Latency and Lifetime-Aware Clustering and Routing in Wireless Sensor Networks

Chuanyao Nie Hui Wu Wenguang Zheng

School of Computer Science and Engineering
The University of New South Wales, Australia
{cnie, huiw, wenguangz}@cse.unsw.edu.au

Technical Report
UNSW-CSE-TR-201614
November 2016

THE UNIVERSITY OF
NEW SOUTH WALES

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

Clustering is an effective technique for improving both the network lifetime and the robustness of a WSN (Wireless Sensor Network). We investigate the following latency and network lifetime-aware clustering problem for data collection: Given a WSN with one base station and a natural number k, construct a set of disjoint clusters for the WSN and a routing tree for inter-cluster communication such that the network lifetime is maximized and the maximum hop distance between each cluster to the base station is at most k. We propose a novel approach to this problem. Our approach consists of a polynomial-time heuristic for constructing clusters, a polynomial-time heuristic and an ILP (Integer Linear Programming) algorithm for constructing a routing tree for inter-cluster communication. We have performed extensive simulations on network instances with 200, 400, 600, 800 and 1000 sensor nodes with the uniform distribution and the random distribution, and compared our approach with two state-of-the-art approaches, namely MR-LEACH and DSBCA. In terms of network lifetime, the average improvements of our approach using the ILP algorithm for constructing a routing tree over MR-LEACH and DSBCA are 35% and 62%, respectively. We have also compared the heuristic and the ILP algorithm for constructing a routing tree. The ratios of the average lifetimes achieved by the heuristic and the ILP algorithm are 93% and 90%, for the uniform distribution and the random distribution, respectively.

1 Introduction

WSN has numerous applications, ranging from border protection, disaster management, security surveillance and combat field reconnaissance, to environment monitoring [15]. A WSN consists of a set of sensor nodes. Each sensor node is typically battery powered, resulting in a limited lifetime. Sensor nodes communicate with each other using radio signals. Experiments show that most energy of a sensor node is consumed in communication [15].

Clustering has been proposed to effectively increase the lifetime and the robustness of a WSN. By clustering, a WSN is partitioned into disjoint clusters. Each cluster has a cluster head. The cluster head of a cluster collects the sensed data from each member, performs data aggregations, and sends the aggregated data to the designated base station via a routing topology. Because of data aggregations, clustering reduces the overall data size and the number of communication sessions required to deliver the sensed data to the base station, effectively reducing the overall energy consumption of the entire sensor network [15]. Furthermore, clustering improves the robustness of a WSN. If a cluster head does not work, another sensor node of the same cluster can be the new cluster head without affecting the functionality of the entire network.

A typical clustering algorithm consists of two phases. In the first phase, it partitions all the sensor nodes of a WSN into disjoint clusters. In the second phase, it constructs a routing topology for inter-cluster communication. Many clustering algorithms have been proposed [1–7, 9, 10, 13, 16, 17, 19, 20]. LEACH [4] is the first clustering approach for WSNs. It uses a probability scheme to select cluster heads. Each cluster head performs data aggregation and sends the aggregated data to the base station directly. Subsequent clustering approaches improve LEACH in clustering and inter-cluster routing. Many clustering algorithms use multi-hop communication which aims at prolonging the network lifetime. Examples are EEUC [1], DSBCA [9], and ACT [7]. A major problem with all the previous clustering algorithms is that they do not have a precise energy model for quantifying the energy consumption of each sensor node when performing clustering and constructing a routing topology, resulting in unbalanced clusters and unbalanced energy consumptions of cluster heads.

In many applications, the sensed data need to be sent to the base station in a timely manner. In such applications, the designers need to ensure that the maximum hop distance for delivering any sensed data to the base station does not exceed a pre-specified value. In this paper, we investigate the following network lifetime and latency-aware clustering problem. Consider a WSN with one base station where all the sensor nodes are identical and have adjustable transmit power. Given a natural number k, partition all the sensor nodes into disjoint clusters and construct a routing tree for inter-cluster communication such that the lifetime of the WSN is maximized and the following constraints are satisfied. Firstly, the hop distance from each cluster to the base station does not exceed k. Secondly, all the members of a cluster can communicate with the cluster head directly. Thirdly, in each cluster, sensor nodes take turns to become a cluster head in an optimal way. Lastly, each cluster head performs perfect data aggregations. In perfect data aggregations, the size of the aggregated data after each aggregation is equal to one of the input data sizes. Examples of such data aggregations include computing the maximum temperature, the minimum temperature, and the average temperature sensed by all the sensor nodes. We make the following major contributions:

1. We propose a novel approach. Our approach consists of a polynomial-time heuristic for constructing clusters, a polynomial-time heuristic and an ILP algorithm for constructing a routing tree for inter-cluster communication. Unlike all the previous algorithms for constructing a routing tree for inter-cluster communication, our heuristic and the ILP algorithm use a precise model for calculating the energy consumption of each sensor node by considering the cluster head rotations within each cluster and the inter-cluster communication cost, making the routing tree more balanced in terms of energy consumption. As far as we know, our approach is the first one for this clustering problem with the maximum latency guarantee.

2. We have performed extensive simulations on network instances with 200, 400, 600, 800 and 1000 sensor nodes with the uniform distribution and the random distribution. The simulation results show that our approach using the ILP algorithm for constructing a routing tree for inter-cluster communication significantly improves the network lifetimes over two state-of-the-art approaches, namely MR-LEACH [3] and DSBCA [9]. The average lifetime improvements of our approach using the ILP algorithm over MR-LEACH and DSBCA are 35% and 62%, respectively. We have also compared our heuristic and our ILP algorithm for constructing a routing tree. The ratios of the average lifetime achieved by the heuristic and the ILP algorithm are 93% and 90%, for the uniform distribution and the random distribution, respectively.

