RESEARCH   **Open Access**

Optimizing the operating time of wireless sensor network

Thanh Tung Nguyen[1*] and Van Duc Nguyen[2]

Abstract

A difficult constraint in the design of wireless sensor networks (WSNs) is the limited energy resource of the batteries of the sensors. This limited resource restricts the operating time that WSNs can function in their applications. Routing protocols play a major part in the energy efficiency of WSNs because data communication dissipates most of the energy resource of the networks. There are many energy-efficient cluster-based routing protocols to deliver data from sensors to a base station. All of these cluster-based algorithms are heuristic. The significant benefit of heuristic algorithms is that they are usually very simple and can be utilized for the optimization of large sensor networks. However, heuristic algorithms do not guarantee optimal solutions. This article presents an analytical model to achieve the optimal solutions for the cluster-based routing protocols in WSNs.

Keywords: Sensor networks, Routing, Cluster networks, Battery, Linear programming, Optimization.

Introduction

There is a common problem in energy efficiency considerations in wireless sensor networks (WSNs): maximizing the amount of data sent from all sensor nodes to the base station (BS) until the first sensor node is out of battery. In sensor networks, sensors send data to each BS periodically during each fixed amount of time. Thus, the problem is the same as maximizing network operation lifetime until the first sensor node run out of battery. Numerous studies have been done on the energy efficiency using cluster-based routing in WSNs [1-5]. Cluster-based routing was originally used to solve the scalability problems and resources-efficient communication problems in wire-line and wireless networks [6,7]. The method can also be used to perform energy-efficient routing in WSNs. In the cluster-based routing, nodes cooperate to send sensing data to a BS. In this routing, a network is organized into clusters and nodes play different roles in the network. A node with higher remaining energy can be elected as the cluster head (CH) of each cluster. This node is responsible to receive data from its members in the cluster and to send the data to the BS.

However, all of the above-mentioned cluster-based routing work is heuristic. The real benefit of heuristic algorithms is that they are usually very simple and can be used for the optimization of large sensor networks. However, in general, heuristic algorithms do not guarantee optimal solutions.

In this article, an analytical model is used to obtain the optimal solutions for the above clustering lifetime problem. The basic idea is to formulate the problem as an integer linear programming (ILP) problem and to utilize ILP solvers [8] to compute the optimal solutions. These solutions are employed to evaluate the performance of previous heuristic algorithms. These analytical models are used to formulate the system lifetime problem into a simpler problem, find the optimum solution for the system lifetime problem, and evaluate the performance of heuristic models.

This article is organized as follow. The following section summarizes previous work in energy efficiency using cluster-based routing. Then, an analytical model of the cluster-based routing is developed. The model is first implemented by an analysis of a simple network with one cluster. After that, the analysis is extended for more complex cases of multiple clusters. A new heuristic cluster-based routing is also proposed. Finally, the simulation results of the analytical model, old heuristic solutions, and the new ones are presented and discussed.

* Correspondence: tungnt@isvnu.vn
[1]International School, Vietnam National University, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam
Full list of author information is available at the end of the article

Previous work in energy efficiency using cluster-based routing

In a cluster-based routing, higher remaining energy nodes can gather data from low ones, perform data aggregation, and send the data to a BS. Nodes in networks are grouped into clusters, and nodes that have higher remaining energy are elected as the CHs. In each cluster, the nominated CH node receives and aggregates data from all sensor nodes in the cluster. Usually, the sizes of the data of all sensors are the same and the aggregated data at the CH node has the same size with the data of every sensor in the cluster. As the data are aggregated in the CH node before reaching a BS, this technique reduces the amount of information sent to the distant BS, hence saves energy. For example, if each sensor in the cluster sends a message of 100 bits to the CH node, then the CH node sends the aggregated message of 100 bits to the BS. Details are given in [2,6,9]. As shown in Figure 1, all nodes in Cluster 1 send data to the CH. The node aggregates the data with its own data and sends the final data to the BS.

In sensor applications, every sensor node sends data periodically to its BS. Initially, every node starts with the initialized battery storage. A round of data transmission is defined as the duration of time to send a unit of data to the BS. At the end of each round, every sensor node loses an amount of energy which is used to send a unit of data to the BS. The lifetime of sensor networks is defined as the total number of rounds sending data to the BS until the first node is off.

Heinzelman et al. [1,2] proposed a Low-Energy Adaptive Clustering Hierarchy (LEACH). In LEACH, the operation of the protocol is divided into rounds. Each round consists of the setup and the transmission phase. In the setup phase, the network is divided into clusters and nodes negotiate to nominate CHs for the round. In more details, during the setup phase, a predetermined fraction of nodes, p, elect themselves as CHs as follows. A node picks a random number, r, between 0 and 1.

If ($r<T(n)$) then

The node becomes a CH for the current round
else
The node remains a non-CH node
where T is a threshold value given by:

$$T(n) = p : n \in G, \tag{1}$$

where G is the set of nodes that are involved in the CH election. The selected CHs for the round advertise themselves as the round's new CHs to the rest of the nodes in the network. All the non-CH nodes decide on the cluster to which they want to belong to. The decision is based on the distance to the closest CH.

In the transmission phase of LEACH, the elected CH collects all the data from nodes in its cluster, aggregates these data, and forwards them to a BS. In the next rounds, the process is repeated and CH positions are reallocated among all nodes in the network to extend the network lifetime.

For examples, as can be seen from Figure 2, the role of CH for Zone 1 is moved from Node 2 to Node 1 and the role of CH for Zone 2 is moved from Node 4 to Node 3 in the next round of data transmission. Therefore, the energy dissipation of these nodes during the network operation is balanced.

The LEACH protocol ensures that every node can become a CH exactly once within $1/p$ rounds. This will not give the optimum network lifetime, as sensor nodes that are far away from the BS will consume more energy than closer nodes to send data to the BS. Therefore, nodes, which are close to BS, need to become CHs more frequently than other nodes.

