Meng JT, Yuan JR, Feng SZ *et al.* An energy efficient clustering scheme for data aggregation in wireless sensor networks. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 28(3): 564–573 May 2013. DOI 10.1007/s11390-013-1356-y

An Energy Efficient Clustering Scheme for Data Aggregation in Wireless Sensor Networks

Jin-Tao Meng[1,2,3] (孟金涛), *Member, CCF, ACM, IEEE*, Jian-Rui Yuan[4] (苑建蕊), *Member, CCF, ACM*
Sheng-Zhong Feng[1,*] (冯圣中), *Member, CCF, ACM, IEEE*, and Yan-Jie Wei[1,*] (魏彦杰)

[1]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*

[2]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China*

[3]*Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

[4]*CYG SUNRI CO., LTD, Shenzhen 518057, China*

E-mail: {meng.jintao, yuan.jianrui}@gmail.com; {sz.feng, yj.wei}@siat.ac.cn

Received November 22, 2012; revised February 28, 2013.

Abstract In wireless sensor networks, a clustering scheme is helpful in reducing the energy consumption by aggregating data at intermediate sensors. This paper discusses the important issue of energy optimization in hierarchically-clustered wireless sensor networks to minimize the total energy consumption required to collect data. We propose a comprehensive energy consumption model for multi-tier clustered sensor networks, in which all the energy consumptions not only in the phase of data transmissions but also in the phase of cluster head rotations are taken into account. By using this new model, we are able to obtain the solutions of optimal tier number and the resulted optimal clustering scheme on how to group all the sensors into tiers by the suggested numerical method. This then enables us to propose an energy-efficiency optimized distributed multi-tier clustering algorithm for wireless sensor networks. This algorithm is theoretically analyzed in terms of time complexity. Simulation results are provided to show that, the theoretically calculated energy consumption by the new model matches very well with the simulation results, and the energy consumption is indeed minimized at the optimal number of tiers in the multi-tier clustered wireless sensor networks.

Keywords wireless sensor network, clustering scheme, data aggregating, energy efficiency

1 Introduction

The technological advances in the micro-electro-mechanical systems and the wireless communications have reduced the costs on the deployment of the small intelligent sensor nodes at homes, in workplaces, supermarkets, streets, plantations, and oceans to monitor the environment[1]. The realization of smart environments to improve the efficiency of nearly every aspect of our daily lives by enhancing the human-to-physical world interactions is one of the most exciting potentials of sensor network applications. However, this objective calls for efficient and application specific communication protocols to assure the reliable and energy-efficient communications of the sensed event features.

The longevity of wireless sensor networks is usually limited by their non-rechargeable energy resources. This then motivated research efforts on such topics as energy efficient medium access control protocols[1-6], localization systems[7-8], time synchronization protocols[9-11], energy efficient routing protocols[12-18], and energy-efficient data gathering schemes[19-22]. In particular, many studies have been made on clustering schemes[12,15-16,19-21] with the aim of balancing and reducing energy consumption over the whole network.

Within the energy efficient communication protocols, many approaches have been proposed, such as Directed Diffusion (DD)[14], Two-Tier Data Dissemination (TTDD)[16], Low-Energy Adaptive Clustering Hierarchy (LEACH)[12], Power-Efficient Gathering in Sensor Information Systems (PEGSIS)[15]. Recent surveys[1,18] provide more information on these communication protocols. LEACH[12] was firstly proposed as a clustering routing protocol. It was validated to be adaptable to large-scale networks and has the capability of prolonging the networks lifetime significantly[18]. In LEACH, each sensor becomes a cluster head (CH) with fixed probability during startup and every non-CH sensor joins the cluster of the nearest CH. The CHs act as

Regular Paper

*Corresponding Author

aggregators. As CHs are likely to consume more energy than non-CHs, LEACH allows rotation of CH status to balance the energy consumption among sensors.

The recent studies in [19-21] have further discussed related issues in hierarchically clustering scheme of sensor networks. Explicitly, [19-20] propose their energy consumption models for multi-hop data collection in which they show that the energy saving increases with the number of tiers based on their energy consumption functions and simulations. However, in those energy consumption models, the energy that consumed in CH's rotation and data aggregation are not taken into account. Paper [21] has considered the aggregation factor into its energy consumption model, but it still does not cover the energy consumption in CH's rotation. The authors of [21] observed that there indeed exists the optimal number of tiers in the multi-tier hierarchical network through simulations. Unfortunately, no guideline is provided there on how to find this optimal solution. Therefore, for wireless sensor networks it still remains as a question that how to calculate the optimal tiers within a comprehensive energy consumption model that accounts for all the energy consumptions in CH rotations, data aggregations, and data communications.

This paper discusses this important issue of energy optimization in hierarchically-clustered wireless sensor networks to minimize the total energy consumption required to collect data. We propose a comprehensive energy consumption model for multi-tier clustered sensor networks, in which all the energy consumptions not only in the phase of data transmissions but also in the phase of cluster head rotations are taken into account. This model will include the energy consumption of protocols DD[14], LEACH[12] to be its special cases. By using this new model, we are able to obtain the solutions of optimal tier number and the resulted optimal clustering scheme on how to group all the sensors into tiers by the suggested numerical method. This then enables us to propose an energy-efficiency optimized distributed multi-tier clustering algorithm for wireless sensor networks. This algorithm is also analyzed theoretically in terms of computation complexity. Simulation results are provided to show that, the theoretically calculated energy consumption by the new model matches very well with the simulation results, and the energy consumption is indeed minimized at the optimal number of tiers in the multi-tier clustered wireless sensor networks.

