A Novel Clustering Algorithm in Wireless Sensor Networks

Jiangpeng Dai
Key Laboratory of Computer Network and Information Integration, Ministry of Education
School of Computer Science & Engineering
Southeast University
Nanjing, P. R. China
jpdaai@seu.edu.cn

Shu Liu
Key Laboratory of Computer Network and Information Integration, Ministry of Education
School of Computer Science & Engineering
Southeast University
Nanjing, P. R. China
s_lau@seu.edu.cn

Abstract—The demands of energy-efficient, distributed and scalable features are critical factors for designing topology control and routing management schemes in wireless sensor networks. To meet these demands, clustering is a good method in sensor networks topology control and routing. But by analyzing former clustering algorithms, it is found that there are deficiencies and limitations because of the usage of random method in cluster head election. In this paper, clustering mechanism is modeled and analyzed theoretically, and then turned to approximate optimization problem. A novel heuristic information based clustering algorithm is proposed to solve the above problem. The performance of this novel algorithm is analyzed and validated by simulation experiments. The results of experiments indicate that the algorithm can conduct clustering in network well and obtain excellent network performance, such as improved robustness, lower global energy cost and longer network lifetime. In particular, the performance is still stable while increasing the nodes scale.

Keywords—wireless sensor network; clustering; heuristic; pheromone; guidance information

I. INTRODUCTION

A wireless sensor network (WSN) [1] is composed of a large number of sensor nodes, enables a wide variety of applications, including environmental monitoring, medical treatment, emergency response, outer-space exploration, etc. To ensure reliable network performance, sensor nodes need to be organized into a robust self-organizing multi-hop wireless network, which poses several challenges in the network design. First, due to the features of sensor networks, the high energy-efficiency should be considered as the most critical factor. Second, the networks should be distributed and scalable. Third, these unattended sensor nodes should be self-organizing and applications oriented.

Facing these challenges and research issues, several hierarchical approaches have been introduced into sensor networks topology control and routing in order to prolong the lifetime of the networks with energy-efficiency. Their methods are essentially partitioning the network into clusters and making sensor nodes only need to communicate within clusters while the inter-cluster communications are handled by cluster heads [2-4]. Many clustering algorithms in various contexts have been proposed in the past such as GAF [5] algorithm and LEACH [6] algorithm. In these algorithms, the process of election is based on stochastic method while some factors are ignored, such as the connectivity of network and local status information of sensor nodes. Our work starts with the aim of generalizing the clustering in a robust and energy-efficient way in order to improve the performance of the whole network.

It has been proved that constructing the optimal clustering for wireless sensor networks with a large number of energy-constrained sensor nodes is NP-hard [7]. Heuristic method is an effective approach to solve the problem [8-9]. In this paper, we focus on research of heuristic information based sensor networks clustering. We generalize the clustering model and propose a novel fast, heuristic, distributed algorithm for cluster head election, to provide the networks better energy efficiency, adaptability and scalability.

II. CLUSTERING MECHANISM IN SENSOR NETWORKS

The clustering problem can be summarized as follows: partitioning the N sensor nodes in the network into K sets, making each set has one leading node and M_i (i=1, 2, ..., K) common nodes which can communicate with leading node at minimum average cost. These sets are called clusters. The leading nodes are called cluster heads while the sensor nodes except heads are called cluster members. The result of a good clustering algorithm should satisfy the following requirements:

(1) Decreasing the energy consumption of sensor nodes and minimizing the communication cost of the whole network.

(2) Balancing the residual energy of nodes all over the network especially for the cluster heads.

There are two main design approaches for clustering in existing algorithms. One is, for some algorithms such as LEACH, TEEN [10], HEED [11], to elect cluster heads first and then make other nodes choose clusters to join in according to some rules; the other is, for algorithms such as GAF and clustering method in [12], to partition the sensing area into several sub-areas as clusters and choose one node as cluster head in each cluster. In most of these clustering algorithms, cluster heads are elected randomly in order to balance the energy consumption of the whole network. However, such stochastic method overlooks the residual energy of nodes and the cost of communication inside cluster. It may cause the distribution of cluster heads unbalanced and thus the communication cost among member nodes and cluster heads would become higher.

Reference:
In general, an optimal clustering solution for wireless sensor networks should be as follows: to identify a subset of nodes within the network by appropriate algorithms as heads of certain nodes in their proximity. In the hierarchy every node is no more than one hop away from a cluster head. The cluster head election algorithm is the key in this kind of solution and our choice is to use heuristic information based method during the election process.