There are some LEACH variants to address the above issues in LEACH protocol [3,10-13]. Saha Misra et al. [3] proposed the energy enhanced-efficient adaptive clustering protocol for distributed sensor networks. CHs can be formed based on the residual energy of each node. The residual energy is calculated for every node after each round of transmission. Every node transmits a code containing the information about its residual
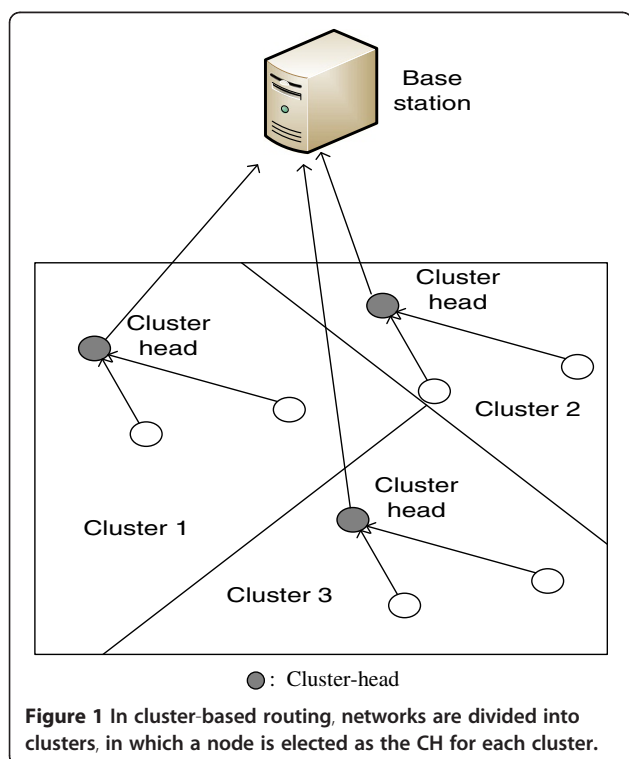
Figure 1 In cluster-based routing, networks are divided into clusters, in which a node is elected as the CH for each cluster.

Figure 2 CHs are reallocated in different rounds of transmission.
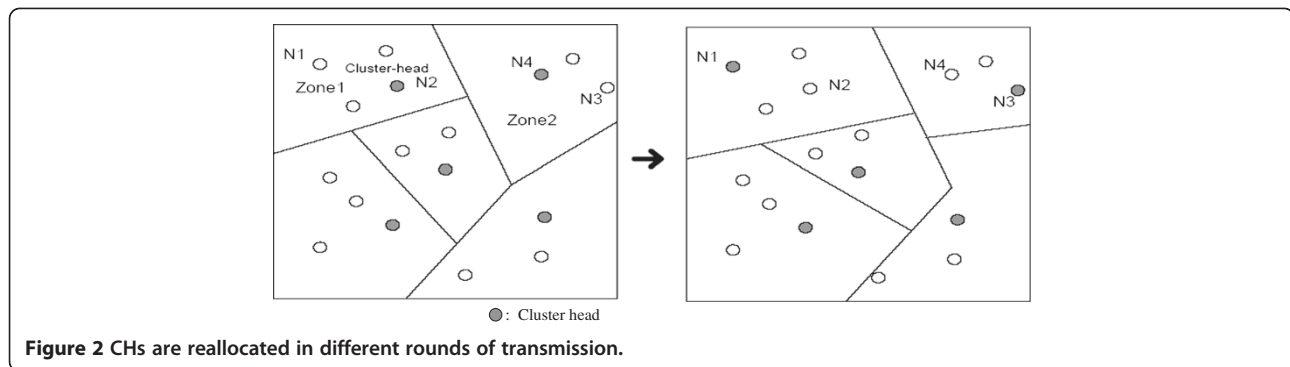
energy and its identification. If this residual energy is more than the ones of all other nodes in the same sub-area, then the node is the CH for that round in this sub-area. Otherwise, it can detect the node that has the maximum residual energy and elects this node as the CH.

A different approach was used by the authors of [4,5] who add the current energy information of sensor nodes into Equation (1).

$$T(n) = \frac{p}{1 - p(r \bmod (1/p))} \times \frac{E_{current}}{E_{initial}}, n \in G, \qquad (2)$$

where $E_{current}$ is the current energy of Node n and $E_{initial}$ is the initial energy of the node.

If $(r < T(n))$ then
The node becomes a CH for the current round
else
The node remains a non-CH node

Simulation results showed that the lifetime of the network with the scheme is improved 30% compared with the LEACH algorithm under the same experiments for LEACH.

After the design of LEACH protocol, these authors further proposed a new centralized version called LEACH_C in [2]. Unlike LEACH, LEACH_C utilizes the BS for creating clusters. During the setup phase, the BS receives the information about the location and the energy level of each node in the network. Using this information, the BS decides the number of CHs and configures the network into clusters. To accomplish this, the BS computes the average energy of nodes in the network, and nodes that have energy storage below this average cannot become CHs for the next round. From the remaining CH nodes, the BS uses the simulated annealing (SA) algorithm to find the k optimal CHs. The selection problem is an NP-hard problem [14,15]. The solution attempts to minimize the total energy required for non-CH nodes in sending data to the corresponding CHs. As soon as the CHs are found, the BS broadcasts a message that contains a list of CHs

for all sensors. If a node CH's ID matches its own ID, the node becomes a CH. Otherwise, the node determines its TDMA slot for its data transmission from the broadcast message and turns off its radio until the transmission phase. The transmission phase of LEACH_C is identical to that of LEACH. Under the same experimental settings, LEACH_C improves LEACH from 30 to 40%.

Besides cluster-based routings [10-13], there is also a chain-based one. Lindsey and Raghavendra [16] proposed one type of chain-based protocol called power-efficient gathering in sensor information systems (PEGASIS), which is near optimal for gathering data in sensor networks. PEGASIS forms a chain among sensor nodes so that each node will receive data from a near neighboring node and transmit data to another near neighbor. Gathered data move from a sensor node to the nearest neighbor, are aggregated with the neighbor's data, and eventually reach a determined CH before finally being transmitted to the BS. Figure 3 illustrates the ideas of the PEGASIS protocol. In this round of data transmission, Node 3 is elected as the CH. Node 5 transmits data to Node 4, and Node 4 fuses the data with its own data and transmits the fused data to Node 3. Similarly, Node 1 transmits data to Node 2, and Node 2 transmits the fused data to Node 3. Finally, Node 3 fuses the data of the other nodes with its own data and transmits the final fused data to the BS. The data fusion function can be any function, e.g., minima, maxima, and average, depending on specific applications. Nodes take turns equally to be the CH so that the energy spent by each node is balanced. In other words, each node becomes a CH once for every n rounds of data transmission, where n is the number of sensor nodes.

The comparison between the chain-based routings and cluster-based routings were done extensively in [9] and this is not mentioned here as this article only focuses on cluster-based routing.

