BDCP: A Backoff-based Distributed Clustering Protocol for Wireless Sensor Networks

Tarachand Amgoth and Prasanta K. Jana
Department of Computer Science and Engineering
Indian School of Mines, Dhanbad
Dhanbad, India
tarachand.ism@gmail.com and prasantajana@yahoo.com

Abstract—Clustering is considered as an efficient technique for improving scalability and conserving energy of the sensor nodes to prolong the life time of wireless sensor networks (WSNs). In this paper, we propose a new backoff based distributed clustering protocol, called BDCP for wireless sensor networks. The algorithm consists of two phases, namely cluster head (CH) selection and cluster formation. For energy saving, the proposed algorithm forms the clusters by taking care of proper distribution of the CHs, their residual energy and the intra-cluster distance. Performance of our protocol is validated by simulation results which show that the BDCP outperforms other well known clustering algorithms.

Keywords—Wireless sensor networks, clustering energy consumption, network lifetime

I. INTRODUCTION

Wireless sensor network (WSN) is a network of autonomous tiny sensor devices with limited battery power, communication, and computation capability. Sensor nodes are deployed in a target area for many applications, e.g., forest fire detection, remote environmental monitoring, flood detection, and homeland security surveillance [1]. The principal job of the sensor nodes is to collect data such as temperature, humidity, pressure etc., and relay them to a special node called sink. The sink acts as a liaison between the network and the outside world such as Internet. In many applications, sensor nodes are deployed in hostile areas, e.g., battlefields, remote geographic region, toxic urban locations, etc., [1]. So, recharging the batteries of the sensor nodes manually in such places is extremely difficult. It also requires that the network should operate autonomously for a long period of time under energy constraints. Therefore, energy consumption of the sensor nodes is the major challenge for long run of the WSN.

Network clustering plays a prominent role in reducing the energy consumption of the sensor nodes. In clustering, sensors nodes are organized into groups and each group is coordinated by a leader called cluster head (CH). Each node in the cluster transmits sensed data to its CH only. Then CH performs the data aggregation and transmits the aggregated data to the sink via single hop or multi-hop communication. Clustering can minimize the data redundancy by incorporating the data aggregation at the CH level in the network. Thus, the amount of data transmitted to the sink is significantly reduced and bandwidth is also conserved.

Clustering can produce better results, if CHs are well distributed in the sensing field (with more residual energy) and intra-cluster distance is minimal. HEED [2] is a popular clustering algorithm that selects CH based on residual energy of the sensor nodes and intra-cluster distance as the second parameter. But HEED introduces extra overhead to compute cost of communication to its neighbors, which may consume more energy. Another clustering algorithm proposed in [3] that uses a backoff strategy to ensure that the selected CHs are well distributed in the sensing field with more residual energy. However, it does not consider the intra-cluster distance which is not justifiable for energy saving of the sensor nodes.

In this paper, we propose a new backoff based distributed clustering protocol for wireless sensor networks. The algorithm not only selects the CHs in well distributed manner but also considers the intra-cluster distance. Moreover, it ensures the load of CHs with respect to their residual energy. In our approach, each node starts the campaign to become a CH by initiating a time delay, which is inversely proportional to its residual energy. To form clusters, we define a new cost function through which sensor nodes join a CH with more residual energy and near one. We perform tests on the proposed method extensively. Experimental results clearly demonstrate that the proposed method outperforms the existing related protocols such as given in [3] and [5].

The remainder of the paper is organized as follows. Section 2 briefly reviews some of the related work. Section 3 presents the network model. Section 4 elaborates our proposed clustering protocol i.e., cluster head selection and cluster formation. We discuss the experimental results in section 5 and conclude the paper in section 6.
II. RELATED WORK

Many clustering algorithms [2-3, 5-15] have been proposed for WSNs. Some of the algorithms are based on centralized approach and some are distributed in nature. Clustering protocols have also been investigated in the recent years which can be found in [16-20]. Since our techniques focus on CH selection and cluster formation, we limit our discussion only to such techniques which are related to our work.

