A New Clustering Algorithm in WSN Based on Spectral Clustering and Residual Energy

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

***Abstract*— Wireless Sensor Networks (WSNs) are composed of large number of sensor nodes that are randomly distributed in a region of interest. The nodes are responsible of the supervision of a physical phenomenon and periodic transmission of results to the sink. Energy saving results in extending the life of the network, which presents a great challenge of WSNs. To deal with this, a hierarchical clustering scheme, called K-Way Spectral Clustering Algorithm in Wireless Sensor Network (KSCA-WSN), is proposed in this paper. This algorithm is based on spectral classification; it also considers the residual energy, as well as some properties of the network nodes. Thus, our approach aims to seek for an ideal distribution of sensor nodes (clusters) and proposes new feautures to elect the appropriate cluster-heads. In term of extending the network lifetime and minimizing the energy consumption, the simulation results show an important improvement of the network performances with KSCA-WSN compared to other existing clustering methods.**

Keywords-*Clustering; Graph theory; Spectral classification; Energy consumption.*

I. INTRODUCTION

A Wireless Sensor Network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions at different locations, such as temperature, sound, and vibration. It contains a large number of nodes, which sense data from an inaccessible area and send their reports towards a processing center called Base Station (BS). WSNs have many advantages, such as the ease deployment and the capacity of self-organization. However, their main challenges are the limited resources in terms of radio communication capabilities and energies. This article addresses this last point.

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 distance transmissions should be minimized to extend the lifetime of the network. One way to reach this result consists in dividing the network into multiple clusters; each cluster has a cluster-head node (C-H) [2]. The C-H collects data from nodes of its same cluster, aggregates them and transmits them to the BS. Nevertheless, the main problem of many proposed clustering protocols 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 the clustering problem and to fairly localize C-Hs, we propose to consider methods of spectral clustering and the sensor's residual energies.

Spectral clustering has become one of the most popular modern clustering algorithms. It is based on graph theory, can be solved efficiently by standard linear algebra software. It also and very often outperforms traditional clustering algorithms such as the k-means algorithm [3]. It includes a variety of methods based on the notion of similarity matrix and use the eigenvectors differently. The k-ways approach is one of these methods. It consists to divide points into K disjoint classes. This method is based on the K eigenvectors related to K largest eigenvalues of Laplacian matrix [4]. Following this approach, we present a K-Way Spectral Clustering Algorithm in Wireless Sensor Network (KSCA-WSN), which has four main goals: (i) extending the network life-time by distributing energy consumption, (ii) running the clustering process within a constant number of iterations/steps, (iii) minimizing control overhead (to be linear in the number of nodes), and (iv) producing well-distributed cluster heads and compact clusters. KSCA-WSN does not make any assumptions about the distribution or density of nodes, neither about node capabilities.

The remaining of this paper is organized as follows: In Section II, we give a brief overview of some related research work. Section III describes the network model and states the addressed problem in this work. Details and properties of the proposed algorithm are given in Section IV while Section V presents the simulations and the results. Finally, conclusion and some perspectives are drawn in Section VI.

II. RELATED WORK

Many routing protocols have been designed based on clustering [5], where C-Hs are elected periodically and alternately. Moreover, the most commonly used approach for clustering is the LEACH algorithm [6]. LEACH is an energy-efficient communication protocol, which employs a hierarchical clustering. Besides, many clustering protocols based on the principle of this algorithm have been developed, such as LEACH-C [7], Energy-Kmeans [8], DECSA [9], EECS [10] and SECA [11].

 ISBN: 978-1-61208-296-7

In LEACH, nodes organize themselves into clusters using a distributed algorithm. Its main idea is to randomly and periodically select the C-Hs. 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 number of node and K is the desired number of clusters. After the election of C-Hs, each ordinary node in the WSN determines its cluster by choosing the C-H that requires the minimum communication energy, based on the received signal strength of the advertisement from each C-H. Each ordinary node will choose to join the C-H which has the highest signal quality. Once the clusters are formed, each C-H node creates Time division multiple access (TDMA) schedule. The C-H collects and aggregates information from sensors in its own cluster and passes on information to the BS. Although LEACH distributed clustering algorithm has advantages, using a central control algorithm to form clusters may produce better clusters by dispersing the cluster head nodes throughout the network [7]. In LEACH-centralized (LEACH-C), the BS determines the C-Hs by computing the average node energy. The node with energy below this average cannot be cluster heads for the current considered round. This ensures that the energy consumption will be effectively distributed among all the nodes.

