EEC-FM: Energy Efficient Clustering based on Firefly and Midpoint Algorithms in Wireless Sensor Network

Ravuri Daniel[1] and Kuda Nageswara Rao[2]
[1]Department of Information Technology, Vignan's Institute of Information Technology
Jawaharlal Nehru Technological University-Kakinada, Visakhapatnam, India
[e-mail: danielravuri@gmail.com]
[2]Department of Computer Science & Systems Engineering
Andhra University Collège of Engineering (A), Andhra University, Visakhapatnam, India
[e-mail: knraoauce@andhrauniversity.edu.in]
*Corresponding author: Ravuri Daniel

Received October 28, 2017; revised January 1, 2018; accepted April 9, 2018; published August 31, 2018

Abstract

Wireless sensor networks (WSNs) consist of set of sensor nodes. These sensor nodes are deployed in unattended area which are able to sense, process and transmit data to the base station (BS). One of the primary issues of WSN is energy efficiency. In many existing clustering approaches, initial centroids of cluster heads (CHs) are chosen randomly and they form unbalanced clusters, results more energy consumption. In this paper, an energy efficient clustering protocol to prevent unbalanced clusters based on firefly and midpoint algorithms called EEC-FM has been proposed, where midpoint algorithm is used for initial centroid of CHs selection and firefly is used for cluster formation. Using residual energy and Euclidean distance as the parameters for appropriate cluster formation of the proposed approach produces balanced clusters to eventually balance the load of CHs and improve the network lifetime. Simulation result shows that the proposed method outperforms LEACH-B, BPK-means, Park's approach, Mk-means, and EECPK-means with respect to balancing of clusters, energy efficiency and network lifetime parameters. Simulation result also demonstrate that the proposed approach, EEC-FM protocol is 45% better than LEACH-B, 17.8% better than BPK-means protocol, 12.5% better than Park's approach, 9.1% better than Mk-means, and 5.8% better than EECPK-means protocol with respect to the parameter half energy consumption (HEC).

Keywords: clustering, firefly, midpoint, unbalanced clusters, overload clusters, cluster heads

1. Introduction

With the recent development of sensors due to advancement of micro-electro-mechanical systems (MEMS) technology wireless sensor network (WSN) [1, 2] have gained recognition use in a variety of applications such as remotely monitoring applications [3,4], localization [5], target tracking [6,7], structural monitoring [8,9], healthcare [10,11], and industrial automation [12,13]. Clustering is an efficient routing mechanism [14] in WSN that aggregates data which is to be sent to BS. Efficient clustering [15–21] can achieve energy savings and extends network lifetime. A significant problem in clustering is to improve unbalance structure of the clusters and optimization of cluster heads (CHs) selection. For many practical applications of WSN is very useful in producing efficient balanced clusters. Though several approaches [22–32] are suffering from the following limitations due to initial centroids of CHs are chosen randomly.

a) An empty cluster may result due to the random selection of initial centroids .
b) May result residual nodes.
c) The number of desired clusters has to be selected by the user as input to the algorithm.
d) Workload of the CHs may not be balanced.
e) Optimum number of CHs are not selected.
f) May not be suitable for multi-hop routing.

Fig. 1. shows the existing algorithms produces unbalanced clusters due to random selection of initial centroids of cluster heads (CHs). Here cluster 2 and cluster 4 contains less than eleven sensor nodes. However, all the other two clusters contain sensor nodes much more than they should contain. As a result, these two CHs become overloaded and get exhausted earlier. The proposed EEC-FM protocol improves the initial centroid CH selection of using midpoint algorithm (MA) resulting in balanced clusters using firefly algorithm (FFA) compared to other protocols. It also optimizes CH selection by considering residual energy and cluster density are the parameters of CH selection in addition to Euclidean distance.

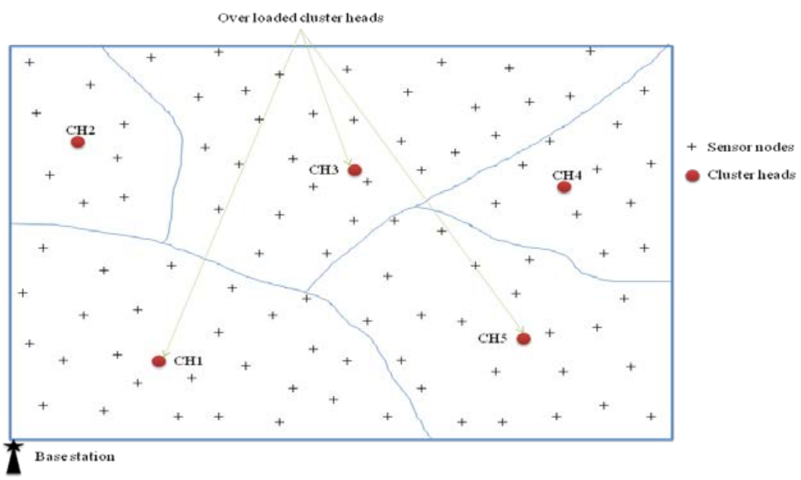


Fig. 1. Unbalanced clusters using existing clustering approaches of 200 x 200 m^2 sensor field.

The rest of the paper is organized as follows. Section 2 discusses related research work. Section 3 describes the basic preliminaries of FFA clustering algorithm and energy models. Section 4 presents the methodology of proposed EEC-FM protocol. Section 5 discusses simulation results and analysis in detail. Finally, section 6 concludes the paper.

2. Related Work

As the energy of each sensor node in a WSN is limited, energy efficient sensor network is an important goal to achieve. Clustering is a major approach to achieve energy efficient in WSN which also provides uniform coverage to a greater extent [14-17]. However, these algorithms suffers from several limitations as discussed in section 1. To deal with the limitation of these algorithms, several approaches are there to compute initial cluster centers [22-32] and calculate the number of clusters which improves energy efficiency. The authors in [22-26] proposed an efficient k-means algorithm which performs clustering without pre-assigning the exact number of clusters. However, these algorithms results in different kinds of clusters in different runs based on the randomly chosen initial centroid and are unable to produce balanced clusters. In [27] an efficient hybrid approach has been proposed based on PSO, ACO and K-means for cluster analysis, but this hybrid approach suffers from balancing of workload of clusters. Clustering based on multiple parameters [28] using K-means provides effective clustering in WSN. However, the limitation is that the number of desired clusters is not calculated accurately and this information has to be given as a user input to the algorithm.

A balanced parallel K-means (BPK-means) clustering protocol was proposed in [29] which employs K-means algorithm to cluster the sensor nodes. CHs are chosen depending on their distance from cluster centre and residual energy. However, threshold distance between the CH and the BS and the energy of sensor nodes are not considered for determining the centroids clusters. Therefore a significant amount of energy reduction takes place for CHs which has an impact on the network lifetime eventually.

In [30] an energy efficient CH selection technique based on K-means clustering algorithm for WSN was proposed, wherein initial centroids are selected randomly. Here residual energy of sensor nodes as a parameter of CH selection, but no estimation is specified regarding energy. A single hop communication between the CH and the BS takes place instead of multihop. Morover, the threshold distance between the CH and the BS is not considered.

In order to enhance network lifetime, a balanced CH selection strategy is proposed using modified K-means [31-32]. Here more than one CH in a cluster is considered so as to achieve reduction in time and energy required for re-clustering. Energy efficient clustering protocol named EECPK-means based on K-means and midpoint algorithm was proposed in [33] which uses mid-point algorithms for initial centroids selection. Though workload of CHs is comparatively better than other approaches, it still has the limitation of unbalanced clusters.

