Equitable Clustering algorithm for Wireless Sensor Networks based on the Spectral Classification

Brahim Elbhiri[1], Sanaa El Fkihi[2], Ali Jorio[3], Driss Aboutajdine[3]
[1]EMSI Rabat, Morocco
[2]ENSIAS Mohammed V University
Souissi Rabat, Morocco
[3]LRIT, Research Unit Associated to the CNRST (URAC 29)
FSR, Med V-Agdal University
Rabat, Morocco
elbhirij@emsi.ac.ma, elfkihi@ensias.ma, jorio.ali@gmail.com, aboutaj@fsr.ac.ma

ABSTRACT: *Wireless sensor network (WSN) consists of low-cost, low power and energy-constrained sensors where the objective is to monitor a physical phenomenon and report the data to a sink node. The key challenge in the design and operation of WSNs is the maximization of system lifetime by saving the energy consumption. The network clustering is considered as one of the most promising techniques largely used for dealing with the given challenge. In this paper we use the spectral clustering methods to propose the Equitable Clustering algorithm for Wireless Sensor Networks based on the Spectral Classification (ECA-WSN-SC). This protocol uses the graph theory techniques and the spectral clustering to separate the network in a fixed optimal number of clusters where each called cluster head collects data and communicates the information to the base station. Defining the optimal number of clusters and enlarging the network lifetime are the strongest characteristics of the ECA-WSN-SC protocol. Simulation results show that the proposed algorithm increases the lifetime of a whole network and presents more energy efficiency distribution compared to the Low-Energy Adaptive Clustering Hierarchy (LEACH) and the Centralized LEACH (LEACH-C) one.*

Keywords: WSN, Clustering, Spectral Classification, Energy Consumption

Received: 3 July 2013, Revised 17 August 2013, Accepted 22 August 2013

1. Introduction

A Wireless Sensor Network (WSN) is a wireless network consisting of distributed autonomous devices using sensors to monitor physical or environmental conditions such as temperature, sound, and vibration. It contains a large number of nodes which can sense data from an inaccessible area and send their reports towards a processing center called Base Station (BS). A WSN has many advantages, such as the ease deployment and the capacity of self-organization. However, it main challenges are the limited resources in terms of communication, computation and energy. In fact, even if the sensors are usually powered by batteries, it is not practical to recharge or replace them because they are often deployed in hostile environments. Furthermore, in a WSN, a large part of energy is consumed when communications are established [1]. Hence, frequent and long distances transmissions should be kept to minimum to extend the network lifetime [2] [3]. To deal with this, an effective approach consists in dividing the network into multiple clusters; each cluster has a clusterhead node (C-H) [4]. The later collects data from its same cluster nodes, aggregates them, and transmits them to the BS. Nevertheless, the main problem of the proposed solution is the random selection of the C-Hs. Indeed, all C-Hs can be located in a small region of the network and some ordinary nodes will be isolated. To tackle this problem and define CHs fairly localized, we propose to take advantage of spectral clustering methods.

Spectral clustering is of the most popular used clustering algorithms. It is based on graph theory, simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm [5]. It included a variety of methods based on some matrices which represent the distance between points or other properties and used them differently. The k-ways approach is one of these methods; it consists to divide a graph to k disjoint classes based on the k eigenvectors related to k largest eigenvalues of Laplacien matrix. Based on this approach, we present an equitable clustering algorithm for wireless sensor networks based on the spectral Classification (ECA-WSN-SC), which aims to extend a network life-time by distributing energy consumption, minimizing control overhead (to be linear in the number of nodes), and producing welldistributed C-Hs and compact clusters. The ECA-WSN-SC does not make any assumptions about the distribution or density of nodes, or about node capabilities.

The remaining of this paper is organized as follows: In Section 2, we provide a brief overview of some related research work. Details and properties of the proposed algorithm are given in Section 3. Section 4 evaluates the performances of ECA-WSN-SC via simulations and compares ECA-WSN-SC with some other commonly used clustering protocols. Finally, Conclusion and some work perspectives are drawn in Section 5.

