一种传感器网络最大化生命周期数据收集算法*

张 卿[1+], 谢志鹏[1], 凌 波[2], 孙未未[1], 施伯乐[1]

[1](复旦大学 计算机与信息技术系,上海 200433)
[2](中国浦东干部学院 信息技术部,上海 200233)

A Maximum Lifetime Data Gathering Algorithm for Wireless Sensor Networks

ZHANG Qing[1+], XIE Zhi-Peng[1], LING Bo[2], SUN Wei-Wei[1], SHI Bai-Le[1]

[1](Department of Computing and Information Technology, Fudan University, Shanghai 200433, China)
[2](Department of Information and Technology, China Executive Leadership Academy Pudong, Shanghai 200233, China)
+ Corresponding author: Phn: +86-21-55074881, E-mail: qzhang79@yahoo.com

Received 2004-11-11; Accepted 2005-03-10

Zhang Q, Xie ZP, Ling B, Sun WW, Shi BL. A maximum lifetime data gathering algorithm for wireless sensor networks. *Journal of Software*, 2005,16(11):1946−1957. DOI: 10.1360/jos161946

Abstract: This paper investigates the maximum lifetime data gathering problem theoretically. Specifically, (1) the simplified static routing scheme where only one routing tree is used to gather data during the lifetime of network is analyzed, (2) the actual dynamic routing scheme where a series of routing trees are used to gather data is analyzed, (3) a near optimal maximum lifetime data gathering and aggregation algorithm MLDGA is proposed, which tries to minimize the total energy consumption in each round and maximize the lifetime of a routing tree used in the round, (4) the MLDGA algorithm is simulated in Java programming language. Comparing with the existing algorithms that are only efficient in some specified conditions, the simulation results show that MLDGA performs well regardless of base station location and initial battery energy levels of sensors.

Key words: wireless sensor network; maximum lifetime; data gathering; data aggregation; MLDGA

摘 要: 从理论上分析了最大化网络生命周期的数据收集问题.主要做了以下 4 项工作: (1) 分析了简化的静态路由模式,其中只有一棵路由树用于收集数据. (2) 分析了真实的动态路由模式,其中有一系列的路由树用于收集数据. (3) 提出了一种近似最优的最大化网络生命周期的数据收集和聚集算法 MLDGA,MLDGA 一方面试图最小化每轮数据收集中所消耗的总能量,另一方面试图最大化每轮数据收集中所使用的路由树的生命周期. (4) 用 Java 语言实验模拟了 MLDGA 算法,并与现有的算法进行比较.实验结果表明,无论基站的位置还是传感

---

* Supported by the National Natural Science Foundation of China under Grant No.60503025 (国家自然科学基金); the Nature Science Foundation of Shanghai of China under Grant No.03ZR14014 (上海市自然科学基金)

ZHANG Qing was born in 1979. He holds a Ph.D. degree from Fudan University. His current research areas are wireless ad hoc network and wireless sensor network. **XIE Zhi-Peng** was born in 1976. He holds a Ph.D. degree from Hefei University of Technology. His current research areas are machine learning, bio-informatics and mobile computing. **LING Bo** was born in 1974. He holds a Ph.D. degree from Fudan University. His current research areas are peer-to-peer computing, information retrieval and economics. **SUN Wei-Wei** was born in 1973. He holds a Ph.D. degree from Fudan University. His current research areas are mobile database and XML. **SHI Bai-Le** was born in 1936. He is a professor at Fudan Unviersity and a CCF senior member. His research areas are database theory and application.

器的初始能量发生变化,MLDGA 都取得良好的性能,而现有的数据收集算法只适应于特定的变化.
关键词: 无线传感器网络;最大化生命周期;数据收集;数据聚集;MLDGA
中图法分类号: TP393 文献标识码: A

1 Introduction

A wireless sensor network is a special kind of ad-hoc network that consists of a large number of small, inexpensive, battery-powered sensors. The sensors can be densely deployed to monitor the environment and collect useful information on their surroundings. During the lifetime of network, the collected information is periodically gathered and transmitted to a base station for further processing. Sensor networks can be used in a wide range of applications from military to civil[1−3].

One of the main constraints on sensor networks is the limited battery energy of sensors. To keep the size of sensors small, a sensor can only be equipped with battery that stores at most 1 Joule of energy[4], and the battery is irreplaceable. For this reason, energy conservation should be the primary concern in data gathering problem. Its aim is to prolong the network lifetime to gather more data. Chang and Tassiulas[5,6] proposed routing algorithms to maximize system lifetime by treating the data transmission process as a maximum flow problem which meets the flow conservation principle. In another work, Bhardwaj *et al*.[7] derived upper bounds on the lifetime of a sensor network that collects data from a specified region.

Data aggregation has now emerged as a particularly useful paradigm for wireless routing in sensor networks to reduce the energy consumption[8,9]. The idea is to combine the data from different sensors enroute to eliminate redundant transmission. This paradigm greatly reduces the amount of data transmitted and thus saves energy. It shifts the focus from address-centric approaches (finding routes between pairs of end-nodes) to data-centric approaches (finding routes from multiple sources to a destination that allows in-network consolidation of redundant data)[10,11]. Several protocols, such as LEACH[12] and PEGASIS[13], used the data-centric approaches to minimize the energy consumed by sensors and increase the network lifetime accordingly. Further, Lindsey et al proposed PEDAP-PA[14], a tree based routing algorithm, which tries to improve the network lifetime. It achieves a good performance comparing with LEACH and PEGASIS in terms of lifetime. However, most of these works lack formal theoretical analysis and their algorithms are not well optimized.

