An Efficient Clustering Algorithm Based on Expectation Maximization Algorithm in Wireless Sensor Network

Mehrsa Moghadaszadeh[1], Hamid Shokrzadeh[2]

[1] Department of Computer Engineering Pardis Branch, Islamic Azad University, Pardis,Iran
moghadaszadeh.6mehrsa@gmail.com
[2] Department of Computer Engineering Pardis Branch, Islamic Azad University, Pardis,Iran
shokrzadeh@gmail.com

Abstract: *Wireless sensor network (WSN) consists of a number of nodes to sense the data that the nodes collect and transmit the messages to Sink or Base Station (BS). The WSN is used in many applications such as defense, industrial services, etc. However, it faces some problems such as energy consumption, an increment network life time, etc. In order to overcome such problems in the WSN during the past years, many routing and data gathering protocols have been introduced such as LEACH, HEED, PEGASIS, GSTEB. In this paper, we propose a new algorithm in which Expectation Maximization (EM) pattern recognition method is used in a clustered base routing protocol. The simulation results illustrate that the proposed algorithm (EM-based clustering) performs better than existing approaches such as K-Means LEACH.*

Keywords: *Expectation Maximization, Clustering, Energy consumption, Network life time.*

1. Introduction

In recent years, the WSN has been introduced as a critical technology. In the single-hop method, each node directly sends data to the BS in which higher energy consumption is in the network. To overcome the aforementioned problem, hierarchical approach [1] has been introduced based on the clusters formed in the network. In other words, sub-groups of the nodes are created where each cluster collects and transmits the data to BS to reserve energy in the network. Thus, there is lower energy consumption and higher lifetime in the network. In the hierarchical approach many protocols have been developed including low-energy adaptive clustering hierarchy (LEACH) introduced as the first hierarchical clustering protocol in the WSN which suffers from limitations such as the following: all of the nodes are chosen with fixed probability as cluster head. In other words, some nodes may be selected as unsuitable cluster head nodes as a result of which energy is wasted in the network. In this protocol, some clusters may be formed only with a node; therefore the node's energy is consumed up early due to direct transmission of information to the BS.

In regard to the aforementioned drawbacks, an incremental, evolving lifetime of the network is required for this protocol and it is necessary to have a replacement role for basic clustering. In order to overcome the aforementioned limitations, a version of the LEACH known as K-Means LEACH which employs K-Means algorithm for cluster formation. The K-Means is a clustering algorithm in data mining that it has been introduced by MacQueen in 1967 in which cluster formation is explained as follows:

Basically, k points are selected randomly as initial centers and then Euclidean distance between each point to the centers is calculated. Finally, the cycle is repeated until it is converged:

Then each point is reassigned to the cluster with minimum distance and the cluster means are updated. Finally, it achieves convergence [2].

However, the K-Means algorithm suffers from some limitations as follows: One drawback is that the result depends on the selection of the initial clusters and it is really sensitive to noise data. Firstly, the number of clusters is specified (assumes k clusters). However, this is not required in many applications.

Many methods have been proposed to improve the K-Means one of which is the EM algorithm. The EM performs better than the K-Means algorithm for the following reasons: Convergence speed of the EM is higher than the K-Means and, therefore, the clusters are formed faster. The EM can consider a large collection of sense nodes which is very important in the design of WSN. The EM selects the best center in each cluster and has higher clustering quality or accuracy than the K-Means. The EM algorithm will be elaborated on in Section 3.

The remaining part of the paper is organized as follows: Section 2 is concerned with related works. The proposed approach and simulation results are discussed in Section 3 – 4. Finally, Section 5 concludes the paper.