The remainder of this paper is organized as follows. In Section 2, assumptions, definitions and notations that are used through the paper are introduced. Section 3 proposes the energy consumption model for data collection in multi-tier clustered sensor networks. Section 4 gives an example to show how to find out the optimal number of tiers and the corresponding optimal number of CHs for each tier. Subsequently, Section 5 proposes a distributed multi-tier clustering algorithm to construct this multi-tier clustered sensor network. In Section 6, we present the simulation results to show that the energy model has predicted the optimal number of tiers exactly and the proposed algorithm has optimized the energy consumption. Finally the conclusions are drawn in Section 7.

2 Assumptions and Notations

In order to build the energy model for the hierarchical-clustered network (a sample of two-tier clustered network can be seen in Fig.1), we make the following assumptions about the sensor network:

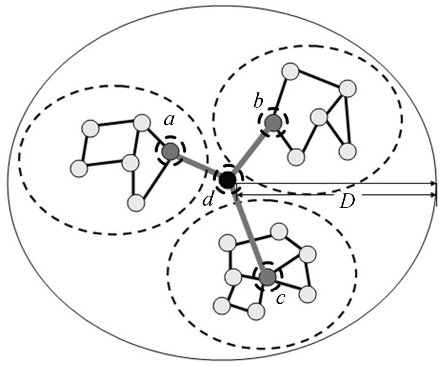


Fig.1. Sample of two-tier multi-hop clustered network. The first tier cluster is consisted of sensors a, b, c, d, here sensor d is the CH, and other three clusters with CHs a, b, c are the second tier clusters.

1) There are N homogeneous sensors that are distributed in a circle with radius D. The location of the sensors are governed by a homogeneous spatial Poisson process and a sink node is located at the center of the circle for data collection.

2) All sensors transmit at the same power level and hence have the same transmission range r.

3) Data exchanged between two communicating sensors not within each other's radio range is forwarded by other sensors. No data aggregation occurs at the forwarding sensors.

4) The cluster heads aggregate the data received by them from the sensors in their cluster before forwarding it further.

5) We introduce the energy consumed in CH's rotation at each round into our energy model, and ignore the energy used to aggregate the information at CHs as the computing energy consumption is always very small compared with CH's rotation in wireless sensor network.

6) Route calculation is carried out during the cluster set-up phase of each round, and the sensors send packets to their respective CHs using multi-hop paths (if necessary).

7) The distance of D between any sensor and its cluster head is equivalent to D/r hops.

8) In each round, the control information broadcast in the cluster set-up phase is m units, and the data sent in the data transmission phase is h units.

9) Each sensor uses one unit of energy to transmit or receive one unit of data.

10) We assume that the maximum number of the tiers will not exceed $K_{\max}$.

11) The communication environment is contention-free and error-free; and hence, sensors do not have to retransmit any data.

A summary list of the notations used in this paper is given in Table 1.

Table 1. Notations

Notation	Description
N	Number of sensors in the network
D	Radius of the network deployment region
r	Radius of sensors transmission range
m	Number of data units in control packet
h	Number of data units in data packet
$K_{\max}$	Maximum number of tiers
$k_{\rm opt}$	Optimal number of tiers
N_i	Number of sensors at the i-th tier (tier-i)
D_i	Radius of clusters at tier-i

3 Energy Models for Multi-Tier Clustering Schemes

In this section, we propose the energy consumption model of multi-tier clustering scheme in wireless sensor networks. For simplicity, we assume that N sensors are deployed in a circular region of radius D with the sink located at the center of the circle. Our model can be easily extended to accommodate other region shapes or sink locations.

3.1 Multi-Tier Clustered Network Topology Analysis

In each round of multi-tier clustering scheme, we construct a k-tier clustering scheme (where $0 \leqslant k \leqslant K_{\max}$) to collect the data packets in the region. The parameter of the i-th tier (tier-i) is (N_i, D_i), where N_i is the number of tier-i CHs, D_i is the radius of tier-i cluster. We set the sink node to be the tier-0 CH with the radius of the cluster being D. That is to say, $N_0 = 1$, $D_0 = D$. An example of three-tier clustered network topology structure is depicted by Fig.2. We also define the number of the common sensors in a k-tier clustering network to be N_{k+1}, and all the number of sensors at each tier N_i, $i = 0, 1, \ldots, k+1$ satisfy the following equation

$$\sum_{i=0}^{k+1} N_i = N + 1, \tag{1}$$

where $N_i \in [1, N]$. Note that for our purpose of optimizing the energy consumption and to obtain the solution of optimal tier, we have allowed $N_i, i = 0, 1, \ldots, k+1$ to be float numbers, though in real networks they are logically needed to be integer numbers.

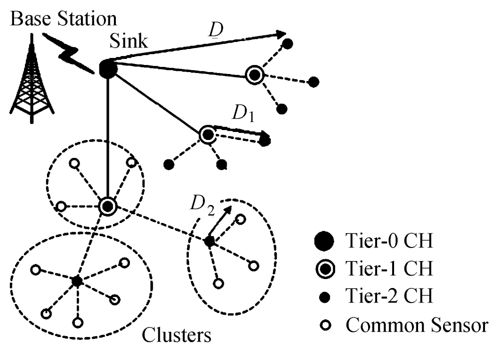


Fig.2. Example of three-tier clustered network topology structure.