III. A NOVEL HEURISTIC CLUSTERING ALGORITHM

A. WSN clustering model

In the sensing area $D$, sensor nodes are distributed to form a wireless sensor network for monitoring. In formalization language, the network can be described as a digraph $G(V, E)$ in two-dimension plane. $V$ is the vertex set, representing all the sensor nodes in WSN, and $|V|=n$. $v_i \in V(1 \leq i \leq n)$ means that element $v_i$ is one of the nodes. $E$ is the edge set, $E=V \times V$, representing all the links among all nodes. $e_{ij} \in E(1 \leq i, j \leq n)$ means that element $e_{ij}$ is the link from $v_i$ to $v_j$. The cluster set is denoted as $C=\{c_1, c_2, ..., c_n\}$ and element $c_j(1 \leq j \leq nc)$ is one cluster. Each node $v_i(1 \leq i \leq n)$ belongs to one of the clusters.

In order to propose a heuristic algorithm to solve the clustering problem, we make the following assumptions:

1. $v_i$ can get the status of itself and neighbor nodes. Based on the status $v_i$ chooses $v_j$ or $v_i$ itself as the cluster head candidate and is willing to join in its cluster.

2. The heuristic information on one node changes if that node is chosen as a cluster head candidate. This change can be got by neighbor nodes. For that node, the status of the heuristic information may influence the chance of being selected as head candidate next time.

B. Algorithm Design

In this section, we first describe the novel algorithm and related parameters, and then give the details and steps of the algorithm.

1) Algorithm description and related parameters

The object of a heuristic clustering algorithm for WSN is $G(V, E)$. Sensor nodes elect the cluster heads according to the heuristic information and join in clusters at the end of this process. The heuristic information includes pheromone and guidance information. After getting the heuristic information about itself and its neighbors, $v_i$ can figure out the probability of electing $v_j$ as cluster head candidate according to the pheromone and guidance information, and then express the will to join in the temporary cluster whose head is $v_j$, denoted as $v_i \in \text{precluster}(j)$. This process leads to the change of pheromone on $v_j$. Also, the pheromone of $v_j$ is volatilizing while accumulating. Then the heuristic information status of $v_i$ is determined by both of the residual pheromone and the guidance information. After the heuristic iteration runs for several rounds, several nodes can be selected as cluster heads and the clusters can be established.

Given the distance between cluster members and cluster head is no more than one hop in the hierarchy, we set one relatively low radio range as the one-hop distance $R_{\text{single}}$. The nodes are neighbors if they can communicate with each other directly within $R_{\text{single}}$. The $v_i$’s neighbors set, denoted as $\text{Neighbor}(i)$, is defined as follows:

Definition 1 (Neighbors Set): \text{Neighbor}(i) is the set of $v_i$’s neighbor nodes in graph $G$.

\[
\text{Neighbor}(i) = \{ v_j \mid \text{dist}(i, j) \leq R_{\text{single}}, (1 \leq i, j \leq n) \}
\]

The amount of nodes in neighbor set is limited.

Definition 2 (Pheromone Content and Increment): $\tau_j(t)$ is the pheromone content of $v_j$ at time $t$, $1 \leq j \leq n$. $\tau_j(t+1)$ is the pheromone content at next time. After the cluster head candidate selection, it changes as follows:

\[
\tau_j(t+1) = (1-\rho) \cdot \tau_j(t) + \Delta \cdot \tau_j(t)
\]

$\rho (\in [0,1])$ is pheromone volatile coefficient, $(1-\rho)$ shows the proportion of residual pheromone on nodes. $\Delta \tau_j(t)$ is the pheromone increment on $v_j$ from time $t$ to next time, defined as:

\[
\Delta \tau_j(t) = \frac{Q \cdot E_{\text{max}}}{\text{precluster}(j) \cdot \text{Neighbor}(i)}
\]

$Q$ is pheromone constant. $E_{\text{residual}}$ and $E_{\text{max}}$ present the residual energy and initial maximum energy of $v_j$ respectively.

Definition 3 (Guidance Information Function): In the procedure of cluster head selection, the function describing the expectation degree of selecting $v_j$ as cluster head candidate is called guidance information function. It is on the basis of the communication cost among nodes, defined as follows:

\[
h_j = \frac{1}{\text{avg}(v_j \in \text{Neighbor}(i)) \cdot \text{precluster}(j) \cdot \text{Neighbor}(i) \cdot \mathcal{E}}
\]

$\text{avg}(v_j \in \text{Neighbor}(i)) \cdot \text{precluster}(j) \cdot \text{Neighbor}(i)$ is the average value of energy cost by communicating among $v_i$ and its neighbors, representing $v_i$’s average communication cost, denoted as $\mathcal{E}$.

