Wireless Sensor Networks and nodes information collecting ability

Elma Zanaj, Indrit Enesi

Department of Electronic and Telecommunication, Polytechnic University of Tirana, Albania

Abstract: **The efficient use of energy in Wireless Sensor Networks represents one of the most important constrains for the management of long term communications. In this paper, we propose a new definition that of node information collecting ability. Numerical simulations will show how this parameter will influence in the network lifetime, reliability and in the quantity of information that the network collects.**

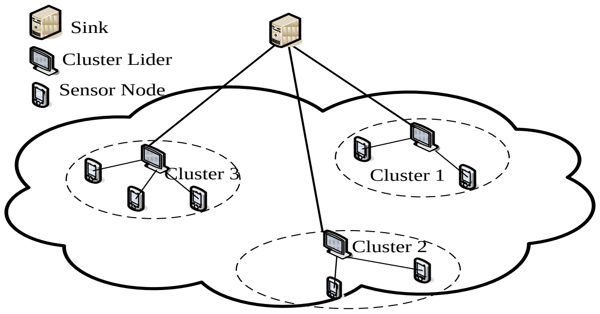
Keywords: wireless sensor networks, nodes information collecting ability.

1. Introduction

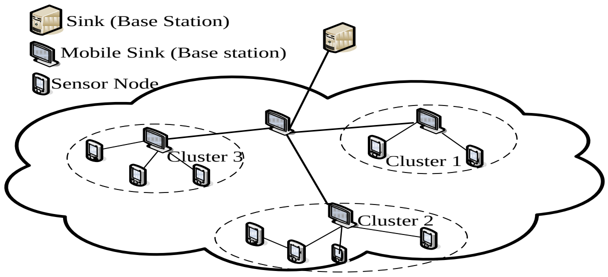
During the use of Wireless Sensor Networks (WSN) the control of efficent use of energy it is very important, because these sensors are supplied by batteries whose quantity of energy is limited and the substitution or recharging of them sometimes may be difficult. That's why the right use of energy will be take in consideration during the network implementation not only for the single node but to perform also the communication in the whole network.

The sensor node contains the units: the sensing (sensor and converters AD/DC), processing (memory and microprocessor), communication (radio) and power (battery) [1,2]. The sensors detect or monitor physical parameters or qualities of environment where are placed [1], like: light, sound, temperature, the quality of the air or water, etc. In different applications results that wireless sensors have more advantages than wired sensors [3]. According to the way the data are collected WSN are classified in three topologies: a) homogenous, b) heterogenous and c) hybrid [4].

Homogenous topology, consists in a base station/sink and sensor nodes. An example of such topology is the hierarchial topology where the sensor are organized in clusters and the cluster lider collects/distributes the data, Fig.1.a.The data collection at the cluster lider reduce the number of messages send to the sink, by improving the efficiency use of the network energy. The heterogenus topology consists in a fix or mobile sink and sensor nodes that have more advanced elaboration and procession abilities than a simple node, Fig.1.b. The data are collected at a base station [5]. Sensors often are distributed far away from each-other and the distance betwen them increase the energy of communication. The nodes at a WSN collect data from the enviroment and communicate them for a long time, and as experimental results show, the data collection at a sink extend the lifetime of the network [6]. At the hybrid networks, some mobile sinks collaborate together for collecting data in real time.



a) homogenous



b) heterogenous

Figura 1. Network topology : a) homogenous; b) heterogenous

The performance of a WSN in general is measured by two parameters: lifetime and quantity of information collected during it. A high value of the lifetime of the network means that the quantity of information collected during this lifetime should be at maximum. Recently network lifetime researches intend rather to provide new concepts which implementation in the network can vary the algorithm behavior than to improve the existing algorithms. As the energy is the main constraint of WSN, some definitions are given for the network lifetime, like : a) the death of network determined by the first node death [7] ; b) the death of the last node determines the network lifetime [8]. The quantity of information collected according to (a), is small, whereas more information is extracted, in the (b). The above definitions are energy-oriented, but as every process is transformed into information in the network, is right to seek definitions or parameters that considerate the information generated. This paper will present a new concept, on the basis of which the network lifetime is judged: the nodes information collecting ability .

