DATA MINING VIA MINIMAL SPANNING TREE CLUSTERING
FOR PROLONGING LIFETIME OF WIRELESS SENSOR NETWORKS

GUANGYAN HUANG†, XIAOWEI LI‡
Advanced Test Technology Lab.,
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, P. R. China 100080
†huanggy@ict.ac.cn
‡lxw@ict.ac.cn

JING HE ∗
Chinese Academy of Sciences Research Center
on Data Technology and Knowledge Economy
Beijing, P. R. China 100080
hejing@ucas.ac.cn

XIN LI
School of Management, Royal Holloway, University of London
Egham, Surrey, UK, TW20 0EX
Xin.Li@rhul.ac.uk

Received (Day Month Year)
Revised (Day Month Year)

Clustering is applied in wireless sensor networks for increasing energy efficiency. Clustering methods in wireless sensor networks are different from those in traditional data mining systems. This paper proposes a novel clustering algorithm based on Minimal Spanning Tree (MST) and Maximum Energy resource on sensors named MSTME. Also, specific constrains of clustering in wireless sensor networks and several evaluation metrics are given. MSTME performs better than already known clustering methods of LEACH and BCDCP in wireless sensor networks when they are evaluated by these evaluation metrics. Simulation results show MSTME increases energy efficiency and network lifetime compared with LEACH and BCDCP in two-hop and multi-hop networks respectively.

Keywords: Clustering; data mining; wireless sensor networks; energy efficiency; minimal spanning tree.

∗Corresponding author.
1. Introduction

1.1. Data mining and wireless sensor networks

Traditional data mining discovers useful information or knowledge from already known databases. Wireless sensor networks comprising hundreds or even thousands of sensors extract physical data from ambient environment. These physical parameters are objective entities which may exist and be kept unknown from people until wireless sensor networks sense and express them by readable data messages. Thus, wireless sensor networks extend the concepts of databases in traditional data mining system.

Like traditional data mining, the goal of wireless sensor networks is to provide useful information or knowledge for people from physical worlds. Wireless sensors are deployed randomly in the physical area to sense sample data from environment and form a collectivity of user-cared knowledge.

1.2. Clustering in wireless sensor networks

Clustering methods used in wireless sensor networks can reduce energy dissipation of the network. This is very important in wireless sensor networks, because sensors have limited battery energy resource and wireless communications consume large amount of energy. The energy consumed by wireless communications is related to the number of data transmitted and transmitting distance. Optimal routes can be chosen to get the minimal average transmitting distance. Furthermore, although they have only limited computing abilities, sensors can fuse data by compressing or getting rid of redundancy to reduce the number of data and then send the user-cared knowledge to Base Station (BS). Thus, sensors are clustered into several clusters and some sensors are chosen as cluster heads (CH) to do the fusing and middle-transferring jobs.

However, different transmitting distances and different jobs of non-CH sensors and CHs bring different energy dissipation in the wireless sensor networks. Thus, all the sensors should work together as a team and evenly distribute energy dissipation to prolong the whole network lifetime that is defined as the Time span from the deployment to the instant when the First sensor Node Dies (TFND) in this paper. Also, the Ratio between TFND and the Time span from the deployment to the instant when All sensor Nodes Die (TAND), RTFTA, is used to evaluate the effectiveness of evenly distributing energy dissipation. Thus, network lifetime is determined by energy efficiency and effectiveness of evenly distributing energy dissipation.

Clustering in wireless sensor networks is expected to group the initially ungrouped sensors according to the proximity of locations to bring the two advantages of reducing both the number of data transmitted and the transmitting distance and distributing energy dissipation evenly in the whole networks, and thus to prolong the network lifetime. Due to the spatial characteristic of the physical environment,
natural physical distance is used as the key parameter to cluster the sensors, for the signal strength increases with the transmission distance in most cases. Like BCDCP, this paper assumes the coordinates of nodes are already known, maybe using GPS.

1.3. Assumptions and definitions for modeling of clustering in wireless sensor networks

There are some specified constraints when clustering algorithms are applied in wireless sensor networks. Some assumptions and definitions are given as follows.

