Efficient Clustering Protocol Using Fuzzy K-means and Midpoint Algorithm for Lifetime Improvement in WSNs

Nandoori Srikanth1* Muktyala Siva Ganga Prasad1

1Department of Electronics & Communication Engineering, KL University, India
* Corresponding author’s Email: srilovesnature@gmail.com

Abstract: Due to shortage of power source, energy is the significant concern part in Wireless Sensor Systems. Wireless Sensor Networks (WSNs) are currently used in numerous kinds of WSN simulators have been established to visually illustrate the coverage, energy consumption status and expected lifetime of WSN sensors. Most of the Researchers have proposed and develop routing algorithm to increase the network lifetime. Among them clustering is the best technique and it is well known for accomplishing energy effectiveness in WSNs. In this paper, Energy Efficient Multi-hop Routing Clustering Protocol (EEMRCP) based on Fuzzy K-Means and Centralized Mid-point Algorithm (FKM-CMA) is proposed for network lifetime improvement. An existing and frequently used algorithm is K-means algorithm in WSN. The main limitation of this K-means algorithm is random initial centroid selection. To improve K-means algorithm, a Midpoint method in initial centroid selection is used. This is named as FKM-CMA. There are two main considerations for cluster head selection in proposed work, one is it’s residual energy, and another one is Euclidean distance used in basic Fuzzy K-means algorithm. Finally, Multi-hop communication is performed for transmitting the packets from CHs to Base Station depending on the distance between them. Furthermore, stimulated outcomes showed that the total network performance of the proposed approach is improved than the other existing approaches.

Keywords: Wireless sensor network, Fuzzy k-means, Midpoint algorithm, Energy consumption, Network lifetime.

1. Introduction

WSNs equipped with a group of various sensor nodes with limited sensing, computing and communication abilities. These sensors nodes are set up over a big area with one or more than one Base Station [1]. Each node contains a detecting part, a processing part, a transmission part, a battery, and a power harvester. WSN has extensive application possibilities, which includes environmental monitoring, health monitoring, terrestrial and underwater habitat monitoring, disaster management, military surveillance, forest fire-tracking, security surveillance, etc. Rapid developments in wireless communications system, low-power electronics, battery technology, and power harvesting abilities have facilitated the improvement of low-cost WSNs [2]. The routing methods selection is a vital problem for the effective delivery of the packets to their target point. Furthermore, in such networks, the applied routing scheme should confirm the lowest of the energy consumption and hence extension of the lifetime of the network [3]. From few years, researchers have demonstrated that the clustering is an effective approach in expanding lifetime and reduces energy consumption of Wireless Sensor Networks. In each cluster, there are several nodes, among them one will be act as a cluster head (CH) and others will be numerous cluster members. Cluster members sense and collects the information from surrounding environment and forward the information to cluster head (CH). Moreover, CHs function as fusion point for data aggregation, so that the actual data transferred to the BS is abridged [4]. Clustered WSNs also utilizes single-hop technique and multi-hop technique for communication between the nodes. The CHs utilizes multi-hopping technique to reach the BS when they are far away...
from the BS. Therefore clustering and multi-hop communication are the effective tools to make network more energy efficient. [5]. By using fuzzy K-means algorithm and centralized midpoint method, k-means algorithm is enhanced and cluster head selection is also improved.

2. Literature review

Pandey A, Muchhal N., Abdullah, M. Mishra, (2014) [4] proposed a clustered WSNs using Multihop transmission. This dynamic algorithm operates spontaneously over no prior distribution of the levels of energy in the nodes and also creates a hierarchy of cluster heads and detects that the energy savings rise with the number of levels in the hierarchy. In this paper, the clustering procedure comprises of setup stage and steady state stage. In setup stage CHs is chosen based on residual energy and the steady state stage involves data aggregation and data reporting procedure. Finally each CH obtains information from its cluster member at the end of sensing process. It combined receive data and direct it to the nearby super CH by means of multi-hop communication. The drawback of this protocol is data aggregation burden on CH increases in Large Scaled WSN.