2. Related Work

Many routing protocols have been designed based on clustering principle, where C-Hs are elected periodically and alternately because otherwise, C-Hs die quickly [4]. The most commonly used approach for clustering is Low Energy Adaptive Clustering Hierarchy (LEACH) [6]. And many clustering protocols based on the principle of this algorithm have been developed, such as LEACH-C [7], EECS [8], and DECSA [9]. LEACH employs a hierarchical clustering. In this protocol, nodes organize themselves into clusters using a distributed algorithm. Its main idea is to randomly and periodically select the C-H. The probability of becoming a C-H for each period is chosen to ensure that every node becomes a C-H at least once within N/K rounds; where N is the sensor number and K is the cluster number. After the election of C-Hs, each non C-H node determines its cluster by choosing the C-H that requires the minimum communication energy, based on the received signal strengths of advertisements from C-Hs. Each node will choose to join the C-H which has the highest signal quality. Then, each C-H node creates time division multiple access schedule. The C-H collects and aggregates information from sensors in its own cluster and transmits data to the BS. Although LEACH algorithm has advantages, this protocol cannot determine the placement of nodes and the number of the needed clusters for a WSN. However, using a central control algorithm to form the clusters may produce better clusters by dispersing the C-H nodes throughout the network [7]. In LEACH-centralized (LEACH-C), the BS determines CHs by computing the average nodes energy, node with energy below this average cannot be C-H. This ensures that the energy consumption is effectively distributed among all nodes.

In [10], the authors proposed a cluster scheme based on residual energy and distance from the BS by modifing the k-means clustering algorithm. In this case, the location information of each sensor node is required. In cluster formation phase, authors start by randomly selecting c C-Hs, each node is assigned to its near C-H that has the highest residual energy. Then, they re-compute the C-H by using centroid method and repeat the process. Finally, for each cluster, one node is elected as a C-H using centroid method; the C-H has the minimum distance between the cluster nodes and the centroid point. Although the algorithm improves the network lifetime and reduces network traffic, it does not consider the distance between nodes in the C-H selection step and cannot define the optimal number of cluster.

Moreover, in [8], the authors present an energy efficient clustering scheme (EECS) for periodical data gathering applications in large scale sensor networks. In the C-H selection phase, several C-Hs are elected by localized competition, which take into account the residual energy and coverage radius of candidate nodes. Then, in the cluster formation step, a weighted function for the plain node is introduced to make decision of which proper cluster should be joined. The function is calculated by the cost of intra-cluster communication and the cost of communication between the C-H and the BS.

In addition, the authors of [9] present distance-energy cluster structure algorithm (DECSA) based on the classic clustering algorithm LEACH. It considers the distance between the network nodes, the position of the BS and the residual energy of nodes. The main idea of the algorithm is to partition the network into three levels of hierarchy to reduce the energy consumption of C-Hs, resulting from the non-uniform distribution of nodes in the network and thus avoid direct communications between the BS and C-H that has minimal energy and are far away from the BS.

Besides, in [11], the authors propose a partitioning algorithm for WSN based on k-nearest neighbor. They propose to partition the network into a number of clusters where each cluster contains one BS and all sensor nodes communicate directly with the BS contained within it.

Finally, in [12], the authors present an approach called the Spectral Classification based on Near Optimal Clustering in Wireless Sensor Networks. This approach is based on spectral bisection for partitioning a network into two clusters and then applies this method recursively to obtain optimal number of clusters. However, recursive spectral bisection finally produces 2^n clusters; where n is the number of iterations. The approach cannot partition the network into any desired number of clusters.

3. ECA-WSN-SC Details

In this section, we give details of the proposed KSCAWSN algorithm. The last consists in partitioning the WSN into a set of clusters based on spectral classification. the spectral classification is widely used to solve graph partitioning problem and it can be applied for partitioning a WSN. The spectral method gets its name to the spectral theorem of linear algebra. This theorem permits to affirm the diagonalization of real symmetric matrices. It also justifies the decomposition of real symmetric matrices into eigenvalues in an orthonormal set of eigenvectors. But the graph partitioning problem can be reduced to the resolution of a numerical system $Mx = \lambda x$. Solve this numerical system is to find an orthogonal base of eigenvectors of the matrix M [13].

