Energy-Efficient Multiple Routing Trees for Aggregate Query Evaluation in Sensor Networks

Yuzhen Liu and Weifa Liang

Department of Computer Science, The Australian National University
Canberra, ACT 0200, Australia
{yliu,wliang}@cs.anu.edu.au

Abstract. In this paper we consider the problem of finding multiple routing trees in sensor networks for the evaluation of a class of aggregate queries including `AVG`, `MIN`, `MAX`, and `COUNT` with an objective to maximizing the network lifetime. Due to the NP hardness of the problem, we instead devise a heuristic algorithm for it. Unlike the previous work that focused on finding a single routing tree for query evaluation, we introduce the concept of multiple routing trees, and use these trees to evaluate aggregate queries, provided that different routing trees are used at different stages of the network lifetime. To evaluate the performance of the proposed algorithm, we conduct extensive experiments by simulation. The experimental results show that the proposed algorithm outperforms existing algorithms based on a single routing tree. We also prove that the approximation ratio of a known approximation algorithm for the identical energy case is a constant, and provide tighter lower and upper bounds on the optimal network lifetime for the non-identical energy case.

1 Introduction

In wireless sensor networks, each sensor periodically measures the physical environment around it and generates a stream of data. The processor within the sensor processes the data and relays the processed result to the other sensors. A wireless sensor network thus can be treated as a *sensor database* [1]. Several sensor database systems like TinyDB [2] have been proposed. Sensor databases allow users to pose queries to a special node (the base station) which then disseminates them over the network. In response to each query, each node evaluates its data against the query and transmits the matched data towards the origin of the query. As the matched data is routed through a routing tree consisting of all the nodes in the network, each relay node in the tree may apply one or more database operators (typically aggregation operators) on the data it received and/or sensed. Such data gathering query processing is referred to as *in-network processing*, which has been shown to be fundamental to achieve energy-efficient communication in data-rich, large-scale, yet energy-constrained sensor databases. The main constraint on the sensors is that they are equipped with energy-limited batteries, which limits the network lifetime and impairs the

J. Harju et al. (Eds.): WWIC 2008, LNCS 5031, pp. 201–212, 2008.

quality of the network. Therefore, energy conservation in sensor networks is of paramount importance. Energy-efficient routing trees built for in-network processing play a central role in such data gathering application [2,3,4,5].

Data gathering in sensor networks aiming to find a routing tree such that the total energy consumption is minimized has been extensively studied in literature [6,7,8,9]. Heinzelman *et al* [6] initialized the study of the problem by proposing a clustering protocol `LEACH`. The nodes in `LEACH` are grouped into a number of clusters in a self-organized way, where a clusterhead serves as a local 'base station' to aggregate data from its members and send the aggregated result to the base station directly. Lindsey and Raghavendra [7] studied the problem by providing an improved protocol `PEGASIS`, in which all the nodes in the network form a chain. One of the nodes in the chain is chosen as the head, and the head is responsible to report the aggregated result to the base station. In both [6] and [7], the measurement of energy consumption is not addressed explicitly. With the assumption that the transmission energy consumption at a node is proportional to the distance between the node and its parent, Tan and Körpeoğlu [8] studied the data gathering problem by proposing another protocol `PEDAP`, based on the above two solutions. `PEDAP` assigns weights to the communication links in the network and finds a minimum spanning tree rooted at the base station in terms of the total transmission energy consumption. Kalpakis *et al* [9] considered a generic data gathering problem with an objective to maximizing the network lifetime, and proposed an integer program solution and a heuristic solution. It should be noticed that all the data gathering methods mentioned above are based on the assumption that the length of the message transmitted by a relay node in a routing tree is independent of the lengths of its children messages, i.e. each node transmits the same volume of data to its parent no matter how much data it received from its children. We refer to this type of data gathering (aggregate) query as the *message-length independent aggregate query*, which is also called *fully aggregate query* in [10]. Such aggregate queries in databases include the popular operations `AVG, MIN, MAX, COUNT`, etc.