If each sensor chooses the closest aggregator as its CH at each tier, the sensors essentially form a Voronoi diagram of the network region where each cluster corresponds to a Voronoi cell[23]. For each tier, a typical cluster can be approximated as a circle with the CH at the center. According to [24], we have the following relationship

$$D_i = \frac{D}{\sqrt{N_i}}, \quad i = 0, 1, \ldots, k+1. \tag{2}$$

Equations (1) and (2) consist of the topology constraints to our energy model which will be used later in deducing the energy consumption model.

3.2 Transmission Energy Model

As in our assumptions, the sensors are distributed according to a homogeneous spatial Poisson process and hence, N sensors are distributed uniformly in a cluster with radius D. Without loss of generality, we assume that the CH is at the center of the circle. Then let d be a random variable that denotes the total length of the segments from the common sensors to the CH (see Fig.3). Therefore,

$$E(d) = \int_0^{2\pi} \int_0^D \frac{x^2 {\rm d}x {\rm d}\theta}{\pi D^2} = \frac{2D}{3}.$$

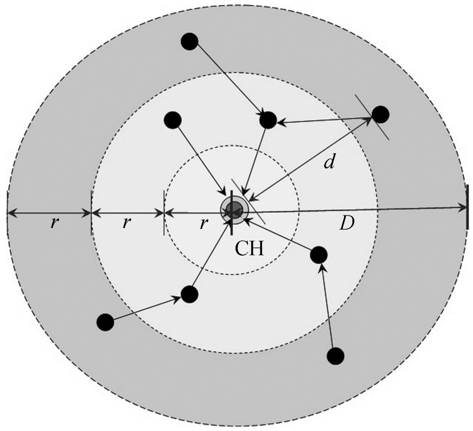


Fig.3. Typical cluster in a multi-hop clustered network.

Then, we define $E_t(D)$ to be the energy used by a sensor in the cluster with radius D to communicate one unit of data to the CH. We have

$$E_t(D) = 2\frac{E(D)}{r} = \frac{4D}{3r}. \quad (3)$$

Equation (3) quantifies the transmission energy that is consumed for a sensor node to communicate one unit of data with its CH.

3.3 Energy Consumption of Tier-i

According to the topology setting of multi-tier clustering scheme in Subsection 3.1, there are N_i clusters with radius D_i in tier-i (see Fig.4 for tier-i topology structure in the multi-tier clustered network). The main source of energy consumption in tier-i is the energy used in the set-up phase and the energy used in the data communication phase at each round.

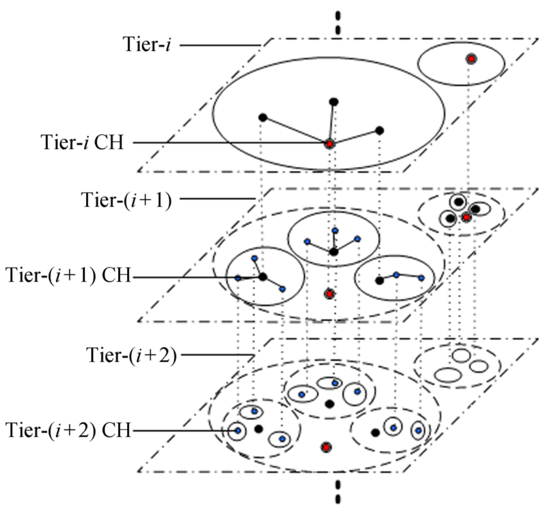


Fig.4. Tier-i network topology structure in multi-hop clustered network.

We define Es_i to be the energy used in the set-up phase for CH rotation of tier-i and we can get it by:

$$Es_i = N_i m + 2m\Big(N + 1 - \sum_{j=0}^{i} N_j\Big), \quad i = 0, 1, \ldots, k. \quad (4)$$

The first part of (4) is the energy consumed by the tier-i CHs broadcasting its control information to all the sensors under tier-i, and the second part of (4) is the energy consumed by all the sensors under tier-i to receive control information and send out the control information to disperse it. The tier-i routing table can be built at the same time, so that every sensor node under tier-i knows how to deliver data packet to reach the tier-i CHs.

We define Ed_i to be the energy used in the data communication phase of tier-i, and we can get it by:

$$Ed_i = N_i \frac{N_{i+1}}{N_i} E_t(D_i) h = \frac{4 D_i N_{i+1} h}{3r}, \quad i = 0, 1, \ldots, k. \quad (5)$$

In tier-i, there are N_i clusters and every cluster has $\frac{N_{i+1}}{N_i}$ sub-CHs at tier-$(i+1)$ or common sensors which will send the data it received or sensed to its CH, so a coefficient of N_i is multiplied. The energy used by $\frac{N_{i+1}}{N_i}$ sensors deployed uniformly in a circle with radius D_i transmitting h units of data to its CH should be $\frac{N_{i+1}}{N_i} E_t(D_i) h$.

Combining the (4) and (5), we can get the energy consumption E_i of tier-i:

$$\begin{aligned} E_i &= Es_i + Ed_i \\ &= N_i m + 2m\Big(N + 1 - \sum_{j=0}^{i} N_j\Big) + \frac{4 D_i N_{i+1} h}{3r}, \end{aligned} \quad (6)$$

where $i = 0, 1, \ldots, k$.