2. Related Work

During the past years, many protocols have been suggested to improve the clustering algorithm in the WSN such as LEACH-C [3], K-Means LEACH, etc. In [4] a clustering technique is used based on Elbow and the K-Means methods where clustering is performed by the K-Means and using the Elbow is selected k value, i.e., the number of clusters. Basically, the Elbow finds some best clusters and then clustering formation is performed by the K-Means. LEACH_CKM algorithm was introduced in 2015. In this approach, two algorithms are combined as K-Means and Minimum Transmission Energy (MTE). Also, the K-Means method is considered instead of Simulated Annealing algorithm in the LEACH-C protocol [5]. In 2015, researchers in [6] presented a seminal approach in which they used Genetic algorithm for cluster formation due to improvements in the LEACH. In [7] the author suggested a method based on DBIK-Means and Gauss algorithms. In this approach, the K-Means is used to create clustering and Gauss algorithm is developed for Cluster Head (CH) selection[8]. Here, Davis-Bouldin Index (DBI) method[9] is used which finds the best number of clusters in the K-Means. In[10], a method is proposed using K-Means algorithm for clustering and two levels of fuzzy logic for cluster head selection based on neighbor nodes, residual energy, distance to the BS and energy spreading parameters. K-Means LEACH and Fuzzy LEACH are proposed as an improvement to the LEACH protocol.

3. EM -Based clustering algorithm

We use the EM clustering algorithm that is introduced in more detail below:

3.1. Expectation Maximization

The EM algorithm is an approach used for parameters estimation of a statistical model such as mean, variance and weight according to maximum likelihood estimate [11]. This algorithm is iterative where each iteration consists of two main phases known as Expectation (E-step) and Maximization (M-step). In other words, the algorithm sequentially switches between the aforementioned steps, i.e., the Expectation and the Maximization steps. Firstly, the E-step (averaging) is calculated for maximum likelihood estimate (MLE) and then the M-step is completed for maximization of the obtained parameters in the E phase; this process continues until convergence is reached for the last values of the parameters [11]. The convergence is determined based on the log-likelihood value (as shown in equation 1) after every iteration, when the algorithm has not changed from one iteration to the next.

$$\log l(\Theta) = \sum_{i=1}^{N} \log p(x_i \mid \Theta) = \sum_{i=1}^{N} \left(\log \sum_{k=1}^{K} \alpha_k \, p_k \left(x_i \mid z_k \, \Theta_k \right) \right) \tag{1}$$

Where parameter vector Θ presents the above-mentoined parameters and $p(x_i \mid \Theta)$ is a mixture model. Here, a data set $X = \{x_1, x_2, \dots, x_N\}$ is introduced where x_i is a d-dimensional vectore measurement $1 \le i \le N$ and $p_k(x_i \mid z_k \, \Theta_k)$ is the Gussian denesity function for the k^{th} mixture component where z_k is K-ary random variable representing the identity of the mixture component that generated x_i. It is worth noting that α_k is introduced as the membership weight for x_i and k is all mixture components $1 \le k \le K$.

3.2. EM-based Clustering

We propose a solution to alleviate the existing problems of the K-Means family used in a number of clustering protocols. Our proposed method is based on the EM clustering algorithm and leads to efficiency in

https://doi.org/10.15242/DiRPUB.DIR1017011

energy and higher network lifetime in the network. It is worth noting that the main phases are divided into two steps 1-setup phase 2-steady state (as shown in Figure 1). During the setup phase, clustering is performed according to the EM algorithm in the Gaussian mixture model. The current paper aims to create clusters on the basis of EM algorithm as defined for Gaussian mixture models (GMM). The EM algorithm finds clusters using the Gaussian mixture model (GMM) as result of which cluster's parameters with maximum likelihood are obtained.

