A Conditional Distributed K-means Cluster-based Routing Scheme for Wireless Sensor Networks

Bouchra Meriem BENMAHDI
STIC Laboratory, Tlemcen University
13000 Tlemcen, Algeria
benmahdibouchra@gmail.com

Mohamed LEHSAINI
STIC Laboratory, Tlemcen University
13000 Tlemcen, Algeria
m_lehsaini@mail.univ-tlemcen.dz

***Abstract*—The main issue in Wireless Sensor Networks (WSNs) is the energy optimization. Clustering routing approach is a promising solution and so much used, which can prolong the network lifetime. In this paper, we aim to improve the dynamic clustering. The proposed solution uses Silhouette method for determining the optimal number of clusters and K-Means approach for the process of forming clusters. The process of determining the number of clusters is triggered when the percentage of alive nodes decreases under a fixed threshold (5%). Simulation results show that our proposal significantly improves network lifetime compared with Distributed_K_Means routing scheme and LEACH protocol especially when the number of nodes increases.**

***Index Terms*—K-Means method; Silhouette method; Dynamic clustering, Clustering;**

I. Introduction

Research in WSNs is more and more active due to their use in many fields such as military field, habitat monitoring, environment conditions observation, agriculture and smart home environment. A wireless sensor network consists of a huge number of sensor nodes able to sense the environment, process, store and transmit the sensed data through wireless channels towards one sink or many sinks [1]. Sensor nodes are generally equipped with low power non-rechargeable battery and their replacement is almost impossible in a harsh environment. Thereby, energy efficiency has become a major issue in WSNs due to the limitation capacity of the battery. In this context, several energy-efficient routing approaches have been proposed in the literature to deal with the limited battery life of sensor nodes in order to increase the network lifetime. Among these approaches clustering is considered as a promising approach to reduce energy consumption and hence increase network lifetime [2].

In clustering process, sensor nodes are grouped into clusters and, from each cluster one node is chosen as cluster-head (CH) according certain criteria. The latter is responsible for inter and intra-cluster data management [3]. CHs are able to receive data packets from cluster members, aggregate them into a single packet and forward it to the base station. Clustering is widely followed by researchers to achieve the network scalability objective, data aggregation/fusion, less load and more robustness while optimizing energy consumption. Two types of clustering algorithms can be used: static clustering and dynamic clustering. In dynamic clustering, clusters are regenerated after each period while keeping the same number of clusters during the life of the network, whereas in static clustering, clusters are formed once and for all during the life of the network and CHs are selected from cluster members [4]–[6].

In this paper, we propose a dynamic clustering where the process of clustering is performed under a predetermined condition. The clustering process is triggered again when the percentage of dead nodes represents less than 5% of the alive nodes and in this case the number of clusters is recalculated using Silhouette method [7] and the clusters are created using the K-Means approach [8].

The remainder of this paper is organized as follows: Section 2 presents the related work, Section 3 describes the approaches used in this paper and Section 4 concerns details of our contribution. The results and discussions are presented in section 5. Finally, the section 6 concludes the paper.

II. Related Work

Several energy-efficient routing approaches have been proposed in the literature to deal with the limited battery life of sensor nodes in order to increase the network lifetime. Among them approaches clustering routing algorithms have gained much attention in WSNs. In this section, we briefly review the existing clustering algorithms based on K-means approach.

In [9], authors proposed a routing protocol called KMV-LEACH protocol in which k-medoids method is used for clustering process and LEACH protocol for cluster-head selection. In order to determine the optimal number of clusters, a comparison between Gap statistic, CalinskiHarabasz and Silhouette method has been done. The results show that the Silhouette method minimizes the energy consumption and extends network lifetime in WSNs compared to the two others validity measures.

In [10], authors presented a distributed K-means clustering approach in which the clustering process is performed by every sensor node in the network. The results show that distributed clustering is more efficient than centralized clustering approach in term of clustering process speed. However, energy consumption is almost the same in both approaches.

In [11], a new strategy for cluster-head election based on K-means method was developed wherein the authors take into account two parameters distance and remaining energy of nodes. Simulation results show that the proposed approach reduces energy consumption and prolongs network lifetime.