The remainder of this paper is organized as follows. Section 2 summarizes the related work. Section 3 describes the network model. Section 4 presents our cluster construction algorithm. Section 5 proposes a polynomial-time heuristic and an ILP algorithm for constructing a routing tree. Section 6 shows the simulation results and analyses. Lastly, Section 7 concludes the paper.

2 Related Work

Many network lifetime-aware clustering approaches have been proposed [1–7, 9, 10, 13, 16, 17, 19, 20]. LEACH [4] is the first clustering approach for WSNs. In LEACH, cluster construction starts with a cluster head selection phase. Cluster heads are selected according to a predetermined probability threshold. A sensor node with a probability above the threshold is selected as a cluster head. Other sensor nodes join the nearest cluster head to form clusters. In the data transmission phase, each cluster head performs data aggregations and send the aggregated data to the base station directly. Direct communication between each cluster and the base station guarantees fast data delivery but results in large energy consumption for data transmission.

Culpepper et al. [2] propose HIT (Hybrid Indirect Transmission). Like LEACH, HIT uses a probability scheme to construct clusters. Unlike LEACH, HIT constructs a routing tree for both intra-cluster and inter-cluster communication to minimize energy consumption and network delay. Mao and Chengfa et al. propose EECS [20] and EEUC [1] clustering algorithms to prolong the network lifetime of a WSN. Since the cluster heads near the base station need to relay more packets from descendant clusters, EECS and EEUC construct clusters with unequal sizes such that the clusters closer to the base station have smaller sizes. EECS uses direct communicate between cluster heads and the base station. EEUC constructs a tree for inter-cluster routing.

Xiangning and Song [19] propose two modified protocols of LEACH, namely Energy-LEACH and Multihop-LEACH. Energy-LEACH improves the cluster head selection by considering the residual energy of each sensor node, and Multihop-LEACH improves the inter-cluster routing by constructing a multi-hop routing tree. Weichao et al. [17] propose LEACH-TM protocol. Unlike LEACH, LEACH-TM constructs a multi-hop routing tree by considering both the residual energy of each sensor node and the hop counts between each cluster to the base station.

Farooq et al. [3] propose MR-LEACH (Multi-hop Routing with Low Energy Adaptive Clustering Hierarchy). MR-LEACH partitions the network into different layers according to the minimum transmission range of all the nodes. Similar to LEACH, MR-LEACH uses a probability scheme to select cluster heads. Unlike LEACH, MR-LEACH constructs a routing tree incrementally by considering the minimum transmission range of each cluster head. Kumar et al. [6] propose the MCR protocol which introduces the concept of advanced and super advanced nodes along with the normal nodes. Advanced nodes are more powerful than the normal nodes in terms of battery capacity and storage capacity. Super advanced nodes are more powerful than the advanced nodes. Cluster heads are selected by applying a probability scheme which assigns different probabilities to normal, advanced and super nodes. Cluster heads send the aggregated data to the base station via a multi-hop routing tree.

Liu et al. [10] propose DEECIC (Distributed Energy-Efficient Clustering with Improved Coverage) which aims to save the energy consumption of each sensor node. It constructs the least number of clusters to cover the whole network. For the inter-cluster communication, DEECIC does not propose any algorithms for constructing a routing topology. Lai et al. [7] propose ACT (Arranging Cluster sizes and Transmission ranges for WSN). Like EEUC, ACT constructs unequal size clusters by reducing the sizes of the clusters near the base station. Unlike EEUC, the cluster size of each cluster in ACT is selected based on the cluster head's distance to the base station together with the cluster head's energy consumption. For data transmission, ACT constructs a multi-hop routing DAG for inter-cluster routing. Ying et al. [9] propose DSBCA (Balanced Clustering Algorithm with Distributed Self-Organization for WSNs). DSBCA takes the node density and the residual energy into account to construct clusters and constructs a routing tree for inter-cluster communication.

Sabet et al. [13] propose a decentralized, energy efficient clustering and routing approach which aims at reducing the overheads caused by extra control message transmissions. The clustering algorithm selects cluster heads based on the energy consumption, the residual energy, and the distance to the base station of each sensor node. The routing algorithm constructs a routing tree for inter-cluster communication. Jorio et al. [5] propose LESCA (Location-Energy Spectral Cluster Algorithm). LESCA uses several factors, including the residual energy, the average energy, the distance to the base station, and the distance to the cluster center to determine clusters and select cluster heads. It does not consider routing topology construction.

Leu et al. [8] propose REAC-IN (Regional Energy-Aware Clustering with

Isolated Nodes) that considers isolated nodes. Isolate nodes are the sensor nodes that do not belong to any clusters. REAC-IN selects cluster heads based on the weight of each cluster. The weight of each sensor node is determined according to its residual energy and the regional average energy of all the sensor nodes in each cluster. To prolong the network lifetime, REAC-IN uses the regional average energy and the distance between sensor nodes and the sink to determine whether an isolated sensor node sends its data to a cluster head or to the base station. In [14], Sabor et al. propose UMBIC (Unequal Multi-hop Balanced Immune Clustering). In the cluster formation phase, UMBIC partitions the network into clusters with unequal sizes based on the distance to the base station and the residual energy of each node. In the routing construction phase, UMBIC constructs a routing tree for inter-cluster communication. Xia et al. [18] propose UCCGRA that forms clusters with unequal sizes taking the average transmission power of each sensor node into account, and constructs a multi-hop routing tree by considering the location and connectivity of each cluster head.

Compared with all the previous approaches, our approach aims at constructing clusters not only having the maximum network lifetime but also meeting the maximum latency requirement. Furthermore, unlike all the previous approaches, our approach constructs a precise model for computing the total energy consumed by each sensor node in both intra-cluster communication and inter-cluster communication.