In this paper, we investigate the maximum lifetime data gathering and aggregation problem theoretically. Tree structure is used as the basic routing scheme for our analysis since it is the minimal graph structure supporting the network connectivity. We analyze the static routing scheme first, where only one tree is used to gather data during the lifetime of network. Then we go into the dynamic routing scheme, where a series of trees are used to gather data. Based on the results of our analysis, we propose a new near optimal maximum lifetime data gathering and aggregation algorithm MLDGA. MLDGA tries to construct a maximum lifetime routing tree for each data gathering round while the tree is energy-efficient. The experimental results show that MLDGA succeeds in achieving longer lifetime and better network utility than several other existing algorithms in various conditions.

The rest of the paper is organized as follows. Section 2 reviews some related works. Section 3 presents the maximum lifetime data gathering problem. Section 4 gives some basic definitions used in this paper. Section 5 investigates the problem with two different routing schemes: the static one and the dynamic one. A near optimal algorithm MLDGA is also proposed. In Section 6, we conduct experiments to compare our algorithm with other known algorithms. We conclude our work in Section 7.

2 Related Work

Several efficient data gathering routing algorithms have been proposed in recent years. We can divide them into three categories approximately: cluster-based routing, chain-based routing, and tree-based routing.

LEACH is a cluster-based distributed routing protocol[12]. In LEACH, each node elects itself as cluster-heads with some probability. The remaining nodes join a cluster that requires minimum communication energy. In the data gathering process, each cluster-head collects data from sensors in its cluster, fuses the data, and then sends the result to the base station. LEACH utilizes the randomized rotation of cluster-heads to evenly distribute the energy load among sensors in the network. Simulation results show that LEACH achieves as much as a factor of 8 reduction in energy dissipation comparing with direct transmission[12]. As an improved version to LEACH, LEACH-C[15] uses a centralized clustering algorithm to produce better clusters, thus achieves better performance.

In PEGASIS[13], sensors are formed by chain. Each sensor communicates only with a close neighbor, and takes turns transmitting to the base station to prevent the failure of network. Only one node is designated to communicate with the base station, consequently the energy dissipation is significantly reduced. PEGASIS achieves better lifetime than LEACH about 100 to 200%[13].

Tan *et al*. proposed two tree based protocols PEDAP and PEDAP-PA[14]. They tried to compute a minimum spanning tree over the sensor network. In PEDAP, the weights of tree edges are the transmission cost between two connected sensors. In PEDAP-PA, the weight of tree edges is the ratio of the transmission cost between two connected nodes to the remaining energy of the sending node. The basic idea is to minimize the total energy expended in a round of communication while balance the energy consumption among sensors. PEDAP prolongs the lifetime of the last node death while PEDAP-PA provides a good lifetime for the first node death. Simulation results show that these two algorithms perform better than LEACH and PEGASIS both in systems that the base station is far away from and inside the field[14].

3 Problem Statement

We consider a wireless sensor network consisting of a group of sensors and a base station that are randomly distributed over a region. The locations of sensors and the base station are fixed. The base station knows the locations of all sensors apriori, which can be obtained by manually entering coordinates or by using GPS-equipped sensors. A sensor can transmit data to any other sensor, and can communicate directly with the base station. The sensors periodically monitor their vicinity and generate monitoring data. The data from sensors are gathered at each time unit and sent to the base station for further processing. The time unit is called *round*. We assume that each sensor has limited battery energy and the action of transmitting or receiving data will consume its battery energy. Our problem is to find a routing scheme to deliver data from all sensors to the base station, which can maximize the lifetime of the sensor network. In the process, data aggregation can be used to reduce the number of messages in the network.

Two points should be pointed out here:

First, the energy model used in our work is based on the first order radio model[12]. In this model, the radio dissipates E_{elec}=50nJ/bit to run the transmitter or receiver circuitry, and E_{amp}=100pJ/bit/m^2 for the transmit amplifier. Therefore, the energy expended to transit a k-bit message to a distance d is

$$E_{Tx}(k,d)=E_{elec}\times k+E_{amp}\times k\times d^2 \tag{1}$$

While the energy expended to receive this message is

$$E_{Rx}(k)=E_{elec}\times k \tag{2}$$

which is a constant for a fixed-size message.

Secondly, we treat the time when the first node dies as the lifetime of the network. Tan *et al*.[14] mentioned two different definitions for lifetime: In applications that the cooperation of all sensors working together is important, lifetime is defined as the time when the first node is drained of its energy; In applications that adjacent sensors record identical data, lifetime is defined as the time when a specified percentage of the sensors die. In this paper, we focus on the applications that need the cooperation of all sensors, and we adopt the first type of definition. Put it another way, we try to maximize the time when the first node dies.

4 Definition

Before diving into the maximum lifetime data gathering problem, we give some fundamental definitions and notations used throughout this paper.