Yong and Pei [9] presented a distance-energy cluster structure algorithm (DECSA) based on the classic clustering algorithm LEACH. This proposed approach considers both the distance between the nodes, the position of the BS and residual energy of nodes. Its main idea is to partition the network into three levels of hierarchy to reduce the energy consumption of C-Hs. this results 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 is far away from the BS.

Elbhiri et al. [12] propose a new approach called the Spectral Classification based on Near Optimal Clustering in Wireless Sensor Networks. This approach is based on spectral bisection for partitioning a sensor network into two clusters. This could be applied recursively to get the desired number of clusters. However, recursive spectral bisection finally produces 2^n clusters where n is the number of iterations. This method cannot partition the network into any desired number of clusters.

III. RADIO MODEL AND PROBLEM STATEMENT

This paper, discusses the periodical data gathering application, for which LEACH is proposed. Thus, the proposed clustering algorithms usually produce clusters of the same size. The C-H consumes the same amount of energy during intra-cluster communications. Moroever, cluster head election method based essentially on residual energy can obtain better energy efficiency than the method in which cluster heads are elected in turns or by probabilities. For this, we use the spectral classification that is widely used to solve graph partitioning problems. The spectral methods get their name from the spectral theories of linear algebra. This theories allows affirming the diagonalization of real symmetric matrices. It also justifies the decomposition of real symmetric matrices into eigenvalues in an orthonormal set of eigenvectors. Besides, the graph partitioning problem can be reduced to the resolution of a numerical system $Mx = \lambda x$. Solving this numerical system consists in seeking an orthogonal base of eigenvectors of the matrix M [4].

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 (let N be the number of nodes). This vector can be represented by an undirected graph $G(V, E)$; where V is the set of vertices (sensor nodes) identified by an index $i \in \{1, \cdots, N\}$ and E is 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 incident edges to node i is given by $d_{ii} = \sum_{j=1}^{j=N} a_{ij}$. 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 :

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

To simplify the network model, we make some assumptions that are:

- All nodes are homogeneous and have the same capacities.
- Each node is assigned a unique identifier (id) that includes the cluster identifier to which the node belongs.
- The network topology remains unchanged over time.
- The BS is placed far away from the network.

In addition, we use the model of the radio hardware energy dissipation presented in [6] and [7], and illustrated by Figure.1.

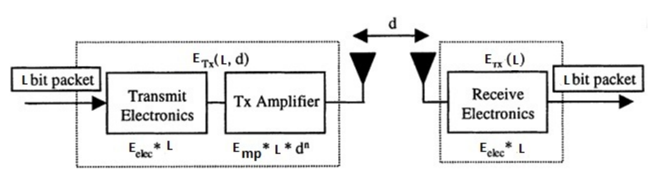

Fig. 1. Radio energy dissipation model.

Thus, to transmit an $L - bits$ message between two nodes separated by a distance d, the spent energy by the radio transmitter is given by the next equation:

$$E_{Tx}(L, d) = \begin{cases} L * E_{elec} + L * E_{fs} * d^2 & if\ d < d_0 \\ L * E_{elec} + L * E_{mp} * d^4 & if\ d \geq d_0 \end{cases} \quad (2)$$

and to receive this message, the spent energy by the radio receiver is:

$$E_{Rx}(L, d) = L * E_{elec} \quad (3)$$

where, E_{elec} is the energy consumption per bit in the transmitter and receiver circuitry, E_{fs} is the free space model's

 ISBN: 978-1-61208-296-7

amplifier energy consumption, and E_{ms} is the multiple attenuation model's amplifier energy consumption. The threshold's value (d_0) is determined by the equation (4):

$$d_0 = \sqrt{\frac{E_{fs}}{E_{ms}}} \quad (4)$$

In addition, once a sensor node runs out its energy, it is considered as died.

IV. KSCA-WSN DETAILS

In this section, we give details of the proposed KSCA-WSN algorithm. The latter consists in partitioning the WSN into a set of clusters based on spectral classification. Besides, it determines the CHs based on the residual energy of each node of the network. Some other properties will be explained in the subsections.

