WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2012; **12**:219–235
DOI: 10.1002/wcm.952

RESEARCH ARTICLE

Aggregate node placement for maximizing network lifetime in sensor networks

Weifa Liang[1]*, Yinlong Xu[2], Jiugen Shi[3] and Junzhou Luo[4]

[1] School of Computer Science, The Australian National University, Canberra, ACT 0200, Australia
[2] School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, P. R. China
[3] School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, P. R. China
[4] School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, P. R. China

ABSTRACT

Sensor networks have been receiving significant attention due to their potential applications in environmental monitoring and surveillance domains. In this paper, we consider the design issue of sensor networks by placing a few powerful aggregate nodes into a dense sensor network such that the network lifetime is significantly prolonged when performing data gathering. Specifically, given K aggregate nodes and a dense sensor network consisting of n sensors with $K << n$, the problem is to place the K aggregate nodes into the network such that the lifetime of the resulting network is maximized, subject to the distortion constraints that both the maximum transmission range of an aggregate node and the maximum transmission delay between an aggregate node and its covered sensor are met. This problem is a joint optimization problem of aggregate node placement and the communication structure, which is NP-hard. In this paper, we first give a non-linear programming solution for it. We then devise a novel heuristic algorithm. We finally conduct experiments by simulation to evaluate the performance of the proposed algorithm in terms of network lifetime. The experimental results show that the proposed algorithm outperforms a commonly used uniform placement schema — equal distance placement schema significantly. Copyright © 2010 John Wiley & Sons, Ltd.

KEYWORDS
aggregate node placement; constraint optimization; data gathering; energy efficiency; network lifetime; sensor network

***Correspondence**
Weifa Liang, School of Computer Science, The Australian National University, Canberra, ACT 0200, Australia.
E-mail: wliang@cs.anu.edu.au

1. INTRODUCTION

Recent advances in micro-electronic technology have made it possible to construct compact and inexpensive wireless sensors. Networks formed by such sensors, referred to wireless sensor networks, have been receiving significant attention due to their potential applications from civil to military domains [1]. While these applications revealed tremendous potential of sensor networks for capturing important environment phenomenon, they also pose certain associated limitations. One of the major limitations is that sensors are powered by energy-limited batteries that impose a severe constraint on the network performance, energy conservation in wireless sensor networks is of paramount importance. To prolong the network lifetime, various energy optimization metrics have been proposed [2]. One optimization metric is minimizing the total energy consumption per operation. However, in many practical applications, the performance measure of actual interest is not to optimize the overall energy consumption, but rather, to maximize the lifetime of individual sensors, because a sensor failure can cause the network partitioned, and any further service may be interrupted. To avoid sensor extinctions due to the exhaustion of their batteries, any energy efficient routing algorithm should balance the energy consumption among the sensors evenly so as to prolong the network lifetime, where the *lifetime* of a wireless sensor network is the time of its first node failure [3]. Data gathering in most sensor networks is a fundamental and frequent operation, the energy gain through optimizing this operation can prolong the lifetime of a sensor network substantially. Energy efficient data gathering thus is one major research focus, which usually is implemented through in-network processing paradigm [4,5,27].

In this paper, we consider a strategy of prolonging network lifetime through the deployment of a very few powerful aggregate nodes into a dense sensor network. This problem is generally NP-hard, because not only are the aggregate nodes needed to be placed at proper positions in the monitored region but also an energy efficient routing tree rooted at the base station and spanning aggregate and sensors needs to be constructed in the resulting heterogeneous sensor network for data gathering.

1.1. Related work

Given a homogeneous sensor network, data gathering that aims to minimize the total energy consumption has been extensively studied in the literature [6–11]. Heinzelman *et al.* [6] initialized the study of this problem by proposing a clustering protocol `LEACH`. The nodes in `LEACH` are grouped into a number of clusters in a self-organizing manner, and a cluster head serves as a local "base station" to aggregate the gathered messages from its members and forward the result to the sink directly. Lindsey and Raghavendra [8] later provided an improved solution `PEGASIS` for the problem, in which all the nodes in the network form a chain and one of the nodes in the chain is chosen as the head to be responsible for reporting the aggregated result to the base station. Kalpakis *et al.* [7] considered this problem by proposing an integer program and heuristic solutions. It should be mentioned that, although these approaches considered the energy consumption issue, none of them incorporated the energy consumption metric explicitly into the problem formulation. Also, an assumption imposed on these algorithms is that the message length transmitted by a relay node is independent of the message lengths of its descendents, i.e., each node transmits the same volume of data no matter how much data it received from its children. Such query operations in databases include `AVG, MIN, MAX, COUNT`, etc. On the other hand, several studies have also been conducted for another kind of data gathering that the message length transmitted by a relay node depends on not only the length of its sensed message but also the message lengths of its descendants [12–17]. Goel and Estrin [14] addressed this latter data gathering problem by minimizing the total transmission energy consumption, assuming that the aggregation function at each relay node is modeled as a given concave, non-decreasing cost function. They proposed a hierarchical matching algorithm for the problem that delivers an approximate solution within a logarithmic factor of the optimum. Cristescu *et al.* [13] studied the data correlation problem with an objective to minimize the total transmission energy consumption by taking data correlation into account. Under the assumption that each node knows which node to be merged to generate a merged message with minimum length and each relay node has data aggregation and compression ability, they showed that the data correlation problem is NP-complete and provided an integer program solution, using the Slepian–Wolf coding approach. Rickenbach and Wattenhofer [16] later provided an approximation algorithm for data correlation problem with approximation ratio of $2(1+\sqrt{2})$. Buragohain *et al.* [12] studied the problem with an objective to maximize the network lifetime. They showed that finding an optimal routing tree is NP-complete. They instead proposed a heuristic for the problem. Liang and Liu [15] also studied the problem independently by showing its NP-completeness and devising various heuristics to tradeoff between different energy optimization metrics to prolong the network lifetime.

Unlike these mentioned works that focus on the construction of routing trees in a given homogeneous sensor network, there are several studies on placing sensors into a region of interest to form a sensor network and building a routing tree jointly in the literature [18–20]. For example, Dasgupta *et al.* [18] considered the sensor placement problem so as to maximize the network lifetime under the constraint that certain points of interest are within the coverage radius of at least one sensor. They proposed a heuristic `SPRING` for such a purpose. Ganesan *et al.* [20] studied the joint optimization of sensor placement and transmission structure so as to minimize the total power consumption of the network, subject to the maximum and average distortion constraints, using different coding schemes and correlation models. Cheng *et al.* [19] considered the node placement problem with an objective to either maximize the network lifetime or minimize the total energy consumption by formulating the problem into a non-linear programming problem.

Unlike these node placement algorithms for homogeneous sensor networks, we here consider a different node placement problem by assuming that a dense sensor network has already been deployed, in which sensors are randomly and redundantly deployed, we now strategically place a few powerful aggregate nodes into the network such that the lifetime of the resulting network (a heterogeneous sensor network) is maximized when performing data gathering, provided that the distortion constraints that both the maximum transmission range of aggregate nodes and the maximum transmission delay between an aggregate node and any of its covered sensors are met. Zhang *et al.* [21] discussed the minimum number of relay node placement problem with an objective to ensure 2-connectivity of the resulting network by proposing several approximation algorithms, they focused on the topological structure of the network by adding minimum number of relay nodes. Wang *et al.* [22] dealt with the replay node placement problem to enhance network connectivity and reliability by proposing approximation algorithms. We concentrate on how to place very few aggregate nodes into the network such that the network lifetime is maximized when performing data gathering, assuming that each aggregate node is responsible to data collection or aggregation for a subset of sensors under the distortion constraints.

1.2. Contributions

In this paper, our major contributions are as follows. We first introduce an aggregate node placement problem in heterogeneous sensor networks, by placing a few powerful aggregate nodes into a dense sensor network such that the lifetime of the resulting network is further maximized when performing data gathering. Due to the NP hardness of the problem, we then devise a novel heuristic for it. We finally conduct extensive experiments by simulation to evaluate the performance of the proposed algorithm in terms of network lifetime. The experimental results show that the proposed heuristic outperforms another commonly used uniform node placement schema — placing aggregate nodes with equal distance in between.

1.3. Paper organization

The rest of the paper is organized as follows. In Section 2, the system model is introduced and notations and notions are given too. In Section 3, a simple heuristic for aggregate node placement in a fan area is proposed. In Section 4, a heuristic placement algorithm for aggregate node placement without any distortion constraint is devised, which then is extended to solve the problem of concern. In Section 5, extensive experiments by simulation are conducted to evaluate the performance of the proposed algorithm against the uniform node placement schema in terms of network lifetime. The conclusions are given in Section 6.

2. PRELIMINARIES

2.1. System model

We consider a heterogeneous sensor network consisting of two types of wireless devices: lots of resource-constrained, cheap *sensor nodes* and a few resource-rich, expensive *aggregate nodes*. The cheap sensor has limited battery power, a short fixed transmission range, a low data rate and a low duty circle. The main tasks performed by a sensor are sensing, data processing, and data transmission and relay. In contrast, the expensive aggregate node has more power reserve, an adjustable transmission range, a higher data transmission rate, and much better data processing and storage capabilities, and its main tasks are to aggregate and/or process the sensed data from sensors and to transmit the aggregated data to the other aggregate nodes. Since the cost associated with aggregate nodes is not cheap, the number of aggregate nodes in a sensor network is very limited. Thus, they need to be placed carefully in order to maximize the lifetime of the resulting network.

To perform data gathering in a heterogeneous sensor network, in-network processing paradigm can be adopted [4,5]. That is, a routing tree rooted at the base station and spanning aggregate nodes and sensors will be used for data gathering. The tree actually is a 2-tier cluster routing tree, in which the aggregate nodes serve as the cluster heads and the sensors serve as the members of clusters. Each sensor has only one cluster head, the sensor is a *member* of the cluster head, and the cluster head *covers* the sensor. Within a cluster, each member sensor can forward its sensing data to the cluster head through multiple-hop member sensor relay. Once the cluster head collects all the sensing data from both its members and its descendant aggregate nodes, it then processes and transmits the collected data to its parent aggregate node. All data will be collected at the base station eventually through multi-hop aggregate nodes relay, using the 2-tier cluster tree. During one data gathering session, each sensor will consume the same amount of transmission energy by transmitting the same length message, since they also have identical transmission ranges. However, the transmission energy consumption of different aggregate nodes by transmitting a unit-length message is various, depending on the distance between the aggregate node and its receiver. For two aggregate nodes u and v with distance $d_{u,v}$, the transmission energy at aggregate node u is modeled to be proportional to $d^{\kappa}_{u,v}$ if a unit-length message is transmitted from u to v, assuming that the SNR at the receiver v meets the given threshold, where κ is a path-loss exponent parameter that typically takes on a value between 2 and 4, depending on the characteristics of the communication medium. Unless otherwise specified, in this paper we assume that $\kappa = 2$ and take into account the transmission energy consumption only by ignoring the other energy consumptions, as the radio frequency (RF) transmission is the dominant energy consumption in wireless communications [23]. It must be mentioned that the energy cost model in this paper can be easily extended to include the reception energy consumption without any difficulty.