3 Network Model

The target WSN consists of a set V of N identical sensor nodes with the same initial energy, deployed on a 2D plane to continuously monitor the environment. The location of each sensor node is known. There is one base station. All the sensor nodes and the base station have a fixed location. The WSN is to be partitioned into disjoint clusters with one cluster head in each cluster. In each cluster, all the members take turns to be a cluster head for a specific time period in a particular order. In each unit time, each sensor node generates one packet of the sensed data, and sends the packet to its cluster head. After collecting the sensed data from all the members, the cluster head of each cluster performs data aggregation, and sends the resulting packet to its parent in the routing tree. We assume perfect data aggregations, that is, the output of each data aggregation performed by a cluster head has the same size as each packet received by the cluster head. The transmit power of each sensor node is adjustable, leading to a variable transmission range. The network lifetime is defined as the number of time units until the first sensor node uses up its energy.

We follow the same energy model functions used in most clustering algorithms [1, 4, 9]. The energy for transmitting a packet of l bits over a distance d is modelled as follows:

$$E_{Tx}(l,d) = lE_{elec} + l\epsilon d^{\alpha} = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, & d < d_0 \\ lE_{elec} + l\epsilon_{mp}d^4, & d > d_0 \end{cases} \tag{3.1}$$

The energy for receiving a packet of l bits is given as follows:

$$E_{Rx}(l) = lE_{elec} \tag{3.2}$$

In the above functions, E_{elec} is the electronic energy that depends on factors such as the digital coding, modulation, and filtering. ϵ_{fs} and ϵ_{mp} are the amplifier energies required to maintain an acceptable signal to noise ratio. The path loss exponent is d^2 if the transmitter and receiver are within certain threshold distance d_0. Otherwise, the path loss exponent is d^4. We assume the amount of energy for a cluster head to perform data aggregations is E_{DA} nJ per bit. We consider the energy consumption of each sensor node consumed in transmitting and receiving data only.

4 Cluster Construction

In order to meet the latency requirement, our clustering heuristic partitions the area where all the sensor deployed into k disjoint layers. All the sensor nodes in each layer forms a group $G_i(i = 1, 2, \cdots, k)$. Our group construction algorithm works as follows:

1. Find a sensor node v_i with the shortest Euclidean distance d_{min} to the base station and a sensor node v_j with the longest Euclidean distance d_{max} to the base station.
2. Draw two circles C_1 and C_{k+1} centred at the base station such that v_i is on the circumference of C_1 and v_j is on the circumference of C_{k+1}.
3. Let r_1 be the radius of C_1 and r_{k+1} be the radius of C_{k+1}.
4. Draw $k-1$ concentric circles $C_i(i = 2, \cdots, k)$ centred at the base stations, each of which has a radius of $r_1 + (i-1) * (r_{k+1} - r_1)/k$.
5. $G_i(i = 1, 2, \cdots, k-1)$ is a set of the sensor nodes in the area between circle C_i and C_{i+1}, excluding the sensors on the circumference of C_{i+1}.
6. G_k is a set of the sensors in the area between C_k and C_{k+1}.

Figure 4.1 shows the groups of a WSN with $k = 3$.

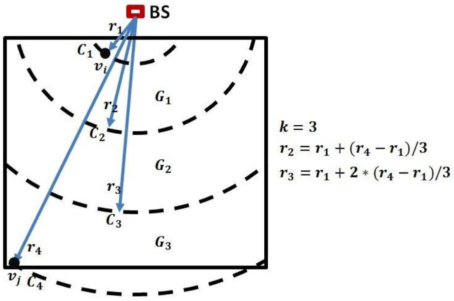


Figure 4.1: The groups of a WSN

Next, our clustering heuristic partitions all the sensor nodes of each group into disjoint clusters. Within each cluster, sensor nodes take turns to be the cluster head for an optimal period. The transmit power of each sensor node is adjusted dynamically such that all the sensor nodes of a cluster can communicate with their cluster head directly.

In the cluster construction phase, we ignore the energy consumption of inter-cluster communication and aim at constructing clusters such that each sensor node consumes the same amount of energy. Our algorithm for constructing clusters works as follows:

For each group $G_i (i = 1, 2, \cdots, k)$,

1. Establish a polar coordinate system with the base station as the pole and an arbitrary direction as the polar axis.
2. Compute the angular coordinate $\alpha_j (0 \leq \alpha_j \leq 360°)$ of each sensor node v_j in G_i.
3. Let L_i be a list of all the sensor nodes in G_i sorted in a non-decreasing order of angular coordinates.
4. Let $L_i[j]$ and $L_i[(j + 1)\ mod\ |G_i|]$ are the two sensor nodes such that the absolute value of the difference of their angular coordinates is the maximum among all the pairs of the adjacent sensor nodes in L_i.
5. Establish another polar coordinate system with the base station as the pole and the vector from the base station to the sensor node $L_i[j]$ as the polar axis.
6. Create a queue Q_i of all the sensor nodes in G_i sorted in a non-decreasing order of their angular coordinates.
7. Repeat the following steps to construct clusters $c_s \in G_i$ until Q_i is empty:
    (a) Create an empty cluster $c_s (s = 1, 2, \cdots)$.
    (b) Keep removing a sensor node from Q_i and add it to c_s until the minimum average energy consumption of all the sensor nodes in c_s exceeds a threshold $\overline{LA}$.

In the next two subsections, we will describe how to compute the minimum average energy consumption of a cluster and the threshold $\overline{LA}$.

4.1 Computing Minimum Average Energy Consumption of A Set of Sensor Nodes

In this section, we describe how to calculate the minimum average energy consumption of a set of sensor nodes by considering only the energy consumption of intra-cluster communication and data aggregation. The energy consumed for inter-cluster routing will be considered when constructing a spanning tree for inter-cluster routing.