Thus, in our approach, the BS constructs the graph corresponding to the WSN based on the spectral clustering principle. Indeed, let x be an observation vector composed of the sensor network nodes. This vector can be represented by an undirected graph G (V, E) where V the set of vertices (sensor nodes) identified by an index $i \in \{1,..., n\}$ and E the set of edges that link each two vertices (Communication link). Let $A \in \Re^{n \times n}$ be the similarity matrix of the graph G.Each value of A is associated to each pair of the graph nodes (i, j). The total weight of edges incident to node i is given by $d_{ii} = \sum_{j=n}^{j=1} a_{ij}$. The degree matrix $D \in \Re^{n \times n}$ of G is a diagonal matrix defined by $D = [d_{ij}]$. Finally, the Laplacian matrix of the graph is calculated by :

$$L = D^{-\frac{1}{2}} A\ D^{-\frac{1}{2}} \qquad (1)$$

The spectral classification algorithm presented here can be thought of a consisting of 3 principal stages :

5.1 The Pre-Processing step

First, each wireless sensor network node determines its position by using a GPS (Global Positioning System) or other tracking device, and transmits it to the base station in a short message. Based on the spectral clustering principle, the BS constructs the graph G and the similarity matrix (A) which represents the network. Each value of (A) is associated to each pair of the graph nodes (i, j). This value is of gaussian type and the matrix (A) is given by equation (5).

$$A = [a_{ij}] = \begin{cases} \exp\left(\frac{1}{2\sigma^2} d^2(i,j)\right) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

d (i, j) is the euclidien distance between nodes i and j.
Thereafter, Base station must deduce the Degree Matrix and Laplacian matrix and compute its eigenvalues and eigenvectors.

5.2 Clustering step

The objectives of the current step are to define the optimal number of clusters and to form them. Based on the Laplacian matrix L defined above, we form a new matrix U composed of the k eigenvectors related to k largest eigenvalues of L. In order to determine the k clusters of the WSN, we apply the classification algorithm k-means to the matrix U. The sensor node i is assigned to cluster C_j if and only if row i of the matrix U was assigned to cluster C_j.

Nonetheless, the most important question raised by the proposed strategy concerns the optimal number of clusters (k) to be used. With the aim to respond to this question, we consider the total consumed energy in each round.We note that by considering k cluster, this energy depends on the distances between the cluster head and the non cluster heads of each cluster.

However, it is known that the amplifier energy in a multipath fading channel models is greater than the amplifier one in a free space model (*i.e.* $E_{fs} = E_{mp}$). In addition, we have $d_j < d_0 = d_i$. Thus, to minimize the energy per round, all non cluster head nodes must operate in a free space model.

The objective function that allows to decide whether to reconsider the partitioning process or not of the WSN, is defined by the distance matrix M_{dis}^{k} ($M_{dis}^{k} = [dis_{ij}^{k}]$); with dis_{ij}^{k} is the distance between the node i and the node j of the cluster labelled k) of each cluster. The allowed threshold to this function is d_0. Hence, if at least one element of any M_{dis}^{k} is greater than d_0, the considered number of clusters will be incremented (k +1) and the k-mean algorithm will be reused. Otherwise, the optimal number of clusters is k.

5.3 Cluster head election and Data Transmission

Once the clusters are determined, the next step consists in defining the cluster heads. the cluster head is responsible for coordination among the nodes within their clusters, aggregation of the received information and communication with the base station.

Hence, for each cluster we use the Low Energy Adaptive Clustering algorithm (LEACH). The main idea is to randomly and periodically select the cluster-head, the probability of becoming a cluster-head for each period is chosen to ensure that every node becomes a cluster-head at least once within N / k rounds, where N is the number of node in the cluster and k is the desired number of cluster head for each cluster. After the election of cluster-heads, each non-cluster-head node in the cluster determines its cluster head by choosing the clusterhead that requires the minimum communication energy, based on the received signal strength of the advertisement from each cluster-head. Each node will choose to join the cluster-head which has the highest signal quality.

Based on the id and the numbers of nodes in the cluster, a schedule TDMA (Time Division Multiple Access) will be created automatically to assign to each node a time when it can transmit his data to cluster head. If we suppose that the node with the $id = i$ is elected as a cluster-head, the node with $id = (i + 1 + |S_k|) \bmod |S_k|$ is assigned the first period to transmit its data.

