Distributed Energy Efficient Clustering Algorithm for Wireless Sensor Networks

Muruganantham Arunraja, Veluchamy Malathi, Erulappan Sakhivel

Department of Electrical and Electronics, Anna University, Regional Centre, Madurai

Abstract: Wireless sensor networks (WSN) are powered by finite energy source like batteries, which imposes stringent energy boundaries. Clustering of nodes avoids redundant message transmissions over the network. Thus conserves energy, communication bandwidth and achieves scalability. Here we propose a Distributed Energy Efficient Clustering Algorithm (DECA), a deterministic clustering approach where Cluster Heads (CHs) are selected based on their residual energy and priority using passive clustering technique. Nodes then associate with the CH with least communication cost and high residual energy. This method achieves longer life span than the previous energy efficient clustering algorithms in both homogeneous and heterogeneous networks. DECA has extended the WSN life span to 30% in the homogeneous environment and 50% in the multi-level heterogeneous environment.

Keywords: Wireless sensor networks; energy conservation; distributed clustering

1 Introduction

As the wireless sensor nodes are self-contained systems with limited energy resource, the life span is finite. By conserving the on board energy, the system life span can be increased, hence ensuring continued service. Clustering is an efficient hierarchical routing approach to conserve energy, where only a set of nodes performs the routing operation and therefore reducing the amount of redundant transmissions and collisions.

Due to its reduced deployment cost, smaller form factor and increased computing power, WSN are an inevitable element of today’s industrial, military and medical establishments. Due to their ease of installation and lesser maintenance they proliferated to various disciplines. WSN consists of autonomous sensor nodes that are spatially distributed over the specific region to report about one or more parameters of interest. In a conventional WSN, densely deployed nodes sense data from the environment, pre-process it and communicate it to the sink node.

As the sensor nodes are self-contained systems with limited energy resource, the life span is finite. Most of the WSNs are deployed in hostile environments, where recharging is a difficult task. By conserving the on board energy, the system life span can be increased, thus ensuring continued service.

The tenacity of WSN is to collect accurate data for longer time span. Data accuracy is improved by increasing sampling frequency. Increased sampling
frequency, results in increased energy cost, thus reduces the life span. This imposes a conflict between data accuracy, and energy efficiency. The solution for the problem is derived by estimating the temporal correlation between the node data in time domain. The temporal correlation intimates the amount of redundant data communicated. By reducing the redundant data both accuracy and energy efficiency can be achieved. Temporal correlation can further be exploited to predict the future data from historical data. By effectively modeling the temporal correlation of a measured signal, a substantial amount of data transmission can be reduced.

In the physical space nearer sensors measure almost equal values due to their close proximity. In a denser network, nodes generate a hefty amount of spatially redundant data. Clustering of nodes can filter out the spatially redundant data locally. Clustering suffers inefficiencies when two or more CHs happened to be in close proximity or when the data load is not equally balanced between CHs. In certain clustering approaches, the clustering process itself consumes significant energy, due to its communication overheads.

The proposed clustering approach addresses the above mentioned problems, thus makes clustering more efficient. Here the clustering is initiated by a node that proclaims itself as CH. The neighbors of the node join the cluster as cluster members. Each node would like to become CH, but the proclamation delay of a node is inversely proportional to its weight. This passive clustering process reduces the clustering overhead, by avoiding data sharing between contending nodes. The cluster size is increased with respect to the distance from the base station, so as to balance the load between near and remote CHs. The one hop neighbors of the CH are disqualified from the contention to become CH, to avoid the formation of nearby CHs. Then from the rest of the nodes highest weight node initiates clustering. This continues until all the nodes are clustered.

Clustering suffers inefficiencies when two or more CHs happened to be in close proximity or when the data load is not equally balanced between CHs. In certain approaches, the clustering process itself consumes significant energy, results in less efficiency. An ideal clustering process should have least overheads and should assure energy efficient data collection along with uniform distribution of CHs.