We divide time into equally lengthed *rounds*. The length of each round is T. Given a set $S = \{v_1, v_2, \cdots, v_s\}$ of s sensor nodes, all the sensor nodes in S take turns to be a cluster head in each round. In each round, assume that each sensor node v_i running as the cluster head for a time period of t_i. Our objective is to find an optimal value of t_i for each sensor node in S such that all the sensor nodes consume the same amount of energy in a round. Let $d_{i,j}$ be the Euclidean distance between a sensor node v_i and a sensor node v_j. T and each t_i satisfy the following constraint:

$$T = t_1 + t_2 + \cdots + t_s \tag{4.1}$$

Assume that each sensor node generates a packet of l bits in a unit time. If a sensor node running as the cluster head, it will receive $s-1$ packets and perform data aggregation, where s is the number of nodes in the cluster. Thus, the energy consumption of a cluster head in a unit time is as follows:

$$E_{CH}(l) = (s-1) * E_{Rx}(l) + s * l * E_{DA} \tag{4.2}$$

For non-cluster head sensor nodes, they only send their packets to the cluster head. Their energy consumption can be calculated by Equation (3.1).

Thus, for a particular sensor node v_i, it takes t_i time running as a cluster head and $T - t_i$ time running as a cluster member. So the total energy consumption E_{T_i} of a sensor node v_i in a round is calculated as follows:

$$E_{T_i} = E_{CH}(l) * t_i + \sum_{v_j \in S, v_j \neq v_i} E_{Tx}(l, d_{i,j}) * t_j \tag{4.3}$$

Our objective is to make all the sensor nodes consume the same amount of energy. Therefore, we have the following constraint:

$$E_{T_1} = E_{T_2} = \cdots = E_{T_s} = LA \tag{4.4}$$

Solving the linear equations (4.1), (4.3) and (4.4), we can obtain the value of LA together with time period of each sensor node running as the cluster head $\{t_1, t_2, ..., t_s\}$.

4.2 Calculating the Threshold $\overline{LA}$

The threshold $\overline{LA}$ is used to determine if a set of sensor nodes form a cluster. $\overline{LA}$ is obtained by constructing a set of tentative clusters of the WSN. It is defined as follows:

$$\overline{LA} = \frac{\sum_{i=1}^{m} LA_i}{m} \tag{4.5}$$

where LA_i is the minimum average energy consumption of a tentative cluster c_i and m is the total number of tentative clusters.

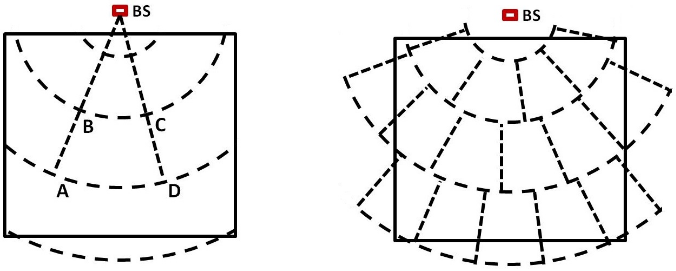
The tentative clusters are constructed as follows:

1. Draw $k+1$ circles C_1, C_2, $\cdots$, C_{k+1} as before.
2. For each region enclosed by two adjacent circles C_i and $C_{i+1} (i = 1, 2, \cdots, k)$, partition it into sampling sections.
3. For each sampling section, if it contains at least one sensor node, all the sensor nodes in this sampling section form a tentative cluster.
4. For each region enclosed by two adjacent circles C_i and $C_{i+1} (i = 1, 2, \cdots, k)$, if a tentative cluster has at most two sensor nodes, merge it with an adjacent tentative cluster.

A sampling section of each region enclosed by two adjacent circles C_i and $C_{i+1} (i = 1, 2, \cdots, k)$ is constructed as follows:

- Draw a line from the base station.
- Let A and B be the two points of intersections of the line and the two circles C_i and C_{i+1}, respectively. Find a point D on the circumference of C_{i+1} such that the Euclidean distance from A to D is equal to $r_{i+1} - r_i$. Then, draw a line from D to the base station. Let C be the point of intersection of the line and C_i. The area enclosed by AB, BC, CD and DA is a sampling section.

Figure 4.2(a) shows an example of a sample section and figure 4.2(b) shows an example of tentative clusters.



(a) Sampling section (b) Tentative clusters

Figure 4.2: A sampling section and tentative clusters

5 Routing Tree Construction

Once the network is partitioned into clusters, we construct a spanning tree rooted at the base station for inter-cluster routing, aiming at maximizing the network lifetime. We propose two approaches, a polynomial-time heuristic, and an ILP (Inter Linear Programming) algorithm.

5.1 Candidate Graph Construction

Both our heuristic and ILP algorithm convert all the clusters into an undirected graph, namely candidate graph. In the candidate graph $G = (V \cup \{BS\}, E)$, the set of vertices consists of all the clusters and the base station (BS). The set E of edges is constructed as follows:

1. For each cluster $c_i \in V$, compute its centroid.
2. For each cluster c_i in group G_1, the edge (c_i, BS) is in E.
3. For each cluster c_i in the group $G_j (j = k, k-1, \cdots, 2)$, all the edges between c_i and clusters in the group G_{j-1} are constructed as follows:
    (a) Let c_s be a cluster in G_{j-1} such that the centroids of c_i and c_s has the shortest Euclidean distance, $h_{i,s}$ the Euclidean distance between the centroids of c_i and c_s, and P_i a set of clusters in the group G_{j-1} such that the Euclidean distance between the centroids of c_i and each cluster in P_i is at most $h_{i,s} * \lambda$, where λ is a pre-defined constant.