The spectral classification algorithm proposed here consists of three steps, as illustrated in the Figure 2.

A. The Pre-Processing step

First, each wireless sensor network node determines its position by using a Global Positioning System (GPS) or other tracking device, and transmits it to the BS 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 eq.(5).

$$A = [a_{ij}] = \begin{cases} exp\left(\frac{-1}{2\sigma^2} \; * d^2(i,j)\right) & if \; i \neq j \\ 0 & otherwise \end{cases} \quad (5)$$

with $d(i, j)$ is the euclidian distance between nodes i and j.

Thereafter, the BS must deduce the Degree Matrix and the Laplacian one. Then, it computes the eigenvalues and the eigenvectors of the last matrix.

B. Clustering step

The objectives of the current step are to define the optimal number of clusters and to form them.

As indicated in [7], the optimum number of clusters can be determined by the next equation:

$$K = \frac{\sqrt{N}}{\sqrt{2\pi}} \; * \sqrt{\frac{E_{fs}}{E_{ms}}} \; * \frac{M}{d_{toBS}^2} \quad (6)$$

where, N is the number of sensor nodes, M is the area of sensor nodes deployment, and d_{toBS} is the average distance of the nodes from the BS.

We construct a new matrix U from the K eigenvectors related to the K largest eigenvalues of the Laplacian matrix. In order to determine the K clusters of the WSN, we apply the classification algorithm k-means to the matrix U. We deal with each row of U as a point in $\Re^K$. We cluster the WSN into K clusters via k-means. 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.

We notice that in the proposed algorithm we determine the clusters before specifying the cluster heads and the optimal number of cluster partitions is as well defined automatically. So, our algorithm is different from the others (such as LEACH, LEACH-C, DECSA, etc.) that determine the C-Hs before the clusters. In addition, in order to define the number of cluster heads, the latter protocols run the same technique in each iteration and by the way consume more energy.

C. Cluster head election step

Once the clusters are determined, the next step of the KSCA-WSN consists to select the C-Hs. The C-Hs are responsible of the coordination among the nodes within their clusters, as well as for the aggregation of the received information and the communication with the BS. In our algorithm, the C-H is determined by taking into account the id of the node in the cluster and its residual energy. The id is defined without considering the position of the node in the cluster; thus, the C-H in each round of communication will be at a random position on the cluster. It is so important that nodes die at random locations of the network.

Indeed, for each round r of the simulation, we use the number $C_k = (r \; mod \, |S_k|)$ to elect a C-H for the appropriate cluster; $|S_k|$ is the total number of nodes in the cluster k. The node with $id = C_k$ and residual energy E_r greater than threshold E_{rmin} will be the C-H of the cluster k in the round r. E_{rmin} is the minimum residual energy required for a given node to be a C-H. It is the summation of the energy needed to receive and process data coming from the appropriate cluster nodes (data aggregation), and to transmit data towards the BS. This energy E_{rmin} is given by the next equation:

$$E_{rmin} = |S_k| \; * (E_{Rx}(L,d) + E_{Aggregation}) + E_{Tx}(L,d) \quad (7)$$

with

- $E_{Tx}(L, d)$is the energy consumed when the C-H transmits $L - bits$ data to the BS by a distance d.
- $E_{Aggregation}$ is the energy needed by the C-H to process data :
  $$E_{Aggregation} = L \; * E_{DA} \quad (8)$$
  E_{DA} is the energy for data aggregation per bit.
- $E_{Rx}(L, d)$is the energy consumed when the C-H receives data from one node.

Hence

$$E_{rmin} = L * \big((|S_k| + 1) \; * E_{elec} + |S_k| \; * E_{DA} + \epsilon_{mp} \; * d_i^4\big) \quad (9)$$

Nevertheless, if the residual energy E_r is less than this threshold E_{rmin} this node must broadcast a short message informing the node with $id = C_k + 1$to its residual energy and so on. Thus, the energy consumption will be distributed with more equitability between all nodes.

