A Novel Cooperative Clustering Approach for Wireless Sensor Networks

Yuning He

(Doctoral Program in Social Systems Engineering)

Advised by Yongbing Zhang  Assoc. Professor

Submitted to the Graduate School of Systems and Information Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Engineering at the University of Tsukuba

January 2006
Abstract

A sensor network consists of many, spatially sensors to monitor environment such as temperature, sound, or pollutants. Each sensor is equipped with a radio transceiver, a small microcontroller, and an energy source for transporting data to a distant base station (BS). Network lifetime, scalability, and load balancing are important requirements for many data gathering sensor network applications. Network clustering is an effective approach for achieving these goals wherein sensors in the network are grouped into clusters.

As introduced in almost conventional clustering approaches, in each cluster, there is a particular node existing in each cluster, which is called head node, while the other nodes are called member nodes. The head node is in charge of collecting data from its member nodes in the cluster, aggregating these data into one single data packet, and transmitting the compressive data to the BS. Since receiving, aggregating, and transmitting to the BS are all accomplished by the head node, energy of the head node is soon depleted.

In this paper, we propose a new data gathering approach wherein both data gathering and data aggregating are performed by the same node but data reporting to the BS may be done by another node. In this way, the energy consumption is balanced over the sensor network. The proposed approach can be directly extended from the traditional approaches without big modification. The simulation results show that the proposed algorithm achieves better performance than other existing algorithms.
### Contents

1 Introduction ............................................................. 1
   1.1 Research Background ........................................... 1
       1.1.1 Sensor Node .............................................. 1
       1.1.2 Sensor Network ........................................... 2
   1.2 Research Objective ............................................. 3
   1.3 Organization of the Thesis .................................... 3

2 Related Works ......................................................... 5

3 Problem Outline ...................................................... 7
   3.1 Network Model .................................................. 7
   3.2 Problem Statement ............................................. 9
   3.3 Data Transmission ............................................. 10

4 Proposed Approach .................................................. 11
   4.1 Analysis of Multi-hop Communication ......................... 11
   4.2 Protocol Architecture ......................................... 15
       4.2.1 Cluster Formation ....................................... 16
       4.2.2 Forwarder Selection ..................................... 17
       4.2.3 Data Communication ..................................... 17
       4.2.4 Pseudo-code of the Proposed Algorithm ................ 19

5 Experiments ............................................................ 21
   5.1 Evaluation Measure ............................................ 21
   5.2 Experiment Environment ....................................... 21
   5.3 Experimental Results .......................................... 23
       5.3.1 Comparisons with HEED and LEACH .................... 23
       5.3.2 Effects of Various Parameter Settings ................. 24

6 Conclusions ............................................................ 30
## List of Figures

1.1 The components of a sensor node. ........................................... 1
1.2 The structure of sensor networks. ........................................... 2

3.1 The traditional clustering algorithm. ....................................... 9
3.2 Single-hop and multi-hop data transmission. ............................. 10

4.1 Minimum transmission range of two neighbor collectors. ............... 12
4.2 Connective graph of collectors. ............................................. 13
4.3 The minimum transmission range to the BS. ............................... 14
4.4 Time line of the proposed approach. ...................................... 15
4.5 Procedure of HEED approach. .............................................. 16
4.6 Procedure of the proposed approach. ..................................... 16
4.7 Data aggregation and reporting to the BS in the proposed algorithm. 19
4.8 Pseudo-code of the proposed algorithm. .................................. 20

5.1 Network lifetime using LEACH vs. proposed approach. ................. 25
5.2 Network lifetime using HEED vs. proposed approach. ................... 26
5.3 Network lifetime with various node numbers. ............................. 27
5.4 Network lifetime with various cluster ranges. ............................. 28
5.5 Network lifetime with various positions of the BS. ...................... 29
Chapter 1

Introduction

1.1 Research Background

1.1.1 Sensor Node

Advantages in sensor technology, low-power electronics, and low-power radio frequency (RF) design have enabled the development of small, relatively inexpensive and low-power sensors, which can be connected via a wireless network [1, 2, 3].

![Diagram of a sensor node components](image)

Figure 1.1: The components of a sensor node.

A sensor node is made up of four basic components, as shown in Figure 1.1: a sensing unit, a processing unit, a communication unit. Some additional application-dependent components such as locating system, power generator, and mobilizer are also available for a sensor node. The sensing unit is responsible for observing phenomenon and feeding into the...
processing unit. The processing unit, which is generally associated with a small storage, takes charge of processing the sensed data and carrying out the assigned sensing tasks. The communication unit is responsible for receiving and sending data and connecting the node to a network. One of the most important components of a sensor node is the power unit. It provides energy for various kinds of duties. The power unit may be supported by a power generator unit to get continuous power supply. There are also other subunits that are application-dependent. The locating system helps a node acquire its location information. A mobilizer may sometimes be needed to move the sensor node when it is required to carry out the assigned tasks.