(b) For any cluster $c_t \in P_i$, the edge (c_t, c_i) is in E.

Figure 5.1 shows a candidate graph, where λ is set to 2. The value of λ affects the performance of the inter-cluster routing tree. Typically, a small value for λ will suffice.

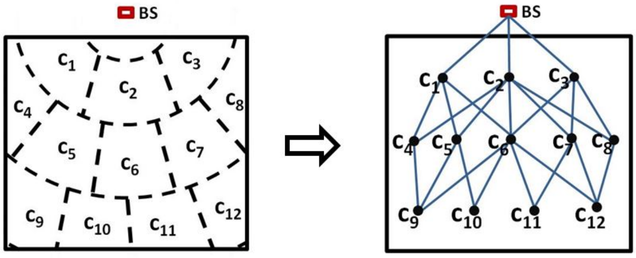


Figure 5.1: Converting clusters into the candidate graph

5.2 Computing Transmit Energy Consumption Between Adjacent Clusters

When considering a cluster c_j as the parent of a cluster c_i, we need to know the additional energy consumption incurred in both c_i and c_j. The cluster head rotation complicates the calculation of the additional energy consumption. Next, we describe how to calculate the additional energy consumption brought to clusters due to inter-cluster communication.

For each cluster c_i, we specify a particular order for each sensor node in c_i running as the the cluster head as follows:

- All the sensor nodes in c_i take turns to be the cluster head in a non-descending order of their Euclidean distances to the base station.

Given two clusters c_i and c_j, let $U_{i(x),j}$ be the energy consumption of a sensor node x when running as the cluster head in cluster c_i and sending packets to cluster c_j in one round, and $Q_{i(x),j}$ the energy consumption of a sensor node x when running as the cluster head in cluster c_i and receiving packets from cluster c_j in one round.

Based on equation 3.2, in a unit time, the energy consumption of x in c_i when receiving a packet from c_j is equal to $E_{Rx}(l)$. The duration of x running as the cluster head in each round is t_x. When x running as the cluster head, it needs to not only receive a packet from each member, but also perform the data aggregation. Therefore, $Q_{i(x),j}$ is calculated as follows:

$$Q_{i(x),j} = (E_{Rx}(l) + E_{DA}) * t_x \tag{5.1}$$

In order to compute $U_{i(x),j}$, we need to construct the cluster head schedules for c_i and c_j. A cluster head schedule specifies the time period of each member acting as the cluster head. Figure 5.2 shows the cluster head schedules of two cluster c_i and c_j in a round, where c_i and c_j have m nodes and n nodes, respectively.

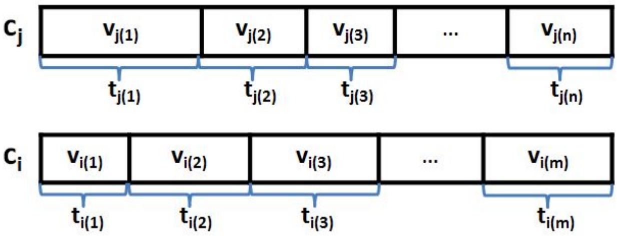


Figure 5.2: An example of cluster head schedules

Notice that the sum of the time periods of all the sensor nodes in c_i is equal to T, the duration of a round. All the time periods t_i are computed by using the approach described in Section 4.1. For each sensor node x in cluster c_i, we need to partition its time period t_x running as the cluster head into multiple sub-periods. Each of the sub-periods represents that a different sensor node in c_j is the cluster head of c_j during the sub-period.

Figure 5.3 shows the sub-periods of each sensor node in c_i. For example, the sensor node $v_{i(2)}$ has two sub-periods, one for the sensor node $v_{j(1)}$, and the other for the sensor node $v_{j(2)}$.

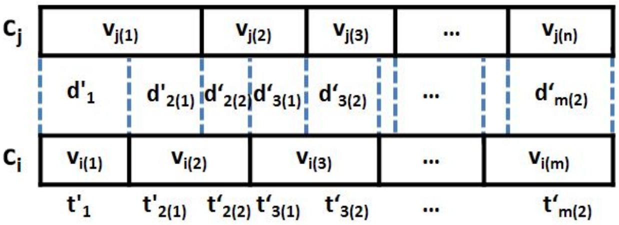


Figure 5.3: A cluster head schedule for c_i with sub-periods

Given the time period t_x of a sensor node x in c_i and a sub-period $t'_s \in t_x$, let d'_s be the Euclidean distance between x and the cluster head of c_j during the sub-period t'_s. $U_{i(x),j}$ is calculated as follows:

$$U_{i(x),j} = \sum_{t'_s \in t_x} E_{Tx}(l, d'_s) * t'_s \tag{5.2}$$

5.3 Heuristic

In this subsection, we present a polynomial-time heuristic for constructing a routing tree for inter-cluster communication. For ease of description, we introduce the following notations:

- W_i: the weight of cluster c_i, which represents the energy consumption of c_i. The initial value of W_i is set to LA_i. W_i will be updated whenever a cluster selects c_i as its parent or c_i is assigned a parent cluster.

- $EW_{i,j}$: the edge weight between c_i and c_j. $EW_{i,j}$ is equal to the maximum energy consumption of c_i and c_j based on the current partial routing tree.

Our heuristic constructs a routing tree in a bottom-up way, starting with the clusters in the group G_k. At each step, for each cluster c_i in the group $G_j(j = k, k-1, \cdots, 2)$, it selects a cluster among all the immediate ancestors in the candidate graph as its parent such that the maximum energy consumption of c_i and its parent cluster is minimized.

