A combination of K-means Algorithm and Optimal Path Selection Method for Lifetime Extension in Wireless Sensor Networks

Wadii Jlassi[1], Rim Haddad[2], Ridha Bouallegue[3], Raed Shubair[4]

[1] Innov'COM Lab/Sup'Com, National Engineering School of Tunis, University Tunis El Manar, Tunis, Tunisia
Wadii.jlassi@gmail.com
[2] Laval University, Quebec, Canada
Rim.haddad@eti.ulaval.ca
[3] Innov'COM Lab/Sup'Com, University of Carthage, Tunis, Tunisia
Ridha.bouallegue@supcom.tn
[4] New York University, Abu Dhabi, UAE
raed.shubair@nyu.edu

Abstract. Wireless Sensor Networks (WSN) are special types of wireless networks where hundreds or thousands of sensor nodes are working together. Since the lifetime of each sensor is equivalent to a battery, the energy issue is considered a major challenge. Clustering has been proposed as a strategy to extend the lifetime of wireless sensor networks. Many clustering algorithms consider the residual energy and distance between the nodes in the selection of cluster heads and others rotate the selection of cluster heads periodically. We propose in this article a CH selection followed by making clusters using the K-means algorithm and we present the PRIM algorithm to transmit the packets in multi-hop transmission between CHs and BS and choose the optimal path. The clustering scheme allows to decrease intra-cluster communications and to gain energy efficiency for sensor nodes. Computer simulation results show that our method aims to extend the lifetime of the wireless sensor network efficiently compared to other existing methods.

Keywords. Wireless Sensor Networks, Clustering, K-means, optimal path.

1 Introduction

Wireless Sensor Network (WSN) consists of a large number of sensor nodes deployed in a hostile environment and self-organized to gather and transmit data from the field to a Base Station (BS).

In a WSN, nodes are grouped into partitions called clusters. Various WSNs are employing the cluster structure, which efficiently allocates the resource and energy and thereby maximizes the network lifetime [1]. Clustering algorithms are energy-efficient methods to collect data to BS [2,3]. The network is divided into clusters. In every cluster, a sensor node is selected as a cluster head. The other members transmit their collected information to the cluster head in a single-hope or multi-hope manner. After gathering information, which are distance separating the nodes and energy for cluster head selection and the optimal path between CH and BS for packets transmission. The cluster head transmits to the base station through the single-stage or multi-stage line [4].

The selection of CH is one of the most critical tasks in the management of WSNs since CHs consume much more energy than other nodes in the network [5]. To elect the cluster heads some protocols use a random number to balance the energy consumption of the sensor nodes over the networks [6]. One of the most recognized protocols in this regard is LEACH [7], it can reduce energy overconsumption by grouping the sensor nodes into clusters to decrease the number of transmitted packets and restrict the direct single-hop communication between the nodes and the base station. Leach-C [8] proposes a centralized clustering algorithm. The K-means algorithm [9] is used to form the clusters such that the distance between the nodes and the CH becomes minimal. The CHs transmit their data to the base station in one-hope. The Hierarchical Agglomerative Clustering (HAC) [10] proposed to group nodes into clusters, then elect CH on a formula that takes into account the position of the nodes and their remaining energy for the CHs packets transmission to the BS. HACMH proposed multi-hop transmission between CHs and BS by applying spanning tree algorithm for choosing the optimal path.

These protocols aim to extend the network lifetime, decrease the number of transmitted packets and reduce direct communication between CHs and BS. This paper introduces the energy usage of sensor networks and how to extend the network lifetime of the sensor nodes by using important criteria.

As an expansion of some previous works in this area, the contributions of this work could be summarized as follows:

- We select CH and optimal path to manage wireless sensor networks efficiently in energy.
- We use K-means algorithm to elect the center of each cluster as a data-gathering point. This algorithm allows to decrease intra-cluster communications.
- We propose a PRIM algorithm for multi-hop transmission between CHs and CH and BS by applying a spanning tree algorithm (PRIM) for choosing the optimal path to avoid the direct transmission between CHs and BS and minimize data transmission time between CHs.

The paper is organized as follows. Section 2 presents the related research, and Section 3 introduces the proposed scheme. The simulation results and discussions are presented in Section 4. Finally, Section 5 concludes the paper and suggests future work.