The remainder of the paper is organized as follows. In Section 2, recall some theoretical analysis regarding: the define of threshold of the maximum network lifetime and the quantity of the information collected: for the non-linear and linear model, will describe our new concepts of information collecting ability at nodes and the network information reliability. Section 3 will be focused on simulated results, thus proving the role of the choice of the threshold. Finally, Section 4 will conclude the paper.

2. Theoretical analysis

2.a. Collection ability of a WSN

The WSN constructed for the data collection periodically consist in nodes that take care of their batteries power constrain. Each of them communicates only with the nearest neighbor and swich off when a message not adressed to it is send. Each sensor, during it lifetime, τ_i , collectes data with an assigned sampling rate, σ_i . Every node of the network has its individual information weight, w_i , that reveals the importance of the sensed data, accordingly to the distance from the event. The information collected from each node during its lifetime is: $w_i \cdot \sigma_i \cdot \tau_i$, while the information collected from the network is the sum of the information at each node, [9]. The network information collecting ability of a WSN is:

$$\sum_{i=1}^{l} w_i \cdot \sigma_i \qquad (1)$$

where l are only the active nodes. We are going to use the normalized information collecting ability of the network, Δ , that at time k , is the collection ability of the network at this time, division with the intial value collection ability of the network. Based on this parameter the network will continue to work or not. The user of an application will set a threshold for the normalized information collecting ability that is $\eta \in [0,1]$. At a time k if $\Delta \geq \eta$, the network doesn't satisfy the request of the applications and it ought to stop. So, the network lifetime depends on the quantity of information that the network generates. This is an information-oriented defintion of the lifetime for the network. There are two models, [9], to define the threshold of the network maximum lifetime and the quantity of the information collected: a) non-linear optimization model, so near the reality but complicated to solve; b) relaxed linear optimization model, that shows a high threshold. In the paper we will introduce a middle way to resolve the non-linearity of (a) and the non precision of (b). The algorithm used here is the asincronous broadcast [10], because WSN are naturally suited to apply broadcast-like algorithms, where the transmission from a sensor can be received by more nodes simultaneously. In a wireless network, when a node transmits some information, all the other nodes in its coverage area are able to receive the transmitted data. At the end of iteration the normalized network information collecting ability is calculated, the simulation continues only if it is over a threshold.

2.b. Non-linear model

The network is compound by sinks and sensor nodes. The nodes collect informations form the enviroment and after send them to the nearest sinks with a constant frequency for a single node σ_i (packets/second). Each sensor contibutes with its information in the network for the time that it is active τ_i or for the time that the network is active, τ_{sys} , because the networks lifetime, may be different from the single node's lifetime. Even if the node's lifetime is more than the network lifetime, the node can not contribute with its information because the network is not active any more. So, the useful working time for a single node is: $\tau_{wi} = \min(\tau_i, \tau_{sys})$, and its information is: $w_i \cdot \sigma_i \cdot \tau_{wi}$. The main goal of this model is that the algorithm will find the maximum of the total information of the network: $\max \sum_{i \in SN} w_i \cdot \sigma_i \cdot \tau_{wi}$ (SN-Sensor Nodes). The node's energy, [9], is divided in: data sensing (β -a factor for each sensed packet); *data receving* (e_{ji}^r - the energy for packet for a node i to receive it from j); *data transmiting* (e_{ij}^t -the energy for a packet for a node i to send it to j). The initial energy of a node i is E_i^0 , so the consumed energy is (for each $i \in SN$):

$$\beta \cdot \sigma_i \cdot \tau_i + \sum_{j \in SN} e_{ji}^r \int_0^{\tau_i} f_{ji}(t) \cdot l_{ji}(t)dt + \\ + \sum_{j \in S} e_{ij}^t \int_0^{\tau_i} f_{ij}(t) \cdot l_{ij}(t)dt \leq E_i^0 \qquad (2)$$

where:

$S = Sensor\_(Nodes + Sinks)$;

f_{ji} and f_{ij} , are the flow rates of each node towards the others;

l_{ji} and l_{ij} , are the link's state between nodes, that is one if there is a connection and zero if it misses.