(i) A round is defined as the process when all the sensors have transmitted their data to BS once. Clusters formed in the current round would be different from those in last round. No fixed optimal clusters are used in the whole process.

(ii) High-energy sensors have energy resource above average level among all sensors in the networks. At least one high-energy sensor is distributed in each cluster in each round as a CH.

(iii) Assuming energy is dissipated mainly on transmitting, receiving and fusing data in wireless sensor networks. All the sensors have the same initial energy resources at the beginning. In a round, non-CH sensors dissipate only transmitting energy. CHs dissipate far more energy than non-CH sensors, considering that they must receive more data, fuse those data, and middle-transmit fused data.

(iv) A sensor is called "dead" when it uses up its energy resource. The goal of the clustering algorithm is to generate different clusters dynamically in each round and ensure that no sensors would become dead long before the others. That means all the sensors in the networks should be dead or alive at the same time.

(v) All sensors are randomly and evenly distributed in the wireless networks.

1.4. Limitations of using traditional clustering algorithms in wireless sensor networks

Traditional clustering algorithms such as DBSCAN\textsuperscript{13}, Chameleon\textsuperscript{14}, CURE\textsuperscript{15}, algorithms in Ref. 1 and Ref. 2 etc., are not good methods to be applied in wireless sensor networks for the reasons as follows:

(i) Although they can find the natural density distribution of the sensors, the number of sensors in each cluster varies significantly. Also, it cannot ensure that at least one high-energy sensor be in each cluster. This would cause the energy dissipation to be distributed in the whole network too unevenly.

(ii) They would produce fixed optimal clusters, and thus unfairness would be generated among the sensors. This would make some sensors dead long before the others.

(iii) Single or few objects as noise would be neglected in traditional data mining system, but in wireless sensor networks, no sensors can be considered as noise.
1.5. How to evaluate the clustering algorithms in wireless sensor networks

Good clustering algorithms in wireless sensor networks would generate clusters in each round satisfying the five optimal rules as follows:

(i) **Proximity**: The average distance between neighbor sensors is the shorter the better in each cluster.

(ii) **Same Number**: The number of sensors in each cluster should be as the same as possible.

(iii) **Maximum Energy**: The energy resource on CHs is the higher the better, and is at least above average level.

(iv) **Even Location**: CHs are distributed evenly.

(v) **Dynamic Change**: Different clusters are generated dynamically in different round.

1.6. Overview of MSTME algorithm

This paper extends the work we have done in Ref. 16 and proposes in detail a novel center-controlled clustering method based on Minimal Spanning Tree (MST) and Maximum Energy resource on sensors (MSTME). MSTME do well in the five optimal rules of Sec. 1.5. The main idea of MSTME is as follows. Firstly, those sensors with energy resource above average level in the network are selected into a CH candidate set, $S$. Then all the candidates in $S$ are connected with a MST with respect to their locations of two-dimension coordinates. Non-candidate sensors become supporters of their closest candidates. At last, a given number, $p$, edges are broken in the MST to divide $S$ into $p+1$ subsets and it makes the sum of supporters in each sub MST (or subset) nearly the same. In each subset, the candidate with maximum energy resource is chosen as CH.

2. Related Work

In cluster-based wireless sensor networks, expected CHs are those sensors that can group approximately same number of closer sensors into clusters, have high energy resource and may be distributed in the network as evenly as possible. Also, they should overcome the limitations of traditional clustering algorithms in data mining system and do better in terms of Maximum Energy and Dynamic Change. That means to satisfy the five optimal rules in Sec. 1.5.

2.1. Clustering in LEACH

Low Energy Adaptive Clustering Hierarchy (LEACH) gives a simple distributed clustering scheme for evenly distributing energy dissipation. In Ref. 17, only sensors that have not yet been CHs recently, and presumably have more energy available than those that have recently performed as CHs may become CHs in
the current round. When the sensors begin with equal energy, a simple probability function is used to rotate positions of CHs in all sensors. Thus, clustering in LEACH satisfies two optimal principles of Maximum Energy and Dynamic Change. However, LEACH does not consider the optimal energy dissipation of each round because the sensors in the same cluster are not near enough. Moreover, the number of sensors varies significantly and the CHs are never distributed evenly in LEACH. The typical clusters of LEACH are shown in Fig. 1.