- **Initialization**

First, the EM algorithm uses initial estimate of vector (θ) including the mean (μ_j) and the covariance matrix (Σ_j) and M parameter which illustrates the total number of clusters [12].

$$\theta(t) = \mu_j(t), \Sigma_j(t), \; j = 1....M \tag{2}$$

- **E-step**

This step is responsible for estimating membership probability to each cluster (C_j) as shown in equation 3.

$$P(C_j \mid x) = \frac{\left|\Sigma_j(t)\right|^{-\frac{1}{2}} \exp^{n \; j} P_j(t)}{\sum_{k=1}^{M} \left|\Sigma_j(t)\right|^{-\frac{1}{2}} \exp^{n \; j} P_k(t)} \tag{3}$$

- **M-step**

Here, parameters are estimated for the iteration $(t+1)$ in the equations below where the Gaussian model parameters are updated.

$$\mu_j(t+1) = \frac{\sum_{k=1}^{N} P(C_j \mid x_k) x_k}{\sum_{k=1}^{N} P(C_j \mid x_k)} \tag{4}$$

$$\sum_j (t+1) = \frac{\sum_{k=1}^{N} P(C_j \mid x_k)(x_k - \mu_j(t))(x_k - \mu_j(t))^T}{\sum_{k=1}^{N} P(C_j \mid x_k)} \tag{5}$$

$$W_j(t+1) = \frac{1}{N} \sum_{k=1}^{N} P(C_j \mid x_k) \tag{6}$$

As shown in equation 7, a convergence test is performed after each iteration. In this step, if the difference of iteration $(t+1)$ from iteration (t) is smaller than an acceptable error tolerance then Q parameter belongs to the inputs of the EM. Thus, convergence is reached; otherwise the previous steps re-performed in this algorithm [12].

$$\left\|\theta(t+1) - \theta(t)\right\| < Q \tag{7}$$

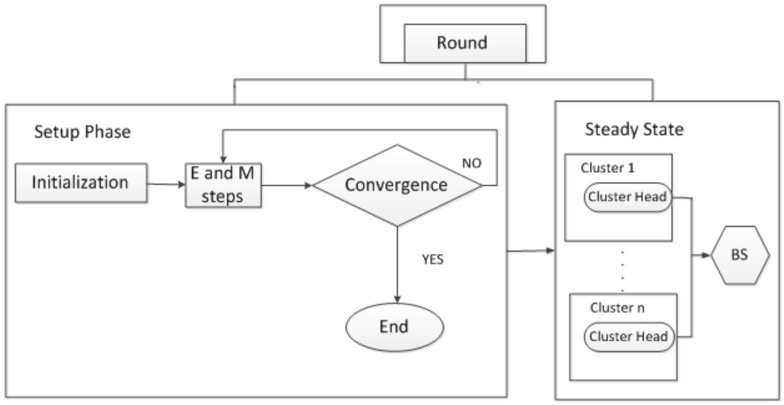

Fig. 1: Diagram of the proposed algorithm

In the steady state, each of the cluster heads creates a Time Division Multiple Access (TDMA) schedule and sends it to its cluster members. Each of the nodes transmits the information to its cluster head in its time slot. As shown in Figure 2, every cluster head sends the data to the BS after receiving the information.

