Optimal Cluster Head Selection in Wireless Sensor Networks using Integer Linear Programming techniques

Zahra Eskandari¹, Seyed Amin Hosseini Seno², Muhamed Shenify³, Rahmat Budiarto¹
¹ Department of Computer Engineering, Quchan University of Advanced Technology, Quchan, Iran
² Network Research Laboratory, Ferdowsi University of Mashhad, Mashhad, Iran
³ College of Computer Science and Information Technology, Al Baha University, Kingdom of Saudi Arabia

ABSTRACT

Wireless sensor network (WSN) consists of sensor nodes which are deployed in the environment densely and randomly. The main constraint of these nodes is limited energy resources; so, the operations which are performed in the network, must be energy efficient. For this reason, routing and data transmission in these networks perform hierarchically and in multi hop manner. One of these hierarchical architectures which have a considerable positive effect on energy consumption is clustering algorithm. But what is important is that the cluster heads election should be done efficiently. Recently some works have focused on optimal cluster head election using Integer Linear Programming techniques. In this paper, Integer Linear Programming techniques are used to formulate the clustering problem. At first, by using Integer Linear Programming techniques, a scalable and multi objective model for optimal cluster head selection is presented and then the distributed clustering algorithm is proposed. As shown in simulation results, the proposed clustering algorithm is more efficient in terms of energy vs. LEACH algorithm.

Copyright © 2014 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:
Seyed Amin Hosseini Seno,
Network Research Laboratory,
Ferdowsi University of Mashhad, Mashhad, Iran.
Email: Hosseini@um.ac.ir

1. INTRODUCTION

Wireless sensor network (WSN) consists of sensor nodes which are deployed in the environment densely and randomly [1]. These nodes sense environmental data such as light, temperature, jitter and so on, and transmit the sensed data to the sink periodically. The main constraint of these nodes is limited energy resources; the small size, low cost, and also inaccessibility of these nodes lead to this constraint. Thus to guarantee the network lifetime, the operations which are performed in the network, must be energy efficient. One of these operations is data transmission. As the power consumption of the radio transmission has a direct correlation with square or higher exponent of the transmission distance [2], so to have energy efficient transmission in wireless sensor networks, data transmission is often performed in multi hop manner that gives energy efficiency and at the same time increases network scalability [3]. In the multi hop manner, sensor nodes transmit sensed data to the intermediate nodes and then these nodes transmit data to the next nodes in the path toward the sink node. One of these hierarchical architectures is clustering algorithm [4].

In clustering algorithms, the network nodes are partitioned into clusters which cluster’s members are managed by a leader node called Cluster Head (CH) [5]. In fact clustering algorithms consist of two parts: cluster head election and cluster formation. The cluster head election part can be done in a centralized or distributed manner, by the sink or locally by the nodes, respectively. After the first part of the algorithm, cluster heads are determined and they broadcast their information in neighboring environment, then in cluster
formation part, each node join to a suitable cluster according to the algorithm parameters. The cluster’s members transmit sensed data to the cluster head and the cluster head aggregates and compresses the data and then transmits aggregated data to the sink in one hop or multi hop manner [6]. Considering the nature of multi hop transmission mechanism in clustering approaches, an accurate clustering mechanism may highly decrease energy consumption and consequently enhance the network lifetime [7].

Until now some clustering algorithms had been proposed. One of the most commonly used approaches is LEACH [8], the basis of many other clustering algorithms such as LEACH-C [9], LEACH-E [10], EECS [11], Energy-Kmeans [12], DECSA [13] and SECA [14]. In fact these algorithms presented new and more efficient structures based on the base clustering algorithm which improve former results. Some works perform efficient clustering using fuzzy tools such as in IFUC [15], EAUCF [16], FLCFP [17], ECPF [18], CHEF [19], CEFL [20], and so forth.