2 Related works

Numerous works done on energy efficient clustering of WSN, yet most of the work concentrates on energy efficiency and data accuracy. Long term coverage of region of interest is not addressed in any of the works. The distributed sensor network is said to be efficient only if it covers a larger portion of the region of interest for a longer time span. The proposed work considers energy efficient coverage as its major point.

Energy conservation in WSN is analyzed by different approaches [1]. Some methods take up energy efficient routing as one of the tools for energy balancing between wireless nodes [2]. Some methods employ MAC based schemes, which ON/OFF the node circuitry in appropriate time slots so that energy expense can be drastically reduced [3].

Spatial correlation among the sensor nodes is used to reduce the data transmission by employing clustered aggregation schemes. Different kinds of clustering methods are listed by Younis et al [4]. Clustering is done in WSN in different application to achieve different goals.

Clustering may be centralized [5], where the base station collects the node parameters and announces the cluster heads and their cluster members. This one is effective for the small scale network. For large scale networks, distributed clustering methods [6] suit well, where the clustering, decision is made by individual sensor nodes by acquiring only local information from one hop neighbors. Clustering may also consider a single hop [7] or multi hop networks [8]. Usage of heterogeneous nodes [9] reduces the overhead during CH election.

LEACH [7] is one of the earliest and most sought clustering methods for energy efficiency WSN. The problem with leach is a random selection of CH which results in proximate CHs and unbalanced load distribution. There are other works [10], which depicts the node degree as one of the selection criteria for CH selection.

In [11], clustering is done by identifying similarity between sensor data. The nodes send their data to the base station, where the nodes are assigned to the appropriate cluster based on data similarity.

Different clustering approaches use different parameters for CH selection, based on the selection criteria, the algorithms can be divided. Deterministic algorithm [12] takes into account of node degree and node id etc., where these parameters are fixed during the deployment itself. Adaptive algorithms [13] consider the residual energy of node, mobility etc. These parameters vary over the operational span of the network. Hybrid algorithms [14] consider both the
deterministic and adaptive parameters for cluster head selection.

Several works done on exploiting temporal correlation of data, where linear regression methods are used to calculate the temporal relation between time series generated by sensors [15]. The schemes suffer from reduced accuracy and lack of adaptability towards dynamic variations in input signal. In [16], ARIMA based methods are used to predict the sensor data from previous values. Saintini et al [17] proposed a dual prediction scheme using LMS based adaptive filters. The advantage of this method is its ability to converge the prediction, without any prior model [18]. The scheme can still be improved by adapting the internal parameters.

Most of the existing literature either concentrates on spatial correlation based data aggregation or on temporal correlation based data reduction. DECA exploits spatio-temporal correlation between sensory data for energy efficient data collection and also it maintains maximum level data integrity. The work involves simple and computationally light weight algorithms, which makes the implementation simpler and less energy consuming in terms of computational cost. To achieve energy efficiency, numerous clustering approaches have been proposed. LEACH [7] selects CH on a random probability, which results in proximate CHs and unbalanced load distribution. SEP [19] assigns a weighted probability to each node based on its initial energy, where advance nodes get more chances to become a CH. In DEEC [20], CH formation is based on the residual energy of the entire network and residual energy of the node that wants to become a CH. Since SEP and DEEC are probabilistic clustering approaches, there might be cases when two CHs are selected in close vicinity of each other, increasing the overall energy expense of the network. Secondly, the number of CH nodes generated is not fixed, hence in some rounds it may be more or less than the preferred value. Deterministic protocols eliminate uncertainty over the number of CHs and CH election. The weighted clustering algorithm [21] selects a node as a CH based on its weight. The weight of an individual node signifies the remaining battery energy, degree, mobility, and distances with the neighbors. Distributed clustering algorithm (DCA) uses application based weights of a node to select it as the CH [22]. Both the algorithms necessitate extensive control messages to construct a cluster, which makes them suitable only for a small wireless sensor network.