2.2. Aggregate node placement problem

Given a sensor network consisting of n cheap sensors that are redundantly, uniformly, and randomly deployed in a region of interest, a base station, and a few expensive aggregate nodes K ($K << n$), *the aggregate node placement problem* is to place the K aggregate nodes into the dense sensor network such that the lifetime of the resulting network is maximized when performing data gathering, subject to the distortion constraints that both the maximum transmission range $R_{\max}$ of an aggregate node and the maximum transmission delay $D_{\max}$ between an aggregate node and any of its covered sensor. Notice that K usually is a constant or no more than the logarithmic of network size n, while n is quite large in comparison with K.

2.3. Data volume *versus* the number of sensors in an area

With the assumption that the sensors in the sensor network are uniformly and densely deployed, we further assume that the data generation rate at each sensor is identical, we can conclude that the total volume of sensed data generated by the sensors in a subregion is proportional to the area of the subregion. Thus, unless otherwise specified, in this paper we abuse the area of a subregion concept, and use the area to represent the total volume of sensed data generated by the sensors and the number of sensors in the area interchangeably.

2.4. Network lifetime

In this paper, we assume that the lifetime of a heterogeneous sensor network is fully determined by the lifetime of the induced subnetwork consisting of aggregate nodes only. Intuitively, the lifetime of a heterogeneous sensor network will be determined solely by the sensors rather than by the aggregate nodes through the following arguments. For a given aggregate node v_i, it serves as a *cluster head* to collect the sensed data generated by its member sensors in an area S through multi-hop sensor relay. Within the 2-tier cluster routing tree rooted at the base station, a subtree rooted at v_i consists of its member sensors in S only. Notice that these member sensors in the subtree near to the root v_i will consume much more energy than the other member sensors. The lifetime of the subtree will be determined by the lifetimes of these member sensors rather than v_i itself, because the power reserve in v_i is much higher than that in any of member sensors. However, the rationale behind our assumption is as follows.

Since lots of cheap sensors are densely (redundantly) deployed in the monitored region, the data generated by the sensors near to each other are highly correlated. Thus, the set of cheap sensors in an area S can be further partitioned into several disjoint subsets such that the sensors in each subset cover the area S and the communication subgraph induced by them is connected. A routing subtree rooted at v_i spanning the sensors in each subset can then be obtained [24,25]. Despite that the lifetime of each such subtree is limited, the sum of the lifetimes of all different subtrees rooted at v_i are comparable to the lifetime of aggregate node v_i. Notice that the aggregate node v_i consumes its energy on not only relaying data for all sensors it covered but also relaying data for its descendant aggregate nodes. Consequently, the lifetime of the routing tree consisting of aggregate nodes will determine the lifetime of the heterogeneous sensor network. Specifically, let $v_1, v_2, \ldots, v_K$ be the K aggregate nodes with each v_i consuming the amounts of energy $ec(v_i)$ per data gathering session, $1 \le i \le K$. Denote by $T(n, K)$ the network lifetime of the resulting heterogeneous sensor network by placing the K aggregate nodes into a sensor network, then

$$T(n, K) = \frac{E_{\mathrm{agg}}}{\max_{1 \le i \le K}\{ec(v_i)\}} \tag{1}$$

where E_{agg} is the initial energy capacity of an aggregate node. Equation (1) implies that the network lifetime is inversely proportional to the maximum energy consumption among the aggregate nodes. Thus, in the rest of this paper we will focus on minimizing the maximum energy consumption among the aggregate nodes.

3. AGGREGATE NODE PLACEMENT IN A FAN

In this section, we assume that the monitored region is a fan of angle θ with fan radius L, and the sensors in the fan are densely deployed, $0 < \theta < \pi$. We assign k aggregate nodes to the fan to collect the sensed data generated by the sensors, with an objective to maximize the lifetime of the resulting network. Finding an optimal solution for this problem is very challenging, because it involves searching through the space of all possible configurations of aggregate node placement and all possible routing trees for each configuration. In fact, the problem is NP hard, which will be shown later. Following the optimal node placement in lines [19,20], we propose a heuristic by placing the k aggregate nodes in the middle ray line of the fan. The aggregate nodes are indexed from k to 1 in decreasing order, started from the base station which is at the center of the fan (see Figure 1). Let r_i be the transmission distance between aggregate node v_i and aggregate node v_{i+1} in this placement schema, $1 \le i < k$. In the following we decide the value of each r_i to maximize the lifetime of the resulting network, $1 \le i \le k$.

Since an aggregate node near to the base station transmits more data than the other aggregate nodes, to prolong the network lifetime by balancing the energy consumption among the aggregate nodes, it is desirable that all aggregate nodes consume the same amount of energy roughly per data gathering session, and in the end they will run out of energy and die at the same time. Obviously, $r_1 > r_2 > \ldots > r_{k-1} > r_k$, because the data volume transmitted by aggregate node v_{i+1} is larger than that by aggregate node v_i, $1 \le i \le k$. Let r_0 be the distance between the other endpoint of the middle ray line and aggregate node v_1. Let F_i be the total volume of data transmitted from aggregate node v_i to its parent aggregate node v_{i+1} in the routing tree, following the assumption in Section 2.3

$$\begin{aligned} F_i &= \frac{\theta}{2\pi}\left(\pi * L^2 - \pi * \left(\sum_{j=i}^{k} r_j\right)^2\right) \\ &= \frac{\theta}{2}\left(L^2 - \left(\sum_{j=i}^{k} r_j\right)^2\right) \end{aligned} \tag{2}$$

We thus have a simple solution to the problem by solving the following multiple-variable r_i non-linear programming, $0 \le i \le k$.

$$\sum_{i=0}^{k} r_i = L \tag{3}$$

subject to

$$r_{i+1}^2 F_{i+1} = r_i^2 F_i, \quad \text{for all } i \text{ with } 0 \le i \le k \tag{4}$$

where $r_i^2 F_i$ is the energy consumption of aggregate node v_i, $1 \le i \le k$.

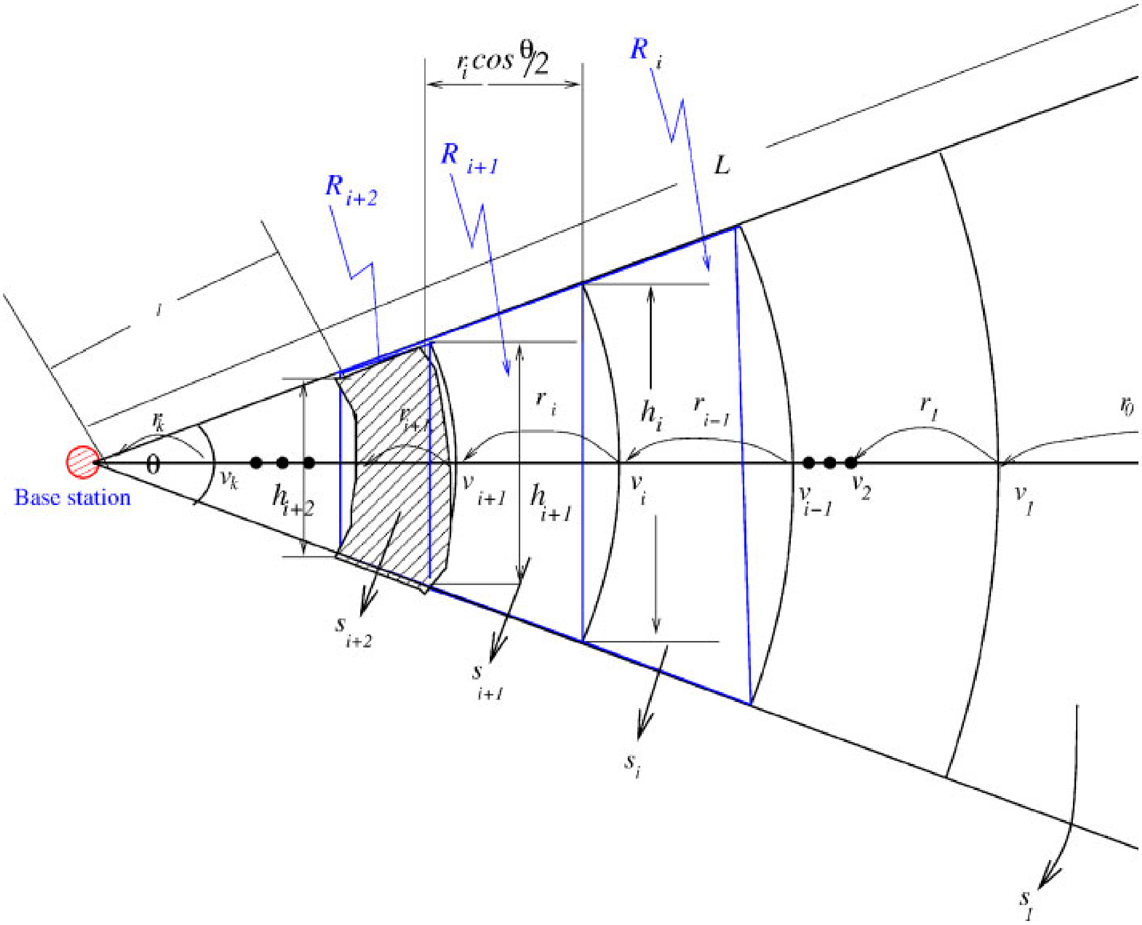


Figure 1. A fan of angle θ and radius L.

It is difficult to find a solution for the above non-linear programming within reasonable time even for a moderate problem size. A brute force approach to it is to enumerate all possible values of each real variable r_i ranged from 0 to L and check whether Equations (3), (2), and (4) are met. Therefore, this simple algorithm essentially is inapplicable in practice.

3.1. Heuristic algorithm

In the following we focus on devising a heuristic algorithm for the problem. Following $r_i < r_{i+1}$, we assume that $r_{i+1} = \sigma_{i+1} r_i$ with $0 < \sigma_{i+1} < 1$. To find a proper value for each r_i, we figure out σ_i first by the following lemma, $1 \le i \le k$.