Recently some works [21] focus on optimal cluster head election using Integer Linear Programming techniques. In these approaches, the process of selecting optimal cluster heads is modeled as an optimization problem with emphasis on maximum network lifetime and ensuring that all the network nodes are covered. Then, the mentioned problem is solved and optimal or near optimal set of cluster heads is selected. In this context, the selection of the optimal cluster heads problem and clustering problem are used interchangeably.

In this paper we try to use Integer Linear Programming to model optimal cluster head selection problem. The important difference between the proposed model in this work and the previous work [21], are as follows: 1) the proposed model has high scalability, and 2) here the clustering problem is modeled as a multi objectives function by considering not only minimum energy consumption, but also maximum network lifetime. In fact in this work, by using most important parameters in wireless sensor networks, energy and transmission distance, the clustering problem is formulated as an optimization problem. The contribution of this paper consists of two phases: at first we model mentioned problem as a multi objective optimization problem which is solved by the sink as a centralized section of the algorithm. Then a distributed clustering algorithm based on the results of the previous phase is proposed.

The rest of the paper is organized as follows: In the next section we study some clustering algorithms especially under the title efficient or optimal clustering. In Section 3, mathematical concepts of Linear Programming and Integer Linear Programming are described. In Section 4, the contribution which consists of two phases, is presented; the Integer Linear programming models of the clustering problem and the proposed clustering algorithm. In Section 5 the proposed models and the clustering algorithm are evaluated, and followed by the presentation of the results. Finally Section 6 concludes the paper.

2. RELATED WORKS

Until now some clustering algorithms have been presented, but in recent years some works focused on optimal or efficient clustering. In this section, at first the basis LEACH algorithm and some of its successors and then optimal or efficient clustering algorithms which are based on Linear Programming techniques are reviewed. It must be mentioned that the reason of the emphasis on LEACH algorithm and its family is that because of the importance of LEACH algorithm, we compare the proposed clustering algorithm results with the LEACH algorithm in the next sections, so the description of these algorithms is given here briefly.

In LEACH, some nodes periodically are selected as cluster heads and after the election of cluster heads, in a distributed manner. Each node selects its cluster by choosing the nearest cluster head. In data transmission phase each cluster head collects data from its cluster’s members and sends aggregated data to the sink in one hop. Distributed manner of LEACH clustering algorithm has some advantages but in [9], LEACH-centralized (LEACH-C) was proposed. In LEACH-C the sink determines that the nodes which have energy less than the average energy of the network cannot be selected as cluster heads for the current round. This mechanism leads toward more effective energy consumption.

In [10] a clustering algorithm which called LEACH-E, was proposed to optimize the selection of cluster heads. In this work, similar to [9], the nodes with low energy are avoided to be selected as the cluster heads, and so they could balance the energy consumption of the nodes. Simulation results showed that their algorithm had better results compared to LEACH algorithm.

In [12] a modified K means clustering algorithm is proposed. In this work by knowing the location of each node, clustering is formed based on the highest residual energy and minimum distance from the sink. In this work, cluster heads are selected in centroid method. Indeed, the minimum distance between the cluster node’s and the centroid point is elected as a cluster head. In [13] a distance-energy cluster structure algorithm (DECSA) is presented which considers both the distance between the nodes and the sink and also the energy of nodes. The main claim of the authors is that by considering these parameters, they try to reduce energy consumption.
In [14], the authors propose a saving energy clustering algorithm to provide efficient energy consumption. The main idea of them is to reduce data transmission distance of nodes by using the uniform cluster. In order to make an ideal distribution of the clusters, they use the average distance between the nodes and the energy for cluster head selection. Simulation results indicate that their proposed algorithm has less energy consumption and higher network lifetime in comparison with other clustering algorithms.