D. Data Transmission

Based on the id and the numbers of nodes in the cluster, a schedule TDMA will be created automatically to assign to each node a time when it can transmit its data to its own C-H. If we suppose that the node with the $id = i$ is elected

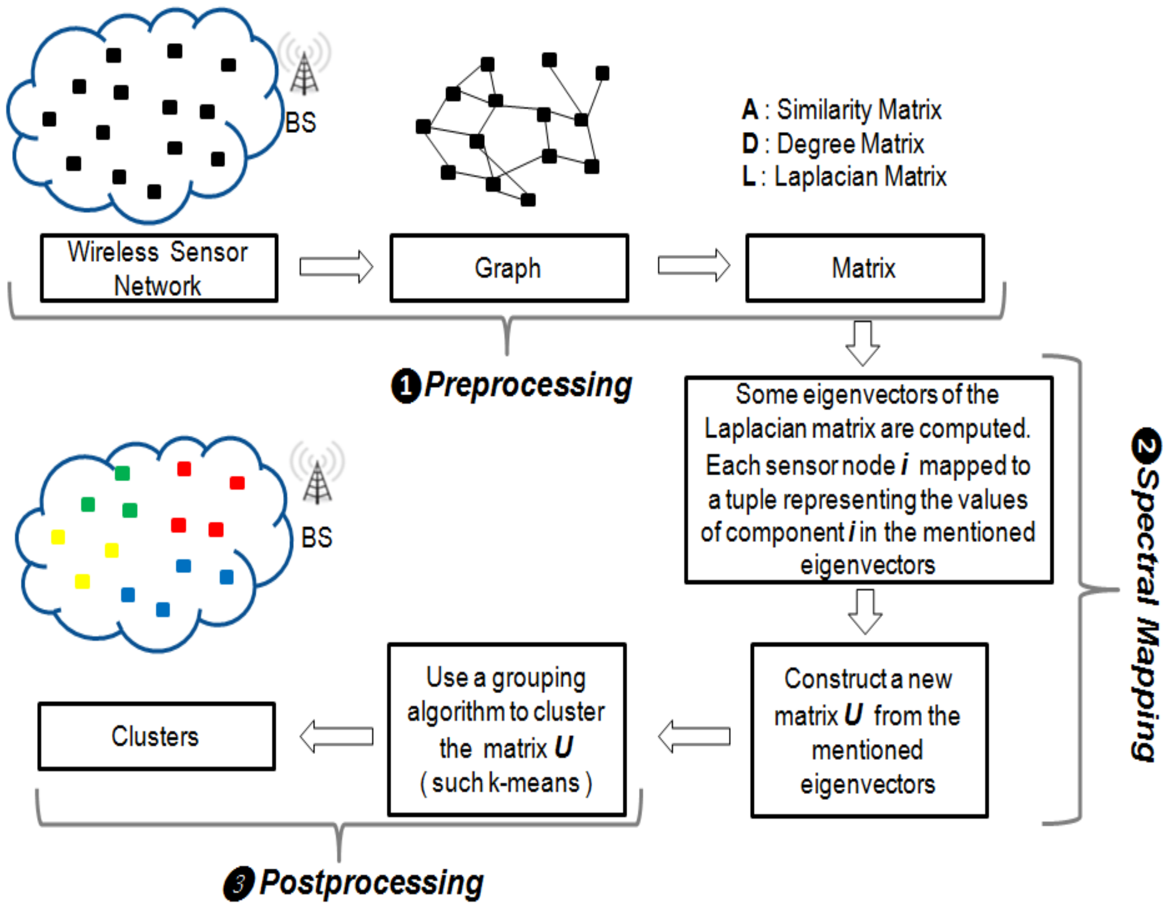

Fig. 2. Spectral clustering scheme of the KSCA-WSN approach.

as a C-H, the node with $id = ((i + 1 + |S_k|)\ mod\, |S_k|)$ 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. Hence, we avoid energy consumption due to synchronization of the cluster nodes when the C-H is elected to assign the TDMA. Indeed, this technique ensures not having collisions between data messages. It also allows the radio components of each non-C-H node to be disabled at any time, except during the time it is allocated for transmitting these data. this reduces the energy consumed by the nodes. However, the C-H must keep its receiver in order to receive data from other nodes. In addition, to save more energy in a wireless sensor network, we assume that if the distance between a node and the BS is lower than the distance between this node and its C-H. Then the node transmits its data directly to the BS. Moreover, each node can move in the standby mode to reduce energy consumption.