2 Related works

Wireless sensor nodes perform the operation of transmitting the data from the source to the destination which should be made in an efficient way so that the data transmission between the sender and the receiver will be in an effective manner. The sensor nodes have limitations in storage, power, latency, constraint bandwidth, and reduced corporal size [11].

In this section, we briefly review some recent researches proposed for clustering in WSNs that are relevant to our approach. Leach protocol [7] is a typical routing scheme based on clustering. The algorithm consists of electing the CHs to distribute the energy consumption evenly among the nodes. The CHs aggregate the data collected from the member nodes and transmit it to the BS. The protocol is based on a probabilistic model that rotates CHs selection in order to balance the energy consumption of nodes in the network. However, Because of the probability model, the Leach protocol has some drawbacks. CHs could be situated very close to each other or in a place where the number of nodes is small. In this case, the protocol doesn't consider the remaining energy of each node. To avoid these drawbacks, the LEACH-C protocol employs a centralized control approach using the location information of the sensor nodes [12]. During the set-up phase, each member node transmits the current location information and energy level to the BS. Then the BS calculates the average energy of the member nodes in the current round and divides the network into a number of clusters. After the selection of a CH in each cluster, the member nodes send the data to the closest CH. Finally, the set-up phase is completed after deciding the routing path. The energy consumption of CH can be reduced by the efficient selection of the CH using the location information of the member nodes. Although, the requirement of getting exact location information of all sensor nodes causes additional energy consumption and the routing path is not always shorter than LEACH protocol.

The protocol HEED [13] chooses CH by referring to the residual energy of each member node. It also considers the inter-cluster communication cost as a secondary clustering parameter. This protocol extends the network lifetime and creates well-distributed clusters. But, the random selection of the cluster heads may cause higher communication between a cluster head and a base station. The ER-HEED [13] (Energy-based Rotated HEED) is considered an improvement of the HEED protocol with the introduction of rotation in equal-sized clusters. This protocol consists of three phases as follows:

- The cluster formation and cluster head selection are performed according to the HEED protocol.
- The CH election is the member of the cluster with the highest energy without the need to perform an election protocol.
- If any cluster member dies, re-clustering is performed by repeating cluster formation and cluster head election step.

The HACSH [11] uses hierarchical agglomerative clustering (HAC) for the formation of clusters from the nodes based on the Euclidean distance (nearest neighbor) between nodes. The HAC algorithm is executed to form k clusters from n nodes with K <= n. After running the HAC algorithm, the distance between node i and the base station (BS) is calculated in Equation (1):

$$d_i = \sqrt{(x_i - x_{sb})^2 - (y_i - y_{sb})^2)} \quad (1)$$

After that, we calculate the distance between node i and the base station (BS) in the cluster and the centroid, the nearest node will be designated as CH. Now, each node is associated with one of the k clusters, we calculate the opti point for each cluster which is given by the following formula opti (xopti; yopti) (2,3) with:

$$X_{opti} = \frac{1}{\sum_{i=1}^{s} E_i} * \sum_{j=1}^{s} E_j x_j \quad (2)$$

$$Y_{opti} = \frac{1}{\sum_{i=1}^{s} E_i} * \sum_{j=1}^{s} E_j x_j \quad (3)$$

Finally, we choose the nearest node of the opti mark (xopti; yopti) as the new CH, always based on the Euclidean distance between the nodes of each cluster and opti point is illustrated in Equation (4):

$$d_i = \sqrt{(x_i - x_{opti})^2 + (y_i - y_{opti})^2} \quad (4)$$

All nodes of the cluster transmit packets to CH, which will handle to transmit them to the base station (BS).

3 Proposed work

In this section we propose a data gathering point selection method which employs the K-means algorithm for CHs selection. To transmit the packets from CHs to BS. We propose a minimum spanning tree.

3.1 Network Model:

The sensor nodes are randomly distributed in the target area and have the same amount of energy when they are initially deployed. Each node has an ID number and is placed stationary after the deployment. The Base Station can be placed anywhere in the area. Sensor nodes can send the data to it and The CHs are aware of their remaining energy.