Consider that our heuristic selects a parent for each cluster c_i in a specific group G_j. We define a priority of each cluster c_i in G_j to be the number of immediate ancestors of c_i in the candidate graph, where a smaller number implies a higher priority. Our heuristic works for all the clusters in G_j as follows:

1. Compute the priority of each cluster in G_j.
2. Repeat the following until each cluster in G_j is assigned a parent:
   (a) Select a cluster c_i which has the highest priority among all the clusters in G_j without a parent.
   (b) Compute the edge weight of each edge between c_i and its immediate ancestors in the candidate graph. The edge weight between c_i and one of its immediate ancestors c_p is computed as follows:
      i. Find the largest $U_{i(max),p}$ of all the sensor nodes in cluster c_i.
      ii. Find the largest $Q_{p(max),i}$ of all the sensor nodes in cluster c_p.
      iii. The edge weight between cluster c_i and c_p is $EW_{i,p} = \max\ \{W_i + U_{i(max),p}, W_p + Q_{p(max),i}\}$.
   (c) Select a cluster c_s in G_{j-1} as the parent of c_i satisfying that c_s is an immediate ancestor in the candidate graph and the edge weight between c_i and c_s is the smallest among the weights of the edges between c_i and all its immediate ancestors in the candidate graph.
   (d) Update the weights of clusters c_i and c_s by $W_i = W_i + U_{i(max),s}$, $W_s = W_s + Q_{s(max),i}$.

5.4 ILP Algorithm

The objective of our ILP algorithm is to construct a shortest path spanning tree rooted at the base station such that the network lifetime is maximized. For each cluster c_i in the candidate graph G, we introduce the following additional notations:

- PS_i: a set of the parents of a cluster c_i in the candidate graph.
- CS_i: a set of the children of a cluster c_i in the candidate graph.

In order to represent each edge uniquely, we stipulate that an edge $E_{i,j}$ denotes the edge between cluster c_i in group G_{s+1} and cluster c_j in group G_s.

For each edge $E_{i,j}$, we introduce a binary decision variable $X_{i,j}$ as follows:

$$X_{i,j} = \begin{cases} 1 & c_j \text{ is selected as the parent of } c_i \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

Therefore, for each cluster c_i, we have the following parent selection constraint:

$$\sum_{c_j \in PS_i} X_{i,j} = 1 \tag{5.4}$$

The above constraint implies that among all the candidate parents of c_i, only one cluster can be selected as the parent.

Next, we derive the energy constraint for each sensor node. For each sensor node in cluster c_i, its intra-cluster energy consumption is LA_i. The energy consumption of a sensor node x in c_i when acting as the cluster head and sending packets to its parents is $\sum_{c_s \in PS_i}(X_{i,s} * U_{i(x),s})$, and the energy consumption of a sensor node x when acting as the cluster head of c_i and receiving packets from children nodes is $\sum_{c_t \in CS_i}(X_{t,i} * Q_{i(x),t})$. Therefore, the total energy consumption W_x of a sensor node x satisfies the following constraint:

$$W_x = LA_i + \sum_{c_s \in PS_i}(X_{i,s} * U_{i(x),s}) + \sum_{c_t \in CS_i}(X_{t,i} * Q_{i(x),t}) \tag{5.5}$$

The maximum total energy consumption W_{max} of all the sensor nodes in V is calculated as follows:

$$W_{max} = \max_{x \in V}\{W_x\} \tag{5.6}$$

Our optimization objective function is shown as follows:

$$\min W_{max} \tag{5.7}$$

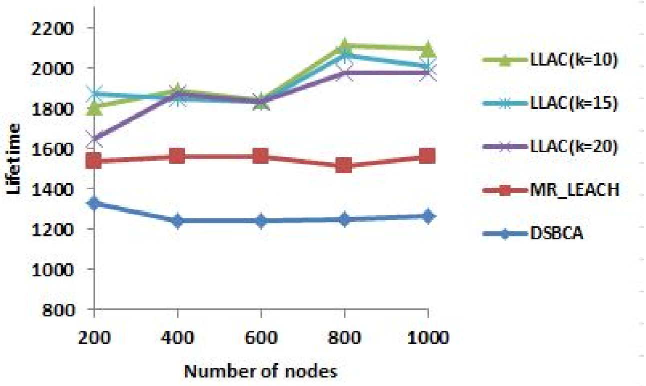
6 Simulation Results

In order to evaluate the performance of our approach, we have implemented our approach, and the two state-of-the-art approaches MR-LEACH [3] and DSBCA [9] that consider the same aggregation model as ours. We generate network instances of 200, 400, 600, 800, and 1000 sensor nodes deployed in a $200m*200m$ square area, with the uniform distribution and the random distribution. For each network instance, we choose 10, 15 and 20 for k, respectively. Each sensor node generates one packet of data in each time unit. Other parameters are given in Table 6.1.

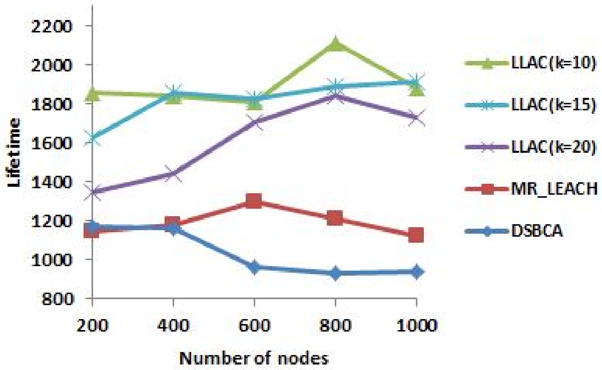
Table 6.1: Simulation Parameters

Parameter	Value
Base station location	(100,0) m
Initial energy	1 J
E_{elec}	50 nJ/bit
ϵ_{fs}	10 $pJ/bit/m^2$
ϵ_{mp}	0.0013 $pJ/bit/m^4$
d_0	87 m
E_{DA}	5 nJ/bit
Data packet size	4000 $bits$

The hardware platform of our ILP solver is Intel Core i5-3470 with a clock speed of 3.20 Ghz, a memory size of 8 GB and a cache size of 8192 MB. The solver for ILP problems is *Intlinprog* Solver of MATLAB. The longest running time of our approach using the ILP algorithm for constructing a routing tree is 43 seconds, which occurs on a network instance of 1000 sensor nodes with $k = 10$. The running time of our approach using the heuristic is negligible.