In [12], authors proposed an energy-efficient hybrid K-means approach for WSNs in which they have combined K-means approach and LEACH protocol. K-means approach is used for clustering process and LEACH protocol is used for cluster-heads selection. The simulation results show that this protocol is better than LEACH in terms of energy consumption and network lifetime but the major weakness of this protocol is the great energy consumption during cluster-head election process.

In [13] an improved clustering routing algorithm called LEACH-KED has been proposed. In this protocol, clustering process is performed after each round and cluster-head selection is based on a weight that represents the ability of a node to become a CH and K-means method is used to create clusters. The simulation results show that the proposed protocol outperforms LEACH [14] and LEACH-C [15] in term of energy consumption.

In [16], authors presented an approach to find the optimal number of clusters in networks with mobile base station. In this approach, they assume that the base station (BS) travels very close to the cluster-heads to collect aggregated data. This results in zero attenuation during data collection from CH by the BS. However, electing CHs throughout the life of the network may have a negative impact on the network lifetime since CHs perform several tasks that can drain their batteries quickly and therefore make the network inoperable.

III. Overview

Before heading into the technical details of our contribution, we first give a brief overview of the approaches used in our contribution.

A. K-means technique

K-Means [8] is one of the simplest unsupervised learning algorithms that solves the well known clustering problem. The procedure used by K-Means follows a simple way to classify a given data set through a predefined number of cluster (assume K clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the closest centroid. When no point is pending, the first step is completed and clusters are formed. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done i.e centroids do not move any more. Finally, this technique aims at minimizing an objective function, in this case a squared error function as presented in equation (1).

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^j - c_j\|^2 \qquad (1)$$

where $\|x_i^j - c_j\|^2$ is a chosen distance measure between a data point x_i^j and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

K-means is composed of the following steps:

1) Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2) Assign each object to the group that has the closest centroid.
3) When all objects have been assigned, recalculate the positions of the K centroids.
4) Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

B. Silouhette method

The concept of silhouette width [7] involves the difference between the within-cluster tightness and separation from the rest. Particularly, it provides a quantitative way to measure how well each item lies within its cluster as opposed to others. The average silhouette for each point is a measure of how similar that point is to points in its own cluster when compared to points in other clusters. The Silhouette value of a data point is defined as:

$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \qquad (2)$$

Where, a(i) is the average distance from the i^{th} point to other points within the same cluster. Let $b(i,k)$ be the average distance from the i^{th} point to the points in the k^{th} cluster. Then *b(i)* is the minimum of all *b(i,k)* over all clusters that the i^{th} point is not assigned to. The value of s(i) is not greater than one, and that s(i) is close to one indicates that the i^{th} point lies well within its own cluster.

The silhouette method is composed of the following steps:

- Execute K-means by varying k number of clusters from 1 to max,
- For each k, calculate the average silhouette of observations,
- Plot the curve of avg.sil according to the number of clusters k,
- The location of the maximum is considered as the appropriate number of clusters.

C. Distributed K-means Method

In [17], the authors proposed a new Distributed K-means scheme in which clustering process is performed in distributed manner and cluster-head election process is based on the node's weight in which distance and remaining energy are involved in calculating of the weight according to equation (3). The proposed routing scheme is carried out in two phases: set-up phase and steady-state phase. In set-up phase K-Means method is used for clustering and steady phase is performed in several rounds. Each round consists of two sub-phases namely election cluster-head sub-phase and communication sub-phase. In the first sub-phase, the election of CH process is performed in distributed manner based on the node's cost by taking into account residual energy and distance to centroids represented by:

$$Cost(Node_i) = \alpha * Energy + \frac{\beta}{dist(Node_i, Centroid)} \quad (3)$$

where $\alpha + \beta = 1 \quad \alpha \in]0.75, 1[\quad \beta \in]0.05, 0.25[$

Each node sends its cost to all the members of its cluster. The node with the greatest cost becomes cluster-head during the actual round. In the second sub-phase each cluster member sends the collected data to its respective CH in one hop according to its TDMA time slot. After receiving all the packets by the CH, the latter aggregates them in one packet and forwards it to the sink in one hop. This proposed clustering scheme based on K-Means approach provides better performance compared to LEACH protocol.