1.1.2 Sensor Network

A sensor network is composed of a large number of sensor nodes, which are deployed randomly and left unattended inside the phenomenon or very close to it. As shown in Figure 1.2, sensor nodes collaborate to observe the surroundings and send the information back to the base station (BS), where network managers can access to the sensor network via internet or other media. The communication between a sensor and the BS is in a multihop communication way, which means data packets originated from a sensor are transmitted to the BS via multiple intermediate sensors.

![Figure 1.2: The structure of sensor networks.](image)

The characteristics of sensor networks ensure a wide range of applications. Some of the application areas are health, military, and home. In military applications, for example, the rapid deployment, self-organization, and fault tolerance characteristics of sensor networks are suitable for military communications, surveillance, targeting systems etc. In health applications, sensor nodes can also be deployed to monitor patients and assist disabled patients. Other commercial applications include managing inventory, monitoring product quality, and monitoring disaster areas.
Because the size of a sensor node may be smaller than even a cubic centimeter, the wireless sensor node can only be equipped with a limited power source. Furthermore, in some application scenarios, such as seismic monitoring, battlefield surveillance, or radiation level control in a nuclear plant, replenishment of power might be impossible. Sensor node lifetime shows a strong dependence on battery lifetime. Therefore power consumption plays an important role in sensor networks.

In multihop communication way, each node plays the dual role of data originator and data router. The malfunctioning of a few nodes can cause significant topological changes and might require rerouting of packets and reorganization of the network. Hence, energy conservation and power management must be considered to prolong the lifetime of sensor networks [4, 5].

Network lifetime can be defined as the time elapsed from the network operation starts until the first node (or the last node) in the network depletes its energy (dies). Energy consumption in a node can be due to either "useful" or "wasteful" operations. The useful energy consumption includes transmitting or receiving data messages, and processing query requests. On the other hand, the wasteful consumption can be due to overhearing, retransmitting because of harsh environment, dealing with the redundant broadcast overhead messages, as well as idle listening to the media. An energy efficient approach should decrease wasteful energy consumption and utilize useful energy as sufficiently as possible.

1.2 Research Objective

Network clustering is an effective approach to conserve the limited energy resources of sensors [6, 7]. Sensor nodes in the network are divided into clusters, and sensors in a cluster communicate information only to their cluster head, which is a particularly selected node in a cluster. The cluster head communicates the aggregated information to the BS. The clustering technique avoids many connections between sensors and the BS so that energy consumption is decreased. However, in conventional clustering approaches, a head node is in charge of data collecting, aggregating, and aggregated data reporting. Because of concentrated energy consumption on a head node, it is much easier for a head node to run out of energy than the other nodes.

In this paper, we propose a new energy efficient clustering algorithm wherein data gathering and aggregating are performed by one node, called the collector, while the reporting to the BS is performed by another node, called the forwarder. Forwarders are determined according to the residual energy of nodes and transmission distance to the BS. In this way, the energy consumption is well balanced over the sensor network. The simulation results show that the proposed algorithm achieves longer network lifetime than other existing algorithms.

1.3 Organization of the Thesis

The paper is organized as follows. In Chapter 2, we review the related research work. In Chapter 3, we introduce the network model used in the paper, and present the weakness of the traditional algorithms. In Chapter 4, multi-hop data transmission from a sensor to the
BS is discussed, and details of our proposed algorithm based on single-hop transmission are also described there. We evaluate the performance of our approach by simulation in Section 5. In Chapter 6, we summarize the concluding remarks of the paper.
Chapter 2

Related Works

Sensor networks have many different application areas, and data gathering application plays an important role in sensor networks. Sensor nodes watch the surroundings and send the information to the BS periodically. According to the unique features of sensor networks, special multihop wireless routing protocols between sensor nodes and the BS are needed. The vital principle of routing protocols is power-aware. In these protocols, energy-efficient routes can be found based on the available power of the nodes and the energy required for transmission along the routes in order to make the energy consumption is low over the whole network. Besides, if a few nodes die quickly, there may cause vacancy area that can not be watched. Hence, energy consumption balanced over the network should also be considered.

There have been various routing protocols proposed for sensor networks. MECN [8] computes an energy-efficient subnetwork, when a communication network is given. A family of adaptive protocols, called Sensor Protocols for Information via Negotiation (SPIN) [9], is designed to address the deficiencies of traditional protocols by negotiation and resource adaptation. Besides the protocols mentioned above, another method of wireless communication is based on clustering. Clustering enables bandwidth reuse and increase system capacity. Using clustering enables better resource allocation and helps improve power control. In sensor networks, data fusion helps to reduce the amount of data transmitted between sensor nodes and the BS. Data fusion combines one or more data packets from different nodes to produce a single packet. Since the BS is far away, the energy consumption for a node to transmit data to the BS directly is high so that the node will die quickly. If all the nodes use direct communication for transmitting, the network lifetime will be quite short. Clustering communication protocols can minimize the amount of data that must be transmitted to the BS to conserve energy. Based on the data fusion technique, many clustering algorithms are proposed for wireless sensor networks.