3.2 Energy Consumption Model:

To evaluate the performance of the proposed method, we will compare its energy model to LEACH.As defined in this model, there are two cases: the free space(fs) and the multi-path (MP). When the distance between a sending node and a receiving node is less than threshold value d0, the free space model, d^2 powerless, is used. Otherwise, multipath
model, d^4 powerless, is used. From the two models above, the energy consumption for transmitting a k-bit packet over a distance d is given by Equation (5):

$$E_{Tx}(K,d) = E_{elec} * K + \mathcal{E}_{fs} * k * d^2, d < d_0$$
$$E_{Tx}(K,d) = E_{elec} * K + \mathcal{E}_{mp} * k * d^4, d \geq d_0 \quad (5)$$

Where E_{elec} is the energy required for processing 1-bit data with the electronic circuits. εfs and εmp is the energy taken for transmitting 1-bit data to achieve an acceptable bit error rate in the case of the free space model and multipath model, respectively. They are dependent on the distance of transmission. Note that energy dissipation of free space and multipath is proportional to d^2 and d^4, respectively. The threshold, d_0, can be obtained by Equation (6):

$$d_0 = \sqrt{\frac{\mathcal{E}_{fs}}{\mathcal{E}_{mp}}} \quad (6)$$

The energy taken to receive a k-bit message is calculated in Equation (7):

$$E_{Rx=E_{elec}} * K \quad (7)$$

3.3 Proposed Scheme:

After deployment of sensor nodes in the sensed area, we divide the area into clusters. The clustering decreases redundant data, reduce the number of inter-node communication by localizing data transmission within the clusters and decrease the overall amount of transmission to the base station.

In this section, we will present our approach. The proposed scheme uses K-means algorithm for clusters creation from the nodes based on the Euclidean distance between them.

The algorithm accepts two inputs: S= {s1, s2, .sn} sensor nodes with location information and k number of clusters. The output is C= {C1, C2…Ck} set of clusters with input data partitioned among them.

K-means consists of three steps as follows.

- Step 1: Initial clustering

K-means algorithm is executed clusters creation in wireless sensor networks. First, k out of n nodes are randomly selected as the CHs. Each of the remaining nodes decides its nearest CH according to Euclidean distance

- Step 2: Re-clustering

After that, each node in the network is assigned to a cluster. The centroid of each cluster is illustrated in Equation (8).

$$Centroid(X,Y) = (\frac{1}{s}\sum_{i=1}^{s} x_i \, , \frac{1}{s}\sum_{i=1}^{s} y_i) \quad (8)$$

Note that the centroid of a cluster is a virtual node located at the center position of the cluster. In this step, the center of each cluster is updated and the new center will be the average location point of all the sensor nodes locations in the cluster. For each node, recalculate the distance between the node and all cluster centers using Euclidean distance and allocate the closest one with the new center (CH) in each cluster, Step 2 is recursively executed until no point switches clusters and the CH is not changed anymore.

- Step 3: Choosing the CH

As soon as clusters are formed, an ID number is assigned to each node of a cluster according to the distance from the centroid, assigning a smaller number to the closer one. The ID number of a node indicates its order to be chosen as a CH. Therefore, the ID number plays an important role in the selection of a node as CH. The residual energy of the CH is checked every round to retain the connectivity of the network. If the energy of the CH is smaller than the preset threshold, the node in the next order is selected as a new CH. The newly elected CH informs the other nodes of the change of the CH. After clustering and CHs selection, in each round, the proposed scheme adopts multi-hop transmission between CHs and BS by applying a spanning tree algorithm (PRIM) for choosing the optimal path.

3.4 PRIM algorithm:

The algorithm starts with a tree consisting of a single vertex, and continuously increases its size one edge at a time. It halts when all the vertices have been reached.

- Input: A non-empty connected weighted graph with vertices V and edges E (the weights can be negative).
- Initialize: $V_{new} = \{x\}$, where x is an arbitrary node (starting point) from V, $E_{new} = \{\}$ and Repeat until $V_{new} = V$. Then, choose an edge {u, v} with minimal weight such that u is in V_{new} and v is not (if there are multiple edges with the same weight, any of them may be picked). Finally, add v to Vnew, and {u, v} to E_{new}.
- Output: V_{new} and E_{new} describe a minimal spanning Tree.