Lemma 1. *Given a linear deployment of aggregate nodes in the middle ray line of a fan of angle θ with radius L, let s_i be the number of sensors that propagate their sensing data to aggregate node v_i directly through multi-hop sensor relay and R_i the trapezium of the area s_i covered by v_i. Then,* $\frac{s_{i+2}}{s_{i+1}} = \frac{R_{i+2}}{R_{i+1}} < \sigma_{i+1}$, $0 \le i \le k-1$.

Proof. Following the illustration in Figure 1, we have

$$s_{i+1} = \frac{\theta}{2\pi}[\pi(l + r_{i+1} + r_i)^2 - \pi(l + r_{i+1})^2]$$

$$= \frac{\theta}{2}[2(l + r_{i+1})r_i + r_i^2],$$

$$s_{i+2} = \frac{\theta}{2\pi}[\pi(l + r_{i+1})^2 - \pi l^2] = \frac{\theta}{2}[2lr_{i+1} + r_{i+1}^2],$$

$$\frac{s_{i+2}}{s_{i+1}} = \frac{2lr_{i+1} + r_{i+1}^2}{2(l + r_{i+1})r_i + r_i^2} \tag{5}$$

On the other hand, the areas of trapeziums R_{i+1} and R_{i+2} are as follows.

$$R_{i+1} = \left[2 * (l + r_{i+1}) \sin\frac{\theta}{2} + 2 * (l + r_{i+1} + r_i) \sin\frac{\theta}{2}\right]$$

$$* \frac{1}{2} r_i \cos\frac{\theta}{2},$$

$$R_{i+2} = \left[2 * (l + r_{i+1}) \sin\frac{\theta}{2} + 2 * l \sin\frac{\theta}{2}\right] * \frac{1}{2} r_{i+1} \cos\frac{\theta}{2},$$

$$\frac{R_{i+2}}{R_{i+1}} = \frac{2lr_{i+1} + r_{i+1}^2}{2(l + r_{i+1})r_i + r_i^2} \tag{6}$$

Thus,

$$\frac{s_{i+2}}{s_{i+1}} = \frac{R_{i+2}}{R_{i+1}} = \frac{\frac{h_{i+2}+h_{i+1}}{2}(r_{i+1}\cos\frac{\theta}{2})}{\frac{h_{i+1}+h_{i}}{2}(r_{i}\cos\frac{\theta}{2})} < \frac{\frac{h_{i+1}+h_{i+1}}{2}(r_{i+1}\cos\frac{\theta}{2})}{\frac{h_{i+1}+h_{i+1}}{2}(r_{i}\cos\frac{\theta}{2})} = \frac{r_{i+1}}{r_i} = \sigma_{i+1} \tag{7}$$

■

Having Lemma 1, the following heuristic will deliver a feasible solution for the problem. Recall that F_i is the volume of data transmitted from aggregate node v_i to its parent aggregate node v_{i+1}, $1 \le i \le k$. Then

$$F_1 = \frac{\theta}{2\pi}[\pi L^2 - \pi(L-r_0)^2] \tag{8}$$

$$\sigma_1 = 1 \tag{9}$$

For all i, $1 < i \le k$, we have

$$F_i = s_i + F_{i-1} \le \sigma_{i-1}s_{i-1} + F_{i-1} \le \Pi_{j=1}^{i-1}\sigma_j F_1 + F_{i-1} = \sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j F_1), \tag{10}$$

let $\rho_i = \frac{s_i}{F_i} = \frac{\Pi_{j=1}^{i-1}\sigma_j F_1}{\sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j F_1)} = \frac{\Pi_{j=1}^{i-1}\sigma_j}{\sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j)}$ (11)

since $\sigma_i\sigma_{i-1}\ldots\sigma_1 < \sigma_{i-1}\ldots\sigma_1$. Clearly $\rho_1 = 1$ and $\rho_2 = \frac{1}{2}$.

To balance the energy consumption among the aggregate nodes, it is desirable that all aggregate nodes consume the same amounts of energy per data gathering session, thereby they will run out of energy simultaneously, i.e., for each i, $1 \le i \le k$, we have

$$r_{i+1}^2 F_{i+1} = r_i^2 F_i \tag{12}$$

Given $r_{i+1} = \sigma_{i+1}r_i$, we have

$$\begin{aligned} r_{i+1}^2 F_{i+1} &= (\sigma_{i+1}r_i)^2 F_{i+1} = (\sigma_{i+1}r_i)^2(s_{i+1} + F_i) \\ &= (\sigma_{i+1}r_i)^2\left(1 + \frac{s_{i+1}}{F_i}\right)F_i \\ &= (\sigma_{i+1}r_i)^2\left(1 + \frac{\sigma_i s_i}{F_i}\right)F_i \le (\sigma_{i+1}r_i)^2(1 + \sigma_i\rho_i)F_i \end{aligned} \tag{13}$$

Combing with Equation (12), we have

$$\sigma_{i+1}^2(1 + \sigma_i\rho_i) = 1, \quad \sigma_{i+1} = \sqrt{\frac{1}{1 + \sigma_i\rho_i}} \tag{14}$$

Given σ_2 and ρ_2, the rest of σ_i and ρ_i can be computed by Equation (14), $i \ge 3$. To determine the value of r_i, $2 \le i \le k$, we have

$$r_i = \sigma_i r_{i-1} = \Pi_{j=1}^{i}\sigma_j r_1 \tag{15}$$

The problem now is reduced to find a solution for both r_1 and r_0. Combined with Equation (15), Equation (3) can be rewritten as follows.

$$\sum_{i=1}^{k}\Pi_{j=1}^{i}\sigma_j r_1 = L - r_0 \tag{16}$$

Meanwhile, Equation (12) implies that aggregate nodes v_1 and v_k consume the same amounts of energy per data gathering session, then

$$r_1^2 F_1 = r_k^2 F_k \tag{17}$$

Equation (17) can be rewritten below.

$$2r_0L - r_0^2 = (\Pi_{i=1}^{k}\sigma_i)^2[L^2 - (\Pi_{i=1}^{k}\sigma_i)^2 r_1^2] \tag{18}$$

The solution to variables r_1 and r_0 then can be obtained, by solving Equation (16) and Equation (18). All other r_is ($i > 1$) can be computed by Equation (15).

4. AGGREGATE NODE PLACEMENT IN DENSE SENSOR NETWORKS

In this section, we first show that the aggregate node placement problem is NP-complete. We then devise a heuristic algorithm for a simplified version of the problem without any distortion constraint. We finally extend the heuristic to solve the problem by incorporating the specified distortion constraints into consideration.

4.1. NP-hardness

The aggregate node placement problem is a joint optimization of aggregate node placement and the communication structure (a routing tree rooted at the base station), subject to the distortion constraints that both the maximum transmission range of an aggregate node and the maximum transmission delay between an aggregate node and any of its covered sensors. The problem is NP-hard by the following lemma.

Lemma 2. *The aggregate node placement problem is NP-hard.*

Proof. We show that a simplified version of the problem is NP-hard. We assume that the aggregate nodes have been placed into the region of interest, i.e., the position of each aggregate node has been fixed already. We further assume that each aggregate node covers the same number of sensors and the distance between any two sensors that are covered by two different aggregate nodes is greater than the sensor transmission range. This means that the total volume of sensed data generated by the sensors covered by each aggregate node is equal and each aggregate node covers the same number of sensors. Then, an optimal 2-tier cluster routing tree rooted at the base station and spanning the aggregate nodes for data gathering is needed to be found, in terms of maximizing the network lifetime. However, finding such a tree in such a network has been shown to be is NP-complete in Reference [15]. Since the aggregate node placement problem is at least as hard as this simplified version, it is NP-hard too. ■

4.2. Overview of the proposed algorithm

Following the same assumption as described in References [19,20,26], we assume that the monitored region is a circle of radius L, suppose that the sensors have been redundantly deployed in the circle already. We propose a heuristic solution to the problem by partitioning the circle into a number of fans of angle θ and allocating each fan with k $(= \lfloor \frac{K\theta}{2\pi} \rfloor)$ aggregate nodes. Assume that the k aggregate nodes assigned to each fan are arranged into the middle ray line of the fan and r_i is the distance between aggregate nodes v_i and v_{i+1}, $1 \leq i \leq k-1$, the problem then is reduced to minimize the maximum energy consumption among the aggregate nodes in the middle ray line, which can be expressed as follows.

$$\text{minimize} \max_{1\leq i\leq k} \left\{ \frac{\theta}{2\pi} r_i^{\kappa} \left(\pi L^2 - \pi \left(\sum_{j=i}^{k} r_j \right)^2 \right) \right\}$$

where $\frac{\theta}{2\pi} r_i^{\kappa} \left(\pi L^2 - \pi\left(\sum_{j=i}^{k} r_j\right)^2\right)$ is the energy consumption of aggregate node v_i, κ is the path loss exponent which is typically between 2 and 4. Recall that in this paper we set $\kappa = 2$. As a result, the problem is to find r_i and θ such that $\max_{1\leq i\leq k}\left\{\theta r_i^2 \frac{L^2-\left(\sum_{j=i}^{k} r_j\right)^2)}{2}\right\}$ is minimized, $1 \leq i \leq k$. The network lifetime $T(n, K)$ of the network thus is

$$T(n, K) = \frac{E_{\text{agg}}}{\max_{1\leq i\leq k} \left\{ \theta r_i^2 \frac{L^2-\left(\sum_{j=i}^{k} r_j\right)^2)}{2} \right\}} \tag{19}$$

Recall that E_{agg} is the initial energy capacity of any aggregate node, which is given in advance.

4.3. Algorithm without any distortion constraint

We here consider a simplified version of the problem without any distortion constraint. We start with the following important theorem.

Theorem 1. *Given a few powerful aggregate nodes K, assume that a circle is partitioned into f fans of angle θ ($\theta * f = 2\pi$) and each fan is assigned with $k = \left\lfloor \frac{K}{f} \right\rfloor$ aggregate nodes, which are placed in the middle ray line of the fan. To maximize the lifetime of the resulting network, the number of fans in the fan partition is no more than four, i.e., $f \leq 4$.*

Proof. Let f be the number of fans in the current fan partition of the circle with $f > 4$. To show the claim, we show that there is always another better fan partition to partition the circle into f' fans with $f' < f$ such that the network lifetime is longer than that by the current fan partition.