We view a sensor network as a directed graph $G(N, A)$, where N is a set of all sensors and the base station in network. Each sensor is labeled with a node $ID\in\{1,2,\ldots,|N|\}$, the base station is labeled with ID 0; A is a set of directed edges connecting two sensors, i.e. $A=\{(i, j)\}$.

We denote the weight of edge $(i, j)\in A$ as w_{ij} ($w_{ij}\neq w_{ji}$).

Definition 1. For a sensor network $G(N, A)$, $T=(N, A')$ is a rooted directed spanning tree of G, where A' is a subset of A(i.e. $A'\subset A$) and T is rooted at the base station (i.e. root[T] = 0). We call T a routing tree for network G.

The tree structure can be used as a basic routing scheme for data gathering in sensor networks. A spanning tree is the minimal graph structure supporting the network connectivity. For each sensor i, we can find a directed path from i to the base station by using the tree structure, and the path does not contain a cycle. Therefore, we base our discussion on the tree structure in the following paper. LEACH (a cluster-based routing scheme) and PEGASIS (a chain-based routing scheme) described in section 2 can also be viewed as a special case of the tree structure based routing scheme.

Definition 2. For a sensor network G, if the underlying routing structure does not change over time and only one routing tree T is used to gather data during the lifetime of network, we say that G employs static routing scheme. If the underlying routing structure of G changes over time and a series of routing trees are used to gather data, we say that G employs dynamic routing scheme. The series of routing trees is denoted as $\{T(t)\}$, which means that $T(t)$ is used as the routing tree at round t.

Here, we also introduce some notations used in the following paper. For a sensor i in routing tree T, the transmit power of i is denoted as $E_{Tx}^{i}(T)$, the number of sensors sending data to i is denoted as $n_i(T)$. Then the received power of sensor i is $n_i(T)\times E_{Rx}$. The tree with minimum total energy consumption is denoted as T^0,

$$T^{o}=\underset{T\subset G(N,A)}{\arg\min}\sum_{i=1}^{|N|}(E_{Tx}^{i}(T)+n_i(T)\times E_{Rx}) \tag{3}$$

The residual battery energy level of a sensor i at round t is denoted as $R_i(t)$. If a network G employs dynamic routing scheme, the residual battery energy level of a sensor i after t rounds is $R_i(t)=R_i(0)-\sum_{\tau=1}^{t}(E_{Tx}^{i}(\tau)+n_i(\tau)\times E_{Rx})$.

Definition 3. We define the lifetime of a sensor i as the duration of the time when i is alive, and denote it as l_i.

If a network G employs static routing scheme and the routing tree for G is T, the lifetime of a sensor i in T is $l_i=\dfrac{R_i(0)}{E_{Tx}^{i}(T)+n_i(T)\times E_{Rx}}$. It is evident that the lifetime of a sensor i is related to the sensor which i sends data to and the number of sensors which send data to i.

Definition 4. We define the lifetime of an edge (i, j) in tree T as the duration of the time when i and j are alive, and denote it as l_{ij}. For an edge (i, j) in tree T, we have $l_{ij}=\min\{l_i,l_j\}$.

If a network G employs static routing scheme and the routing tree for G is T, the lifetime of G is

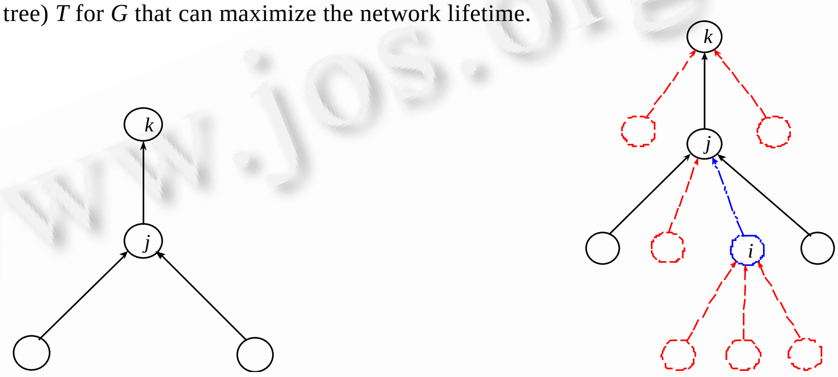
$L(G)=L(T)=\min_{i\in N}\{l_i\}=\min_{(i,j)\in A'(T)}\{l_{ij}\}$, where $A'(T)$ is the edge set induced by T. It means that the sensor that has minimum lifetime, which in turn depends on the edge that has minimum lifetime, determines the lifetime of G.

5 Maximum Lifetime Data Gathering: Static Scheme and Dynamic Scheme

In this section, we investigate the maximum lifetime data gathering and aggregation problem in sensor networks theoretically. We first examine the case when a sensor network employs static routing scheme, then we go into the dynamic routing scheme. The dynamic scheme corresponds to the actual scenario of our problem. In the dynamic scheme, we propose an algorithm MLDGA for the maximum lifetime data gathering problem. The implementation details of MLDGA are also discussed.

5.1 Static scheme

We first explore the simpler case when a sensor network G employs static routing scheme. The task is to find a routing tree (static tree) T for G that can maximize the network lifetime.