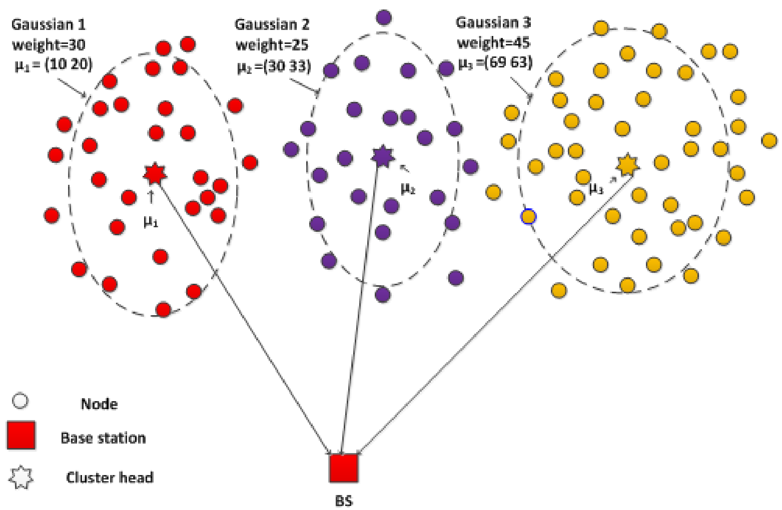

Fig. 2: The data transmission to the BS

4. Results and Analysis

In order to examine the proposed algorithm, the experiments on the proposed algorithm are simulated using MATLAB. To simulate the proposed method, we have considered 100 sensor nodes that are randomly distributed over a field of (100×100) square unit area in the network. We use the same radio and energy models as in [13]. Network lifetime and efficiency energy are important considerations in the WSN. The proposed algorithm is compared with the LEACH and K-Means LEACH protocols in the charts below. Figure 3 shows the number of live nodes in the proposed algorithm and the K-Means LEACH and the LEACH algorithms in 9999 iterations. In regard to the figure, the proposed algorithm has the highest value compared to other protocols. Here, lifetime is defined as the failure time of the first node in the network. Figure 4 compares the network lifetime of the three algorithms in which nodes die in the K-Means LEACH and the LEACH protocols earlier than in the proposed algorithm. Therefore, the proposed algorithm lifetime performs better than the protocols for simulation rounds. There is a significant difference between the protocols as the proposed algorithm has increased the network lifetime in comparison with other protocols. The total residual energy of the nodes in the networks has been shown in Figure 5 where the proposed algorithm shows the highest residual energy in comparison to other protocols. Therefore, the proposed algorithm performs better compared to other protocols.