Unlike the previous work in [6,7,8,9,10] that either a dedicated routing tree is built for each individual query or a single shared routing tree is built for all the queries, we aim to build a series of routing trees for query evaluation such that the network lifetime is maximized, where network lifetime is referred to the time of the first node failure in the network [11]. Consider a sensor network with base station a, illustrated by Fig. 1(a). We assume that each node is assigned 1,500 units of initial energy. We further assume that each node only transmits a unit-length of data and the energy consumption in both transmitting and receiving a unit-length of data is one unit of energy for each aggregate query. The minimum degree spanning tree of Fig. 1(a) is shown in Fig. 1(b). If the optimal routing tree in Fig. 1(b) is used to evaluate aggregate queries until some node in the network runs out of energy, then the (optimal) maximum network lifetime is $\frac{1500}{1*4+1} = 300$. However, if the entire network lifetime is partitioned into several stages, then the network lifetime can be further prolonged, provided that a different routing tree is employed at each different stage. For the example in Fig. 1(a), we assume the network lifetime consists of two stages. We use another routing tree shown in Fig. 1(c) for the rest of aggregate queries after the current routing tree in Fig. 1(b) has been used for the first 240 aggregate queries, then the network lifetime is $240 + \frac{1500-240*(3+1)}{1*5+1} = 330$, which is longer than that delivered by using just a single spanning tree during the entire network lifetime.

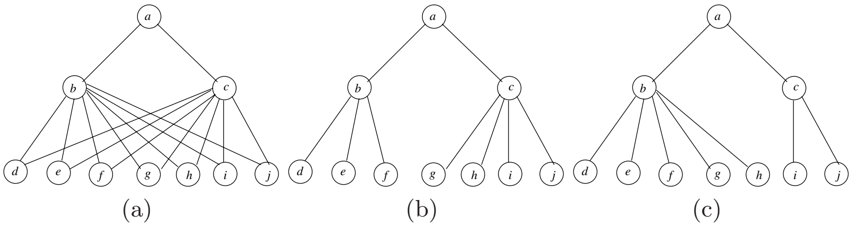


Fig. 1. (a)a sensor network (b)the optimal routing tree and (c)another routing tree

Motivated by the above example, we tradeoff the overhead of building routing trees and the difference of residual energies among the nodes, and introduce the concept of multiple routing trees. Instead of building a routing tree either for each query or for all queries, a routing tree just serves for a certain period of the network lifetime. Thus, a number of routing trees will be built for the evaluation of aggregate queries in sensor networks, and the energy overhead on building these trees can be reduced or alleviated, in comparison with that on building a dedicated routing tree for each individual query. On the other hand, we also balance the residual energies among the nodes by building another routing tree for the later queries, after a certain number of queries have been evaluated using the current routing tree.

2 Preliminaries

2.1 System Model

We consider a wireless senor network consisting of $n-1$ stationary homogeneous *sensor nodes* and a base station s distributed arbitrarily in a two dimensional region of interest. The network can be modelled by an undirected graph $G = (V, E)$, where V is the set of nodes with $|V| = n$ and there is an edge (u, v) in E if nodes u and v are within the transmission range of each other. The base station s is assumed to have unlimited energy supply. Each sensor node is equipped with an omni-directional antenna and can transmit messages with a fixed power level (its maximum power level). We assume that generating a unit-length of data consumes s_e amount of energy, transmitting a unit-length of data consumes t_e amount of energy, while receiving a unit-length of data takes r_e amount of energy. Clearly, $t_e \geq r_e$, and $s_e << \min\{t_e, r_e\}$. The transmission energy consumption of a sensor with transmission range R for transferring one unit-length data is R^2. Following the same assumption as in [10], there is time synchronization for evaluating aggregation queries. Time is divided into equal sized periods called *epochs*. At each epoch a leaf node transmits one unit of data to its parent. After a parent node has received the data from all its children, it aggregates the data and transmits the result to its parent in the next epoch.

2.2 Problem Definition

Given a wireless sensor network $G = (V, E)$, we assume that each sensor has a fixed transmission range and produces the sensed data as it monitors its vicinity periodically. Each sensor consumes the same amount of energy for one unit-length of data transmission, whereas its energy consumption of receiving data from its children is proportional to the number of its children. Specifically, we consider a class of message-length independent aggregate queries by finding multiple routing trees for them, with an objective to maximizing the network lifetime. For the sake of simplicity, we assume the message length transmitted by each node in the tree is one unit-length of data. The basic operation is to build one or multiple routing trees rooted at the base station and spanning all the other nodes. Each non-leaf node performs aggregation and forwards the partial aggregated result to its parent, the final query result will be available at the base station. It can be seen that the proposed algorithm can be easily to extended to the case where the message length transmitted by each node is an arbitrary l units of data with a minor modification, $l \geq 1$.