(a) A sub-routing-tree T_0 (b) A new link (i, j) is built between i and j

Fig.1 An example which shows difficulty of the problem

Definition 5. The optimum network lifetime L^* of a sensor network G employing static routing scheme is defined as:

$$L^*=\max_{T\subset G(N,A)}\{L(T)\}=\max_{T\subset G(N,A)}\left\{\min_{(i,j)\subset A'(T)}\{l_{ij}\}\right\} \tag{4}$$

It is evident that this problem can be viewed as a max-min bottleneck optimization problem[9]. The bottleneck edge is the one with the minimum lifetime in routing tree $T\subset G(N, A)$, which in fact determines the lifetime of network. Let the weight of edge be the lifetime of edge, i.e. $w_{ij}=l_{ij}$. The peculiarity of our problem is that the edge weight w_{ij} cannot be determined beforehand. Figure 1 shows an example. A sub-routing-tree T_0 for G has been built. j is a sensor in T_0, and i is a sensor not in T_0. Now, a new link is built between i and j. We cannot determine the weight of edge (i, j) at this time because the edge weight w_{ij} should be

$$\begin{aligned} w_{ij} &= \min\{l_i, l_j\} \\ &= \min\left\{\frac{R_i(0)}{E^i_{Tx}(T)(k,d_{ij})+n_i(T)\times E_{Rx}(k)}, \frac{R_j(0)}{E^j_{Tx}(k,d_{jk})+n_j(T)\times E_{Rx}(k)}\right\} \end{aligned} \tag{5}$$

where T is the final routing tree for G. The final number of sensors sending data to i or j (i.e. $n_i(T)$ or $n_j(T)$) cannot be determined at this time. Therefore, we cannot get the lifetime of sensor i or j when the final tree T has not been completely constructed, to say nothing of w_{ij}. It is just this uncertainty makes the difficulty of our problem.

Here, we present a heuristic greedy strategy to solve this max-min bottleneck optimization problem. We assume that T_0 is a sub-routing-tree that has already been built for G, j is a sensor in T_0, and i is a sensor not in T_0. If a new link will be built between *i* and *j*, we make

$$l_i' = \frac{R_i(0)}{E_{Tx}^i(k, d_{ij})} \tag{6}$$

as an estimation of the lifetime of *i*, make

$$l_j' = \frac{R_j(0)}{E_{Tx}^j(T_0)(k, d_{jk}) + (n_j(T_0) + 1) \times E_{Rx}(k)} \tag{7}$$

as an estimation of the lifetime of *j*, where $n_j(T_0)$ is the current number of sensors sending data to *j*, and make $w_{ij}' = \min\{l_i', l_j'\}$ as the estimation of w_{ij}. We select the edge (i_0, j_0) that has the maximum weight in all (i, j) pairs and add the edge into T_0. The process continues until all sensors have been added to T_0, then the static routing tree for *G* is constructed.

5.2 Dynamic scheme

In this subsection, we investigate the case when a sensor network *G* employs dynamic routing scheme. The task is to find a series of routing trees (dynamic trees) $\{T(t)\}$ for *G* which can maximize the network lifetime.

Definition 6. For a sensor network *G* employing dynamic routing scheme, let the series of routing trees used by *G* be $\Gamma=\{T(t)\}$. Then the network lifetime of *G* is

$$L(\Gamma) = L(T(1), T(2), T(3), \ldots) = \min_{i \in N}\{l_i\} \tag{8}$$

where l_i is the maximum value which satisfies $\sum_{\tau=1}^{l_i}\left(E_{Tx}^i(\mathrm{T}(\tau)) + n_i(\mathrm{T}(\tau)) \times E_{Rx}\right) \le R_i(0)$. The optimum network lifetime L^o of *G* is defined as

$$L^o = \max_{\Gamma \in \Gamma^*}\{L(\Gamma)\} \tag{9}$$

where Γ^* is the set of all possible routing tree series for *G*.

We have following theorem for the optimum network lifetime L^o.

Theorem 1. For a sensor network *G* employing dynamic routing scheme, its optimal network lifetime L^o is upper bounded by $L_U^o = \frac{\sum_{i=1}^{|N|} R_i(0)}{\sum_{i=1}^{|N|}(E_{Tx}^i(T^o) + n_i(T^o) \times E_{Rx})}$, i.e. $L^o \le L_U^o$.

Proof. We cannot spend more energy than the total energy of *G*, which is $\sum_{i=1}^{|N|} R_i(0)$. In every round, at least $\sum_{i=1}^{|N|}(E_{Tx}^i(T^o) + n_i(T^o) \times E_{Rx})$ amount of energy should be spent no matter what kind of tree we use. Therefore, the network lifetime cannot exceed L_U^o.

Based on Theorem 1 and the static scheme discussed in Section 5.1, we present two strategies to deal with the maximum lifetime data gathering problem when *G* employs dynamic routing scheme. (1) Try to minimize the total energy consumption of a routing tree in each round. From Theorem 1, we know that if the total energy consumed in a round is minimized, we will have more energy to be used in the following data gathering process, thus the lifetime of the network can be increased. (2) Try to maximize the lifetime of a routing tree constructed in a round. The dynamic trees can be viewed as a series of static trees (a static tree in a round). By maximizing the lifetime of a routing tree used in a round, we will have a cumulative effect to improve the lifetime of the network.