The center-controlled version of LEACH is called LEACH-C, whose performance is nearly the same as LEACH, so like in BCDCP, we only consider LEACH in this paper.

2.2. Clustering in BCDCP

Base Station Controlled Dynamic Clustering Protocol (BCDCP) is a center-controlled method to improve LEACH by considering both energy efficiency and effectiveness of even distribution of energy dissipation. To reduce energy dissipation of each round, BCDCP reduces the transmitting distances of both non-CH sensors and CHs. Clusters are formed by iterative cluster splitting algorithm. Sensors that have energy resource above average level are chosen as candidate sensors. Then two candidate sensors with longest distance are chosen as CHs and the remaining sensors would be affiliated to their nearer CHs. Thus, all the sensors in the network are split into two subsets and the subsets would be split iteratively by the same way. At last, CHs are connected with a MST, and one leader is chosen to send data to BS. To evenly distribute energy dissipation, BCDCP adopt balanced clustering technique in Ref. 20 to make the clusters have approximately the same number of sensors. Clusters in BCDCP (shown in Fig. 3) do better in terms of both Proximity and Same Number than in LEACH (shown in Fig. 1).

Compared with traditional clustering algorithm in data mining system, classic clustering methods applied in wireless sensor networks generally can satisfy rules of Maximum Energy and Dynamic Change. Clustering algorithms in both LEACH and BCDCP firstly consider the sensors with more energy resource to be chosen as CHs. Also, they generate different clusters in each round. Dynamic changes of clusters during three consecutive rounds in LEACH and BCDCP are shown in Fig. 1 (a)-(c) and Fig. 3 (a)-(c) respectively. However, both LEACH and BCDCP cannot do well in Even Location. Therefore, MSTME is given to improve the clusters further in terms of Proximity, Same Number and Even location. Also, clusters in MSTME are compared with those in LEACH and BCDCP by evaluating them with a new clustering evaluation model.

3. MSTME: Clustering Using Minimal Spanning Tree

MST on a finite set of points \((X_1, X_2, ..., X_n)\) in \(R^2\) is the connected graph with these points as vertices and with the minimum total edge length. Applications of MST in data clustering are called single-linkage cluster analysis. In wireless
sensor networks, $X_i$ denotes two-dimension coordinates of the locations of sensors. Also, each edge represents the physical distance between two sensors. The same as clustering in LEACH and BCDCP, high-energy sensors are considered prior as CHs. Because high-energy sensors vary in different rounds, Dynamic Change rule is also satisfied. Therefore, the main goal of MSTME algorithm is to choose those expected CHs, which should be distributed evenly in the sensor networks to reduce transmitting energy dissipation of CHs and also should group approximately the same number of closer sensors into clusters to reduce transmitting energy dissipation of non-CH sensors and evenly distribute the energy dissipation of CHs.

3.1. **MSTME algorithm**

In MSTME, sensors with energy above average level are chosen as CH candidates. A MST is created to describe the closeness of the CH candidates. Non-candidate sensors become supporters of their closest candidates. A expected number, $N_{CH}$, of CHs are chosen by MSTME algorithm, which satisfy all of the five optimal principles. CH choosing algorithm of MSTME is given in Table 1.

| Step 1 | Sensors with more energy resource than average level are selected into CH candidates set, $S$. |
| Step 2 | A MST, $T$, is used to connect all the items in $S$. |
| Step 3 | Supporters of a CH candidate $x$ are those non-candidate sensors that are nearest to $x$ among all CH candidates. Compute the number of supporters for each CH candidate including candidate itself. |
| Step 4 | Suppose supporters are just around their candidates and thus the latter can delegate the former to decide which edge would be split. |
| Step 5 (Initialization) | Let the number of already split edge $nSplit=0$, $T'=T$ and $S'=S$. |
| Step 6 (Loop) | Find an edge, which breaks $T'$ into two sub MSTs of $T1$ and $T2$ and at the same time the nodes in $S'$ are grouped into two subsets of $S1$ and $S2$ respectively with the nearest number of supporters in both subsets. Then let $nSplit=nSplit+1$. |
| Step 7 (Termination test) | If $nSplit=N_{CH}$ go to Step 8. Otherwise, go on splitting $S1$ and $S2$ in turn. If the number of supporters in $S1$ (or $S2$) is more than $N/N_{CH}$, then let $S'=S1$ and $T'=T1$ (or $S'=S2$ and $T'=T2$). Go to Step 6. |
| Step 8 | The CH candidate with the most energy resource among CH candidates in each subset is chosen as real CH. If more than one CH candidates have maximum energy resource in their subset, then randomly chose one as real CH. |