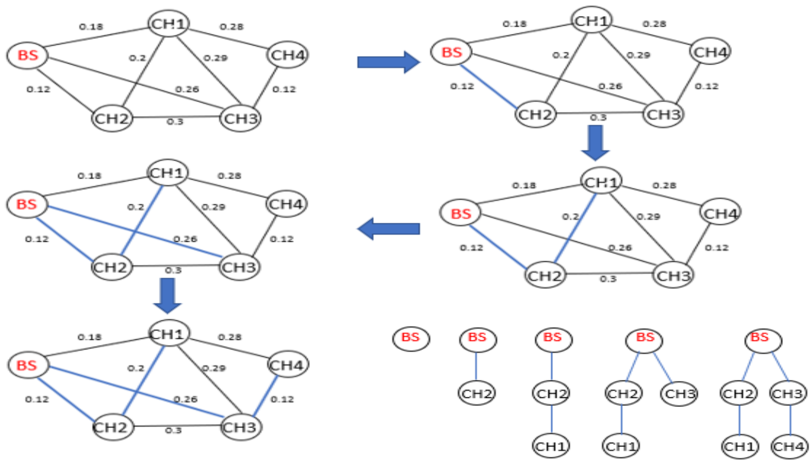

Fig. 1. Steps of the algorithm

In this example, we present and suppose 4 clusters with one BS in figure1. We calculated the distance between all CHs and BS. Initially, the root of the tree is always BS. The weight between nodes in the graph is represented by the distance between CHs, CH, and BS. After building the minimum spanning tree using the PRIM algorithm, each CH sent packets to the CH or directly to the BS.

4 Simulations Results

In the previous section, we analyzed and evaluated the simulation results with different numbers of clusters. Then, we compared the results with other algorithms namely LEACH-C, ER-HEED, HACMH, and k-means under different scenarios to validate the network performance. The proposed method has been implemented in MATLAB R2015a. Here, we consider a network of 200 sensor nodes randomly distributed in a field of 200m * 200m, and the values used in the first model are described in Table 1.

Table 1. Simulation Parameters

Parameter	Value
Size of the network	200 m*200 m
N (Number of deployed nodes)	200
E_0 (initial energy of nodes)	1 J
Eelec	50 nJ/bit
Ecpu	7 nJ/bit
Ɛfs	10 pj/bit/m2
Ɛmp	0.0013 pj/bit/m4
Position of base station	0 m*0 m
Packet size	4000 bits

The proposed method is analyzed with the different number of clusters in the network. In Fig.2, First, when k=3 clusters, the network lifetime is estimated as 1020 rounds. Then, when the number of clusters increases to 10 clusters the network lifetime increases to 1376 rounds. As the cluster number increases, the size of the cluster decreases. So, the energy consumption decreases when the network is divided from 3 clusters to 10 clusters, and the simulation results show the same performance when the number of clusters is 15, 20, and 30.

Next, the proposed scheme is compared with related works in terms of network lifetime and remaining energy. Network lifetime is defined as the number of rounds when all the nodes run out of energy. In Table 2, We compare the behavior of the network in terms of the First Node Dies FND, the Half of the Nodes Alive HNA and the Last Node Dies LN. The configuration parameters for the simulation are presented in Table 1 with the number of clusters is k=15.