Definition 4 (Selection Probability): It shows the probability of selecting one node as cluster head candidate for $v_i$. That node may be one of $v_i$’s neighbors $v_j$ or $v_i$ itself. It is calculated based on heuristic information around $v_i$ and denoted as $p_{vj}^i$.

\[
p_{vj}^i = \frac{\frac{Q \cdot E_{\text{max}}}{\text{precluster}(j) \cdot \text{Neighbor}(i)}}{\text{avg}(v_j \in \text{Neighbor}(i)) \cdot \text{precluster}(j) \cdot \text{Neighbor}(i)}
\]

$\alpha$ is pheromone heuristic factor, while $\beta$ is guidance information heuristic factor.

2) Algorithm Steps

The running process of the algorithm is:

Step 1: Initialization of the nodes in graph $G(V, E)$. Confirming the neighborhood according to one-hop distance $R_{\text{single}}$. Setting the pheromone initial value $\tau_j(0)$ with constant $C$. Calculating guidance information value $\eta_j$ based on $\mathcal{E}$, as in (5). Every node broadcasts these values to its neighbors.
Step 2: Calculating selection probability \( p_{ij} \) for every node \( v_i \), according to (6) \( (v_j \in \text{Neighbor}(i) \) or \( j=i) \).

Step 3: Every node \( v_i \) selects cluster head candidate \( v_j \) according to \( p_{ij} \) and is maintained as the member of precluster('j'). \( v_i \) updates its pheromone content according to (2)(3)(4), and broadcasts it to all neighbors.

Step 4: If \( N_i \) is less than predetermined iteration number, and the result does not converge, turning to step 2; otherwise turning to step 5.

Step 5: Every node \( v_i \) selects one of its neighbor nodes \( v_j \), which contains the biggest value of pheromone content, to be cluster head (if there are several nodes with the same maximum pheromone content value, choose the one whose index value \( j \) is minimum). If the head is \( v_i \) itself, \( v_i \) broadcasts elected message to its neighbors; otherwise, it sends affiliation message to the elected cluster head \( v_j \) as its member node.

Step 6: As a node which receives the affiliation message, \( v_j \) broadcasts cluster head elected message. If \( v_j \) selected other node as cluster head in step 5, it will drop the affiliation to that node by sending affiliation dropping message, and modify its role as cluster head.

Step 7: Conducting sensing and data collecting, until next clustering cycle, then turning to step 1.

C. Analysis of algorithm

We analyze the novel algorithm theoretically on the aspects of correctness, complexity, parameters and performance.

(1) Analysis of correctness and complexity

**Property 1**: After the running process of the algorithm, every node is either a cluster head or a member node in one cluster. Every cluster will have and only have one cluster head.

Proof: In step 5 of algorithm, every node selects one of its neighbor nodes or itself to be its cluster head. This is an enforced process. So there is no such node working as neither a cluster head nor a member node belonging to a cluster. In addition, assuming there is a cluster which does not contain a cluster head. That means none of nodes in this cluster select itself to be cluster head. They all select some neighbor node as their head. So, in this cluster, there must be some node receiving cluster affiliation message. According to step 7, it will change its role to be cluster head. This result conflicts with assumption. So the assumption is wrong.

Property 1 shows that the clustering result of the algorithm is one kind of complete partition of network topology.

**Property 2**: The time complexity of algorithm is \( O(n) \); the message exchange complexity is \( O(n) \), the message exchange complexity of every node is \( O(1) \).

Proof: In every step of algorithm, the time complexity can be determined clearly. Assume the number of sensors is \( n \). It is \( O(n) \) in step 1, \( O(N_{\text{sel}} \times N_{\text{iter}} \times n) \) in step 2, \( O(N_{\text{iter}} \times n) \) in step 3 and 4. In remaining steps, although role modification may happen, the time for clustering is no more than \( O(2n) \). \( N_{\text{sel}} \) and \( N_{\text{iter}} \) are the numbers of neighbors and iterations respectively. According to their definitions, they are both constant values. So, the time complexity for all running process of algorithm is \( O(n) \).

Meanwhile, no matter in which period of running process of the algorithm, the complexity of sending message from one node to another is not more than \( O(1) \). So the message exchange complexity of every node is \( O(1) \), and for the entire network, this kind of complexity is \( O(n) \).