The authors in [11] proposed a distributed weight-based energy efficient hierarchical clustering protocol (DWEHC). The iterative nature of DWEHC is similar to HEED and produces relatively high control message overhead. Another energy efficient clustering scheme (EECS) proposed in [12]. In EECS, the nodes decide to join a CH based on a weighted cost factor during cluster formation. Although EECS introduces efficient cluster formation, it produces high control message overhead which is higher than HEED. A modified version of EECS is proposed in [13] called EEUC (energy efficient unequal clustering). Although it forms good clusters but each node needs to keep track information about all the CHs in the network and each CH needs to transmit advertisement message at high transmission power to cover the whole deployment area. Therefore, it results in exchange of large number of control messages and extra energy wastage. Recently, an energy efficient distributed clustering protocol has been proposed in [16], called GESC (Geographic sensor clustering protocol). GESC exploits the local network characteristic and the residual energy of the neighboring nodes to achieve network lifetime prolongation. Therefore, GESC exchanges high control message with the neighbors and it is quite similar to HEED. A new distributed energy efficient clustering protocol for WSNs has been proposed in [17], in which CHs are elected by three-way message exchanges between each sensor node and its neighbors. Sensor nodes are elected as cluster heads based on their residual energy and node degree. In [20], a multi-criterion optimization technique for energy efficient cluster formation was presented, called MOECS. Although MOECS is efficient in cluster formation, it suffers from the cost of huge control messages and high computation complexities.

Cao et al. [3] proposed a backoff based distributed algorithm which we refer as DCA in this article. In their approach, each node starts campaigning to become a CH by initiating a timer which is inversely proportional to its residual energy. Once the timer of a node expires, the node selects itself as a CH and then it broadcasts a CH claim message. If a node receives the claim message, then it switches off its timer and joins the cluster of the node. This approach ensures that CHs are well distributed with respect to sensing area. Unlike HEED, the message complexity of this algorithm is O(1) per node. Therefore, this technique outperforms both the HEED and LEACH [5] in enhancing the network lifetime. However, intra-cluster communication in this approach is not minimized as each node joins a CH based on receiving a claim message irrespective of any distance.

The algorithm presented in this paper performs clustering with even distribution of the CHs. It also considers the load balancing of the CHs with respect to their residual energy and most importantly it takes care of the intra-cluster distance to minimize the overall energy consumption of the network.

III. SYSTEM MODEL

In this paper, we assume that a set of homogenous sensor nodes are deployed randomly in the area of interest. The sink is one and static, located outside the area under observation. Nodes are left unattended after deployment. Therefore, battery recharge is not possible. Initially, all the sensor nodes are charged with same amount of energy. The Link is symmetric and bidirectional. All the sensor nodes are organized into distinct clusters. The data sensed by sensor nodes transmits to their corresponding CH. Then CH sends the aggregated data to the sink directly. The function of our proposed algorithm is divided into rounds. Each round begins with CH selection and cluster formation followed by data transmission.

The energy needed to send β-bit data from node i to node j can be calculated by using the path loss model [4]. We also calculate the transmission energy by using free space (fs) and the multipath fading (mp) channel model as follows [5]

\[ E_{tx}(i, j) = \left( \alpha + \varepsilon_{fs}d^2 \right) \beta \\
E_{rx}(i, j) = \left( \alpha + \varepsilon_{mp}d^4 \right) \beta \]

(1)

where, \( d \) is the transmission distance between the node i and j, \( \varepsilon_{fs} \) and \( \varepsilon_{mp} \) are the energy required by amplifier in free space and multipath respectively and \( \alpha \) is the total energy per bit consumed by the transmitter and the receiver. The energy consumed in receiving β-bit data is given by

\[ E_{rx}(j) = \alpha \beta \]

(2)

Therefore, the total energy consumed in sending β-bit data is given as follows

\[ E_{tx}(i, j) = E_{tx}(i, j) + E_{rx}(j) \]

(3)

IV. PROPOSED PROTOCOL

Clustering protocol is divided into two phases, namely, CH selection and cluster formation. They are discussed successively in the following sections.

A. Cluster head (CH) Selection

In cluster head selection, a node i initializes its timer \( t(i) \) which is derived as follows

\[ t(i) = (1 - \lambda(i)) \times T_A \]

(4)

\[ \lambda(i) = \frac{E_{tx}(i)}{E_{tx}(i)} \]

(5)
where, $T_A$ is the maximum time allotted for CHs selection, $E_r(i)$ and $E_m(i)$ are the residual energy and initial maximum energy of sensor node $i$ respectively. A node with high residual energy is more likely to become a CH, since its timer expires first. When the timer of node $i$ is expires, then the node $i$ selects itself as a CH for the upcoming communication round and broadcasts a CH announcement message in its range, which includes its $E_r(i)$, identity, and location information. When a node $j$ receives the message, then switch of its timer and records the message by including senders’ information into its set $N_{CH}(j)$, where $N_{CH}$ is the set of neighbor CHs of node $j$. From now onwards if the node $j$ receives such messages from other nodes, then it simply updates the set $N_{CH}(j)$. Node $j$ uses this set to decide its cluster membership after completion of CH announcement time. Node $j$ cancelling its timer indicates that the node $j$ withdraw itself from CH competition and become a non-CH for the upcoming communication round.