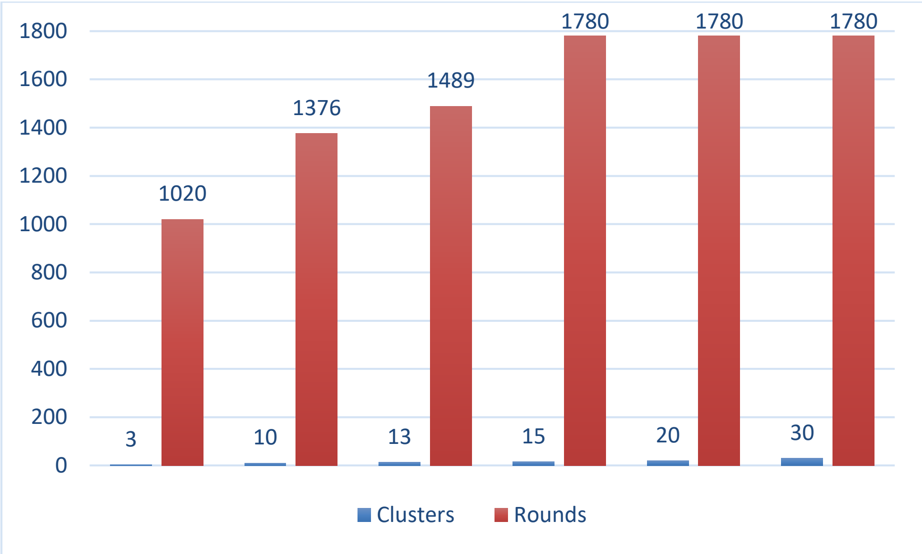

Fig. 2. Number of clusters and rounds

The number of alive nodes is checked to assess the lifetime of the network. In figure 3, we note that the proposed scheme has the highest alive node from the beginning to the total depletion of nodes. After 1500 rounds, almost no alive node is left with Leach-C, ER-HEED, HACMH, and k-means but about 25 nodes are still alive with the proposed scheme. This is since communication overhead is lower than that with compared protocols. Note that the proposed scheme forms the clusters so that the distance between the CH and the member's nodes are minimized also the communication between CHs and CH and BS is optimal.

Table 2. Values of FND, HNA and LND metrics for each algorithm

Algorithm	FND	HNA	LND
LEACH-C	149	488	1032
ER-HEED	183	599	1120
HACMH	191	655	1185
K-MEANS	300	742	1495
PROPOSED WORK	500	1050	1780

Figure 4 shows the remaining energy of the network as the round proceeds. The proposed scheme is always better than Leach-C, ER-HEED, K-means, and HACMH. We observe in Figure 4 that little energy is left after 1050 rounds with Leach-C and ER-HEED and after 1200 rounds with HACM and K-means. However, the network still has some residual energy until 1780 rounds with the proposed scheme.

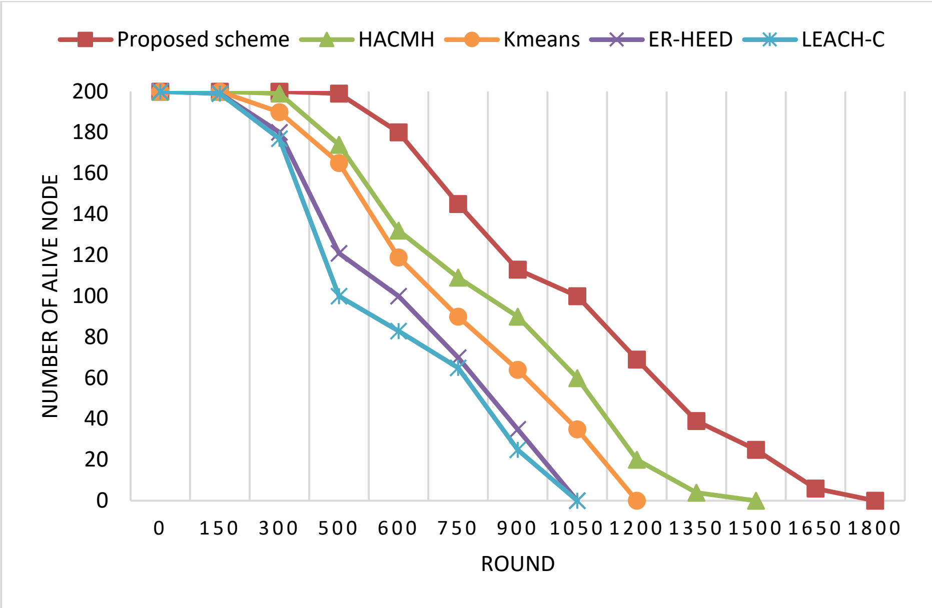

Fig. 3. Distribution of alive sensor nodes according to the number of rounds for each algorithm