In related to optimization problems, the earliest contribution [22] formulated a non Integer Linear Programming (non-ILP) algorithm clustering mechanism called Virtual Grid Architecture (VGA) and a corresponding ILP formulation. The objective was to find the minimum set of connected CHs. But due to the limitation of the solver, maximum number of nodes is restricted to 30 nodes. In [23], the authors proposed three different ILP formulations, which not only focused on the minimum number of CHs, indeed they specified number of CHs and then after formulating the clustering problem such that the maximum possible network lifetime was obtained, they select optimal CHs. However, due to the complexity of the ILP formulations, only up to 9 nodes could be solved in their modeling.

As the latest work that use these techniques, in [21] the authors proposed an ILP formulation of the clustering problem, building on the model presented in [23], which improves its weaknesses. Indeed, in [21] an improved ILP formulation for the Clustering Problem was presented which try to minimize energy consumption as objective function. In their model, the coverage of the network nodes meets a set of constraints in the model. It is notable that in this model, the number of the cluster heads should be predetermined in trial and error manner.

3. LINEAR PROGRAMMING AND INTEGER LINEAR PROGRAMMING

Linear Programming (LP) [24] deals with the problem of minimizing or maximizing a linear objective function subject to linear equality and linear inequality constraints. These constraints form a convex polyhedron region which called feasible region. Indeed this convex polyhedron is the intersection of some half spaces, each of which is defined by a linear inequality. The Linear Programming algorithm finds a point in the polyhedron where has minimum or maximum value provided that such a point exists. A minimization linear problem is in the form of:

$$\text{min}\{Cx \mid Ax \leq b, x \in \mathbb{R}^n\}$$

The set \( S = \{x \in \mathbb{R}^l, Ax \leq b\} \) sets up the feasible region and the element \( x \) is called a feasible solution. In addition to mentioned constraints above, it has some integrality restrictions on some or all of the variables [25, 26]. In general, the ILP is in the form of:

$$\text{min}\{Cx \mid Ax \leq b, x \in \mathbb{Z}^k \times \mathbb{R}^l\}$$

This problem is called the mixed Integer Linear Programming problem but when all decision variables must be integers, it is called a pure Integer program. The problem of this special form:

$$\text{min}\{Cx \mid Ax \leq b, x \in \mathbb{Z}^n, x = 0, 1\}$$

is called Binary Integer Linear Programming (BILP) which all variables must be binary. In this work, due to the the essence of the cluster head selection problem, this form of the optimization problem is used. In clustering problem, the selection of which nodes as the cluster heads is considered. So, the variables are the status of a node as a cluster heads which is represented as a Boolean variable with value 0 or 1. Since the variables are binary values, so the problem is a BILP problem.

To solve a MIP problem, there are different possible techniques. Although the Simplex method [24] is a simple and effective technique for solving linear programs, but there is no unique technique for solving Integer programming problems. Some other techniques such as branch-and-bound procedure, cutting-plane techniques and group-theoretic techniques are available to be applied to the integer programming problems. However, it must be mentioned that the performance of any technique is highly problem-dependent. In addition, several composite procedures have been proposed, which combine several of these techniques; some solvers such as CPLEX [27], SCIP [28] use these composite procedures as well.

But what is important is that the processing time of a Linear Programming problem depends on the size of the problem; the number of variables, especially the number of variables with restrictions; and the number of constraints, particularly the complexity and density of the constraints. Thus, the formulation of the problem is very important.
4. PROPOSED CLUSTERING ALGORITHM

In this section, how to model clustering problem as an ILP problem is described and then the clustering algorithm details are considered.

4.1. Modeling Clustering Problem

To modeling a problem as an ILP problem, the problem’s variables, constraints and the objective function should be determined. As mentioned above, in clustering problem, the variables are the status of a node as a cluster heads which is shown as a Boolean variable. Since the variables are binary values, so the problem is a BILP problem and the variable is defined as follows:

\[ x_i = \begin{cases} 
1 & \text{if node}_i \text{ is selected as CH} \\
0 & \text{else} 
\end{cases} \]

With regards to the constraints, the coverage of the nodes in the network is considered, it means that each node in the network must be placed in a cluster as a cluster member and be connected with a cluster head, provided that the distance between them should be less than efficient communication range. Otherwise the clustering may not be energy efficient.

