Geographical Information Based Clustering Algorithm to Equalize Cluster Lifetime throughout Wireless Sensor Networks

Dali Wei, H Anthony Chan and Shaun Kaplan
Dep. of Electrical Engineering, University of Cape Town, South Africa
E-mails: dlwei@erg.ee.uct.ac.za, h.a.chan@ieee.org, shaunkaplan@gmail.com

Abstract—Wireless sensor networks (WSNs) usually have a high node density, which makes it unnecessary to require all nodes to be continually active. Scheduling different nodes to be awake at different times saves energy for WSNs. However, traditional scheduling algorithms do not consider the directional data traffic towards the sink and maintaining coverage of the entire network, making them fail to optimize the energy efficiency and the network performance. This paper proposes a clustering algorithm, which combines scheduling techniques, to split the network into clusters based on the geographical information of nodes and the directional data traffic. The clusters are organized in such a way that all clusters have similar lifetimes. Only one node in each cluster is scheduled to be active at any given time to monitor that cluster. Simulation results show that this algorithm not only improves energy efficiency for WSNs, but also improves the performance of delivering data to the sink.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) usually have a higher node density than mobile ad hoc networks. In a WSN, an area that could, almost, be monitored by only one node may in fact have several sensor nodes. It is therefore unnecessary to require that all nodes in this area be active at any given time [1]. Furthermore, having all these nodes continuously active could be disadvantageous for the following reasons: (1) Though the data generated by these nodes can be locally aggregated before being sent to the sink, the source node may still waste significant energy in transmitting these data to the next hop. (2) The large number of nodes participating in relaying data makes the network topology more complicated. This complicated topology increases flooding in a network that uses a reactive routing protocol or overhead in a network that uses a proactive routing protocol. (3) The resultant high-density data traffic increases the collisions and congestions, thereby increasing loss of data packets.

Scheduling algorithms have been proposed to extend the node lifetime by allowing only some nodes to be active while others are in energy saving mode [2-12]. However, most scheduling algorithms do not consider the directional data traffic towards the sink. In most WSNs, the processing capability of sensor nodes is extremely limited. A data sink is therefore necessary to collect all data or to manage the network in some cases. All data are relayed to the sink making the data traffic directional towards the sink, which burdens the nodes differently. Those which need to relay more data to the sink have higher burdens and will die prematurely. Furthermore, most scheduling algorithms do not consider maintaining maximum coverage of the network as they usually schedule the nodes to be active without the consideration of the network coverage. Joint clustering and scheduling in such a network has some advantages. First, in a WSN with high node density, the nodes can be clustered in such a way that one active node in a cluster can almost cover the entire cluster. This means that almost the entire network will be covered by the active nodes. The active node in each cluster is referred to as the cluster head (CH) of that cluster. Secondly, by joint clustering and scheduling, it is possible to allocate energy to each cluster in accordance with its burden, making the equalization of cluster lifetime throughout the network feasible.

This paper is organized as follows: after investigating the related work in Section II, the proposed algorithm, which combines clustering and scheduling techniques, is introduced in detail in Section III; the performance evaluation is conducted in Section IV to show the contributions and the correctness of the new design; the paper is finally concluded in Section V.

II. RELATED WORK

GAF [13] combines clustering and scheduling techniques to improve energy efficiency. GAF separates the network into identical squares and only one node in each square is scheduled to be active at any given time to relay data for others while others are into low power consumption mode for energy saving. GAF, however, is proposed for MANETs and therefore the directional data traffic towards the sink is not considered. It therefore fails to optimize energy efficiency in WSNs.

Deng et al. [14, 15] propose an algorithm which applies a scheduling scheme into a clustered WSN. This algorithm organizes the nodes into one-hop clusters. It considers that the nodes that are further away from the CH will have higher burdens as they need higher transmission power to reach the CH. By scheduling these nodes into sleep state more frequently than others, this higher burden of the nodes can be compensated. The node lifetime in the same cluster can then be equalized. However, the algorithm only considers the different burdens of the nodes in the cluster. It does not take the directional data traffic towards the sink into account. This algorithm therefore fails to equalize node lifetime throughout the network, thus failing to optimize the energy efficiency in WSNs.
III. PROPOSED CLUSTERING ALGORITHM

A. Network model and energy model

A data sink in a WSN has sufficient resources, such as memory, processing capability and energy. It therefore is capable of performing cluster organization [16].