(a) Uniform distribution



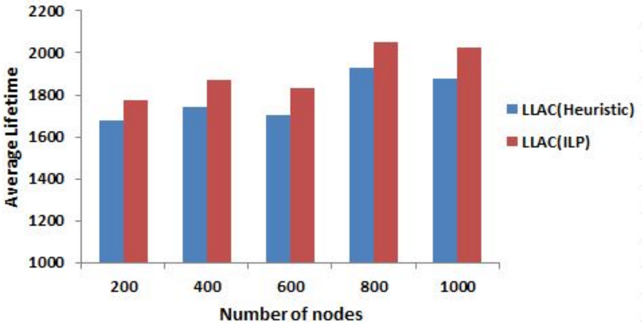
(b) Random distribution

Figure 6.1: Comparison of network lifetimes

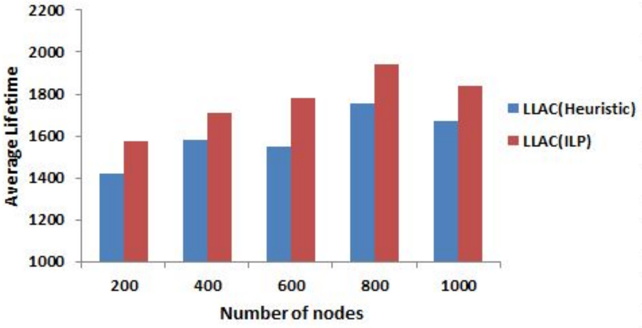
Figure 6.1 shows the network lifetimes achieved by MR-LEACH, DSBCA, and our approach using the ILP algorithm for constructing a routing tree, where LLAC (Latency and Lifetime-Aware Clustering algorithm) denotes our approach. We can see that, for both distributions, LLAC achieve better performance. Compared to DSBCA, the maximum improvement, the minimum improvement, and the average improvement of our approach are 126%, 6%, and 62%, respectively. Compared to MR-LEACH, the three improvements are 74%, 8%, and 35%, respectively.

The key reason why our approach significantly outperforms both MR-LEACH and DSBCA is that our approach uses a precise energy consumption model when constructing clusters and a routing tree for inter-cluster communication.

Whereas, both MR-LEACH and DSBCA consider neither the energy consumption of each sensor node on intra-cluster communication when constructing clusters nor the energy consumption of each sensor node on inter-cluster communication when constructing a routing tree.



(a) uniform distributions



(b) random distributions

Figure 6.2: Comparison of network lifetimes between our heuristic and ILP algorithm

We have also compared our heuristic and the ILP algorithm for constructing a routing tree. Figure 6.2 shows the average network lifetimes of all the network instances using the heuristic and the ILP algorithm, where the vertical axis denotes the average network lifetime when k is set to 10, 15, and 20, respectively. We can see that the performance of our heuristic is competitive with the ILP algorithm. The average ratio between the network lifetime achieved by our approach using the heuristic and the network lifetime achieved by our approach using the ILP algorithm is 93% for the uniform distribution, and 90% for the random distribution.

We have also evaluated the impact of k on the network lifetime for our approach using the ILP algorithm for constructing a routing tree. We generate a network of 400 sensor nodes in the uniform distribution and choose k from 2 to 24 with an increment of 1. Other experimental settings are the same as before. Figure 6.3 shows how the network lifetime changes with k. Initially, the network lifetime increases as k increases. After the network lifetime reaches its

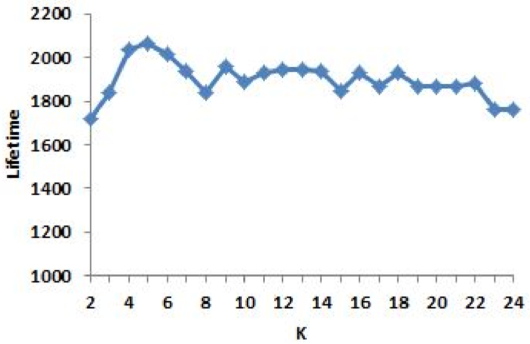


Figure 6.3: Network lifetimes with different values of k

maximum at $k = 5$, it decreases until $k = 9$. After that, the network lifetime fluctuates without significant changes.

The reasons are analysed as follows. For $k \leq 5$, increasing k will decrease the size of each cluster dramatically. So the average distance from each sensor node to its cluster head decreases, reducing the energy consumption of each sensor node in transmitting its own data to its cluster head. Furthermore, the energy consumption on intra-cluster communication dominates the network lifetime. As a result, the network lifetime increases as k increases. For $k > 5$ and $k \leq 9$, the first dead sensor node is in the first layer closest to the base station. Increasing k reduces cluster sizes of the first layer so that each sensor node spends more time running as the cluster head. For intra-cluster communication, decreasing cluster size does not have significant impact on the network lifetime as the number of packets received by each cluster head also decreases. For inter-cluster communication, the distance of each sensor node to base station does not change. Therefore, increasing the time for a sensor node running as the cluster head results in more energy consumption of each sensor node in the first layer. Consequently, the network lifetime decreases. For $k > 9$, neither intra-cluster communication nor inter-cluster communication dominates the network lifetime any more. The fluctuation is caused by the uncertainty of the number of clusters in each layer. Increasing k will result in the decrease of the sample area in the cluster formation phase. At the same time, the number of sensor nodes in each layer also decreases. Hence, the number of children of the cluster with the highest workload may increase or decrease, resulting in the network lifetime fluctuation.