This model respects the flow conservation constrain for the packets that a single node sends and receives, [9], is (for each $i \in SN$):

$$\sigma_i \cdot \tau_i + \sum_{j \in SN} \int_0^{\tau_i} f_{ji}(t) \cdot l_{ji}(t)dt = \sum_{j \in S} \int_0^{\tau_i} f_{ij}(t) \cdot l_{ij}(t)dt \tag{3}$$

So, then the normalized information collecting ability, is:

$$\Delta = \frac{\sum_{i=1}^{n} w_i \cdot \sigma_i \cdot H(\tau_i, k)}{\sum_{i=1}^{n} w_i \cdot \sigma_i}, \tag{4}$$

where: $H(\tau_i, k) = \begin{cases} 0 \rightarrow k \geq \tau_i \\ 1 \rightarrow k \leq \tau_i \end{cases}, \forall i \in SN$

$H(\tau_i, k)$ reflects the fact that the topologies change, and the information collecting ability of the inactive nodes is not taken in consideration. After an interval of time the decreasing of Δ is influenced by the inactive nodes, so for a network that fulfills the required features : $\Delta \geq \eta$. As nodes become inactive at different times, for obtain a linear system the traffic's rate and the link's state must be stable for all the time. This means that the topology ought to be constant during the network lifetime, and this can be possible only if all the nodes become inactive at the same time or have the same lifetime with the network.
So, this is way we obtain the *relaxed linear model.*

2.c. Relaxed linear model

The relaxed linear optimization model must respects the flow conservation constrain with the non linear model, when topologies change. The lifetime of a single nodes changes, but it conserves the flow conservation constrain by fulfill the relation:

$\sigma_i \cdot \tau_{w_i} = \sigma_i^{'} \cdot \tau_{w_i}^{'}$, as: $\tau_{wi} = \tau_{sys}$, then: $\sigma_i \cdot \tau_{w_i} = \sigma_i^{'} \cdot \tau_{sys}$,

where: $\sigma_i^{'}$, is the frequency changed at each node, due to the flow conservation constrain, the lifetime changes, too. There are two cases: 1) if the node lifetime is small and it increases, the frequency will decrease; 2) if the node lifetime is high and it decreases then the frequency will increase. In this model the topology does not change, also the links state and the total packets number, so the integral can be neglected, from the (2) and (3) of the flow and energy conservation, so for each $i \in SN$:

$$\sigma_i^{'} + \sum_{j \in SN} f_{ji} \cdot l_{ji} = \sum_{j \in S} f_{ij} \cdot l_{ij} \tag{5}$$

$$(\beta \cdot \sigma_i^{'} + \sum_{j \in SN} e_{ji}^{r} \cdot f_{ji} \cdot l_{ji} + \sum_{j \in S} e_{ij}^{t} \cdot f_{ji} \cdot l_{ij}) \cdot \tau_{sys} \leq E_i^0 \tag{6}$$

While the normalized information collecting ability is simplified in:

$$\Delta = \frac{\sum_{i \in SN} w_i \cdot \sigma_i^{'}}{\sum_{i \in SN} w_i \cdot \sigma_i} \geq \eta \tag{7}$$

2.d. Information collection node ability

Above a relation between the normalized network information collecting ability and the network lifetime is described. So, for a given threshold if this value is lower than it, the network doesn't worth to work. Some numerical examples will be given now to make it more clear.

So, for a WSN with 50 nodes, where: $\eta = 0.1$, $w_i = 1$, $\sigma_i = 1$ packets/sec and 5 nodes become inactive each t seconds. The normalized information collecting ability required is 0.1, which happens when only five nodes are active in the network. If every t seconds 5 nodes die then 45 nodes will die during $45/5 = 9 \cdot t$ seconds. If $\eta = 0.3$, for the same considerations, only 15 nodes are active, so for each t seconds 5 nodes die then 35 nodes will die at $7 \cdot t$ seconds. So, the network lifetime and the quantity of information collected will increases, if η decreases.