(2) Analysis of parameters and performance

Pheromone content \( \tau \) is the basis of the algorithm. During the process of cluster head candidate selection, the \( \tau \) value on each node is keeping changing. The increment is \( D \tau \). In our design, \( D \tau \) is related to the proportion of residual energy of node, as in (3)(4). It reflects the tendency to select a node in better energy status to be cluster head.

The guidance information \( \eta \) is to guide the process of solving problem to a right direction. It reduces the performance fluctuation of algorithm, and improves the global convergence. In the algorithm, \( \eta \) is related to the average communication cost \( ACC \), as in (5). To each node, the smaller value of \( ACC \) means the larger value of \( \eta \).

Heuristic factors \( a \) and \( \beta \) are very critical parameters to determine the weights of \( \tau \) and \( \eta \) in selection probability calculation. \( a \) shows the emphasis on the usage of pheromone and \( \beta \) shows the importance of guidance in the algorithm. It is very critical to set suitable values for \( a \) and \( \beta \) in order to make sure that the algorithm performs well.

The parameter \( \rho \) is to prevent pheromone from accumulating unlimited. The value of \( \rho \) determines the impact of residual pheromone in cluster head candidate selection and influences the process and performance of algorithm.

IV. SIMULATION AND EVALUATION

For the purpose of analyzing and evaluating the novel clustering algorithm based on heuristic information, we have used OMNeT++ platform to conduct a series of simulation experiments, and compared the algorithms’ performance.

A. Simulation Environment

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{\text{max}} )</td>
<td>Maximum communication distance</td>
<td>150m</td>
</tr>
<tr>
<td>( R_{\text{rel}} )</td>
<td>Single hop communication distance</td>
<td>75m</td>
</tr>
<tr>
<td>( k )</td>
<td>Energy cost coefficient</td>
<td>( 1 \times 10^{-3} ) J/kb*m²</td>
</tr>
<tr>
<td>( \tau )</td>
<td>Pheromone initial value</td>
<td>15</td>
</tr>
<tr>
<td>( Q )</td>
<td>Pheromone constant</td>
<td>3</td>
</tr>
<tr>
<td>( E_{\text{max}} )</td>
<td>Initial maximum energy</td>
<td>100 J</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Pheromone heuristic factor</td>
<td>5</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Guidance information heuristic factor</td>
<td>1</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Pheromone volatile coefficient</td>
<td>0.2</td>
</tr>
<tr>
<td>( I )</td>
<td>Iteration number of algorithm</td>
<td>5</td>
</tr>
</tbody>
</table>

We assume that \( N \) sensor nodes are randomly dispersed into a field with dimensions 400m \( \times \) 400m. The coordinates of nodes obey two-dimensional uniform distribution. One node is selected randomly as sink node to be responsible for gathering all data collected by other
nodes. For simplicity, we ignore the energy cost of receiving packets and assume the absence of some factors such as noise and physical obstacles in our experiments. Table 1 shows the major parameters in experiments.

B. Evaluation and Comparison

We evaluate the performance of the novel algorithm and compare it with LEACH and GAF. Fig.1 shows the comparison of their performance curves while increasing the nodes scale. The curves labeled by HC represent the performance of the novel heuristic clustering algorithm. Fig.1a illustrates the curves of average communication cost among nodes in cluster. From this figure, we can see that, compared to LEACH and GAF, the performance of the novel algorithm is more stable. That is because we use heuristic information based algorithm in cluster head election, which makes the distribution of heads more uniform and the organization of cluster more reasonable.

![Graph of average communication cost](image)

![Graph of average cluster head residual energy](image)

Fig. 1b shows the curves of average residual energy of cluster heads. It is obvious that the performance of the novel algorithm is much better than those of LEACH and GAF. All these are due to using heuristic method in cluster head election instead of using stochastic method. It is no doubt that this performance increment will prolong life time of sensor networks.

V. CONCLUSIONS

In this paper, we discuss the former clustering algorithms and their deficiencies, then model and analyze the WSN clustering problem theoretically. In order to get the approximate optimal solution to this discrete NP-hard optimization problem, we put forward a heuristic information based clustering algorithm by using heuristic method. We analyze and evaluate the performance of this novel algorithm by simulation experiments. The experiment results indicate that this algorithm can conduct clustering in network well and obtain excellent network performance, such as high robustness, low global energy cost and prolonging network lifetime. In particular, the performance is still stable while increasing the nodes scale.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under Grants No. 60603067 and 90604003.

REFERENCES