IET Wireless Sensor Systems

Research Article

ISSN 2043-6386
Received on 8th June 2015
Revised on 24th April 2016
Accepted on 27th May 2016
doi: 10.1049/iet-wss.2015.0087
www.ietdl.org

Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network

Anindita Ray[1,2], Debashis De[2] ✉

[1]*Department of Computer Application, B.P. Poddar Institute of Management and Technology, Poddar Vihar, 137 V.I.P Road, Kolkata-52, West Bengal, India*
[2]*Department of Computer Science and Engineering, West Bengal University of Technology, BF-142, Sector 1, Salt Lake City, Kolkata 700064, West Bengal, India*
✉ *E-mail: dr.debashis.de@gmail.com*

Abstract: Wireless sensor networks (WSNs) consist of an enormous number of tiny sensor nodes deployed in huge numbers which are able to sense, process and transmit environmental information to the base station (BS) for a variety of applications. Energy efficiency is one of the primary concerns for maintaining WSN in operation. In this study, an energy efficient clustering protocol based on K-means algorithm named EECPK-means has been proposed for WSN where midpoint algorithm is used to improve initial centroid selection procedure. The proposed approach produces balanced clusters to ultimately balance the load of cluster heads (CHs) and prolong the network lifetime. It considers residual energy as the parameter in addition to Euclidean distance used in basic K-means algorithm for appropriate CH selection. Multi-hop communication from CH nodes to BS takes place depending on their distances from BS. Simulation result shows that the proposed approach outperforms LEACH-B, balanced parallel K-means (BPK-means), Park's approach and Mk-means with respect to network lifetime and energy efficiency. Simulation result also demonstrates that the proposed approach can reduce the energy consumption at most 50% compared to LEACH-B, 14% compared to BPK-means protocol, 10% compared to Park's approach and 6% compared to Mk-means.

1 Introduction

In recent days, wireless sensor network (WSN) has gained popularity due to its potential use in a variety of applications such as target tracking [1, 2], localisation [3], environment monitoring, healthcare and industrial automation [4]. Clustering is an effective mechanism of data aggregation in WSN [5]. Efficient clustering [6–14] and relay node placement [15] can achieve high energy utilisation and extends network lifetime of WSN. One of the most important problems in clustering is to improve cluster structure and optimise the selection of cluster heads (CHs). K-means algorithm is very useful in producing clusters for many practical applications including WSN. Several approaches are there based on K-means algorithm for better clustering [16–28]. However, K-means algorithm suffers from some limitations which are as follows [16–29]:

(i) In K-means algorithm, the initial centroids are chosen just randomly out of the input data set. During each run the algorithm results in different kinds of clusters according to the different randomly chosen initial centroids. Thus random selection of initial centroids leads the algorithm into local optima.
(ii) In worst case K-means algorithm may produce an empty cluster due to the random selection of initial centroids.
(iii) Applying K-means algorithm each time the end clustering result will come out to be different. Then analysis is required to find out the most appropriate result. Thus there is no guarantee that the K-means algorithm will converge into best results.
(iv) The number of desired clusters is not calculated thoroughly and is therefore required to be inserted as a user input to the algorithm.
(v) Moreover, the time complexity of traditional K-means clustering algorithm is very high because the data points are reassigned as number of times as the iterations of the loop run.

Due to random selection of initial centroids K-means algorithm produces unbalanced cluster as shown in Fig. 1. Here cluster 4 contains only three sensor nodes. However, all the other three clusters contain sensor nodes much more than it should contain. As a result, these three CHs become overloaded and get exhausted earlier. Our proposed EECPK-means protocol improves the initial centroid selection procedure of K-means using midpoint algorithm [17] which produces balanced cluster compared to K-means. It also optimises CH selection method by considering residual energy as a parameter of CH selection in addition to Euclidean distance used in basic K-means algorithm.

The remaining paper is organised as follows. Section 2 presents the related work. Section 3 illustrates the motivation of our proposed work. Section 4 describes the basic K-means clustering algorithm. Section 5 presents our proposed EECPK-means protocol. Section 6 contains simulation result and analysis in details. Finally, Section 7 concludes our paper.

2 Related work

The main aim in WSN is to achieve energy efficient network as the energy of each sensor node is limited. Relay node placement [15], tree-based data aggregation [30] and efficient sensor movement [31] can optimise energy consumption of sensor network. Clustering the entire network in an efficient way is one of the major goals in WSN since it not only can reduce energy consumption, but also can provide uniform coverage to a greater extent [4, 5]. K-means clustering scheme is widely used in many applications including the field of WSN for better clustering [16–29]. Traditional K-means algorithm suffers from several limitations as discussed in Section 1. To deal with the limitation of K-means algorithm, several approaches are there to compute initial cluster centres [17, 18] and calculate the accurate number of clusters [19]

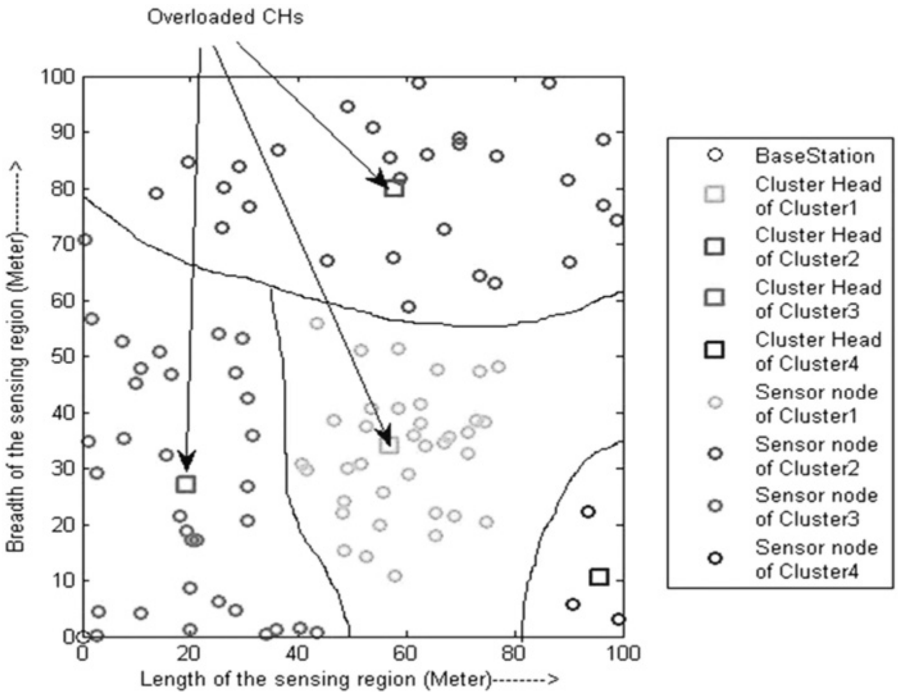

Fig. 1 *Unbalanced cluster using K-means algorithm*

which improves K-means algorithm. Some of the existing approaches for clustering based on K-means algorithm are discussed as follows:

(i) Performance of K-means algorithm is to some extent dependent on the initial cluster centres. In [18] an algorithm has been proposed to compute initial cluster centres. It has been found that, using this algorithm the computed cluster centres are very close to the desired cluster centres. However, this approach is unable to produce balanced cluster.