Jeong.Y, Chung. Y, & Park.J (2011) [6] proposes a coverage visualization technique for a sensor node using heat map for increasing network lifetime. This method depends on the binary model that utilizes the detection rate of nodes depicted on a heat map. In this paper, the simulator validates visualization methods for effective coverage visualization and energy lifetime prediction has been designed and implemented. The simulator have named as current reduction state observer on WSN device (CROWD), which implements simulation by a virtual sensor network and multi-mobile objects. This CROWD also demonstrates the detection rate of sensors on a heat map.

Singh.S, Chand.S, Malik.A & Kumar.B, (2016) [7] proposed a new effective clustering algorithm for extending lifetime of WSN. This novel energy efficient clustering protocol (NEECP) chooses the cluster heads in an effective way by an adaptable sensing range and achieves data aggregation by means of chaining approach. It also evades communication of redundant data by using a redundancy check function for enhancing the network lifetime. This method is implemented with aggregation and without aggregation of data. The experimental results show that this algorithm rises the lifetime of a network by 59.76% and 7.17% without aggregation as compared to the HEED and IBLEACH, respectively. Moreover increases the network lifetime by 92% and 49.53% over the HEED and IBLEACH, respectively with data aggregation. The drawback of this protocol is Communication Overhead increases due to chain based Data Aggregation.

Agarwal.N, & Agarwal.K. (2012), [8] proposed a Clustering approach based on K-means method for data mining. In k-means clustering algorithm, the total centroid is equivalent to the total clusters in which data has to be divided which in turn is taken as an input parameter. The first centroids in original k-means are selected arbitrarily from the given dataset. Then various clustering outcomes are formed with different randomly selected initial centroids for the same dataset. Finally the author provides a solution to this restriction of the original K-means Algorithm.

Singh.J, Singh.B, & Shaw.S., (2014), [9] proposed a modified LEACH for energy optimization in WSN. In this method, clusters are created intermittently based on remaining energy and distance. Re-clustering allocates the workload among various nodes and in turn increases the network lifetime by revolving the cluster head. But the sensor nodes remain in active state only during its transmission period and rest of the time it remains in sleep state to save energy. The outcome infers that this algorithm executes superior than the LEACH and also modified LEACH protocol in terms of network lifetime and also provides more throughput than LEACH.

Periyasamy.S, Khara.S., & Thangavelu.S (2016) [10] proposed a distributed WSN by modified k-means algorithm. This method considers three CHs for each cluster. The CHs perform a load distribution technique to rotate as the active CH, which saves the nodes remaining energy, thus prolonging network lifetime. Furthermore, it decreases the number of times clustering has to be done and considerably rises the number of data packets transmits during network operation. The outcome demonstrates that Modified k-means algorithm was found to outperform the existing clustering algorithms due to its exclusive multiple cluster head methodology.

Kumar (2013), [12] proposes two clustering protocols. In single-hop energy efficient clustering protocol (S-EECP), the CHs are chosen by a weighted likelihood according to the relation between remaining energy of every node and mean energy. Therefore, the nodes having maximum beginning energy and remaining energy selected as Cluster Heads than nodes with minimum energy while in multi-hop energy efficient clustering
protocol (M-EECP), the chosen CHs transfer the data packets to the BS through multi-hop communication method. The authors consider three sensor nodes provided with dissimilar battery energy. Finally, this approach is mainly proposed for increasing the network lifetime of the WSNs.

Nirmala Hiremani, Tiptur Gangaraju Basavaraju, [13] proposed Efficient Routing Protocol Adopting Enhanced Cluster Formation accompanied by Fuzzy K –means algorithm. This protocol chooses cluster member selection based on particle swarm optimization and authors proposed a super cluster head selection based on Mamdani’s rule, and nearest distance to sink. It increases the Lifetime of the network, and enhances the throughput of the Network. The Drawback of this protocol is, Due to selection of SCHs which are near to sink; The Count CHs have been greatly reduced in large scaled WSNs.