Objectives of Proposed Work

As discussed above, the most important issues regarding clustering are to improve cluster structure, optimise the selection of CHs and reduce energy consumption for data transmission. Hence, the main objectives of the proposed energy efficient clustering protocol EEC-FM are as follows:

i. To calculate the optimum number of desired clusters based on the size of the sensing region and the number of sensors present in it.

ii. To apply midpoint method for initial CHs selection, instead of choosing initial CHs randomly.
iii. To consider residual energy of sensor nodes as the parameters of CH selection in addition to the Euclidean distance used in existing algorithms and the threshold distance between the CH and the BS considered, firefly algorithm is used for cluster formation in such way that the CHs can effectively deliver the aggregate data to the BS.
iv. To avoid residual nodes.
v. To reduce energy consumption of CHs for data communication.

3. System Models

A wireless sensor network consists of spatially distributed autonomous sensor nodes to cooperatively monitor physical or environmental conditions. The nodes communicate wirelessly and are self-organized after being deployed in an ad-hoc fashion. The system model consists of the following sub models:

1. Network Model
2. Radio Energy Dissipation Model and
3. Energy Model For Clustering

3.1 Network model for WSN

The wireless sensor network basically consists of N homogeneous and/or heterogeneous sensor nodes deployed in the monitoring area which can sense, do little process and communicate. The following assumptions are made for modelling the proposed EEC-FM.

- The nodes are randomly deployed in a uniform manner within an M×M square region.
- All the nodes in the network are stationary and have uniform energy at the time of deployment.
- A fixed base station (BS) can be presented outside the sensor fields and this is assumed to have infinite power source and finite storage.
- Without GPS usage, BS knows the complete geographical information of sensor nodes
- Depending on the distance from the BS, CHs follow either single-hop or multi-hop communications.

3.2 Radio Energy Dissipation Model

The radio model used in this work [15] has been used in earlier works. This model is first order energy model for energy consumed when communication occurs between two nodes. Here, both the free space and the multipath fading channel models were used depending on distance between the transmitter and the receiver. Thus, if a node transmits ‘l’ number of bits, the energy used in transmission will be for the distance ‘d’ between two nodes, the transmitter energy consumption for transmitting ‘l’ bit is calculated using equations (1) and (2).

$$E_{TX}(l,d) = E_{elec}.l + E_{amp}(l,d) \tag{1}$$

$$E_{TX}(l,d) = \begin{cases} lE_{elec} + l.\varepsilon_{fs} d^2 & if\, d < d_0 \\ lE_{elec} + l.\varepsilon_{amp} d^4 & if\, d > d_0 \end{cases} \tag{2}$$

The threshold values is calculated by

$$d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{amp}}} \tag{3}$$

The energy consumed by the receiver is given by

$$E_{RX}(l) = E_{elec}.l \tag{4}$$

E_{elec}., ε_{fs}, ε_{amp} represent the coefficients of energy consumption for different channel propagation models. d_0 is a threshold value to distinguish free-space path loss model from a multipath fading model.

3.3 Energy Model For Clustering

Since the distance between the communicating CHs and the BS is assumed to be less than the threshold distance mentioned in (3), here it follows free space radio energy model for energy consumption mentioned in (1). BS calculates the number of sensor nodes n_c in each cluster after cluster formation phase. CH which sends data via nearby cluster heads (nCHs), consumes energy per round as follows

$$E_{nCH} = (n_c - 1)E_{elec}.l + n_c l.E_{DA} + (l.E_{elec} + l.\varepsilon_{fs} d_{BS}^2) \tag{5}$$

$$n_c = \frac{N}{k_{opt}} \tag{6}$$

CHs whose $d_{toBS} \leq d_{threshold}$ can send the data of its own clusters as well as the data of the clusters whose CH cannot send data directly to the BS, consumes energy per round as follows

$$E_{CH} = ((n_c - 1) + \frac{k_n}{k_{opt} + k_n})E_{elec}.l + (n_n + \frac{k_n}{k_{opt} - k_n}).l.E_{DA} + (l.E_{elec} + l.\varepsilon_{fs} d_{BS}^2) \tag{7}$$

where, n_c is the number of sensor nodes in that cluster, k_n is the number of CHs which are unable to send data directly to the BS and k_{opt} is the total desired number of CHs. Here the value of k_n ranges from 0 to ($k_{opt} - 1$). The energy dissipation of each non-CH node per round is

$$E_{non-CH} = l.E_{elec} + l.\varepsilon_{fs} d_{CH}^2 \tag{8}$$

Therefore total energy dissipation in a round using EEC-MP protocol is calculated as

$$E_{round} = \sum_{K_c} E_{nCH} + \sum_{k_{opt} - k_c} E_{CH} + (N - k_{opt})E_{non-CH} \tag{9}$$

Here, N is the total number of sensor nodes deployed in the sensing region.

4. Proposed EEC-FM protocol

The most important issues regarding clustering are to improve cluster structure, optimise the selection of CHs and reduce energy consumption for data transmission. This motivates us to propose an energy efficient clustering protocol EEC-FM to resolve these issues. The proposed protocol, EEC-FM operation model is shown in **Fig. 2**, is to build up an improved clustering for lifetime enhancement in WSN. Midpoint method described in algorithm 1 has been applied for initial CHs selection, instead of choosing initial CHs randomly. The firefly optimization algorithm described in algorithm 2 has been applied for energy efficient cluster formation. The main contributions of the proposed protocol are as follows:

a) It calculates the optimum number of desired clusters based on the size of the sensing region and the number of sensors present in it. Suppose N is the total number of sensor nodes uniformly distributed in an M×M square sensing region. The optimum number of clusters k_{opt} can be obtained as follows [32]

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{fs} \quad M}{\varepsilon_{mp} \quad d^2_{CH\ to\ BS}}} \tag{10}$$

Here, $d_{CH\ to\ BS}$ is the distance from CH to BS, ε_{fs} is the parameter for free space mode and ε_{mp} is the parameter for multipath model.

b) Midpoint algorithm has been applied for initial CHs selection, instead of choosing initial CHs randomly. It obtains balanced cluster where CHs are uniformly distributed and each cluster contains an almost equal number of sensor nodes. As a result, the load of the CHs becomes balanced, which ultimately prolong the network lifetime.

c) The proposed protocol considers residual energy of sensor nodes as the parameters of CH selection in addition to the Euclidean distance, so that the CHs can successfully deliver the aggregate data to the BS. If the node's residual energy is less than the threshold value, it cannot be selected as CH. In this protocol, we have given an estimation of the threshold residual energy, which is the amount of energy required to receive, aggregate and transmit the average number of sensor nodes in the cluster.

d) It reduces energy consumption of CHs for data communication. It is achievable by keeping the distance between the communicating CHs and the BS short. To keep it short, if the distance between selected CH node and the BS is greater than some threshold distance, the CH will not communicate to the BS directly. In this case, multi-hop communication will occur via another CHs. As a result, it provides enhanced network lifetime in WSN.