7 Conclusion

We investigate the lifetime and latency-aware clustering problem in WSNs and propose a polynomial-time heuristic for constructing clusters, a polynomial-time heuristic and an ILP algorithm for constructing a lifetime and latency-aware, shortest path routing tree for inter-cluster communication such that the maximum hop distance between any cluster to the base station is at most k. We have evaluated our approach using network instances with 200, 400, 600, 800 and 1000 nodes in the uniform distribution and the random distribution, and two different values, 10 and 20, for k. The simulation results show that our approach significantly outperforms MR-LEACH and DSBCA. To our knowledge, this is the first approach to this problem.

We assume that each cluster head performs perfect data aggregation. In some applications, the data sensed by each sensor node must be delivered to the base station. For those applications, different energy consumption model must be derived. Hence, our approach may not be efficient. We will investigate the clustering and routing tree construction problems for those applications in the future research. Furthermore, the assumption on perfect link quality made in this paper is not realistic. The transmission between two sensor nodes may not be always reliable. Another future research problem is to construct lifetime and latency-aware clusters and a routing tree by considering link quality.

References

[1] Guihai Chen, Chengfa Li, Mao Ye, and Jie Wu. An unequal cluster-based routing protocol in wireless sensor networks. *Wireless Networks*, 15(2):193–207, 2009.

[2] Benjamin J Culpepper, Lan Dung, and Melody Moh. Design and analysis of hybrid indirect transmissions (hit) for data gathering in wireless micro sensor networks. *Mobile Computing and Communications Review (SIGMOBILE)*, 8(1):61–83, 2004.

[3] Muhamnmad Omer Farooq, Abdul Basit Dogar, and Ghalib Asadullah Shah. Mr-leach: multi-hop routing with low energy adaptive clustering hierarchy. In *Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM)*, pages 262–268. IEEE, 2010.

[4] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *33rd annual Hawaii international conference on system sciences*, pages 1–10. IEEE, 2000.

[5] Ali Jorio, Sanaa El Fkihi, Brahim Elbhiri, and Driss Aboutajdine. An energy-efficient clustering routing algorithm based on geographic position and residual energy for wireless sensor network. *Journal of Computer Networks and Communications*, 2015, 2015.

[6] Dilip Kumar, Trilok C Aseri, and RB Patel. Multi-hop communication routing (mcr) protocol for heterogeneous wireless sensor networks. *International Journal of Information Technology, Communications and Convergence*, 1(2):130–145, 2011.

[7] Wei Kuang Lai, Chung Shuo Fan, and Lin Yan Lin. Arranging cluster sizes and transmission ranges for wireless sensor networks. *Information Sciences*, 183(1):117–131, 2012.

[8] Jenq-Shiou Leu, Tung-Hung Chiang, Min-Chieh Yu, and Kuan-Wu Su. Energy efficient clustering scheme for prolonging the lifetime of wireless sensor network with isolated nodes. *Communications Letters*, 19(2):259–262, 2015.

[9] Yifan Liao, Hongsheng Qi, and Wenyuan Li. Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. *Sensors Journal*, 13(5):1498–1506, 2013.

[10] Zhixin Liu, Qingchao Zheng, Liang Xue, and Xinping Guan. A distributed energy-efficient clustering algorithm with improved coverage in wireless sensor networks. *Future Generation Computer Systems*, 28(5):780–790, 2012.

[11] Sabbir Mahmud and Hui Wu. Lifetime aware deployment of k base stations in wsns. In *International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 89–98. ACM, 2012.

[12] Sabbir Mahmud, Hui Wu, and Jingling Xue. Efficient energy balancing aware multiple base station deployment for wsns. In *European Conference on Wireless Sensor Networks*, pages 179–194. Springer, 2011.

[13] Maryam Sabet and Hamid Reza Naji. A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks. *AEU-International Journal of Electronics and Communications*, 69(5):790–799, 2015.

[14] Nabil Sabor, Mohammed Abo-Zahhad, Shigenobu Sasaki, and Sabah M Ahmed. An unequal multi-hop balanced immune clustering protocol for wireless sensor networks. *Applied Soft Computing*, 43:372–389, 2016.

[15] Santar Pal Singh and SC Sharma. A survey on cluster based routing protocols in wireless sensor networks. *Procedia Computer Science*, 45:687–695, 2015.

[16] Sudhanshu Tyagi and Neeraj Kumar. A systematic review on clustering and routing techniques based upon leach protocol for wireless sensor networks. *Journal of Network and Computer Applications*, 36(2):623–645, 2013.

[17] Wang Weichao, Du Fei, and Xu Qijian. An improvement of leach routing protocol based on trust for wireless sensor networks. In *5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, pages 1–4. IEEE, 2009.

[18] Hui Xia, Rui-hua Zhang, Jia Yu, and Zhen-kuan Pan. Energy-efficient routing algorithm based on unequal clustering and connected graph in wireless sensor networks. *International Journal of Wireless Information Networks*, pages 1–10, 2016.

[19] Fan Xiangning and Song Yulin. Improvement on leach protocol of wireless sensor network. In *International Conference on Sensor Technologies and Applications*, pages 260–264. IEEE, 2007.

[20] Mao Ye, Chengfa Li, Guihai Chen, and Jie Wu. Eecs: an energy efficient clustering scheme in wireless sensor networks. In *Performance, Computing, and Communications Conference*, pages 535–540. IEEE, 2005.