Let $E(\theta, k, L, f)$ be the maximum energy consumption among the aggregate nodes by the proposed algorithm in the previous section through partitioning the circle of radius L into f fans of angle θ, and k aggregate nodes are deployed into the middle ray line of each fan. There is another fan partition that partitions the circle into f' fans. Each fan is assigned $2k$ aggregate nodes at least, where $f' = f/2$ and the fan angle is 2θ if f is even; otherwise, $f' = \frac{f-1}{2}$ and the fan angle is $\frac{4\pi}{f-1}$ $(= 2\theta + \frac{2\theta}{f-1})$. We claim that this latter fan partition leads to a longer network lifetime. In other words, (i) $\frac{E(\theta,k,L,f)}{E(2\theta,2k,L,f/2)} \geq 4/3 = 1.33$ if $f > 2$ is even; otherwise, (ii) $\frac{E(\theta,k,L,f)}{E(\frac{4\pi}{f-1},2k,L,\frac{f-1}{2})} \geq 16/15 \approx 1.05$ when $f \geq 5$. We first show Claim (i) through the following digram (see Figure 2).

Assume that there are two neighboring fans of angle θ, and each is assigned k aggregate nodes to aggregate the sensed data from its fan area. A merged fan of angle 2θ is formed by merging the two fans. We reallocate the $2k$ aggregate nodes in the middle ray line of the merged fan as follows.

The first k aggregate nodes are placed in the line exactly at the same positions as they were placed in the middle ray line of the fan of angle θ. For each of the remaining k aggregate nodes, insert it into the half way of the two corresponding placed neighboring nodes in the line. As a result, for each aggregate node v_i in this new placement schema, the transmission distance between v_i and its parent aggregate node is a half in the original placement. Meanwhile, the total amount of data transmitted by each v_i varies, which is analyzed as follows.

Let r_i and r'_i be the transmission distances between aggregate node v_i and its parent v_{i+1} before and after fan merging, and F_i and F'_i the amounts of data transmitted by v_i to its parent v_{i+1} before and after fan merging, respectively. Following the fact that the energy consumption at a node is super-linear to the transmission distance between the sender and its receiver, there are two cases to be dealt with.

Case 1: the distance from v_i to the base station in the new placement is identical to the original one, but its transmission distance to its parent is half of its original one, the amount of data transmitted by v_i is twice as much as its original one. The energy consumption at v_i thus is

$$\begin{aligned} E(2\theta, 2k, L, f', v_i) &= r'^2_i F'_i \leq \left(\frac{r_i}{2}\right)^2 (2F_i) \\ &= \frac{1}{2} r_i^2 F_i = \frac{1}{2} E(\theta, k, L, f, v_i) \end{aligned} \tag{20}$$

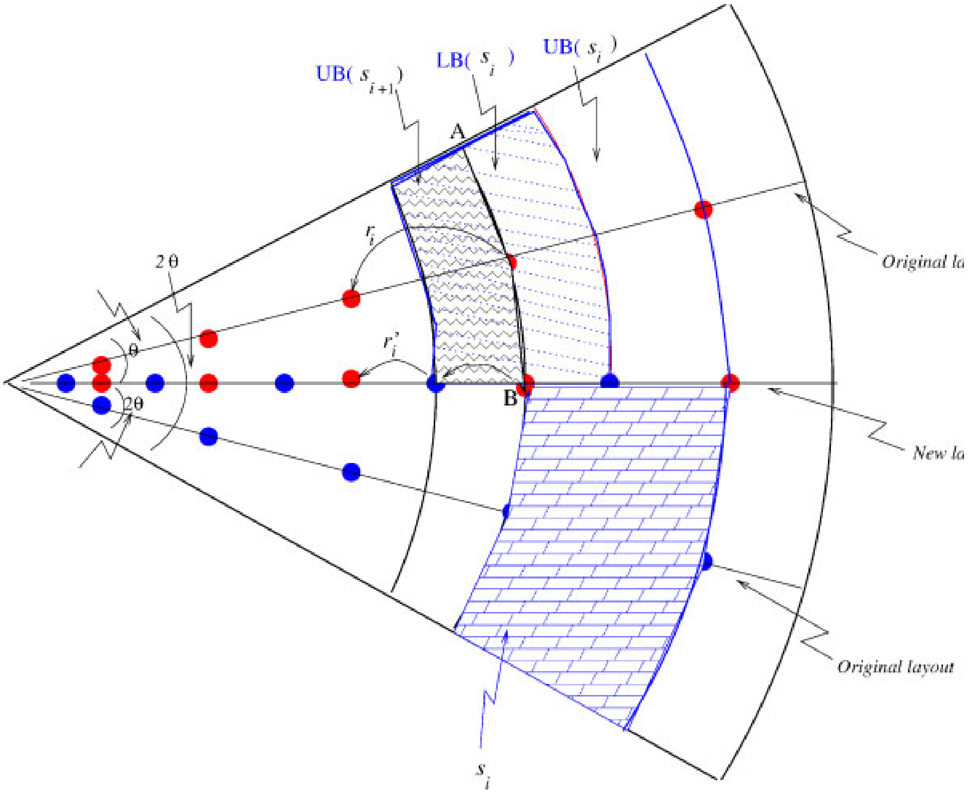


Figure 2. Fan merging with each being angle θ and k aggregate nodes.

Case 2: the distance from v_i to the base station in the new placement is equal to its original distance minus $r_i/2$ (v_i is placed in the half way of two neighboring nodes), i.e., $r'_i = r_i/2$. Let $UB(s_i)$ and $LB(s_i)$ be the larger and the smaller curve areas of s_i by partitioning the area using its middle curve line $\widehat{AB}$ (see Figure 2). Clearly, $UB(s_i) > \frac{s_i}{2}$ while $LB(s_i) < \frac{s_i}{2}, 2 \le i \le k$. To compute F'_i, s'_i needs to be computed, where s'_i is the number of sensors in an area that are covered by v_i in the new placement. It is obvious that s'_i is no greater than $2UB(s_{i+1})$, while $UB(s_{i+1}) \le LB(s_i) \le \frac{s_i}{2}$, illustrated by Figure 2, we have

$$\begin{aligned} F'_i &= s'_i + 2F_i \le 2UB(s_{i+1}) + 2F_i \\ &\le s_i + 2F_i \le 3F_i, \quad s_i \le F_i \qquad (21) \end{aligned}$$

$$\begin{aligned} E(2\theta, 2k, L, f', v_i) &= r'^2_i * F'_i \le \frac{r_i^2}{4} * 3F_i \\ &= \frac{3}{4} E(\theta, k, L, f, v_i) \qquad (22) \end{aligned}$$

When f is even, we have

$$\frac{E(\theta, k, L, f)}{E(2\theta, 2k, L, f/2)} = \frac{\max_{1 \le i \le k}\{E(\theta, k, L, f, v_i)\}}{\max_{1 \le j \le 2k}\{E(2\theta, 2k, L, f/2, v_j)\}} \ge \min\{2, 4/3\} = 1.33$$

Claim (ii) can be shown with similar arguments as for Claim (i) by the following observation.

Observation 1 *Assume that the circle has been partitioned into f fans of angle θ, let $S(\theta)$ be the area of a fan. The fan area $S(\frac{4\pi}{f-1})$ in the new fan partition of the circle into f' fans of angle $\frac{4\pi}{f-1}$ $(= 2\theta + \frac{2\theta}{f-1})$ is $(2 + \frac{2}{f-1})S(\theta)$ when $f \ge 2$.*

We place the $2k$ aggregate nodes into each of the f' fans by two cases.

Case 1: the distance from v_i to the base station in the new placement is identical to the original one, but the transmission distance between v_i and its parent is a half of its original one, the amount of data transmitted by v_i is $2 + \frac{2}{f-1}$ times as much as its original one by Observation 1. The energy consumption at v_i in the new node placement thus is

$$\begin{aligned} E\left(\frac{4\pi}{f-1}, 2k, L, \frac{f-1}{2}, v_i\right) &= r'^2_i F'_i \le \left(\frac{r_i}{2}\right)^2 \\ \times \left(2 + \frac{2}{f-1}\right) F_i &= \frac{1}{2}\left(1 + \frac{1}{f-1}\right) r_i^2 F_i \\ &= \frac{1}{2}\left(1 + \frac{1}{f-1}\right) E(\theta, k, L, f, v_i) \end{aligned}$$

$$
\begin{aligned}
&= \frac{f}{2(f-1)} E(\theta, k, L, f, v_i) \\
&< E(\theta, k, L, f, v_i), \quad if \quad f \geq 3 \qquad (23)
\end{aligned}
$$

Case 2: the distance from v_i to the base station in the new placement is equal to its original distance minus $r_i/2$ (v_i is placed in the half way of two neighboring nodes), i.e., $r'_i = r_i/2$. To compute F'_i, s'_i needs to be computed. However, s'_i is no greater than $(2 + \frac{2}{f-1})s_{i+1}$, following Observation 1, we thus have

$$
\begin{aligned}
F'_i = s'_i + \left(2 + \frac{2}{f-1}\right) F_i
&\leq \left(2 + \frac{2}{f-1}\right) UB(s_{i+1}) + \left(2 + \frac{2}{f-1}\right) F_i \\
&\leq \left(1 + \frac{1}{f-1}\right) 2LB(s_i) + \left(2 + \frac{2}{f-1}\right) F_i \\
&\leq \left(1 + \frac{1}{f-1}\right) s_i + 2\left(1 + \frac{1}{f-1}\right) F_i \\
&= \left(1 + \frac{1}{f-1}\right) (s_i + 2F_i) \\
&\leq \frac{3f}{f-1} F_i, \quad s_i \leq F_i \qquad (24)
\end{aligned}
$$

Then,

$$
\begin{aligned}
E\left(\frac{4\pi}{f-1}, 2k, L, \frac{f-1}{2}, v_i\right) &= r'^2_i * F'_i \\
&\leq \frac{r_i^2}{4}\left(\frac{3f}{f-1}\right) F_i \\
&= \frac{3f}{4(f-1)} r_i^2 F_i \\
&= \frac{3f}{4(f-1)} E(\theta, k, L, f, v_i) \\
&< E(\theta, k, L, f, v_i), \text{ if } f \\
&\geq 5 \qquad (25)
\end{aligned}
$$

We thus have

$$
\begin{aligned}
&\frac{E(\theta, k, L, f)}{E\left(\frac{4\pi}{f-1}, 2k, L, \frac{f-1}{2}\right)} \\
&= \frac{\max_{1\leq i\leq k}\{E(\theta, k, L, f, v_i)\}}{\max_{1\leq j\leq 2k}\left\{E\left(\frac{4\pi}{f-1}, 2k, L, \frac{f-1}{2}, v_j\right)\right\}} > 1, \\
&\text{if } f \geq 5 \qquad (26)
\end{aligned}
$$