One of the most important challenges in our approach consists in reducing the total consumed energy of each round. As well, we avoid energy consumption due to synchronization of the cluster nodes when the cluster-head is elected for assign the TDMA. Indeed, this technique ensures not having collisions between data messages and also allows the radio components of each non-cluster-head node be disabled at any time except during the time it is allocated for transmitting these data, which reduces the energy consumed by the nodes, however, the cluster-head must keep its receiver on order to receive data from other nodes. To save more energy in a wireless sensor network, we assume that if the distance between a node and the base station is lower than the distance between this node and its cluster-head then, the node transmits its data directly to the base station. Also, each node can move in the standby mode to reduce energy consumption.

4. Simulation And Results

In this section, we present the results of the experiments that we have done to evaluate ECA-WSN-SC algorithm. All algorithms are simulated, an analyzed and compared using MATLAB software. Parameters used in our simulations are shown in Table 1.

We refer readers to [7], for more details from different radio energy dissipation model parameters used in our simulations.

To simplify the network model, we adopt a few reasonable assumptions as follows:

- All nodes are homogeneous and have the same capacities.
- Each node is assigned as a unique identifier (ID) and includes the cluster identifier, to which it belongs.
- Network topology remains unchanged over time.
- Base station is placed far away the network.

• Nodes say that position by a GPS (Global Positioning System) or other tracking device.

• Nodes transmit at the same fixed power levels, which is dependent on the transmission distance.

A node has run out of energy is considered dead and cannot transmit or receive data. For these simulations, the energy of a node decreases each time it sends, receives or aggregate the data according to the model Radio parameters. Also, we ignore the effect caused by the signal collision and the interference in the wireless channel.

Besides, each simulation result shown below is the average of 100 independent experiments where each experiment uses a different randomly-generated uniform topology of sensor nodes.

As illustrated in Figure 1, an example of wireless sensor networks with $N = 200$ nodes randomly distributed in a $200m * 200m$ area. The Base Station is located far away of the sensing area ($x_{SB} = 100m$; $y_{SB} = 250m$).

Parameter	Value
E_{elec}	50 nJ/bit
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
Initial energy E_0	0.5 J
E_{DA}	5 nJ/bit/message
Area of Network	$200m * 200m$
Sink coordination	$(100m, 250m)$
d_0	$88m$
Message Size	4000 bytes
Number of Nodes	200

Table 1. Experimental Simulation Parameters

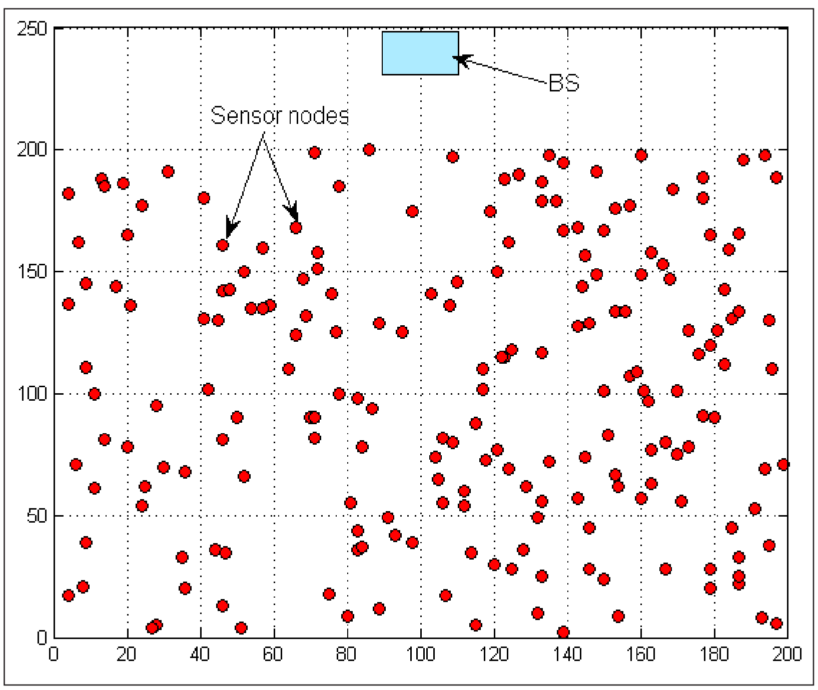

Figure 1. Distribution of Wireless Sensor Network

In order to evaluate the performances of the new proposed protocol, we propose to compare it to the LEACH distributed and LEACH-C centralized [6] [7]algorithms.

