Query based Dynamic Clustering in Wireless Sensor Networks

T. Sujithra
Department of Computer Science and Engineering
PSG College of Technology, Coimbatore, India

R. Venkatesan
Professor
Department of Computer Science and Engineering
PSG College of Technology, Coimbatore, India

ABSTRACT
Cluster based data gathering is one of the approaches to reduce the energy consumption in wireless sensor networks. In this dissertation, we propose query based dynamic clustering algorithm to improve the lifetime of the sensor network. The main idea of this paper, is to increase the network lifetime by clustering the nodes in dynamic basis. The nodes who satisfy the query sent by the base station that are only allowed to remain in the transmit mode and to participate in the clustering process. Remaining nodes go to the sleep mode immediately. Thus transceiver scheduling increases the network lifetime. In which relay nodes are used as routing element, it collects the aggregated information from the cluster head and transmit to the base station via other relay node in a multi-hop fashion. It balances the load of the network. It also reduces the packet loss because of data traffic. From the simulation results, we show that the proposed approach outperforms the existing protocol in terms of increased lifetime and decreased energy consumption.

Keywords
Multi-hop communication, cluster centroid, network lifetime, cluster head, relay node.

1. INTRODUCTION
Sensors are used to measure physical parameters of the environment. It comprises of sensor and actuator, processing equipment, small size memory and transceiver [1]. In a wireless sensor network, as sensor nodes are smaller in size they are densely deployed to cover a large geographical area. For example forest fire monitoring, habitat monitoring, landslide monitoring, etc. Lifetime of the sensor node is the time at which the first sensor node in the network runs out of energy [2]. The lifetime of the wireless sensor network can be improved by various ways, such as efficient routing, effective clustering, cluster head election strategy and efficient sleep scheduling algorithm, etc. Sleep scheduling algorithm deals with the scheduling of the transceiver states [3]. The transceiver has three states, namely idle, transmit/receive and sleep state. In an idle mode, most of the parts in the transceiver are switched off and only significant parts of the transceiver are in active mode. It consumes only less amount of energy than the transmit/receive mode. In sleep mode, all the parts of the transceiver are switched off. Hence there is no energy loss at all. Meticulous scheduling of transceiver states will improve the lifetime of the sensor network. Energy efficiency plays a pivotal role in wireless sensor network as it affects the lifetime of the sensor network. Furthermore, once the sensor nodes are deployed, it is hard to recharge due to either the huge volume of sensor node deployment or human non-intervention characteristics of the deployment area. In a wireless sensor network, most of the energy is consumed for communication purpose. The amount of energy consumed in wireless communication is directly proportional to the communication distance. Therefore, the long distance communication is not preferred. Usually, the data collection in large geographical area is accomplished via multi-hop communication. Furthermore, the expenditure of energy in multi-hop communication is heterogeneous as nodes near the base station (BS) drain out very quickly. Because, these nodes are forced to forward the data of other nodes in the network to BS. In order to address the non-uniform consumption problem among the sensor nodes we propose a query based dynamic clustering approach. The main idea is to

- Reduce the energy consumption by controlling the cluster count and its density based on the query sent by the BS.
- Improve the network lifetime by using relay nodes as routing element.

The rest of the paper is organized as follows i) Section 2 describes the related work of this paper ii) Section 3 discusses about the proposed scheme iv) Section 4 deals with the experimental results and discussion and v) Section 6 gives the conclusion.