In the next section, an analytical model is presented to achieve the optimal solutions for the frequency of CHs of sensor nodes. The basic idea is to formulate the problem as an ILP problem and to utilize ILP solvers [8] to
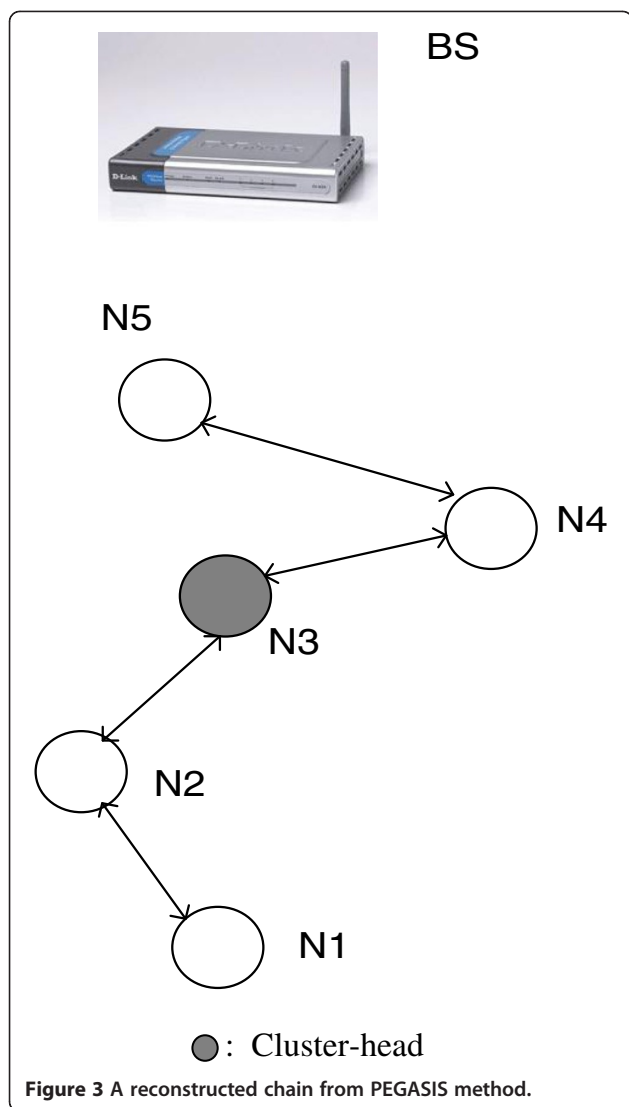
Figure 3 A reconstructed chain from PEGASIS method.

compute the optimal solutions. These solutions are employed to evaluate the performance of previous heuristic algorithms.

Analytical model for optimizing the lifetime of sensor network with one CH

In order to minimize the complexities of the clustering problem, the wireless radio energy dissipation model is not used. This assumption does not change the validation of any simulation result. A very simple energy usage model is given as

$$E(S) = \alpha d^2, E(D) = 0, \text{ for } \alpha > 0$$

where S denotes a source node, D denotes a destination node, $E(S)$ is the energy usage of node S, and d is the distance from S to D. This formula states that the energy required to transmit a unit of data is proportional to the square of the distance to a destination, and there is no energy spent at the destination. In this section, α is set to 1.

Let us analyze a very simple network to establish a general method that can be applied for any complicated problem. Figure 4 shows a simple network topology in which there are five nodes that lie on a line. The nodes are located equally from position 0 to position 80 m and the BS is located on the position 175 m. In sensor applications, every sensor node sends data periodically to the BS. A round of data transmission is defined as the duration of time to send a unit of data to the BS. Therefore, the lifetime of sensor networks is defined as the total number of rounds of sending data to the BS until the first node is off. It is assumed that every node starts with the equal initial battery storage of 500,000 units. The problem is maximizing the total the number of rounds of sending data to the BS until the first sensor node runs out of battery.

In each round of operation, every node must transmit a unit of data to the BS. It is also assumed that only one node acts as the CH in each round of transmission and the role is reallocated among all nodes so the system lifetime is maximized. The analytical model needs to compute the optimal usage of nodes as CHs under the battery constraint of every sensor.

Let us denote x_j, $\forall j \in [1\ldots5]$ to be the number of rounds, which Node j becomes a CH and c_j^i be the energy consumption of Node i, to deliver a unit of data in each round, when Node j becomes a CH, $\forall i, j \in [1\ldots5]$. As there are five nodes and only one CH, there are five possible choices for the CH in each round and there are also five energy usages for these five sensor nodes, respectively. This is shown in Table 1. For example, the energy dissipation of Node 1 when Node 5 becomes a

Figure 4 A simple network topology of five nodes on a line.
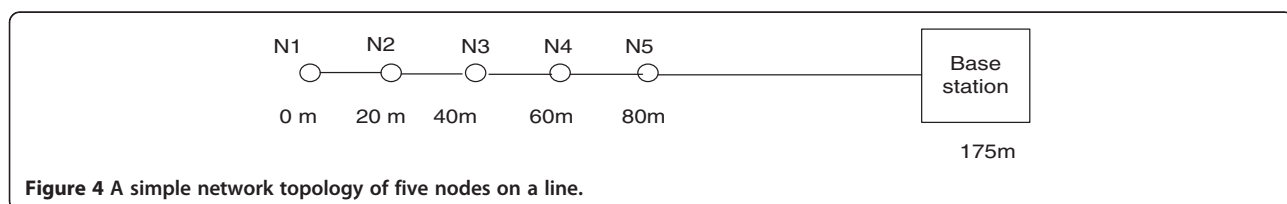
Table 1 The energy dissipated c_j^i (units) per round of node i when node j becomes a CH

	Node 1	Node 2	Node 3	Node 4	Node 5
CH1	**30625**	400	1600	3600	6400
CH2	400	**24025**	400	1600	3600
CH3	1600	400	**18225**	400	1600
CH4	3600	1600	400	**13225**	400
CH5	6400	3600	1600	400	**9025**

CH, c_5^1 is $(80 - 0)^2 = 6400$, the energy dissipation of Node 1 when Node 1 becomes a CH, c_1^1 is $(175 - 0)^2 = 30625$. The optimum number of transmission rounds (or system lifetime) for the network is written as the following ILP problem.

Maximize: $\sum_{j=1}^{5} x_j$

Subject to:

$$\sum_{j=1}^{5} c_j^i x_j \le E_i : \forall i \in [1 \ldots 5] \, x_j \in Z^+ : \forall j \in [1 \ldots 5] \qquad (3)$$

where E_i is the initial battery storage of node i. Formulation (3) states that the total number of rounds must satisfy the battery storage constraint of every sensor node.

Table 2 shows the optimum result obtained from (3) when the battery capacity increases from 125,000 to 50 million units. When the battery size is large enough (greater than 1 million units), the number of rounds that each node becomes a CH increases almost linearly with the battery capacity (e.g., the number of rounds of each node is nearly doubled when the battery capacity is increased from 1 to 2 million).