DECA is an energy aware passive clustering algorithm, where the clustering is performed without communicating nodal parameters. The DECA algorithm considers only self-estimating parameters for CH selection, thus a single announcement is enough to select a CH. The selection parameters are highly dynamic and support in better distribution of the load among sensor nodes. This forms energy efficient clusters with minimal clustering overheads.

3 Deca

The proposed DECA constructs the energy efficient clusters in two phases. The DECA consists of the following functional modules:
- An energy aware passive clustering approach that reduces clustering overheads and assures uniform energy distribution.
- A node association approach based on residual energy and communication cost of a CH.

The proposed work makes the following assumptions before designing the energy efficient protocol.
- Topology is static
- All nodes are aware of their location
- All the cluster members can reach CH in one hop
- CH can reach the base station in one hop or multiple hops

The actual purpose of creating clusters is to reduce the energy consumption of the sensor nodes and the bandwidth requirement for the network. The clusters in the network are attributed by a single CH, connected to multiple sensor nodes nearby. The sensor nodes transmit their data to the cluster head, which aggregates the data and forwards it to the base station. This process moderates the energy expense of the nodes and decreases the probability of data collisions.

The energy model of the conventional cluster is given as

\[ \sum_{i \in N} \sum_{j \in N} (l y_{ij} (D_j + E + E_{DA}) + l_{ag} F_j x_i) \]  

Where \( y_{ij} = 1 \), if node \( j \) is a cluster member of the CH node \( i \).
\( y_{ij} = 0 \), otherwise.
\( l \), the length of data packet from cluster member to CH.
\( E \), energy spent for receiving a data bit.
\( E_{DA} \), data aggregation energy
\( l_{ag} \), length of aggregated data packet
\( x_i = 1 \), if node \( i \) is a CH.
\( x_i = 0 \), otherwise.
The data transmission energy from the cluster member to CH is given as

\[ D_{ij} = \begin{cases} E + \varepsilon_f d_{ij}^2 & \text{if } d_{ij} < d_0 \\ E + \varepsilon_{mp} d_{ij}^4 & \text{if } d_{ij} > d_0 \end{cases} \] (2)

Where \( d_{ij} \) the distance from sensor node ‘i’ to sensor node ‘j’.
\( d_0 \), the threshold distance.
\( \varepsilon_f \), amplifier energy
\( \varepsilon_{mp} \), amplifier energy

The data transmission energy from CH to Base Station is given as

\[ F_i = \begin{cases} E + \varepsilon_f f_i^2 & \text{if } f_i < d_0 \\ E + \varepsilon_{mp} f_i^4 & \text{if } f_i > d_0 \end{cases} \] (3)

Where \( f_i \) the distance from CH node ‘i’ to Base station.

### 3.1 Energy Efficient CH election

In the proposed work, higher energy nodes are elected as CH to attain energy efficiency. To distribute the load uniformly among the nodes, node priority is considered. Both the components constitute for the weight of the node. To reduce the energy consumption during clustering process, the passive clustering method has been proposed, where the proclamation delay is defined as the function of node's weight.

#### 3.1.1 Weight Calculation of Node

In this work, each node is assigned with specific weight based on its suitability to become CH. The election of the CH is done on the basis of the largest weight among the neighbors. This means that a node becomes a CH or a cluster member, depending on its own and one hop neighbor’s weights.

In WSN, there are several heuristics for selecting CHs. Node energy, node degree, distance from the Base Station(BS), node ID and cumulative time for which the node acted as CH (node priority) are the prominent heuristics. Since CHs are overloaded with multiple tasks, the rate of energy depletion is also high. Hence the CH elected should have high residual energy. Thus, for an energy efficient clustering approach, residual energy is the major indicator for the dominant set of nodes in both homogeneous and heterogeneous (energy) networks. Acquiring residual energy is again an internal task, doesn't need any communication.