Then we will use this energy model to calculate the energy consumption of each tier in the multi-tier clustering scheme.

3.4 Energy Model of Multi-Tier Clustering Scheme

According to the topology construction of k-tier clustering scheme, we obtain the energy consumption Et_k of k-tier clustering scheme as follows:

$$Et_k = \sum\nolimits_{i=0}^{k} E_i, \quad (7)$$

where $k = 0, 1, \ldots, K_{\max}$.

By substituting (2) and (6) into the (7) we can get:

$$Et_k = 2km(N+1) - m\sum_{i=0}^{k}(2k+1-2i)N_i + \frac{4Dh}{3r}\sum_{i=0}^{k}\frac{N_{i+1}}{\sqrt{N_i}}, \tag{8}$$

where $k = 0, 1, \ldots, K_{\max}$.

Equation (8) is the energy model of k-tier clustering scheme.

3.5 Energy Model Analysis

Our purpose here is to find the optimal number of tier k_{opt} and the corresponding solution $N^{k_{\text{opt}}} = \{N_0, N_1, \ldots, N_{k_{\text{opt}}}, N_{k_{\text{opt}+1}}\}$ to reach the smallest energy consumption $\min(Et_{k_{\text{opt}}})$ among all the minimized energy consumption $\min(Et_k)$ of k-tier clustering scheme, where $k = 0, 1, \ldots, K_{\max}$.

The method we use here to get the optimal number of tiers is to find out all the minimized energy consumption $\min(Et_k)$ of k-tier clustering scheme first, and then we select the tier number which reaches the smallest energy consumption among all the minimized energy consumption $\min(Et_k)$ of k-tier clustering scheme as the optimal tier.

Now the only problem left is that for a k-tier clustering scheme whether there exist a optimal solution which minimizes the energy consumption of each k-tier clustering scheme and if it does exist, how to obtain the corresponding optimal solution.

We first analyze the energy model (8) of k-tier clustering scheme (where $k = 0, 1, \ldots, K_{\max}$) to get its first order derivative and second order derivative of (8) as below.

$$\frac{\partial Et_k}{\partial N_i} = -m(2k+1-2i) + \frac{4Dh}{3r}\Big(\frac{1}{\sqrt{N_{i-1}}} - \frac{N_{i+1}}{2N_i^{\frac{3}{2}}}\Big),$$

$$\frac{\partial^2 Et_k}{\partial^2 N_i} = \frac{DhN_{i+1}N_i^{-\frac{5}{2}}}{r}, \tag{9}$$

where $i = 1, 2, \ldots, k$.

As the energy consumption Et_k is continuous, and the second order derivative (9) is larger than 0, there exists a solution $N^k = \{N_0, N_1, \ldots, N_k, N_{k+1}\}$ which reaches the minimum energy consumption to this k-tier clustering scheme.

Now we set all the first order derivatives to be 0, and yield the following equations:

$$-m(2k+1-2i) + \frac{4Dh}{3r}\Big(\frac{1}{N_{i-1}^{1/2}} - \frac{N_{i+1}}{2N_i^{3/2}}\Big) = 0, \quad i = 1, 2, \ldots, k. \tag{10}$$

Therefore, we have k conditions here. Together with (1) and the setting $N_0 = 1$, we are able to obtain the optimal solution from these $k+2$ equations theoretically. We herein give a recursion method on how to obtain the required optimal solutions.

Note that (10) is reduced to the following equation by some mathematical manipulations:

$$N_{i+1} = 2N_i^{\frac{3}{2}}\Big(\frac{1}{N_{i-1}^{1/2}} - \frac{3mr(2k+1-2i)}{4Dh}\Big), \quad i = 1, 2, \ldots, k. \tag{11}$$

Given $N_0 = 1$, and if we also know the number of tier-1 CHs N_1, we can get all the number of other tiers' CH (i.e., $N_2, N_3, \ldots, N_{k+1}$) by (11), whereas (1) also must be satisfied. So a method of binary search on adjusting N_1 can be used to find a value for N_1 satisfying (1) and then by this iterating manner we can get the optimal solution for k-tier clustering scheme.

From the above discussions, we can see that the complexity of computing the optimal solution to minimize the energy consumption of a k-tier clustering scheme is $O(k\log(N))$ where N is the number of sensors. Hence the complexity of obtaining the optimal tier of multi-tier clustering scheme is $O(K_{\max}^2\log(N))$, where $K_{\max}$ will not exceed 10 even for a network consisting of million of sensors displayed in a vast area according to our experience of simulations.

4 Example of Finding Optimal Tiers of Multi-Tier Clustering Scheme

In this section, an example is given. We first describe the energy consumption of multi-tier clustering scheme by the proposed energy model and then we calculate, by the proposed numerical method, the corresponding optimal solution $N^k = \{N_0, N_1, \ldots, N_{k+1}\}$, where $k = 0, 1, \ldots, K_{\max}$.

The parameters of this example are set according to Table 2. Substituting these parameters into (8), and then the energy consumption model of this network is

$$Et_k = 2002k - \sum_{i=0}^{k}(2k-2i+1)N_i + \frac{200}{3}\sum_{i=0}^{k}\frac{N_{i+1}}{\sqrt{N_i}}, \tag{12}$$

where $k = 0, 1, \ldots, K_{\max}$.