IV. Contribution

A. Assumptions

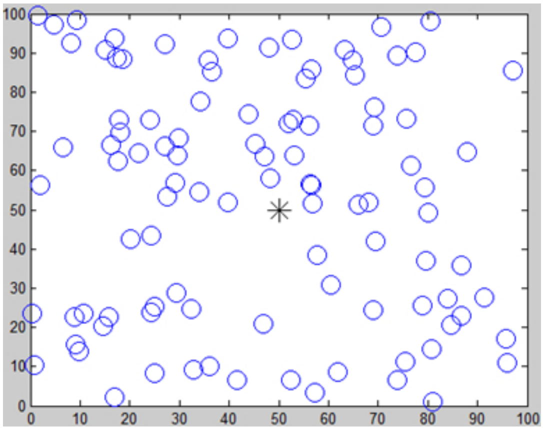
In this paper, we assume a sensor network model with the following properties:

- Nodes are randomly deployed in the environment and are static,
- Base station is located at the center of the network area,
- Sensor nodes are able to adjust the transmission power according to the distance that separates it to the desired node,
- Initially, all sensor nodes have equal energy,
- Location and ID of all nodes are known by the base station,
- Each node knows its own position.

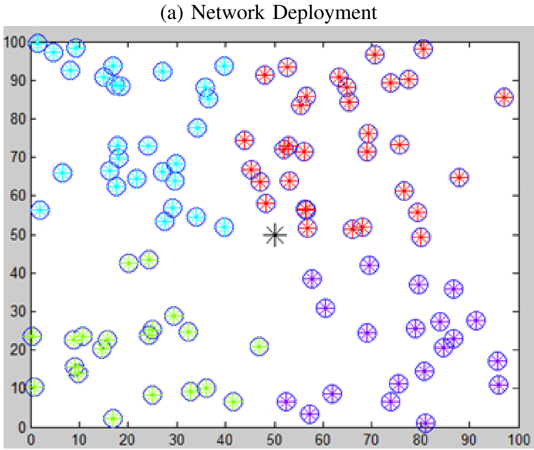
B. Proposed work

In this section, we proposed a dynamic distributed routing scheme based on Silhouette method [7] to determine the appropriate number of clusters and K-Means method [8]to create clusters. This proposed scheme consists of two phases: set-up phase, steady-state phase.

During the set up phase, the base station collects the positions of all the nodes of the network using a localization system such as GPS (Global Positioning System) and performs K-means clustering approach after determining the number of clusters with silhouette method and the positions of the centroids. When the clustering process is completed the base station sends an advertisement packet to each sensor node that is affiliated to a cluster. Figure 1(a) illustrates the random deployment of nodes in an area of interest and the Figure 1(b) shows the formation of four clusters after running of Silhouette method and K-means clustering approach.



(a) Network Deployment



(b) Clusters generated by K-Means

Fig. 1: Clusters formed by the proposed contribution

The proposed routing scheme is carried out in several rounds which represent the duration of the set-up and the steady phases. Set-phase consists of three sub-phases namely determining the number of clusters by Silhouette method sub-phase, re-clustering sub-phase by K-Means method and cluster-head selection sub-phase. During cluster-head selection sub-phase, in each cluster, a cluster-head is elected in a distributed manner as follows: each member calculates its weight according to its residual energy and the distance that separates it from the centroid as presented in equation (3), and sends it to all the members of cluster to which it belongs. The node with the greatest weight becomes cluster-head during this round. Then, each CH sends a TDMA schedule to all members in its corresponding cluster. The set-up phase is triggered when the percentage of alive nodes decreases compared to a fixed threshold and during which Silhouette and K-Means methods are performed. When the clustering process is completed the base station sends an advertisement packet to each sensor node that is affiliated to a cluster. Figure 2 shows the formation of six clusters after running of the first and the second sub-phases.

Finally, the steady phase starts and during which each

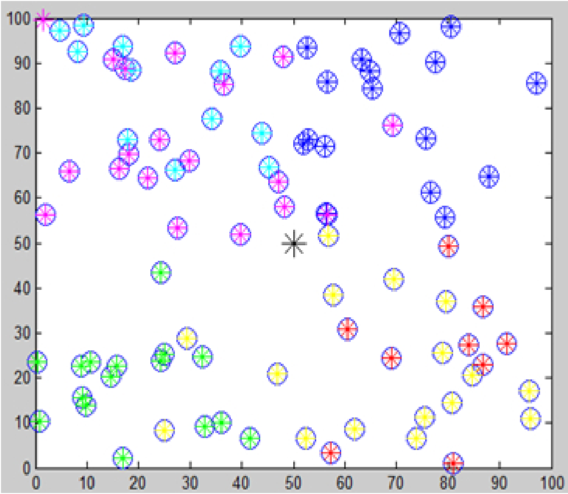


Fig. 2: Clusters generated in re-clustering phase

member sends the collected data to its respective CH according to its TDMA time slot in one hop. After receiving all the packets by the CH, the latter aggregates them in one packet and forwards it directly to the sink.