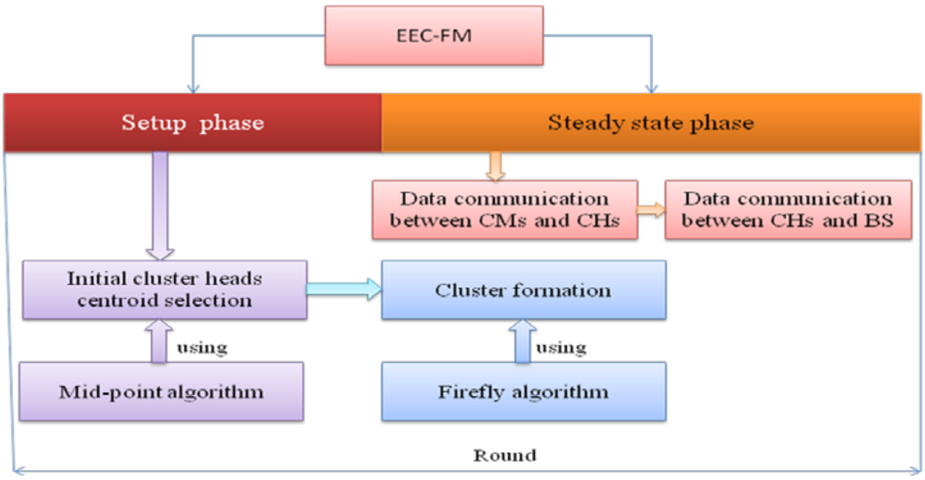


Fig. 2. Proposed EEC-FM protocol working model

The working process of the proposed protocol has divided into two sections, setup phase and steady state phase and it is divided into several rounds. The process of proposed work as follows:

Phase 1: Setup Phase

- Step 1: Initial CH selection and
- Step 2: Balanced cluster formation

Phase 2: Steady state Phase

- Step 3: Data communication from CMs to CHs and
- Step 4: Data Communication between CHs and BS

These two phases are described in Algorithm 1, Algorithms 2, and Algorithms 3 respectively.

Step 1: Initial CH selection

The midpoint algorithm [34] which has been used for initial CH selection assuming that the data points contain only positive values is described in Algorithm 1. Here the desired number of clusters k_{opt} is calculated from (10).

The process of initial cluster heads of a particular cluster of ten nodes are selected based on midpoint algorithm which is shown **Fig. 3**. Here the centroid of this cluster is a virtual node located at the centre position of the cluster. To maintain the connectivity of the network, residual energy of the CH is checked every cycle of operation. If the energy of the CH is less than the threshold energy, the node having the next ID number is selected as a new CH as in [31]. The newly elected CH then informs other nodes about the change of the CH.

Algorithm 1: Initial CH selection using midpoint algorithm

Input:

- N = set of number of sensor nodes.
- k_{opt} = optimum number of desired clusters

Output:

- k optimum (k_{opt}) number of initial centroids.

Process:

Step 1: The distance from each sensor nodes position to origin is computed.

Step 2: Sort the sensor nodes point in accordance with the distances obtained in step 1.

Step 3: Group and Partition the sorted data points of sensor nodes into k_{opt} equal sets.

Step 4: In each set, take the middle point sensor as the initial centroid.

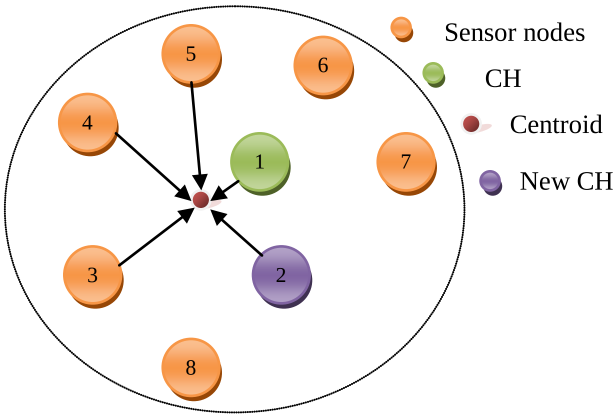


Fig. 3. Centroid of the CH selection in proposed EEC-FM protocol

Step 2: Balanced cluster formation

The balanced formation of the clusters are formed using firefly algorithm (FFA). The FFA is a recently developed swarm intelligence method that was inspired by the social behaviour of fireflies, in which the brighter firefly attracts other darker fireflies, as shown in **Fig. 4** [35]. In general, the FFA incorporates three strategies: attractiveness, distance between fireflies, and firefly movement.

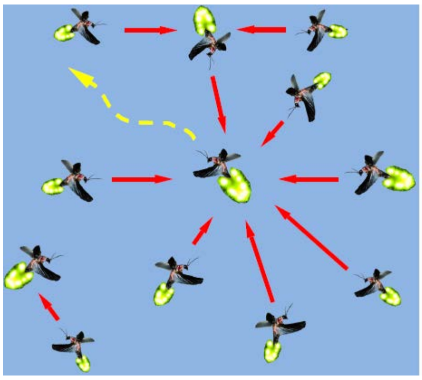


Fig. 4. Schematic of the firefly algorithm.

Algorithm 2: Pseudo-code of Firefly Algorithm (FFA)

Input:

- Create an initial population of fireflies n within d-dimensional search space
- x_{ik}, i = 1, 2, . . . , n and k = 1, 2, . . . , d.
- Evaluate the fitness of the population $f(x_{ik})$ which is directly proportional to light intensity I_{ik}.
- Algorithm's parameter—β_0, γ

Output:

- Obtained minimum location: x_i min

```
BEGIN
    repeat
        for i = 1 to n
            for j = 1 to n
                if (I_j < I_i)
                    Move firefly i toward j in d-dimension
                end if
                Attractiveness varies with distance r via exp[−r^2]
                Evaluate new solutions and update light intensity
            end for j
        end for i
        Rank the fireflies and find the current best until stop condition true
END
```

Proposed Cost Function for Cluster Formation

In proposed approach, cost is calculated with respect to the total spatial distance of clusters for cluster formations. Cost function [18, 21] is defined as follows:

$$F(x) = \sum_{j=1}^{k_{opt}} \sum_{s_i \in C_j} d(s_i, CH_j) + d(CH_j, BS) \tag{11}$$

Where, $d(s_i, CH_j)$ measures the Euclidian distance between the sensor node s_i and its corresponding CH_j in cluster C_j, k_{opt} is the number of clusters in the sensing region, and $d(CH_j, BS)$ measures the distance between CH_j and base station (BS). A sensor node can be assigned to the CH based on the minimum distance from that node to CH and CH to BS.

Algorithm 3 describes the balanced cluster formation phase. In the proposed approach, we have given an estimation of threshold residual energy which was not addressed in [31]. Here the threshold energy is the amount of energy required to receive, aggregate and transmit the average number of sensor nodes in the cluster. Therefore the threshold energy is given by

$$E_{threshold} = \left(\frac{N}{k_{opt}} - 1\right).E_{elec}.l + \frac{N}{k_{opt}} E_{DA}.l + E_{TX}(l,d) \quad (12)$$

where, N is total number of sensor nodes and k_{opt} is the optimum number of desired clusters.

Anothe important factor in Algorithm 3 is cluster formation. The cluster formation is done based on the centroid calculation of each cluster by using quetion (13) [36].

$$centroid(X,Y) = \left(\frac{1}{s}\sum_{i=1}^{s} x_i, \frac{1}{s}\sum_{i=1}^{s} y_i\right) \quad (13)$$

where, centroid (X, Y) is the average point between the coordinates locations for cluster heads, x_i is the summation of x-axis coordinates for all cluster heads, y_i is the summation of y-axis coordinates for all cluster heads, and S is the number of sensor nodes,

Algorithm 3: Balanced cluster formation using FFA

Input:
S = set of n data points of sensor nodes
K_{opt} = optimum number of desired clusters
$E_{threshold}$ = threshold energy

Output:
A set of effective structure of k_{opt} clusters

Process:
Step 1: Apply midpoint method presented in Algorithm 2 to choose k_{opt} out of S sensor nodes as initial cluster heads.
Step 2: Repeat
Step 3: Each of the remaining nodes decides to join its nearest CH according to the Euclidean distance.
Step 4: Centroid of each cluster is calculated using equetion (13)
Step 5: After cluster formation, based on the distance from the centroid, an ID is allotted to each of a cluster, assigning the smaller number to the closer one.
Step 6: for (All selected cluster heads)
    If (residual energy of the cluster head >= $E_{threshold}$)
    - the node will remain as cluster head

    else
    - check ID number of all the nodes in that cluster
    - the node in the next order of ID is selected as a new cluster head

    endif
  endfor
Step 7: the newly elected CHs inform other nodes about the cluster CHs changes
Step 8: until the cluster heads are not changed any more

Step 3:Data communication between cluster members (CMs) **and CHs**
Here, TDMA protocol is used to coomunicate between CMs and CH with a single hop distance.
Step 4:Data communication between CHs and BS
CHs follows single-hop or multi-hop communication depending on the distance from the BS from CHs. The distance between selected CH node and the BS is calculated using equation (14),

$$d_{threshold} = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{amp}}} = 87.7 \text{ mts} \quad (14)$$

If the distance between CH and BS (d_{BS})is less than threshold distance($d_{threshold}$) then cluster heads directly communicate to BS using single-hop. Otherwise, they follow multihop routing. The routing is based on the selection of the nearest neighbor cluster head whose d_{BS} is less then $d_{threshold}$ to communicate to BS.