Integrating the two strategies discussed above, we propose a new maximum lifetime data gathering and aggregation algorithm MLDGA. In each round, MLDGA constructs a routing tree for this round. The tree construction procedure is described as follows:

(1) Initialize variables. The procedure maintains a tree structure T_0 that represents the sub-routing-tree that has been built for *G* at a round. T_0 is initialized with one element, the base station. The procedure also maintains two sets: the selected sensors set *S* which contains the sensors that have already been selected as part of T_0 (includes the base station), and the unselected sensors set *U* which contains the sensors that have not been inserted into T_0

($U=N-S$). S is initialized to the base station, and U is initialized to all the sensors.

(2) Estimate the edge weight. For each node i in U and each node j in S, we make the estimation of edge weight w_{ij} at round t as

$$\begin{aligned} w'_{ij} &= \frac{\min\{l'_i(t), l'_j(t)\}}{E^i_{Tx}(k, d_{ij})} \\ &= \frac{1}{E^i_{Tx}(k, d_{ij})} \times \min\left\{ \frac{R_i(t)}{E^i_{Tx}(k, d_{ij})}, \frac{R_j(t)}{E^j_{Tx}(T_0) + (n_j(T)+1) \times E_{Rx}} \right\} \end{aligned} \quad (10)$$

If only one element (the base station) in T_0, we treat the lifetime of the base station as infinite, then

$$w'_{ij} = \frac{\min\{l'_i(t), \infty\}}{E^i_{Tx}(k, d_{ij})} = \frac{1}{E^i_{Tx}(k, d_{ij})} \times \frac{R_i(t)}{E^i_{Tx}(k, d_{ij})} \quad (11)$$

(3) Add the edge with maximum weight to T_0. Set S and U are changed correspondingly.

(4) Repeat (2) and (3) until all sensors in U have been added to S. The routing tree T for G in current round is constructed ($T_0=T$).

From the definition of edge weight, we can get following two points that fit our strategies about the maximum lifetime data gathering problem:

- If the transmission cost between the sending sensor i and the receiving sensor j is low, the edge weight w_{ij} tends to be high. Thus the sensor i has high possibility to be added to T_0 and connects to j. It meets the strategy to minimize the total energy consumption in each round;
- If the lifetime of edge (i,j) is high, the edge weight w_{ij} tends to be high. Then edge (i,j) has high possibility to be added to T_0. It meets the strategy to maximize the lifetime of a routing tree constructed in a round.

Here, we compare MLDGA algorithm with other existing algorithms through analysis. In LEACH[12] , each sensor sends data to the nearest cluster head, then the cluster heads send the fused data to the base station. If the cluster heads are far from the base station, it will cost a lot of energy since there are several cluster heads that should send data to the base station. PEGASIS notices this point. In PEGASIS[13], each sensor sends data to a close neighbor, and only one sensor sends the fused data to the base station. PEGASIS saves great energy compared with LEACH since only one sensor sends data to the base station. However, if the base station is inside the field, both LEACH and PEGASIS perform poor. The reason is that both of them do not take the exact cost of sending data to base station and the total energy consumed per-round into account[14]. In PEDAP-PA[14], the authors used Prim's algorithm to construct a minimal spanning tree. The edge weight in PEDAP-PA relates to the transmission cost and the remaining energy of sending node. However, PEDAP-PA only considers the effect of the sending node, and the edge weight keeps unchanged during the construction of routing tree at a round. In MLDGA, we tries to construct a maximum lifetime routing tree for each data gathering round. Our definition of edge weight comes from the analysis of a network employing static routing scheme. It considers not only the transmission cost of sending node, but also the combined effect of sending node and receiving node, which can be got from the definition of edge lifetime ($l_{ij}=\min\{l_i, l_j\}$). We think our definition of the edge weight better reflects the essence of the maximum lifetime data gathering problem, and can achieve a longer lifetime comparing with other existing algorithms, which is confirmed by experiments.

5.3 Implementation

The implementation of MLDGA algorithm is similar to LEACH[12] and PEDAP-PA[14], where each round begins with a set-up phase, then followed by a steady-state phase.

The set-up phase constructs a routing tree for data gathering in the steady-state phase. We assume that the location and initial energy of all sensors is known by the base station apriori. The base station can compute a maximum lifetime routing tree according to the location and initial energy of all sensors by using MLDGA algorithm. At the same time, it can also estimate the remaining energy of all sensors since it knows how much energy a sensor expends in a round, which will be used in the computation of next round. After the computation, the base station compares the tree constructed in this round with the tree constructed in previous round. If these two trees are different, the base station broadcasts the necessary information for the sensors in the network. For each sensor *i*, the information contains the parent sensor of *i* in the tree, the time slot when the sensor sends its data to its parent, and the time slot when the sensor receives its children's data etc.

In the steady-state phase, sensors transfer their collected data to the base station along the routing tree constructed in the set-up phase. Beginning from the leaf nodes, each node sends its fused data to its parent. To prevent the collision of messages from children with the same parent, the parent node applies TDMA multiple access scheme among its children. To prevent the collision of messages from children belonging to several different parents, each parent node uses different CDMA codes and the children of a parent send their messages with its parent CDMA code. The process continues until the base station is arrived, then the data gathering process in this round is over.

6 Experiments

In this section, we present the performance analysis of MLDGA by using simulation programs written in Java programming language. We compared MLDGA with seven other different data gathering algorithms: Direct, MTE, LEACH, LEACH-C, PEGASIS, PEDAP and PEDAP-PA. In Direct[5], each sensor sends its data directly to the base station. In MTE[5], data traverses along the minimum energy consumption path to the base station.