Using algorithm in Table 1, given number of CHs are chosen, and then non-CH
sensors are affiliated to their closest CHs and thus clusters are formed in each round. However, we must ensure that the assumption in Step 4 is valid. This guarantees that CH candidates can represent their supporters to give decisions for edge splitting.

**Theorem 3.1.** In wireless sensor networks, assume all the sensors have the same initial energy resource and sensors are deployed evenly and randomly in the network area, then non-candidate supporters are approximately around their CH candidates in each round.

**Proof.** With the same initial energy resource, all the sensors in the networks should be connected by a MST, and then $N_{CH}$ sub trees are formed by MST-based methods in Table 1. Because all the sensors are distributed in the network area evenly and randomly, CHs chosen by MSTME method are distributed evenly and randomly in the first round. The principle of nearly the same number of supporters in MSTME ensures that each split sub tree covers nearly the same size of network area. Also, according to Step 8 in Table 1, CHs are randomly chosen. Therefore, CHs are distributed evenly and randomly in the whole network area in the first round. This provides basic conditions for the later rounds to choose CH candidates. In the second round or later rounds, all the sensors generally have different energy resources. Because CHs and non-CH sensors that are far from CHs consume more energy according to the energy dissipation model in Sec. 1.3, those sensors nearer to latest CHs would become CH candidates in the current round. Therefore, in the local area of the networks, non-candidate supporters are approximately around CH candidates.

MST is reasonable for clustering while evenly distributing energy dissipation in wireless sensor networks. Firstly, MST connects all CH candidates with the minimal edges. Wherever the edge is broken, the nodes in the same split sub trees are closer. Because non-CH candidate sensors support the nearest CH candidates, CH candidates may delegate their supporters in terms of location. Thus, sensors in the same clusters formed by MSTME are closer to each other. Also approximately the same number of supporters in each CH candidate subset ensures nearly the same number of sensors be closest to the CH in this subset. At the same time, because the sensors are evenly and randomly distributed in the network area, the nearly same number of sensors may cover the same size areas. Each area includes one CH, that is, CHs are distributed evenly in the whole networks. At last, any nodes in MST are high-energy sensors among all sensors in the networks. Therefore, MSTME synthetically does better in the five optimal principles.

Like BCDDCP, MSTME algorithm runs on the BS. Therefore MSTME is a center-controlled clustering method, assuming that all the locations of sensors have already known.
3.2. Discussions of optimal number of CHs

However, optimal number of CHs is determined by topology of the networks. In this paper, six CHs are used in two-hop networks of both LEACH and MSTME according to Ref. 17. But in multi-hop topology like in BCDCP, the optimal number of tree-connecting CHs is not the same as LEACH. In this paper, nine CHs in multi-hop networks are used in both BCDCP and MSTME.

In multi-hop networks, the number of CHs affects average transmitting distance of each sensor and the number of data to be fused. Firstly, the greater the number of CHs, the less the average square transmitting distance of non-CH sensors, but the greater the average square transmitting distance of CHs. Moreover, the MST used to connect CHs produces more data to be fused. Thus, the greater the number of CHs, the more energy dissipated on data fusion. The optimal trade-offs on transmitting energy dissipation of non-CH sensors and energy dissipation of both CH transmitting and data fusion are determined by the number of CHs.

In any case, given optimal number of CHs, MSTME is better than both LEACH and BCDCP in both two-hop and multi-hop networks respectively.