5. Simulation Results and Analysis

The proposed EEC-FM protocol was simulated in NS2 and simulation parameters are summarised in **Table 1**. We have considered 100 sensor nodes in 200 × 200 m^2 network region. If we consider $d_{CH\ to\ BS}$ = 100, we get the number of desired CH = 4 and considering $d_{CH\ to\ BS}$ = 85, we get the number of desired CH = 5. WSN simulation was performed on both 4-cluster and 5-cluster networks. The proposed EEC-FM protocol has been compared with LEACH-B [28] and other K-means based approaches [29-33] used in WSN with respect to cluster formation. Existing approaches [28-33] were also compared with respect to different network parameters.

5.1 Comparison of four cluster structure of proposed EEC-MP protocol with EECPK-means protocol and Park's approach

Fig. 5 compares the cluster formation of the 4-cluster sensor network using Park's approach [31] and EECPK-means [33] with EEC-MP protocol.

It has been observed that in Park's approach to form 4 clusters 100 sensor nodes a large variation in the cluster formation occurs . Out of eight observations it is found that in 4th observation, cluster 2 contains 39 nodes, which is much higher than the average number of nodes (25). However, in 5th observation, cluster 4 contains only 11 nodes, which is lower than the average number of sensor nodes. As a result, CH of the heavily loaded cluster, which contains 39 nodes will be exhausted much earlier than the other clusters as shown **Fig. 5a**.

The second EECPK-means protocol, where the midpoint algorithm was used for initial CH selection it is found that a particular cluster contains maximum 29 nodes and minimum 21 nodes which is also relatively closer to the average number of nodes to be present (25) in a particular cluster. Here the resultant clusters have less uneven number of sensor nodes, which ultimately leads to unbalanced clusters which is shown in **Fig. 5b**.

Table 1. Simulation parameters

Parameter	Value
number of sensor nodes (N)	100
network size	200 x 200 m^2
base station's location	(0,0)
number of clusters (k_{opt})	4, 5
initial energy of node	1 J
data packet	3200 bits
E_{elec}	50 nJ/bit
ε_{mp}	0.0013 pJ/bit/m^4
ε_{fs}	10 pJ/bit/m^2
energy for data aggregation (EDA)	5 nJ/bit/signal
$d_{CH\ to\ BS}$	85–100 m
d_{nCH}	dBS/2
$d_{threshold}$	87.7 mts

By applying the proposed EEC-FM protocol, where midpoint algorithm is used for initial CH selection and FFA is used for cluster formation, it is found that a particular cluster contains maximum 27 nodes and minimum 23 nodes, which is very close to the average number of nodes to be present (25) in a particular cluster. Hence the resultant clusters have almost equal number of sensor nodes, which ultimately leads to balanced cluster and results are plotted in **Fig. 5c**. Overall simulation results of four clustered networks are summarised in **Table 2**

Table 2. Number of nodes present in four clusters using Park's approach, EECPK-means protocol and proposed EEC-FM protocol

Number of observations	Number of nodes present in 4 clusters											
	Park's approach				EECPK-means protocol				Proposed EEC-FM protocol			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
1	24	18	38	20	25	25	23	27	25	26	25	24
2	15	22	35	28	26	25	24	25	26	25	24	25
3	31	30	26	13	28	26	23	23	27	26	24	23
4	19	39	24	18	23	28	24	25	26	26	24	24
5	29	27	33	11	26	24	28	22	26	24	27	23
6	30	20	18	32	22	26	27	25	24	24	27	25
7	33	17	22	28	23	27	21	29	25	26	23	26
8	29	24	25	22	26	26	24	24	25	27	24	24

5.2 Comparison of five cluster structure of proposed EEC-MP protocol with EECPK-means protocol and Park's approach

Considering $d_{CH\text{-}BS}$ = 85, we get the number of desired CH = 5 out of 100 sensor nodes in 200 × 200 m^2 sensing region. This time also we get unbalanced cluster formation like 4-clustered network. It has been observed that in Park's approach amongst the eight observations, it is found that a particular cluster contains 33 nodes in cluster 4, which is much higher than the average number of sensor nodes (20) to be present in a cluster. At the same time, cluster 5 contains only 11 sensor nodes, which is much lower than the average number of sensor nodes. The detals are shown in **Fig. 6a**

Using EECPK-means protocol, amongst the eight observations a particular cluster contains maximum 24 nodes and minimum 16 nodes, which is just higher to the average number of nodes to be present (20) in a particular cluster showed in **Fig. 6b**. Therefore this approach is also produces unbalanced clusters comparatively better than Park's approach.

Using proposed EEC-FM protocol as shown in **Fig. 6c**, we find amongst the eight observations a particular cluster contains maximum 22 nodes and minimum 18 nodes, which is much closer to the average number of nodes to be present (20) in a particular cluster. Therefore our proposed approach produces balanced cluster compared to Park's approach and EECPK-means. Simulation results of all the above three approaches are summarised in **Table 3**.

Table 3. Number of nodes present in Five clusters using Park's approach, EECPK-means protocol and proposed EEC-FM protocol

Number of observations	Number of nodes present in 5 clusters														
	Park's approach [22]					EECPK-means protocol					Proposed EEC-FM protocol				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	13	18	33	25	11	21	19	24	18	18	20	20	22	19	19
2	23	24	15	19	19	18	21	22	18	21	19	20	21	19	21
3	18	26	26	18	12	20	19	21	20	20	19	20	21	20	20
4	17	32	16	17	18	19	21	19	21	20	19	20	20	21	20
5	10	30	24	14	22	24	20	23	16	17	22	21	20	19	18
6	28	12	16	23	21	21	19	16	22	22	22	20	18	20	22
7	15	25	17	18	25	24	17	18	20	21	22	19	18	20	21
8	25	15	25	15	20	22	20	19	18	21	21	21	19	18	21

5.3 Measure of scattering of the number of nodes present in 4-cluster and 5-cluster networks using EECPK-means, Park's approach and proposed protocol with respect to the standard deviation (σ)

The deviation in cluster formation using Park's approach, EECPK-means and our proposed approach from an ideal one can be tested using the parameter standard deviation using equation (15).

$$\sigma = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2} \tag{15}$$

The number of nodes present in 4-cluster and 5-cluster using Park's approach, EECPK-means and EEC-FM protocol are summarised in **Table 2** and **Table 3** respectively. Considering 100 sensor nodes deployed in WSN, for 4-cluster network the value of average x = 25 and for 5-cluster network the value of average x = 20. **Table 4** shows the measure of dispersion of the number of nodes present in different clusters using Park's approach, EECPK-means and proposed approach with respect to standard deviation.