To measure the performance of these algorithms, we consider three metrics. (1) The round when the first node dies (RFND). We consider it in the first place because it meets our definition of network lifetime. (2) The round when the last node dies (RLND). Besides trying to maximize the time when the first node dies, we also expect to prolong the time when the last node dies. (3) The ratio of the round when the first node dies to the round when the last node dies. We call the ratio as *network utility*. If the network utility is high, it means that the network is fully utilitized at most of its living time.

Table 1 summarizes the parameters used in our experiments. 100 sensors were randomly distributed over a region of 50m×50m (Fig.2) and 100m×100m. Each sensor had an initial energy level of 0.5J or uniform distribution between 0.2J and 0.8J. The base station was located far away from the field (at point (0, –100)) or in the center of the field (at point (25,25)). In each round, a sensor sent a 2000-bit message to the base station. The base station computed a routing scheme to deliver data in the set-up phase of each round. The energy model is based on the first

Table 1 Parameters and their values

Parameter	Value
Number of sensors	100
Network size	50m×50m, 100m×100m
Initial energy level	0.5J, 0.2-0.8J
The location of base station	(0,–100), (25, 25)
Message size	2000-bit
Energy model	First order radio model
E_{elec}	50nJ/bit
E_{amp}	100pJ/bit/m^2

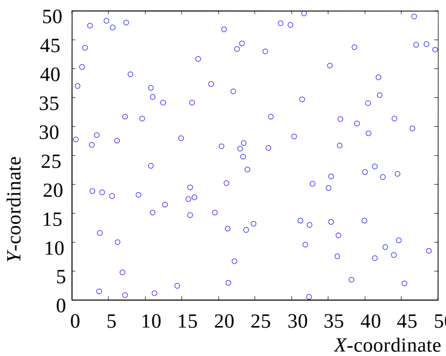


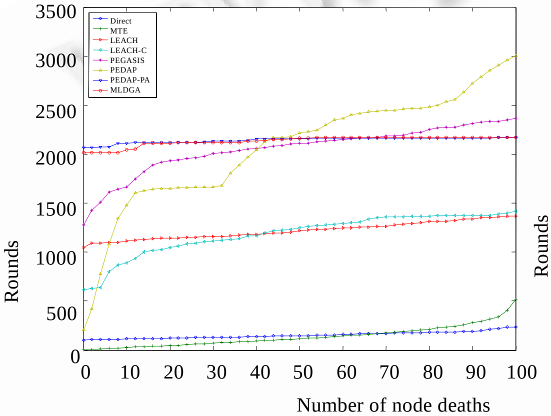
Fig.2 100-Node random sensor network

order radio model (E_{elec}=50nJ/bit and E_{amp}=100pJ/bit/m^2).

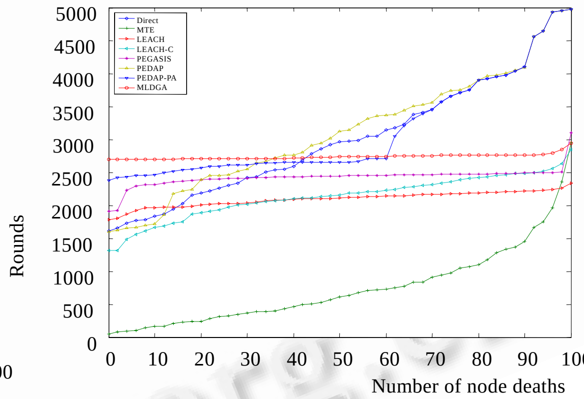
6.1 Effect of base station location

We investigate the effect of base station location first. The authors of PEDAP-PA[14] mentioned that PEDAP-PA performs well both in systems that the base station is far away from the field and the base station is in the center of the field, while LEACH and PEGASIS perform poor when the base station is inside the field since they do not take the cost of sending data to base station into account. Following this observation, we examined the effect of base station location on our algorithm first.

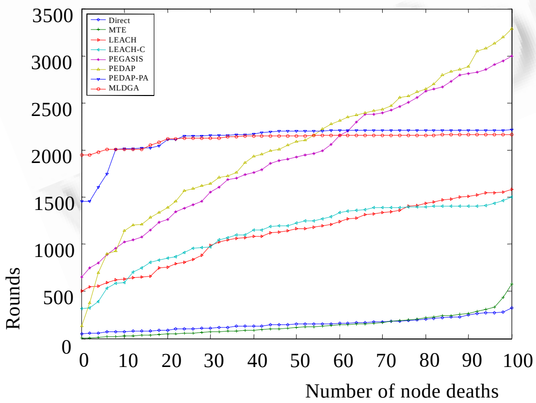
Figure 3(a) presents the situation that the base station was located far away from the field and all sensors had equal initial energy levels (F-EE). It can be observed from Fig.3(a) that both MLDGA and PEDAP-PA achieve a good RFND comparing with other algorithms. The lifetime of Direct and MTE is far from optimal. PEDAP provides a good RLND but has a bad RFND, since it constructs a minimum energy consuming routing tree for each round of communication but pays no attention to balance the load among sensors. LEACH-C uses a centralized cluster formation algorithm to minimize the total energy spent by the non-cluster-head nodes. It achieves a better RLND comparing with LEACH. PEGASIS provides both a good RFND and a good RLND. However, MLDGA and PEDAP-PA further improves the RFND about 60% comparing with PEGASIS, and about 100% comparing with LEACH.