Also, we use two metrics to analyze and compare the performance of the protocols :

• **Network lifetime:** She can be defined as the time elapsed until the first node (or the last node) in the network depletes its energy (dies). This paper uses First Node Dies(FND) metrics which were presented as an approach to define lifetime of WSN in literature.

• **Energy consumption:** Uniform energy consumption is very important for network load balancing. More uniform energy consumption, less possibility for node premature death. And less energy consumption per round, more network performance.

The Fig. 2 presents an example of clustering for the ECAWSN-SC algorithm. We note that the network is subdivided into fifteen clusters and the nodes are correctly distributed over the sensing area. Also, there is no intersection between the different clusters.

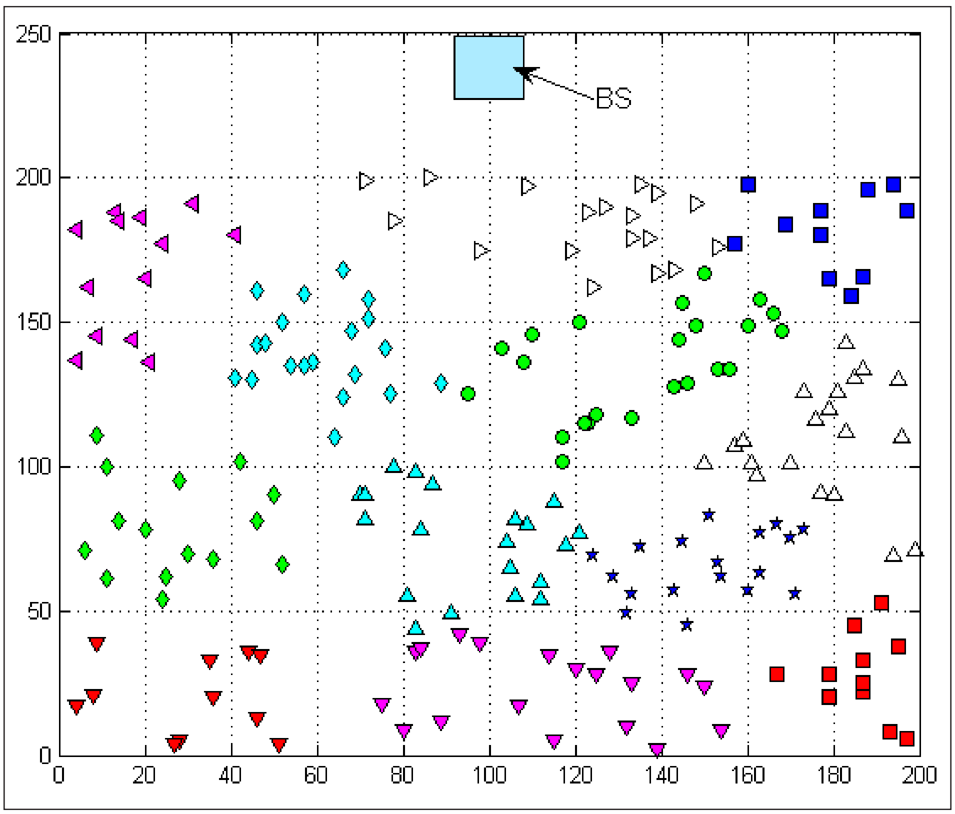

Figure 2. Clustering results with N = 200 nodes

In Figure 3 shows a significant improvement for ECA-WSNSC in terms of numbers of periods relating to death the first node. In the proposed method, death of the first node occurs at the period 343 while it occurs at the 180 period for LEACH and 21 for LEACH-C.

Figure 4 gives the total network remaining energy in every transmission round. The network remaining energy decreases more rapidly in LEACH-C and LEACH protocols than in the ECA-WSN-SC algorithm. The Figure 5 shows the effects of the node density on the compared clustering techniques as well as on the network's stable regions (First Node Dead '*FND*').