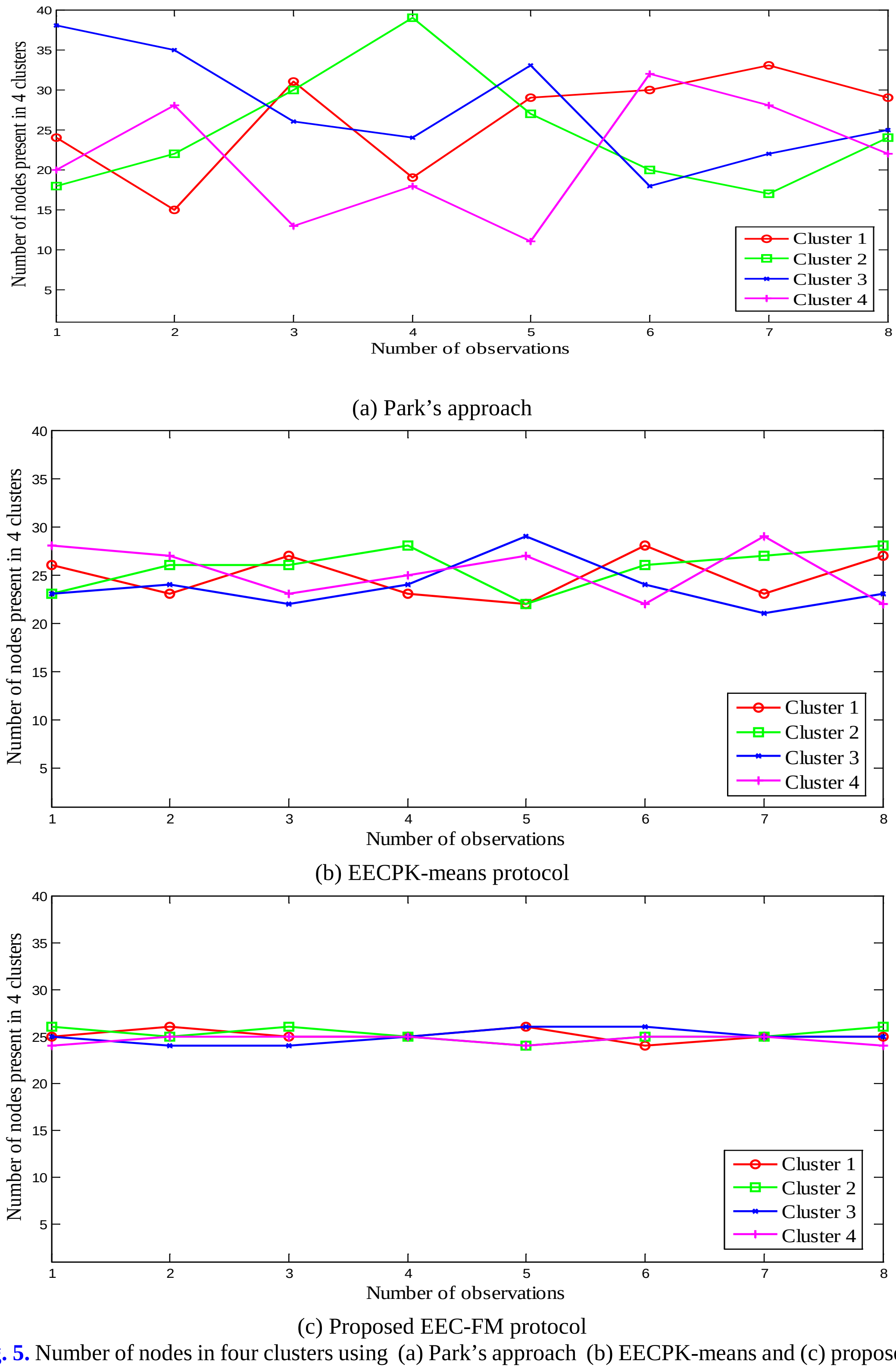
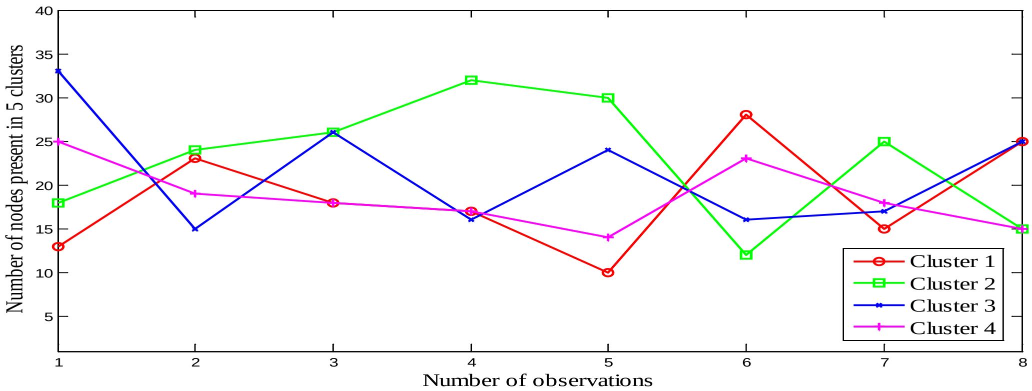
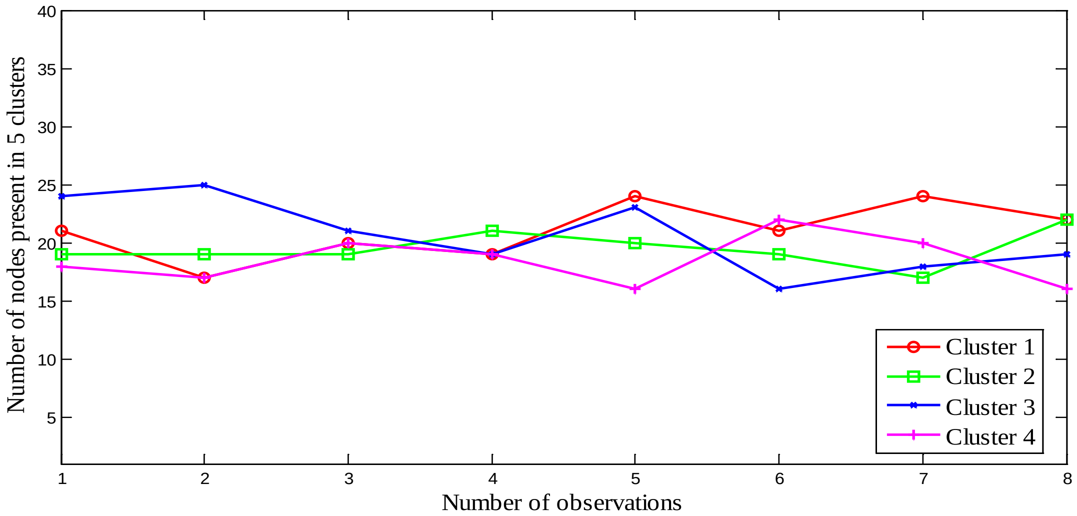


