithm which chooses the best next relay with probability p and any other neighbor node with equal probability. Accordingly, the routing matrix of a p-random routing protocol is

$$[Q]_{i,j} = \begin{cases} p & \text{if } n_j = \psi(n_i) \text{ and } \phi(n_i) > 1 \\ \frac{(1-p)}{\phi(n_i)-1} & \text{if } n_j \neq \psi(n_i) \text{ and } \phi(n_i) > 1 \\ 1 & \text{if } n_j = \psi(n_i) \text{ and } \phi(n_i) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Note that if p is equal to 1, then random routing becomes a shortest path routing.

3 Performance Analysis

We define and derive the location privacy loss when p-random routing is applied in Sections 3.1. Then, in Section 3.2, we will derive the corresponding energy consumption. Such performance metrics will be evaluated as a function of the probability p. This allows us to evaluate appropriate tradeoffs between privacy loss and energy consumption.

3.1 Privacy Loss

A measure of privacy is crucial to the evaluation of privacy enhancement solutions. Accordingly, in the recent past some research effort has been devoted to the definition of an appropriate privacy metric. In [6] an overview of the most interesting solutions is provided. Here we extend the definition proposed in [1] to the location privacy case.

Let S be the random variable representing the current position of the panda. We identify the location with the sensor node that detects the presence of the panda. Accordingly, at any time the random variable S can assume a value in the set $\{n_0, n_1, ..., n_{M-1}\}$.

Now suppose that following an attack, the hunter can observe a variable X which is correlated to S and can assume one of the following N values: $\{x_0, x_2, \cdots, x_{N-1}\}$.

The loss of privacy is related to the amount of information gained by the hunter following the attack. Such information is given by the difference between the uncertainty on S before and after knowing X. In the context of the information theory the measure of uncertainty of a random variable can be evaluated as the entropy of such a variable, $H(S)$.

In [1] the loss of privacy is calculated as:

$$\rho = 1 - 2^{-I(S,X)} \tag{3}$$

where $I(S, X)$ is the mutual information between S and X and is given by $I(S, X) = H(S) - H(S|X)$.

The uncertainty on S is (see [7] for example) defined as:

$$H(S) = -\sum_{m=0}^{M-1} p_S(n_m) \log_2 [p_S(n_m)] \tag{4}$$

where $p_S(n_m)$ represents the probability that the source node is n_m, whereas the uncertainty on S given X is

$$H(S|X) = -\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} p_{SX}(n_m, x_n) \log_2 [p_S(n_m|x_m)] \tag{5}$$

where $p_{SX}(n_m, x_n)$ represents the joint probability that S assumes the value n_m and X assumes the value x_n, whereas $p_S(n_m|x_n)$ represents the probability that S assumes the value n_m given that X assumes the value x_n.

Obviously, the probability $p_S(n_m|x_n)$ can be calculated as

$$p_S(n_m|x_n) = \frac{p_{SX}(n_m, x_n)}{p_X(x_n)} = \frac{p_{SX}(n_m, x_n)}{\sum_{i=0}^{M-1} p_{SX}(n_i, x_n)} \tag{6}$$

Suppose that all locations are equiprobable, i.e., $p_S(n_m) = 1/M$ for any n_m. Accordingly, the uncertainty on S given in eq. (4) can be calculated as $H(S) = \log_2 M$.

Also, suppose that the hunter attacks the WSN at node n^*. Following the attack, the hunter detects radio activity if the path between the source node and the sink passes through the node n^* and viceversa. Accordingly, X can assume only two values:

$$X = \begin{cases} 0 & \text{if there is no radio activity at node } n^* \\ 1 & \text{if there is radio activity at node } n^* \end{cases} \tag{7}$$

As a consequence, we can rewrite eq. (5) as

$$H(S|X) = \sum_{m=0}^{M-1}\sum_{x=0}^{1} p_{SX}(n_m, x) \log_2 \left(\frac{1}{p_S(n_m|x)}\right) \tag{8}$$

In eq. (8) we need to calculate the probability $p_{SX}(n_m, x)$ which can also be used in eq. (6) to calculate $p_S(n_m|x_n)$. The probability $p_S(n_m|x_n)$ is given by

$$p_{SX}(n_m, x) = p_X(x|n_m) \cdot p_S(n_m) = p_X(x|n_m)/M \tag{9}$$

Observe that $p_X(x|n_m)$ represents the probability that a packet generated by node n_m does not pass through n^*, if $x = 0$, and that such packet passes at least ones through n^*, if $x = 1$.