■

Theorem 1 implies that given the value of K, the larger the value of fan angle θ, the longer the network lifetime will be. To maximize the network lifetime, the maximum number of fans obtained by the circle partition is no more than four. We thus have the following heuristic algorithm.

```
Algorithm Lifetime_without_const($G, L, K$)
begin
    $max\_energy\_consu \leftarrow \infty$; $net\_lifetime \leftarrow 0$;
    /* the maximum energy consumption among
    the aggregate nodes */
    /* to achieve the network lifetime $net\_lifetime$ */
    $f \leftarrow 5$; $\theta \leftarrow \frac{2\pi}{f}$; $k \leftarrow \lfloor \frac{K}{f} \rfloor$; /* $k$ is the number
    of nodes assigned to each fan area */
    $control \leftarrow' true'$;
    $\rho_2 \leftarrow 1/2$; $\sigma_2 \leftarrow \frac{1}{\sqrt{2}}$;
    while $control$ do
        for $i \leftarrow 3$ to $k$ do
            $\sigma_i \leftarrow \sqrt{\frac{1}{1+\sigma_{i-1}\rho_{i-1}}}$; $\rho_i \leftarrow \frac{\Pi_{j=1}^{i-1}\sigma_j}{\sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j)}$;
        endfor;
        compute $r_0$ and $r_1$, using Equation (16) and
        Equation (18)
        for $i \leftarrow 2$ to $k$ do
            $r_i \leftarrow \Pi_{j=1}^{i}\sigma_j r_1$;
        endfor;
        $E(\theta, k, L, f) \leftarrow max_{1\leq i\leq k}\{\theta r_i^2 \frac{L^2-(\sum_{j=i}^{k} r_j)^2)}{2}\}$
        if $max\_energy\_consu > E(\theta, k, L, f)$ then
            $max\_energy\_consu \leftarrow E(\theta, k, L, f)$;
            $\theta_0 \leftarrow \theta$; $k_0 \leftarrow k$; $f_0 \leftarrow f$;
            $f \leftarrow \lfloor f/2 \rfloor$; $\theta \leftarrow \frac{2\pi}{f}$; $k \leftarrow \lfloor \frac{K}{f} \rfloor$;
        else $control \leftarrow' false'$;
        endif;
    endwhile;
    $max\_energy\_consu \leftarrow E(\theta_0, k_0, L, f_0)$;
    $net\_lifetime \leftarrow \frac{E_{agg}}{max\_energy\_consu}$;
end.
```

If taking into account the distortion constraints, the above approach suffices a serious shortcoming. Consider a scenario where the circle is partitioned into two fans of fan angle π, within each fan, the $k = K/2$ aggregate nodes are arranged into the middle ray line of the fan, and the lifetime of the resulting network will be maximized, following Theorem 1. However, under such a placement schema, the maximum transmission delay between a sensor and the aggregate node that covers the sensor is as large as $\sqrt{L^2 + (L - r_0)^2} \approx \sqrt{2}L$, which is unacceptable for many real applications, because the sensor requires to take $\lceil \frac{\sqrt{2}L}{t_x} \rceil$ hops sensor relays to reach its cluster head, where t_x is the fixed transmission radius of a sensor. As each sensor is severely energy constrained and the message length transmitted by a sensor is at least proportional to the number of hops, to perform message transmission to a cluster head, a member sensor near the cluster head will run out of its battery very quickly.

4.4. Algorithm with distortion constraints

We consider the aggregate node placement problem under distortion constraints that the maximum transmission delay between an aggregate node and its covered sensor is bounded by a given value $D_{\max}$ and the maximum transmission range of each aggregate node is bounded by a value $R_{\max}$. To meet the distortion constraints, the minimum number $k_{\min}$ of aggregate nodes in the middle ray line of a fan is at least $\lceil \frac{L-D_{\max}}{R_{\max}} \rceil$, since the following inequality must be met.

$$k_{\min} R_{\max} + D_{\max} \geq L \tag{27}$$

Otherwise, the communication graph induced by the aggregate nodes is disconnected. Consequently, such an aggregate node placement schema is useless because it cannot bring any benefit to the network in terms of prolonging the network lifetime.

Following Theorem 1, if the aggregate nodes deployed into the circle are organized as a star structure centered at the circle center (i.e., they are deployed in the middle ray lines of fans of angle θ), then, the larger the θ value, the longer the network lifetime will be. On the other hand, with the growth of the value of θ, the transmission delay becomes bigger and bigger, and ultimately it will beyond the maximum transmission delay bound $D_{\max}$. The value of θ thus cannot be arbitrarily large, the rest is to choose a proper value for θ to meet the distortion constraints.

Following aggregate node placement into the middle ray line of a fan with fan angle θ, the sensing area s_1, covered by aggregate node v_1, is the largest one among the subareas covered by the k aggregate nodes in the fan area. We have the following observation.

Observation 2 *If the transmission delay (distance) between the farthest corner sensor in s_1 and aggregate node v_1 is no greater than $D_{\max}$, then, there is no any other sensor in s_1 whose distance to v_1 is larger than $D_{\max}$. Furthermore, for any other area s_i, the maximum transmission delay between a sensor in s_i and the aggregate node v_i covering the sensor is no more than $D_{\max}$ too,* $2 \leq i \leq k$.

Having Observation 2, we can determine the value of θ and the value of each r_i to meet the maximum transmission delay constraint, by solving the following equations and inequalities: Equation (18), Inequality (30), Inequality (31), and Equation (33), $0 \leq i \leq k$. Figure 3 illustrates the distortion constraints.

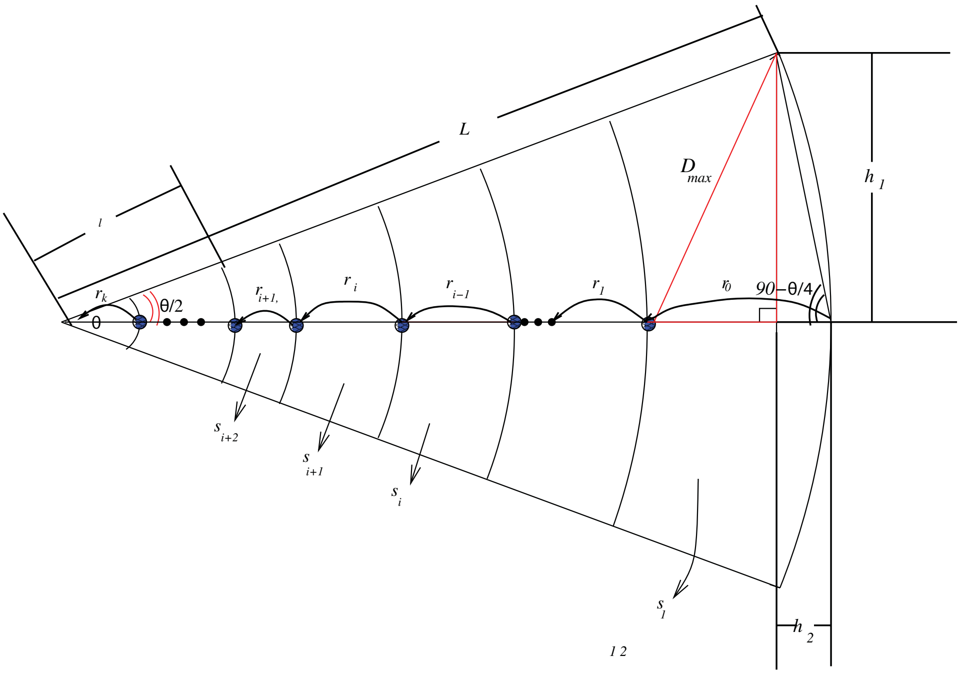


Figure 3. A circle is partitioned into a number of fans of angle θ with the maximum transmission delay $D_{\max}$.

$$h_1 = L \sin \frac{\theta}{2} \tag{28}$$

$$h_2 = h_1 / \tan\left(90^o - \frac{\theta}{4}\right) \tag{29}$$

and

$$(r_0 - h_2)^2 + h_1^2 \leq D_{\max}^2 \tag{30}$$

We thus have

$$r_0 \leq \frac{L \sin \frac{\theta}{2}}{\tan(90^o - \frac{\theta}{4})} + \sqrt{D_{\max}^2 - L^2\left(\sin\frac{\theta}{2}\right)^2} \tag{31}$$

$$r_0 \leq D_{\max}, \quad k = \left\lfloor \frac{K\theta}{2\pi} \right\rfloor \tag{32}$$

Given $\rho_2 = 1/2$ and $\sigma_2 = \frac{1}{\sqrt{2}}$, the rest of σ_i and ρ_i can be computed by Equation (14) for all $i \geq 3$.

$$L - r_0 = \sum_{i=1}^{k} r_i = \sum_{i=1}^{k} \Pi_{j=1}^{i} \sigma_j r_1 \tag{33}$$

Given r_0 and r_1, we then compute the value of $r_i = \Pi_{j=1}^{i}\sigma_j r_1$ for all i, $2 \leq i \leq k$.

We then incorporate the maximum transmission range $R_{\max}$ of an aggregate node into the design of the proposed algorithm. It is obvious that the maximum transmission range between any two neighboring aggregate nodes in a middle ray line of a fan is the one between aggregate nodes v_1 and v_2, i.e., r_1, since $r_{i+1} < r_i$ for all i, $1 \leq i \leq k$. Therefore,

$$r_1 \leq R_{\max} \tag{34}$$

In summary, we have the following theorem.