Simplification of formulation (3)

Formulation (3) can be converted to a linear programming (LP) formulation as given below:

Maximize: $\sum_{j=1}^{n} x_j$

Table 2 The number of rounds that each node i is a CH over the number of initial battery E (units) of each node

E	Node 1	Node 2	Node 3	Node 4	Node 5
125,000	0	2	5	8	11
250,000	0	5	11	17	13
500,000	1	11	22	34	44
1,000,000	3	23	44	68	88
2,000,000	7	46	89	135	176
50 millions	180	1155	2241	3391	4404

Subject to:

$$\sum_{j=1}^{n} c_j^i x_j \le E_i : \forall i \in [1 \ldots n] \, x_j \ge 0 : \forall j \in [1 \ldots n] \qquad (4)$$

where the condition of variables being integers is removed. There are two cases to use the formulation to obtain the optimization solutions:

(1) $E_i \to \infty$ then the solution of (4) becomes the solution of (3)
(2) $E_i \ne \infty$ then the solution of (4) is the approximation of the solution of (3)

Formulation (4) can remove the NP-hard characteristic of the ILP formulation (3). Therefore, the optimization solution can be solved by the simplex method [8,9]. In the next section, we will verify the solutions obtained from both formulations. A simple network topology of 11 nodes is given in Figure 5. All nodes are located equally on the line. The nodes are located equally from position 0 to position 100 m (separated each 10 m) and the BS is located on the position 175 m.

In the simulation, each node starts with an equal amount of initial energy of 500 million units. The lifetime problem for the network is first formulated as an ILP problem using (3). Then the LP formulation as in (4) is used to calculate the approximate solutions. Table 3 shows that the solutions given by both methods are almost identical. Therefore, the formulation of (4) can be an approximating solution of (3). Also, Nodes 10 and 11 never become a CH as they are too far from other nodes. Node 1 will never become a CH as it is too far from the BS.

Analytical model for optimizing the lifetime of sensor network with multiple CH

The previous section assumes a very simple case when there is only one CH. It is obvious that for the simple network of Figure 4, too many CHs will drain the energy of all sensor nodes very quickly as the nodes have to send data to the distant BS. This is not true for the other network topologies. The network considered in the analysis section has 20 nodes. The network topology is given in Figure 6. All nodes are located equally on the two lines.

For the network, one CH could not be enough, as other non-CH nodes would consume energy significantly to deliver a unit of data to the CH in each round. Table 4 shows the performance of the network with a variable number of clusters. The simulation result shows that two CHs will minimize the total energy consumption to send data to the BS.
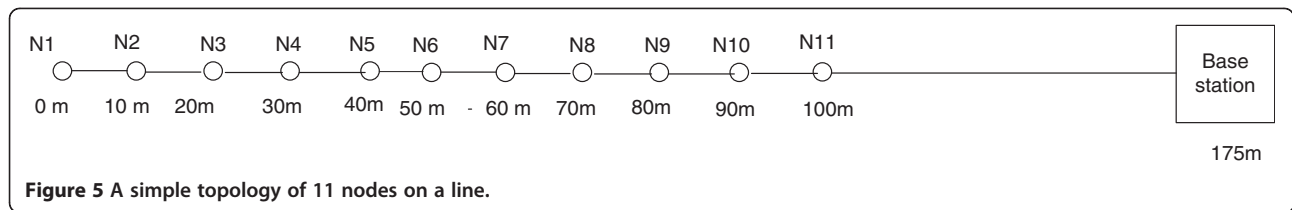
Figure 5 A simple topology of 11 nodes on a line.

When the number of CHs is more than one, it is much more complicated to obtain optimum solutions. The number of possible combinations of CHs is $O(n^k)$, where n is the number of sensor nodes and k is the number of CHs. Furthermore, with a selected solution of CHs, each sensor has k choices to select its CH. Therefore, the method of finding the optimum solution includes two optimization processes: optimization of the position of CHs and optimization of gathering traffic to the CHs.

In order to design an analytical model for complex cases with multiple CH in sensor networks, Theorem 1 is stated and proved.

Theorem 1: Consider two ILP problems with the same objective function and the same variables, if the set of coefficients of ILP problem 2 is smaller than the set of coefficients of ILP problem 1, respectively, for all of these coefficients, then the optimal solution of Problem 2 is higher than that of Problem 1.

Consider two ILP problems:

Problem 1:

Maximize: $\sum_{j=1}^{n} x_j$

Subject to:

$$\sum_{j=1}^{n} c_j^i x_j \leq E_i : \forall i \in [1 \ldots m] x_j \in Z^+ : \forall j \in [1 \ldots n] \quad (5)$$

Problem 2:

Maximize: $\sum_{j=1}^{n} x_j$

Subject to:

$$\sum_{j=1}^{n} c_j'^i x_j \leq E_i : \forall i \in [1 \ldots m] x_j \in Z^+ : \forall j \in [1 \ldots n] \quad (6)$$

Definition: O_1 is the optimal solution of Problem (5). O_2 is the optimal solution of Problem (6)

If $c_j'^i \leq c_j^i \forall i \in [1 \ldots m], \forall j \in [1 \ldots n]$, then $O_2 \geq O_1$

Proof: Since $c'^i_j \leq c_j^i \forall i \in [1 \ldots m], \forall j \in [1 \ldots n]$ and O_1 is the optimal solution of Problem 1, then O_1 is a feasible solution of Problem 2 because O_1 satisfy all constraints of (6). Since O_2 is the optimal solution of Problem 2, $O_2 \geq O_1$ ∎

To illustrate Theorem 1, let us consider two simple ILP problems:

Simple problem 1:

Maximize $x_1 + x_2$

Subject to:

$$\begin{matrix} 2x_1 + 3x_2 \leq 20 \\ 3x_1 + 4x_2 \leq 20 \end{matrix} x_1, x_2 \in Z^+ \quad (7)$$

Simple problem 2:

Maximize $x_1 + x_2$

Subject to:

$$\begin{matrix} x_1 + 2.5x_2 \leq 20 \\ 2.5x_1 + 3.5x_2 \leq 20 \end{matrix} x_1, x_2 \in Z^+ \quad (8)$$

Applying Theorem 1 for two simple problems (1) and (2), as the coefficients of the constraint functions (7) are all higher than those of (8) respectively, the optimal solution

Table 3 The number of rounds each node i becomes a CH solved by formulations (2) and (3)

Node i	Formulation (2)	Formulation (3)
1	0	0
2	569	569.6
3	1152	1152.3
4	1737	1737.5
5	2307	2307.2
6	2831	2831.2
7	3258	3258.7
8	3503	3503.3
9	1290	1289.1
10	0	0
11	0	0
Total	16647	16646