Now we will calculate $p_X(1|n_m)$; once this is known, $p_X(0|n_m)$ can be easily evaluated as

$$p_X(0|n_m) = 1 - p_X(1|n_m) \tag{10}$$

Recall that $p_X(1|n_m)$ is the probability that a packet generated by node n_m passes through node n^* at least once before reaching the sink n_0. Let V be the random variable representing the hop at which the packet visits for the first time node n^*. Applying the theorem of the total probability, the probability $p_X(1|n_m)$ can be calculated as the sum of probabilities that a packet generated by node n_m visits at the V-th hop node n^*, for any value of V, i.e.,

$$p_X(1|n_m) = \sum_{v=0}^{\infty} p_{XV}(1, v|n_m) \tag{11}$$

The probability in the sum in the right handside of eq. (11) is the probability that the packet generated by n_m does not visit node n^* and does not reach the sink until hop $(v-1)$, and, finally, at the v-th hop visits node n^*.

This can be calculated as:

$$p_{XV}(1, v|n_m) = w^{(m)} \cdot G^v \cdot [w^{(n^*)}]^T \tag{12}$$

where:

- $w^{(j)}$ is an array of $M-1$ elements, $w^{(j)} \in \Re^{M-1}$, all set equal to zero, with the exception of the j-th element which is equal to 1, i.e.,

$$[w^{(j)}]_i = \begin{cases} 0 & \text{if } i \neq j \text{ and } 1 \leq i \leq M-1 \\ 1 & \text{if } i = j \text{ and } 1 \leq i \leq M-1 \end{cases} \tag{13}$$

- G is an $[M-1] \times [M-1]$ matrix, $G \in \Re^{[M-1]\times[M-1]}$, and its generic element $[G]_{[i,j]}$ represents the probability that a packet received by node n_i will be relayed to node n_j, with $n_j \neq n^*$, and is not relayed by node n^*. This can be obtained as follows:

$$[G]_{[i,j]} = \begin{cases} [Q]_{[i,j]} & \text{if } j \neq n^* \\ 0 & \text{if } j = n^* \end{cases} \tag{14}$$

- $[w]^T$ represents the transponse of the array w.

Substituting eq. (12) in eq. (11) we can easily obtain:

$$p_X(1|n_m) = w^{(m)} \cdot \left[\sum_{v=0}^{\infty} G^v\right] \cdot [w^{(n^*)}]^T \tag{15}$$

By applying the spectral decomposition to matrix $G = D \cdot B \cdot D^{-1}$, where B is a diagonal matrix containing the eigenvalues β_i of G and D is the matrix whose columns are the corresponding eigenvectors, we can rewrite eq. (15) as follows:

$$p_X(1|n_m) = w^{(m)} \cdot D \cdot \left[\sum_{v=0}^{\infty} B^v \right] \cdot D^{-1} \cdot [w^{(n^*)}]^T \tag{16}$$

We call K the sum in the right hand side of eq. (16). We can easily obtain that K is a diagonal matrix whose generic element is

$$[K]_{[i,j]} = \begin{cases} 1/(1-\beta_i) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Accordingly, eq. (16) can be rewritten as

$$p_X(1|n_m) = w^{(m)} \cdot D \cdot K \cdot D^{-1} \cdot [w^{(n^*)}]^T \tag{18}$$

where K has been calculated in eq. (17).

Once $p_X(1|n_m)$ has been calculated we have all the parameters required for the calculation of the uncertainty on S given X, i.e., $H(S|X)$.

Note that the value of privacy loss ρ depends on n^*. Since the hunter knows the structure and protocols of the network, the node n^* which maximizes the privacy loss will be selected. As a consequence, the WSN gives a privacy loss γ given by $\gamma = \max_{n^*}\{\rho\}$.