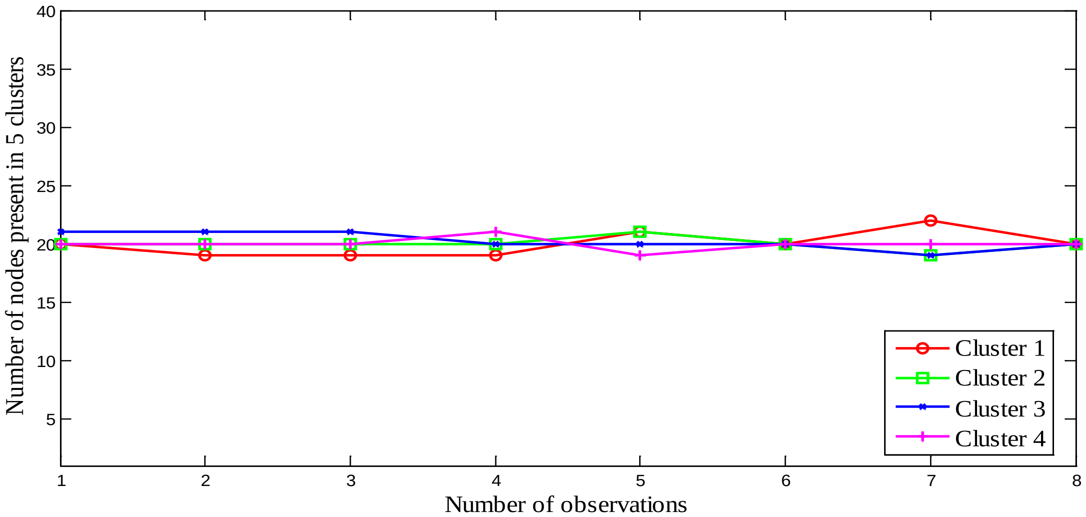
Fig. 5. Number of nodes in four clusters using (a) Park's approach (b) EECPK-means and (c) proposed EEC-FM protocol.



(a) Park's approach



(b) EECPK-means Protocol



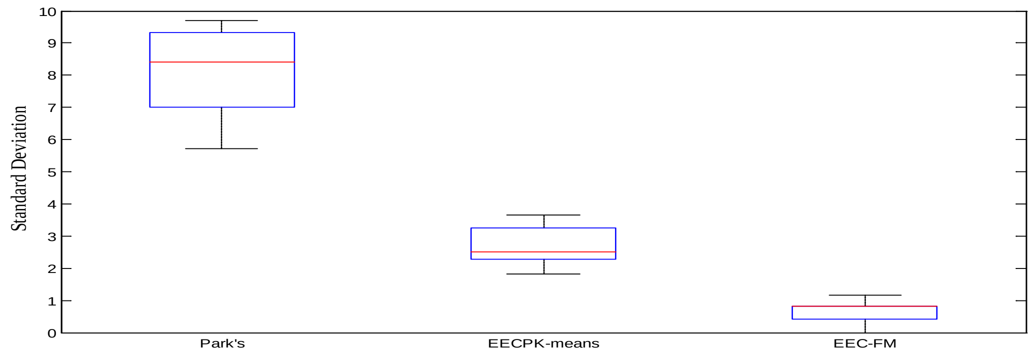
(c) Proposed EEC-FM protocol

Fig. 6. Number of nodes in five clusters using (a) Park's approach (b) EECPK-means and (c) proposed EEC-FM protocol

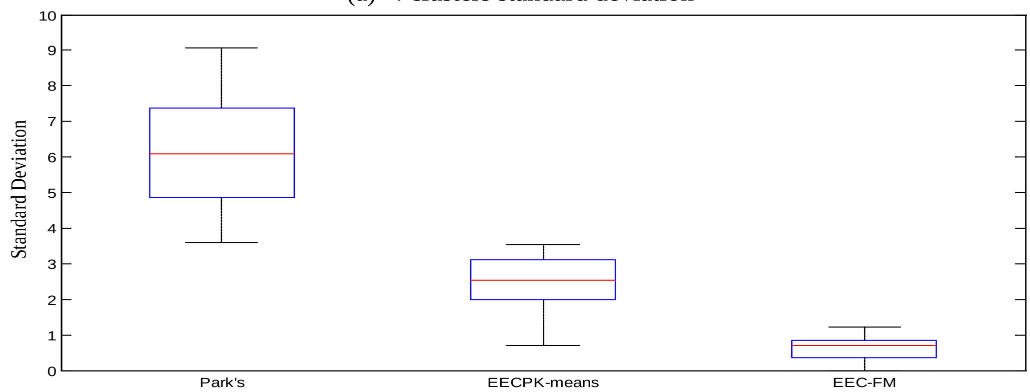
From **Fig. 7**, it is clear that compared to our proposed EEC-FM protocol, the dispersion in Park's approach is much higher and EECPK-means protocol's dispersion relatively higher. So for both 4-cluster and 5-cluster network it is evident that the proposed approach performs better compared to the other two approaches in providing balanced clusters. As a result the proposed approach ultimately balances the load of CHs and prolongs network lifetime.

Table 4. Comparison of dispersion of the number of nodes present in different clusters with respect to the parameter standard deviation

Observations	Standard Deviation (σ)					
	Park's approach		EECPK-means		EEC-FM	
	4 Clusters	5 Clusters	4 Clusters	5 Clusters	4 Clusters	5 Clusters
1	9.02	9.06	1.63	2.55	0.82	1.22
2	8.52	3.61	0.82	1.87	0.82	1.00
3	8.29	6.00	2.45	0.71	1.83	0.71
4	9.70	6.75	2.16	1.00	1.15	0.71
5	9.66	8.00	2.58	3.54	1.83	1.58
6	7.02	6.20	2.16	2.55	1.41	1.67
7	6.98	4.69	3.65	2.74	1.41	1.58
8	5.72	5.00	1.15	1.58	1.41	1.41



(a) 4 clusters standard deviation

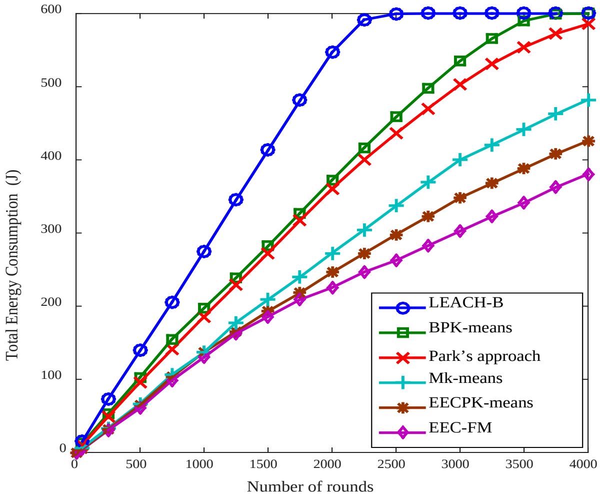


(b) 5 clusters standard deviation

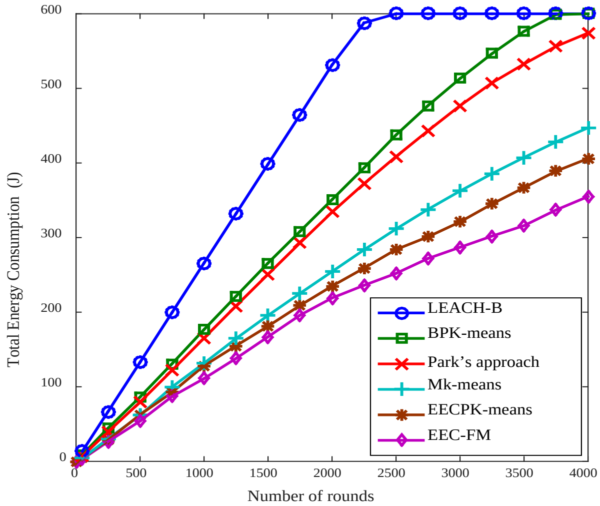
Fig.7. Comparison of dispersion of the number of nodes present in different clusters with respect to the parameter standard deviation

5.4 Comparison of proposed EEC-FM protocol with EECPK-means, LEACH-B, BPK-means, Park's approach and Mk-means with respect to energy efficiency

Energy consumption is defined as the amount of energy consumption of the network over a number of rounds, in each of which, CHs collect data, aggregate and route it to the BS. We have extensively tested performance of energy consumption of proposed protocol over various scenarios with varying number of cluster heads. The comparision of energy consumption of proposed EEC_FM protocols with of standard protocols like LEACH-B [28], BPK-means [30], Park's approach [31], Mk-means [32], and EECPK-means protocol [33] shown in **Fig. 8**. It can be observed that the proposed protocol outperforms with existing protocols in terms of total energy consumption.