Figure 6 A simple network topology of 20 nodes on 2 lines, where each line has 10 nodes. The BS is at (50, 175).
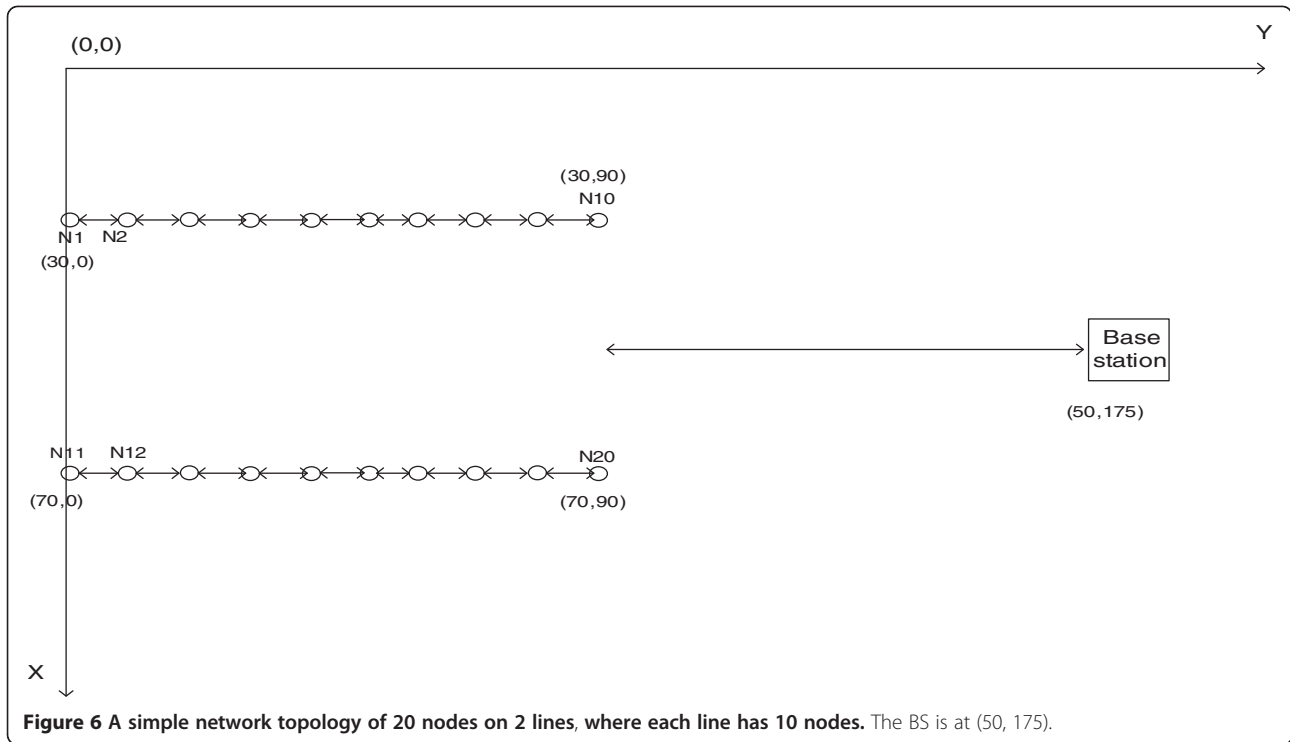
of (7) must be smaller than that of (8). This result is verified by using the ILP solver in [8]. The optimal solution of Simple problem (1) is 6 while the optimal solution of Simple problem (2) is 8.

This theorem is important because in many cases, this is very hard to calculate O_1. One of the reasons is that working out all coefficients c_j^i is impossible. Based on the theory, we know that O_2 can be an upper bound of O_1, or all the feasible solutions of Problem 1 are bounded by O_2.

Theorem 2: Given a clustering sensor network with k CHs, connection from non-CH nodes to the closest CH node of the k CHs provides the optimal lifetime for the clustering network.

In more detail, we are given a set of n sensors located in two-dimensional space R^2. Let us define S as the set of ways to select k CHs in the given set of n sensors. If every CH is different to the remaining $k - 1$ CHs, the number of elements in S is $\binom{n}{k}$. However, in the theorem, some CHs might be the same and these same CHs are considered as one CH. Therefore, the number of elements in S is n^k elements. Let us define $s_n^k(i)$ as the

ith element in S where i in $(1 \ldots n^k)$. Let us define c_i^j as the energy usage of Node j consumes, when the ith element in S is selected as the CHs. Let us define n_i as the number of rounds, which the ith element in S is selected as the CHs. Let us define E_j as the initial energy of Node j and O as the optimal solution of the following ILP problem:

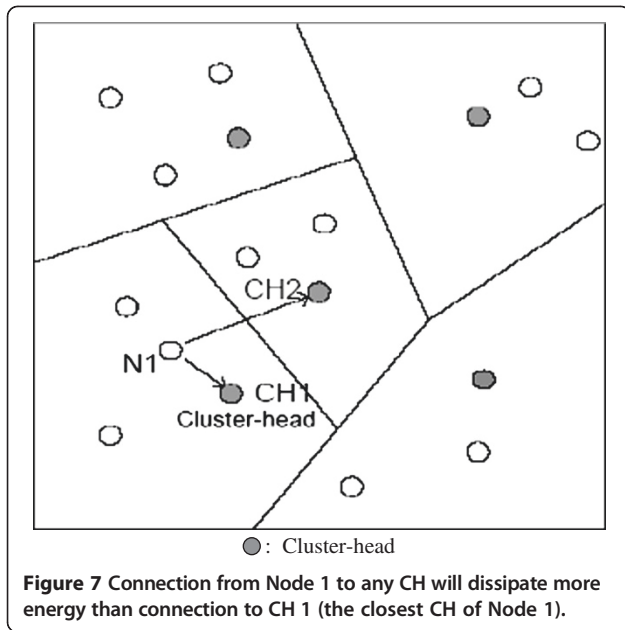Maximize:

$$\sum_{i=1}^{n^k} n_i \tag{9}$$

Subject to:

$$\sum_{i=1}^{n^k} n_i c_i^j \le E_j : \forall j \in [1 \ldots n] n_i \in Z^+ : \forall i \in [1 \ldots n^k]$$

The energy c_i^j is equal to the energy dissipation of Node j to send a unit of data to the closest sensor node in the ith element in S. Then, O is the optimal lifetime for the sensor network with k CHs.

Proof: Let us denote c'^j_i as the energy usage in any arbitrary way to send a unit of data from sensor node j to the ith element in S, $\forall i \in S$, $\forall j \in [1 \ldots n]$. The optimum selection of CHs of S is found by solving the mixed integer programming (MIP) problem below:

Table 4 The average energy dissipated (units) per round over the number of CHs

	1 CH	2 CHs	3 CHs
Energy per round (units)	65933	62016	69560

Figure 7 Connection from Node 1 to any CH will dissipate more energy than connection to CH 1 (the closest CH of Node 1).