V. Simulations and Results

In this section, we present the results of the evaluation experiments of the proposed KSCA-WSN algorithm. We used the MATLAB software to simulate and analyze this algorithm. Besides, we consider the radio model presented in Figure 1 and the different parameters used in our simulations are shown in TABLE I.

TABLE I
Experimental Simulation parameters.

Parameter	value
E_{elec}	$50nJ/bit$
E_{fs}	$10pJ/bit/m^2$
E_{mp}	$0.0013pJ/bit/m^4$
Initial energy E_0	$0.5J$
E_{DA}	$5nJ/bit/message$
Area of Network	$200m * 200m$
Sink coordination	$(100m, 250m)$
d_0	$88m$
Message Size	$4000bytes$
Number of Nodes	500

A node that runs 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 aggregates data according to the model radio parameters. Also, we ignore the effect caused by the signal collision and the interference in the wireless channel. Furthermore, each simulation result

 ISBN: 978-1-61208-296-7

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 3, an example of wireless sensor networks with $N = 500$ nodes randomly distributed in a $200m*200m$ area. The BS is located far away from the sensing area $(x_{SB} = 100m\ ;\ y_{SB} = 250m)$.

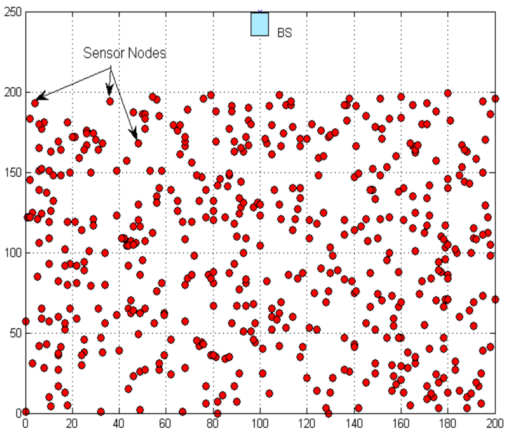

Fig. 3. Distribution of wireless sensor network.

In order to evaluate the performances of the new proposed protocol, we propose to compare it to the Spectral Classification based on Near Optimal Clustering (SCNOC) [12], the DECSA [9] and the LEACH-C [7] algorithms. Also, we use two metrics to analyze and compare the performances of the protocols:

- Network lifetime: It 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 a 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.

As illustrated in Figure 4, the main problem of LEACH-C and DECSA protocols are the random selection of the CHs. Indeed, all CHs can be located in a small region of the network. Hence, some ordinary nodes would be out of reach. Also, SCNOC produces 2^n clusters where n is the number of iterations; it cannot partition the network into any desired number of clusters.

The Figure 5 presents an example of clustering for the KSCA-WSN algorithm. We note that the network is subdivided into six clusters, and the nodes are correctly distributed over the sensing area. Also, there is no intersection between the different clusters.

Figure 6 shows a significant improvement for the KSCA-WSN approach in terms of numbers of periods relating to

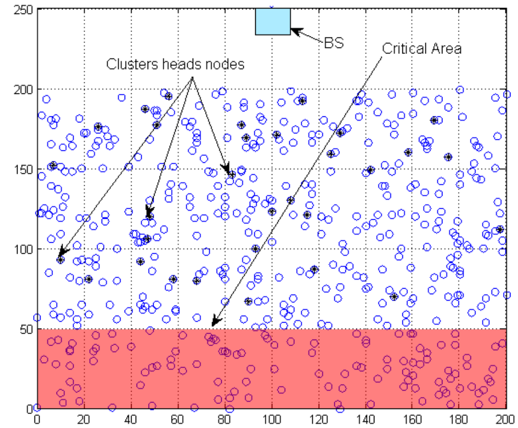

Fig. 4. Illustration of the critical area in the LEACH-C clustering algorithm with N=500 nodes

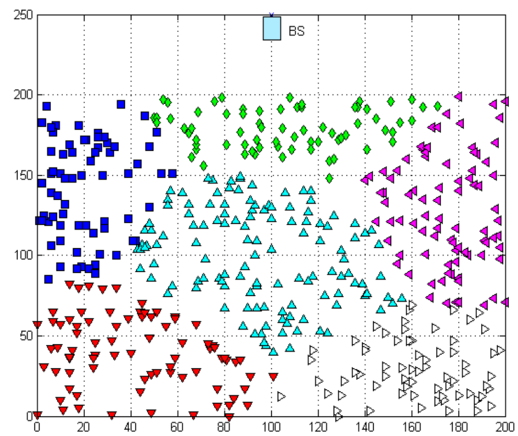

Fig. 5. Clustering results of the KSCA-WSN algorithm with N=500 nodes.