(a) Energy consumption for four clusters network



(b) Energy consumption for five clusters network

Fig. 8. Energy consumption of four and five clustered networks

From **Table 5**, it is observed that our proposed EEC-FM protocol is 45% better than LEACH-B, 17.8% better than BPK-means protocol , 12.5% better than Park's approach, 9.1%

better than Mk-means, and 5.8% better than EECPK-means protocol with respect to the parameter half energy consumption (HEC).

Table 5. Number of rounds with respect to HEC

Clustering algorithm	Number of rounds with respect to HEC
LEACH-B [28]	1150
BPK-means [30]	1720
Park's approach [31]	1830
Mk-means [32]	1900
EECPK-means [33]	1970
Proposed EEC-FM	2090

5.5 Comparison of proposed EEC-FM protocol with EECPK-means, LEACH-B, BPK-means, Park's approach and Mk-means with respect to network lifetime

The life time of the network can be defined as the as death of first node (FND), half dead node (HND) and last dead nodes with respect to number of rouds. As the lifetime of the network increases the better is the performance of the network. The number of nodes which are alive using LEACH-B [28], BPK-means [30], Park's approach [31], Mk-means [32], EECPK-means protocol [33] and our proposed EEC-FM protocol are compared with respect to the number of rounds as shown in **Fig. 9**. NS2 simmulation are clearly shows that EEC-FM protocol provides better network lifetime compared to the above mentioned algorithms with respect to the parameters FND, HND and LND summarised in **Table 6** and helps to provide enhanced network lifetime to a great extent.

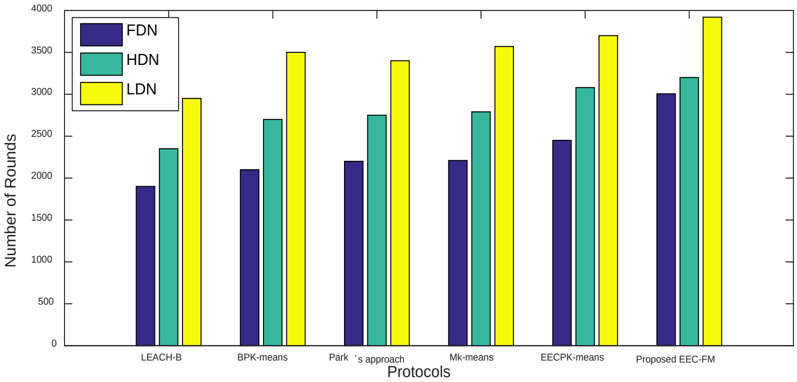


Fig. 9. Performance of network lifetime with respect to FDN, HND and LND

Table 6. Comparison of network lifetime with respect to FDN, HND and LND

Clustering Algorithm	Round first node dies (FND)	Round half node dies (HND)	Round last node dies (LND)
LEACH-B [7]	1900	2350	2950
BPK-means [21]	2100	2700	3500
Park's approach [22]	2200	2750	3400

Mk-means [23]	2210	2790	3570
EECPK-means [32]	2450	3080	3700
Proposed EEC-FM	3005	3200	3920

5.6 Comparison of proposed EEC-FM protocol with EECPK-means, LEACH-B, BPK-means, Park's approach and Mk-means with respect to packets receipt by BS

Packets received by BS is defined as the total number of packets received by the base station (BS) over the span of network lifetime. **Fig. 10** shows the performance analysis of the packets received by base station and the proposed protocol EEC-FM outperforms with existiong protocols. The proposed protocol receipt more number of data packets because it consumes less energy and has more network lifetime. The proposed protocol is decline more number of packets received by BS coparitively with the existing protocols due to the fact that the proposed protocol takes care of proper selection of the CHs, proper formation of cluster and proper routing among the CHs to BS taken care with the help of efficient cost function.

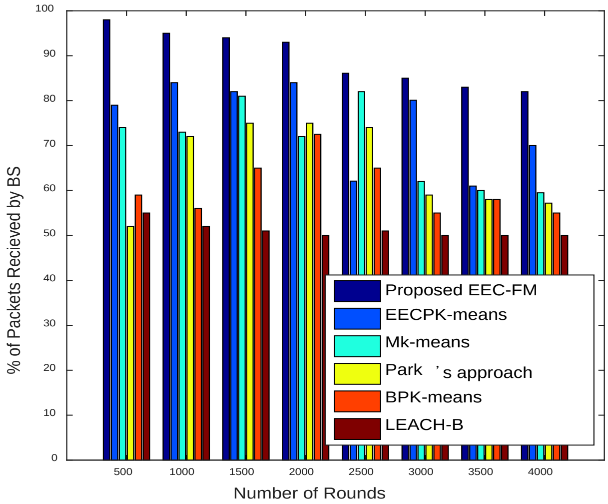


Fig. 10. The performance analysis of the packets received by BS

6. Conclusion

Clustering is one of the solutions for aggregation of data which reduces energy consumption and extend network life time. Though K-means is a commonly used clustering algorithm in various fields including WSN but it is not capable of getting best results due to its random initial centroids selection. The proposed EEC-FM protocol provides better optimal results than K-means related algorithms as it employs midpoint method for initial centroids selection and firefly algorithm used for cluster formation. In this proposed EEC-FM protocol the a new cost function is formulated by considering intra and inter communication distance as a parameters for cluster formation. It also optimises CH selection method by considering residual energy as the parameters of CH selection in addition to Euclidean distance. Multi-hop communication is provided between CHs and BS so that the CHs which are far away from the sensing region does not exhaust much energy and can successfully deliver the aggregate data to the BS. We have shown simulation results along with their comparisons lik structure of four and five clustered network, Measure of scattering of the number of nodes present in 4-cluster and 5-cluster networks using the standard deviation (σ), energy efficiency of four and five clustered networks, network lifetime, and packets received by by BS with various existing protocols, namely, LEACH-B, BPK-means, Park's approach, Mk-means, and EECPK-means protocol. Simulation results demonstrated that the proposed EEC-FM protocol is 45% better than LEACH-B, 17.8% better than BPK-means protocol, 12.5% better than Park's approach, 9.1% better than Mk-means, and 5.8% better than EECPK-means protocol with respect to the parameter half energy consumption (HEC). Future work relates to more hybridisation of protocols based on optimisation techniques for solving clustering and routing problems.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, 393-422, 2002. Article (CrossRef Link)

[2] Yick, J., Mukherjee, B., Ghosal, D., "Wireless sensor network survey," *Comput. Netw.*, 52, (12), pp. 2292–2330, 2008. Article (CrossRef Link)

[3] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. of Proceedings of the ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, 2002. Article (CrossRef Link)

[4] D. Steere, A. Baptista, D. McNamee, C. Pu, and J. Walpole, "Research challenges in environmental observation and forecasting systems," in *Proc. of Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2000. Article (CrossRef Link)

[5] Pagano, S., Peirani, S., Valle, M., "Indoor ranging and localisation algorithm based on received signal strength indicator using statistic parameters for wireless sensor networks," *IET Wirel. Sens. Syst.*, 5, (5), pp. 243–249, 2015. Article (CrossRef Link)