Theorem 2. *Given a dense sensor network consisting of n sensors for monitoring a circle of radius L with the base station at the circle center, assume that the circle is partitioned into a number of fans of angle θ, $1 \leq \theta \leq \pi/2$. Let $R_{\max}$ be the maximum transmission range of each aggregate node and $D_{\max}$ the maximum transmission delay between a sensor and the aggregate node that covers the sensor. Let K ($K << n$) be the number of aggregate nodes to be placed into the circle. To maximize the lifetime of the resulting network by placing the K aggregate nodes into the sensor network, there is an approach of placing the K aggregate nodes into the middle ray lines of fans to form a star structure such that both constraints are met, the values of θ and r_i can be found by solving the following equations and inequalities as well as Equation (18).*

$$r_0 \leq \frac{L \sin \frac{\theta}{2}}{\tan(90^o - \frac{\theta}{4})} + \sqrt{D_{\max}^2 - L^2\left(\sin\frac{\theta}{2}\right)^2}, \tag{35}$$

$$r_0 \leq D_{\max}, \quad r_1 \leq R_{\max}, \quad k = \left\lfloor \frac{K\theta}{2\pi} \right\rfloor, \tag{36}$$

$$L - r_0 = \sum_{i=1}^{k} \Pi_{j=1}^{i} \sigma_j r_1 \tag{37}$$

and

$$r_i = \Pi_{j=1}^{i} \sigma_j r_1, \quad 2 \leq i \leq k \tag{38}$$

In the following a greedy algorithm is proposed, based on Theorem 2. Let f be the number of fans. We partition the circle into a number of fans of identical angle, and each fan is assigned $k_{\min}$ aggregate nodes initially. It is obvious that $f = \lfloor \frac{K}{k_{\min}} \rfloor$, where $k_{\min}$ is defined by Inequality (27). We further assume that $K \geq 2k_{\min}$. The algorithm proceeds in iterations. Let f be the number of fans of angle θ at the current iteration. If there is a feasible solution to meet all constraints, we then check whether f is even. If it is, following Theorem 2 we merge every two neighboring fans into a merged fan. As a result, there are $f/2$ fans of angle 2θ, and each fan is assigned $\lfloor \frac{2K}{f} \rfloor$ aggregate nodes at the next iteration. Otherwise, there are $\frac{f-1}{2}$ fans of angle $\frac{4\pi}{f-1}$ ($= 2\theta + \frac{2\theta}{f-1}$), and each fan is assigned $\lfloor \frac{2K}{f-1} \rfloor$ aggregate nodes at the next iteration. If there is no solution to meet the constraints at the current iteration, then, the setting value of fan angle θ at the current iteration is too large, and the fan angle θ' in the potential solution is such a maximum value between $\theta/2$ and θ that meets the constraints. Instead of working on the value of θ' directly, we use the number of aggregate nodes assigned to a fan to derive the value of θ'. The detailed algorithm is described in Page 12.

5. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed algorithm against another commonly used uniform placement schema that places sensors with equal distance in between. It is noticed that the uniform placement schema has been used as the benchmark of performance evaluation in literatures [19,20]. In all our experiments, we assume that the radius of the monitored circle $L = 250$ m and $n = \pi L^2 = 196\,250$ cheap sensors have been uniformly deployed in the circle. The value at each figure is the average of 15 results of applying each algorithm to different topological networks with the same network size. Recall that the network lifetime is proportional to the initial energy capacity of aggregate node and inversely proportional to the maximum energy consumption among aggregate nodes per data gathering session. Thus, we focus on minimizing the maximum energy consumption of aggregate nodes per data gathering session to maximize the network lifetime.

```
Algorithm Max_Net_Lifetime($G$, $L$, $K$, $D_{max}$, $R_{max}$)
begin
    $max\_energy\_consu \leftarrow \infty$; $net\_lifetime \leftarrow 0$;
    $f \leftarrow \lfloor \frac{K}{k_{min}} \rfloor$; /* the circle is partitioned into f fans */
    /* and each fan is assigned $k_{min}$ aggregate nodes initially */
    $\theta \leftarrow \frac{2\pi}{f}$; /* the value of fan angle $\theta$, following Equation (27) */
    $k \leftarrow \lfloor \frac{K}{f} \rfloor$; /* k is the number of aggregate nodes assigned to each fan */
    $control \leftarrow' true'$; $\rho_2 \leftarrow 1/2$; $\sigma_2 \leftarrow \frac{1}{\sqrt{2}}$;
    while   control do
        for   $i \leftarrow 3$ to $k$ do
            $\sigma_i \leftarrow \sqrt{\frac{1}{1+\sigma_{i-1}\rho_{i-1}}}$; $\rho_i \leftarrow \frac{\Pi_{j=1}^{i-1}\sigma_j}{\sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j)}$;
        endfor;
        compute $r_0$ and $r_1$, using Equation (18) and Inequalities from (35) to (38).
        if   there is a solution satisfying the constraints then
            for   $i \leftarrow 2$ to $k$ do
                $r_i \leftarrow \Pi_{j=1}^{i}\sigma_j r_1$;
            endfor;
            $E(\theta, k, L, f) \leftarrow max_{1\leq i\leq k}\{\theta r_i^2 \frac{L^2-(\sum_{j=i}^{k} r_j)^2)}{2}\}$;
            if   $max\_energy\_consu > E(\theta, k, L, f)$ then
                $max\_energy\_consu \leftarrow E(\theta, k, L, f)$;
                $f_0 \leftarrow f$; $\theta_0 \leftarrow \theta$; $k_0 \leftarrow k$;
                $f \leftarrow \lfloor \frac{f}{2} \rfloor$; $\theta \leftarrow \frac{2\pi}{f}$; $k \leftarrow \lfloor \frac{K}{f} \rfloor$;
            else $control \leftarrow' false'$;
            endif;
        else
            $k_L \leftarrow k_0$; $k_H \leftarrow k$;
            while   $(k_L \neq k_H)$ do
                $k \leftarrow \frac{k_L+k_H}{2}$; $f \leftarrow \lfloor \frac{K}{k} \rfloor$; $\theta \leftarrow \frac{2\pi}{f}$;
                for   $i \leftarrow 3$ to $k$ do
                    $\sigma_i \leftarrow \sqrt{\frac{1}{1+\sigma_{i-1}\rho_{i-1}}}$; $\rho_i \leftarrow \frac{\Pi_{j=1}^{i-1}\sigma_j}{\sum_{l=1}^{i-1}(\Pi_{j=1}^{l}\sigma_j)}$;
                endfor;
                compute $r_0$ and $r_1$, using Equation (18) and Inequalities. from (35) to (38).
                if   there is a solution to meet the constraints then
                    for   $i \leftarrow 2$ to $k$ do
                        $r_i \leftarrow \Pi_{j=1}^{i}\sigma_j r_1$;
                    endfor;
                    $f_0 \leftarrow f$; $\theta_0 \leftarrow \theta$; $k_0 \leftarrow k$;
                    $k_L \leftarrow k$; /* looking for the next possible larger k */
                else   $k_H \leftarrow k$;/* looking for the next possible smaller k */
                endif;
            endwhile;
            $control \leftarrow' false'$;
        endif;
    endwhile;
    $max\_energy\_consu \leftarrow E(\theta_0, k_0, L, f_0)$;
    $net\_lifetime \leftarrow \frac{E_{agg}}{max\_energy\_consu}$;
end.
```

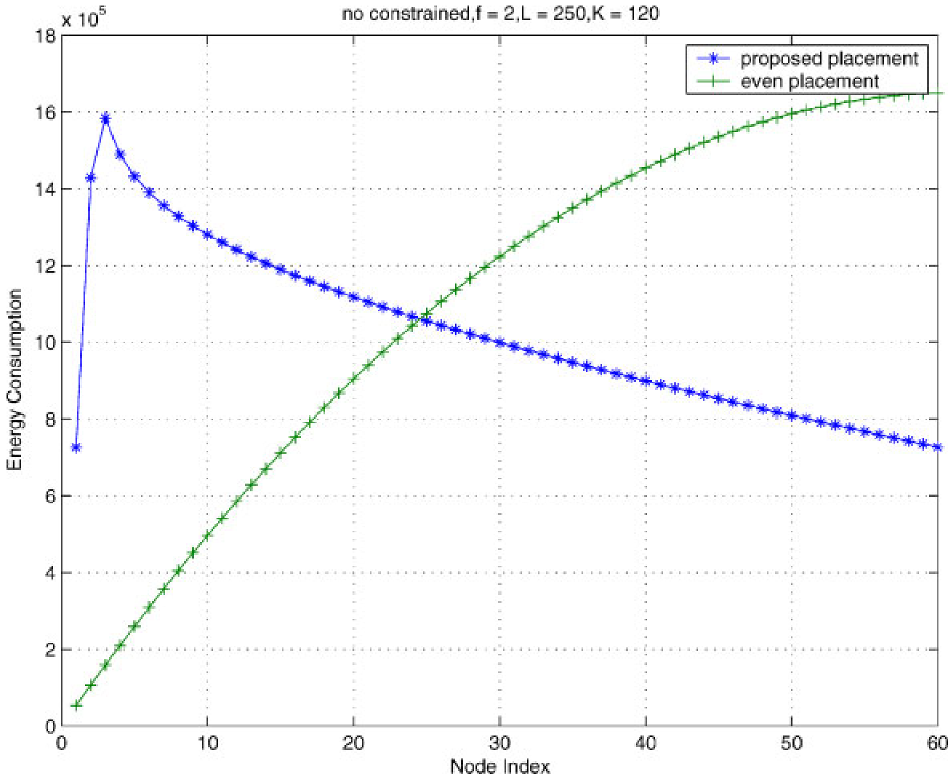


Figure 4. The energy consumption at individual aggregate nodes between the proposed algorithm and the uniform placement schema when $K = 120$ aggregate nodes are placed into two fans, and each fan has 60 aggregate nodes. The energy consumption at each aggregate node indexed from 1 to 60.

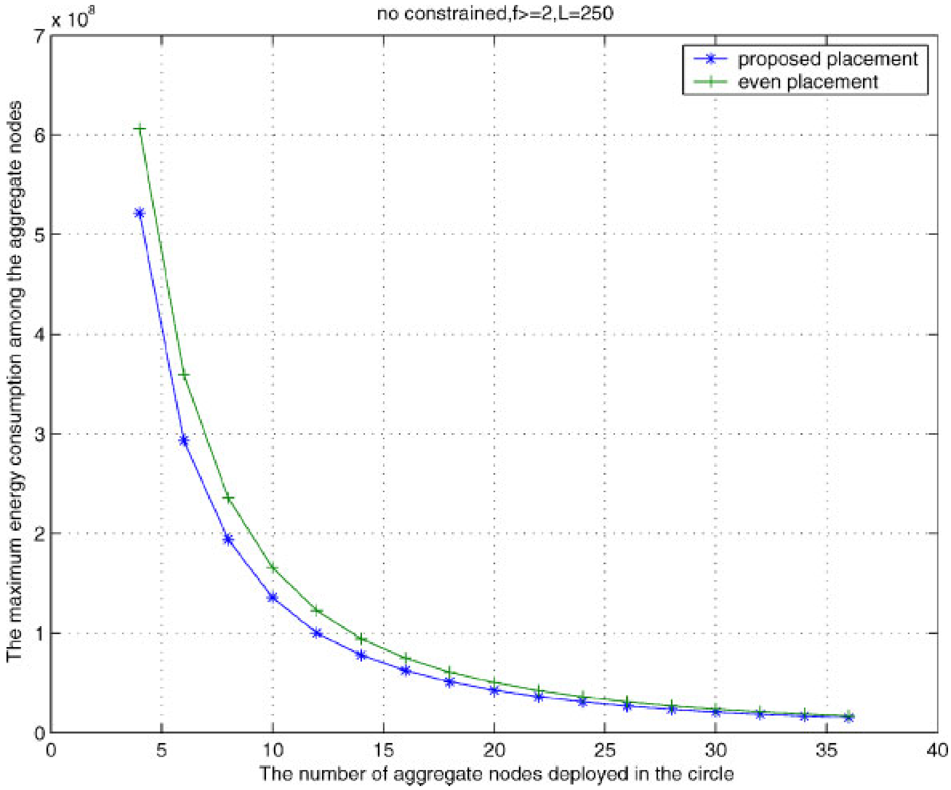


Figure 5. The maximum energy consumption among the aggregate nodes with different *K*s.