The energy-limited identical sensor nodes, which have the capacity to vary transmission power, are uniformly distributed in the network. The maximum transmission range of the nodes is limited and only some nodes, which are nearer the sink, can reach the sink over a single-hop route. However, to achieve higher energy efficiency, these nodes can use a multi-hop route to send data to the sink if this multi-hop route is more energy efficient than the single-hop one.

The sensor nodes are aware of their two-dimensional location information. With this information, the nodes can be geographically grouped into clusters with the expected sizes and shapes. The location information can be achieved by global positioning system (GPS) technology.

A sensor node can be a member node or a CH. A CH needs to monitor the cluster whereas a member node turns off its radio components for energy saving.

The network model is similar to the models in [16-19], which is described in (1). The meaning of each parameter can be found at Table I.

\[
E = E_r + E_p = \left( e_c + e_{amp}d^4 \right) \times a + e_r \times a
= \begin{cases} 
  e_c \times a + e_{amp}d^2 \times a + e_r \times a & (d < d_{crossover}) \\
  e_c \times a + e_{amp}d^4 \times a + e_r \times a & (d \geq d_{crossover})
\end{cases}
\]  

\( E \) Total energy dissipated during the communication
\( E_r \) Energy dissipated in the transmitter of the source node
\( E_p \) Energy dissipated in the receiver of the destination
\( e_c \) Radio dissipation of the transmitter circuit
\( e_r \) Radio dissipation of the receiver circuit
\( e_{amp} \) Parameter of transmitter amplifier
\( d \) Transmission range of the hop
\( a \) Data amount

Table I. Definitions of the parameters of the energy model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E )</td>
<td>Total energy dissipated during the communication</td>
</tr>
<tr>
<td>( E_r )</td>
<td>Energy dissipated in the transmitter of the source node</td>
</tr>
<tr>
<td>( E_p )</td>
<td>Energy dissipated in the receiver of the destination</td>
</tr>
<tr>
<td>( e_c )</td>
<td>Radio dissipation of the transmitter circuit</td>
</tr>
<tr>
<td>( e_r )</td>
<td>Radio dissipation of the receiver circuit</td>
</tr>
<tr>
<td>( e_{amp} )</td>
<td>Parameter of transmitter amplifier</td>
</tr>
<tr>
<td>( d )</td>
<td>Transmission range of the hop</td>
</tr>
<tr>
<td>( a )</td>
<td>Data amount</td>
</tr>
</tbody>
</table>

In this algorithm, the same parameters applied in [16-19] are used: for distance \( d < d_{crossover} \) (where \( d_{crossover} = 87m \)), \( e_{amp} = 10pJ/\text{bit/m}^2 \); for distance \( d \geq d_{crossover} \), \( e_{amp} = 0.0013pJ/\text{bit/m}^4 \); and \( e_c = e_r = 50 \text{nJ/bit} \).

B. Cluster organization and scheduling algorithm

During the cluster setup, each node sends its location information to the sink. On reception of the location information, the sink divides the network into geographically based rectangular clusters, with a pre-set number of clusters and cluster sizes. The pre-set cluster sizes serve to equalize cluster lifetime throughout the network, which will be presented in Section III D. The sink broadcasts the cluster information, including the cluster ID number, the cluster geographical information and the number of nodes in each cluster, to the sensor nodes in the network.

On reception of the partition information, each node knows the cluster ID by comparing its own location to the received geographical information of the cluster. The nodes can now recognize other nodes in the same cluster.

After the clusters have been formed, the next step is to elect a CH in each cluster to monitor that cluster and to relay data from other clusters. The selection of the CH is based on the residual node energy. However, should the network commence when all nodes have the same residual energy, the node at the centre of the cluster is selected to be CH.

The scheduling scheme used for the competition of being a CH is contention-based. To reduce the flooding caused by exchanging residual energy information, only the CH will broadcast its residual energy information to the member nodes when each round expires. This is because, only the CH consumes energy during each round (other nodes are in sleep state with negligible power consumption) and thus only its residual energy information needs to be updated. Each time residual energy information is updated; the node with the highest residual energy becomes a CH for the next round. The sink assigns the CHs a number of frames. This number is preset according to the requirement of the network. If a network needs a high quality of service (QoS) and needs all information carried by the entire network, the active CH needs to be rotated frequently. This frequent rotation, however, may consume some energy, increase topology overhead and flooding in routing discovery. The tradeoff between the QoS and the power consumption can be achieved by adjusting the rotation period.