In order to get the optimal solution to minimize the energy consumption of k-tier clustering scheme, the $k+2$ conditions are worked out as below:

Table 2. Parameters Setting in the Example

Parameter	Value
N	1 000
D	10
r	1
h	5
m	1
$K_{\max}$	5

$$\begin{cases} N_0 = 1, \\ \sum_{i=0}^{k+1} N_i = 1001, \\ N_2 = 2N_1^{3/2}\Big(\frac{1}{N_0^{1/2}} - \frac{3(2k-1)}{200}\Big), \\ N_3 = 2N_2^{3/2}\Big(\frac{1}{N_1^{1/2}} - \frac{3(2k-3)}{200}\Big), \\ \quad\vdots \\ N_{k+1} = 2N_k^{3/2}\Big(\frac{1}{N_{k-1}^{1/2}} - \frac{3}{200}\Big), \end{cases} \quad (13)$$

where $k = 0, 1, \ldots, K_{\max}$.

Now, for every k-tier clustering scheme, by solving these $k+2$ equations, we will get the minimized energy consumption and the corresponding optimal solution. This procedure is broken down as following.

When $k = 0$, there are no CHs in the network, which is our energy model in the special case of DD[14], and the optimal solution calculated by (13) is $N^0 = \{1, 1\,000\}$. Then the energy consumption of this scheme can be calculated directly by (12) to be $\min(Et_0) = 66\,664.67$.

When $k = 1$, there is only one tier of CHs in this network, which is our energy model in the special case for LEACH[12]. Certain numerical calculations by (13) yield the optimal solution: $N^1 = \{1, 60.4, 939.6\}$, then the optimal energy consumption of one-tier clustering scheme obtained from (12) is $\min(Et_1) = 14\,024.83$.

When $k \geqslant 2$, this is the energy model of multi-tier clustering scheme, and the optimal energy consumption of those clustering schemes together with the corresponding solution $N^k = \{N_0, N_1, \ldots, N_k, N_{k+1}\}$ can also be computed out. The optimal energy consumption and the corresponding solution of all k-tier clustering schemes are given in Table 3.

Now we have obtained the minimum energy consumptions when $k = 0, 1, 2, 3, 4$. Then we know that Et_2 consumed the minimum energy, so the optimum number of tier is $k_{\text{opt}} = 2$, and the optimal solution is $N^2 = \{1.0, 17.4, 145.0, 837.6\}$. Subsequently, the energy consumption will be minimized by constructing the network to be a two-tier clustering scheme with one tier-0 CHs (or sink node precisely), 17.4 tier-1 CHs, 145.0 tier-2 CHs and 837.6 tier-3 CHs (or common sensors precisely).

Table 3. Optimal Energy Consumption and Corresponding Solution of Clustering Scheme in the Example

k	N_0	N_1	N_2	N_3	N_4	N_5	$\min(Et_k)$
0	1	1 000.0					66 665
1	1	60.4	939.6				14 025
2	1	17.4	145.0	837.6			11 916
3	1	7.2	38.8	179.8	774.2		12 884
4	1	3.4	12.5	48.1	188.7	747.2	14 605

5 Distributed Multi-Tier Cluster Algorithm

Now, we will present a distributed multi-tier cluster algorithm to generate a k-tier clustering scheme for the multi-hop sensor network. By this we mean that there are k tiers of CHs, with tier-0 being the highest and tier-k being the lowest tier. In this clustered environment, the common sensors deliver the gathered data to the tier-k CHs. The CHs in tier-k aggregate this data and deliver the aggregated data to the tier-$(k-1)$ CHs and so forth. Finally, the CHs in tier-1 deliver the aggregated data to the sink node which is the tier-0 CH.

5.1 Algorithm Description

The algorithm works in an up-bottom fashion, i.e., it first elects the tier-1 CHs, then tier-2 CHs, and so on. The tier-1 CHs are chosen as follows. Each sensor decides to become a tier-1 CH with probability $\frac{N_1}{N}$ and broadcast a message to advertise itself as a tier-1 CH to the sensors nearby. This advertisement is forwarded to all the sensors without hop limitation which is not similar to the algorithm proposed in [19-20]. Each sensor that receives an advertisement will join the cluster by registering the route to the CH into their routing table and relay the advertisement to its neighbors.

The rest of sensors then elect N_2 sensors as tier-2 CHs with a probability $\frac{N_2}{N-N_1}$ and broadcast their decision of becoming a tier-2 CHs. Each sensor that receives an advertisement will join the tier-2 cluster by registering the route to the tier-2 CHs into their routing table and relay the advertisement to its neighbors. All the tier-2 CH will route data according to the tier-1 routing table to reach the tier-1 CHs. CHs at the tier-i are chosen in the similar fashion with probability $\frac{N_i}{N-\sum_{i=1}^{i-1} N_i}$, where $i = 3, 4, \ldots, k$. This algorithm generates k-tier clustering scheme.