2.3 Algorithm for Finding a Single Routing Tree

We first briefly re-produce an approximation algorithm for finding such a routing tree that will be used for all aggregate queries during the entire network lifetime [10]. We then prove that the approximation ratio of the approximation algorithm for the identical energy case is a constant. We finally propose an approach to determine the lower bound and upper bound on the optimal network lifetime for the non-identical energy case.

Identical initial energy. If the initial energy at each node is identical, the network lifetime is determined by the maximum degree node in the routing tree, because the energy consumption of each node in the tree is determined by its transmission and reception energy consumptions, while the transmission energy consumption at each node is identical and the reception energy consumption of a node depends on how many children it has. Buragohain *et al* [10] concluded that finding a single optimal routing tree is equivalent to finding a Minimum Degree Spanning Tree (`MDST`) in the network, while the latter problem is to find a spanning tree such that the maximum node degree in the tree is minimized. The approximation ratio of the approximation algorithm given by Buragohain *et al* [10] is at least $\frac{1}{1+r_e}$, which depends on not only the transmission energy consumption ($t_e = 1$) but also the reception energy consumption r_e for a unit-length data. We now prove that the approximation ratio of their approximation algorithm is a constant by the following lemma.

Lemma 1. *Given a sensor network G in which each sensor has identical initial energy and transmission range, there is an approximation algorithm for the message-length independent aggregate query problem, and the network lifetime through the use of the routing tree delivered by the algorithm is at least 2/3 of the maximum (optimal) network lifetime.*

Proof. Let Δ^* be the maximum degree of the nodes in a minimum degree spanning tree in G and IE the initial energy at each sensor. Since $r_e \leq t_e$ and $\Delta^* \geq 2$, $\Delta^* + t_e/r_e \geq 3$. Following an algorithm due to Fürer and Raghavachari [12], referred to as MDST, there is an approximation solution for the minimum degree spanning tree problem, and the maximum node degree in the tree delivered by their algorithm is no more than $\Delta^* + 1$. Consequently, the maximum energy consumption of the maximum degree node in the routing tree is at most $r_e\Delta^* + t_e$ for each query, assuming one unit-length message will be transmitted, since the nodes in the tree have at most Δ^* children in the worst case. Let τ be the network lifetime delivered by the approximation algorithm MDST and τ_{opt} the maximum (optimal) network lifetime. We thus have

$$\frac{\tau}{\tau_{opt}} \geq \frac{\frac{IE}{r_e\Delta^*+t_e}}{\frac{IE}{r_e(\Delta^*-1)+t_e}} = \frac{r_e(\Delta^*-1)+t_e}{r_e\Delta^*+t_e} \geq 1 - \frac{1}{3} = 2/3.$$

Non-identical initial energy. It is reasonable to assume that the initial energy of each sensor is identical in the initial deployment of a sensor network. However, after a certain period of time, some sensors consume more energy than the others since they have been used as relay nodes to relay data for the others. As a result, the residual energies of different nodes will be different. This implies that the algorithm for the identical energy case is no longer applicable. For this case, Buragohain *et al* [10] proposed a novel reduction, which reduces the non-identical energy case to the identical energy case as follows.

Assume that the network lifetime τ is given in advance. Let $b(v)$ be the degree of node v in the tree. Then, the energy consumption at v for such an aggregate query is $r_e(b(v)-1)+t_e$ (assume a unit-length of data transferred by per node). Let $RE(v)$ be the residual energy at v at this moment. If the routing tree will be used for evaluating the rest of other aggregate queries, then, the maximum lifetime $\tau(v)$ of node v is bounded by $\tau(v) = \lfloor \frac{RE(v)}{r_e(b(v)-1)+t_e} \rfloor$. Obviously, $\tau \leq \min_{v \in V}\{\tau(v)\}$. Thus, the degree $b(v)$ of v in the routing tree is bounded by $b(v) = \lfloor \frac{RE(v)}{r_e\tau} - \frac{t_e}{r_e} + 1 \rfloor$.

Given the sensor network $G(V,E)$ and the network lifetime τ with $|V| = n$, the algorithm in [10] first calculates $b(v)$ for each node $v \in V$. It then constructs an auxiliary graph $G' = (V \cup V_1, E \cup E_1)$ that is defined as follows. For each node $v_i \in V$, add $n - b(v_i)$ new nodes $v_{i_1}, v_{i_2}, \cdots, v_{i_{n-b(v_i)}}$ into V_1 and add an edge between v_i and v_{i_j} into E_1, $1 \leq j \leq n - b(v_i)$. Thus, the degree of each node $v \in V$ in G' is n, while the degree of every newly added node is one. An approximate, minimum degree spanning tree in G' is found afterwards. It finally prunes those nodes and edges incident to the nodes from the tree if the nodes are not in V. The resulting tree, referred to as *the degree-constrained spanning tree*, will be used as the routing tree. We refer to this algorithm as NMDST. However, in reality τ is unknown in advance. Buragohain *et al* [10] instead proposed an approach to find the optimal network lifetime by using the binary search on an interval of reasonable size. We here propose an approach to determine a narrow interval in which the optimal network lifetime τ_{opt} falls by the following lemma.