Syambabu Karthick [16] proposed a secured Energy aware Routing protocol, and this protocol is named as Trust Destruct Protocol (TDP). In this protocol an improved k- means algorithm is used to enhance topology management. Every Node energy is calculated by sending and receiving a set of sample packets (or hello packet), Every Node is assigned with a grade according to its Energy. The main drawback of this protocol is every node has to calculate its Trust and distrust values before transmission to other node, due to this Computation and communication cost may increases.

Theagarajan Senthil, Balasubramanian Kannapiran [18] proposes an energy efficient, cluster based and trustworthy Energy Conserving Trustworthy Topology Management (ECTTM) algorithm. The area of the network is divided into number of grids of equal in size, and the sensor nodes are randomly deployed in entire grids. Location coordinates are the major criteria to form clusters and selection of chief node. The major duty of chief node is to compute the trust degree and to manage the state of the nodes. The reason for minimal energy consumption is the control of functioning nodes. Here BS is a highly energized mobile unit, and sensed nodes send their data to the BS through Chief Nodes. The Main drawback of this protocol is BS could not able to cover all areas in the network due to unexpected shape of the grids, and BS could not survive for several Large Number of Rounds.

3. Problem statement

Energy efficiency is one Key research area in WSN. Greater energy is used in transmission of data. The energy efficiency and life time of sensor networks, both have linear in relationship. Hence, to improve the life time of the network is performed by reducing the power consumption of sensor nodes. Numerous research works have performed to maintain the sensor network up to a long time, but till it is major issues in WSN. Several efficient energy routing and clustering protocols have been proposed such as LEACH, TEEN, APTEEN, PEGASIS, HEED, IBLEACH, and T-LEACH, etc. in the existing system to reduce the energy consumption and also to improve lifetime of network. This existing protocol is unable to produce balanced cluster and reduces energy consumption only to a minimum level. Therefore, the proposed EEMRCP methodology improves the structure of the cluster formation and also minimizes energy consumption for data communication.

4. Proposed methodology

In the proposed method, the data transmission is performed using Fuzzy K-Means and centralized mid-point algorithm. This proposed EEMRCP enhances K-means algorithm by using midpoint method for initial centroid selection. There are two main considerations for cluster head selection in proposed work, one is it’s residual energy, and another one is Euclidean distance used in basic Fuzzy K-means algorithm.

4.1 Objective of proposed approach

- To design an Efficient Energy Clustering protocol using Fuzzy K-means and centralized midpoint algorithm.
- Evaluate the performance of proposed protocol.
- Compare the results of proposed protocol with existing methods.
- Increases the lifetime of the networks with low energy consumption.

4.2 Network model

In proposed methodology, the sensor network model is built upon it with the following properties. The following assumption of the network model are given below:

(a) The Base station is highly energized and it is placed far away from the sensor network area.
(b) Every sensor nodes are equally spread inside a square field with a uniform initial residual energy.
(c) The sensor node continuously sense from the environment and forward sensed data to the BS at required regular time intervals.

(d) The sensor nodes should perform computational functions and collects data from environment individually.

(e) The Base station and sensor node are left unattended after deployment.

(f) Each and every node should have to communicate with CH and transmit data to the CH by using either single Hop or Multi Hop communication, which depends on the threshold distance value.

(g) Instead of transmitting individual data to the BS, the CH should aggregate the data collected from all neighbour nodes and then forward aggregated data to BS.

(h) By using Data compression techniques the CH transmits compressed data to the BS at high data rates have been achieved.

(i) The BS specifies every node to reinitiate clustering when all the CH nodes in the network have inadequate energy.

