A low coverage first classifying algorithm for wireless sensor networks

GUO Jian*, SUN Lijuan, WANG Ruchuan, XIAO Fu

College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing, 210003, China

Abstract

The issue of node scheduling for wireless sensor networks (WSNs) is studied in this paper. In WSNs with node scheduling, nodes may be classified into different sets and each set is activated in turn. All nodes in one set will continue working until another is awakened. Hence the power of the network is saved and the conflicts in the channel are reduced. In this paper, classification-based scheduling algorithms are researched and a low coverage first classifying algorithm (LCFC) is proposed. Comparison experiments showed that LCFC achieved good results.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: Wireless Sensor Networks; Node Scheduling; Low Coverage First

1. Introduction

Wireless sensor networks (WSNs) are the hot of current researches in computer networks [1]. WSNs are the combination of many technologies, such as wireless communication, sensors, embedded computing and distributed information processing, and they are widely used in military, environmental, medical, transport and other fields. Coverage control is one of the key technologies for applications of WSNs. It affects the monitoring quality as well as the energy consumption greatly.

When deploying WSNs, high coverage rate is a common requirement. To this end densely deployment is often adopted, which causes the phenomenon of coverage overlap in the network. The more nodes are used, the more serious the situation is. When all nodes deployed are in the state of working, collisions in
channel and retransmission of data will be inevitable, because multiple nodes may catch up a same event at the same time and submit a report synchronously. It causes many problems, including increase of communication delay, decrease of network throughput, and waste of the node energy. To solve these problems, scheduling algorithms are introduced, in which nodes are activated in turn, while the monitoring quality is guaranteed at a high level still.

By now, many scheduling algorithms have been proposed, such as random scheduling algorithms [2], location-based scheduling algorithms [3-4], cluster based scheduling algorithms [5], and neighbour information-based scheduling algorithms [6-7]. Among these algorithms, _K_-set based scheduling is a classical idea, in which nodes are classified into _K_ sets and activated alternately. The value of _K_ is associated with the application requirements, which all sets should satisfy. A typical example of _K_-set based scheduling problem is the set _K_-cover problem [8], in which full coverage of target area is required. Of course, this demand is not very common in most applications.

Because of its simplicity, _K_-set based scheduling problem attracts much attention. Abrams et al [9] proposed a random scheduling algorithm and two greedy scheduling algorithms to solve this problem. In random scheduling algorithm, each node randomly selects a set to joins in. In greedy scheduling algorithms, each node chooses a set minimizing the blind area. These three algorithms achieved some results. Since the problem of dividing nodes into _K_ sets is a NP hard problem [8], many approximate algorithms are used in recent researches. Deshpande et al [10] simplified the set _K_-cover problem into a max _K_-cut problem, and gave an approximate solution. Hu et al [11] adopted genetic algorithms to solve this problem. However, these methods have problems of high complexity and premature convergence, and are not practical.

In this paper, _K_-set based scheduling problem is studied and a low coverage first classifying algorithm (LCFC) is proposed. In order to ensure the coverage rate of each set, regions covered by few nodes are always considered firstly in LCFC. In this way, the monitoring capacity of all nodes is fully utilized, and coverage rate of each set can reach a high level. The effectiveness and superiority of LCFC were demonstrated through experiments. Contrast tests showed that, performance of LCFC was better than other algorithms, and its average coverage rate was always the highest.

2. _K_-set based scheduling problem

Suppose there are _n_ sensor nodes in the network. For a node _i_, 1 ≤ _i_ ≤ _n_, the point set it covers is _s_i_. And assume that the set of all points in the target area is _s_sum_.

Definition 1: Network coverage _P_ _area_. _P_ _area_ is the ratio between the monitoring area of all active nodes and the region area, that is

\[
P_{area} = \frac{|(s_1 \cup s_2 \cup ... \cup s_n) \cap s_{sum}|}{|s_{sum}|}
\]  (1)
From Formula (1), we may get $p_{area} \leq 1$. When $p_{area} = 1$, the problem is referred as the full coverage problem.

Definition 2: K-set based scheduling problem. For a given sensor set $S$ and number $K$, all nodes in $S$ are expected to be assigned into $K$ subsets, denoted as $C_1$, $C_2$, $\ldots$, $C_K$. These $K$ subsets should satisfy the following conditions:

1. $C_1 \cup C_2 \cup \ldots \cup C_K = S$
2. $C_i \cap C_j = \emptyset$ when $i \neq j$
3. $\forall i$, maximize $p_i$ as much as possible

Where, $1 \leq i \leq K$, $1 \leq j \leq K$, and $p_i$ is the coverage rate of subset $C_i$.

3. Network model

In this paper, WSNs are supposed to follow these assumptions.

(1) Nodes are randomly deployed in a square region with sides of $a$.
(2) The sensing range of a sensor node is a disc, whose radius is $r$ and center is the coordinate of the node, as shown in Figure 1.

For a point $A$ on the plane, it is thought to be covered by a sensor $O$ when Formula (2) is satisfied.

\[ \text{dist (O, A)} \leq r \quad (2) \]

Where, $\text{dist (O, A)}$ is the Euclidean distance between $A$ and $O$, and $r$ is the perception radius.

(3) All nodes have the same sensing radius.