Lemma 2. *Let τ_{low} and τ_{upp} be the lower and upper bounds of the optimal network lifetime τ_{opt}. Then, $\tau_{low} = \min_{v \in V}\{\frac{RE(v)}{r_e(d(v)-1)+t_e}\}$ and $\tau_{upp} = \frac{\max_{v \in V}\{RE(v)\}}{r_e(\Delta^*-1)+t_e}$.*

Proof. Since the degree of any node in a spanning tree is no greater than its physical degree in the network, the lower bound then follows. We now deal with the upper bound on τ_{opt}. Let $b(v)$ be the degree of node v in a minimum degree spanning tree and $b(v_0) = \Delta^*$ where Δ^* is the maximum degree of nodes in the minimum degree spanning tree, $v_0 \in V$. Then

$$\tau_{opt} \leq \min_{v \in V}\{\frac{RE(v)}{r_e(b(v)-1)+t_e}\} \leq \frac{RE(v_0)}{r_e(\Delta^*-1)+t_e} \leq \frac{max_{v \in V}\{RE(v)\}}{r_e(\Delta^*-1)+t_e}.$$

3 Heuristic Algorithm for Finding Multiple Routing Trees

In this section we propose a novel approach for the problem of concern to further prolong the network lifetime, if multiple rather than a single routing tree is employed during the different stages of the network lifetime. We start with two stages, and then consider K stages with a given $K (K \geq 2)$.

3.1 Finding Two Routing Trees

Assume that the initial energy at each node is identical, and the entire network lifetime consists of at most two stages. The proposed algorithm proceeds as follows.

In the first stage, an approximate, minimum degree spanning tree T_1 in G can be found using algorithm `MDST`, which will be used as the routing tree. After T_1 has been used for a certain period of time τ_1, a new spanning tree T_2 will be built, using algorithm `NMDST`, based on the current residual energy of each node. If the use of T_2 instead of T_1 will prolong the network lifetime further, the second stage proceeds, and T_2 will be used for evaluating the rest of aggregate queries. Otherwise, T_1 continues to be used until the end of network lifetime.

One fundamental issue related to this two-stage approach is to find the shifting time point τ_1, which is also the duration of the first stage. To find such a shifting time point, we check every τ' in the interval $[0, \tau]$ to see whether the inequality $\min_{v \in V}\{\frac{IE-[(d_{T_1}(v)-1)r_e+t_e]\tau'}{(d_{T_2}(v)-1)r_e+t_e}\} > \frac{IE}{(\Delta_{T_1}-1)r_e+t_e} - \tau'$ holds, where Δ_{T_1} is the maximum degree of nodes in T_1 and $d_{T_i}(v)$ is the node degree of v in T_i, $i = 1, 2$. If there is no such a τ', then, a single stage suffices. Otherwise, we select a τ'_0 that results in the longest network lifetime. In order to minimize the energy overhead on finding a shifting time point, we can use the binary search to find a shifting time point in the interval $[0, \tau]$. Thus, only $\log \tau$ routing trees are needed to be built.

3.2 Algorithm for Finding Multiple Routing Trees

In the following we propose an approach for finding $K \geq 2$ routing trees to prolong the network lifetime, which avoids the energy overhead on finding shifting time points of each stage as the above two-stage case. The idea is that the network lifetime is partitioned K stages. At each potential stage, the *quota of energy* assigned to each node is $1/K$ of the initial battery capacity. A routing tree in which each node is assigned at least the quota of energy will be used at each stage.