5.1. Maximum energy consumption among aggregate nodes

We first examine the energy consumption at each individual aggregate node between the proposed algorithm and the uniform placement schema. Figure 4 depicts the energy consumption of each aggregate node where $K = 120$ aggregate nodes are placed into the circle of consisting of 196 250 sensors, and the circle is partitioned into two fans of angle π and each fan is deployed $k = K/2 = 120/2 = 60$ aggregate nodes. From this figure we can see that aggregate node v_k in the uniform placement schema has the maximum energy consumption, because it not only transmits the sensed data by the sensors in area s_k but also relays the sensed data for the other aggregate nodes in the fan. The experimental results indicate that aggregate node v_3 in the proposed algorithm has the maximum energy consumption. From this figure it can be seen that the energy consumption of all aggregate nodes in the proposed algorithm are more evenly distributed than that of the uniform placement. The difference of consumed energy among different aggregate nodes in the uniform placement is much bigger than that of the proposed algorithm. In terms of the maximum energy consumption among the aggregate nodes, the solution by the proposed algorithm consumes about 5–18% less energy than that by the uniform placement schema. In other words, the network lifetime delivered by the proposed algorithm is around 105–122% of the network lifetime delivered by the uniform placement schema. Experiment results also imply that both algorithms favor placing as many aggregate nodes as possible in a fan to reduce the maximum energy consumption among the aggregate nodes, i.e., both algorithms aim to partition the circle into as few fans as possible, provided that the number of aggregate nodes is given. This is due to the fact that for a given aggregate node, the transmission distance between the aggregate node and its parent aggregate node outweighs the data volume transmitted by the aggregate node significantly, in terms of the transmission energy consumption of the aggregate node.

We then analyze the impact of different values of K on the network lifetime between the proposed algorithm and the uniform placement schema without any distortion constraint for the given sensor network. Figure 5 plots the maximum energy consumption among the aggregate nodes for various values of K. It clearly shows that the proposed algorithm outperforms the uniform placement schema when K is ranged from 4 to 25. The difference of the maximum energy consumption between them begins to diminish, with the growth of the number of aggregate nodes K, and the energy savings by the proposed algorithm is no longer significant in comparison with that by the uniform placement schema when $K \geq 33$. This is consistent with our initial assumption that $K << n$. For this case, $n = \lceil \pi L^2 \rceil = 196\,250$ and $K \leq \lceil \log_2 n \rceil \approx 33$.

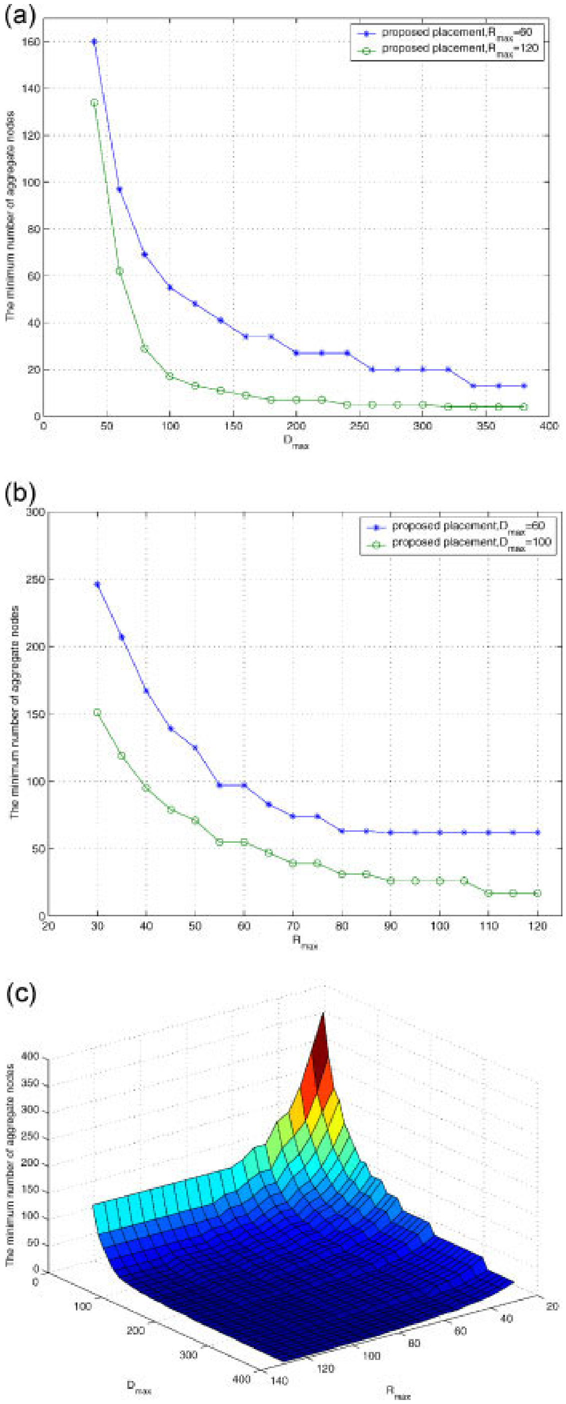


Figure 6. The minimum number of aggregate nodes $K_{\min}(D_{\max}, R_{\max})$ is needed by the proposed algorithm to meet different distortion constraints of $D_{\max}$ and $R_{\max}$: (a) Minimum number of aggregate nodes when $R_{\max} = 60\,m$, $120\,m$ is fixed, (b) Minimum number of aggregate nodes when $D_{\max} = 60\,m$, $100\,m$ is fixed, and (c) Minimum number of aggregate nodes $K_{\min}$ needed with different distortion constraints of $D_{\max}$ and $R_{\max}$.

5.2. Minimum number of aggregate nodes to meet the distortion constraints

What followed is to investigate the minimum number of aggregate nodes needed by the proposed algorithm to meet the distortion constraints of $D_{\max}$ and $R_{\max}$. Figure 6(a) illustrates that the minimum number of aggregate nodes is significantly reduced, when the value of $R_{\max}$ is doubled from 60 m to 120 m, whereas the value of $D_{\max}$ is still between 70 m and 250 m. Figure 6(b) implies that there is a big gap in terms of the number of aggregate nodes needed when $D_{\max}$ is fixed at either 60 m or 100 m, while $R_{\max}$ is ranged from 30 m to 60 m. The experimental results indicate that the proposed placement schema requires more aggregate nodes when $R_{\max}$ is small, while Figure 6 (c) plots the minimum number of aggregate nodes needed for different combinations of $D_{\max}$ and $R_{\max}$. From Figure 6, we can see that the minimum number of aggregate nodes $K_{\min}(D_{\max}, R_{\max})$ is required to meet distortion constraints of both $R_{\max}$ and $D_{\max}$, where $K_{\min}(D_{\max}, R_{\max})$ is the minimum number of aggregate nodes needed by the proposed algorithm. Otherwise, the subnetwork induced by the aggregate nodes will be disconnected.

5.3. Performance evaluation with the distortion constraints

We finally evaluate the proposed algorithm against the uniform placement schema in terms of the maximum energy consumption among the aggregate nodes (or the network lifetime) under different distortion constraints.

Figure 7 plots the maximum energy consumption among the aggregate nodes for both algorithms with various values of $R_{\max}$ and $D_{\max}$. Clearly, the maximum energy consumption among the aggregate nodes by the proposed placement

