Modified Clustering Algorithm for Energy Efficiency Utilizing Fuzzy Logic in WSN (MCF)

Tuba Firdaus¹, Meenakshi Yadav²
Dept. of Computer Sc. & Engg, MMMUT Gorakhpur, U.P., India, Dept. of Computer Sc. & Engg, MMMUT Gorakhpur, U.P., India
tubafirdaus@gmail.com; cycuspinus@gmail.com

Abstract

A number of research have taken place in the field of Wireless Sensor Networks (WSN) as there is continuous need of advancement in the field of wireless communication, digital technology and micro-electro-mechanical systems (MEMS). So the need of growth of low cost, low power, multifunctional sensor nodes have been required. A Wireless Sensor Network is a collection of sensor nodes that have the capability of sensing any environmental phenomenon, processing that information and then sending that data to the base station. A single sensor node is not capable of capturing desired information from a particular region so a collection of nodes are arranged to get accurate and sufficient result. This collection of sensor nodes along with a base station will collaboratively form a network that is known as Wireless Sensor Network.

As limited energy is one of the most important constraint of WSN so it must be assured that it is utilized in most efficient way. Clustering is best approach to remove redundant data transmission to base station. Each cluster has a cluster head that is responsible for transmitting data to base station for that cluster members. Cluster head (CH) collect the data from all members of its cluster and perform aggregation on these data to remove redundancy then send it to base station. So appropriate CH election is very important for improving efficiency.

In this thesis we have presented a clustering approach that has taken a heterogeneous environment and uses fuzzy logic to elect CHs more efficiently. We have combined two parameters Distance and Residual Energy and apply fuzzy rules on that to find the priority of a node for being a CH. Simulation shows that using fuzzy logic in SEP (Stability Election Protocol) will improve the energy efficiency by providing better load distribution and utilizing the benefits of heterogeneity of network. We have shown our analysis on two parameters- Number of dead nodes and average energy of nodes

Keywords

Sensor Nodes; Clustering; Energy Efficiency; Fuzzy Logic; Residual Energy Distance

INTRODUCTION

Wireless Sensor Networks [1][2] is an emerging technology in the field of communication. Wireless sensor nodes provide real observed data from the different environmental conditions and provide data for monitoring and controlling the systems.

In Wireless Sensor Network sensor nodes are most important which are responsible for gathering sensory information, processing it and then forwarding that information to other nodes [3][4]. Wireless distributed sensor and actor network consists of randomly deployed wireless sensors nodes in densely or sparsely manner and a base station. Wireless Sensor Network uses the sensor nodes for sensing the environmental entities like humidity, noise level etc. periodically or on demand and propagates these sensed data to the base station. These sensor nodes are equipped on an on board processor. Each sensor node consists of transceiver with an antenna, processing unit (processor and storage), and sensors for sensing the desired features, detecting events, sensing location of event ID
and controlling actuators locally[5]. Sensors have analog to digital converter and some optional parts like location finding system, mobilizer and actuator.

The prime idea about such kind of WSN is anytime and anywhere computing paradigm with sensing, computing, communication etc. elements i.e.


Energy consumption for transmitting 1-bit data [6] over a distance $d$ in WSNs can be easily understood by following formula:

$$E = \begin{cases} 
E_{elec} \cdot 1 + E_{fs} \cdot d^2 & d < d_0 \\
E_{elec} \cdot 1 + E_{mp} \cdot d^4 & d => d_0 
\end{cases}$$

(1)

Where $d_0$ is threshold value, we use free space ($E_{fs}$) model if receiver and transmitter lies between distance $d_0$ i.e. we consider direct communication between sensor node and base station else we use multipath fading ($E_{mp}$) model i.e. we consider energy dissipation due to presence of multiple path for propagation and reflections of another nodes in network. Once the network is deployed, nodes continue sensing the information and the battery power decreases exponentially. Sensor node uses exterior power source which has restricted energy. So to increase the lifetime of overall network it is required that each nodes utilize their power in most efficient way.

Clustering [7] provide one of the best ways to achieve better utilization of energy of sensory nodes. As given in Fig [1] the entire network is divided into certain number of clusters and for each cluster we select a leader called as CH whose job is to aggregate overall cluster information and then forwarding it to base station. Users can access network information from base station via internet which is center of network [8]. As each CH is located at different distances from base station they cause imbalance in energy consumption. Imbalance energy consumption creates network partitions and which causes some nodes to die soon due to hot-spot problem.