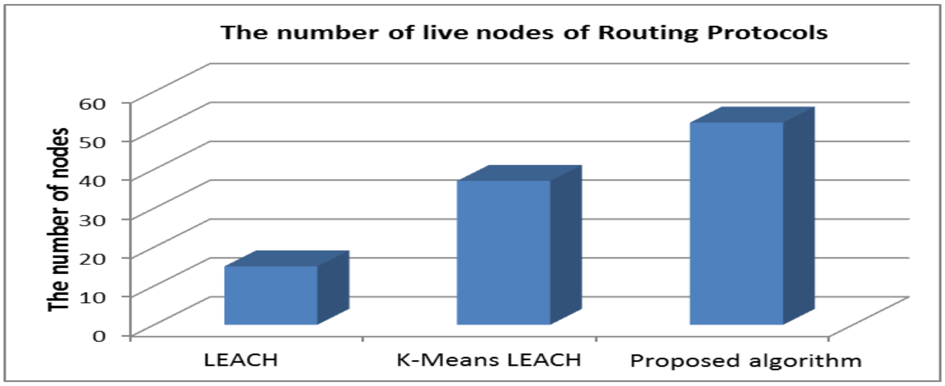

Fig. 3: the number of live nodes compared in routing protocols

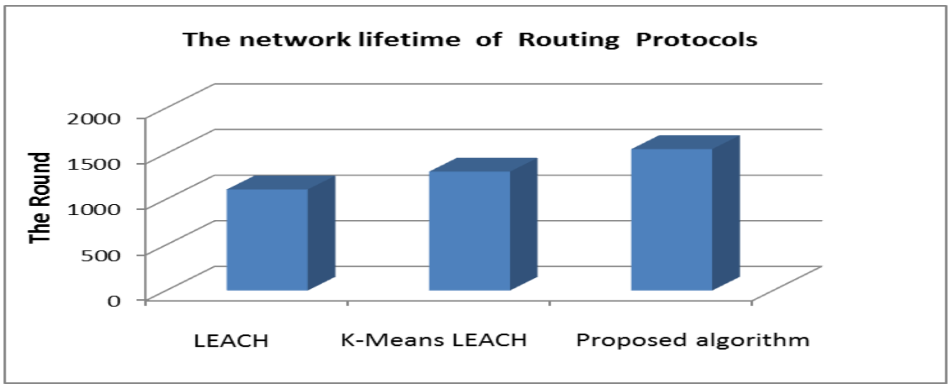

Fig. 4: the network lifetime compared in Routing Protocols

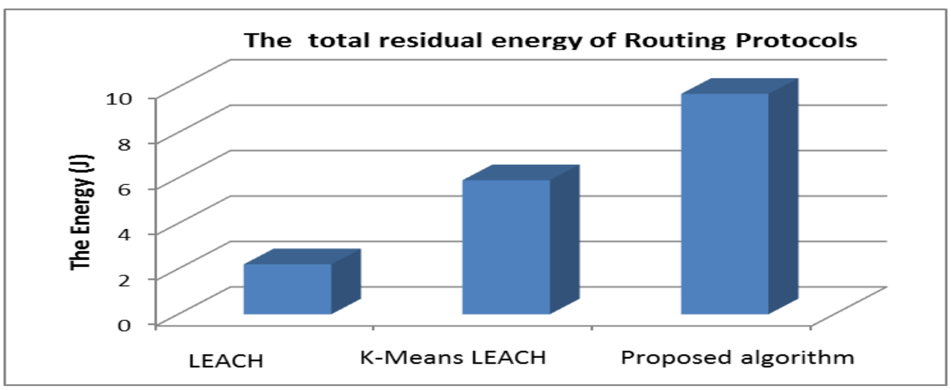

Fig. 5: the total residual energy compared in Routing Protocols

Figure 6 compares the residual energy average in the networks where the average is higher in the proposed algorithm in comparison to the LEACH and the K-Means algorithms. Energy consumption is total energy consumed by the nodes for receiving and transmission information in the network. As shown in Figure 7, the graphical results indicate that energy consumption of the proposed algorithm is less than the other protocols. Since energy consumption is an important factor in the network design; therefore, the proposed algorithm performs better compared to other protocols.