(ii) In [19] the authors propose an efficient K-means algorithm which performs accurate clustering without pre-assigning the exact number of clusters. However, initial centroids are selected randomly. Therefore the algorithm results in different kinds of clusters in different runs based on the randomly chosen initial centroids.

(iii) Multi parameter-based clustering [20] using K-means provides effective clustering in WSN. It considers different parameters in cluster formation. By varying different parameters, it can produce cluster with uniformly distributed nodes, optimise intra- and inter-cluster distances and is capable of selecting less power consuming CHs. However, the number of desired clusters is not calculated thoroughly and is therefore required to be inserted as a user input to the algorithm.

(iv) In [21] the authors have proposed a balanced parallel K-means (BPK-means)-based clustering protocol for WSN. In this approach, K-means algorithm is used to cluster the sensor nodes and CHs are chosen depending on their distance from cluster centre and residual energy. It balances clusters so that the energy consumption due to intra-cluster communication reduces. However, clustering does not consider threshold distance between the CH and the BS. Moreover, using K-means it does not consider the energy of sensor nodes for determining the centroids. Therefore still a significant amount of energy depletion happens for CH nodes which ultimately has an impact on the network lifetime.

(v) In [22] the authors have proposed an energy efficient CH selection technique based on K-means clustering algorithm for WSN. Here initial centroids are selected randomly. Therefore the algorithm results in different kinds of clusters in different runs based on the randomly chosen initial centroids. In this approach, though they have considered residual energy of sensor nodes as a parameter of CH selection, but they have not specified any proper estimation regarding this. Moreover, clustering does not consider threshold distance between the CH and the BS. It follows a single hop communication between the CH and the BS.

(vi) To enhance the network lifetime, a balanced CH selection strategy is proposed using modified K-means [23]. In this approach, more than one CH in a cluster is considered to reduce the time and energy required for re-clustering.

(vii) In [24] the authors propose a hybrid algorithm based on the combination of ant colony optimisation, fuzzy adaptive particle swarm optimisation and K-means algorithm which provides better clustering.

3 Motivation and contributions of this paper

As discussed in Section 2, 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 EECPK-means to resolve these issues. The main contributions of our proposed approach are as follows:

(i) 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 $M \times M$ square sensing region. The optimum number of clusters k_{opt} can be obtained as follows [7]

$$k_{\text{opt}} = \frac{\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{\text{fs}}}{\varepsilon_{\text{mp}}}} \frac{M}{d_{\text{BS}}^2} \tag{1}$$

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

(ii) Midpoint method described in Algorithm 2 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.

(iii) Our proposed approach considers residual energy of sensor nodes as the parameters of CH selection in addition to the

Euclidean distance used in basic K-means algorithm, 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 our approach, 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.

(iv) 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.

4 K-means clustering algorithm

The basic idea of K-means clustering algorithm is to classify a given set of data items into k number of disjoint clusters where the value of k is predefined. Algorithm 1 [18–20, 22] describes the basic K-means algorithm which mainly consists of two phases as follows.

Algorithm 1: The K-means clustering algorithm

Input:
D = set of n data items
k = number of desired clusters
Output:
A set of k clusters.
Steps:
1: Randomly select k data-items from D as initial centroids;
2: Repeat;
3: Allot each data item to its closest centroid;
4: Compute new mean for each cluster;
5: Until convergence criteria is met.

---

5 Proposed approach

5.1 Network model

We have the following assumptions regarding the network model:

(i) The sensor nodes and BS are all static after deployment.
(ii) There is only one BS far from the sensing region.
(iii) Sensors are homogeneous having the same initial energy.
(iv) After deployment BS knows the geographical information of all sensor nodes.
(v) Data aggregation happens in each of the CH. The CH ultimately sends aggregated data to the BS.
(vi) CHs follow single-hop or multi-hop communications depending on the distance from the BS.

5.2 Radio energy dissipation model

In order to anticipate the performance of our proposed approach, radio hardware energy dissipation model is used as in [8]. Therefore, to convey a l-bit message to a distance d, the radio spends energy as follows

$$\begin{aligned} E_{\text{Tx}}(l, d) &= E_{\text{Tx-elec}}(l) + E_{\text{Tx-amp}}(l, d) \\ E_{\text{Tx}}(l, d) &= E_{\text{elec}} \times l + \varepsilon_{\text{fs}} \times l \times d^2 \quad \text{If } d < d_0 \end{aligned} \tag{2}$$

$$E_{\text{Tx}}(l, d) = E_{\text{elec}} \times l + \varepsilon_{\text{mp}} \times l \times d^4 \quad \text{If } d \geq d_0 \tag{3}$$

where the threshold

$$d_0 = \sqrt{\frac{\varepsilon_{\text{fs}}}{\varepsilon_{\text{mp}}}} \tag{4}$$

To receive this message, the radio expends

$$E_{\text{Rx}}(l) = E_{\text{Rx-elec}}(l) = E_{\text{elec}} \times l \tag{5}$$

5.3 Proposed EECPK-means protocol

The working principle of our proposed EECPK-means protocol is divided into three phases as follows:

Phase 1: Initial CH selection.
Phase 2: Balanced cluster formation.
Phase 3: Data communication.

These three phases are described in Algorithm 2, Algorithms 3 and 4 (see Figs. 2 and 3), respectively.

Algorithm 2: Midpoint algorithm for initial CH selection

Input:
D = set of n data points.
k_{opt} = optimum number of desired clusters
Output:
k_{opt} number of initial centroids.
Steps:
1: Compute the distance from origin for each data point.
2: Sort the distances obtained in step 1. Sort the data points in accordance with the distances.
3: Partition the sorted data points into k_{opt} equal sets.
4: In each set, take the middle point as the initial centroid.

Algorithm 2 describes the midpoint algorithm [17] which has been used for initial CH selection assuming that the data points contain only positive values. Here the desired number of clusters k_{opt} is obtained from (1). Fig. 4 shows an example of a particular cluster of ten nodes where initial CHs have been selected through the midpoint algorithm. Here the centroid of a cluster is a virtual node locating at the centre position of the cluster. In this figure, initial CH is denoted by encircled sensor node. To maintain the connectivity of the network, residual energy of the CH is checked every round. If the energy of the CH is smaller than the threshold energy, the node having the next ID number is selected as a new CH as in [22]. The newly elected CH informs other nodes about the change of the CH.