As shown in Figure 5, for different values of N equal to 200, 300 and 400, our algorithm presents an improvement of the performance compared to the others algorithms. It follows that even if the node density increases the new proposed approach still gives best results compared to the others.

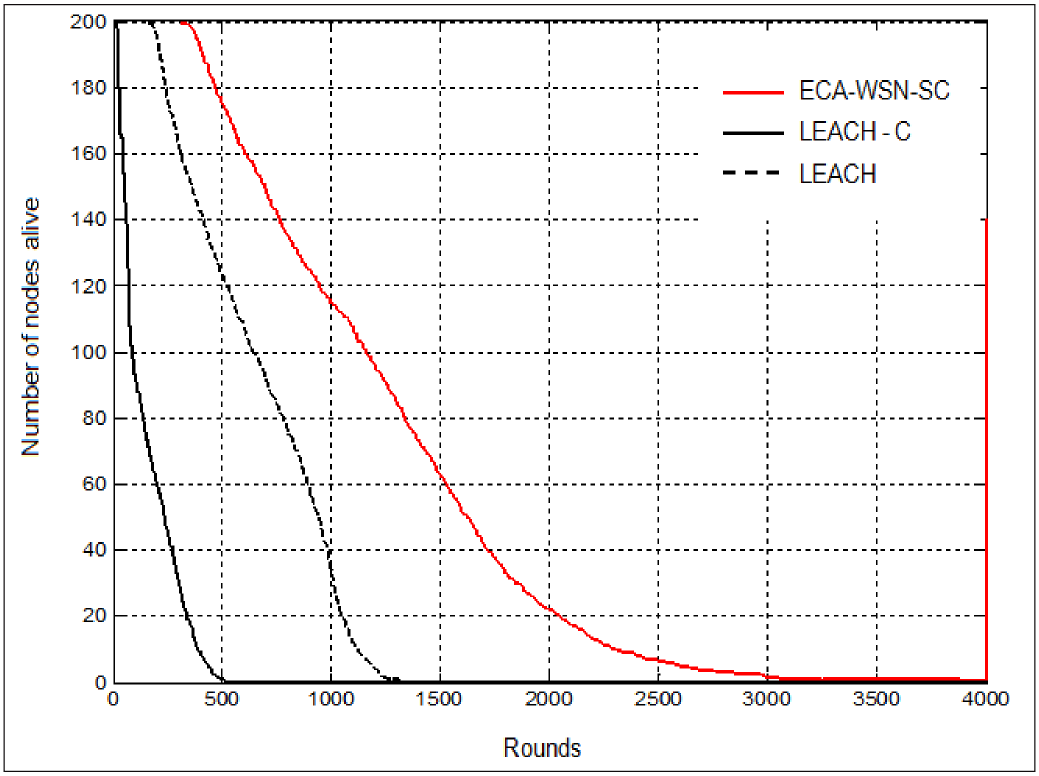

Figure 3. Number of nodes alive over time of the compared protocols

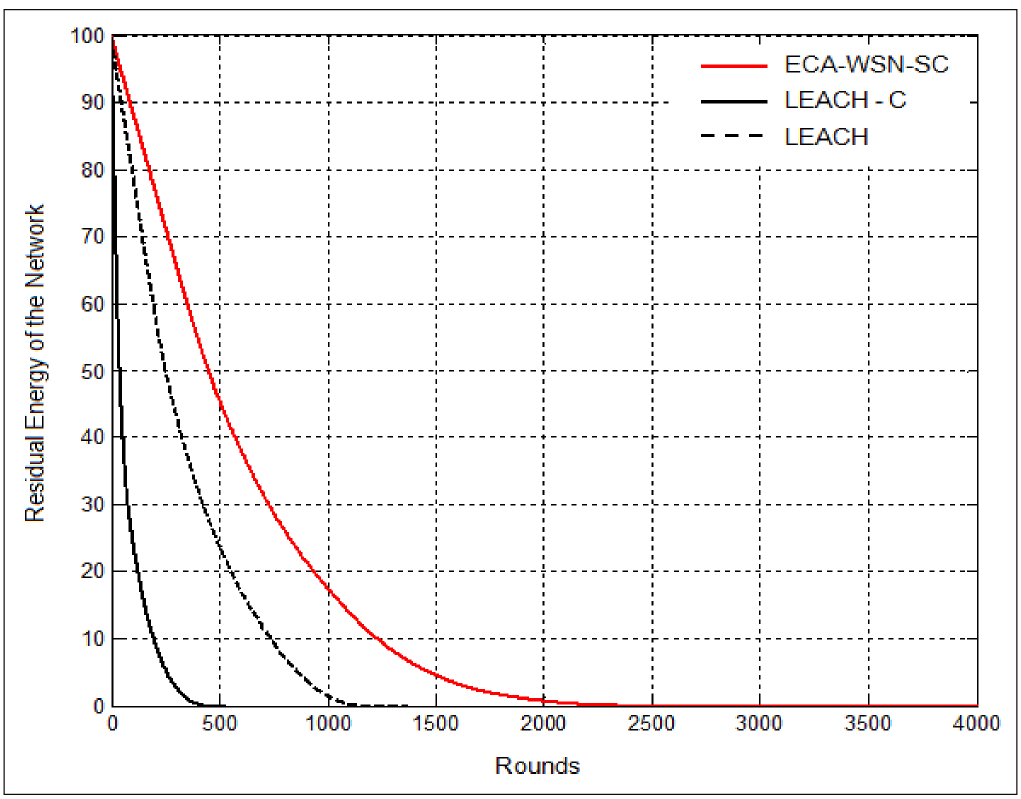

Figure 4. Evolution of the remaining energy in the network when the transmission rounds