Maximize:

$$\sum_{i=1}^{n^k} n_i \qquad (10)$$

Subject to:

$$\sum_{i=1}^{n^k} n_i c'^j_i \leq E_j : \forall j \in [1 \ldots n] \ \ n_i \in Z^+ : \forall i \in \left[1 \ldots n^k\right]$$

As $c'^j_i \geq c^j_i \forall i \in S, \forall j \in [1\ldots n]$, since c^j_i is equal to the energy dissipation of Node j to send a unit of data to the closest sensor node in the ith element in S, any optimum solution O' of (10) is smaller than the optimum solution O obtained by (9) as Theorem 1. This statement is illustrated in Figure 7. As the result, O is the global optimum solution for maximizing the operation time with k CHs. ∎

Calculation of coefficients for Problem (9)
The energy coefficients c^j_i of formulation (9) for a network of n nodes with k CHs can be calculated as follows:

For every combination of k CHs from the n nodes
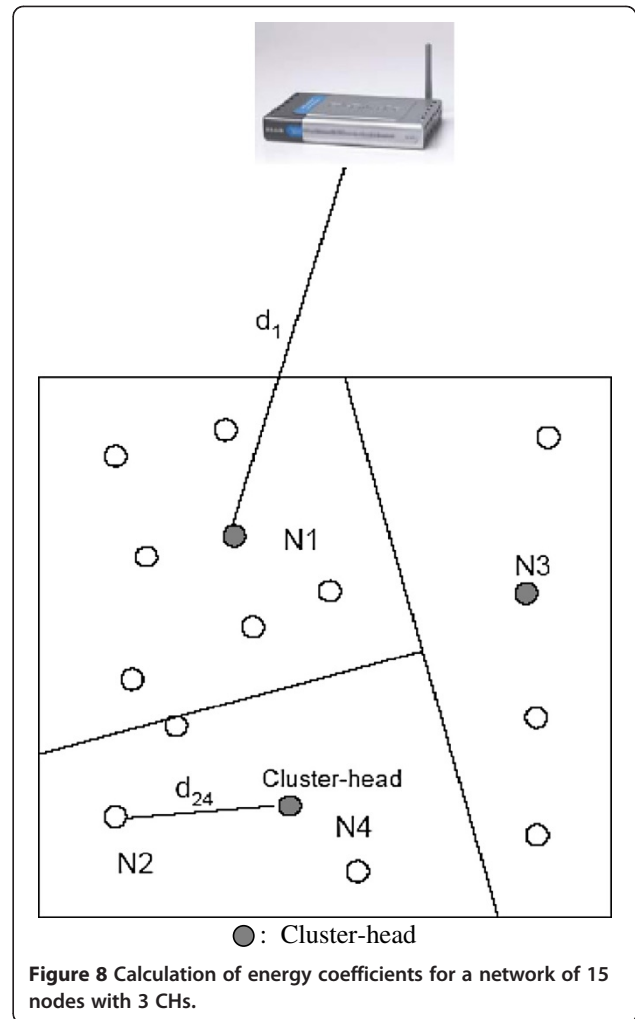   **For** every node from the n nodes
        If (the node is a CH) **then**

$$c^j_i = d^2_{toBS}$$

        else

$$c^j_i = d^2_{toCH}$$

End of code

Figure 8 Calculation of energy coefficients for a network of 15 nodes with 3 CHs.

where d_{toCH} is the distance from the sensor node to the closest CH from the k CHs, d_{toBS} is the distance from the sensor node to the BS.

Figure 8 shows that for the current selection of $k = 3$ CHs and $n = 15$ nodes, the energy coefficient of Node 2 is equal to d^2_{24}, and the energy coefficient of Node 1 is equal to d^2_1.

Theorem 3: The problem formulation in (9) provides the optimum solution for maximizing the operation time for any clustering network with the number of CHs smaller than or equal to k.

Proof: As stated in Theorem 2, S is the set of ways to select k CHs in the given set of n sensors. In each

Table 5 The average energy dissipated (units) per round and the number of rounds over the number of CHs

	1 CH	2 CHs	3 CHs
Energy per round (units)	65933	62016	69560
Number of rounds	332	377	364

combination selection, some CHs might be identical and these identical CHs are considered as one CH. In this case, the number of CHs is less than k. Therefore, any network of less than k CHs is a special element in S, where some CHs are the same. ∎

It is of interest to know the optimum solution of the network topology in Figure 6. Every sensor node begins with 1 million units of energy and the above-mentioned simple energy model is used. Table 5 shows the optimum system lifetime versus the number of CHs. The results show that the network achieves the optimum solution at the number of two CHs.

It is also of interest to see the distribution of optimums CHs among the 20 sensor nodes in Figure 6. The distribution depends on the position of sensors. The energy model used is d^2 energy model (gamma = 2).

Figure 9 shows the five pairs that are chosen as CHs most frequently. The results show that the pair of nodes (7, 17) is the most preferred CHs. This is due to the fact that the nodes are not very far from the BS as well as the rest of other nodes. As such, they can become intermediate CHs to deliver data to the BS. The five pairs are selected as CHs for 56% of the total number of rounds.

The same experiments are carried out on the same network over the "power 4" (gamma = 4) model. The model is given below:

$$E(S) = \alpha d^4, E(D) = 0, \text{ for } \alpha > 0$$

where S denotes a source node, D denotes a destination node, $E(S)$ is the energy usage of node S, and d is the distance from S to D. This formula states that the energy required to transmit a unit of data is proportional to the

Figure 10 Percentage of the total number of rounds that each pair of nodes is a pair of CHs for d^4 energy model.
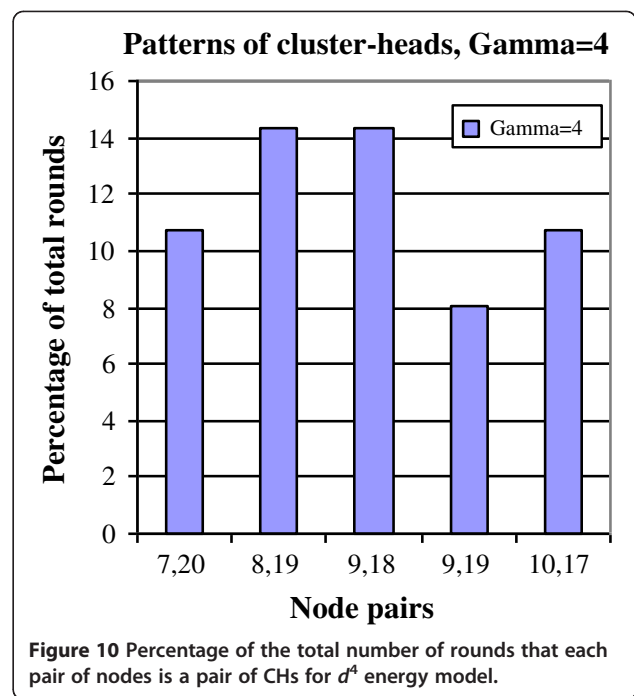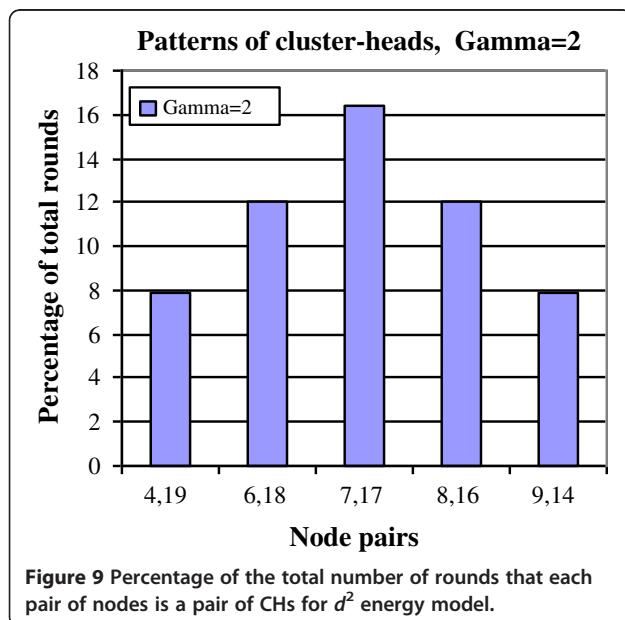
"power 4" of the distance to a destination, and there is no energy spent at the destination. For the rest of this section, α is set to 1.