Algorithm 3 (Fig. 2) describes the balanced cluster formation phase. In our approach, we have given an estimation of threshold residual energy which was not addressed in [22]. 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

$$E_{\text{threshold}} = lE_{\text{elec}}\left(\frac{N}{k_{\text{opt}}} - 1\right) + lE_{\text{DA}}\frac{N}{k_{\text{opt}}} + lE_{\text{elec}} + l\varepsilon_{\text{fs}}d_{\text{toBS}}^2 \tag{6}$$

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

5.4 Energy consumption in EECPK-means protocol

Since the distance between the communicating CHs and the BS is assumed to be less than the threshold distance mentioned in (4), here it follows free space radio energy model for energy consumption mentioned in (2). BS calculates the number of sensor nodes n_{c} in each cluster after cluster formation phase. CH which sends data via intermediate CHs, consumes energy per round as

Algorithm 3: Balanced Cluster formation

Input:

D = set of n data items

k_{opt} = Optimum number of desired clusters

$E_{threshold}$ = threshold Energy

Output:

A set of k_{opt} clusters.

Steps:

1: Apply Midpoint method presented in **Algorithm 2** to choose k_{opt} out of D sensor nodes as initial cluster heads;

2: Repeat

3: Each of the remaining nodes decides to join its nearest CH according to the Euclidean distance.

4: Centroid of each cluster is calculated as

$$Centroid(X,Y) = \left(\tfrac{1}{S}\sum_{i=1}^{s} x_i, \tfrac{1}{S}\sum_{i=1}^{s} y_i \right)$$

5: After cluster formation, based on the distance from the centroid, an ID number is allotted to each node of a cluster, assigning smaller number to the closer one.

6: For all selected Cluster Heads

7: **If** (Residual energy of cluster head $\geq E_{threshold}$)

8: **then**

9: The node will remain as cluster head

10: **else**

11: Check ID number of all sensor nodes in that cluster.

12: The node in the next order of ID number is selected as a new CH.

13: **End if**

14: End for

15: The newly elected CHs inform other nodes about the CHs change.

16: Until the cluster heads are not changed any more.

Fig. 2 *Balanced cluster formation*

Algorithm 4: Data Communication

Input:

D = set of n data items

$\{CH_1, CH_2, \ldots CH_{k_{opt}}\}$ = A set of k_{opt} clusters

$d_{threshold}$ = threshold distance $= \sqrt{\frac{fs}{mp}} = 87.7\,meter$

Steps:

1: Sensor nodes send data packet to their CHs.

2: Calculate the distance between each elected CH and BS (d_{BS}).

3: If ($d_{BS} < d_{threshold}$)

4: **then**

5: Cluster Head directly communicate to the BS

6: else

7: It selects the nearest neighbour cluster head whose d_{BS} is less than $d_{threshold}$ to communicate to the BS.

8: End if

Fig. 3 *Data communication*

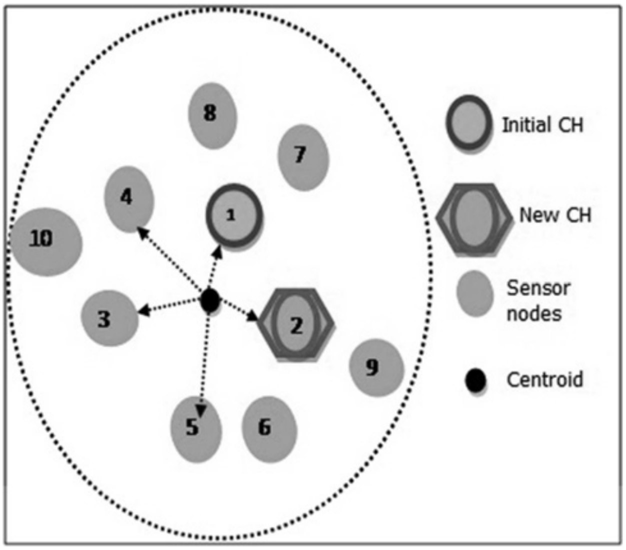

Fig. 4 *ID number assignment based on Euclidian distance of nodes from centroid in EECPK-means protocol*

follows

$$E_{\rm ICH} = lE_{\rm elec}(n_{\rm c} - 1) + lE_{\rm DA}n_{\rm c} + lE_{\rm elec} + l\varepsilon_{\rm fs}d^2_{\rm toICH} \quad (7)$$

CHs whose $d_{\rm toBS} \le d_{\rm threshold}$ 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

$$E_{\rm CH} = lE_{\rm elec}\left((n_{\rm c} - 1) + \frac{k_{\rm c}}{k_{\rm opt} - k_{\rm c}}\right) + lE_{\rm DA}\left(n_{\rm c} + \frac{k_{\rm c}}{k_{\rm opt} - k_{\rm c}}\right) + lE_{\rm elec} + l\varepsilon_{\rm fs}d^2_{\rm toBS} \quad (8)$$

where $n_{\rm c}$ is the number of sensor nodes in that cluster, $k_{\rm c}$ is the number of CHs which are unable to send data directly to the BS and $k_{\rm opt}$ is the total desired number of CHs. Here the value of $k_{\rm c}$ ranges from 0 to $(k_{\rm opt} - 1)$.

The energy dissipation of each non-CH node per round is

$$E_{\text{non-CH}} = lE_{\rm elec} + l\varepsilon_{\rm fs}d^2_{\rm toCH} \quad (9)$$

Therefore total energy dissipation in a round using EECPK-means protocol is calculated as

$$E'_{\rm round} = \sum_{k_{\rm c}} E_{\rm ICH} + \sum_{k_{\rm opt}-k_{\rm c}} E_{\rm CH} + \left(N - k_{\rm opt}\right)E_{\text{non-CH}} \quad (10)$$

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

6 Simulation result and analysis

For simulating our proposed EECPK-means protocol, MATLAB 7.7 and C language have been used. Simulation parameters are summarised in Table 1. Considering $d_{\rm BS} = 100$, we get the number of desired CH = 4 and considering $d_{\rm BS} = 85$, we get the number of desired CH = 5, having 100 sensor nodes in 100×100 m^2 sensing region. Simulations have been performed for both 4-cluster and 5-cluster networks. Our proposed EECPK-means protocol has been compared with K-means based Park's approach [22] used in WSN with respect to cluster formation. As well as, our proposed approach has been compared with existing approaches [7, 20–23] with respect to different network parameters.