Let's consider two networks with different number of nodes, but with: $\eta = 0.6$, $\sigma_i = w_i = 1$ and every t seconds 25 nodes die.

1) $n = 50$. The normalized information collection ability at a time τ, is: $\Delta = 25/50 = 0.5 < 0.6$, the requested feature isn't satisfied this time.

2) $n = 250$. The normalized information collection ability at a time τ, is: $\Delta = 225/250 = 0.9 > 0.6$. Now, the network satisfies the requested feature, and it will work until Δ becomes less than 0.6.

The network lifetime at the second case will be longer than the one's with 50 nodes, also the quantity of information. Before we described two models and both of them have some drawbacks. The non linear model uses a dynamic topology, near the reality but difficult for the simulations implementation, while the relax model is easy for the implementation but far away from the reality. For the resolving the situation an intermediate solution that improves the dynamic model will be given during this study. At the dinamic model each node has its individual lifetime that depends on its communications and other parameters mentioned before. The normalized information collection ability is compared with η (whose choice depends on the application), at a time k, after the function H must be determined. It depends on node's lifetime, τ_i, that is defined during the algorithm's simulation, so a single value is decided by the integral, this implies the non linearity mentioned before. What we propose is that the information collection ability to be not a global term but a local parameter, at each node. At the time k, the value of H is generated, by comparing the lifetime of a node with k. Once decided the value of η, it will be used as a local parameter, furthermore a new parameter will be introduced: the probability that a node i is active P_i, at time k, as:

$$P_i = \frac{E_i^k}{E_i^0} \qquad (8)$$

where: E_i^k is the actual value of node's energy, while E_i^0 its initial energy, [11]. The probability for each node to be active will compare with η, if it is smaller than η, its $H(\tau_i, k)$ is zero, otherwise is one. This parameter reflects topology changes and results in a model near the reality. Now we will show how much this new parameter influeces the optimization of the network.

Simulations are done over a Random Geometric Graph (RGG) topology, where n nodes are randomly distributed in the simulation surface and the network taken in consideration is fully conected. The algorithm applied is the asincronous broadcast [10]. At time k, randomly a node i becomes *leader*, and delivers its information to others j by using for each a weight of communication w_{ij}. The w_{ij} are elements of a communication's weights matrix W that is a stochasic matrix and is construced as follows. When an event happens at a network all the nodes will try to take this information, but only the node with the lowest distance from it will be definite as the base. It will be the element for pondering the distance of the others node from the event. So, the weight of information for each node i of the network, is proportional with its distance from the event divide the distance of the base. While the weight of information for a couple of node that will communicate, will be the average of their information's weight. So this way all the elements w_{ij} of the stochasic matrix W are calculated.

Another characteristic that will be described here is the predefined sampling rate σ_i, of each node i. The nodes nearest the event have larger weights of information than the others. As is our interes that these nodes to be active for a long time, they will have a lower sampling rate. If R is the maximal distance that the farest away node has from the event, then this distance is divided in three zones, Fig.2. The smallest distance corresponds to the lowest sampling rate so: a) for $r = R/3$ it is $\sigma_i = 1$;b) for $R > r > 2R/3 \Rightarrow \sigma_i = 2$ and for $r = R \Rightarrow \sigma_i = 3$.

Another parameter that depends on the distance is also the node reliability, the reliability of the information that a single node generates, is inverse propotional with the distance.

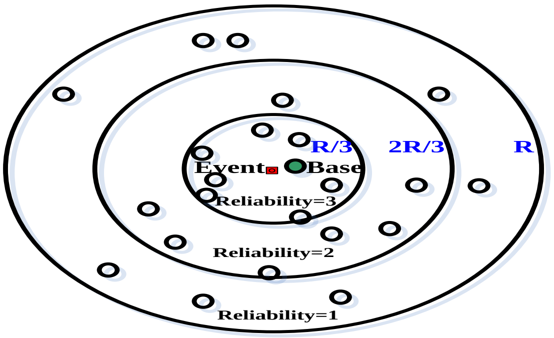


Figure 2. The reliabilty zones based on the distance from the event.