Figure 10 shows the simulation results when α is set to 1. Compared to the previous results, the CHs move closer to the BS. This is because when the "power 4" model is used, the energy of CH nodes is drained quickly. As such, the nodes need to be closer to the BS. The five pairs are selected as CHs for 58% of the total number of rounds.

A simplified LEACH_C protocol (AVERA)

As mentioned in the Section "Previous work in energy efficiency using cluster-based routing", LEACH_C utilizes the BS for creating clusters. During the setup phase, the BS receives information about the location and the energy level of each node in the network. Using this information, the BS decides the number of CHs and configures the network into clusters. To do so, the BS computes the average energy of nodes in the network. Nodes that have energy storage below this average cannot become CHs for the next round. From the remaining possible CH nodes, the BS uses the SA algorithm to find the k optimal CHs. The selection problem is an NP-hard problem.

If the BS is also far away from main power sources and is energy-limited and processing-limited, it is impractical for the BS to run LEACH_C as it creates significant delay and requires significant computation. In this case, we modify LEACH_C algorithm by removing

Figure 9 Percentage of the total number of rounds that each pair of nodes is a pair of CHs for d^2 energy model.

the SA algorithm process. In more details, our algorithm AVERA is implemented as below.
   AVERA:

In every round, select k CHs randomly from m sensor nodes that have their energy level above the average energy of all nodes.

   Given:
   N: The number of sensor nodes indexed from 1 to N
   s: The current CH solution
   m: The number of sensor nodes that have energy above the average energy of all sensors
   For every round of data transmission
   s=k sensors in Random[1…m]
   Result: s is the CH solution for the round obtained from the AVERA algorithm. (End of code)

Simulation and comparison

Most of previous work on WSN lifetime [1-5] used the energy consumption model and the energy dissipation parameters given in [9]. The data are kept the same in our experiments to make the comparison between our proposed algorithms and previous ones feasible. The power transmission coefficients for free space and multi-path are given below.

$$\varepsilon_{FS} = 10pJ/b/m^2$$

$$\varepsilon_{MP} = .0013pJ/b/m^4$$

   From the parameters, the output power of a transmitter over a distance d is given by

$$P_{amp}(d) = \{\varepsilon_{FS}kd^2, d < d_o$$
$$P_{amp}(d) = \{\varepsilon_{MP}kd^4, d > d_o$$

where d_o is set to 82.6 m. The value of E_{elec} follows the experiments in [1,2,17-19] and is set to 50 nJ/bit.

   In summary, the total transmission energy of a message of k bits in sensor networks is calculated by

$$E_t = kE_{elec} + \varepsilon_{FS}kd^2, \text{ when } d < d_o$$
$$E_t = kE_{elec} + \varepsilon_{MP}kd^4, \text{ when } d > d_o$$

and the reception energy is calculated by

$$E_r = E_{elec}k$$

where E_{elec}, ε_{FS}, ε_{MP}, and d_o are given above.

   First, the optimum number of CHs of these networks is studied. In the experiments, 100 random 80-node sensor networks are generated. Each node begins with 1 J of energy. The network settings for the simulations are given

Network size $(100m \times 100m)$

Base station $(50m, 175m)$

Number of sensor nodes   80   nodes
Data message size: 4000 bits
Broadcast message: 200 bits
Energy message: 20 bits
Position of sensor nodes: Uniform placed in the area

Energy   model: $E_{elec} = 50*10^{-9}$ J, $\varepsilon_{fs} = 10*10^{-12}$ J/bit/m2 and

$\varepsilon_{mp} = 0.0013*10^{-12}$ J/bit/m4

Figure 11 Average energy dissipation per round (units) over the number of CHs.

below. The sensor positions and the BS position are defined as below . This is the same settings used in [1-5,9,18,19].
   Network size $(100m \times 100m)$
   Base station $(50m, 175m)$
   Number of sensor nodes 100 nodes
   Data message size: 4000 bits
   Broadcast message: 200 bits
   Energy message: 20 bits
   Position of sensor nodes: Uniform placed in the area
   Energy model: $E_{elec} = 50* 10^{-9}$ J, $\varepsilon_{fs} = 10* 10^{-12}$ J/bit/m^2 and $\varepsilon_{mp} = 0.0013* 10^{-12}$ J/bit/m^4

   During the sensor operation, every sensor node sends data periodically to the BS. A round of data transmission is defined as the duration of time to send a unit of data (4000 bits) to the BS. Each round consists of a setup and a transmission phase. In the setup phase, the network is divided into clusters and nodes negotiate to nominate CHs for the round. In the LEACH_C and AVERA protocols, each node sends its energy level message to the BS (20 bits). The BS decides the CHs for the round and sends a broadcast message (200 bits) about the decision for the round to all sensor networks.
   In the transmission phase, the elected CH collects all data from nodes in its cluster and forwards the data to a BS. After each round, every sensor node loses an
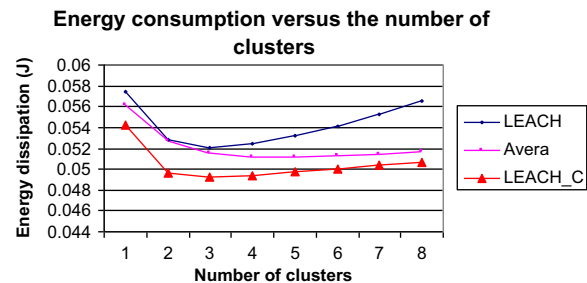
Figure 12 Ratio of the number of rounds between RE, LEACH, and the optimum solution.

amount of energy for the data transmission in the round. The amount depends on the distance from the sensor to its CH or to the BS. The lifetime of sensor networks is measured as the total number of rounds sending data to the BS until the first node is off.