Table 1 Simulation parameters

Parameter	Value
number of sensor nodes (N)	100
network size	100×100 m^2
base station's location	(0,0)
number of clusters ($k_{\rm opt}$)	4, 5
initial energy of node	1 J
data packet	3200 bits
$E_{\rm elec}$	50 nJ/bit
$\varepsilon_{\rm mp}$	0.0013 pJ/bit/m^4
$\varepsilon_{\rm fs}$	10 pJ/bit/m^2
energy for data aggregation ($E_{\rm DA}$)	5 nJ/bit/signal
$d_{\rm BS}$	85–100 m
$d_{\rm ICH}$	$d_{\rm BS}/2$
$d_{\rm threshold}$	88 m

6.1 Comparison of cluster structure of our proposed EECPK-means protocol with Park's approach

In Park's approach [22], the authors applied basic K-means algorithm for clustering. Fig. 5 compares the cluster formation of the 4-cluster sensor network using Park's approach [22] and our proposed approach in a 100×100 m^2 sensing region. It has been observed that in Park's approach a large variation in the cluster formation occurs to form 4 clusters from 100 sensor nodes. Here we have taken seven observations where it is found that in observation 4, cluster 2 contains 39 nodes, which is much higher than the average number of nodes (25). However, in observation 5, cluster 4 contains only 11 nodes, which is much lower than the average number of sensor nodes. Using Park's approach the resultant cluster shows uneven number of sensor nodes as shown in Fig. 5*a*. As a result CH of the heavily loaded cluster, which contains 39 nodes will be exhausted much earlier than the other clusters. After applying our proposed EECPK-means protocol, where the midpoint algorithm has been used for initial CH selection, we get cluster formation as shown in Fig. 5*b*. Amongst the seven observations, it is found that a particular cluster contains maximum 29 nodes and minimum 21 nodes, which is much closer to the average number of nodes to be present(25) in a particular cluster. Here the resultant clusters have almost equal number of sensor nodes, which ultimately leads to balanced cluster. Simulation results are summarised in Table 2 which is shown in Fig. 6.

In Fig. 7*a*, it shows communication between the CHs and the BS using Park's approach [22] and our proposed approach. In Park's approach single-hop communication happens between the CHs and the BS. This approach does not check whether the CHs are beyond the threshold distance from BS or not. As a result, the farthest CH consumes much more energy than others and get exhausted earlier in Park's approach [22]. However, in our approach the distance between the CH and the BS is checked. If this distance 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 CH. As a result, it provides enhanced network lifetime in WSN.

Considering $d_{\rm BS} = 85$, we get the number of desired CH = 5 having 100 sensor nodes in 100×100 m^2 sensing region. This time also we get unbalanced cluster formation like 4-clustered network. Here amongst the seven 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 for same observation 1, cluster 5 contains only 11 sensor nodes, which is much lower than the average number of sensor nodes. Using our proposed EECPK-means protocol, we find amongst the seven observations a particular cluster contains maximum 24 nodes and minimum 16 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 [22]. Simulation results are summarised in Table 3 and are shown in Fig. 8.

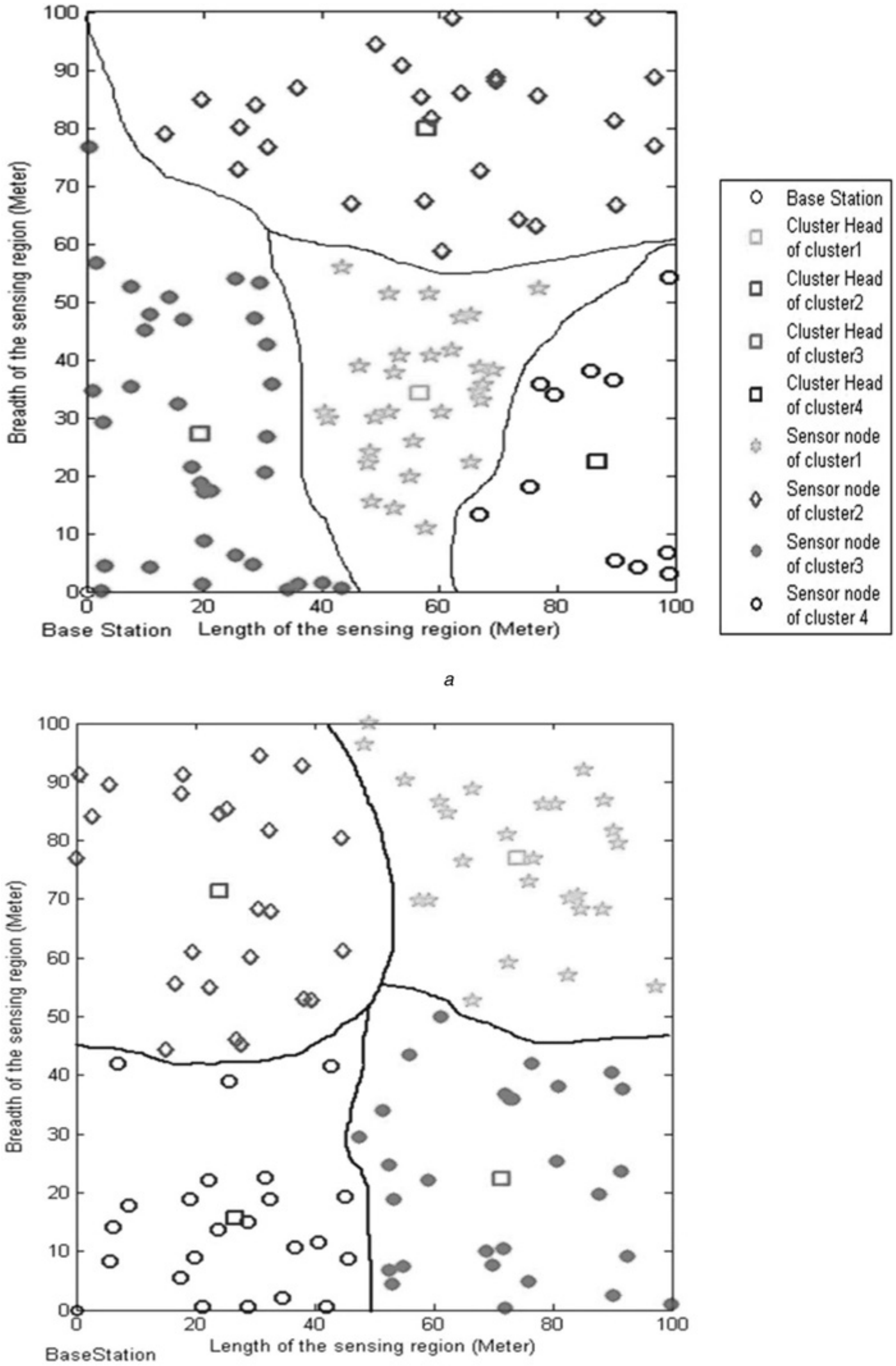

Fig. 5 *Cluster formation of sensor nodes*
a Using Park's approach [22]
b Using proposed approach