It means that regarding to the Fig.2, for nodes with distance $r = R$ reliability is 1, while for $R > r > 2R/3$ it is 2, and for $R/3$, 3. The network reliability is the average of all nodes reliability.

3. Simulations in Matlab

The first purpose is to demostrate how the Network Lifetime (N.L) decreases with the increasing value of η. This occurs because the network stops at the moment when the normalized information collection ability is less then a threshold, depending also on the number of active nodes at that time. For a low number of active nodes in the network the value of the normalized information collection ability will be also low and vice versa. However is not the value of the normalized information collection ability that determines the death of the network but the threshold η. If the value of η is low, it means that the value of the normalized information collection ability at network "death" time will be low, so there are few active nodes in the network. So it has been a lot of time since the network was active. While a high value of η at the network "death" time requires a high value of the normalized information collection ability, means a considerate number of nodes to be active. The choice of the η depends on the application, so in the case when is required at the moment of the information collection, a high number of nodes is required to be active then we only ought to choose the right value of η. Fig.3, indicates the behaviour of the N.L for a WSN with $n = 50, 250, 500$ nodes and the simulation of the network in two other cases: First Node Dies (F.N.D), [12,13], and Last Node Dies (L.N.D), [8]. In our simulations: $e^R = 0.005$; $e^T = 0.3$; $E^0 = 100$. As expected, it demostrates that the network lifetime decreases with the increasing value of η, and the band created by L.N.D and F.N.D increases with the number of nodes, [14]. The normalized information collection ability doesn't increase or decrease the network lifetime, but it influences in a better usage of the network sources. For the same: topology, nodes and algorithm, but for a different choice of η, the results are different. However the optimal value it is not decided only from the required network lifetime but also some other parameters that will be mention in the following.

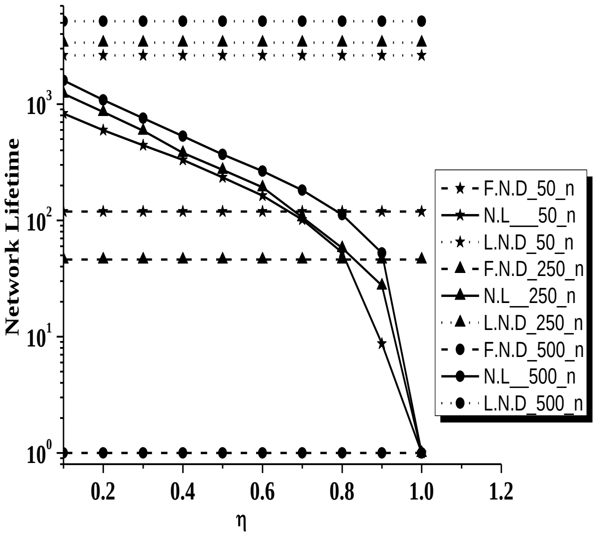


Figure 3. Network .Lifetime for different $n = 50; 250; 500$.

The Network Lifetime (N.L), determined by the time where the first node die, means a low network lifetime. In such network, the sources of energy will be not efficently used, because may be some nodes will not communicate. Fig.3, demonstrates that for: $\eta < 0.65$ (for $n = 50$), $\eta < 0.8$, (for $n = 250$) and $\eta < 1$ (for $n = 500$) the network lifetime of the simulated broadcast is higher than the case where the first node dies, whereas for $\eta > 0.65 \Rightarrow n = 50$ the network lifetime is lower than in the case of F.N.D (the same thing for the other number of nodes). Another definition of the network lifetime, is the time where the last nodes dies, which is the extreme case of the network definition for the first node dies. The received network lifetime values are unobtainable even when we applied the node information collection ability, as in Fig.3. In this case the quantity of collected information will be maximal. Anyway, this case has a problem, because the last active node can be far away from the event and its information weight will be low. In both cases a node is dead only if its energy is lower than the energy required to communicate with its nearest neigbour. In both simulations of the death of the first and the last node algorithm, Fig.3, the network lifetime doesn't depends on the value of η, but being the simulations of two extreme cases, give us an idea how the network lifetime can be in these particular cases.