4. Low coverage first classifying algorithm

4.1. Design and implementation of LCFC

Since the monitoring field is a continuous area, it is difficult to calculate the coverage rate directly with Formula (1). Hence, discretizing the monitoring area is implemented. Discrete points with equal interval $\Delta d$ in the field are selected and the coverage rate of these points is taken as the approximation of the real coverage rate. Obviously, the smaller the value of $\Delta d$ is, the higher accuracy the algorithm gets.

For these points, some may be covered by many sensor nodes, while others are relatively less. We call the latter as the low coverage points. The main idea of LCFC is that, when deciding which class to join in, a node chooses the set which can cover most low coverage points. In this way, the coverage overlap of the network is reduced effectively and the monitoring capacity of each node is fully utilized.

The storage structure of LCFC is discussed below.

$C_i$: $C_i$ is a set, $1 \leq i \leq K$. In the initial stage, $C_i$ is empty. After the algorithm is run, it stores the ID of the sensor nodes assigned to it.

Cover: $\text{Cover}_i$, $1 \leq i \leq K$, and its size is $(a/\Delta d+1) \times (a/\Delta d+1)$. For an element $c_{mn}$ in $\text{Cover}_i$, $1 \leq m$, $n \leq a/\Delta d+1$, assuming its corresponding discrete point is $x$, the value of $c_{mn}$ can be calculated by Formula (3).

\[ c_{mn} = f_p(x) = \begin{cases} \infty & \exists s, s \in C_i \text{ and dist}(s, x) \leq r \\ \frac{1}{S_{\text{unassigned}}} & \text{otherwise} \end{cases} \quad (3) \]
where, $S_{\text{unassigned}} = \{s \mid s \in \{S - C_1 \cup C_2 \cup ... \cup C_K \}, \text{and } s \text{ satisfying dist}(s, x) \leq r \}$. With $C_i$ and $\text{Cover}_i$, the profit of a node $s_j$ joining in a set $C_i$ can be calculated by Formula (4).

$$f(s, C_i) = \sum_{\text{dist}(s_j, x) \leq r} \frac{1}{f_p(x)}$$

Formula (4) indicates the idea of priority of low coverage. For a node $s_j$, if it could cover more low coverage points in set $C_i$, the value of $f(s_j, C_i)$ will be larger, and the possibility of it joining in $C_i$ will be higher. Hence, the coverage overlap will be reduced greatly.

![Sensing model of a sensor node](image)

Fig. 1. Sensing model of a sensor node

Pseudo codes of LCFC are shown in Figure 2.

```plaintext
/* Initialization phase */
Initialize $S$ and $K$
for $i = 1$ to $K$
    Begin
        Initialize $C_i$ and $\text{Cover}_i$
    End
/* Classification phase */
for $j = 1$ to $n$
    Begin
        for $i = 1$ to $K$
            Begin
                Calculate the profit of $s_j$ joining in $C_i$
            End
        Find the set $C_m$ with max profit
        $C_m = C_m + \{j\}$
        $S = S - \{j\}$
    End
    for $i = 1$ to $K$
        Begin
            Updating $\text{Cover}_i$
        End
End
```

Fig. 2. Pseudo codes of LCFC
5. Simulations and tests

5.1. Experiment setting

The LCFC algorithm was realized and tested in the Matlab 7.4 environment. In addition, we also implemented three strategies, namely random classifying strategy [9] (Random), greedy classifying strategy [9] (Greedy) and genetic algorithm-based strategy (GA-based). These four algorithms were compared under several scenarios. Test parameters are shown in Table 1.

5.2. Results

The average coverage rates of all classes in the network are shown from Figure 3 to 4, corresponding to 800, 2400 nodes respectively. It can be seen from the figures that, with the number of nodes increases, the coverage rate of all algorithms are improved. In these four algorithms, the performance of LCFC is the best, whose average coverage rate is always the highest. In contrast, the effect of random is the worst in all scenes.

Table 1. An example of a table

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployment</td>
<td>Random</td>
</tr>
<tr>
<td>Scene size</td>
<td>400 × 400 (m²)</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>800, 1600, 2400</td>
</tr>
<tr>
<td>Sensing radius</td>
<td>30 (m)</td>
</tr>
<tr>
<td>Number of K</td>
<td>5, 10, 15, 20, 25, 30, 35</td>
</tr>
</tbody>
</table>

Fig. 3. Average coverage of all sets (n = 800)

Fig. 4. Average coverage of all sets (n = 2400)
6. Conclusion

When there are many nodes in a network, classifying them and scheduling alternately is a rational approach. It can save the network energy and reduce the channel conflicts. LCFC proposed in this paper uses a strategy of low coverage first, which can balance the coverage rate of each set. Its performance is verified well by our experiments.

Acknowledgements

We are grateful that the subject is sponsored by the National Natural Science Foundation of P. R. China (No. 60973139, 60773041, 61003039, 61003236), Scientific & Technological Support Project (Industry) of Jiangsu Province (No.BE2010197,BE2010198), The Special Foundation for Development of Modern Service Industry of Jiangsu Province, Project sponsored by Jiangsu provincial research scheme of natural science for higher education institutions(10KJB520013, 10KJB520014), Scientific Research & Industry Promotion Project for Higher Education Institutions(JH10-14), Science & Technology Innovation Fund for higher education institutions of Jiangsu Province(CX10B-196Z, CX10B-197Z, CX10B-200Z), the six kinds of Top Talent of Jiangsu Province (2008118), Doctoral Fund of Ministry of Education of China (20103223120007) and key Laboratory Foundation of Information Technology processing of Jiangsu Province (KJS1022).

References