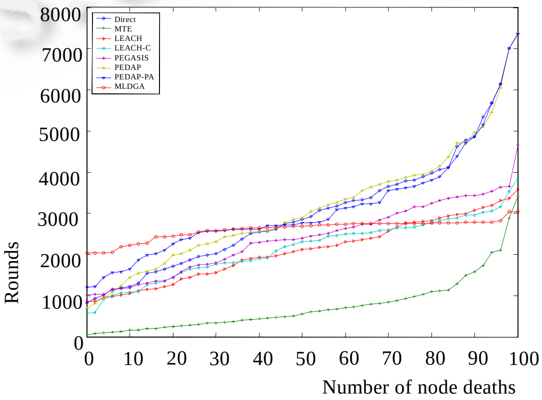
(a) F-EE



(b) C-EE



(c) F-UE



(d) C-UE

Fig.3 Timings of node deaths in 4 situations

Figure 3(b) presents the situation that the base station was located in the center of the field and the sensors had equal initial energy levels (C-EE). It can be seen from Fig.3(b) that MLDGA also achieves a good RFND. The RFNDs of MLDGA, PEDAP-PA, PEDAP, LEACH and PEGASIS are 2698, 2375, 1603, 1782 and 1909 respectively. MLDGA is the best one which achieves the longest RFND in all 8 algorithms. It improves the RFND about 40% comparing with PEGASIS, and about 50% comparing with LEACH.

From Fig.3(a) and (b), we can conclude that MLDGA has a good RFND (i.e. lifetime) regardless of the base station location. It distributes the load evenly among sensors and tries to minimize the total energy consumed in a round, so it achieves a good RFND and a RLND near to RFND.

6.2 Effect of initial energy level

To check the ability of balancing the load among sensors further, we investigate the situations when the sensors have unequal initial energy levels. We distributed the initial energy levels of all sensors between 0.2J and 0.8J uniformly. The average energy level of a sensor is 0.5J, which is the same as energy level in Fig.3(a) and (b).

Figure 3(c) presents the situation that the base station was located far away from the field and the sensors had unequal initial energy levels (F-UE). Comparing with Fig.3(a), we see from Figures 3(c) that the RFNDs of all algorithms decrease significantly except MLDGA. It means that MLDGA is robust to balance the energy consumption among sensors even the initial energy levels of all sensors are different greatly. MLDGA improves the RFND about 35% comparing with PEDAP-PA, about 200% comparing with PEGASIS, and about 290% comparing with LEACH.

Figure 3(d) presents the situation that the base station was located in the center of the field and the sensors had unequal initial energy levels (C-UE). Again, MLDGA is the best one which achieves the longest RFND in all algorithms. It improves the RFND about 70% comparing with PEDAP-PA, and about 100% comparing with PEGASIS, and about 140% comparing with LEACH.

From Figure 3(c) and 3(d), we can conclude that MLDGA has a good RFND (i.e. lifetime) regardless of the initial energy levels of all sensors. It achieves the longest RFND comparing with other existing algorithms. Its abilities to distribute the load evenly among sensors and to minimize the total energy consumed in a round are further confirmed since it achieves a good RFND and a RLND near to RFND even when all sensors have unequal initial energy levels.

6.3 Network utility performance

Figure 4 shows the network utility of all these 8 algorithms in 4 different situations. The network utility of MLDGA is the highest in all 4 situations (except a little lower than PEDAP-PA in F-EE). It reaches 90% in F-EE, C-EE and F-UE, and near 70% in C-UE. It means that the network is fully utilized at most of its living time by using MLDGA. This feature is important for applications that need the cooperation of all sensors working together, since the quality of the system will decrease dramatically after the first node dies. From another aspect, the values of network utility imply MLDGA distributes the load evenly among sensors (The energy levels of all sensors are near to zero when the first node dies).

Table 2 summarizes the results for 4 situations (F-EE, C-EE, F-UE and C-UE) when the network size is 100m×100m (where NU represents network utility). Compared with the setting that the network size is 50m×50m, the distance between two sensors tends to be longer. RLND and RLND usually decrease as a result of the increase in cost of sending data. We can still find that MLDGA performs well in all four situations, which is consistent with the results getting from the setting that the network size is 50m×50m. The outstanding performance of MLDGA in C-UE especially shows its abilities to balance the load evenly among sensors since all sensors have unequal initial energy levels and the sensors are close to base station.