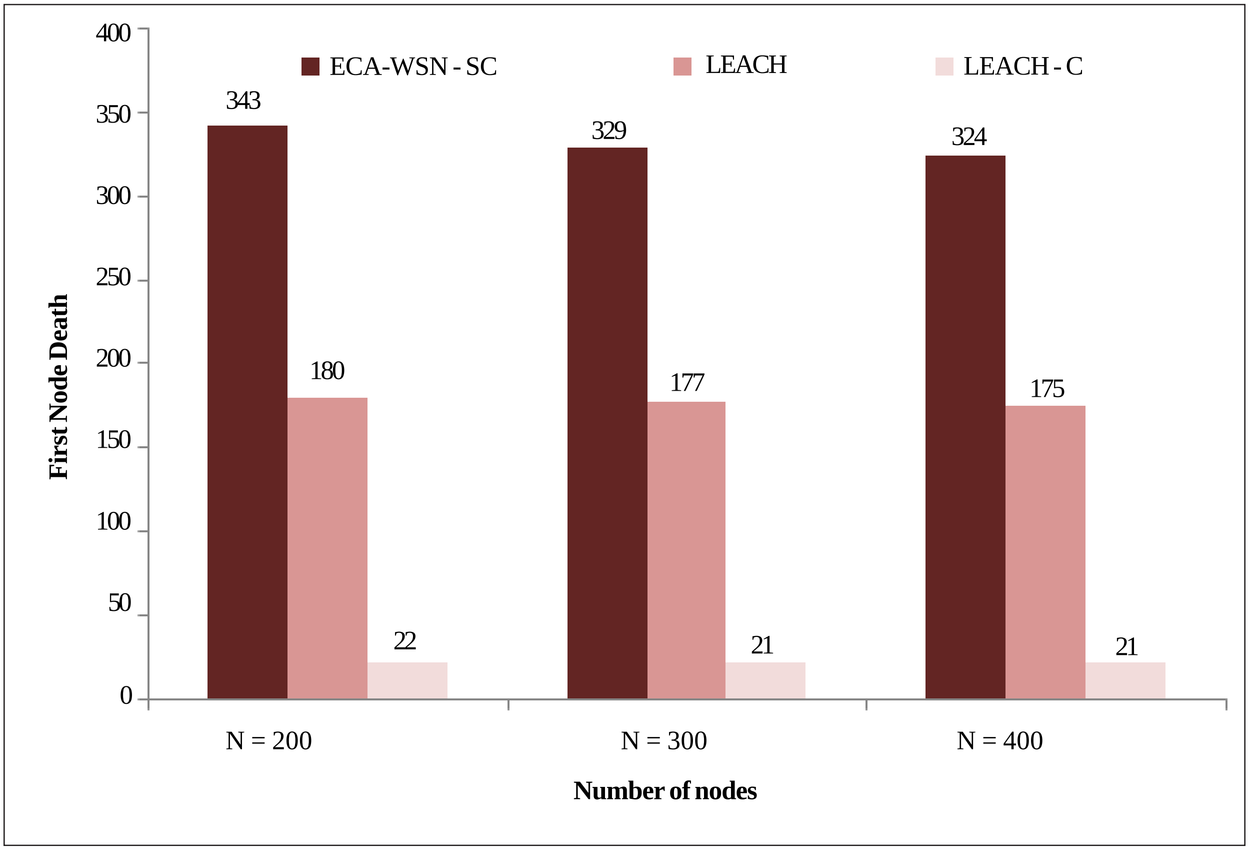

Figure 5. Impact of the node density N on the performances of the compared

We conclude that the ECA-WSN-SC algorithm gives a significant performance improvement in terms of energy and lifetime gains, compared to LEACH-C and LEACH protocols.

5. Conclusion

In this paper we have presented and detailed a new method of clustering based on spectral classification. The proposed approach called ECA-WSN-SC optimizes the energy consumption in a wireless sensor network. In particular, it uses the method of K-ways and a equitable technique to select C-Hs. In addition, we have measured and compared performances between our algorithm and two others. The experiment results have shown that the former approach presents a significant performance improvement in terms of energy and lifetime gains, compared to the others. Further works remain for studying other spectral classification techniques which may be more efficient in this kind of applications and determining other properties to improve the C-H election process.

References

[1] Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E. (2002). A survey on sensor networks, *Communications Magazine*, IEEE, 40 (8) 102–114.

[2] Al-Karaki, J., Kamal, A. (2004). Routing techniques in wireless sensor networks: a survey, *Wireless Communications, IEEE*, 11 (6) 6–28.

[3] Estrin, D., Girod, L., Pottie, G., Srivastava, M. (2001). Instrumenting the world with wireless sensor networks, in Acoustics, Speech, and Signal Processing. *In*: Proceedings. (ICASSP '01). IEEE International Conference on, 4, p. 2033–2036.

[4] Boyinbode, O., Le, H., Takizawa, M. (2011). A survey on clustering algorithms for wireless sensor networks, *International Journal of Space- Based and Situated Computing*, 1 (2) 130–136.