The LEACH protocol, presented in [10], is the first clustering protocol, which minimizes energy dissipation in sensor networks. In LEACH, a small number of clusters are formed in a self-organized manner. A particular node, called head, plays an importance role in the routing protocol. A head node collects data from the other nodes in its cluster, aggregates the received data into a single data packet using data fusion technique, and transmits the commpressive data directly to the distant BS. The network is re-clustered automatically
periodically, and each node may become a cluster head according to a certain probability which is assigned in advance. In this way, the high energy dissipation in communicating with the base station is spread to all sensor nodes in the sensor network. The energy consumption is balanced over the network. Non-head nodes choose the closest cluster to join to conserve energy. LEACH assumes that any two nodes can communicate with each other directly, and also that any node can communicate with the distant BS directly.

HEED [11] extends LEACH by using cluster range limits and cost information. In HEED, the initial probability for each sensor to become a cluster head is dependent on its residual energy. Then, sensors that are not covered by any cluster heads double their probability of becoming a cluster head. This procedure iterates until all the sensors are covered by at least one head. Finally, sensors join cluster heads that have the lowest cost within their range. HEED assures that a node with larger energy has more opportunity to be chosen as a cluster head, as well as cluster heads are well-distributed over the area.

VCA [12] is a voting-based clustering algorithm for maximizing the lifetime of quasi-stationary sensor networks. The approach lets sensor note for their neighbors to elect suitable cluster heads. VCA combines load balancing, energy and topology information together by using very simple voting mechanism. Therefore, this approach is energy-efficient, location-unaware, and fully distributed.

Other clustering communication algorithms, such as EECS that is proposed in [13] and a distance-based energy efficient clustering protocol, proposed in [14] can also be found in the literature.
Chapter 3

Problem Outline

Data gathering is a typical application in wireless sensor networks. Sensor nodes monitor the environment and periodically transmit the sensed data to the BS. The BS gathers the messages from the sensors to draw some conclusions about the conditions of the service area, and broadcasts control messages to sensors in the network. In this section, we introduce the sensor network model under consideration.

3.1 Network Model

The network model is basically similar to those in LEACH [10] and HEED [11]. The nodes are grouped into several clusters. The sensed data in a cluster are gathered at a particular node and then sent to the BS directly or via multiple intermediate nodes. Unlike existing algorithms, data gathering and aggregating in the proposed algorithm are performed by the same node, called the cluster collector (or simply collector), while reporting to the BS is performed by another node, called the cluster forwarder (or simply forwarder). The network model used in this paper can be summarized as follows:

1. Nodes are uniformly dispersed in the service area of the network.
2. The nodes are quasi-stationary in the sense that they do not change their locations once deployed.
3. The nodes are location-unaware.
4. All the nodes have similar characteristics.
5. The nodes are left unattended after deployment.
6. Data fusion technique is available.

Table 3.1 lists the parameters and their definitions used in this paper. We use a simple model for the radio hardware energy dissipation. The energy consumptions for sending and for receiving an $l$-bit message are denoted by $E_t(l)$ and $E_r(l)$ respectively, and they are
assumed to be equal and determined by the hardware factors, that is, $E_t(l) = E_r(l) = Y$. The energy consumption required to amplify the radio signal in order to send an $l$-bit message to a location with distance $d$ away is denoted by $E_a(l, d)$. It is well-known that the transmission power depends on the environment and the transmission distance. In this paper, two parameters $\varepsilon_1$ and $\varepsilon_2$ are used to indicate the amplification factors for different transmission distance. When the transmission distance $d$ is less than a given threshold $d_0$, that is, if $d < d_0$, $\varepsilon_1$ is used, otherwise $\varepsilon_2$ is used. Additionally, the energy consumption for data aggregation is denoted by $E_g$.

Furthermore, energy needed to transmit an $l$-bit message to a location with a distance $d$ away is given by:

$$E_t(l, d) = E_t(l) + E_a(l, d) = \begin{cases} l(Y + \varepsilon_1 d^2) & d < d_0, \\ l(Y + \varepsilon_2 d^4) & d \geq d_0. \end{cases}$$

(3.1)

The energy consumption for receiving an $l$-bit message is given as follows.

$$E_r(l) = lY.$$  

(3.2)

<table>
<thead>
<tr>
<th>Table 3.1: Meanings of the parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
</tr>
<tr>
<td>$E_t$</td>
</tr>
<tr>
<td>$E_r$</td>
</tr>
<tr>
<td>$d_0$</td>
</tr>
<tr>
<td>$E_a$</td>
</tr>
<tr>
<td>$E_g$</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>$v_i$</td>
</tr>
<tr>
<td>$E_i$</td>
</tr>
<tr>
<td>$e_i$</td>
</tr>
<tr>
<td>$R_c$</td>
</tr>
<tr>
<td>$R_l$</td>
</tr>
</tbody>
</table>
3.2 Problem Statement