The pseudo-code of this clustering algorithm is given below:

//Initialize
Calculate the optimal solution $N^k = \{N_0, N_1, \ldots, N_k, N_{k+1}\}$ for the k-tier clustering scheme
//Set up routing table
If (sensor has not received a tier-i message and has not been elected as a CH at this round)
  Listen for message
  If (a message (k, i, id, hop) is received)
    /*Register route to the CH's tier-i routing table*/
    $Routing\_table[i] = id$
    Broadcast message $(k, i, my\_id, hop + 1)$
//Elect tier-$(i + 1)$ CH
If ($i < k$ and the sensor has not been elected as a CH)
  $rand$ = generate a random number between 0 and 1
  if $\left(rand < \dfrac{N_{i+1}}{N - \sum_{j=1}^{i} N_j}\right)$
    $ch\_ther = i + 1$
    $routing\_table[i + 1] = -1$
    Calculate t = time taken by one message to pass through one hop
    /*A period of $2 \times \left\lceil \dfrac{D}{r} \right\rceil \times t$ time will be used to build each tier */
    Wait for $\left(2 \times \left\lceil \dfrac{D}{r} \right\rceil - hop\right) \times t$ time
    Broadcast message $(k, i + 1, my\_id, 0)$

In this algorithm, we have not limited the diffuse hop of the advertisement message broadcasted by the CH as [19-21] do, as the limitation of diffuse hop will lead to many forced CHs at each tier, which will disturb our energy model, and in our simulation the forced CH will waste more energy by forcing itself to be a CH, instead of joining a cluster.

At the beginning of each round, the sink node will broadcast a message $(k, 0, sink\_id, 0)$ to start the algorithm to generate a k-tier clustering scheme. If the sink node broadcast a message $(k_{\rm opt}, 0, sink\_id, 0)$ then the algorithm will generate a $k_{\rm opt}$-tier clustering scheme which will minimize the energy consumption among all multi-tier clustering networks.

5.2 Time Complexity of the Clustering Algorithm

The algorithm can take advantage of opportunities for concurrency in communication. The run-time of the algorithm is the time to transmit a message from all CHs at all levels in the cluster hierarchy to the members of their clusters. Since the cluster members in each tier are at most $2 \times \lceil \frac{D}{r} \rceil$ hops away from the CH at the worst case, a period of $2 \times \lceil \frac{D}{r} \rceil$ hops deliver time can be used to build each tier. The time complexity of the proposed algorithm to generate a multi-tier clustering scheme with k tiers is $O(\frac{kD}{r})$ in a contention-free environment, where D is the radius of the sensor area, r is the radius of sensor's transmission range. This will make the proposed algorithm more suitable for large networks.

6 Simulation Results and Model Validation

In this section, we will validate our energy model proposed in Section 3 and simulate the algorithm described in Section 5. We first perform some simulations on the energy consumption versus the number of CHs in the one-tier case in order to show that the energy consumption is minimized at the optimal number of CHs calculated from our model. Then we verify that the energy consumption is also minimized at the optimal number of tiers which is also calculated from our model. Finally, we study the optimal tiers and minimum energy consumption of multi-tier clustering network by adopting the distributed multi-tier clustering algorithm.

In all the simulations, we use a network of N sensors uniformly deployed in a circular region of radius D, and the communication range of each sensor is assumed to be 1. One unit of control packet will be diffused and five units of data packet will be collected in each round for each sensor. The system specification we use in this simulation is the same as the parameters setting in Table 2.

We now measure the energy consumed in a single-tier clustering network, which is similar to LEACH[12]. We select different numbers of tier-1 CHs, and then simulate the cluster setting up, data collection, packet transmission, data aggregation and transmitting the data to the sink node. Fig.5 shows the energy consumption versus the number of CHs in the one-tier clustering

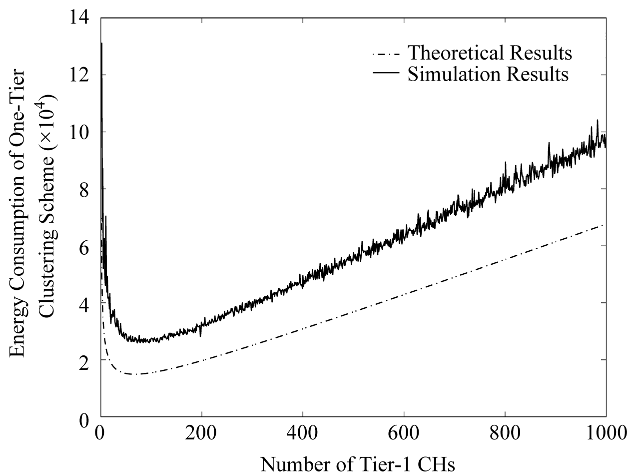


Fig.5. Energy consumption versus the number of CHs in the one-tier clustering scheme.

scheme. Observations from Fig.5 suggest that the total energy consumption appears to be minimized when the number of tier-1 CHs is between 50 and 80. Noting that this value predicted by our model is 60, thus it is evident that the energy is indeed minimum at the theoretically predicted optimal number of tier-1 CHs.

We now consider the energy consumption in a multi-tier clustering network. We perform simulations for each k-tier clustering scheme for $0 \leqslant k \leqslant 5$. We compute the optimal solution N^k for each k-tier clustering scheme according to (8), and the algorithm proposed in Section 5 is used to generate the corresponding k-tier clustering scheme. For each k-tier clustering scheme, we run 100 simulations. The averaged values of these simulation results and the theoretical results for each k-tier clustering scheme are plotted into Fig.6. From these experimental results, we conclude that for 1 000 sensors deployed in the circle with radius 10, the optimal tiers of the multi-tier clustering scheme used to minimize the energy consumption is 2 tiers. The theoretical results of our energy model also predict the same result as suggested by Fig.6. This then indicates that our energy model of (8) is suitable to be used in real situations.