[5] Von Luxburg, U. (2007). A tutorial on spectral clustering, *Statistics and Computing*, 17 (4) 395–416.

[6] Heinzelman, W., Chandrakasan, A., Balakrishnan, H. (2000). Energyefficient communication protocol for wireless microsensor networks, in System Sciences. *In*: Proceedings of the 33rd Annual Hawaii International Conference on, 2. IEEE, p. 10.

[7] Heinzelman, W. B., Chandrakasan, A. P., Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks, IEEE Transactions on Wireless Communications, 1 (4) 660–670.

[8] Ye, M., Li, C., Chen, G., Wu, J. (2005). Eecs: an energy efficient clustering scheme in wireless sensor networks, in Performance, Computing, and Communications Conference. IPCCC 2005. 24th IEEE International. IEEE, p. 535–540.

[9] Yong, Z., Pei, Q. (2012). A energy-efficient clustering routing algorithm based on distance and residual energy for wireless sensor networks, Procedia Engineering, 29, p. 1882–1888.

[10] Jerusha, S., Kulothungan, K., Kannan, A. (2012). Location aware cluster based routing in wireless sensor networks, *International Journal of Computer Communication Technology*, 3 (5) 0975–7449.

[11] Rehena, Z., Roy, S., Mukherjee, N. (2011). Topology partitioning in wireless sensor networks using multiple sinks, in Computer and Information Technology (ICCIT), 14th International Conference, p. 251–256.

[12] Elbhiri, B., El Fkihi, S., Saadane, R., Aboutajdine, D. (2010). Clustering in wireless sensor networks based on near optimal bi-partitions, in Next Generation Internet (NGI), 6th EURO-NF Conference on. IEEE, p. 1–6.

[13] Ng, A., Jordan, M. (2002). Weiss et al. On spectral clustering: Analysis and an algorithm, *Advances in Neural Information Processing Systems*, 2, p. 849–856.