The scheduling scheme is shown in Fig. 1. At each round (or slot), the node with the highest residual energy in the cluster becomes the CH. After the CH is determined, the number of frames is assigned to each CH. A CH is allowed to send a data packet within each frame.

Figure 1. Scheduling scheme for ECL.

C. Data communication phase

When the number of frames is assigned to each CH, the CH can send data to the sink. This algorithm selects the routes based on the consideration of residual node energy. After the
available routes are found by an on-demand routing discovery process, the minimum residual energy of the CH in the route is set as a comparative parameter. The route with the maximum value is selected to relay the data. This selection exempts the CHs with lower residual energy from relaying data and better balances power consumption among the CHs and routes [1, 20, 21].

D. Equalizing cluster lifetime

This subsection discusses equalizing cluster lifetime using the model shown in Fig. 2.

![Figure 2. WSN to explain cluster-lifetime equalization.](image)

The clusters are named according to their geographic positions. The identical sensors with the initial energy of \( E_{\text{initial}} \) are uniformly distributed in the network which has a node density of \( n_d \). To simplify the analysis, it is assumed that each cluster selects the node at the centre of the cluster to be CH. For cluster \( a_ib_j \), the transmission range of \( d_{ij} \) is assigned. This transmission range, however, is adjustable in both realistic applications and simulations. To equalize the cluster lifetimes, (2) must be observed.

\[
\sum P_{r}(i) + \sum P_{t}(i) = \text{const} \quad (2)
\]

where \( E_{\text{total}}(i) \) is the total energy store in cluster \( a_ib_j \), and \( \sum P_{r}(i) \) and \( \sum P_{t}(i) \) are respectively the total CH power consumption of receiving and transmitting data in cluster \( a_ib_j \).

There are four neighboring clusters as shown in Fig. 3. Each CH generates a \( a \)-bit data message. Only the CHs nearer the sink can relay the route requests. It is assumed that the probability of a neighboring CH nearer the sink of cluster \( a_ib_j \) to relay the data is \( \beta \). \( \beta \) is a variable in realistic applications and the simulation as the selection of a route is based on the residual CH energy. The total data of the CH in cluster \( a_ib_j \) (Data\((a_ib_j)\)) that need to be relayed is separated into the data that the CH receives from others \( (R(a_ib_j)) \) and the data that the CH transmits to the next hop \( (T(a_ib_j)) \), as shown in (3).

\[
\text{Data}(a_ib_j) = R(a_ib_j) + T(a_ib_j) \quad (3)
\]

![Figure 3. Different data traffic in neighboring clusters.](image)

in which

\[
R(a_ib_j) = \sum_{(i+j-l-m)} (i+j-l-m) \times a_i a_m \times \beta^{i+j-l-m} \quad (4)
\]

\[
T(a_ib_j) = a + \sum_{(i+j-l-m)} (i+j-l-m) \times a_i a_m \times \beta^{i+j-l-m} \quad (5)
\]

The power consumption of the CH in cluster \( a_ib_j \) is therefore determined in (6).

\[
P_{r}(a_ib_j) = R(a_ib_j) \times \epsilon_r + T(a_ib_j) \times \left( \epsilon_t + e_{\text{amp}} \times d_{ij}^{\alpha} \right) \quad (6)
\]

The total initial energy store of cluster \( a_ib_j \) is (7):

\[
E_{\text{initial}}(a_ib_j) = a_i \times b_j \times n_d \times E_{\text{initial}} \quad (7)
\]

According to (2), by acquiring (8), the cluster lifetimes are equalized.

\[
\frac{E_{\text{initial}}(a_ib_j)}{P_{r}(a_ib_j)} \approx \text{const} \quad (8)
\]

IV. PERFORMANCE EVALUATION

The performance of the proposed algorithm to equalize cluster lifetime (ECL) throughout the network for WSNs is compared to a traditional clustering algorithm, which organizes clusters with similar sizes, such as in [16-19], and sends data to the sink over multi-hop routes, such as in [16, 18]. To render these two algorithms comparable, the scheduling algorithm being implemented makes only the CH in the selected traditional clustering algorithm active. This algorithm, which is now similar to GAF, is referred to as Similar Size with only CH Active (SSCA).

Homogeneous sensor nodes, 400 in total, are uniformly distributed within an area of 100m×50m, with the data sink located 10m away from the network, as shown in Fig. 4. Each node has 2\( J \) initial energy with a maximum transmission range of 70m. The operation of both algorithms is broken up into rounds. Each round assigns 20 frames to a CH and a CH sends
a 500 byte packet to the sink over a multi-hop route during each frame. The nodes are organized into 8 clusters.