We now analyze the improvement of the network lifetime through the use of multiple routing trees in comparison to the use of a single routing tree. For simplicity, we assume that each query session only transfers a unit-length message, and the initial energy IE at each node is identical. Then, $\tau = \lfloor \frac{IE}{(\Delta-1)*r_e+t_e} \rfloor$, where Δ is the maximum degree of nodes in the approximate, minimum degree spanning tree. For convenience, in the rest of discussion we assume that the network lifetime is always an integral value and we ignore the floor of the computed value of the network lifetime. Clearly, the duration of the first stage, in which the approximate, minimum degree spanning tree T_1 in the network will be used, is $\tau_1 = \frac{\tau}{K}$, because the quota of energy assigned to each node v at this stage is $E_1(v) = IE/K$. After the first stage, a degree-constrained spanning tree T_i is constructed, using algorithm NMDST for each i of the remaining stages, assuming that the same quota of energy is assigned to each node at each stage, $i \geq 2$. If the duration of T_i is less than τ_1, T_1 will be used at stage i because the quota of energy of each node is at least IE/K and the duration of T_1 is at least τ_1. Otherwise, T_i should be used at stage i.

Let $b_i(v)$ be the degree of node v in the routing tree built at stage i and τ_i the duration of stage i. We have $\tau_i \geq \tau_1 \geq \frac{\tau}{K}$. Note that although the minimum quota of energy among the nodes at stage i is $\frac{IE}{K}$, the available energy at node v is actually $E_i(v) = \frac{IE}{K} + \Delta E_i(v)$, where $\Delta E_i(v)$ is the residual energy inherited from stage $(i-1)$, $2 \leq i \leq K$. Obviously, if $i = 2$, $\Delta E_2(v) = ((\Delta - b_1(v)) * r_e) * \tau_1$, which is the difference of energy consumption between the maximum degree node and node v in the routing tree T_1. Otherwise, $\Delta E_i(v) = E_{i-1}(v) - ((b_{i-1}(v) - 1) * r_e + t_e) * (1/K + \delta_{i-1})\tau$, assuming that the duration of stage i is $\tau_i = (1/K + \delta_i)\tau$, $0 \leq \delta_i < 1$. Suppose that $\frac{E_i(v_{i_0})}{(b_i(v_{i_0})-1)*r_e+t_e} = \min_{v \in V}\{\frac{E_i(v)}{(b_i(v)-1)*r_e+t_e}\}$, we have $(1/K + \delta_i) * \tau = \frac{E_i(v_{i_0})}{(b_i(v_{i_0})-1)*r_e+t_e}$. Then,

$$\begin{aligned}
\delta_i &= \frac{1}{\tau}\left(\frac{E_i(v_{i_0})}{(b_i(v_{i_0})-1)*r_e+t_e}\right) - \frac{1}{K} \\
&= \frac{(\Delta-1)*r_e+t_e}{IE} * \frac{IE/K+\Delta E_i(v_{i_0})}{(b_i(v_{i_0})-1)*r_e+t_e} - \frac{1}{K} \\
&= \frac{(\Delta-1)*r_e+t_e}{(b_i(v_{i_0})-1)*r_e+t_e} * \left(\frac{1}{K} + \frac{\Delta E_i(v_{i_0})}{IE}\right) - \frac{1}{K} \\
&= \frac{1}{(b_i(v_{i_0})-1)*r_e+t_e} * \left(\frac{(\Delta-b_i(v_{i_0}))*r_e}{K} + \frac{((\Delta-1)*r_e+t_e)*\Delta E_i(v_{i_0})}{IE}\right) \\
&\geq \frac{1}{(d_G(v_{i_0})-1)*r_e+t_e} * \left(\frac{(\Delta-d_G(v_{i_0}))*r_e}{K} + \frac{((\Delta-1)*r_e+t_e)*\Delta E_i(v_{i_0})}{IE}\right) \qquad (1)
\end{aligned}$$

where $d_G(v)$ is the physical degree of node v in G. Thus, the value of δ_i depends on the network connectivity, the number of stages K, the initial battery capacity

IE, the transmission energy t_e and the reception energy r_e of each sensor. In particular, when the transmission range is reduced, the node degree $d_G(v)$ of v will decrease. This results in the reduction of the network connectivity and increase of the maximum degree Δ of nodes in the approximate, minimum degree spanning tree. Therefore, the gain of network lifetime at stage i is positive, and so is the entire network lifetime. Our later simulations in Section 4 indicates that the network lifetime delivered by using multiple routing trees is much longer than that by using a single routing tree when the transmission range is relatively smaller.