B. Cluster Formation

In cluster formation, if node $j$ receives a single CH advertisement message, then it simply joins the cluster, which includes its ID. When node $j$ receives more than one CH advertisement messages, node $j$ decides its cluster membership depending residual energy of a CH and distance with the CH. To be energy aware, node $j$ simply joins a CH with highest residual energy and belongs to the set $N_{CH}(j)$, even though it may not be the nearest one. To be energy efficient, the node $j$ joins the nearest CH and belongs to the set $N_{CH}(j)$ in order to minimize its communication energy. To be both energy efficient and power aware, node $j$ uses the following cost function to decide its cluster membership.

$$CF(i) = w(i) \times E_r(i), \forall i \in N_{CH}(j)$$  \hspace{1cm} (6)

where, $w(i)$ is dimensionless co-efficient, and is used as a weighting factor for computing the cost function. It is derived as follow

$$w(i) = \frac{E_m(i) - E_r(i)}{E_r(i)}$$  \hspace{1cm} (7)

A bigger value of cost function is due to either high communication energy consumption by the node $i$ or lower residual energy of the cluster head $i$, or both. Then node $j$ computes the cost function of each of the CH belonging to the set $N_{CH}(j)$ and joins the CH for which cost function returns minimum value. The pseudo code of BDCP is shown in Fig. 1.

C. Complexity Analysis

**Observation 1.** The proposed algorithm is both power aware and energy efficient. CHs are selected on the basis of their residual energy and remaining sensor nodes join the clusters by considering both the residual energy of the CHs and communication energy of the sensor nodes. Therefore, the proposed clustering protocol not only minimizes the energy consumption of the sensor nodes but also balances it.

**Observation 2.** The proposed protocol is a distributed protocol since each sensor node makes decision independently whether to select itself as CH or join as a member of a cluster.

**Lemma 1.** The message complexity of our proposed algorithm is $O(1)$ per node.

**Proof.** During the CH selection or cluster formation of proposed algorithm, each sensor node either sends a CH announce message or cluster joins message only. Therefore, the message complexity of our algorithm is $O(1)$.

**Lemma 2.** Time complexity of our proposed algorithm is $O(n)$ for $n$ number of sensor nodes.

**Proof.** Fig. 1 presents the pseudo code of proposed protocol. In BDCP, each sensor node independently initiates their timer (line 03), which can be done in $O(1)$. In phase 2, each non-CH node keeps track of CH announcement messages it receives during CH selection. In worst case, a sensor node receives maximum $n$ messages (line 09). Therefore, to join a cluster, a sensor node needs to compute the cost function $n$ times (line 17) to decide its cluster membership. Hence, the time complexity of our proposed algorithm is $O(n)$.

**Lemma 3.** Cluster head selection lasts for at most $T_A$.

**Proof.** Note that the value of $\lambda$ is lies between 0 and 1. By using the equation 4, the timer of sensor node ranges from 0 to $T_A$.

---

**Algorithm : BDCP**

/* Phase 1: CH Selection */
(01) for each node $i$
(02) $\lambda(i) = E_r(i)/E_m(i);$  \hspace{1cm} (03) $t(i) = (1 - \lambda(i)) \times T_A$;
(04) end for
(05) if ($t(i)$ is expired) then  
(06) broadcast (CH advt.);
(07) end if
(08) if (node $j$ receives CH advt.) then
(09) $t(j)$ is switched off and updates $N_{CH}(j);$  \hspace{1cm} (10) end if
(11) if ($t(j)$ is switched off and receives CH advt.) then
(12) update $N_{CH}(j);$  \hspace{1cm} (13) end if
(14) /* Phase 2: Cluster Formation */
/* $C_{set}(i)$: set of member sensor nodes of cluster $i$ */
(15) for each node $j$
(16) if ($t(j)$ is switched off and $|N_{CH}(j)| = 1$) then
(17) $C_{set}(i) = \bigcup \{C_{set}(i), j\}$;  \hspace{1cm} (18) end if
(19) end if
(20) if ($t(j)$ is switched off and $|N_{CH}(j)| \geq 2$) then
(21) if ($i \in N_{CH}(j)$) & $& (CF(i)$ is minimal) then
(22) $C_{set}(i) = \bigcup \{C_{set}(i), j\}$;  \hspace{1cm} (23) end if
(24) end if
(25) end for

**Fig. 1.** Pseudo code of BDCP
The experiments were carried out on Intel® Core™ i7 running on 64-bit MS Window 7 O.S. The simulation program was developed using Dev C++ and matlab. The parameters and their values used in the simulation are given in Table I.