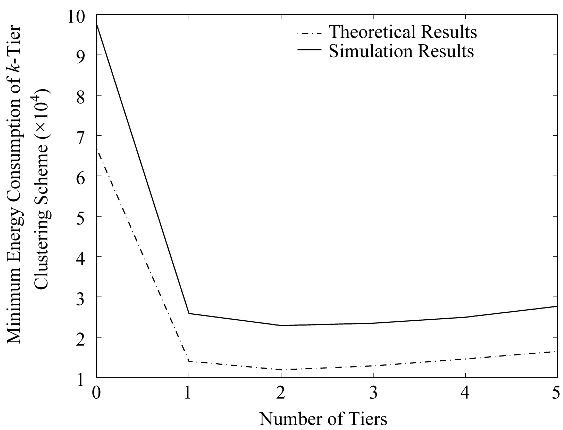


Fig.6. Energy consumption versus the number of tiers in the multi-tier clustering scheme.

Finally, we provide results of simulation experiments in the global situation of variable number of sensors N and variable radius of the sensor area D. The optimal number of tiers for each pair of (N, D) calculated by the energy model is plotted into Fig.7; the minimized theoretical energy consumption and the averaged energy consumption using the multi-tier clustering algorithm to generate the optimal clustered network in the simulation are plotted into Fig.8. Fig.7 shows that the optimal number of tiers of the multi-clustering scheme decreases with the number of sensors N, and increases with the radius of sensor area D. That is to say, the optimal number of tiers of the multi-clustering scheme increases with the density of network decreasing. Fig.8 shows the minimized energy consumption increases with the number of sensors N and radius of senor area D.

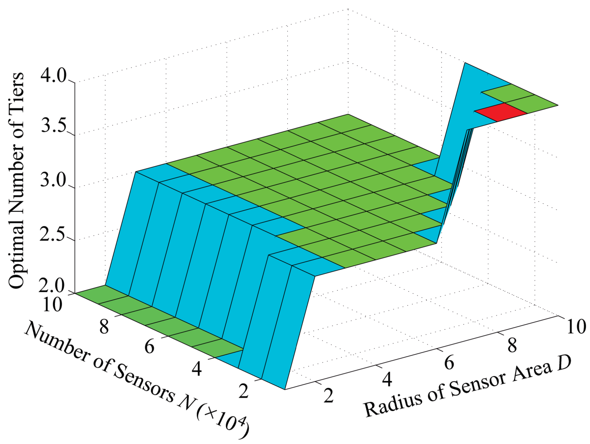


Fig.7. Optimal number of tiers $k_{\rm opt}$ versus the number of sensors N and the radius of sensors area D.

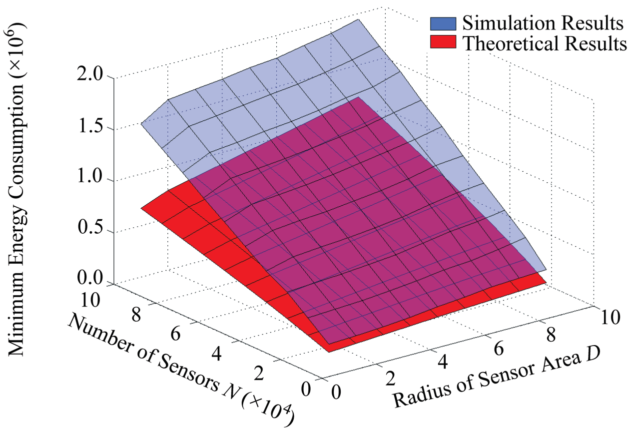


Fig.8. Minimum energy consumption versus the number of sensors N and the radius of sensor area D.

The simulation results in this section provide a useful insight into the energy consumption of the sensor network. First, our energy model has described the energy consumption of the network, as verified by the simulation results. Second, there exist optimal tiers which minimize the energy consumption of the multi-tier clustering network, and this solution can be found by our energy model. This is also verified in our simulation by adopting the multi-tier clustering algorithm proposed in Section 5.

7 Conclusions

In this paper, we have focused on finding the optimal tiers in the multi-tier clustered scheme to minimize the energy consumption for data collection in wireless sensor networks. In particular, we first proposed an energy consumption model for the multi-tier clustered network. With distinction to the reported energy models, in this model we have accounted for the energy consumption of the CH rotations. We then proposed an approach to find out the optimal number of tiers to minimize the energy consumption by numerical computing.

We presented a distributed multi-tier clustering algorithm to generate the energy efficiency optimized multi-tier clustering scheme for wireless sensor networks. Simulation results validate the proposed energy model. The optimal solution of tier number can be worked out in a computing efficient way by our energy model as demonstrated by an example.

In this paper, we have assumed that the sensors are distributed according to a homogeneous Poisson process, but the local difference of the sensor's density also affects the energy consumption. Therefore, extending the multi-tier clustering algorithm to cover the heterogeneous clustering situation according to the local density will bring the algorithm to be more suitable for large-scale networks. This deserves further studies.

References

[1] Akyildiz I F, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: A survey. *Computer Networks*, 2002, 38(4): 393-422.

[2] Ye W, Heidemann J, Estrin D. An energy-efficient MAC protocol for wireless sensor networks. In *Proc. the 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, June 2002, Vol.3, pp.1567-1576.

[3] Dam T, Langendoen K. An adaptive energy-efficient MAC protocol for wireless sensor networks. In *Proc. the 1st international Conference on Embedded Networked Sensor Systems*, November 2003, pp.171-180.