### 4.3 Radio energy model

The energy dissipation model is used for predicting performance of the proposed technique. The energy required for transmitting H bits from transmitter to the receiver is expressed as,

\[
\begin{align*}
e_T(H, r) &= \begin{cases} 
H \times e_c + H \times e_d \times r^2 & \text{if } r \leq r_0 \\
H \times e_c + H \times e_m \times r^4 & \text{if } r \geq r_0 
\end{cases}
\end{align*}
\]

(1)

Where \( r \) is the interval between sender and receiver and \( e_c \) is the amount of energy consumption for transmitting and receiving data. By equating these two expressions at \( r = r_0 \), then \( r_0 = \sqrt{\frac{e_m}{e_s}} \). The parameters \( e_s \) free space and \( e_m \) multi path fading channel models differ based on the transmitter and receiver distance. The free space model is used if the distance is smaller than a threshold \( r_0 \), and multi path model is used if the distance is greater than \( r_0 \). The cost of energy essential for getting the data is expressed as,

\[
e_g = H \times e_c
\]

(2)

In this method, consider an area \( L \times L \) square meters. The Base station is placed at the middle of the network for simplicity. Every non-channel head node transmits R bits to the elected Channel Head node. Therefore, the energy dissipated in the Channel Head node is expressed by

\[
e_{ch} = \left( \frac{T}{K} - 1 \right) \times H \times e_c + \frac{T}{K} \times H \times e_d + H \times e_c + e_s \times r_d^2
\]

(3)

Where,

- \( T \) is the number of nodes equally dispersed over the square area.
- \( e_d \) is the processing cost of a bit report to the Base Station.
- \( r_d \) is the average distance between a Channel Head and Base Station.

The energy used in a non-CH is expressed as

---

International Journal of Intelligent Engineering and Systems, Vol.11, No.4, 2018  
DOI: 10.22266/ijies2018.0831.07
\[ e_{\text{non-ch}} = H \times e_{\text{c}} + H \times e_{\text{r}} \times r_h^2 \]  
\[ \text{Where}, \quad r_h^2 = \frac{L^2}{2\pi K} \]

\[ r_h \] is the average distance between a cluster member and CHs.

Now adding Eqs. (3) and (4), get the total energy dissipated during a round is given by

\[ e_{\text{total}} = H \times (2 \times T \times e_{\text{c}} + T \times e_{\text{r}} \times (K \times r_h^2 + T \times r_h^2)) \]

The e_{\text{total}} is differentiated with respect to K and equating to zero, the optimal number of cluster can be estimated is given by

\[ K_o = \sqrt{\frac{T}{2\pi}} \sqrt{\frac{e_{\text{r}} L}{e_{\text{m}} r_h^2}} \]

Where, \[ r_p = 0.765 \times \frac{L}{2} \]

The optimal probability of a node to become CHs, \( G_{\text{opt}} \) is given by

\[ G_{\text{opt}} = \frac{1}{0.765} \times \frac{2}{L \pi} \times \frac{e_{\text{r}}}{e_{\text{m}}} \]

\( G_{\text{opt}} \) is very significant parameter because for formation of cluster and entire energy spent during a process is increased exponentially.

5. **Fuzzy k-means and centralized midpoint clustering algorithm**

Fuzzy K-means Clustering algorithm is mainly used for calculating mean distance from the origin. This clustering algorithm classifies the data set into R clusters using the Euclidian distance mean. The distance is calculated between the data points by Euclidean distance such as,

\[ \text{Dis}(S_1, S_2) = \sqrt{\sum_{i=1}^{n} (S_{1i} - S_{2i})^2} \]

This K-means clustering algorithm is performed by using fuzzy set. Fuzzy set is one of the simple models in fuzzy logic which can be characterized by some linguistic terms like ‘near’, ‘few’, and ‘high’.