V. Evaluation and Simulation Results

In this section, we evaluated the proposed distributed clustering routing scheme and compared their performance to Distributed K-Means routing scheme [17] and LEACH protocol [14] in terms of energy consumption and network lifetime.

A. *Working environment*

All simulations were carried out using MATLAB simulator. We used two scenarios to evaluate the performance of our proposal. The first scenario consists of a network with 100 nodes distributed randomly in an area of 100m x 100m and the base station is placed at (50,50) point. The second scenario consists of 500 nodes distributed randomly in an area of 200m x 200m and the base station is located at the center of the area i.e (100,100) point. The second scenario is used to prove the robustness of our proposal when the number of nodes increases in the network. In our simulations we assumed that every node transmits a data packet per round to its respective cluster-head and the number of clusters in LEACH represents 5% of the total number of nodes in the network (P = 5%). The parameters used in the simulation are summarized in the table I.

B. *Radio Energy Model*

We used the same energy parameters and radio model presented in [14], wherein energy consumption is mainly divided into two parts: receiving and transmitting messages. The transmission energy consumption requires additional energy to amplify the signal according to the distance from the destination. To transmit a k-bit message to a distance d, the radio expends energy as described by equation (4), where ε_{elec} is the energy consumed by radio electronics, $\varepsilon_{friss-amp}$ and $\varepsilon_{two-ray-amp}$ for amplifier. The reception energy consumption is expressed by equation (5).

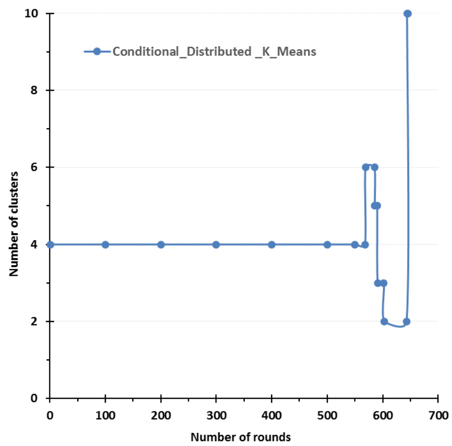
TABLE I: Simulation Parameters

Parameters	Values
Network Grid	(0,0)x(100,100) and (0,0)x(200,200)
Base Station	(50,50) and (100,100)
ε_{elec}	50 nJ/bit
ε_{friss_amp}	$10pJ/bit/m^2$
$\varepsilon_{two\_ray\_amp}$	$0.0013pJ/bit/m^4$
d_0	87 m
Transmission Range	30 m
Initial energy per node	0.5 J
Number of nodes	100, 500
P	5%

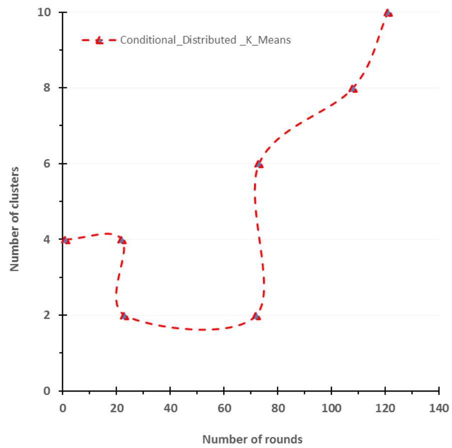
$$E_{Tx}(k,d) = \begin{cases} k * E_{elec} + \varepsilon_{fs} * d^2, & if\ d < d_0 \\ k * E_{elec} + \varepsilon_{amp} * d^4, & if\ d \geq d_0 \end{cases} \tag{4}$$