**FIGURE 1 CLUSTERING IN WIRELESS SENSOR NETWORK**

But our desired network should be in such a way that most of the nodes die together so that it can be replaced together. Other desirable characteristics of this network are fault tolerance, stabilized network topology, and reliability etc. By using clustering energy efficient routing can be performed for inter or intra communication in WSNs.

**BENEFITS OF CLUSTERING:**

- **a)** Decreases the size of routing table by restricting the route arrangement in a cluster.
- **b)** Consumption of less energy in communication
- **c)** Enhance lifetime of particular sensor
- **d)** Do not require fixed infrastructure

In past decades many approaches of clustering has been proposed and main focus of each approach is to improve energy efficiency. Clustering approaches can be classified by considering overall network topology or operational objectives or desired number of generated clusters etc.
**Fuzzy logic**: Fuzzy logic is basically many valued logic where variables can be any number between 0 and 1 unlike Boolean logic where values of variables can only be 0 or 1. Fuzzy logic is useful to handle the concept of partial truth where it is not necessary that value is either true or false there is a possibility of being partially true. Such concepts can easily be handled by fuzzy logic and in WSN fuzzy logic can be used to find the probability of being CH y considering many parameters where a node need not fall in any one category it may partially belong to two or more categories. Complete fuzzification method is explained in fig[2].

In section II various clustering algorithms proposed for WSN has been explained. In section III a comparative analysis of each algorithms have been done based on different parameters. In section IV we conclude current clustering technologies and give some future research options.

**RELATED WORK**

**LEACH (Low Energy Adaptive Clustering Hierarchy)** The first effort in the area of hierarchal clustering the sensor nodes in WSNs is given in LEACH presented by Hein Zelman [9] which uses TDMA-based MAC protocol. Main idea behind LEACH is to move the cluster head over whole network for better load distribution. In LEACH data fusion is also performed to reduce data transmission load. Cluster head selection for each group is totally based on probability.

In LEACH whole operation is divided into rounds and in every round the algorithm execute two phases first is setup phase and second is steady state phase. In setup phase each node calculates its probability by generating a random number within the range of 0 and 1. If a nodes random number is less than a threshold the node is selected as CH and send CH-ADV message to all other nodes. The nodes that have value higher than threshold are not selected as CH and it send join message to the nearest CH. LEACH protocol ensures that each node will become CH once in 1/P rounds, where p is the required number of cluster heads in the network.

All messages in LEACH use CSMA (Carrier Sense Multiple Access) for transmitting data either from node to CH or from CH or base station. LEACH has the chief priority because of its capability of providing load distribution, energy efficient solution and scalability. But LEACH protocol has some critical issues as it is totally based on probability. Residual energy of a node is not considered so node having low energy can be selected as CH. It is also possible that elected CHs are very close to each other which cause congestion in particular areas. LEACH uses single hop communication that’s why it is not very useful in large networks.

**PEGASIS (Power-Efficient Gathering in Sensor Information System)** is a Chain-based protocol given by Lindsey et al. [10] which is an advancement of the LEACH protocol. It focused on data gathering and gives the idea that energy saving can also be achieved without forming clusters. In this algorithm a chain is established from farthest node to nearest node to the base station. Each node sense data merge it with its own data and create a packet of same size then it will forward the data to its nearest neighbor. Finally the node nearest to the base station will send that processed data to the base station.
PEGASIS performs better than LEACH as it eliminates the requirement of forming dynamic clusters and also it provide better load balancing as fusion is done at each node rather than creating overhead at a single node which tends to that all nodes will die nearly at the same time. But PEGASIS has some limitations as it requires prior knowledge of network topology which is quite impractical in large distributed systems and also if a single node dies it will break the whole network which will reduce reliability.

**M-LEACH (Multi-hop LEACH)** LEACH is not very feasible when the network diameter is very large as it uses single hop to communicate with base station which causes high energy dissipation but Multi hop LEACH proposed by Aslam et al. [11] uses multi hop path for communicating with base station that reduces energy consumption at cluster heads. M-LEACH is made to use in heterogeneous environment where sensor nodes are of different capacity level. Except the communication method between base station and cluster heads rest of the protocol is same as LEACH. Like LEACH this algorithm is divided into rounds and each round has two phases. In setup phase appropriate nodes elect itself as cluster heads and other nodes join nearest cluster head.