Table 2 Number of nodes present in four clusters using Park's approach [22] and our proposed approach

Number of observations	Number of nodes present in four clusters							
	Using Park's approach [22]				Using proposed EECPK-means protocol			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	24	18	38	20	25	25	23	27
2	15	22	35	28	26	25	24	25
3	31	30	26	13	28	26	23	23
4	19	39	24	18	23	28	24	25
5	29	27	33	11	26	24	28	22
6	30	20	18	32	22	26	27	25
7	33	17	22	28	23	27	21	29

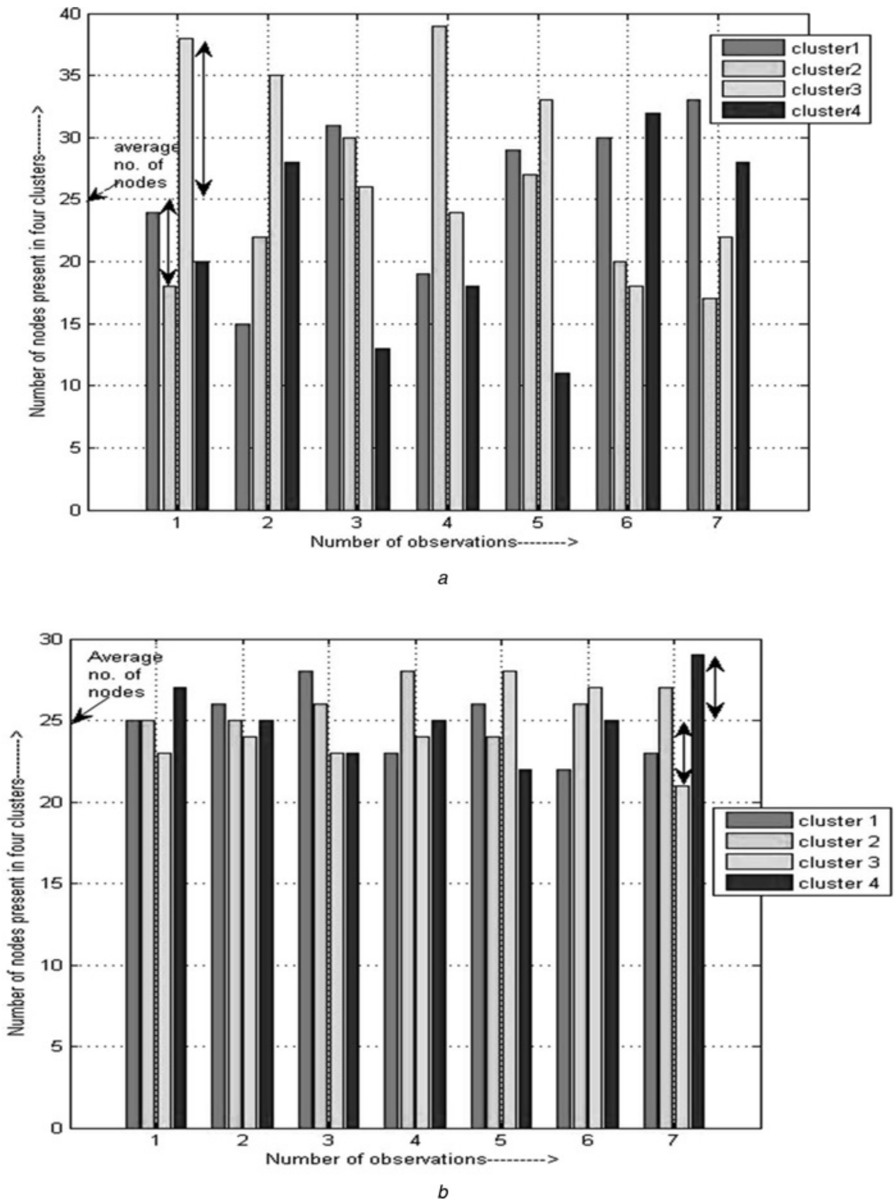

Fig. 6 *Number of nodes in four clusters using*
a Park's approach [22]
b Our proposed approach

6.2 Measure of dispersion of the number of nodes present in different clusters using Park's approach and our proposed approach with respect to the parameter standard deviation (σ)

Standard deviation is the measure of dispersion, which actually shows the degree of heterogeneity in the data. The standard deviation for a set of n numbers x_1, x_2, x_3, …, x_n having their arithmetic mean $\bar{x}$ is as follows

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum\left(x_i - \bar{x}\right)^2} \tag{11}$$

The deviation in cluster formation using Park's approach and our proposed approach from an ideal one can be tested using the parameter standard deviation. The number of nodes present in 4-cluster and 5-cluster using Park's approach [22] and EECPK-means protocol are summarised in Tables 2 and 3, respectively. Since total 100 sensor nodes are deployed, so for 4-cluster WSN the value of $\bar{x} = 25$ and for 5-cluster WSN the value of $\bar{x} = 20$. Table 4 shows the measure of dispersion of the number of nodes present in different clusters using Park's approach [22] and proposed approach with respect to standard deviation. From Table 4, it is obvious that the dispersion in Park's approach [22] is much higher than our proposed EECPK-means protocol. So from both 4-cluster and 5-cluster network it is obvious that our proposed approach performs better compared to Park's approach [22] in providing balanced clusters. As a result our proposed approach ultimately balances the load of CHs and prolong network lifetime.

6.3 Comparison of our proposed EECPK-means protocol with LEACH-B, BPK-means, Park's approach and Mk-means with respect to network lifetime and energy efficiency

The number of nodes which are alive using LEACH-B [7], BPK-means [21], Park's approach [22], Mk-means [23] and our proposed EECPK-means protocol are compared with respect to the number of rounds as shown in Fig. 9. It clearly shows that