A typical clustering algorithm consists of four components: network clustering, channel assignment, data collection, and data report. The network clustering component divides nodes in the network into clusters. The channel assignment component determines the channel used for data transmission in each cluster. The data collection component is used for gathering and aggregating data in a cluster. Finally, the data report component is used for sending the compressive data to the BS. Most of the existing algorithms focus on the first component. Specifically, they consider the way to choose cluster heads as well as achieve the clusters well-balanced over the network. The differences among the clustering algorithms (such as LEACH [10], HEED [11], and VCA [12]) are only in the cluster forming period. After the formation of clusters, all the algorithms execute the same following components. In this paper, we keep the cluster forming method nearly the same as the existing approaches, but develop a data reporting method to the BS to prolong the network lifetime.

Figure 3.1: The traditional clustering algorithm.

In most distributed clustering algorithms, a cluster head node does not only receive and aggregate data from member nodes, but also transmits the compressive data to the BS. Since the energy consumption is quite concentrated on the head node, it causes traditional cluster algorithms unsatisfied in some conditions. For example, as shown in Figure 3.1, four sensors have the same cluster radius. Both A and C have 1J energy, while B and D have 2J and 3J energy, respectively. Nodes A, C and D are within the communication range of B. However, A, C and D can not communicate with each other directly. In this case, node B is the most suitable to be the head, because data fusion can be performed by B and only one piece of message needs to be delivered to the distant BS. However, compared to node D, node B is not suitable for transmitting the compressive message, because D has more residual energy and is nearer to the BS.
3.3 Data Transmission

We can consider two distinct approaches for forwarders to transmit the collected data to the BS as shown in Figure 3.2. In the figure, the collectors and forwarders of clusters are denoted by the gray and the black nodes, respectively, while the other nodes are denoted by the white nodes. One of the communication methods is single-hop transmission, that is, each forwarder transmits the collected data directly to the BS. This approach is simple to implement but clusters far from the BS have to consume much energy. The other is multi-hop transmission, that is, a forwarder can transmit the collected data to the BS via multiple intermediate forwarders. In the next chapter, we will first discuss the multi-hop transmission by mathematic analysis, and then propose our approach based on single-hop transmission method.
Chapter 4

Proposed Approach

4.1 Analysis of Multi-hop Communication

We consider a sensor network with \( N \) nodes. We derive the necessary condition under which any sensor node can communicate with the BS by using multi-hop transmission. In HEED [11], Younis et al. considered the case in which the network is divided into \( N \) square cells each with a cell side length of \( R_c/\sqrt{2} \), where \( R_c \) denotes the transmission range of a cluster. They derived a theorem for multi-hop network connectivity on condition that the following lemma [11, 15] holds.

Lemma 1 Assume that \( n \) nodes are uniformly and independently dispersed at random in an area \( R = [0, L]^2 \). Also assume that the area is randomly divided into \( N \) square cells with length \( R_c/\sqrt{2} \). If \( R_c^2 n = aL^2 \ln L \), for some \( a > 0 \), then \( \lim_{n,N \to \infty} E[\eta(n, N)] = 1 \), where \( \eta(n, N) \) is a random variable that denotes the minimum number of nodes in a cell (i.e., each cell contains at least one node a.a.s., or the expected number of empty cells is zero a.a.s.).

From the above lemma, we can see that if the node density of the network area is large enough, there exists at least one node in any square cell with side length of \( R_c/\sqrt{2} \). Furthermore, we assume that collectors are well-distributed, and no collector is located in the coverage of another collector. Most clustering approaches such as HEED and VCA all meet the requirement that collector nodes are well-distributed. In the following, we provide the necessary conditions for multi-hop network connectivity.

Lemma 2 If the inter-cluster transmission range satisfies the condition \( R_t \geq 2.8 R_c \), a cluster collector can find at least one other collector within the transmission radius \( R_t \).

Proof. As described above, we suppose there is at least one node in any square cell with side length of \( R_c/\sqrt{2} \). Let us consider a situation shown in Figure 4.1 where there are two cluster collectors, \( A \) and \( B \). Let a square cell tangent to the cluster area centered at \( A \) with a side length of \( R_c/\sqrt{2} \) as shown in the figure. According to [16], there must exist a node in the square cell. Assume that node \( v \) belongs to the cluster centered at \( B \) and is located at the boundary of cluster \( B \). Let us consider the case where node \( v \) is located at the right bottom of the cell, which is the farthest point in the cell from collector \( A \). Then, the Euclidean distances among \( A, B, \) and \( v \) satisfy the following relation.