[6] Bhatti, S., Xu, J., Memon, M., "Clustering and fault tolerance for target tracking using wireless sensor networks," *IET Wirel. Sens. Syst.*, 1, (2), pp. 66–73, 2011. Article (CrossRef Link)

[7] Li, W., Zhang, W.: 'Sensor selection for improving accuracy of target localisation in wireless visual sensor networks', IET Wirel. Sens. Syst., 2012, 2, (4), pp. 293–301. Article (CrossRef Link)

[8] S. Intille, "Designing a home of the future", IEEE Pervasive Computing, 1(2):76–82, April 2002. Article (CrossRef Link)

[9] C. Kidd et al. "The aware home: A living laboratory for ubiquitous computing research," in *Proc. of Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild)*, 1999. Article (CrossRef Link)

[10] D. W. Kumar, "Healthcare Monitoring System Using Wireless Sensor Network," *Intr. Journal of Advanced Networking and Applications*, vol. 4, no. 1, pp. 1497-1500, 2012. Article (CrossRef Link)

[11] S. Mukherjee, K. Dolui, S. K. Datta, "Patient health management system using e-health monitoring architecture," in *Proc. of IEEE International Conference on Advance Computing(IACC)*, pp. 400-405, 2014. Article (CrossRef Link)

[12] M. Yamaji, Y. Ishii, T. Shimamura, and S. Yamamoto, "Wireless Sensor Networks for Industrial Automation," in *Proc. of 3rd International Conference on Networked Sensing System*, 2006.

[13] Vehbi C. Gungor and Gerhard P. Hancke, Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches, IEEE Transactions On Industrial Electronics, VOL. 56, NO. 10, 2009.

[14] Abbasi, A.A., Younis, M., "A survey on clustering algorithms for wireless sensor networks," Comput. Commun., 30, (14), pp. 2826–2841, 2007. Article (CrossRef Link)

[15] Heinzelman,W.B., Chandrakasan, A.P., Balakrishnan, H., "An application-specific protocol architecture for wireless micro-sensor networks," *IEEE Trans. Wirel. Commun.*, 1, (4), pp. 660–670, 2002. Article (CrossRef Link)

[16] Ray, A., De, D., "Energy efficient clustering hierarchy protocol for wireless sensor network," in *Proc. of IEEE Int. Conf. on Communication and Industrial Application(ICCIA)*, pp. 1–4, December 2011. Article (CrossRef Link)

[17] Ray, A., De, D., "Energy efficient cluster head selection in wireless sensor network," in *Proc. of IEEE Int. Conf. on Recent Advances in Information Technology (RAIT)-2012, ISM*, Dhanbad, Jharkhand, pp. 306–311, March 2012. Article (CrossRef Link)

[18] Ray, A., De, D., "Energy efficient clustering algorithm for multi-hop green wireless sensor network using gateway node," *Adv. Sci. Eng. Med.*, 5, (11), pp. 1199–1204, 2013. Article (CrossRef Link)

[19] Ray, A., De, D.: 'Level wise initial energy assignment in wireless sensor network for better network lifetime', *Proc. Adv. Comput. Netw. Inf.*, 2014, 2, pp. 67–74. Article (CrossRef Link)

[20] Guo, P., Jiang, T., Zhang, K., et al., "Clustering algorithm in initialization of multi-hop wireless sensor networks," *IEEE Trans. Wirel. Commun.*, 8, (12), pp. 5713–5717, 2009. Article (CrossRef Link)

[21] Kumar, R., Malik, A., Kumar, B., "NEECP: a novel energy efficient clustering protocol for prolonging lifetime of WSNs," *IET Wirel. Sens. Syst.*, 2016

[22] Cao, F., Liang, J., Jiang, G., "An initialization method for the K-means algorithm using neighbourhood model," *Comput. Math*, 58, pp. 474–483, Appl, 2009. Article (CrossRef Link)

[23] Khan Shehroz, S., Amir, A., "Cluster centre initialization algorithm for K-means clustering," *Pattern Recognit. Lett*., 25, (11), pp. 1293–1302, 2004. Article (CrossRef Link)

[24] Sasikumar, P., Khara, S., "K-means clustering in wireless sensor networks," in *Proc. of IEEE Fourth Int. Conf. of Computational Intelligence and Communication Networks (CICN)*, pp. 140–144, November 2012. Article (CrossRef Link)

[25] Hansen, P., Ngai, E., Cheung, B.K., et al., "Analysis of global K-means, an incremental heuristic for minimum sum-of-squares clustering," *J. Classif.*, 22, (2), pp. 287–310, 2005. Article (CrossRef Link)

[26] Napoleon, D., Ganga Lakshmi, P., "An enhanced K-means algorithm to improve the efficiency using normal distribution data points," *Int. J. Comput. Sci. Eng.*, 2, (7), pp. 2409–2413, 2010.

[27] Niknam, T., Amiri, B., "An efficient hybrid approach based on PSO, ACO and K-means for cluster analysis," *Appl. Soft Comput.*, 10, (1), pp. 183–197, 2010. Article (CrossRef Link)

[28] Tong, M., Tang, M., "LEACH-B: an improved LEACH protocol for wireless sensor network," in *Proc. of Int. Conf. Wireless Communications Networking and Mobile Computing(WiCOM)*, pp. 1–4, September 2010. Article (CrossRef Link)

[29] Khan, A., Tamim, I., Ahmed, E., et al., "Multiple parameter based clustering (MPC): prospective analysis for effective clustering in wireless sensor network (WSN) using K-means algorithm," *Wirel. Sens. Netw*., 4, (1), pp. 18–24, 2012. Article (CrossRef Link)

[30] Tan, L., Gong, Y., Chen, G., "A balanced parallel clustering protocol for wireless sensor networks using K-means techniques," in *Proc. of IEEE Second Int. Conf. On Sensor Technologies and Applications*, pp. 300–305, August 2008. Article (CrossRef Link)

[31] Park, G.Y., Kim, H., Jeong, H.W., et al., "A novel cluster head selection method based on K-means algorithm for energy efficient wireless sensor network," in *Proc. of IEEE 27th Int. Conf. on Advanced Information Networking and Applications Workshops*, 2013, pp. 910–915

[32] Periyasamy, S., Khara, S., Thangavelu, S., "Balanced cluster head selection based on modified k-means in a distributed wireless sensor network," *Int. J. Distrib. Sens. Netw.*, 2016, pp. 1–11, Article ID 5040475, 2016.

[33] Anindita Ray, Debashis De, "Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wirel. Sens. Syst*., pp. 1–11, 2016.

[34] Aggarwal, N., Aggarwal, K.A., "A mid–point based k-mean clustering algorithm for data mining," *Int. J. Comput. Sci. Eng*., 4, (6), pp. 1174–1180, 2012.

[35] Yang XS, "Multiobjective firefly algorithm for continuous optimization. Engineering with Computers," 29(2),175–184, 2013. Article (CrossRef Link)

[36] M. A. Mizher, Saleh H. Al-Sharaeh, Mei Choo Ang, Ayman M. Abdalla, Manal A. Mizher, “Centroid dynamic sink location for clustered wireless mobile sensor networks,” *Journal of Theoretical and Applied Information Technology*, Vol.73 No.3, 2015.

Ravuri Daniel, research scholar in the Department of Computer Science and Engineering, JNTUK-Kakinada, Indian. His interest of research includes wireless sensor networks, computer networks, IoT and nature inspired optimization techniques.

Dr. Kuda Nageswara Rao, Professor, department of Computer Science & Systems Engineering, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India. His research interests includes Computer Networks, IoT, TCP/IP, Internet Technologies, Wireless and Wireless Sensor Networks. Life member in ISPS and ISTE, member in IETE, ISCA, CSI etc.,