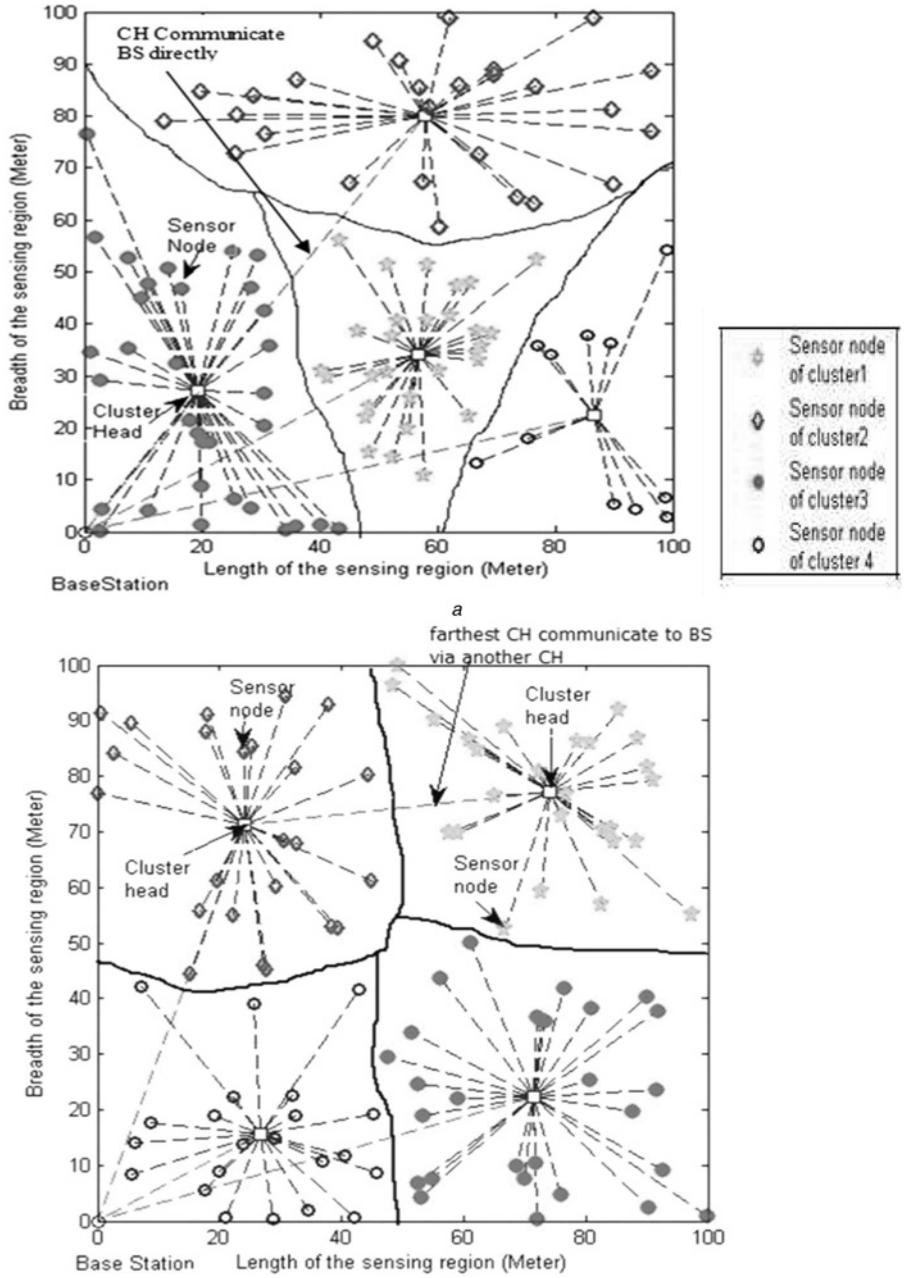

Fig. 7 *Communication between the CHs and the BS*

a Using Park's approach [22]
b Using proposed EECPK-means protocol

EECPK-means protocol provides better network lifetime compared to the above mentioned algorithms and helps to provide enhanced network lifetime to a great extent. Simulation result shows our proposed approach out performs LEACH-B, BPK-means, Park's approach [22] and Mk-means algorithm [23] with respect to the parameters FND, HND and LND as summarised in Table 5.

Table 3 Number of nodes present in five clusters using Park's approach [22] and our proposed approach

Number of observations	Number of nodes present in five clusters									
	Using Park's approach [22]					Using proposed EECPK-means protocol				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	13	18	25	33	11	19	21	24	18	18
2	23	24	19	15	19	21	18	22	21	18
3	18	26	18	26	12	19	20	21	20	20
4	17	32	17	16	18	21	19	19	20	21
5	10	30	14	24	22	20	24	23	17	16
6	28	12	23	16	21	19	21	16	22	22
7	15	25	18	17	25	17	23	18	22	20

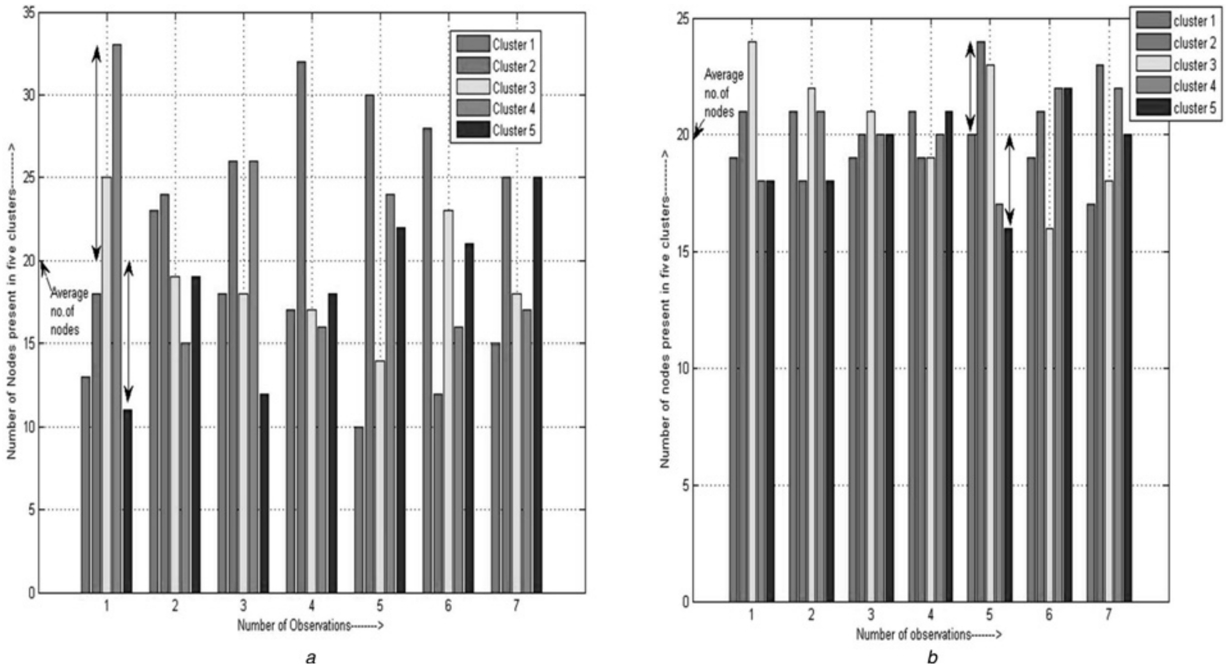

Fig. 8 *Number of nodes in five clusters*

a Using Park's approach [22]
b Proposed EECPK-means Protocol

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

From Table 6, it is observed that our proposed EECPK-means protocol is 50% better than LEACH-B [7], 14% better than BPK-means protocol [21], 10% better than Park's approach [22] and 6% better than Mk-means [23] with respect to the parameter half energy consumption (HEC).