11
\[ \|A - B\| \leq \|A - v\| + \|v - B\| \]
\[ = \sqrt{(1 + \frac{1}{\sqrt{2}})^2 + \left(\frac{1}{2\sqrt{2}}\right)^2} \times R_c + R_c \]
\[ \approx 2.8R_c. \]

We see that the distance of 2.8Rc indicates the farthest distance between the neighboring cluster collectors, A and B. Therefore, if \( R_t \geq 2.8R_c \), there is no isolated cluster collector in the network. □

**Lemma 3** The network graph of the cluster collectors is connected if \( R_t \geq 2.8R_c \).

**Proof.** If we can prove that any two collectors are connected when the transmission radius between two cluster collectors \( R_t \) is just equal to 2.8Rc, then we can prove the lemma. Now, we choose two collectors arbitrarily, \( H_1 \) and \( H_2 \). If \( \|H_1 - H_2\| \leq 2.8R_c \), \( H_1 \) and \( H_2 \) are connected directly, since \( R_t = 2.8R_c \). Let us consider the case when \( \|H_1 - H_2\| > 2.8R_c \) as shown in Figure 4.2. We can draw a square area with side length of \( R_c/\sqrt{2} \), which is not tangent to but very close to the area centered at \( H_1 \). Therefore, there must exist a node \( v \) in the square area and it does not belong to \( H_1 \) or \( H_2 \). Assuming that node \( v \) belongs to the cluster centered at \( H_3 \), we have \( \|H_1 - H_3\| \leq 2.8R_c \). Therefore, \( H_1 \) is connected with \( H_3 \). Since no two collectors are within each other’s range, we have \( H_3 \) as shown in Figure 4.2 that indicates the farthest point from \( H_2 \). We see that \( \|H_2 - H_3\| \approx 2.7R_c \), when \( \|H_1 - H_2\| = 2.8R_c \). Therefore, we can find a collector \( H_3 \) for \( H_1 \) that is connected to \( H_3 \) and is nearer to \( H_2 \). By using the similarly way, we can find another collector that is connected to \( H_3 \) and is closer to \( H_2 \) if \( \|H_2 - H_3\| > 2.8R_c \). Since the distance between \( H_1 \) and \( H_2 \) is limited, we can finally find a collector that is connected with \( H_1 \), and directly connected to \( H_2 \). Therefore, \( H_1 \) and \( H_2 \) is connected. □

In order to realize the multi-hop transmission from a cluster to the BS, the network graph of the cluster forwarders should be connected. We define \( R'_t \) as the transmission
radius between two forwarders. In the following theorem, we give the condition under which the forwarders in the network are connected.

**Theorem 1** The network graph of the cluster forwarders is connected if the transmission range between forwarders satisfies the condition $R'_t \geq 4.8R_c$.

**Proof.** In Lemma 3 we see that all the collectors are connected if $R_t \geq 2.8R_c$. The distance between a collector and its corresponding forwarder is shorter than $R_c$, therefore, the forwarders are connected if

$$R'_t \geq R_t + 2R_c \geq 2.8R_c + 2R_c = 4.8R_c.$$ 

The theorem is proved. □

Let $m$ denote the shortest distance from the BS to the nearest point of the sensor service area. In the following, we derive the conditions under which sensors are able to reach to the BS.

**Theorem 2** When the maximum transmission power level of sensor nodes is stronger than $\max\{4.8R_c, m + 2.8R_c\}$, all the nodes can deliver the data to the BS using multi-hop communication.

**Proof.** In Figure 4.3, we see that there is at least one node in the semicircle area, or else a blank ($R_c/\sqrt{2} \times R_c/\sqrt{2}$) size of cell exists. The maximum distance from the BS to the nearest forwarder is given by:

$$\| BS - forwarder \| \leq m + \frac{\sqrt{10}}{4}R_c + 2R_c \approx m + 2.8R_c.$$  \hspace{1cm} (4.1)

That is, if the transmission range is longer than $m + 2.8R_c$, there is at least a forwarder $w$ which can communicate with the BS. Moreover, according to Theorem 1, if $R'_t \geq 4.8R_c$, 

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{figure4.2.png}
\caption{Connective graph of collectors.}
\end{figure}
Figure 4.3: The minimum transmission range to the BS.

any forwarder can reach forwarder $w$. Therefore, the transmission range must be longer than both $m + 2.8R_c$ and $4.8R_c$. \hfill \qed
4.2 Protocol Architecture

In this section, we describe the proposed approach in details. Our approach is extended from the traditional approaches, such as LEACH and HEED, which are based on single-hop communication. For comparison, in this paper, we only focus on the single-hop transmission way. That is, forwarders transmit data to the BS directly without using any intermediate nodes. As shown in Figure 4.4, Data collection is performed periodically, and each period is denoted by a round, each of which has the same time length. Each round begins with a clustering-period phase during which the clusters are organized. Following the clustering-period, sensor nodes begin to transmit data. The transmitting-period is divided into several slots, and a sensor node can send data once during one slot. The network is reclustered during a period of time to avoid draining the energy of any one sensor in the network. The time of transmitting-period should be much longer than the time of clustering-period to reduce overhead for transmissions.

Figure 4.4: Time line of the proposed approach.