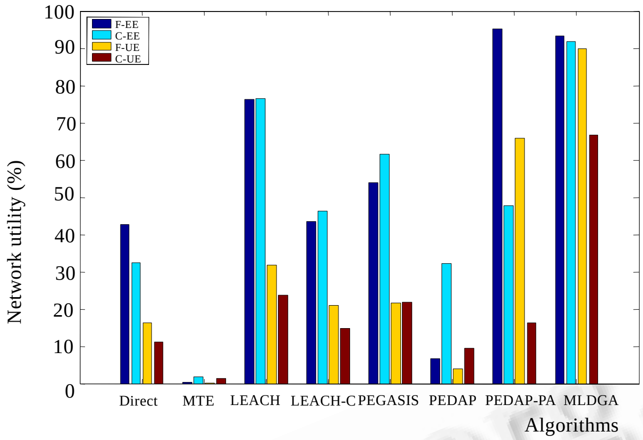


Fig.4 Network utility

Table 2 Performance of algorithms in 4 situations when network size is 100m×100m

Protocol	F-EE			C-EE			F-UE			C-UE		
	RFND	RLND	NU (%)	RFND	RLND	NU (%)	RFND	RLND	NU (%)	RFND	RLND	NU (%)
Direct	55	231	24	470	4647	10	26	272	10	229	6258	4
MTE	2	352	1	52	1607	3	2	337	1	56	1690	3
LEACH	558	946	59	861	1655	52	293	1108	26	438	2685	16
LEACHC	472	956	50	710	2062	34	337	1022	33	452	3248	14
PEGASIS	583	2067	28	1007	2589	39	183	2456	8	954	3412	28

7 Conclusions

In this paper, we investigated the maximum lifetime data gathering and aggregation problem in wireless sensor networks. Two types of routing scheme were discussed: the static routing scheme and the dynamic routing scheme. In the static routing scheme, only one tree is used to gather data during the lifetime of the network, which can be viewed as a max-min bottleneck optimization problem. We presented a heuristic greedy strategy to deal with this problem. The strategy makes an estimation of the weight edge and selects the maximum weight edge in each iteration to construct a routing tree for network. In the dynamic routing scheme, a series of trees are used to gather data. We proposed a near optimal algorithm MLDGA to construct the series of routing trees. In MLDGA, the edge weight is estimated as the ratio of the edge lifetime to the remaining energy of sending node, which meets the requirement to construct an energy-efficient maximum lifetime routing tree for each data gathering round. We conducted experiments to evaluate the performance of MLDGA. The experimental results show that MLDGA achieves a good lifetime performance and a good network utility performance regardless of the base station location and the initial battery levels of sensors.

References:

[1] Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: a survey. Computer Networks, 2002,38(4): 393–422.

[2] Li JZ, Li JB, Shi SF. Concepts, issues and advance of sensor networks and data management of sensor networks. Journal of Software, 2003,14(10):1717–1727 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/1717.htm

[3] Ren FY, Huang HL, Lin C. Wireless sensor networks. Journal of Software, 2003,14(7):1282–1291 (in Chinese with English abstract) . http://www.jos.org.cn/1000-9825/14/1282.htm

[4] Kahn JM, Katz RH, Pister KSJ. Next century challenges: Mobile networking for smart dust. In: Proc. of the 5th Annual ACM/IEEE Int'l Conf. on Mobile Computing and Networking. Seattle: IEEE Computer Society, 1999. 263−270.

[5] Chang JH, Tassiulas L. Maximum lifetime routing in wireless sensor networks. IEEE/ACM Trans. on Networking, 2004,12(4): 609−619.

[6] Chang JH, Tassiulas L. Energy conserving routing in wireless ad-hoc networks. In: Proc. of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 22−31.

[7] Bhardwaj M, Chandrakasan A, Garnett T. Upper bounds on the lifetime of sensor networks. In: IEEE Int'l Conf. on Communications. Helsinki: IEEE Computer Society, 2001. 785−790.

[8] Considine J, Li F, Kollios G, Byers J. Approximate aggregation techniques for sensor databases. In: Proc. of the Int'l Conf. on Data Engineering. Boston: IEEE Computer Society, 2004. 449−460.

[9] Kang I, Poovendran R. Maximizing static network lifetime of wireless broadcast ad-hoc networks. In: Proc. of the IEEE Int'l Conf. on Communications. Alska: IEEE Computer Society, 2003. 2256−2261.

[10] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: Proc. of the ACM/IEEE Int'l Conf. on Mobile Computing and Networks. Boston: ACM Press, 2000. 56−67.

[11] Krishnamachari B, Estrin D, Wicker S. The impact of data aggregation in wireless sensor networks. In: Proc. of the 22nd Int'l Conf. on Distributed Computing Systems Workshops. Vienna: IEEE Computer Society, 2002. 575−578.

[12] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient communication protocol for wireless microsensor networks. In: Proc. of the 33rd Annual Hawaii Int'l Conf. on System Sciences. Maui: IEEE Computer Society, 2000. 3005−3014.

[13] Lindsey S, Raghavendra CS. Pegasis: Power-Efficient gathering in sensor information systems. In: Proc. of the IEEE Aerospace Conf. Montana: IEEE Computer Society, 2002. 1−6.

[14] Tan HO, Korpeoglu I. Power efficient data gathering and aggregation in wireless sensor networks. SIGMOD Record, 2003,32(4): 66−71.

[15] Heinzelman W, Chandrakasan A, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. on Wireless Communications, 2002,1(4):660−670.

附中文参考文献:

[2] 李建中,李金宝,石胜飞.传感器网络及其数据管理的概念、问题与进展.软件学报,2003,14(10):1717−1727. http://www.jos.org.cn/1000-9825/14/1717.htm

[3] 任丰原,黄海宁,林闯.无线传感器网络.软件学报,2003,14(7):1282−1291. http://www.jos.org.cn/1000-9825/14/1282.htm