To apply the constraints, the coverage matrix of the network is formed as follows: a matrix that the rows demonstrate the nodes in the network and the columns indicate the candidate cluster heads is considered. The value of each entry \( m_{ij} \) is determined in this way: if the node \( i \) is covered by candidate cluster head \( j \) in less than efficient communication range, the \( m_{ij} = 1 \), otherwise \( m_{ij} = 0 \).

\[ m_{ij} = \begin{cases} 
1 & \text{if node}_i \text{ is covered by CH}_j \\
0 & \text{else} 
\end{cases} \]

So, the constraints are defined as follows:

\[ \sum_{j=1}^{k} m_{ij} \ast x_j \geq 1 \quad \text{for all } i = 1..N \]

It must be mentioned that in this formula, \( k \) is the number of candidate cluster heads. These constraints mean that each node \( i \) must be covered by at least one cluster head. Since all these inequalities are linear, so they form a convex polyhedron which is a precondition to the feasibility of the optimization problems.

As regards to the objective function, to compare the performance of different modeling method, two models of clustering problem are presented here. In objective function which called MIN-CH-NUMS model, the objective function is defined in the following way:

\[ \min \sum_{i=1}^{k} x_i \]

Indeed we try to have a minimum number of cluster heads to avoid overlapping of the clusters and also having a redundant cluster head. So, the clustering model MIN-CH-NUMS is as follows:

\[ \min \sum_{i=1}^{k} x_i \]

subject to:
And in another objective function with the name MIN-DIST-CHS model, having some cluster heads with minimum distances to the network nodes is highlighted. By this way, the nodes are selected as cluster heads having minimum transmission distances which lead to efficient energy consumption in data transmissions.

\[
\min \sum_{i=1}^{k} \sum_{j=1}^{N} \text{distance}(\text{node}_j, \text{node}_i) \cdot x_i
\]

And the clustering model MIN-DIST-CHS is as follows:

\[
\min \sum_{i=1}^{k} \sum_{j=1}^{N} \text{distance}(\text{node}_j, \text{node}_i) \cdot x_i
\]

subject to:

\[
\sum_{j=1}^{k} m_{ij} \cdot x_j \geq 1 \quad \text{for all } i = 1..N
\]

### 4.2. Proposed Clustering Algorithm

The proposed algorithm is presented in Figure 1. The clustering algorithm consists of two sections: centralized section and distributed section. To do clustering in the network, as a cluster head’s energy reaches a predefined threshold, the information of nodes is transmitted to sink. In centralized section, sink calculates the average energy of the alive nodes (Eave) then forms cluster head candidate set (CHCs) by considering the nodes which have higher energy than Eave as a cluster head candidate. By performing such procedure, low energy nodes are prevented from being cluster head candidates. The procedure guarantees that the cluster heads which transmit data in the far distance, have high energy and low energy nodes can participate in network activity as cluster members and transmit data in the efficient transmission distance, and can be alive for a longer time. Thus, it provides network lifetime enhancement. Consequently, based on discussion in Section 3, the coverage matrix, \( M \) is formed as follows for the ILP models.

\[
M = \begin{pmatrix}
m_{1,1} & \cdots & m_{1,n} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,n}
\end{pmatrix}
\]

by using the ILP models, the sink sets up the constraints and objective function and then solves the optimization problem with based on the solution of the model. Hence, real cluster heads are determined.

Note that \( n \) is the number of nodes in the network. After this step, the sink performs Inter cluster heads routing and creates a spanning tree between cluster heads. Using the spanning tree, the cluster heads transmit their data in a multi hop manner to the sink. For each node, its parent is selected based on the below formula:

\[
\frac{\text{energy}(CH_i)}{\text{distance}(CH_i, CH_j) \cdot \text{hop - count} - \text{to} - \text{sink}(CH_j)}
\]
In this formula for each cluster head \( j \), among other cluster heads, cluster head \( i \), a cluster head with highest value in the above formula is selected as the parent in the spanning tree. Each cluster head \( j \) selects its parent as follows: a cluster head which has higher energy and lower hop count to the sink, and also be the nearest cluster head to node \( j \) is selected as its parent.