The nodes are organized into 8 clusters. Data sink (0,0) 10 35 60 85 110

**Figure 4. Simulation Scenario.**

Fig. 5 shows the comparison at 1072 rounds, which is the first-node lifetime (when any node in the network dies) of SSCA. Up to 1072 rounds, SSCA consumes 187J energy whereas ECL consumes 181J energy. When any node dies in cluster 1 in SSCA, other clusters still have much energy left. The directional data traffic makes cluster 1 in SSCA die much earlier than the others. However, the average residual energy in ECL is almost the same across clusters. ECL thus presents a far better balance in power consumption than does SSCA.

**Figure 5. Average residual node energy at 1072 rounds in ECL and SSCA.**

Fig. 6 shows the average residual node energy at 2298 rounds, which is the first-node lifetime (when any node in the network dies) of SSCA. As more and more nodes die in SSCA, the area covered by nodes that have already used up their energy will finally exceed the maximum transmission range of the nodes that still have residual energy. They can then no longer relay data to the sink. The connectivity between the sink and the network is therefore lost. Up to 2298 rounds, SSCA consumes 389J energy whereas ECL consumes 383J energy. As the network cannot send any data to the sink, the network can be considered to be dead and the residual energy in remaining areas of the network is wasted. The directional data traffic makes ECL waste more than 50% of the network’s energy. At 2298 rounds, ECL still has 1J energy left in each node.

**Figure 6. Average residual node energy at 2298 Rounds in ECL and SSCA.**

Fig. 7 compares the residual node energy in ECL and SSCA at their respective first-node lifetimes. At the first-node lifetime, SSCA consumes only 187J energy whereas ECL consumes 780J energy. The lower total residual energy of the network at the first node lifetime means there is a better balance in power consumption. ECL performs far better in balancing power consumption throughout the network than does SSCA.

**Figure 7. Average residual node energy at first-node lifetime of ECL and SSCA.**

Fig. 8 compares the residual node energy in ECL and SSCA at their respective connectivity lifetimes. The connectivity lifetime in SSCA is only 2298 rounds whereas it is 4736 rounds in ECL. Up to connectivity lifetime, SSCA consumes 389J energy whereas ECL consumes 789J energy. The wasted energy in SSCA and ECL is therefore 411J and 11J respectively. ECL performs far better in using the energy than does SSCA.

**Figure 8. Average residual node energy at connectivity lifetime of ECL and SSCA.**

The distribution of the residual node energy at connectivity lifetime in SSCA and ECL is presented in Fig. 9. All nodes in ECL have less than 5% energy left. However, in SSCA, many...
nodes have a good proportion of residual energy: about 13% of nodes have 50% energy left and about 50% of nodes have more than 75% energy left. The wasted residual energy in SSCA is therefore quite significant. This means that the balancing of power consumption among the clusters in typical clustered WSNs is actually important for increasing WSN lifetime [20].

![Distribution of residual node energy at connectivity lifetime](image)

Figure 9. Distribution of residual node energy at connectivity lifetime.

Fig. 10 shows the total number of data packets delivered to the sink up to connectivity lifetime. This illustrates the effectiveness of ECL in delivering data. Up to connectivity lifetime, SSCA only delivers 337240 packets whereas ECL delivers 757040 packets to the sink. ECL offers improvements in data delivery by 124.5% over SSCA. Among these network data, 99.0% and 50.9% packets are delivered to the sink before the first-node lifetime, which are from the entire network, in ECL and SSCA respectively. Other packets delivered to the sink in both algorithms are from a partial network as some nodes have died and their areas are no longer monitored. ECL not only improves the energy efficiency but also the network performance of delivering data to the sink.

![Number of packets delivered to the sink](image)

Figure 10. Number of packets delivered to the sink.

V. CONCLUSION

This paper proposes to combine scheduling and clustering technologies to improve energy efficiency for homogeneous sensor networks. In this algorithm, only the CHs are active. The clusters are organized in such a way that their burdens are relative to their energy stored, so that the cluster lifetimes are equalized. The performance of the proposed algorithm is evaluated by simulation and by comparing it to a traditional design. Simulation results show that the proposed ECL algorithm balances power consumption throughout the network, extends both first-node lifetime and connectivity-lifetime, and improves performance of delivering data to the sink. Further work will be conducted on the investigation of coverage issues.

REFERENCES