#### 3.1.2 Passive Clustering

In passive clustering, Cluster head is selected based on “first declaration wins” rule, therefore the node that first proclaims becomes the CH. In the earlier works, the proclamation delay is random. Random delay may results in selection of low energy nodes as CH, hence decreases the energy efficiency of the system. In the proposed work, to elect most appropriate nodes, the proclamation delay is made inversely proportional to the node’s weight. Once the proclamation delay expires, the node proclaims itself as CH. If a node hears a proclamation before the expiration of its proclamation delay, it refrains itself from the contention to become CH. The waiting time \( T_w \) of node \( n \) is given as

\[ T_w (n) = \frac{K}{E_{res}(n)} \] (4)

Where \( k \) is a constant;

### 3.2 Node Association

When the node proclaims as CH, it advertises its residual energy and location information along with the proclamation. If a non CH node receives multiple proclamations, the advertisement serves as a key factor for selecting the most appropriate CH. The proclamation message format is indicated in table.1. Based on this information, a non-CH node estimates CH’s compatibility towards it.

#### Table.1: CH proclamation format

<table>
<thead>
<tr>
<th>CH ID</th>
<th>CH Residual Energy</th>
<th>CH Location</th>
</tr>
</thead>
</table>

Due to diverse communication overheads and functionality, there is an asymmetry of energy between various nodes. Hence a good design for data collection should not put heavy burdens on low energy sensor nodes. Instead, the heavy burdens should be assigned to the high energy sensor nodes. The CH residual energy component helps in identifying the CH with highest energy near the sensor nodes.

If the CH is situated between the sensor node and BS, the communication cost will be reduced considerably. The CH location component helps in calculating the communication cost of the node to forward the data to BS through that CH. Thus the CHs location information has a crucial role in node association. Based on this information a non-CH node selects a CH, so that the total data transmission distance is minimized. The communication cost through a CH is estimated from the node to CH distance \( d_{loc} \) and CH to BS distance \( d_{Ch2BS} \). The
goal of the node associated process is to minimize the total distance. \[ \text{Min} \left( d_{\text{nCH}} + d_{\text{CHtoBS}} \right) \]

The node selects CH with highest residual energy and lowest communication cost. The CHs aptness is the ratio between the residual energy of the CH and communication cost through this CH to BS.

If a node ‘i’ receives ‘n’ CH proclamations, the node estimates the weight of each CH as

\[
W_{\text{CH}}(n_i) = \frac{E_{\text{res}}(n)}{F(n_i)}
\]

Where \( F(n_i) \), communication cost of node ‘i’ through CH (n) to BS.

\[ F_{n_i} = LE_{\text{elec}} + L \cdot E_{\text{RF}} d^2 \]

\( L \): size of data packet \\
\( E_{\text{elec}} \): Energy consumed by RF Module \\
\( E_{\text{RF}} \): RF Amplifier energy \\
\( d \): Total distance from the node to BS

The node then associates with the highest weight CH in its vicinity. The CH with minimum residual energy is elected by very few nodes. This avoids repeated loading of few CH nodes.

4 Results and discussion

Here we evaluate the performance of DECA protocol using MATLAB. The proposed work has been simulated on a network consists of 100 randomly distributed nodes in a 100m X 100m field. We assume the base station is in the center of the sensing region. The radio parameters used in our simulations are shown in Table 2. The proposed DECA approach is compared with other energy efficient clustering protocols like LEACH, SEP and DEEC. The life time of nodes and total packets transmitted are estimated on both homogeneous and heterogeneous nodal energy levels. The life time is classified into stable period (until first node dies) and lasting period. The effectiveness of the clustering is measured by the ratio between stable period and unstable period.