We assumed that the sensor nodes are randomly deployed in an area of 500x500 square meter. The initial energy of sensor nodes was set to 0.5 Joules. We assumed that the round in the experiments is same as LEACH. We evaluate the proposed algorithm with three performance metrics, namely, network lifetime, energy consumption and amount of data sent to the sink. The network lifetime is taken as the number of sensor nodes that die for a given instant of time. Communication energy reflects the amount of energy consumed by the sensor nodes during a given instant of round. We also study the amount of data sent to the sink during each round of communication.

Table I. Parameters used in simulation

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sink location</td>
<td></td>
<td>(50,175)</td>
</tr>
<tr>
<td>Sensor nodes</td>
<td>N</td>
<td>200</td>
</tr>
<tr>
<td>Transmission range</td>
<td>r</td>
<td>25m</td>
</tr>
<tr>
<td>Transmission energy</td>
<td>( \alpha )</td>
<td>50 nJ/bit</td>
</tr>
<tr>
<td>Amplifier energy 1</td>
<td>( \epsilon_1 )</td>
<td>100 pJ/bit/m(^4)</td>
</tr>
<tr>
<td>Amplifier energy 2</td>
<td>( \epsilon_2 )</td>
<td>0.0013 pJ/bit/m(^4)</td>
</tr>
<tr>
<td>Data packet size</td>
<td>( D_p )</td>
<td>0.5 Kbits</td>
</tr>
<tr>
<td>Control message size</td>
<td>( C_p )</td>
<td>0.1 Kbits</td>
</tr>
</tbody>
</table>

Fig. 2 and Fig. 3 compare the network lifetime and total energy consumption of BDCP with those of LEACH and DCA respectively. The results show that BDCP outperforms others.

Probabilistic based CH selection in LEACH does not guarantee that a CH is selected with high residual energy and as a result a CH with less residual energy may die quickly. LEACH does not guarantee the well distribution of CHs in the sensing field. Therefore, intra-cluster communication cost is very high in comparison to BDCP and DCA. In DCA, CH selection is similar to our CH selection but intra-cluster communication is not minimized.

Furthermore, our cluster formation is more efficient than DCA in which both the residual energy of CH and communication energy of non-CH are taken care. By adopting backoff strategy in our proposed clustering protocol results in better distribution of CHs in the sensing field and with high residual energy. We also studied the amount of data transmitted to the sink, as shown in Fig. 4. Note that, Fig. 2 and Fig. 3 show that the BDCP performs better than others. Obviously, our proposed protocol sends more data than others.

**Remark:** Although our BDCP performs better than others but it sends the data to the sink directly which may not be the realistic assumption in all the cases. To overcome this, BDCP adopts multi-hop commutation to reach the sink and we refer it as M-BDCP. In this protocol, to obtain the routing path, the sink first broadcast a route discovery message in the network. Upon
receiving the message, each node introduces a delay proportional to its $1/z(i)$ before it forwards the message in its range. In this way a message arrives at each node along the desired path, which is called energy aware route. Then the CHs route the data in the same reverse path.

![Fig. 5. Number of live sensor nodes per round](image)

For the simulation and comparison purpose, we modified the DCA and LEACH protocol into multi-hop based communication by applying the same techniques used in M-BDCP and we referred them as M-DCA (Multi-hop DCA) and LEACH-M (LEACH with Multi-hop) respectively. The result in Fig. 5 shows that M-BDCP performs better than LEACH-M and M-DCA in terms of network lifetime.

VI. CONCLUSION

In this paper, we have proposed a new backoff based distributed clustering protocol for wireless sensor networks. The algorithm has been shown to have $O(1)$ message complexity and $O(n)$ time complexity for a wireless sensor networks of $n$ nodes. The simulation results have shown that the algorithm performs better than LEACH and DCA in terms total energy consumption, network lifetime and amount of data sent to the sink. We have also introduced multi-hop based BDCP (M-BDCP) and compared with LEACH-M and M-DCA. The result shows that M-BDCP performs better than LEACH-M and M-DCA. Our future endeavor will be to explore M-BDCP in a more elaborate and comprehensive way.

REFERENCES