3.2 Energy Consumption

The energy consumption for routing a packet from its source to the sink can be calculated as the product of the energy cost for a single hop transmission, c, and the number of hops between the source node and the sink[2]. We call Z the random variable representing the number of hops needed for a packet to reach the sink. The average energy consumption, ϵ, needed to route a packet to the destination can be calculated as

$$\epsilon = c \cdot E\{Z\} = c \cdot \sum_{z=1}^{\infty} z \cdot p_Z(z) \tag{19}$$

where $E\{Z\}$ represents the average value of Z and $p_Z(z)$ represents the probability that the number of hops between the source and the destination is equal to z. The probability $p_Z(z)$ is the probability that a packet does not reach the sink in $(z-1)$ hops and finally arrives at the sink at the z-th hop. Therefore, it is easy to show that $p_Z(z)$ can be written in compact form as

$$p_Z(z) = \pi^{(S')} \cdot P^{z-1} \cdot \omega^T \tag{20}$$

where

[2] Observe that c can also take possible retransmissions into account. In this sense, analysis of c is simple and not reported in this paper for space constraints.

- $\pi^{(S')}$ is an array of $(M-1)$ elements, $\pi^{(S')} \in \Re^{M-1}$. Its generic element is given by:
$$\left[\pi^{(S')}\right]_m = p_S(n_m) = 1/M \quad \text{with } 1 \leq m < M. \tag{21}$$
- P is an $[M-1] \times [M-1]$ matrix, i.e., $P \in \Re^{[M-1]\times[M-1]}$. Its generic element $[P]_{[i,j]}$ represents the probability that a packet received by node n_i is transmitted to node n_j, with $n_j \neq n_0$. Accordingly, the generic element of P is given by:
$$[P]_{[i,j]} = [Q]_{[i,j]} \quad \text{if } i \text{ and } j \in \{1, 2, ..., M-1\} \tag{22}$$
- ω is an array of $M-1$ elements, i.e., $\omega \in \Re^{M-1}$. Its generic element $[\omega]_m$, with $1 \leq m < M$ represents the probability that a packet is relayed by node n_m to the destination. Accordingly,
$$[\omega]_m = [Q]_{[m,0]} \tag{23}$$

Applying the spectral decomposition of P and following a procedure analogous to that presented in Section 3.1, we can rewrite eq. (19) as

$$\epsilon = c \cdot \pi^{(S')} \cdot T \cdot H \cdot T^{-1} \cdot \omega^T \tag{24}$$

In eq. (24) the matrix H is a diagonal matrix and its generic element is

$$[H]_{[i,j]} = \begin{cases} 1/[(1-\lambda_i)]^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

where λ_i is the i-th eigenvalue of P and T is the matrix of the eigenvectors of P.

4 Numerical Examples

In this section we apply the proposed analytical framework to describe how this can be used to evaluate the tradeoffs between location privacy and energy efficiency in WSN.

We consider a network of M sensor nodes uniformly distributed on a squared area of size 1 km × 1 km. We assume that all sensor nodes have coverage radius equal to $R = 200$ m. Once position of sensor nodes is set and the value of the parameter p, characterizing the random routing, is known, it is possibile to construct the routing matrix Q as given in eq. (2).

Starting from the routing matrix Q it is possible to evaluate the privacy loss γ and the average energy consumption ϵ as reported in Section 3.

All values in the following figures have been evaluated as the average of the results obtained in 20 cases. For each case, a new distribution of sensor nodes has been generated. Moreover, for each case individual routes are chosen considering the same sink node and a source node chosen in a random fashion based on uniform distribution.

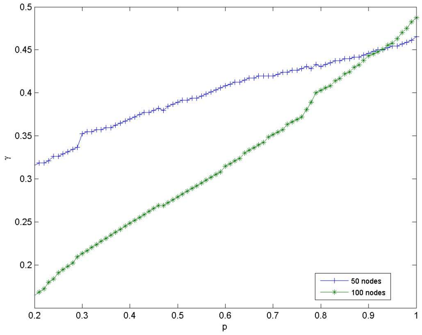

Fig. 3. Privacy loss, γ, versus the probability p for different values of the number of sensor nodes, i.e., $M = 50$ and $M = 100$

In Figure 3 we show the privacy loss, calculated as described in Section 3.1, versus the value of the probability p for two different values of the number of nodes, i.e., $M = 50$ and $M = 100$.

In Figure 3 the privacy loss increases as the probability p becomes higher. This is an expected result. Indeed, using low values of p makes the routing behavior fuzzy and therefore, the hunter cannot obtain significant information attacking the network.