FND. In the proposed method, FND occurs at the period 520 while it occurs at the 378 period for spectral classification based on Near Optimal Clustering, at 21 periods for LEACH-C, and at 74 for DECSA protocols.

Figure 7 gives the total network remaining energy in every transmission round. The network remaining energy decreases more rapidly in the spectral classification based on Near Optimal Clustering, the LEACH-C, and the DECSA protocols than in the KSCA-WSN algorithm.

The Figure 8 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 8, for different values of N equal to 100, 200, 300, and 400, our algorithm presents an improvement of performances compared to the other algorithms. It follows that even if the node density increase the new proposed approach still gives best results compared to the other ones.

Figure 9 shows the performances of the different compared

 ISBN: 978-1-61208-296-7

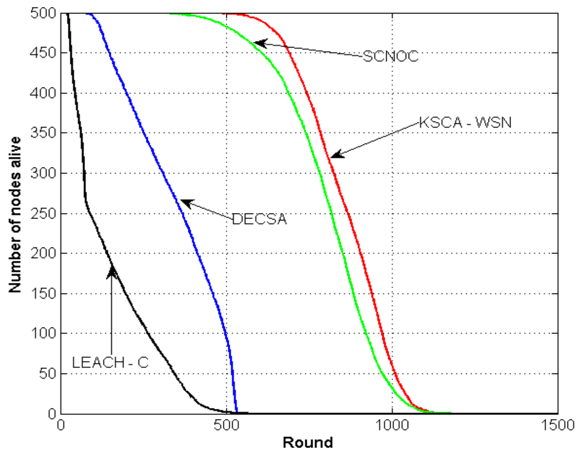

Fig. 6. Number of nodes alive over time of the compared protocols

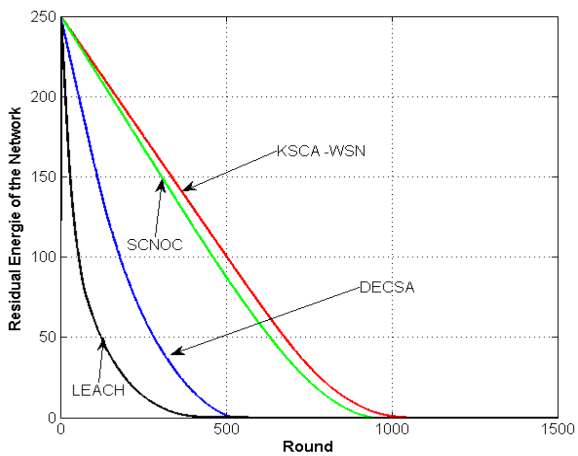

Fig. 7. Evolution of the remaining energy in the network when the transmission rounds.

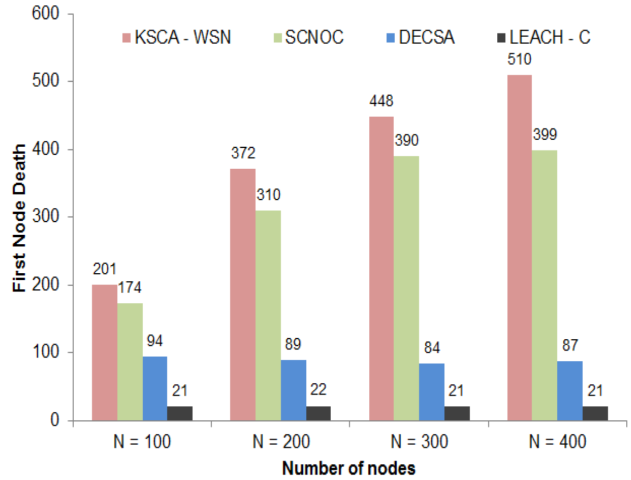

Fig. 8. Impact of the node density N on the performances of the compared algorithms