**Fuzzy set F can be expressed by a set of ordered pairs in universe of discourse U.**

\[ F = \{(p, \mu_R(x)) | x \in U\} \]

Where \( \mu_R(x) \) denotes the membership function of \( x \) in F set.

The implementation of Fuzzy K-Means algorithm involves four steps such as

(i) Partition the inputs object into R- number of clusters as the centroid such as \( \{S_1, S_2, S_n\} \) and association of \( S_i \) will be done only to one cluster.

(ii) Place R centroids in random places \( \{D_1, D_2, \ldots, D_R\} \)

(iii) Determine the new cluster centres \( D_j \), \( j=1,2,\ldots,R \) which includes following steps such as

1. Find the nearest centroid for each point \( S \)

\[ D(i) = \arg \min \| S_i - D_k \|^2, i = 1,2,\ldots,M \]  

2. Assign the point \( S_i \) to centroid \( D_j \) which is at the least distance.

3. The new cluster Centre for each cluster \( j=1,2,\ldots,R \) is given by

\[ D_j(z) = \frac{1}{w_j} \sum_{S_i \in D_j} S_i(z) \quad \text{For } z=1,2,\ldots,a \]

Where \( z \) is the attribute value of the data set and \( w_j \) is the number of sample in \( D_i \).

(iv) If distance is greater than threshold value, packets will be directly transmitted to BS, otherwise the neighbouring cluster head will transmits the packet using multi-hop routing.

(v) Go back to step (ii), stop if \( D_j(z) = D_R \) for \( j=1,2,\ldots,R \) then the algorithm has converged and the process is ended.

Or else go to stage (ii).
Algorithm 1: Fuzzy K Means Clustering Algorithm

N = Number of nodes set
H = Number of considered cluster head
Calculate distance from origin to each node
Sort the mean value 0<20<50, upto 100
Consider one centroid for each groups
Apply Euclidian distance to each node from centroid
If D > threshold, cluster head is considered
Otherwise go for step 5
Stop

Algorithm 2: Centralized Midpoint algorithm with multi-hop routing

H = Number of considered cluster head
Calculate distance from each cluster head to base station.
Sort the distance value for 4 clusters
If D > threshold, It will transmit the packets to base station directly.
Otherwise it will select the neighbours cluster head as multi-hop and then it will transmit the packets using the selected multi-hop routing.
Stop

The cluster formation of proposed approach was explained in above steps with many stages.

6. Simulation and analysis of results

The proposed EEMRCP was implemented using Network Simulator (NS-2) software. NS-2 is one of the best simulator for WSNs. It consists of number of protocols in each layer. It uses OTcl as configuration and script interface. Here, ns-allinone version 2.33 are used to simulate the parameters. The proposed EEMRCP with FKM-MA algorithm are compared with the existing EECPK-means, LEACH-B, BPK-Means and Park’s approach[11] in terms several parameters.

The parameters used in simulation are listed in table 1. The network size of the proposed simulation model in terrain area is 100 m x 100 m using Destination-Sequenced Distance Vector (DSDV) routing protocol for monitoring the important parameters of the networks.

Fig. 2 shows the initial result of node deployment and Fig. 3 states that the cluster formation of selected nodes based on the distance calculated distance values.

<table>
<thead>
<tr>
<th>Table 1. Simulation parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARAMETERS</td>
</tr>
<tr>
<td>Number of nodes</td>
</tr>
<tr>
<td>Network size</td>
</tr>
<tr>
<td>Node placement</td>
</tr>
<tr>
<td>Node mobility</td>
</tr>
<tr>
<td>MAC layer protocol</td>
</tr>
<tr>
<td>Routing Protocol</td>
</tr>
<tr>
<td>Time simulation</td>
</tr>
<tr>
<td>Dimension of Topography (x,y)</td>
</tr>
<tr>
<td>Interface Queue Type</td>
</tr>
<tr>
<td>Antenna Type</td>
</tr>
<tr>
<td>Application Layer Protocol</td>
</tr>
<tr>
<td>Channel</td>
</tr>
<tr>
<td>Radio Propagation Model</td>
</tr>
<tr>
<td>Network Interface</td>
</tr>
</tbody>
</table>