2. RELATED WORK
This section discusses about the works that are under study leading to fresh estimates. Based on the sensor node deployment, wireless sensor network is classified into two, namely homogeneous network and heterogeneous network. Sensor nodes deployed in a homogeneous network having the identical communication capability and equal amount of energy, in which some of the nodes are selected as cluster head, because of energy insufficiency role of cluster head is rotated, however energy consumption is eventually distributed among the sensor nodes. Frequent cluster head re-election gives additional overhead [4] [5]. Heterogeneous network composes of small number of resource rich nodes and large number of energy limited basic sensor nodes. The basic sensor node has only limited communication capability and energy. Resource rich nodes are equipped with powerful transceivers and high battery. In which resource rich nodes are acting as cluster head and basic sensor nodes are acting as cluster members. However, cluster head re-election frequency is minimized by excessive energy. Placing resource rich nodes at appropriate location is difficult [6]. Event driven, query based, time driven and hybrid are possible ways of initiating the data transfer from sensors to the sink [7]. For any data gathering scheme, sensors should send their readings to the sink, it may be either static sink or mobile sink. Mostly in static sink approach, data are transferred from sensors to the sink via multi-hop communication. Due to this, nodes near the static sink deplete its energy quickly. It makes the static sink unreachable. Clustering is a standard approach for achieving
efficient and scalable performance in wireless sensor networks. Some of the clustering approaches are discussed as follows. Sangho Yi et al. [8] proposed power efficient and adaptive clustering hierarchy protocol for wireless sensor networks (PEACH) in 2007. The goal of this paper is to elect the cluster head (CH) based on the characteristics of overhearing. It is an adaptive multi level clustering algorithm. During data transmission, packet is received by all the nodes in the cluster. But the node which is the owner of the packet will be elected as cluster head. Remaining nodes will become a cluster member, even they hear the packet. By this way, cluster head election is carried out. Heinzelman et al. [9] introduced centralized low adaptive clustering hierarchy (LEACH-C) in 2002. In which, BS acts as central coordinator. It takes the responsibility of cluster formation. Cluster head is elected based on the energy level of the sensor node. Node having maximum energy gets a higher chance to get elected as cluster head. The BS knows about the entire details about the network. Jau Yang Chang et al. [10] introduced an efficient cluster based power saving scheme for wireless sensor networks (SECA) in 2012. The main idea is to extend the network lifetime by using uniform cluster locations and balancing the network loading among the clusters. It partitions the entire network into equal sized clusters using K-Means algorithm. Cluster head is elected based on both hybrid energy level of the sensor node and distance between the sensor nodes. Smaragdakis G et al. [11] presented a stable election protocol for clustered heterogeneous wireless sensor networks (SEP) in 2004. It handles the heterogeneous nodes in terms of energy. Cluster head is elected based on the residual energy of the nodes. Two kinds of nodes play an important role in the clustering process, namely advanced nodes and normal nodes. Nodes with highest energy are advanced nodes, whereas nodes with a nominal energy are normal nodes. Advanced node is having higher opportunity to get elected as cluster head than the normal node. Ye M et al. [12] presented an energy efficient clustering scheme in wireless sensor networks (EECS) in 2005. It is mainly designed for periodical data gathering application. Node having maximum residual energy is elected as cluster head in an autonomous manner. It uses distance based method to balance the load. Heinzelman et al. [4] introduced low energy adaptive clustering hierarchy (LEACH) in 2000. Clustering process is done in two phases, namely set up phase and steady state phase. Cluster formation and cluster head election is carried out in the set up phase. Data transmission is carried out in the steady state phase. Role of cluster head is rotated to balance the energy consumption. Steady state phase uses Time Division Multiple Access (TDMA) schedule for synchronizing the sensor nodes and cluster head which gives an effective data communication. During the cluster head election, the energy level of the node and distance between the CH and BS is not taken into account. Parul Bakarania et al. [13] proposed an improved version of the LEACH protocol for lifetime improvement in wireless sensor network (K-LEACH) in 2013. The main idea is to find the center position of cluster by using K-Medoids clustering algorithm. A node which is closer to the center position of the cluster is elected as CH. It avoids the problem of the unreachable node. Sasikumar et al. [14] presented K-Means clustering in wireless sensor networks (K-Means) in 2012. It focuses on finding centroids of the cluster using K-Means algorithm. The cluster head election process uses Euclidean distance and hybrid energy level of the node as selection criteria. Node with minimum distance to the centroid and maximum energy is appointed as CH. At the time of cluster head gets depleted, the node which is nearest to the depleted cluster head with maximum energy is elected as CH.

3. PROPOSED QUERY BASED DYNAMIC CLUSTERING APPROACH

Query based dynamic clustering approach has two fold objective: reducing the energy consumption by controlling cluster count and its density, and increasing the network lifetime by using relay element based routing technique. Below we discuss about the problem identification, network model, energy consumption model and proposed system details. The proposed approach is designed based on the following observations. WSN has three basic components: basic sensor nodes, resource rich nodes and base station. Some of their characteristics in WSN.

- Sensing environment comprises of a large number of basic sensor nodes and limited number of resource rich nodes called relay node.
- The basic sensor node has only limited energy and communication capabilities.
- The basic sensor nodes are acting as data sources. It collects the information from the sensing environment.
- The resource rich node has a powerful transceiver and communication capabilities. It acts as routing element.
- The sensor node cannot be recharged after deployment.
- Base station is stationary. It is fixed at the centre of the sensing environment.
- Location of the relay nodes is known.
- SECA [10] and LEACH [4] are taken as the benchmark system.