$$E_{Rx}(k) = k * E_{elec} \tag{5}$$

C. *Simulation Results*



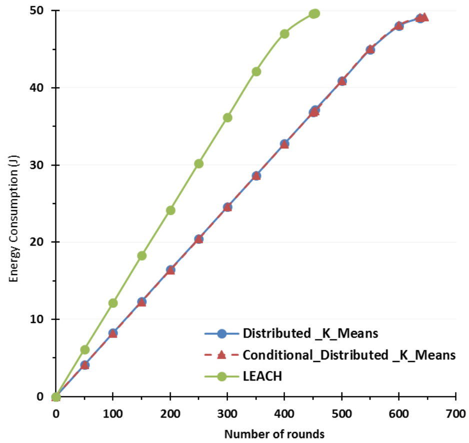
(a) Network with 100 nodes



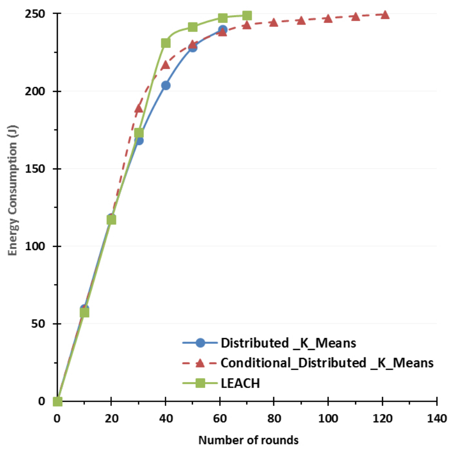
(b) Network with 500 nodes

Fig. 3: Number of clusters formed by the proposed approach

Figures 3(a) and 3(b) illustrate the number of clusters respectively in the network with 100 nodes and the network with 500 nodes. We notice when the number of nodes decreases the number of clusters increases this is due to the dispersion of the nodes in the zone of deployment.



(a) Network with 100 nodes

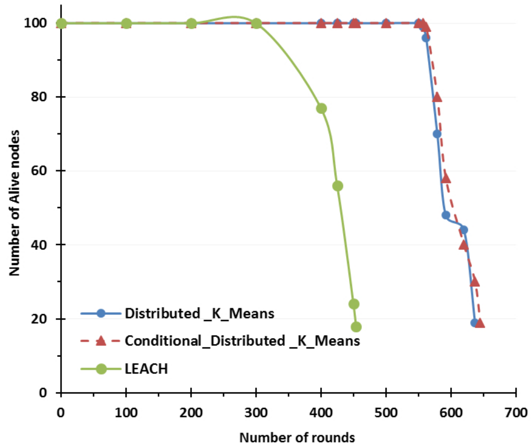


(b) Network with 500 nodes

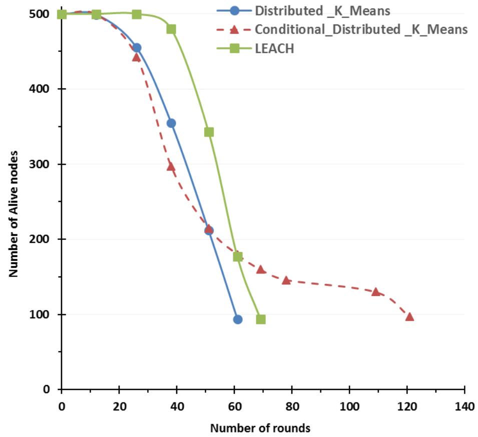
Fig. 4: Energy Consumption

Figures 4(a) and 4(b) show energy consumption during network lifetime. The proposed routing scheme provides better results compared to Distributed K-Means routing scheme and LEACH protocol due to the good distribution of energy consumption during network lifetime. Moreover, in our contribution, the number of clusters formed according to the number of alive nodes had a positive impact on the performance of our proposal while in LEACH and distributed K-Means routing scheme the number of clusters is independent of the number of alive nodes in the network. In LEACH, this number is equal to 5 for a network of 100 nodes and 20 for a network of 500 nodes and in distributed K-Means routing scheme it still keeps its initial value.

Figures 5(a) and 5(b) show the number of alive nodes during network lifetime. We assume that when the number of alive nodes reaches 20 in the first network this latter is considered



(a) Network with 100 nodes



(b) Network with 500 nodes

Fig. 5: Number of Alive nodes

as no functional network. So in a similar way for the network with 500 nodes where when the number of living nodes is less than 100 we consider that it is not functional. In our contribution, the network restructuring based on the number of alive nodes has increased the network lifetime compared to the other two protocols as illustrated by Figure 5.