Having done the cluster head election, the sink broadcasts the information about the cluster heads to the network. Then distributed phase is started. Each cluster head broadcasts a control packet of its information in its surroundings, and then other nodes select the nearest cluster head which has higher energy as its cluster head, and send a request to join message to the selected cluster head. It is notable that in both phases after determining the parent of each node, the node adjusts the transmission signal strength based on the distance between himself and its parent.

5. PERFORMANCE EVALUATION

As mentioned in Section 3, the way to formulate a problem affects the size of a problem and the size of the problem in turn determines the processing time. So, to prove the scalability of the proposed model, the comparison results of this work and the previous work which uses the ILP technique [21] are performed. In Table 1 the number of variables and the constraints of these models are presented.

As shown in Table 1, the number of the variables and constraints of the proposed algorithm has order of complexity of \( N \), so it is expected that the proposed model has high scalability compared to the precious model which has complexity \( N^2 \). In Table 2 the processing time of these models are presented under different \( N \). The commercial solver IBM Ilog Cplex 12.1 available under academic license is used here to measure the processing time. As expected before, processing time of the proposed algorithm has a considerable improvement compared to the previous work.

<table>
<thead>
<tr>
<th>model</th>
<th>number of variables</th>
<th>number of constraints</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR model</td>
<td>( 2N^2+2N )</td>
<td>( 2N^2+8N+7 )</td>
<td>[21]</td>
</tr>
<tr>
<td>MIN-DIST-CHS</td>
<td>( N )</td>
<td>( N )</td>
<td>this paper</td>
</tr>
</tbody>
</table>

Figure 1. The proposed clustering algorithm
To present the processing time of the proposed models in Section 3, Figure 2 shows the solution time of the models per different node number (N).

![Figure 2. Processing time of models](image)

To evaluate the performance of the proposed clustering algorithm, the proposed clustering algorithm is simulated and compared to the LEACH. The simulation network environment is considered as a 25m*25m square area. The number of nodes (N) varies between 50 and 500. The initial energy of each node is a random variable less than 200 joule. The communication range of all nodes is set to 4 meters.

We suppose at each node \(i\) the variables \(ER_i\) and \(EA_i\) represent the required energy for data receiving and data aggregation, respectively. Furthermore, suppose \(dist_{ij}\) and \(ET_{ij}\) are the distance and required transmission energy between node \(i,j\), respectively. Using the energy model given in [29], the following equations are obtained where \(c_1, c_2\) and \(c_3\) are three fixed variables:

\[
ER_i = c_1 \times K
\]
\[
ET_{ij} = c_2 \times K \times dist_{ij}^2
\]
\[
EA_i = c_3 \times K
\]

Furthermore, the performance of the proposed algorithm is evaluated following some evaluation parameters: network energy, alive nodes number and network lifetime. Some events in the network simulation are produced as follows: When a node detects an event, it forwards the sensed data to the sink using spanning tree. All nodes that sense an event and forward their data to the sink, consumes energy for only data transmission (\(ET_i\)). CH nodes consume energy for both data receiving and data transmission (\(ER_i\) and \(ET_{ij}\), respectively). In addition, they also consume energy for data aggregation (\(EA_i\)).