The proposed approach was implemented in two stages such as direct connection and indirect connection. CH4 and CH2 is directly transmitting the packets to base station and CH3 & CH1 is indirectly transmitting the packets to base station through multi-hop routing. Here, CH4 and CH2 will act as a multi-hop to reduce the energy consumption. The simulation output was graphically explained in Figs. 4 and 5.
a) **Connectivity**

Though connectivity is a critical thing for sensing the coverage area, it is an important parameter for wireless sensor networks. Fig. 6 shows the connectivity graph which specifies that the proposed algorithm achieves high connectivity than the existing approaches.

b) **Delay**

Delay is one important factor in wireless networks and it is calculated as the time it takes to transmit data packets from the source to the destination. It is confirmed that the FKM-CMA has minimum end to end delay when compared to other existing approach which is shown in Fig. 7.

c) **Energy Consumption**

The total energy utilized by the nodes in every moment in a network is considered as Energy Consumption. The energy utilized by the nodes for transmitting (Pt), reception (Pr), and initialization (Pi) is called as energy consumption. The average amount of energy utilization is proportional to the average amount of number of parcels delivered in the system. The energy consumption for the proposed network was represented graphically in Fig. 8. Here the proposed technique consumes less amount of energy compared with all existed protocols.

d) **Throughput**

Throughput is an important parameter and it reflects on system performance. It is nothing but
amount of data forward to receiver from sender within a time period successfully. The proposed FKM-CMA achieves higher throughput than existing schemes which is represented in Fig. 9.

e) Network Lifetime

Network lifetime depends on the lifetimes of every single node that established WSNs. Therefore, the proposed method improves the network lifetime compared to the existing algorithm [7] which is graphically represented in Fig. 10.

The proposed EEMRCP with FKM-CMA algorithm was compared with existing protocol such as EECPK-means [11] and PDORP [30]. The simulation results were stated in Tables 1 and 2. The results shows that the proposed protocol was efficient than existing protocol.

Table 1. Comparison of network lifetime, energy consumption and connectivity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Energy Consumption</td>
<td>42</td>
<td>43</td>
<td>68</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>Network Lifetime</td>
<td>55</td>
<td>48</td>
<td>39</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>Connectivity</td>
<td>68</td>
<td>64</td>
<td>44</td>
<td>53</td>
</tr>
</tbody>
</table>

Figure. 9 Throughput

Figure. 10 Network lifetime

Figure. 11 Comparison of network lifetime, energy consumption and connectivity

Figure. 12 Comparison of throughput and delay
Table 2. Comparison of throughput and delay

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Throughput</td>
<td>89</td>
<td>8</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Delay</td>
<td>3</td>
<td>4</td>
<td>33</td>
<td>14</td>
</tr>
</tbody>
</table>

7. Conclusion

This paper presents an EEMRCP using Fuzzy K-means and centralized mid-point algorithm for network lifetime improvement in WSNs. The CHs communicate with the BS directly or indirectly by means of multi-hop transmission. The network lifetime is one of the significant factors in WSN due to limited power energy of sensors. The Performance estimation of the proposed approach attains higher throughput, network lifetime and connectivity as well as reduced energy consumption and delay as compared to the existing protocol such as EECPK-means [11] and PDORP [30] respectively. Therefore, the proposed approach enhances cluster formation and minimizes energy consumption. The simulation output infers that proposed EEMRCP provides better network lifetime and reduces energy consumption to a great extent. The extension of present work (or) fixed sensor networks will be introduced with some mobile nodes in each cluster to reduce energy up to great extent. The proposed algorithm will be incorporated with optimization techniques and inclusion of security in future.

References