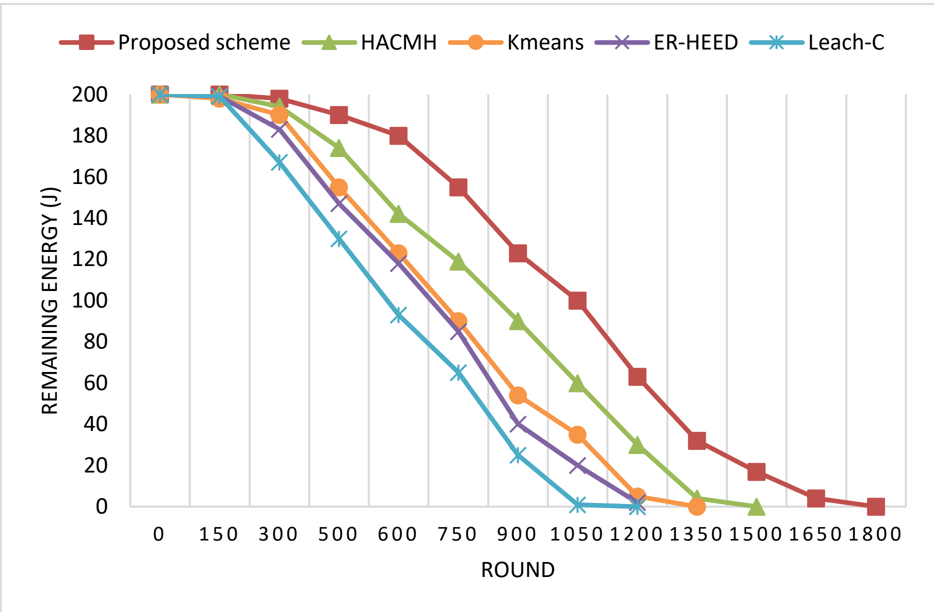

Fig. 4. Residual energy of the network as the round proceeds

5 Conclusion

One of the most challenging issues in Wireless Sensors Networks (WSN) is severe energy restrictions. In this study, we use the center of each cluster as a data gathering point. K-means algorithm is used to find the optimal data gathering point and make clusters. It allows to decrease intra-cluster communications and to increase energy efficiency for sensor nodes. To choose the optimal path, we proposed the PRIM algorithm. The simulation results show that the proposed scheme provides better energy efficiency and a longer network lifetime than the other exits algorithms.

Acknowledgments. We are grateful for the support of the Department of Electrical and Computer Engineering at the New York University of Abu Dhabi (NYU).

References

1. Q. Zytoune, M. El. Aroussi, M. Rziza, D. Aboutajdine: StochasticLow Energy Adaptive Clustering Hierarchy, (2008), 47-51.
2. Ma, J., Shi, S., Gu, X., & Wang, F.: Heuristic mobile data gathering for wireless sensor networks via trajectory control. International Journal of Distributed Sensor Networks, (2020).
3. Tashtarian, F., Moghaddam, M. H. Y., Sohraby, K., & Efati, S.: On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks. IEEE Transactions on Vehicular Technology, (2015), 3177–3189.
4. Gajjar, S., Sarkar, M., & Dasgupta, K. FAMACROW: Fuzzy and ant colony optimization based combined mac, routing, and unequal clustering cross-layer protocol for wireless sensor net-works. Applied Soft Computing (2016).
5. M.C.M. Thein and T. Thein: An energy efficient cluster-head selection for wireless sensor networks, ISMS, (2010), 287-291.
6. Sert, S. A., Bagci, H., & Yazici, A.: MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks. Applied Soft Computing, (2015), 151-165.
7. Ghosh, N., Banerjee, I., & Sherratt, R. S: On-demand fuzzy clustering and ant-colony optimization based mobile data collection in wireless sensor network. Wireless Networks, (2019), 1829–1845.
8. F. Xiangning, S. Yulin: Improvement on LEACH Protocol of Wireless Sensor Network, International Conference on Sensor Technologies and Applications, (2007) 260-264.
9. Raval, G., Bhavsar, M., & Patel, N.: Enhancing data delivery with density-controlled clustering in wireless sensor networks. Microsystem Technologies, (2016), 1-19.
10. Abidoye, A. P., & Kabaso, B.: Energy-efficient hierarchical routing in wireless sensor networks based on Fog Computing (2020).
11. birami, K. Shanthi, Sybil attack in Wireless Sensor Network, (2013).
12. F. Xiangning, S. Yulin: Improvement on LEACH Protocol of Wireless Sensor Network, International Conference on Sensor Technologies and Applications, (2007), 260-264.
13. Zaib Ullah, Leonardo Mostarda: A comparison of HEED based clustering algorithms - introducing ER-HEED: International Conference on Advanced Information Networking and Applications, (2016) 339-345.