In steady state phase nodes in every cluster send data to their CHs, cluster heads aggregate that data and then forward it to base station either directly or using other intermediate cluster head. In multi hop leach two types of communication takes place first intra-cluster communication and second is inter-cluster communication. In first type of communication nodes communicates with cluster heads and cluster head forward cluster data to the base station using single hop communication. In second type of communication when base station and CHs are at large distance, CHs finds best path that requires minimum hop-count between first cluster head and base station. Multi-hop LEACH performs better than LEACH in large distributed system as energy dissipation for communication with base station is reduced by transmitting through minimum cost path.

**TEEN (Threshold Sensitive Energy Efficient Sensor Network Protocol)** given by Manjeshwarand et al. [12] is an event driven approach in which data to the base station is send only if some event occurs. TEEN uses two types of threshold namely soft and hard. When a node sensed soft threshold that may be because of some change in attribute or no change it switches to transmitter without having any report for cluster head. When node achieves hard threshold it switches to transmitter and forward the data to the cluster head. In TEEN hierarchy of cluster head is formed which reduces number of transmissions to the base station. TEEN is good to use in applications which are time-concerned. But it has some flaws too as node respond only if they achieve the threshold but if a node dies the user will not be aware of it and forced to wait to get response of that node. So TEEN is not good to use in applications that require periodic update.

**APTEEN (Adaptive Threshold Sensitive Energy Efficient Sensor Network)** given by the authors of TEEN Manjeshwar et al. [13] is an algorithm that provides improvement over LEACH and TEEN. APTEEN combines both periodic approach of LEACH and event driven approach of TEEN to provide solutions to the problem occurring in TEEN. In this algorithm CHs are elected by base station. As soon as CHs are selected these CHs broadcast a message to its members that contains four parameters- thresholds, schedules, count time and attributes. Based on this information nodes sends its sensed data to the base station only if it is satisfying hard threshold. Nodes that have not forwarded data for pre-specified time as informed by CHs will immediately sense the environment and then forward the data to the CHs. These four parameters make this algorithm very flexible as we can get our requirement by adjusting these parameters. APTEEN is good for the applications that require periodic update but it increases the complexity as it imposes additional threshold function and count time.

**SEP (A Stable Election Protocol for clustered heterogeneous wireless sensor networks):** Proposed by Georgios Smaragdakis et. al.[14] is an algorithm that utilizes the heterogeneity of network and make CHs among the nodes which are equipped with some extra power. LEACH protocol does not consider the heterogeneity of nodes that is energy of nodes are different initially so energy consumption cannot be optimized in the case where heterogeneity exists. So in this paper author proposed an approach where heterogeneity of node is considered and based on that nodes are divide into two parts namely- Normal nodes and advanced nodes. Normal nodes have lesser energy as compared to advanced nodes. Advanced nodes will more often get selected as CH as compared to normal nodes and all the nodes will generate random number to calculate threshold value. Author presents comparison with
LEACH protocol considering the factor First Node Dies (FND) which is improved over LEACH as normal nodes will die soon due to lesser energy and in last only advanced nodes will be left in network.

**PROPOSED WORK**

In this chapter, we propose an algorithm for cluster-head election in wireless sensor networks that is designed for heterogeneous network. We also provide a comparative analysis of our proposed approach with SEP which is one of the best protocols in the field of clustering for heterogeneous network.

Our algorithm is implemented in a heterogeneous environment where all nodes will not initially have same amount of energy. There are two characteristics parameters of heterogeneity i.e. namely the fraction of advanced nodes (m) and the additional energy factor between advanced and normal nodes (α) i.e. advanced nodes will have more energy than normal nodes by a factor of α. Like SEP our algorithm attempts to maintain the constraint of well-balanced energy consumption [14]. Advanced nodes have to become cluster heads more often than the normal nodes, which is equivalent to a fairness constraint on energy consumption. Initially each node can become a cluster head with a probability \( P_{\text{opt}} \) i.e. our protocol guarantees that every one of them will become a cluster head exactly once in every \( \frac{1}{P_{\text{opt}}} \) rounds. Suppose that each normal node will have energy \( E_0 \) and advanced nodes will have energy \( E_0 \cdot (1+\alpha) \). Then the total energy of this heterogeneous WSN will be equal to:

\[
E = n \cdot (1-m) \cdot E_0 + n \cdot m \cdot E_0 \cdot (1 + \alpha) = n \cdot E_0 \cdot (1 + \alpha \cdot m)
\]

Where \( n \) is the total number of nodes and \( m \) is the fraction of advanced nodes. So fraction of normal nodes can be obtained by \( n(1-m) \). Each normal node will become CH once in \( \frac{1}{P_{\text{opt}}} \) rounds while each advanced node becomes a cluster head exactly \( 1 + \alpha \) times in every \( \frac{1}{P_{\text{opt}}} \) rounds.