In Fig. 10, it shows the comparison of the energy consumption of EECPK-means protocol with LEACH-B [7], BPK-means [21], Park's approach [22] and Mk-means algorithm [23] with respect to the number of rounds. It shows that EECPK-means protocol can reduce energy consumption significantly compared to above mentioned algorithms which eventually construct a green WSN.

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

Number of clusters	Observations	Standard Deviation (σ)	
		Park's approach [22]	EECPK-means
4	1	1.56	0.282
	2	1.48	0.141
	3	1.435	0.424
	4	1.68	0.374
	5	1.67	0.447
	6	1.22	0.374
	7	1.21	0.634
	average of seven observations	1.465	0.382
5	1	1.81	0.509
	2	0.721	0.374
	3	1.2	0.141
	4	1.35	0.200
	5	1.6	0.707
	6	1.241	0.509
	7	0.938	0.509
	average of seven observations	1.266	0.421

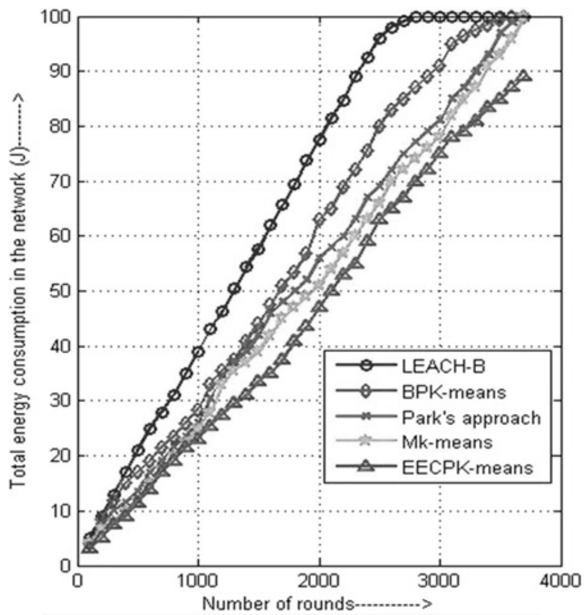

Fig. 9 *Number of alive nodes with respect to number of rounds*

Table 5 Comparison of network lifetime with respect to 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	2450	3080	3700

Table 6 Number of rounds with respect to HEC

Clustering algorithm	Number of rounds with respect to HEC
LEACH-B [7]	1350
BPK-means [21]	1780
Park's approach [22]	1830
Mk-means [23]	1900
EECPK-means	2020

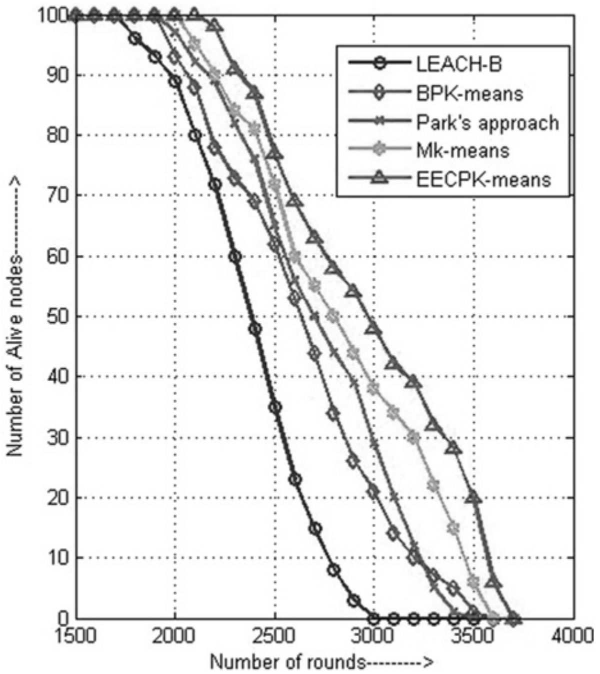

Fig. 10 *Energy consumption with respect to number of rounds*

Table 8 Comparison of our proposed approach with Park's approach and Mk-means algorithm with respect to cost function

Number of observations	Cost function		
	Park's approach [22]	Mk-means [23]	Proposed EECPK-means
1	3105.6	2670	2400.9
2	3138.6	2700	2692.7
3	3134.6	2800	2618.1
4	2942.4	2580	2613.1
5	3028.1	2900	2731.6
6	2986.4	2970	2639.9
7	3154.7	2990	2545.6

The contribution and novelty of our proposed approach are presented in Table 7.

6.5 Cost function

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

$$\text{Cost} = \sum_{j=1}^{k} \sum_{s_i \in C_j} \text{dist}\left(s_i, \text{CH}_j\right) \tag{12}$$

In (12), dist(s_i, CH$_j$) measures the Euclidian distance between the sensor node s_i and its corresponding CH CH$_j$ in cluster C_j. Here k is the number of clusters in the sensing region. Table 8 shows the comparison of our proposed approach with Park's approach [22] and Mk-means algorithm [23] with respect to cost function. In this case, our proposed approach shows better performance than both Park's approach and Mk-means algorithm.

7 Conclusion

Clustering is an effective means of data aggregation, which not only reduces energy consumption, but also provides uniform coverage profoundly. Though K-means is a widely used clustering algorithm in various fields including WSN but it is unable to ensure best result due to its random initial centroids selection. Our proposed EECPK-means protocol improves K-means algorithm by incorporating midpoint method for initial centroids selection. It also optimises CH selection method by considering residual energy as the parameters of CH selection in addition to Euclidean distance used in basic K-means algorithm. At the same time multi-hop communication occurs depending on the distance of the CHs from the BS so that the CHs which is far away from the sensing region does not exhaust so much and can successfully deliver the aggregate data to the BS. At the same time EECPK-means protocol is 50% better than LEACH-B [7], 14% better than BPK-means protocol [21], 10% better than Park's approach [22] and 6% better than Mk-means algorithm [23] with respect to the parameter half energy consumption. Thus, green WSN is achieved.