Before we mentioned that there are some parameters that will influence the values of the normalized information collection ability, and one of them is the Network Reliability (N.R). Fig.4 illustrates the results of the normalized N.L and the N.R as a function of η, for different number of nodes, where the N.R increases with η, while the network lifetime decreases. A high value of η requires that a large number of nodes will be active, and this influence N.R, because for a large number of nodes the N.R will be high and vice versa.

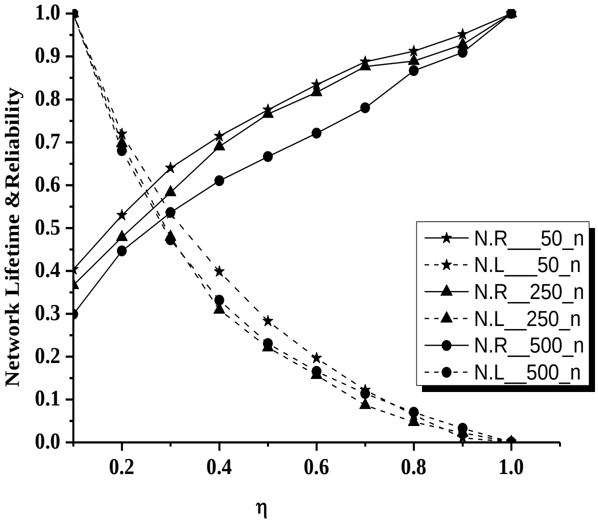


Figure 4. Normalized N.R and N.L in function of η , for different n .

A high influence at the N.R has also the nodes position at the time or to be exact the level of N.R of the single node, that depends on the nodes distribution. So, it is important which nodes are active at a time in the network. Let's make it clear with a simple example. If a network has only 5 active nodes at a time t , and both of them are placed at the $R/3$ zone, then the value of N.R will be higher than the case when only one of them will be placed at $R/3$. This is the main reason why our simulations are taken as an average of 100 interations, because as mention before nodes are randomly distributed over a RGG topology. Let's make some consideration for the value of the N.R in two cases: F.N.D and L.N.D. It is clear that N.R is higher at the F.N.D than in L.N.D, because at a first case network dies when many of nodes are active.

It is important that the choice of η satisfies a precise value of N.L and N.R. So, N.R curve at Fig.4, intersects N.L curve at $\eta = 0.26$ (for n=50), $\eta = 0.27$ (for n=250), $\eta = 0.28$ (for n=500), but it doesn't mean that these are the optimal choices, because they depends on application considerations.

The simulations at Fig.4 take in consideration the value of N.L and N.R in function of η. However there are many applications that are not interest for a high value of N.R when N.L is low, but require a high quantity of information even for not a high value of N.R. That is the case with of data mining process where a large quantity of information is more requested then a high level of reliability. The value of N.L and N.R, help us to find indirectly the quantity of information that a network collects. The Quantity of Information (Q.I) is the number of packets generated at the network during its lifetime, [9].

The N.L for a small value of η, for e.g.: $\eta = 0.2$ is higher than for a high value for e.g.: $\eta = 0.9$ (for all the simulated curves with different number of nodes), so a priory the quantity of information at $\eta = 0.2$ will be also high. However our simulated model has the particularity that the nodes have different sampling rates and individual information weights. So, two packets generated at a node with $w_i = 1$, don't have the same worth with two packets generated at a node with $w_i = 3$.

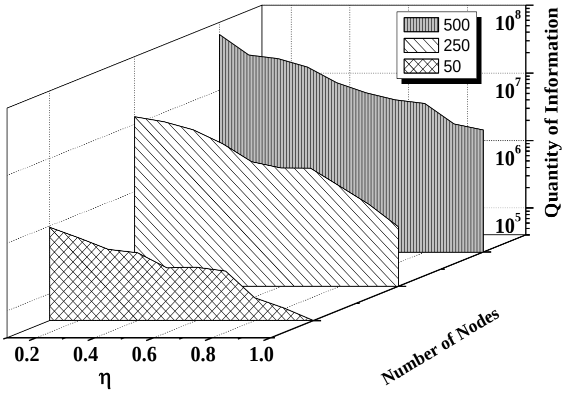


Figure 5. Network Q..I in function of η, for different n.