[4] Lu G, Krishnamachari B, Raghavendra C. An adaptive energy efficient and low-latency MAC for data gathering in wireless sensor networks. In *Proc. the 18th International Parallel and Distributed Processing Symposium*, April 2004.

[5] Ye W, Heidemann J, Estrin D. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Trans. Networking*, June 2004, 12(3): 493-506.

[6] Rhee I, Warrier A, Aia M, Min J, Sichitiu M. Z-MAC: A hybrid MAC for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 2008, 16(3): 511-524.

[7] Bulusu N, Estrin D, Girod L, Heidemann J. Scalable coordination for wireless sensor networks: Self-configuring localization systems. In *Proc. the 6th International Symposium on Communication Theory and Applications*, July 2001, pp.1-6.

[8] Bulusu N, Heidemann J, Estrin D. Adaptive beacon placement. In *Proc. the 21st IEEE International Conference on Distributed Computing Systems*, April 2001, pp.489-498.

[9] Elson J, Girod L, Estrin D. Fine-grained network time synchronization using reference broadcasts. In *Proc. the 5th Symposium on Operating Systems Design and Implementation*, December 2002, pp.147-163.

[10] Elson J, Estrin D. Time synchronization for wireless sensor networks. In *Proc. the 15th International Parallel and Distributed Processing Symposium*, April 2001, pp.1965-1970.

[11] Syed A, Heidemann J. Time synchronization for high latency acoustic networks. In *Proc. the 25th IEEE Int. Conf. Computer Communications*, April 2006, pp.1-12.

[12] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks. In *Proc. the 33rd Annual Hawaii International Conference on System Sciences*, January 2000.

[13] AI-Karaki J, Kamal A. Routing techniques in wireless sensor networks: A survey. *IEEE Wireless Communications*, 2004, 11(6): 6-28.

[14] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proc. the 6th Annual Int. Conf. Mobile Computing and Networking*, August 2000, pp.56-67.

[15] Lindsey S, Raghavendra C. PEGASIS: Power-efficient gathering in sensor information systems. In *Proc. IEEE Aerospace Conference*, March 2002, Vol.3, pp.1125-1130.

[16] Ye F, Luo H, Cheng J, Lu S, Zhang L. A two-tier data dissemination model for large-scale wireless sensor networks. In *Proc. the 8th Annual International Conference on Mobile Computing and Networking*, Sept. 2002, pp.148-159.

[17] Bandyopadhyay S, Tian Q, Coyle E. Spatio-temporal sampling rates and energy efficiency in wireless sensor networks. *IEEE/ACM Transactions on Networking*, 2005, 13(6): 1339-1352.

[18] Akkaya K, Younis M. A survey on routing protocols for wireless sensor networks. *Ad Hoc Networks*, 2005, 3(3): 325-349.

[19] Bandyopadhyay S, Coyle E. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proc. the 22nd Annual Joint Conference of the IEEE Computer and Communications*, March 30-April 3, 2003, Vol.3, pp.1713-1723.

[20] Bandyopadhyay S, Coyle E. Minimizing communication costs in hierarchically clustered networks of wireless sensors. *Computer Networks*, 2004, 44(1): 1-16.

[21] Chen Y, Liestman A, Liu J. A hierarchical energy-efficient framework for data aggregation in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 2006, 55(3): 789-796.

[22] Krishnamachari B, Estrin D, Wicker S. The impact of data aggregation in wireless sensor networks. In *Proc. the 22nd International Conference on Distributed Computing Systems Workshops*, July 2002, pp.575-578.

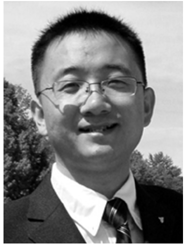
[23] Foss S, Zuyev S. On a Voronoi aggregative process related to a bivariate Poisson process. *Advances in Applied Probability*, 1996, 28(4): 965-981.

[24] Xue F, Kumar P. The number of neighbors needed for connectivity of wireless networks. *Wireless Networks*, 2004, 10(2): 169-181.

Jin-Tao Meng received his B.E. and M.S. degrees in computer science from the Department of Computer Science in Central China Normal University, Wuhan, in 2005 and 2008, respectively. He now is an engineer in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), and also a Ph.D. candidate in Institute of Computing Technology, CAS, Beijing. His research interest includes algorithm analysis, parallel and distributed computing, network protocol design, and bioinformatics.

Jian-Rui Yuan received her B.E. and M.S. degrees in electronic and communication engineering from the School of Information Science and Engineering in Central South University, Changsha, in 2009 and 2012 respectively. She now is an engineer in CYG SUNRI CO., LTD. Her research interest includes electricity networks, smart grid, sensor networks, and distributed computing.

Sheng-Zhong Feng received his B.E. degree from University of Science and Technology of China, Hefei, 1991 and his Ph.D. degree in computer science from Beijing Institute of Technology, in 1997. He now is the director of Institute of Advanced Computing and Digital Engineering in Shenzhen Institutes of Advanced Technology, CAS. His research interest includes high performance computing and bioinformatics.



Yan-Jie Wei received his B.S. degree in applied physics from Sichuan University, Chengdu, in 2004 and received his Ph.D. degree in computational biophysics from Michigan Technological University, USA, in 2007. Then he joined the CASL lab at Princeton University in March 2008 as a research associate. Since Aug. 2011 he has been working as an associate professor at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interest includes protein folding, protein structure prediction, gene assembly and gene classification, etc.