In this paper, we describe our approach based on HEED, since HEED has all typical features of a clustering protocol as introduced previously. Also, the approach can be extended from other clustering protocols. The determination of the collectors in our approach is similar to the method to determine cluster heads in HEED. However, our approach use two kinds of nodes, collectors and forwarders to collaborate with data communication. Figure 4.5 and Figure 4.6, describe the procedures of the two approaches. In the following, we divided the whole approach into three parts for description: clusters formation, forwarders selection, and data communication.
4.2.1 Cluster Formation

This part is responsible for cluster construction. The method for selecting collectors is the same as selecting head nodes in HEED, which is introduced in [11]. Because this part is totally the same as HEED, we only describe it generally.

Each node executing HEED proceeds through three phases: (i) Initialization, (ii) Main Processing, and (iii) Finalization. In the "Initialization" phase, each node sets its probability of becoming a cluster collector to: $CH_{prob} = C_{prob} \times \frac{e_i}{e_{max}}$, where $e_i$ is the residual energy of node $i$, $e_{max}$ is a reference maximum energy (corresponding to a fully charged battery) which is typically identical for all nodes, and $C_{prob}$ is a fixed small probability (in this paper, $C_{prob}=0.05$) used to limit the initial number of nodes competing to become cluster collectors. It is proved that varying $C_{prob}$ does not have a significant effect on protocol performance in [11]. Meanwhile, each node computes its cost according to node degree.

In the "Main Processing" phase, a node considers itself "covered" if it can find a collector node within its cluster range. A collector has two status: tentative or final. A final collector is the real collector node after clustering-period ends, while a tentative collector can become a regular node at a later iteration if it finds a lower cost cluster collector. Every "uncovered" node elects to become a cluster collector with probability $CH_{prob}$. If a node elects to become
a collector, it broadcasts a message to its neighbours, where the selection status is set to *tentative collector*, if its $CH_{prob}$ is less than 1, or *final collector*, if its $CH_{prob}$ has reached 1. After a sensor determines itself whether to be a collector or not, it doubles its $CH_{prob}$ and executes this phase again. When $CH_{prob} = 1$, the node finishes iteration and goes to the next phase.

During the ”Finalization” phase, final collectors are determined, and a non-collector node joins the cluster, where the collector has the minimum cost. After that, the network is clustered, and in each cluster there is a collector node.

### 4.2.2 Forwarder Selection

It is assumed that the BS broadcasts a ”hello” message periodically with a certain given power level. Then, each node computes the approximate distance to the BS based on the strength of signal received. For example, node $i$ can determine the value of parameter $E_i$ shown in Table 3.1 using the distance from node $i$ to the BS.

The forwarder of each cluster is selected according to the report cost of nodes. The node report cost is determined by the current remaining energy and the distance to the BS. The remaining energy of node $i$ in a cluster is denoted by $e_i$. Furthermore, the energy required for transmission from node $i$ to the BS is denoted by $E_i$. We consider that the report cost is proportional to the energy consumption from node $i$ to the BS, $E_i$, and inversely proportional to the current remaining energy of node $i$, $e_i$. That is, the report cost of node $i$ in a cluster, denoted by $c_i$, is given by:

$$c_i = \frac{E_i}{e_i}. \quad (4.2)$$

Before a node joins a cluster, the report cost should be computed. For node $i$, when node $i$ decides which cluster to join, node $i$ mixes the joining message and $c_i$ information together and sends it to the collector. Compared with HEED, there is no much overhead increased. After all the member nodes join the cluster, the collector node selects the member node which has the minimum report cost to be the forwarder. Note that the collector can also be the forwarder itself in its cluster if its report cost is the minimum. Then, the collector sets up a TDMA \(^1\) schedule and transmits this schedule to the nodes in the cluster. This ensures that there are no collisions among data messages and also allows the radio components of each member node to be turned off at all time except during their transmit time. In this way, energy consumption is reduced. After the TDMA schedule is received by all nodes in the cluster, data transmission starts.

### 4.2.3 Data Communication

The duty assignment of data communication is controlled by the TDMA schedule. During the data communication period, an ordinary node which is not a collector or a forwarder

---

\(^1\)TDMA (time division multiple access) is a digital transmission technology that allows a number of transmitters to access a single radio-frequency (RF) channel without interference by allocating unique time slots to each transmitter within each channel.
sends data to its corresponding collector. A collector gathers the data from its members except the forwarder, aggregates the collected data into one data packet, and then transmits it to the forwarder. A forwarder receives the compressive data from the collector, aggregates the compressive data with its own data, and then transmits to the BS directly. Compared to Figure 3.1, the transmission condition is described in Figure 4.7 under the same assumption introduced previously.
4.2.4 Pseudo-code of the Proposed Algorithm