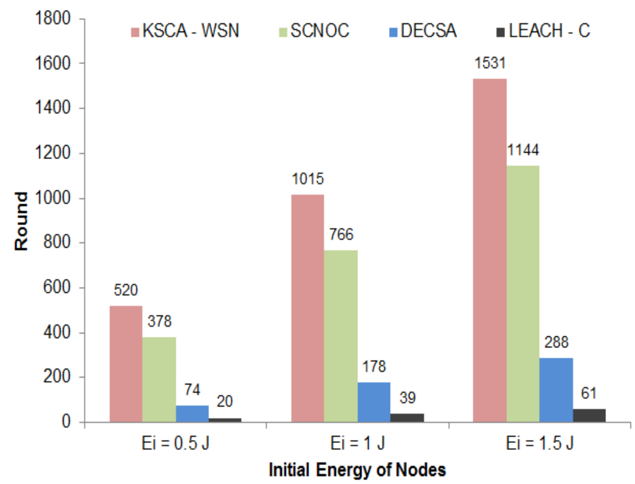

Fig. 9. Impact of the initial energy quantity on the performances of the four compared algorithms

protocols by using different initial energies. It gives the FND round depending on the quantity of the node initial energy. Once more, it is shown that for different values of the energy, the new proposed approach presents a significant improvement compared to the others.

We conclude that the KSCA - WSN algorithm gives a significant performance improvement; in terms of energy and lifetime gains, compared to the Spectral Classification based on Near Optimal Clustering, the LEACH-C, and the DECSA protocols. The best results of the KSCA-WSN approach can be explained by the three points: (i) the proposal starts by selecting the clusters (similar nodes) before the election of the C-Hs. (ii) the approach considers the WSN nodes residual energies when electing C-Hs. And (iii) a C-H is elected at any position of the wireless sensor network.

VI. Conclusion and Future Work

Energy saving is an important challenging issue in a wireless sensor network. To ensure more effective energy distribution and to extend the life time of a network, new and efficient energy saving scheme is developed in this paper. We have proposed a new method of clustering, based on spectral classification. The latter has been widely used to solve graph partitioning problem. In particular, we use the method of k-ways. Thus, we start by determining similar nodes (clusters) before selecting C-Hs. Furthermore, we take into account the node residual energy when selecting the appropriate C-H nodes. Indeed, simulation results show that our approach ensures the low energy consumption and improves the network lifetime. Promising results are obtained by several simulations, and exceed those obtained by others algorithms.
Further works need to determine other properties to improve both the clustering and the cluster-head election processes.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," Communications magazine, IEEE, vol. 40, no. 8,

2002, pp. 102−114.

[2] J. Al-Karaki and A. Kamal, ”Routing techniques in wireless sensor networks: A survey,” Wireless Communications, IEEE, vol. 11, no. 6,2004, pp. 6−28.

[3] U. Von Luxburg, ”A tutorial on spectral clustering,” Statistics and computing, vol. 17, no. 4, 2007, pp. 395−416.

[4] A. Ng, M. Jordan, Y. Weiss, ”On spectral clustering: Analysis and an algorithm,” Advances in neural information processing systems, vol. 2, 2002, pp. 849−856.

[5] O. Boyinbode, H. Le, and M. Takizawa, ”A survey on clustering algorithms for wireless sensor networks,” International Journal of SpaceBased and Situated Computing, vol. 1, no. 2, 2011, pp. 130−136.

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

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, ”An application-specific protocol architecture for wireless microsensor networks,” IEEE Transactions on wireless communications, vol. 1, no. 4, 2002, pp. 660−670.

[8] S. Jerusha, K. Kulothungan, and A. Kannan, ”Location aware cluster based routing in wireless sensor networks,” International Journal of Computer and Communication Technology, vol. 3, no. 5, 2012, pp. 0975−7449.

[9] Z. Yong and Q. Pei, ”A energy-efficient clustering routing algorithm based on distance and residual energy for wireless sensor networks,” Procedia Engineering, vol. 29, 2012, pp. 1882−1888.

[10] M. Ye, C. Li, G. Chen, and J. Wu, ”Eecs: an energy efficient clustering scheme in wireless sensor networks,” in Performance, Computing, and Communications Conference, IPCCC 2005, IEEE, 2005, pp. 535−540.

[11] J.-Y. Chang and P.-H. Ju, ”An efficient cluster-based power saving scheme for wireless sensor networks,” EURASIP Journal on Wireless Communications and Networking, vol. 2012, no. 1, 2012, pp. 172.

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