The algorithm will be divided into rounds and in these rounds CHs will be elected by using the threshold value and the priority value. Threshold will be calculated by the following formula:

\[
Th = \frac{P_{\text{opt}}}{1 - P_{\text{opt}}(r \mod \frac{1}{P_{\text{opt}}})}
\]

Where \( r \) is the current round number and \( P_{\text{opt}} \) is different for normal nodes and advanced nodes and this can be calculated by following formula:

\[
P_{\text{norm}} = \frac{P_{\text{opt}}}{1 + \alpha \cdot m}
\]

\[
P_{\text{adv}} = \frac{P_{\text{opt}}}{1 + \alpha \cdot m} \cdot (1 + \alpha)
\]

The motivation behind our protocol is that using important and effective parameters for selection of CHs will lead to provide energy efficient protocol. For selecting a CH we will combine several factors to find the probability of each node being CH. In this protocol we will prove that using fuzzy logic can reduce gathering data and calculation overheads. Thus, the lifetime of the sensor network can be extended. Our clustering algorithm is aimed to overcome the significant limitation of SEP as it is pure probabilistic approach.

Cluster-heads produced by SEP may be located at the edges of network and may be possible that CHs are too close to each other[14]. To overcome shortcomings of SEP we propose a protocol that uses three parameters to elect a CH. Like SEP in our protocol there are two types of nodes in the network i.e. advanced nodes and normal nodes.

In our proposed work we are implementing fuzzy logic with SEP (Stable Election Protocol). Our algorithm will use two parameters: Distance and Residual energy for calculating the chance value of a node for being CH for that particular round.
Distance to base station will be calculated by using the mathematical formula of distance between two points.

\[ d_{\text{to BS}} = \sqrt{(x - x_b)^2 + (y - y_b)^2} \]  

(6)

where \( x \) is X coordinate of node and \( y \) is Y coordinate of node and \( x_b \) is X coordinate of base station and \( y_b \) is Y coordinate of base station.

Residual energy of nodes will be calculated by the formula given in equation (1) by replacing the variables with the values. Now a set of fuzzy if-then rules will be defined to find the optimal \( P \) value that will be used for threshold calculation in further rounds. Rules that we are using in our simulation are given below:

<table>
<thead>
<tr>
<th>Distance</th>
<th>ResEnergy</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>close</td>
<td>low</td>
<td>average</td>
</tr>
<tr>
<td>close</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>mid</td>
<td>medium</td>
<td>average</td>
</tr>
<tr>
<td>mid</td>
<td>high</td>
<td>Low</td>
</tr>
<tr>
<td>mid</td>
<td>high</td>
<td>Average</td>
</tr>
<tr>
<td>far</td>
<td>low</td>
<td>High</td>
</tr>
<tr>
<td>far</td>
<td>medium</td>
<td>High</td>
</tr>
<tr>
<td>far</td>
<td>high</td>
<td>High</td>
</tr>
</tbody>
</table>

The membership function we use for the distance variable and residual energy is Gaussian membership function and for output variable priority we use triangular membership function. The numeric values corresponding to each membership functions is taken as follows-

![Figure 3: Membership Function for Input Residual Energy](image3.png)

**Figure 3** Membership Function for Input Residual Energy

![Figure 4: Membership Function for Input Distance](image4.png)