In Figure 3 and Figure 4, the value of N, the number of nodes in network, is 100 and 450, respectively. To evaluate the performance of the proposed model, in the end of each round the energy of alive nodes is determined as the total energy of the network. As the MIN-DIST-CHS model considers distance in addition to the coverage of the nodes, so the solver tries to minimize distance between cluster members and their cluster heads. And also in the distributed phase the cluster members select the nearest cluster head. So, the transmission distance decreases which in turn, leads to less energy consumption in data transmission. As a result, in MIN-DIST-CHS model the energy of the network is saved better than MIN-CH-NUMS model.
Here the energy consumption of the proposed clustering algorithm by MIN-DIST-CHS model and LEACH clustering algorithm is analyzed under different scenarios with N=100, 300 and 450, and the results are shown in Figure 5, Figure 6 and Figure 7, respectively. As presented in the figures, the proposed clustering algorithm outperforms the LEACH algorithm considerably. As described previously, in order to achieve energy efficiency, the proposed algorithm selects the cluster heads which provide minimum transmission distance in the network, so the energy consumption in the network is decreased in data transmission phase. As presented in the figures, although the number of nodes increases, this property is maintained. The proposed algorithm in addition to scalability, it has energy efficiency property. Whereas in the LEACH, cluster head selection is done based on probabilities so it could not be efficient in terms of energy.

Another parameter to evaluate the performance of the proposed algorithm is the number of the alive nodes. After each simulation round, the alive nodes in the network are counted, as in the proposed algorithm, only nodes with high energy level are selected as a cluster head candidate, it causes that nodes with low energy remains longer time in the network. Additionally as mentioned in the the energy of the network comparison results during the cluster head selection phase, the cluster heads are selected by considering its energy and
distance, so the nodes select a suitable cluster head which led to less energy consumption. Efficient energy consumption has a direct effect on the lifetime of the nodes. The results of the comparison are shown in Figure 8, Figure 9 and Figure 10 under different scenarios with N=100, 300 and 450, respectively.

![Figure 8. Number of network nodes at N=100](image)

![Figure 9. Number of network nodes at N=300](image)

![Figure 10. Number of network nodes at N=450](image)

![Figure 11. Life time of the network, life time is defined as deaths of at least 12% of the nodes](image)

In Figure 11 and Figure 12, for clustering algorithm the lifetime is plotted versus the number of nodes. As the proposed algorithm consumes less energy and has more alive nodes, so its lifetime is longer than the LEACH. In these figures, network lifetime is defined as the longevity of the network in terms of the number of rounds performed before the failure of 12% and 20% of the total nodes, respectively.

6. CONCLUSION AND FUTURE WORKS

Clustering algorithm is a hierarchical architecture to route and transmit data in WSN. An efficient cluster head selection in clustering algorithms has a significant impact on the performance of the network.
This paper proposed ILP mathematical models as a basis to develop an optimize cluster heads selection mechanism for WSN nodes clustering. The proposed clustering algorithm has efficient performance in network energy consumption. Through simulations with different scenarios and comparison to the LEACH algorithm this paper concludes the claim.

As a future work, we consider to investigate other ILP techniques to be combined to provide better complexity degree in WSN routing and transmission problems.

REFERENCES

BIOGRAPHIES OF AUTHORS

Zahra Eskandari received her MS degree in computer engineering, with concentration in Wireless Sensor Networks from Ferdowsi University of Mashhad, Iran. She has published some conference and journal papers. Her research interests are wireless sensor networks and optimization using linear and nonlinear programming.

Seyed Amin Hosseini Seno received his B.Sc and M.Sc. degree in Computer Engineering from Ferdowsi University of Mashhad, Iran in 1990 and 1998 respectively and his Ph.D. from Universiti Sains Malaysia, Penang, Malaysia in 2010. He is currently Head of Virtual Learning Center at the Ferdowsi University of Mashhad. His research interests include Wireless and Sensor Networks, Network protocols, QoS and Network Security.


Rahmat Budiarto received B.Sc. degree from Bandung Institute of Technology in 1986, M.Eng, and Dr.Eng in Computer Science from Nagoya Institute of Technology in 1995 and 1998 respectively. Currently, he is a professor at College of Computer Science and IT, Albaha University, Saudi Arabia. His research interests include IPv6, network security, Wireless sensor networks and MANETs.