The proposed algorithm consists of four procedures: initialization, cluster construction, TDMA slot allotment, and data transmission. The pseudo-code of the proposed algorithm is shown in Figure 4.8. In the proposed algorithm, the variable $v_i.status$ indicates the status of node $i$, which may be a collector, a non-collector, or a forwarder (a forwarder is also a non-collector). The variable $v_i.cluster$ indicates the set of member nodes of a cluster centered at collector $v_i$. $v_i.collector$ and $v_i.forwarder$ denote the collector and the forwarder of the cluster which contains node $v_i$. The pseudo-code is based on HEED algorithm, but it can also be varied based on other clustering algorithms without big modification. We only need change the the first line in the cluster construction sub-procedure of the code.
• **Initialization**
  1. Receive "hello" message from BS
  2. Compute $E_i$

• **Cluster construction**
  1. Construct clusters using HEED algorithm
  2. If ($v_i\text{.status}=\text{collector}$)
  3. Wait for the joining messages from members
  4. If ($v_i\text{.status}=\text{non-collector}$)
  5. Join a cluster with $E_i$ information

• **TDMA allotment**
  1. If ($v_i\text{.status}=\text{collector}$)
  2. $v_i\text{.forwarder} \leftarrow \text{least\_cost}(v_i\text{.cluster})$
  3. Assign TDMA slots to all the members
  4. Else
  5. Receive TDMA information from $v_i\text{.collector}$

• **Data transmission**
  1. If ($v_i\text{.status}=\text{collector}$)
  2. If ($v_i\text{.forwarder}=v_i$)
  3. Receive all the data from members
  4. Aggregate and transmit to the BS
  5. Else
  6. Receive all the data from members
  7. Aggregate and transmit to $v_i\text{.forwarder}$
  8. If ($v_i\text{.status}=\text{non-collector}$)
  9. If ($v_i\text{.status}=\text{forwarder}$)
  10. Receive data from $v_i\text{.collector}$
  11. Aggregate and transmit to the BS
  12. Else
  13. Transmit to $v_i\text{.collector}$

Figure 4.8: Pseudo-code of the proposed algorithm.
Chapter 5

Experiments

In this chapter, we evaluate the performance of the proposed approach by simulations. Firstly, we define the evaluation standard of the performance. Secondly, we describe the experiment environment for simulations. Finally, we compare our proposed approach with the traditional approaches, and then investigate the performance of the proposed approach by varying simulation parameters.

5.1 Evaluation Measure

We take the network lifetime as the evaluation measure, because prolonging network lifetime is especially important for unattended networks. The network lifetime is defined by round, which is introduced in Figure 4.4. A round is the time interval, which begins from clustering-period starts until transmitting-period ends. The transmitting-period is individed into several slots, and sensor nodes transmit data once per slot. Because of special features of sensor networks, sensor nodes should monitor surroundings in real-time way, so even a small vacancy area that is not watched can not be tolerated. For this reason, we define the network lifetime as the time elapsed from the network operation starts until the first node dies. Since if a node dies, it may cause a vacancy area which can not be covered by any sensor nodes.

5.2 Experiment Environment

In our experiments, we consider a network topology where nodes are uniformly distributed in a square field between \((x = 0, y = 0)\) and \((x = 100, y = 100)\). The energy consumption for wireless transmission is dictated by power model introduced previously. For two sensor nodes, the required transmission power level from one node to the other is determined by the distance between them. Practically, other factors may also affect the received power, such as noise or physical obstacles. For simplicity, we assume the absence of these factors in our experiments. In these simulations, energy is consumed whenever a node transmits or receives or performs data fusion. We repeat the simulation 100 times for accuracy, and the results shown are the average of 100 experiments. Each experiments uses a different randomly
generated topology, where each node is assigned an initial energy 0.5J. The parameters used in the simulation are given in Table 5.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network grid</td>
<td>from (0,0) to (100,100)</td>
</tr>
<tr>
<td>Base station</td>
<td>At (50,200)</td>
</tr>
<tr>
<td>Threshold distance ($d_0$)</td>
<td>100m</td>
</tr>
<tr>
<td>Cluster radius</td>
<td>25m</td>
</tr>
<tr>
<td>$E_t$, $E_r$</td>
<td>50nJ/bit</td>
</tr>
<tr>
<td>$\varepsilon_1$</td>
<td>10pJ/bit/m$^4$</td>
</tr>
<tr>
<td>$\varepsilon_2$</td>
<td>0.0013pJ/bit/signal</td>
</tr>
<tr>
<td>$E_g$</td>
<td>5nJ/bit/signal</td>
</tr>
<tr>
<td>Round</td>
<td>1 clustering-period and 5 communication slots</td>
</tr>
<tr>
<td>Initial energy</td>
<td>0.5J/battery</td>
</tr>
</tbody>
</table>

Energy consumption can be separated from two parts: “useful consumption” and “wasteful consumption”. “Useful consumption” is used for transmitting, receiving and aggregating sensed data originated from sensor nodes. “Wasteful consumption” is used for transmitting and receiving broadcast overhead information. We consider two scenarios in Table 5.2 for simulation, each of which has a distinct ratio of broadcast packet size to the whole packet size, to observe the impact of the overhead data. The ratio of overhead information is about 5% in Scenario 1, and 25% in Scenario 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data packet (byte)</td>
<td>500</td>
<td>150</td>
</tr>
<tr>
<td>Broadcast packet( byte)</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>
5.3 Experimental Results

5.3.1 Comparisons with HEED and LEACH

In this experiment, we compare network lifetime of our approach with LEACH and HEED approach respectively. The approach extended from HEED had been proposed in Chapter 4. The approach extended from LEACH is nearly the same as the pseudo-code in Figure 4.8, except that we only change the the first line into ”Construct clusters using LEACH algorithm” in the cluster construction part of the Pseudo-code.