LEACH, LEACH_C, and AVERA are used over 100 network topologies while varying the number of CHs from 1 to 8, and the system lifetime and the energy dissipation per round are recorded for these numbers of CHs.

Figure 11 shows that the energy dissipation per round is minimized for LEACH, LEACH_C, and AVERA at the number of CHs from 3 to 4. The result agrees well with the analytical model and the results are presented in [1,2,17].

Validation of the analytical model

In this section, the performance of LEACH, LEACH_C, and AVERA and the optimum solution from the analytical model is verified. The number of CHs is set to three in all methods. All methods are run over the above 100 random 80-node network topologies and the ratio between the lifetime of the three protocols and the optimum are recorded. For the calculation of the optimum solution, we use the GNU Linear Programming Kit (GLPK) and the MIP solver. GLPK is a free GNU LP software package for solving large-scale LP, MIP [8].

GLPK provides two methods to solve LP and MIP problems:

(1) Create a problem in C programming language that calls GLPK API routines
(2) Create a problem in a text editor and use a standalone LP/MIP solver to solve it.

We use method 2 to calculate the optimum solution. Figure 12 shows that both AVERA and LEACH_C perform very closely to the optimum solution while LEACH is only 70% of the optimum solution.

The computation time for all three protocols is also recorded on the 100 network topologies. The computational time for LEACH, AVERA, and LEACH_C are 1.6,2.5, and173.2 s, respectively. This shows that the new protocol AVERA provides a reasonably good operation time while guarantees less processing from the BS.

Conclusion

This article has presented some energy-efficient cluster-based routing protocols. In sensor networks, BSs only require a summary of the events occurring in their environment, rather than the sensor node's individual data. To exploit the function of the sensor networks, sensor nodes are grouped into small clusters so that CH nodes can collect the data of all nodes in their cluster and perform aggregation into a single message before sending the message to the BS. Since all sensor nodes are energy-limited, CH positions should be reallocated among all nodes in the network to extend the network lifetime. The determination of adaptive clusters is not an easy problem. We start by analyzing simple networks with one CH first to be able to obtain an effective solution for the problem. Then the model is extended to networks with multiple CHs.

Heuristic algorithms are also proposed to solve the problem. Simulation results show that LEACH solution performs quite far from the optimum solution as it does not directly work on the remaining energy of all sensor nodes. At the same time, both AVERA and LEACH_C solutions perform very closely to the optimum solution. Note that the computational time for AVERA is also 1.4% of LEACH_C.

Author details
[1]International School, Vietnam National University, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam. [2]Faculty of Electronics and Telecommunications, Hanoi University of Science and Technology, Dai Co Viet Str. 1, Hanoi, Vietnam.

References
1. WB Heinzelman, AP Chandrakasan, H Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *33rd Hawaii International Conference Systems Sciences*, 2000, p. 10. Vol.2. ISBN: 0-7695-0493-0
2. WB Heinzelman, AP Chandrakasan, H Balakrishnan, An application specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun **1**(4), 660–670 (2002)
3. I Saha Misra, S Dolui, A Das, Enhanced-efficient adaptive clustering protocol for distributed sensor networks, in *IEEE ICON*, 2005, p. 6. doi:10.1109/ICON.2005.1635525. Vol.1, ISBN:1-4244-0000-7
4. J Zhang, B Huang, L Tu, F Zhang, A cluster-based energy-efficient scheme for sensor networks, in *The Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'05)*, 2005, pp. 191–195. doi:10.1109/PDCAT.2005.3. ISBN: 0-7695-2405-2
5. R Chang, C Kuo, An energy efficient routing mechanism for wireless sensor networks, in *The 20th International Conference on Advanced Information Networking and Applications (AINA'06)*, 2006. Vol.2, p. 5. doi:10.1109/AINA.2006.86. ISSN: 1550-445X; ISBN: 0-7695-2466-4
6. JN Al-Karaki, AE Kamal, Routing techniques in wireless sensor networks: a survey. IEEE Wirel. Commun. **11**(6), 6–28 (2004)
7. J Deng, YS Han, W Heinzelman, Balanced-energy sleep scheduling scheme for high density cluster-based sensor networks. Comput. Commun. **28**, 1631–1642 (2005)
8. GLPK programming, 2011. http://www.gnu.org/software/glpk/
9. T Nguyen Thanh, *Energy-efficient routing algorithms in wireless sensor networks*. PhD thesis (Monash University, Australia, 2008)
10. L Guo, Y Xie, C Yang, Z Jing, Improvement on LEACH by combining adaptive cluster head election and two-hop transmission, in *9th International Conference on Machine Learning and Cybernetics*, 2010, pp. 1678–1683

11. T Qiang, W Bingwen, W Zhicheng, MS-LEACH: a routing protocol combining multihop transmission and single-hop transmission, in *Pacific-Asia Conference on Circuits, Communications and Systems*, 2009, pp. 107–110

12. M Farooq, AB Dogar, GA Shah, MR-LEACH: multi-hop routing with low energy adaptive clustering hierarchy, in *4th International Conference on Sensor Technologies and Applications*, 2010, pp. 262–268

13. J Chen, H Shen, MELEACH-L: more energy efficient LEACH for large scale WSNs, in *Wireless Communications, Networking and Mobile Computing*, 2008, pp. 1–4

14. SG Nash, A Sofer, *Linear and Nonlinear Programming* (McGraw-Hill, New York, 1996)

15. T Kanungo, DM Mount, NS Netanyahu, A local search approximation algorithm for k-means clustering, in *The Eighteenth Annual Symposium on Computational geometry*, 2002. Volume 28, Issues 2–3, pp. 89–112

16. S Lindsey, C Raghavendra, Power-efficient gathering in sensor information systems, in *IEEE Aerospace Conference*, 2002. Vol.3, p. 3-1125-3-1130. doi:10.1109/AERO.2002.1035242.0-7803-7231-X

17. WB Heinzelman, *Application-specific protocol architectures for wireless networks*. PhD dissertation (Massachusetts Institute of Technology, 2000)

18. T Nguyen Thanh, V Phan Cong, *The Energy-Aware Operational Time of Wireless Ad-hoc Sensor Networks. ACM/Springer Mobile Networks and Applications (MONET) Journal*, Volume 17, August, 2012. doi:10.1007/s11036-012-0403-1

19. T Nguyen Thanh, *Heuristic Energy-Efficient Routing Solutions to Extend the Lifetime of Wireless Ad-Hoc Sensor Networks* (Springer, LNCS 7197, 2012), pp. 487–497. ISBN: 978-3-642-28489-2