3.1 Problem Identification

In existing system clustering approach, all the nodes in the sensing environment are taken into account while forming cluster. IE., the nodes that are close to each other are grouped together. It increases the cluster count and its density in large scale data gathering and reduces the lifetime of the sensor network. In our algorithm, the number of nodes that are participating in the clustering process is controlled based on the query sent by BS. The node who satisfies the query which is only allowed to participate in the clustering process. Others go to the sleep mode immediately. It reduces count of clusters formed and increase the network lifetime. The cluster head election is very important because inappropriate cluster head election leads to re-clustering. It affects the lifetime of the sensor network. Our algorithm elects the cluster head based on the distance and energy level of the node. Node having maximum energy level and closer distance to the centroid of the cluster is elected as cluster head. K-Means algorithm is used to find the centroid of the cluster. The cluster head is in the proximity of BS is taken as first level cluster head, and it uses a free space model to transmit data to BS. The cluster heads, which are out of BS proximity is taken as second level cluster head, it transmits data to BS with the help of relay node. The relay node transmits the data to BS via other relay node.
3.2 Network Model

In our scenario, consider PXQ sensing area comprises of a BS, ‘n’ number of sensor nodes that are randomly distributed over the sensing environment and ‘m’ number of relay nodes are eventually distributed at the regular intervals. We assume that, BS is positioned at the centre of the sensing environment. The relay nodes are always to be in active mode. The nodes in the sensing environment are clustered based on the query sent by BS. The nodes who satisfy the query are only allowed to remain in the transmit mode and actively participating in the clustering process. The node, who does not satisfy the query go to the sleep state immediately. It is not allowed to participate in the clustering process. Cluster head collects the information from its cluster members. Relay node collects the information from its immediate neighborhood CH and forwards it to BS via other relay node in a multi-hop fashion. The relay node acts as routing element. AODV algorithm is used to find the shortest path.

3.3 Energy Consumption Model

According to the energy consumption model [4] [10], the energy is needed for transmitting ‘M’ bits to a distance ‘d’ is expressed as follows.

\[ E_{TX}(L, d) = E_{elec} + E_{amp} \]  

(1)

The energy is needed for receiving data is denoted as

\[ E_{RX}(L) = E_{elec} \]  

(2)

Where \( d \) is the distance between the two sensor nodes, \( E_{TX}(L,d) \) is the transmitter energy consumption, and \( E_{RX}(L) \) is the receiver energy consumption. \( E_{elec} \) is the electronics energy consumption per bit in the transmitter and receiver sensor nodes. \( E_{amp} \) is the amplifier energy consumption. It is calculated by using

\[ E_{amp} = \begin{cases} \epsilon_{fs} \times d^d, & \text{when } d \leq d_0 \\ \epsilon_{mp} \times d^4, & \text{when } d > d_0 \end{cases} \]  

(3)

Where \( d_0 \) is a threshold value. If the distance \( d \) is less than \( d_0 \), the free space propagation model is used. Otherwise multipath propagation channel model is used. \( \epsilon_{fs}, \epsilon_{mp} \) are communication energy parameters. Suppose that a non cluster head node \( N \) transmits \( L_N \) bits to the BS. Let \( d_{N,CH} \) be the distance between the non-cluster head node \( N \) and its cluster head node \( CH \). Let \( d_{CH,RN} \) be the distance between the cluster head and its immediate neighborhood relay node \( RN \). Let \( d_{RN,CH} \) be the distance between the relay node \( RN \) and the BS. Due to effective cluster head election, multi-hop communication is reduced. Only the minimal amount of non-cluster head nodes are using multi-hop communication. Others using free space model for communicating with the CH. The residual energy of the non-cluster head node \( N \) is expressed as

\[ E_{res,N} = E_{TX}(L_N, d_{N,CH}) \]  

(4)

In addition, the residual energy of the cluster head node CH is denoted as

\[ E_{res,CH} = E_{TX}(L_N, d_{CH,RN}) \]  

(5)
In order to reduce the energy hole problem in the static sink approach, Cluster head transmits its data to its immediate neighborhood relay node instead of using other sensor nodes in the sensing environment. The residual energy of the relay node is expressed as

\[ E_{R_{ch}} - E_{R_{ch}}(L_{N}) - E_{M_T}(L_{N}, d_{R_{ch}BS}) \]  \hspace{1cm} (6)

The relay node acting as routing element, it collects the aggregated information from its immediate neighborhood CH and transfer it to the BS via other relay node.

4. PROPOSED PROTOCOL

In this algorithm, the operation includes two phases, setup and steady state phase.

4.1 Setup Phase

The main goal of this phase is to create clusters and find cluster heads. During the setup phase, the BS collects the information about the position and energy level of all the sensor nodes in the network. It focuses on the dynamic cluster formation, effective cluster head election and positioning of the base station. It is discussed as follows.