We are interest to find how the choice of η influences at the value of Q.I. The simulations, Fig.5, for different number of nodes $n = 50, 250, 500$, indicates an inverse relationship between the value of η and the Q.I collected, during the network lifetime.

Let's make again some consideration for the value of the Q.I in two cases: F.N.D and L.N.D. It is clear that Q.I is higher at the L.N.D than in F.N.D, because at a first case N.L will be high also the Q.I.

When as N.L is consider F.N.D, the N.L will be minimal and even the Q.I be low. From our simulations results that Q.I, for all the number of nodes taken in consideration, decreases when the value of η increases, and this is make sense because when the network starts all the nodes are active while with the passing of time some of them become inactive.

4. Conclusions

The node collection ability as a new definition helps to provide an important result: for each value of η exists an optimal value of nodes for a precise N.L, N.R and Q.I. This is an important conclusion specially when we are interest in the cost of the implementation of the network. As a further work to develop better solutions, others parameters that can change the behaviour of existing algorithms can be taken in considerations.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey". Computer Networks, 38, 2002, pp. 393 – 422.

[2] V. Raghunatan, C. Schurgers, S. Park, and M.B. Srivastava, " Energy-aware wireless Microsensor networks", Proc. IEEE Signal Processing Magazine, 19, 2002, pp. 40 – 50.

[3] F. Zhao and L. Guibas, "Wireless Sensor Networks: An Information Processing Approach" Morgan Kaufmann Publishers, San Francisco, CA. 2004.

[4] H. Nakayama, N. Ansari, A. Jamalipour, Y. Nemoto, and N. Kato, "Fault – resilient sensing in wireless sensor networks", Proc. Computer Communications, 30, 2007, pp. 2375 – 2384.

[5] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling and analysis of a three - tier architecture for sparse sensor networks", Proc. Ad Hoc Networks, 1, 2003, pp. 215 – 233.

[6] I. Chatzigiannakis, A. Kinalis and S. Nikoletseas, "Sink mobility protocols for data collection in wireless sensor networks", in Proc. of the 4th ACM Workshop on Mobility Management and Wireless Access (MobiWac' 06),2006, pp. 52 – 59.

[7] Y. Zhang, M. Ramkumar, N. Memon, "Information flow based routing algorithms for wireless sensor networks" , in Proc. IEEE Global Telecommunications Conf (GLOBECOM'04), vol. 2, Nov. 2004, pp. 742-747.

[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", in Proc. 33rd Hawaii Int. Conf. on System Sciences, Jan. 2000.

[9] Q. Wang, T. Zhang, and S. Pettersson, "Bounding the information collection performance of wireless sensor network routing" , in Proc. 5th Annual Conf. on Communication Networks and Services Research (CNSR'07) , May. 2007, pp. 55-62.

[10] M. Baldi, F. Chiaraluce, E. Zanaj, "Fault Tolerance in Sensor Networks: Performance Comparison of Some Gossip Algorithms", Proc. 7th Workshop on Intelligent Solutions in Embedded Systems (WISES 09), Sep.2009.

[11] H. Hassaniasl, A. Rahmani, M. A. Dezfuli, A.N.Eghbali "A Novel Score-Aware Routing Algorithm in Wireless Sensor Networks", Proc. Interantional Journal of Computer Science and Security, vol.3, Nov. 2009 pp.397-404.

[12] M. Bhardwaj, A.P.Chandrakasan, "Bounding the lifetime of sensor networks via optimal role assignments", in Proc. 21st Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM'02), vol.3, Jun.2002, pp.1587-1596.

[13] J.H.Chang, L.Tassiulas, "Maximum lifetime routing in wireless sensor networks", Proc. IEEE/ACM Transactions on Networiking, vol.12, Aug.2004, pp. 609-619.

[14] E. Zanaj, I.Enesi, E.Zenelaj, "Improving parameters for Wireless Sensor Networks", Proc. NBiS 2011, Sep.2011.