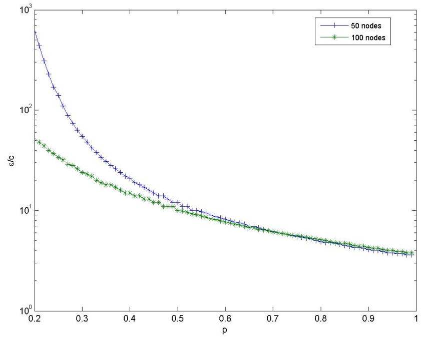

Fig. 4. Normalized energy consumption ϵ/c versus the probability p for different values of the number of nodes, i.e., $M = 50$ and $M = 100$

In Figure 4 we show the average energy consumption to deliver a packet to the sink, ϵ, versus the probability p for $M = 50$ and $M = 100$. More specifically, in Figure 4 we show the values of the ratio ϵ/c. We present normalized energy consumption values because c depends on the specific communication technology utilized, and not on the routing algorithm. As expected, the energy consumption decreases as the value of the probability p increases; furthermore, the higher the

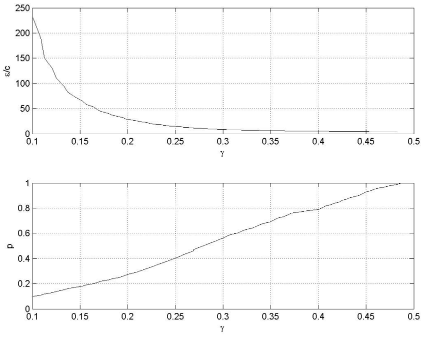

Fig. 5. (Upper plot:) Average energy consumption ϵ versus privacy loss γ and **(Bottom plot:)** The value of the probability p versus the corresponding privacy loss γ

number of sensor nodes M, the lower the energy consumption. This is because, if there are more nodes, it is likely to find better next relays than in case there are few nodes.

To highlight the tradeoff between privacy loss and energy efficiency, in Figure 5 we show two plots. In the upper plot we represent the normalized energy consumption, ϵ/c, versus the corresponding value of the location privacy loss, γ. As expected, the privacy loss increases as the energy consumption decreases. This figure has been obtained considering $M = 100$ nodes and can be utilized by the designer to select an appropriate tradeoff between energy efficiency and privacy. Once a point in the curve is chosen, the designer can use the bottom plot to obtain the corresponding value of p that gives the selected performance.

5 Conclusion

In this paper we have presented an analytical framework for the evaluation of the tradeoff between location privacy and energy efficiency in a WSN applying random routing to increase privacy protection. The proposed framework is based on a Markov-based modeling of the random routing behavior. The framework can be used by network designers to evaluate the most appropriate value of the probability p characterizing the random routing behavior in accordance with the application requirements.

Acknowledgments

This paper has been partially supported by European Commission under contract DISCREET (FP6-2004-IST-4 contract no. 27679).

References

1. D. Agrawal, C. Aggarwal. On the Design and Quantification of Privacy Preserving Data Mining Algorithms. *Proc. of the Twentieth ACM SIGACT-SIGMOD-SIGART*, Santa Barbara, California, USA. May 2001.
2. M. Anand, Z. G. Ives, I. Lee. Quantifying Eavesdropping Vulnerability In Sensor Networks. *Department of Computer & Information Science, University of Pennsylvania*, 2005.
3. P. Kamat, Y. Zhang, W. Trappe, C. Ozturk. Enhancing Source-Location Privacy in Sensor Network Routing. *Proc. of International Conference on Distributed Computing Systems (ICDCS 2005)*, Columbus, OH, USA. June 2005.
4. C. Ozturk, Y. Zhang, W. Trappe, M. Ott. Source-Location Privacy for Networks of Energy-Constrained Sensors. *In Proc. of IEEE IEEE Workshop on Software Technologies for Embedded and Ubiquitous Computing Systems (WSTFEUS)*, Vienna, Austria. May 2004.
5. Y. Xi, L. Schwiebert, W. Shi. Preserving Source Location Privacy in Monitoring-Based Wireless Sensor Networks. *Department of Computer Science, Wayne State University*, 2006.
6. DISCREET Project, State of the art Deliverable. *http://www.ist-discreet.org/Deliverables/D2103.pdf*
7. S. Haykin. Communication Systems, 4th edition.