4.1.1 Dynamic Cluster Formation

In existing approaches [4] [8] [9] [10] [11] [12], all the nodes in the sensing area are taken into account during the cluster formation. It results in an increased cluster count, cluster density, data traffic, energy consumption and burden of the aggregation process. In our paper, states of the transceiver are scheduled based on the query sent by the BS. A query is a requirement that needs to measure from the environment. The BS transmits its query to all of its sensor nodes during its allotted time slot. The node who satisfying the query is only allowed to remain in the transmit mode and allowed to participate in the clustering process. Others go to the sleep mode immediately. Thus transceiver scheduling increase network lifetime. Because the energy consumption in sleep mode is less compared to the energy consumption in the transmit/receive mode. It also reduces the count of the cluster formed.

4.1.2 Cluster Centroid

K-Means algorithm is one of the prominent algorithm used to find the centroid of the cluster [15]. The centroid of the cluster \((\lambda)\) is calculated by using

\[ \lambda = \left( \frac{1}{n} \right) \times \sum_{i} \min_{j} d(X_{i}, m_{j}) \]  \hspace{1cm} (7)

Where \(d(X_{i}, m_{j})\) denotes Euclidean distance between the \(X_{i}\) and \(m_{j}\). \(m_{j}\) denotes the mean value of the cluster and \(X_{i}\) denotes the position of the nodes in that cluster.

4.1.3 Cluster Head Election Strategy

Cluster head election plays a vital role in conserving the energy level of the sensor node. Because poor cluster head election leads to unnecessary re-clustering. It affects the lifetime of the sensor network. The position of the cluster head is a very important factor as long distance communication consumes more energy than the short distance communication. We consider both location and the energy level of the sensor node as criteria for electing a cluster head. Nodes that are closer to the cluster centroid have a maximum energy level is elected as cluster head. It enables the cluster head to reach all of its cluster members with minimum energy consumption. It avoids multi-hop communication and reduce cluster head re-election frequency.

4.1.4 Positioning of Base Station

Proper positioning of the base station also increases the lifetime of the sensor network. It enables the BS to reach all of its sensor nodes with minimum energy consumption. In our paper, The BS is placed at the center of the sensing area. It increases the network lifetime. Center position \((x_c, y_c)\) of the sensing area is calculated by using

\[ (x_c, y_c) = (x_i + x_j)/2, (y_i + y_j)/2 \]  \hspace{1cm} (8)

Where \((x_i, y_i)\) and \((x_j, y_j)\) are diagonal coordinates of the sensing area.

4.2 Steady state phase

Once the clusters are created and the TDMA schedule is fixed, data transmission can begin. The cluster members, send data to CH during their allocated transmission time. When all the data have been received, the cluster head performs aggregation and produce a single signal. It reduces the amount of information transmitted to the BS. Cluster head is always in an active mode. It delivers the aggregated data to either directly to BS or via relay node based on the cluster head classification. It is discussed as follows.

4.2.1 Efficient cluster head classification

In order to simplify the data gathering process, our algorithm classifies the cluster head based on the Euclidean distance between the cluster head and the BS into two namely, first level cluster head and second level cluster head. If the BS is in the communication range of cluster head is taken as first level cluster head. Otherwise, it is taken as the second level cluster head. It is represented as

\[ C = \{c_1, c_2, c_3, ...., c_k\} \]  \hspace{1cm} (9)

\[ D = \{d_1, d_2, d_3, ...., d_s\} \]  \hspace{1cm} (10)

Where \(c_1\), \(c_2\), \(c_3\), \(c_s\) denotes cluster heads in the sensing environment. \(d_1\), \(d_2\), \(d_3\), \(d_s\) refers distance between the cluster head and the BS.

\[ f_1 = \{c_i : d(c_i, BS) \in \gamma, 3\gamma, 1 \leq i \leq n\} \]  \hspace{1cm} (11)

\[ s_1 = \{c_i : d(c_i, BS) \notin \gamma, 3\gamma, 1 \leq i \leq n\} \]  \hspace{1cm} (12)

\(f_1\) and \(s_1\) denotes the first level and second level cluster heads. \(\gamma\) is the maximum communication range of the cluster head. \(d(c_i, BS)\) denotes the distance from the cluster head to the base station.

4.2.2 Intra cluster communication

In our approach data is gathered from the cluster head based on the cluster head classification. First level cluster head transmits its information to BS directly because of its closeness and Second level cluster transmits its data to the BS with the help of relay node as in fig.1. It reduces the energy hole problem in the static sink approach. AODV Algorithm is used to find the shortest path.