On the other hand, our objective is to maximize the entire network lifetime $\sum_{i=1}^{K} \tau_i$ rather than maximize the duration of each individual stage, where

$$\sum_{i=1}^{K} \tau_i = \tau_1 + \sum_{i=2}^{K}(\tfrac{1}{K} + \delta_i)\tau = (1 + \sum_{i=2}^{K} \delta_i)\tau. \quad (2)$$

Eq. (2) implies that the value of K should be as large as possible in order to maximize the network lifetime. However, there is a constraint on K from Inequality (1). It is obvious that δ_i is inversely proportional to the number of stages, and thus K is required to be as small as possible in order maximize the network lifetime at each stage. In addition, a larger K means that more frequent scheduling of using different routing trees is needed, which will incur an extra overhead on energy consumption. Therefore, there is a tradeoff between the choice of K and the prolonged network lifetime. The later experimental simulations confirm that K decreases with the growth of transmission range, since a longer transmission range implies a better network connectivity, and thereby reducing the maximum degree of the nodes in the approximate, minimum degree spanning tree and the difference of residual energy among the nodes become insignificant. The detailed algorithm for finding multiple routing trees is given below.

Algorithm. `Multiple_Routing_Trees`(G, IE, K)
/* G is the sensor network with initial battery capacity IE at each node and */
/* K is the number of potential stages */
begin
1. $\tau_1 \leftarrow$ `MDST`$(G, IE/K)$;
  /* τ_1 is the network lifetime if an approximate, minimum degree spanning */
  /* tree is used at stage 1. */
2. for $i \leftarrow 2$ to K do
3. $\tau_i \leftarrow 0$; /* τ_i is the duration of stage i */
4.   for each $v \in V$ do
5.   compute its residual energy $\Delta E_i(v)$;
     /* after the current tree has been used for τ_{i-1} units */
6.   endfor;
7. $\tau_i^1 \leftarrow$ `NMDST`$(G, IE/K + \Delta E_i())$;
   /* τ_i^1 is the network lifetime delivered by a degree-constrained spanning */
   /* tree, using algorithm `NMDST` based on the quota of energy plus $\Delta E_i(v)$ */
   /* of each node v at stage i */
8.   if $\tau_1 < \tau_i^1$ then
   /* check whether the degree-constrained spanning tree will result in a */
   /* longer duration of stage i */

$\tau_i \leftarrow \tau_i^1$;
else $\tau_i \leftarrow \tau_1$;
endif;
9. endfor;
end.

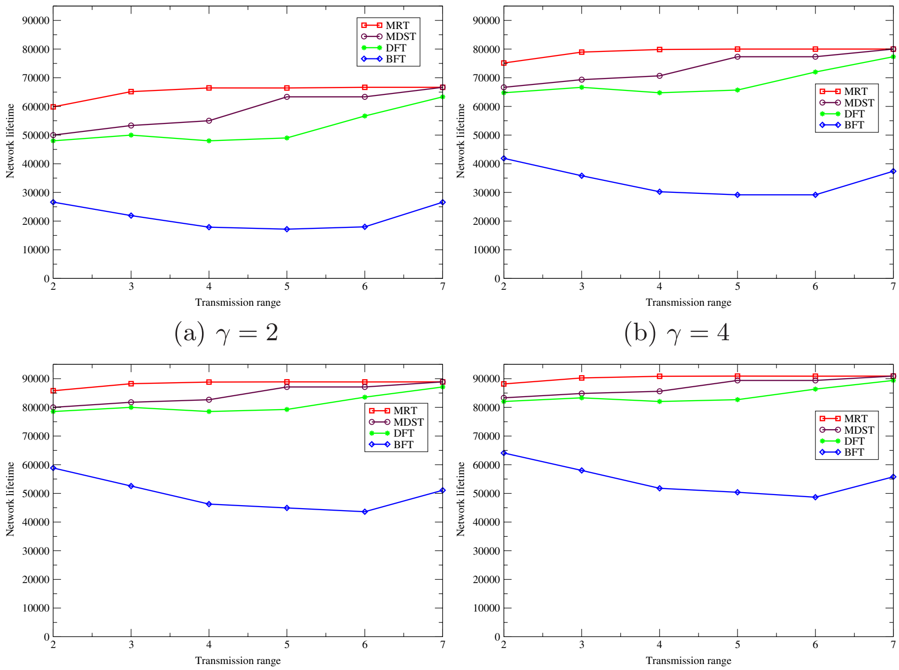
We refer to the above algorithm as algorithm MRT and have the following theorem.