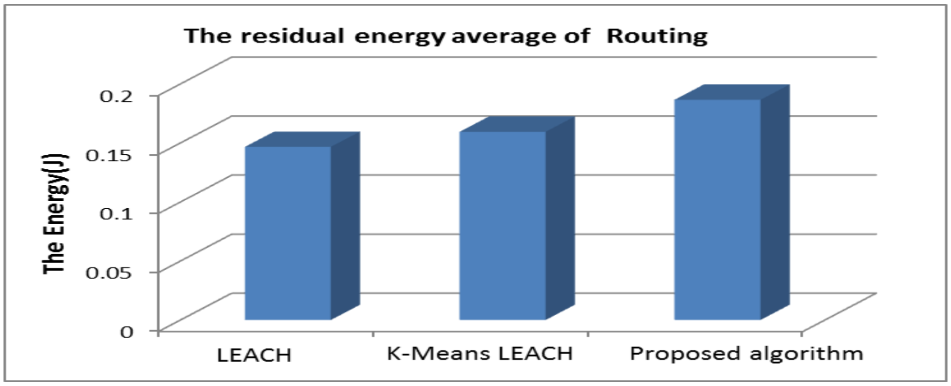

Fig. 6: Residual energy average in the Routing Protocols

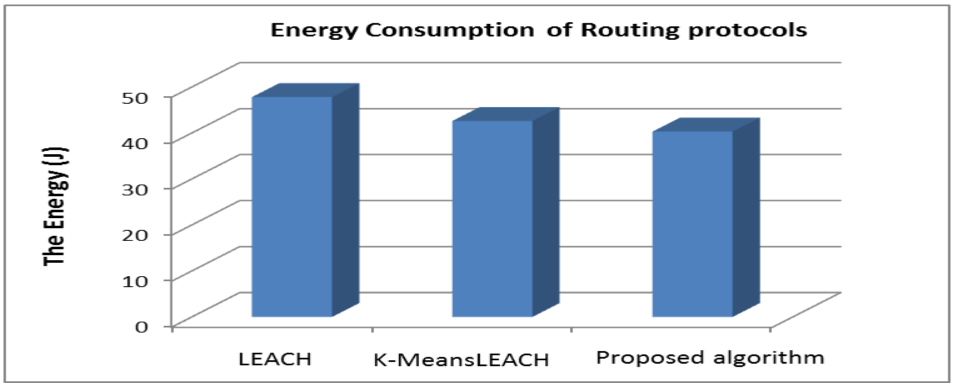

Fig. 7: Energy Consumption in Routing Protocols

5. Conclusion

In this paper, clustering routing protocols using the K-Means model was enhanced by the EM algorithm in the WSN. The EM performed better than the K-Means in some cases such as convergence speed, data set size, accuracy, etc. Thus, we introduced the EM-based clustering algorithm showing better results than the aforementioned protocols. In other words, the proposed algorithm can improve significant factors such as network lifetime, the total residual energy of nodes, the energy average of the network, etc. In further search, our algorithm extension can continue with Distributed Expectation Maximization (DEM), Genetic Algorithm-Expectation Maximization (GA-EM) and Particle Swarm Optimization-Expectation Maximization (PSO-EM) methods for lower energy consumption and higher lifetime in the network.

6. References

[1] H.Shokrzadeh, and M. Moghadaszadeh "A STUDY AND COMPARISON OF GSTEB METHODS WITH THE GOAL TO DATA GATHERING FROM HIERARCHICAL SENSORS," in *Proc. 2016 CSME Conf.*, 2016, pp. 1-5.

[2] M. Pc, S. Meenu, and L. R. P, "Fault Prediction using Quad Tree and Expectation Maximization Algorithm," vol. 2, no. 4, pp. 36–40, 2012.

[3] A. F. S. D. Muruganthan, D. C. F. Ma, B. Rollyi, "A centralized energy-efficient routing protocol for wireless sensor networks,"," *IEEE Radio Commun.*, vol. 43, pp. 8–13, 2005.

[4] P. Bholowalia, and A.Kumar "EBK-Means : A Clustering Technique based on Elbow Method and K-Means in WSN," vol. 105, no. 9, pp. 17–24, 2014.

[5] D. Mechta, S. Harous, I. Alem, and D. Khebbab, "LEACH-CKM: Low Energy Adaptive Clustering Hierarchy protocol with K-means and MTE," *2014 10th Int. Conf. Innov. Inf. Technol. IIT 2014*, pp. 99–103, 2014.

[6] J. Kaur and A. Kumar. (Jun-Dec 2015). GA Based Balanced Clustering Approach for Energy Efficiency in WSN. 9(1). pp. 24-31. Available: http://www.csjournalss.com

[7] E. Rabiaa, B. Noura, and C. Adnene, "Improvements in LEACH based on K-means and Gauss algorithms," *Procedia Comput. Sci.*, vol. 73, no. Awict, pp. 460–467, 2015.

[8] S. a. Nikolidakis, D. Kandris, D. D. Vergados, and C. Douligeris, “Energy efficient routing in wireless sensor networks through balanced clustering,” *Algorithms*, vol. 6, pp. 29–42, 2013.

[9] D. L. Davies. and D .W. Bouldin., “A Cluster Separation Measure,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no.2, pp. 224–227, 1979.

[10] K.E.Kannammal, T.Purusothaman, and M.S.manjusha, “An Efficient Cluster Based Routing in Wireless Sensor Networks,” *Technology, Applied Information*, vol. 59, no.3, pp. 683-689, 2014.

[11] H. Shokrzadeh, S. Khorsandi, and A. T. Haghighat, “Optimized query-driven appointment routing based on Expectation- Maximization in wireless sensor networks,” *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1749–1761, 2012.

[12] A. A. Adebisi, O. E. Olusayo, and O. S. Olatunde, “An Exploratory Study of K-Means and Expectation Maximization Algorithms,” *Br. J. Math. Comput. Sci.*, vol. 2, no. 2, pp. 62–71, 2012.

[13] W. B. Heinzelman, A. P. Chandrakasan, S. Member, H. Balakrishnan, S. Member, and H. Balakrishnan “An application-specific protocol architecture for wireless micros sensor networks,” *IEEE Trans. Wirel. Commun.*, vol. 1, no.4, pp. 660–670, 2002.