4. Comparison of Sample Clusters in MSTME with LEACH and BCDCP

A sample wireless sensor network with \( N = 100 \) nodes randomly deployed in \( M \times M \) (\( M = 100 \text{m} \)) area is adopted to evaluate clusters generated by MSTME, LEACH and BCDCP. In fact, the three clustering methods do not depend on the absolute distance of the wireless sensor networks and also they do not require application-related parameters. These are their advantages compared with the clustering methods in traditional data mining system. Moreover, the three clustering methods in wireless sensor networks satisfy optimal cluster rules of both Maximum Energy and Dynamic Change. The first principle of choosing CH is based on high-energy sensors in each clustering methods and it is also ensure that the latest CHs cannot be chosen as CHs again in the current round. Some examples of clusters in three consecutive rounds are shown in Fig. 1-Fig. 4, and they show that clusters even in consecutive rounds are so different. Therefore, clusters in them are analyzed mainly from the remaining three optimal cluster rules of Proximity, Same Number and Even Location.

Also, distributed LEACH and center-controlled BCDCP are selected to compare with center-controlled MSTME, because LEACH is the most popular routing protocols in wireless sensor networks and BCDCP published in 2005 is one of those that delegate the most resent development level of clustering methods in wireless sensor networks.

Clusters in the 100th round to 102nd round are chosen accidentally as examples to evaluate the performance of MSTME in both single static round and dynamic changes in consecutive rounds. Actually clusters in arbitrary rounds have the same effectiveness. These Arbitrarily selected consecutive rounds are used to evaluate the
above three optimal rules statically in this section. However, to evaluate the whole performance of the clustering methods, a serial of consecutive rounds from the first round to the last round should be considered, and thus the detail simulation results are given in Sec. 5.

4.1. Modeling of evaluation rules

4.1.1. Proximity

In wireless sensor networks, energy dissipation can be cut down by reducing the average square transmitting distance. It is the Proximity rule’s goal. Thus, closeness of intra-clusters are computed by the average distance between non-CH sensors and their CHs as follows:

$$E_C = \frac{\sum_{i=1}^{N - N_{CH}} d_i^2}{N - N_{CH}}$$  \hspace{1cm} (1)

where $N$, $N_{CH}$, $d_i$ are the total number of sensors in the networks, the number of clusters, and the distance between the $i$-th non-CH sensor and its CH.

4.1.2. Same Number

The difference of the number of sensors in clusters is evaluated by

$$E_{SN} = \frac{\sum_{j=1}^{N_{CH}} |N_j - \bar{N}|}{N_{CH}}$$  \hspace{1cm} (2)

where $N_j$, $\bar{N}$ and $N_{CH}$ denote the number of sensors in $j$-th clusters, the average number of sensors in a cluster and the number of clusters in the whole network respectively.

4.1.3. Even Location

The distance between any CH and its nearest CH cannot be too long or too short, thus a MST is used to connect all the CHs, and the difference between each edge of the MST and the ideal average distance between two close CHs. Also, geometrical center of all the CHs should be near to the center of the network area. Thus, the Even Location metric is computed as follows:

$$E_{EL} = \left( \frac{\sum_{j=1}^{N_{CH} - 1} |e_j - M|}{N_{CH} - 1} \right) \times \left( \frac{\sum_{i=1}^{N_{CH}} x_i}{N_{CH} - x} + \frac{\sum_{i=1}^{N_{CH}} y_i}{N_{CH} - y} \right)^2$$  \hspace{1cm} (3)

where $e_j$, $M$, $(x_i, y_i)$ and $(x, y)$ are the length of the $j$-th edge of the MST, the length or width of the network area, the two-dimension coordinate of $i$-th CH and the center of the network area, respectively.
The values of $E_C$, $E_{SN}$ and $E_{EL}$ are smaller, the performances are better.

4.2. MSTME vs. LEACH

In all the four figures from Fig. 1 to Fig. 4, solid dots denote CHs and different types of items denote different clusters. Fig. 1 and Fig. 2 show the distribution of both six CHs and six clusters in LEACH and MSTME respectively during three consecutive rounds. Compared clusters of LEACH in Fig. 1 and clusters of MSTME in Fig. 2, CHs and clusters shown in Fig. 2 are distributed more evenly than those in Fig. 1 at a glance.

Fig. 1. Six clusters in LEACH.

Fig. 2. Six clusters in MSTME.