Table 2: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E )</td>
<td>50nJ/bit</td>
</tr>
<tr>
<td>( E_{\text{elec}} )</td>
<td>10 pJ/bit/m²</td>
</tr>
<tr>
<td>( E_{\text{RF}} )</td>
<td>0.0013 pJ/bit/m⁴</td>
</tr>
<tr>
<td>( E_0 )</td>
<td>0.5 J</td>
</tr>
<tr>
<td>( E_{DA} )</td>
<td>5 nJ/bit/message</td>
</tr>
<tr>
<td>( d_0 )</td>
<td>70 m</td>
</tr>
<tr>
<td>Message size</td>
<td>4000 bits</td>
</tr>
<tr>
<td>( P_{\text{opt}} )</td>
<td>0.1</td>
</tr>
</tbody>
</table>

In a homogeneous environment all the nodes are assigned with equal energy. DECA elects CHs based on residual energy and cumulative time the node acted as CH. Thus, in the homogeneous environment, DECA efficiently balances the load between multiple nodes and achieves longer stable period (Fig.1). The stability period is elongated to 1110 rounds, compared to 910 rounds of LEACH protocol that has the second longest stable period.

Figure 1: Node lifetime in homogeneous network

We then observe the performance of LEACH, SEP, DEEC and DECA under two different two-level heterogeneous networks. Fig.2 shows the performance when fraction of advanced nodes ‘m’ kept at 0.2 and additional energy factor ‘a’ set as 3. In another scenario ‘m’ is set at 0.3 and ‘a’ as 2. The unstable region of SEP and LEACH are also larger than our DECA protocol; it is because the advanced nodes die more slowly than the normal nodes in SEP and LEACH.

Figure 2: Stability period comparison in two-level heterogeneous network

For multi-level heterogeneous networks, the initial energy of nodes is randomly distributed in \([E_0,4E_0]\).
The DECA protocol exhibits longer life span in the heterogeneous energy environment than all the protocols compared. The stability period is elongated from 1500 rounds (DEEC) to 2200 rounds as shown in fig.3. This is because LEACH treats all the nodes without discrimination. SEP has longer stability period than LEACH just because of discriminating nodes according to their initial energy. DEEC take initial energy and residual energy into account at the same time. The longer life span of DECA is attributed to the energy efficient selection of CHs and minimized data transmission distance.

Figure 3: Node lifetime in multi-level heterogeneous network

We also examine the sensitivity of DECA to the degree of heterogeneity in the network by increasing the ‘m’ value from 0.1 to 0.9 and ‘a’ from 0.5 to 5. Being an energy-aware deterministic protocol, DECA outperforms other energy efficient clustering protocols. The difference between the stability period of DECA and other protocols increases with an increased fraction of advanced nodes ‘me’ as indicated in Fig.4.

Figure 4: Round of First Node Dies for Different m when α= 2

The performance of DECA is observed to be close to that of an ideal upper bound obtained by distributing the additional energy of advanced nodes uniformly over all nodes in the sensor field. DECA is more resilient than LEACH in judiciously consuming the extra energy of advanced nodes. Thus DECA yields a longer stability period for higher values of extra energy as in Fig. 5.

Figure 5: Round of First Node Dies for Different awhen m=0.3

Fig.6. Shows the comparison of every protocol for number of packets that are sent to BS. Result shows that DECA has highest successful data rate, as compare to other routing protocols. DECA achieves highest data rate along with longer stability period.

5 Conclusion

Our contribution involves a deterministic CH election protocol (DECA) that holds the distributed property of probabilistic models. Secondly, a novel node association technique is introduced through which...
nodes join with the most appropriate CHs that are having least communication cost and high energy. The node association part, assures load is taken up by only high energy nodes. The network's energy expenses are reduced by shorter data paths. Thus DECA achieves uniform distribution of CH and longer stability period. DECA has extended the WSN life span to 30% in the homogeneous environment and 50% in the multi-level heterogeneous environment. The work shows good results in various kinds of heterogeneous environments. Future work involves construction of multi-level clusters using temporal correlations between sensor data.

6 References