Theorem 1. *Given a sensor network $G = (V, E)$ with identical initial battery capacity, assume that the transmission range of each sensor is identical. There is a heuristic algorithm for finding multiple routing trees for message-length independent aggregate queries. The network lifetime delivered by the proposed algorithm is longer than but at least as long as the one delivered by the algorithm for finding a single routing tree.*

4 Performance Evaluation

In this section we evaluate the performance of the proposed algorithm against existing algorithms through experimental simulations in terms of network lifetime. We assume that the monitored region is a 10×10 m^2 square in which 50 homogeneous sensor nodes are randomly deployed, by the NS-2 simulator. We also assume that the transmission range R of each sensor is from 2 to 7 with increment of 1. Two nodes can communicate with each other if and only if they are within the transmission range of each other, i.e., the Euclidean distance between them is no greater than R. We further assume that all the nodes have identical initial energy capacity $IE = 10^5$. Unless otherwise specified, we assume that the transmission energy consumption t_e for one unit-length of message is 1. Let $\gamma = t_e/r_e$, which is the *energy ratio* of the transmission energy consumption to the reception energy consumption for a unit-length data with $2 \leq \gamma \leq 10$. In the simulations, queries arrive one by one, and once a query arrives, it must be responded by the system, using the established routing tree. For simplicity, we assume that the answer to each query is a unit-length data as well. For each size of the network instance, the value shown in figures is the mean of 10 values obtained by running each algorithm on 10 randomly generated network topologies. In algorithm MRT, we limit the maximum number of various stages is $\lceil \sqrt{n} \rceil$, where n is the number of nodes in the sensor network and $n = 50$.

We first study the performance of the proposed algorithm MRT against the performance of the other three algorithms MDST, BFT and DFT by varying transmission ranges and energy ratios, where algorithms MRT, MDST, BFT and DFT are Multiple Routing Trees, Minimum Degree Spanning Tree, Breadth-First-Search Tree and Depth-First-Search Tree rooted at the base station respectively. As shown in Fig. 2(a), when the energy ratio $\gamma = 2$, algorithm MRT outperforms the others in terms of the network lifetime. For algorithm MRT, there is insignificant difference in the network lifetime when the transmission range R varies from 3 to 7, which implies that algorithm MRT can balance the energy consumption



(a) $\gamma = 2$ (b) $\gamma = 4$

(c) $\gamma = 8$ (d) $\gamma = 10$

Fig. 2. Comparison of network lifetime delivered by various algorithms

among the nodes in the network. Particularly, when $R = 7$, the network lifetime delivered by it is $\frac{10^5}{(0.5+1)} = 66666$, which is the optimal. This is because that the network connectivity is improved by increasing its sensor transmission ranges and the routing tree obtained is almost a chain. The network lifetime delivered by algorithm `MDST` increases with the growth of the transmission range R from 2 to 7, due to the fact that it tries to minimize the maximum degree of the nodes in the spanning tree, while the maximum degree node is the bottleneck of energy consumption among the nodes. With the improvement on network connectivity, the maximum degree of the nodes in the minimum degree spanning tree will reduce and thus the network lifetime is prolonged. Fig. 2(a) also shows that the difference of network lifetime delivered by algorithm `MRT` and `MDST` will diminish when the transmission range increases. The reason is that better connectivity improves the performance of algorithm `MDST` by reducing the maximum degree of nodes in the minimum degree spanning tree, and thus makes the network lifetime longer. When the transmission range R is 7, the maximum degree of nodes in minimum degree spanning tree becomes 2, and the network lifetime delivered by `MDST` reaches the maximum that is consistent with the network lifetime delivered by algorithm `MRT`. Meanwhile, it is observed that algorithm `BFT` is the worst among the algorithms, since the degrees of some nodes near to the tree root is maximized, while these nodes become the bottlenecks of energy consumption, thus they shorten the network lifetime. Similar to the performance of algorithm MDST, the degree of nodes of tree DFT drops with the improvement of network connectivity. Thus the network lifetime delivered by DFT increases with the growth of the transmission range.

When the energy ratio γ is 4, 8 or 10, the similar performance as the case where $\gamma = 2$ can be obtained, which is plotted in Fig. 2(b), (c) and (d). It can be seen that the performance of algorithm MRT is much better than that of algorithm MDST, especially for low connectivity sensor networks. The network lifetime delivered by algorithm MRT is almost the maximum one, which is $\frac{10^5}{(0.25+1)} = 80000$, $\frac{10^5}{(0.125+1)} = 88888$ or $\frac{10^5}{(0.1+1)} = 90909$ respectively when the transmission range R is no less than 3, whereas the network lifetime delivered by algorithm MDST is almost the optimal when the transmission range R is 7.