Using (1) - (3) to evaluate the three metrics of **Proximity**, **Same Number** and **Even Location** respectively in both LEACH and MSTME in each round, numeric results are shown in Table 2. Table 2 shows clusters of MSTME in Fig. 2 (b) and (c) are better than those of LEACH in Fig. 1 (b) and (c), because the value of the three metrics of $E_C$, $E_{SN}$ and $E_{EL}$ are smaller. Although Table 2 also shows that
the values of both $E_C$ and $E_{EL}$ in MSTME in Fig. 2 (a) are a little more than those in LEACH in Fig. 1 (a), the average effectiveness in MSTME is better than LEACH.

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Closer ($E_C$)</th>
<th>Same Number ($E_{SN}$)</th>
<th>Even Location ($E_{EL}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEACH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>647</td>
<td>7.7</td>
<td>150.3</td>
</tr>
<tr>
<td>(b)</td>
<td>903.6</td>
<td>9.7</td>
<td>215.5</td>
</tr>
<tr>
<td>(c)</td>
<td>681.4</td>
<td>10.3</td>
<td>354.5</td>
</tr>
<tr>
<td>MSTME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>685.9</td>
<td>7.3</td>
<td>166.1</td>
</tr>
<tr>
<td>(b)</td>
<td>488.4</td>
<td>6</td>
<td>108.1</td>
</tr>
<tr>
<td>(c)</td>
<td>620.6</td>
<td>8</td>
<td>223.3</td>
</tr>
</tbody>
</table>

In a word, numerical results in Table 2 are approximately consistent with sen-

suous estimation shown in Fig. 1 and Fig. 2 that MSTME outperforms LEACH in terms of **Proximity**, **Same Number** and **Even Location**.

### 4.3. MSTME vs. BCDCP

Typical clusters of BCDCP and MSTME are shown in Fig. 3 and Fig. 4. According to Sec. 3.2, nine clusters are adopted in these examples. Also at a glance, Fig. 4 shows better distribution of CHs and clusters than Fig. 3.

Numeric results of $E_C$, $E_{SN}$ and $E_{EL}$ are given in Table 3 according to (1) -(3). Comparing the values in MSTME with those in BCDCP, the former are smaller than the latter at the most of time, except that $E_{SN}$ in Fig. 4 (a) and $E_{EL}$ in Fig. 4 (c) are a little more than those in Fig. 3 (a) and (c) respectively. This means clustering algorithm in MSTME are better than in BCDCP when they are evaluated by the three metrics of **Proximity**, **Same Number** and **Even Location**.

In wireless sensor networks, suppose the BS knows the locations of all sensors and MSTME runs on the BS. Both six and nine clusters formed by MSTME in two-hop and multi-hop topology networks respectively are better than their counterparts by evaluating the metrics of **Proximity**, **Same Number** and **Even Location**. Also, there is an interesting thing that numeric data in Table 3 are always smaller than in Table 2, because there are more clusters in multi-hop networks than in two-hop networks.

### 5. Performance Analysis

Clusters of sample rounds produced by MSTME method are analyzed with optimal rules in Section 4 and evaluated by the numeric results in Table 2 and Table 3. However, the goal of MSTME is to improve the energy efficiency of wireless sensor networks. So, the whole performance (all the rounds) of MSTME is evaluated by
Table 3. Numeric results for evaluating nine clusters in multi-hop networks.

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Closer ($E_C$)</th>
<th>Same Number ($E_{SN}$)</th>
<th>Even Location ($E_{EL}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCDCP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>750.3</td>
<td>1.9</td>
<td>46.3</td>
</tr>
<tr>
<td>(b)</td>
<td>422.6</td>
<td>2.8</td>
<td>35.2</td>
</tr>
<tr>
<td>(c)</td>
<td>376.9</td>
<td>3.9</td>
<td>46.7</td>
</tr>
<tr>
<td>MSTME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>293.2</td>
<td>3.4</td>
<td>39.3</td>
</tr>
<tr>
<td>(b)</td>
<td>300.6</td>
<td>2.3</td>
<td>30.6</td>
</tr>
<tr>
<td>(c)</td>
<td>275.7</td>
<td>2.3</td>
<td>87.4</td>
</tr>
</tbody>
</table>

simulating average energy dissipation and network lifetime in the wireless sensor networks. Two same two-hop and multi-hop networks as Fig.1 (or Fig.2) and Fig.3 (or Fig.4) respectively are adopted to evaluate energy efficiency of MSTME. In MSTME, clusters and CHs are generated by algorithm in Table 1. Then CHs collect data from their cluster members. At last CHs send their data directly to BS in two-hop networks like LEACH, and the multi-hop transmission is the same as BCDCP, that is, firstly to choose a leader among CHs, and then for other CHs to transmit
their data along the MST that connects all the CHs.