**Figure 4** Membership Function for Input Distance
FIGURE 5 MEMBERSHIP FUNCTION FOR OUTPUT PRIORITY

**ALGORITHM**

1: START
2: Initialize parameters
   - \( n \leftarrow \) number of nodes
   - \( m \leftarrow \) fraction of advanced nodes
   - \( \alpha \leftarrow \) additional energy factor between advanced and normal nodes
   - \( P_{opt} \leftarrow \) Optimal value of cluster-heads
3: Define \( E_0, d_0, E_{tx}, E_{rx}, E_{fs}, E_{DA} \) energies.
4: \( \text{countCH} = 0 \)
5: \( r = 0 \)
6: \( \text{ResEnergy for normal nodes} = E_0 \)
   - \( \text{ResEnergy for advanced nodes} = E_0(1 + \alpha) \)
7: for \( \text{round} = 0 \) to \( r_{max} \)
8:   Calculate threshold \( (Th) \) using equation 3, 4 and 5.
9:   \( \mu \leftarrow \text{rand}(0,1) \)
10: if\( (\mu < = Th) \)
11:     \( r = \sqrt{A/(\pi \times n \times P_{opt})}) \)
12:   Calculate amount of energy dissipated by using equation 1
13:   \( \text{ResEnergy} = \text{ResEnergy} – \text{Energy Dissipated} \)
14:   Calculate distance parameter of nodes by using equation 6
15:   Calculate Priority using fuzzy if-then rules
16:   if\( (\text{Priority} > = \text{adj_priority}) \)
17:     \( \text{countCH} = \text{countCH} + 1 \)
18: endif
19: endif
20: endfor
21: END

**SIMULATION RESULT**

In this chapter, we present the results of the experiments that we have done to evaluate our algorithm. We compare our clustering algorithm MCF with SEP algorithm. We have implemented a wireless sensor network clustering simulator to evaluate our algorithm. We have run several experiments on this tool to evaluate our algorithm. We
have executed our algorithm for 1000 rounds and take 200 nodes in our network. Experimental results have shown that our algorithm performs better than SEP protocol on two parameters- Average energy and Number of Dead nodes. We have calculated the average of energy and number of dead nodes in each 250 rounds.

**TABLE 2. SIMULATION PARAMETERS**

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network size</td>
<td>100x100</td>
</tr>
<tr>
<td>Base Station Location</td>
<td>(100, 50)</td>
</tr>
<tr>
<td>Number of Sensor nodes</td>
<td>200</td>
</tr>
<tr>
<td>Initial Energy</td>
<td>$E = 0.1$ joule</td>
</tr>
<tr>
<td>Data Packet Size</td>
<td>4000 bits</td>
</tr>
<tr>
<td>Transmit/Receive Energy</td>
<td>$E_{\text{elec}} = 50\text{nJ/bit}$</td>
</tr>
<tr>
<td></td>
<td>$E_{\text{mp}} = 0.0013\ \text{pJ/bit/m}^4$</td>
</tr>
<tr>
<td></td>
<td>$E_{\text{fs}} = 10\ \text{pJ/bit/m}^2$</td>
</tr>
</tbody>
</table>

In our simulation initial energy of a normal node is set to $E_0 = 0.4$ Joules, $m=0.2$ i.e. 20% of the nodes are advanced and $\alpha = 3$ i.e. advanced nodes have 30% more energy than normal nodes.

**FIGURE 6 AN INSTANCE OF THE CLUSTERED NETWORK**

**FIGURE 7 DISTRIBUTION OF NUMBER OF DEAD NODES AFTER 500 rounds**
CONCLUSION AND FUTURE WORK

Conclusion

In this thesis we have proposed a heterogeneous clustering algorithm that uses fuzzy logic to combine two parameters i.e. distance and residual energy. Here we have observed various hierarchal clustering algorithms considering in which conditions these algorithms are suitable to use and where it is not providing efficient result. In wireless sensor network limited energy is a crucial issue. Many researchers have been done to give different
approaches of reducing energy consumption in various networking environment. Clustering is one of these approaches that achieve better energy utilization by dividing the whole network into various clusters. Hierarchical clustering involves re-clustering of nodes in various time interval which improves reliability of nodes. Re-clustering let the base station aware of nodes capacity and provide more reliable network by informing dead and low energy nodes. We have taken heterogeneous environment where some nodes called advanced nodes are equipped with some extra energy so that we can get better distribution of nodes after some rounds when nodes started dying. We have implemented fuzzy logic along with the heterogeneous clustering protocol and get better distribution of nodes and compare the algorithm considering two parameters average energy and number of dead nodes.

Future Work

From this thesis it has been concluded that implementing fuzzy logic for clustering is a good way to use effective parameters for the selection of CHs. Heterogeneity of nodes most cogently exist in the network so it is better to use heterogeneous network for your work to deal with the energy efficiency problem. Power utilization of sensor nodes can further be optimized using other important parameters for cluster head selection. Another parameters that can be used in fuzzy logic are node centrality, node concentration, communication cost etc. Using these parameters will optimize the energy efficiency in clustering.

REFERENCES