4.3 Algorithm

1. Initially ‘n’ number of sensor nodes are distributed randomly in the sensing area PXQ.

2. Initialize the transceiver states of all the nodes in the sensing environment to the sleep mode.

3. \(s_1\) = sensing value of the node. \(q_1\) = value of the query, \(1 = total number of clusters in the sensing area, \(f_1\) = first level cluster head, \(s_2\) = second level cluster head, \(c\) = cluster head in the sensing environment, \(\gamma\) = maximum coverage area of the CH, \(e\) is the energy level of the sensor node and \(m\) is total number of nodes within the cluster.

4. Scheduling the transceiver states of the sensor nodes
for (i=0;i<n;i++)
if (s_i>=q_i)
keep the node to be in transmit mode and allow to
participate in the clustering process
else
set the transceiver state of the node from active to
sleep mode.
5. Calculate the centroid of the cluster
\[ \lambda = \frac{1}{n} \times \sum (\text{min}_j \, d'(X_c, m_j)) \]
6. Elect the cluster head
The node having maximum energy and closer to the
centroid is elected as cluster head.
7. Classify the cluster head
for (i=0;i<l;i++)
if (d(c_i, BS)<\gamma)
\text{add } c_i \rightarrow f_i
else
\text{add } c_i \rightarrow s_i
8. Allow first level cluster head to transmit its data to the BS
directly
9. Allow second level cluster head to transmit its data to BS
via relay node using AODV algorithm.

5. EXPERIMENTAL RESULTS
In this section, we evaluate the performance of our proposed
query based dynamic clustering algorithm using the
simulation model. We design a simulation environment by
using C++. The assumption of our simulation study is as
follows.

Table – 1 Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>50,100</td>
</tr>
<tr>
<td>Node deployment</td>
<td>Random</td>
</tr>
<tr>
<td>Mac</td>
<td>802.11</td>
</tr>
<tr>
<td>Total sensing area</td>
<td>100x100m</td>
</tr>
<tr>
<td>Position of BS (X, Y)</td>
<td>(50,50)</td>
</tr>
<tr>
<td>Initial energy</td>
<td>2J</td>
</tr>
<tr>
<td>Energy needed for data</td>
<td>5nJ/bit/signal</td>
</tr>
<tr>
<td>aggregation (E_{data})</td>
<td></td>
</tr>
<tr>
<td>Electronics energy (E_{elec})</td>
<td>50nJ/bit/signal</td>
</tr>
</tbody>
</table>

5.1 Performance Metrics
1. Round
The round is defined as the receiving the data from
all sensor nodes to the base station.

2. Total energy
It is defined as the sum of residual energy at all
sensor nodes.

5.2 Results Discussion
Fig. 2.a show the total network energy when number of sensor
nodes are 50 and the sensing area is 100X100m. It is very
obvious that the total network energy of our proposed
algorithm is higher than that of LEACH and SECA. The
reason is, In the proposed approach, nodes who satisfy
the query are only allowed to participate in the clustering process.
Others go to the sleep mode immediately. It reduces the
cluster count and its density. Thus, the energy consumption is
reduced. Fig 2.b, 2.c shows the first sensor node dead and half
of the sensor nodes alive for three methods, when the number
of sensor nodes is 50 and the sensing area is 100X100m. It is
evident that the lifetime of the sensor node of our approach is
higher than that of LEACH, and SECA. The reason is, in
existing approaches, cluster head forwards its data via other
sensor nodes in the sensing environment. It makes the nodes
near the sink unreachable. In our approach, cluster head
forward its data to the base station via relay nodes. It results in
increased network lifetime. Fig 2.d plots the performance of
benchmarking system and the proposed system as a function
of rounds in terms of total network energy. From the Fig. 2.d
We can see, when the density of the sensor nodes becomes
larger, the advantage of the proposed system over
benchmarking system becomes more because of dynamic clustering.
6. CONCLUSION

The energy conservation is a challenging issue in the wireless sensor networks. To increase the network lifetime and reduce the energy consumption, we propose query based dynamic clustering approach, in which the nodes who satisfy the query are only allowed to participate in the clustering process. Others go to the sleep mode immediately. It reduces the cluster count and its density. It results in increased network lifetime. The node which is closer to the centroid of the cluster and having maximum energy is elected as cluster head. It enables CH to reach all of its cluster members with minimum energy consumption. Finally, relay nodes are used to reduce the energy hole problem in the static sink approach. The simulation results show that our proposed algorithm achieves low energy consumption and better network lifetime in the wireless sensor network.

7. REFERENCES