We have implemented an energy simulator of wireless sensor network programmed in C/C++. To ensure the simulation results of this paper correct, we use our energy simulator to do the same experiments done by LEACH in Ref. 17 and BCDCP in Ref. 19, and get the approximately same results.

5.1. Radio power model

In this paper, we use a radio energy model in Ref. 17, in which the energy dissipation, $E_T(k, d)$, of transmitting $k$-bit data between two nodes separated by a distance of $d$ meters is given as follows:

$$E_T(k, d) = \begin{cases} 
  k \times (E_{elec} + \varepsilon_{FS} \times d^2) (d < d_0) \\
  k \times (E_{elec} + \varepsilon_{MP} \times d^4) (d > d_0)
\end{cases}$$  (4)

where $E_{elec}$, $\varepsilon_{FS}$ and $\varepsilon_{MP}$ denote electronic energy, transmit amplifier parameters corresponding to the free-space model and the multi-path fading model respectively, and $d_0 = \sqrt[4]{\varepsilon_{FS}/\varepsilon_{MP}}$. The energy cost, $E_R(k)$, incurred in the receiver of the destination sensor node is given as follows:

$$E_R(k) = k \times E_{elec}.$$  (5)

All the experiments in this paper use the same wireless sensor networks with the sensor distribution shown in Fig. 1-Fig. 4. Initial energy resource on each node is 2J. The number of transmission data message for each node in each round, or one packet size, is 20kbits. Let $E_{elec} = 50nJ$, $\varepsilon_{FS} = 10pJ/bit/m^2$ and $\varepsilon_{MP} = 0.0013pJ/bit/m^4$. Energy dissipation of fusing 1-bit data message is 5nJ. Position of BS is $(25m, 150m)$.

5.2. Average energy dissipation

In two-hop (or multi-hop) wireless sensor networks, the energy dissipation of the network comprises the energy dissipations of transmitting, receiving and fusing data. Because energy dissipations of both receiving and fusing are only related to the number of data packets, they are the same in both LEACH (or BCDCP) and MSTME. MSTME reduces the average transmitting distance of all sensors by forming clusters with closer sensors and choosing CHs with both high energy resource in their own clusters and better locations than LEACH and BCDCP. Thus MSTME reduces the average energy dissipation of transmitting.

Fig. 5 plots average energy dissipation varies with number of rounds in two-hop wireless sensor networks. Fig. 5 is also used to evaluate the closeness of clusters because energy dissipation of transmitting data on non-CH sensors is reduced a lot by using MSTME to replace LEACH. Fig. 5 shows average energy dissipation in MSTME is less than in LEACH.

Also, average energy dissipation varying with number of rounds in multi-hop wireless sensor networks is shown in Fig. 6. Fig. 6 plots MSTME outperforms
Fig. 5. Average energy dissipation in two-hop wireless sensor networks.

BCDCP. MSTME performs better in multi-hop topology than in two-hop topology, because the average transmitting distance is reduced further by reducing the average length of edges in MST that connects CHs in MSTME.

Fig. 6. Average energy dissipation in multi-hop wireless sensor networks.
5.3. Network lifetime

The most important advantage of MSTME is to cluster better clusters to reduce the average energy dissipation while choosing CHs with higher energy resource, and thus to prolong network lifetime. Network lifetime is defined in Sec. 1 as TFND. Because the total energy resource of all sensors is constant, reducing the average energy dissipation can put off the deaths of most sensors and thus prolong TAND. However, TFND is determined by the first dead sensor. Therefore, all sensors should work together as a team to help each other to put off TFND. RTFTA can be used to evaluate the properties of even energy distribution. Network lifetime is just used to evaluate the two aspects of average energy dissipation and evenly distributing energy dissipation.