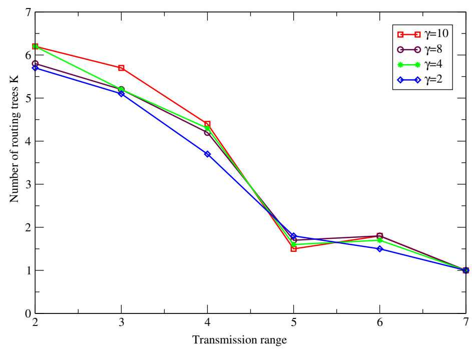


Fig. 3. The optimal number of routing trees in algorithm MRT

We then study the optimal number K of routing trees in algorithm MRT, and these trees will deliver the best possible network lifetime through experimental simulations. Recall that each value of K in Fig. 3 is the average of 10 different numbers of routing trees built for 10 different network topologies, given the transmission range R and the energy ratio γ. Fig. 3 shows that the optimal number of routing trees will decrease, when the transmission range R varies from 2 to 7 and the energy ratio γ is fixed. When the transmission range R is 2, the network lifetime is maximized if its network lifetime is partitioned upto six stages. This implies that more routing trees are needed for a low connectivity sensor network in order to maximize its network lifetime $\sum_{i=1}^{K} \tau_i$. When the transmission range is 7, the number of routing trees becomes 1, since the approximate, minimum degree spanning tree delivered by algorithm MDST now is a chain, when the nodes in the sensor network are highly connected with each other.

5 Conclusions

We have considered the evaluation of a class of message-length independent aggregate queries in sensor databases with an objective to maximizing the network lifetime. We first showed that the approximation ratio of a known approximation algorithm for the identical energy case is a constant, and provided the tighter lower and upper bounds on the optimal network lifetime for the non-identical energy case. We then introduced the concept of multiple routing trees to prolong the network lifetime further and devised a heuristic algorithm for finding such multiple routing trees. We finally conducted extensive experiments by simulation. The experimental results demonstrated that the performance of proposed algorithm significantly outperforms the existing ones that use only one single routing tree for the evaluation of such queries.

Acknowledgment. It is acknowledged that the work by the authors is fully funded by a research grant No:DP0449431 by Australian Research Council under its Discovery Schemes.

References

1. Govindan, R., Hellerstein, J.M., Hong, W., Madden, S., Franklin, M., Shenker, S.: The sensor network as a database. Technical Report 02-771, Computer Science Department, University of Southern California (September 2002)
2. Madden, S., Szewczyk, R., Franklin, M.J., Culler, D.: Supporting aggregate queries over ad hoc wireless sensor networks. In: Proc. 4th IEEE Workshop on Mobile Computing and System Applications, pp. 49–58. IEEE, Los Alamitos (2002)
3. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: TAG: a tiny aggregation service for ad hoc sensor networks. ACM SIGOPS Operating Systems Review 36(SI), 131–146 (2002)
4. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: The design of an acquisitional query processor for sensor networks. In: Proc. SIGMOD 2003, pp. 491–502. ACM Press, New York (2003)
5. Yao, Y., Gehrke, J.: The cougar approach to in-network query processing in sensor networks. ACM SIGMOD Record 31(3), 9–18 (2002)
6. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proc. the Hawaii International Conference on System Sciences, pp. 3005–3014. IEEE, Los Alamitos (2000)
7. Lindsey, S., Raghavendra, C.S.: PEGASIS: Power-efficient gathering in sensor information systems. In: Proc. Aerospace Conference, pp. 1125–1130. IEEE, Los Alamitos (2002)
8. Tan, H.Ö., Körpeoğlu, İ.: Power efficient data gathering and aggregation in wireless sensor networks. ACM SIGMOD Record 32(4), 66–71 (2003)
9. Kalpakis, K., Dasgupta, K., Namjoshi, P.: Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. Computer Networks 42, 697–716 (2003)
10. Buragohain, C., Agrawal, D., Suri, S.: Power aware routing for sensor databases. In: Proc. INFOCOM 2005, pp. 1747–1757 (2005)
11. Chang, J.-H., Tassiulas, L.: Energy conserving routing in wireless ad hoc networks. In: Proc. INFOCOM 2000, pp. 22–31. IEEE, Los Alamitos (2000)
12. Fürer, M., Raghavachari, B.: Approximating the minimum-degree Steiner tree to within one optimal. J. Algorithms 17, 409–423 (1994)