Table 7 Comparison of our proposed approach with existing approaches

Features	Existing approaches				Proposed EECPK-means
	MPC [20]	BPK-means [21]	Park's approach [22]	Mk-means [23]	
based on	K-means algorithm	K-means algorithm	K-means algorithm	K-means algorithm	midpoint algorithm and K-means algorithm
initial CHs are selected	randomly	randomly	randomly	randomly	using midpoint algorithm
produces balanced cluster	×	✓	×	✓	✓
calculates optimum number of CHs	✓	✓	×	×	✓
clustering considers minimal distance between the sensor nodes and CH	✓	✓	✓	×	✓
clustering considers threshold distance between the CH and the BS	✓	×	×	×	✓
CH selection considers residual energy as a factor	✓	×	✓	✓	✓
specified threshold residual energy of CH	×	✓	×	✓	✓
CHs are uniformly distributed in the sensing region	✓	×	×	×	✓
supports multi-hop communication between the CH and the BS	×	×	×	×	✓
improves network lifetime	✓	×	✓	✓	✓

IET Wirel. Sens. Syst., 2016, Vol. 6, Iss. 6, pp. 181–191

8 Acknowledgments

The authors are grateful to Department of Science and Technology (DST) for sanctioning a research Project entitled 'Dynamic Optimisation of Green Mobile Networks: Algorithm, Architecture and Applications' under Fast Track Young Scientist scheme reference no.: SERB/F/5044/2012-2013 and DST FIST Reference no.: SR/FST/ETI-296/2011 under which this paper has been completed.

9 References

1 Bhatti, S., Xu, J., Memon, M.: 'Clustering and fault tolerance for target tracking using wireless sensor networks', *IET Wirel. Sens. Syst.*, 2011, **1**, (2), pp. 66–73
2 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
3 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.*, 2015, **5**, (5), pp. 243–249
4 Yick, J., Mukherjee, B., Ghosal, D.: 'Wireless sensor network survey', *Comput. Netw.*, 2008, **52**, (12), pp. 2292–2330
5 Abbasi, A.A., Younis, M.: 'A survey on clustering algorithms for wireless sensor networks', *Comput. Commun.*, 2007, **30**, (14), pp. 2826–2841
6 Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Trans. Wirel. Commun.*, 2002, **1**, (4), pp. 660–670
7 Tong, M., Tang, M.: 'LEACH-B: an improved LEACH protocol for wireless sensor network'. Proc. Int. Conf. Wireless Communications Networking and Mobile Computing (WiCOM), September 2010, pp. 1–4
8 Ray, A., De, D.: 'Energy efficient clustering hierarchy protocol for wireless sensor network'. Proc. of IEEE Int. Conf. on Communication and Industrial Application (ICCIA), December 2011, pp. 1–4
9 Ray, A., De, D.: 'Energy efficient cluster head selection in wireless sensor network'. Proc. of IEEE Int. Conf. on Recent Advances in Information Technology (RAIT)-2012, ISM, Dhanbad, Jharkhand, March 2012, pp. 306–311
10 Ray, A., De, D.: 'Energy efficient clustering algorithm for multi-hop green wireless sensor network using gateway node', *Adv. Sci. Eng. Med.*, 2013, **5**, (11), pp. 1199–1204
11 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
12 Guo, P., Jiang, T., Zhang, K., *et al.*: 'Clustering algorithm in initialization of multi-hop wireless sensor networks', *IEEE Trans. Wirel. Commun.*, 2009, **8**, (12), pp. 5713–5717
13 Deng, S., Li, J., Shen, L.: 'Mobility-based clustering protocol for wireless sensor networks with mobile nodes', *IET Wirel. Sens. Syst.*, 2011, **1**, (1), pp. 39–47
14 Kumar, R., Malik, A., Kumar, B.: 'NEECP: a novel energy efficient clustering protocol for prolonging lifetime of WSNs', *IET Wirel. Sens. Syst.*, 2016
15 Lu, K., Liu, G., Mao, R., *et al.*: 'Relay node placement based on balancing power consumption in wireless sensor networks', *IET Wirel. Sens. Syst.*, 2011, **1**, (1), pp. 1–6
16 Cao, F., Liang, J., Jiang, G.: 'An initialization method for the K-means algorithm using neighborhood model', *Comput. Math. Appl.*, 2009, **58**, pp. 474–483
17 Aggarwal, N., Aggarwal, K.A.: 'A mid–point based k-mean clustering algorithm for data mining', *Int. J. Comput. Sci. Eng.*, 2012, **4**, (6), pp. 1174–1180
18 Khan Shehroz, S., Amir, A.: 'Cluster center initialization algorithm for K-means clustering', *Pattern Recognit. Lett.*, 2004, **25**, (11), pp. 1293–1302
19 Žalik, K.R.: 'An efficient k′-means clustering algorithm', *Pattern Recognit. Lett.*, 2008, **29**, (9), pp. 1385–1391
20 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.*, 2012, **4**, (1), pp. 18–24
21 Tan, L., Gong, Y., Chen, G.: 'A balanced parallel clustering protocol for wireless sensor networks using K-means techniques'. Proc. of IEEE Second Int. Conf. on Sensor Technologies and Applications, August 2008, pp. 300–305
22 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'. Proc. of IEEE 27th Int. Conf. on Advanced Information Networking and Applications Workshops, 2013, pp. 910–915
23 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, **2016**, pp. 1–11, Article ID 5040475
24 Niknam, T., Amiri, B.: 'An efficient hybrid approach based on PSO, ACO and K-means for cluster analysis', *Appl. Soft Comput.*, 2010, **10**, (1), pp. 183–197
25 Ferreira, L.N., Pinto, A.R., Zhao, L.: 'QK-means: a clustering technique based on community detection and K-means for deployment of cluster head nodes'. Proc. of IEEE Int. Joint Conf. on Neural Networks (IJCNN), June 2012, pp. 1–7
26 Sasikumar, P., Khara, S.: 'K-means clustering in wireless sensor networks'. Proc. of IEEE Fourth Int. Conf. of Computational Intelligence and Communication Networks (CICN), November 2012, pp. 140–144
27 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.*, 2005, **22**, (2), pp. 287–310
28 Napoleon, D., Ganga Lakshmi, P.: 'An enhanced K-means algorithm to improve the efficiency using normal distribution data points', *Int. J. Comput. Sci. Eng.*, 2010, **2**, (7), pp. 2409–2413
29 Jain, A.K.: 'Data clustering: 50 years beyond K-means', *Pattern Recognit. Lett.*, 2010, **31**, (8), pp. 651–666
30 De, D., Ray, A., Mukherjee, S., *et al.*: 'Performance analysis of different tree structures in heterogeneous wireless sensor network', *IETE J. Res.*, 2016, pp. 1–9
31 Ray, A., De, D.: 'An energy efficient sensor movement approach using multi-parameter reverse glowworm swarm optimization algorithm in mobile wireless sensor network', *Simul. Modelling Pract. Theory*, 2016, **62**, pp. 117–136