In two-hop wireless sensor networks, Fig. 7 shows TFND in MSTME increases 3.2% of that in LEACH and also TAND of MSTME outperforms 3.6% that of LEACH. MSTME excels LEACH in terms of both TFND and TAND in two-hop networks. RTFTA of 96% in MSTME is nearly the same as RTFTA of 96.4% in LEACH. It means that MSTME gets better trade-offs on clusters with even numbers of closer sensors and CHs with more energy resources.

Fig. 8 shows network lifetime of both BCDCP and MSTME in multi-hop wireless sensor networks. TFND in MSTME increases 6.4% of that in BCDCP. TAND in MSTME increases 5.6% of that in BCDCP. RTFTA of 98.7% in MSTME is nearly the same as RTFTA of 98% in BCDCP. Also MSTME outperforms BCDCP in terms of network lifetime.

MSTME excels LEACH and BCDCP because MSTME reduces average energy
dissipation of the networks while it keeps RTFTA approximately the same as both LEACH and BCDCP. In addition, it seems that MSTME performs better in Fig. 8 than in Fig. 7. This is because MSTME improves CH distribution in multi-hop network better by reducing average transmitting distance of CHs further.

6. Conclusions

MSTME proposes a good solution to cluster sensors in wireless sensor networks optimally. Based on MST, members in any sub sets of the CH candidate set are closer. Also because CH candidates have energy above average level, the sensors with most energy resources in any sub sets have higher energy resources among the sensors in the whole network. Nearest numbers of supporters in each sub sets make the final clusters have the number of sensors as near as possible, and thus energy dissipations of CHs are approximately the same. CHs with more energy resources also prolong TFND. In addition, closer sensors in clusters reduce the whole energy dissipation of networks. Simulation results show network lifetime of MSTME excels those of LEACH in two-hop and BCDCP in multi-hop networks.

Acknowledgement

This paper is supported in part by the National Basic Research Program of China (No. 2005CB321604 and No. 2004CB720103), and the National Natural Science Foundation of China (No. 90207002). The authors are grateful to the referees for their useful and constructive comments, which have resulted in an improved paper.
References
1. S.-J. OH, J.-Y. Kin, A sequence-element-based hierarchical clustering algorithm for
categorical sequence data, International Journal of Information Technology Decision
2. S.-G. Lee, D.-K. Yun, Clustering categorical and numerical data: a new procedure
using multidimensional scaling, International Journal of Information Technology De-
3. J. He, X. Liu, Y. Shi, W. Xu, N. Yan, Classifications of credit cardholder behavior
by using fuzzy linear programming, International Journal of Information Technology De-
W. Rabiner and A. Wang, Design considerations for distributed microsensor networks,
6. M. Dong, K. Yung and W. Kaiser, Low power signal processing architectures for
173–177.
7. D. Estrin, R. Govindan, J. Heidemann and S. Kumar, Next century challenges: scal-
8. J. Kulik, W. Rabiner and H. Balakrishnan, Adaptive protocols for information dis-
9. Y. Xu, J. Heidemann and D. Estrin, Geography-informed energy conservation for Ad
11. I. F. Akyildiz, Su Weilian, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor
12. W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishnan, Energy efficient communi-
13. M. Ester et al., A density-based algorithm for discovering clusters in large spatial
databases with noise, In Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining,
14. G. Karypis, E.-H. Han, and V. Kumar, Chameleon: hierarchical clustering using dy-
15. S. Guha, R. Rastogi and K. Shim, CUR: an efficient clustering algorithm for large
16. G. Huang, X.-W. Li, and J. He, Clustering versus evenly distributing energy dissi-
pation in wireless sensor routing for prolonging network lifetime, Lecture Notes in
17. W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishnan: An application-specific pro-
tocol architecture for wireless microsensor networks. IEEE Transactions on Wireless
19. S. D. Muruganathan, D. Ma, R. I. Bhasin, A. O. Fapojuwo, A centralized energy-
efficient routing protocol for wireless sensor networks, IEEE Communications Maga-
18. G. Huang, X.-W. Li, J. He & X. Li