We consider 100 sensor nodes are in the network area. The network environment is the same as written in Table 5.1. Additionally, the percentage of cluster collectors assigned in LEACH is 10%. That is to say, there will be about ten clusters formed under this consideration of network condition.

Figure 5.1 and Figure 5.2 compare the proposed approach with LEACH and HEED in two scenarios respectively. The figures show the total number of nodes that remain alive over the simulation time. We can see that the proposed approach achieves longer network lifetime than the traditional approaches in both of two scenarios. This is because the approach reduces the burden of collector nodes so that the energy consumption is balanced over the network area. We conclude that our approach is an effective method to prolong the network lifetime.
5.3.2 Effects of Various Parameter Settings

In the experiments, we investigate the impacts of various parameters on the overall approach performance, and our approach is extended from HEED for comparison in these experiments. We make experiments by varying number of nodes, cluster radius, and position of the BS.

In Figure 5.3, the network environment is the same as Table 5.1. We examine the performance for cases where the number of nodes are from 50 to 550. The figure shows us that the proposed approach clearly improves network lifetime over the traditional one. Besides, node density of the network has an important impact on network lifetime. The network design should be considered a suitable node density, which is neither too large nor too small.

In Figure 5.4, we consider the network area contains 100 sensor nodes, and we do the research with various cluster transmission ranges from 15m to 65m. The other parameters are kept unchanged as written in Table 5.1. From Figure 5.4, we can see that cluster range has a strong impact on network lifetime. When cluster range is short, the number of clusters will be large and many direct communication connects from clusters to the BS will be created, which causes much energy to be consumed. On the other hand, when cluster range is long, the number of clusters is small and a collector will have many member nodes, therefore a collector node will consume much energy for receiving data, which causes that its energy is run out quickly. In order to optimize the energy consumption, we should choose a suitable cluster range that can not be too short or too long.

In Figure 5.5, there are 100 nodes in the network area. We keep the x-coordinate of the BS as 50, which is introduced in Table 5.1, and vary the y-coordinate of the BS. We change the y-coordinate of the BS from 150 to 300 to study the effect of the distance between the BS and the network. Figure 5.5 shows that the proposed approach prolongs network lifetime compared to HEED. Network lifetime severely deteriorates when using single-hop communication as the distance increases, which emphasizes the advantages of network clustering.

Moreover, the ratio of overhead information should be considered for the network design. From the simulation figures, it is evident that the ratio of overhead also has a remarkable impact. The optimal values of parameters for prolonging network lifetime are difficult to decide, since the network lifetime is influenced by node density, cluster range, quantity of overhead information and other factors.
Figure 5.1: Network lifetime using LEACH vs. proposed approach.
Figure 5.2: Network lifetime using HEED vs. proposed approach.
Figure 5.3: Network lifetime with various node numbers.
Figure 5.4: Network lifetime with various cluster ranges.
(a) Scenario 1.

(b) Scenario 2.

Figure 5.5: Network lifetime with various positions of the BS.
Chapter 6

Conclusions

In this paper, we proposed a new clustering approach for sensor networks wherein the data gathering and data aggregation are separated from the data report to the BS, and these tasks can be assigned to distinct nodes in a cluster. The proposed algorithm can be easily extended from the previous algorithms but provides much better performance. Simulation results show that our proposed algorithm prolongs the network lifetime compared with the traditional approaches.

In large scale sensor networks, multi-hop communication becomes important for energy conservation. In this paper, we showed the necessary conditions for multi-hop data transmission. We are now working on designing the multi-hop data transmission protocol to achieve better performance. Since it is generally difficult to have the optimal solution, we are trying to develop a heuristic algorithm for multi-hop data transmission.
Bibliography


Acknowledgements

I would like to express my gratitude to my advisor, Yongbing Zhang who gave me the possibility to complete this thesis. He gave me help, stimulating suggestions and encouragement to help me in all the time of research. Moreover, he were very kind in reading the manuscript carefully and giving many useful suggestions to improve the paper.

I would like to thank my fellow students in Lab. E311 for all their help, support, interest and valuable hints. I aslo would like to show my appreciation to everyone who has, in some way, contributed to this work.

Especially, I would like to give my special thanks to my parents who gave me confidence to complete this work.