VI. Conclusion

In this paper, we proposed a new routing scheme for wireless sensor networks based on Silhouette method for finding optimal number of clusters and K-Means approach for clustering process. The proposed routing uses a dynamic clustering at every decreasing of the percentage of alive nodes under a fixed threshold (5%) and distributed manner to select cluster-heads. In clusters, cluster members sent their collected data to their respective CHs which in turn they sent aggregated packets directly to the base station.

Simulation results shown that our proposed routing scheme achieves lower energy consumption and longer network lifetime compared to LEACH protocol and Distributed-k-Means routing scheme due to the formation of clusters according to the number of alive nodes as well as their positions. Moreover, the performance of the proposed routing scheme is still better even for networks with a large number of nodes.

REFERENCES

[1] I. F. Akyildiz and I. H. Kasimoglu, "Wireless Sensor and Actor Networks: Research Challenges", Ad Hoc Networks Journal, vol. 2, no. 4, pp. 351–367, 2004.

[2] A. Gachhadar, O. N. Acharya, "K-means Based Energy Aware Clustering Algorithm in Wireless Sensor Network", International Journal of Scientific & Engineering Research, vol. 5, no. 5, pp. 156-161, 2014.

[3] C. Sivaraj , P. J. A. Alphonse and T. N. Janakiraman, "Energy-efficient and Load Distributed Clustering Algorithm for Dense Wireless Sensor Networks", International Journal of Intelligent Systems and Applications, vol. 5, pp. 34–42, 2017.

[4] M. Busse, T. Haenselmann and W. Effelsberg, TECA: A topology and energy control algorithm for wireless sensor networks, In Proceedings of the 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Malaga, Spain, October 2006.

[5] Y. Chen and S. H. Son, A fault tolerant topology control in wireless sensor networks, In Proceedings of ACS/IEEE International Conference on Computer Systems and Applications, Cairo, Egypt, January 2005.

[6] R. Ghosh and S. Basagni, Napping backbones: Energy-efficient topology control for wireless sensor networks, In Proceedings of IEEE Radio and Wireless Symposium, San Diego, USA, January 2006.

[7] L. Kaufman and P. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis", New York: J. Wiley & Son, 1990.

[8] J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, pp.281-297, 1967.

[9] A., Mahboub, E. M. En-Naimi, M. Arioua and A. Hatim, Performance Evaluation of Cluster Validity Methods an Energy Optimization in Wireless Sensor Networks Using Hybrid K-Medoids Algorithm, In Proceedings of BDCA'20717, March 2017, Tetouan, Morocco.

[10] P. Sasikumar, S. Khara, K-means clustering in wireless sensor networks, In Proceedings of 4th International Conference on Computational Intelligence and Communication Networks, pp.140-144, 2012, .

[11] S. Randhawa and S. Jain, "Performance Analysis of LEACH with Machine Learning Algorithms in Wireless Sensor Networks", International Journal of Computer Applications, vol. 147, no. 2, pp.712, 2016.

[12] A. Mahboub, M. Arioua, and E. M. En-Naimi, Energy-Efficient Hybrid K-Means Algorithm for Clustered Wireless Sensor Networks, International Journal of Electrical and Computer Engineering (IJECE), 7(4), p.2054-2060, 2017.

[13] J. Yunjie, L. Ming, Z. Song and D. Pengtao, A clustering routing algorithm based on energy and distance in WSN, IEEE, International Conference on Computer Distributed Control and Intelligent Enviromental Monitoring, 2012.

[14] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", IEEE Trans. Wireless Communications, vol. 1, n. 4, pp. 660–670, 2002.

[15] V. Geetha et al., "Clustering in wireless sensor networks: Performance comparison of LEACH and LEACH-C protocols using NS2", Procedia Technology, vol. 4, pp. 163-170, 2012.

[16] C.P. Gupta and A. Kumar, "Optimal Number of Clusters in Wireless Sensor Networks with Mobile Sink", International Journal of Scientific & Engineering Research, Vo. 4, no. 4, pp.1706–1710, August 2013.

[17] M. Lehsaini and M. B. Benmahdi, An improved K-means cluster-based routing scheme for wireless sensor networks, International Symposium on Programming and Systems (ISPS), pp. 1–6, Algiers, 2018.