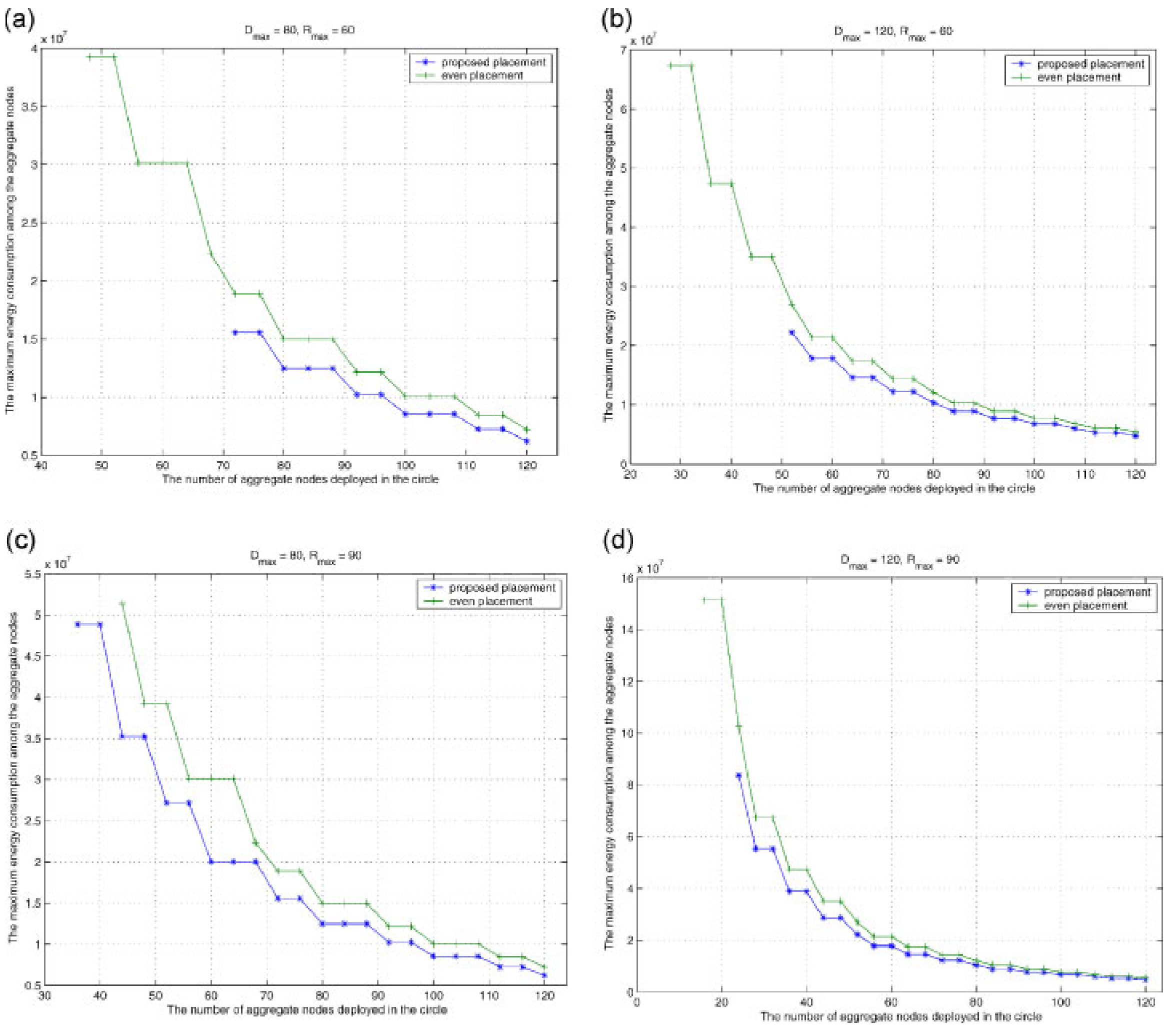


Figure 7. The maximum energy consumption among the aggregate nodes for different combinations of $R_{\max} = 60\,m$ and $90\,m$, $D_{\max} = 80\,m$ and $120\,m$: (a) $R_{\max} = 60\,m$ and $D_{\max} = 80\,m$, (b) $R_{\max} = 60\,m$ and $D_{\max} = 120\,m$, (c) $R_{\max} = 90\,m$ and $D_{\max} = 80\,m$, and (d) $R_{\max} = 90\,m$ and $D_{\max} = 120\,m$.

schema is always less than that by the uniform placement schema, if the number of aggregate nodes is no less than $K_{\min}(D_{\max}, R_{\max})$. On other words, the network lifetime delivered by the proposed algorithm is always longer than that of the uniform placement.

6. CONCLUSIONS

In this paper, we studied the aggregate node placement problem by placing a few powerful aggregate nodes into a dense sensor network such that the lifetime of the resulting network is maximized when performing data gathering. Due to the NP hardness of the problem, we devised a novel heuristic algorithm for it. We also conducted experiments by simulation to evaluate the performance of the proposed algorithm against a commonly used uniform placement schema. The experimental results show that the proposed heuristic significantly outperforms the uniform placement schema in terms of the prolongation of network lifetime when $K << n$.

ACKNOWLEDGMENTS

Authors appreciate Mr Guanjun Ma for implementing the proposed algorithms and evaluating their performance. Also, We thank the anonymous referees for their constructive comments and valuable suggestions which have helped improve the quality and presentation of the paper. It is acknowledged that Weifa Liang's research was funded by the Australian Research Council Discovery grant DP0449431, Yinlong Xu's research is partially supported by the Natural Science Foundation of China under grant no. 60773036, and the work by Junzhou Luo is supported by National Natural Science Foundation of China under grants no. 60903161 and 60903162, National Key Basic Research Program of China under grants no. 2010CB328104, Jiangsu Provincial Natural Science Foundation of China under grants no. BK2008030, Jiangsu Provincial Key Laboratory of Network and Information Security under grants no. BM2003201, and Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under grants no. 93K-9.

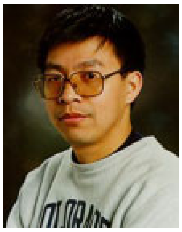
REFERENCES

1. Mainwaring A, Polastre J, Szewczyk R, Culler D, Anderson J. Wireless sensor networks for habitat monitoring. *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, ACM, 2002; 88–97.
2. Singh A, Woo M, Raghavendra CS. Power-aware routing in mobile ad hoc networks. *Proceedings of MobiCom*, ACM/IEEE, 1998; 181–190.
3. Chang J-H, Tassiulas L. Energy conserving routing in wireless ad hoc networks. *Proceedings of INFOCOM'00*, IEEE, 2000.
4. Madden S, Szewczyk R, Franklin MJ, Culler D. Supporting aggregate queries over ad hoc wireless sensor networks. *Proceedings of the 4th IEEE Workshop on Mobile Computing and System Applications*, IEEE, 2002.
5. Yao Y, Gehrke J. The cougar approach to in-network query processing in sensor networks. *ACM SIGMOD Record* 2002; **31**: 9–18.
6. Heinzelman WR, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of the Hawaii International Conference on System Sciences*, IEEE, 2000.
7. Kalpakis K, Dasgupta K, Namjoshi P. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Computer Networks* 2003; **42**: 697–716.
8. Lindsey S, Raghavendra CS. PEGASIS: power-efficient gathering in sensor information systems. *Proceedings of the Aerospace Conference*, IEEE, 2002; 1125–1130.
9. Cha SH, Lee JE, Jo M, Youn HY, Kang S. Policy-based management for self-managing wireless sensor networks. *IEICE Transactions on Communications* 2007; **E90B**: 3024–3033.
10. Lee D-W, Kim J-H. Analysis of optimized aggregation timing in wireless sensor networks. *KSII Transactions on Internet and Information Systems* 2009; **3**: 209–218.
11. Yeo M, Seo D, Yoo J. Data correlation-based clustering algorithm in wireless sensor networks. *KSII Transactions on Internet and Information Systems* 2009; **3**: 331–343.
12. Buragohain C, Agrawal D, Suri S. Power aware routing for sensor databases. *Proceedings of INFOCOM'05*, IEEE, 2005.
13. Cristescu R, Beferull-Lonzano B, Vetterli M. On network correlated data gathering. *Proceedings of INFOCOM'04*, IEEE, 2004.
14. Goel A, Estrin D. Simultaneous optimization for concave costs: single sink aggregation or single source buy-at-bulk. *Proceedings of SODA'03*, ACM-SIAM, 2003; 499–505.
15. Liang W, Liu Y. Online data gathering for maximizing network lifetime in sensor networks. *IEEE Transactions on Mobile Computing* 2007; **6**: 2 –11.
16. von Richenbach P, Wattenhofer R. Gathering correlated data in sensor networks. *Proceedings of 2nd DIALM-POMC*, ACM, October 2004.
17. Sharaf MA, Beaver J, Labrinidis A, Chrysanthis PK. Balancing energy efficiency and quality of aggregate data in sensor networks. *VLDB Journal*, 2004.
18. Dasgupta K, Kukreja M, Kalpakis K. Topology-aware placement and role assignment for energy-efficient information gathering in sensor networks. *Proceedings of the*

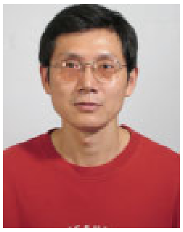
8th IEEE Symposium on Computers and Communications, IEEE, 2003.

19. Cheng P, Chuah C-N, Liu X. Energy-aware node placement in wireless sensor networks. *Proceedings of Globecom'04*, IEEE, 2004; 3210–3214.
20. Ganesan D, Cristescu R, Beferull-Lonzano B. Power-efficient sensor placement and transmission structure for data gathering under distortion constraints. *ACM Transactions on Sensor Networks* 2006; **2**: 155–181.
21. Zhang W, Xue G, Misra S. Fault-tolerant relay node placement in wireless sensor networks: problems and algorithms. *Proceedings of INFOCOM'07*, IEEE, 2007.
22. Wang G, Huang L, Xu H, Wang Y. Reliable relay node placement in wireless sensor network. *Proceedings of the 3rd International Conference on Communications and Networking in China*, IEEE, 2008; 371–375.
23. Pottie GJ. Wireless sensor networks. *Proceedings of Information Theory Workshop*, 1998; 139–140.
24. Gupta H, Das SR, Gu Q. Connected sensor cover: self-organization of sensor networks for efficient query execution. *Proceedings of MobiHoc'03*, ACM, 2003.
25. Awada W, Cardei M. Energy-efficient data gathering in heterogeneous wireless sensor networks. *Proceedings of 2nd International Conference on Wireless and Mobile Communications*, IEEE, 2006; 53–60.
26. Luo J, Hubaux J-P. Joint mobility and routing for lifetime elongation in wireless sensor networks. *Proceedings of INFOCOM'05*, IEEE, 2005.
27. Madden S, Franklin MJ, Hellerstein JM, Hong W. The design of an acquisitional query processor for sensor networks. *Proceedings of SIGMOD'03*, ACM, 2003; 491–502.

AUTHORS' BIOGRAPHIES



Weifa Liang (M'99–SM'01) received his Ph.D. from the Australian National University in 1998, the M.E. degree from the University of Science and Technology of China in 1989, and the B.Sc. degree from Wuhan University, China in 1984, all in Computer Science. He is currently an associate professor in the School of Computer Science at the Australian National University. His research interests include design and analysis of energy-efficient routing protocols for wireless *ad hoc* and sensor networks, information processing in wireless sensor networks, routing protocol design for WDM optical networks, design and analysis of parallel and distributed algorithms, query optimization, and graph theory. He is a senior member of the IEEE.



Yinlong Xu received his B.S. in Mathematics from Peking University in 1983, and M.S. and Ph.D. in Computer Science from University of Science and Technology of China(USTC) in 1989 and 2004, respectively. He is currently a professor with the School of Computer Science and Technology at USTC. Prior to that, he served the Department of Computer Science and Technology at USTC as an assistant professor, a lecturer, and an associate professor. Currently, he is leading a group of research students in doing some networking and high performance computing research. His research interests include network coding, wireless network, combinatorial optimization, design and analysis of parallel algorithm, etc. He received the Excellent Ph.D. Advisor Award of Chinese Academy of Sciences in 2006.

Jiugen Shi received his Ph.D. and M.Sc. in Electrical Engineering from University of Science and Technology of China in 2004 and 2000, respectively, and B.Sc. in Physics from Anqin Normal College in China in 1984. He is currently an associate professor in the School of Computer and Information at Hefei University of Technology in China. His main research interests include networked control systems, wireless ad hoc and sensor networks, and vehicular communication networks.

Junzhou Luo received his B.S. degree in Applied Mathematics from Southeast University in 1982. From then on he has been a faculty member at the School of Computer Science and Engineering, Southeast University. He got his M.S. and Ph.D. degree in Computer Network both from Southeast University in 1992 and in 2000, respectively. He now is a full professor and Dean at the School of Computer Science and Engineering in Southeast University in China. He is an IEEE member and co-chair of IEEE SMC Technical Committee on Computer Supported Cooperative Work in Design. His research interests are the next generation network architecture, protocol engineering, network security and management, wireless LAN, and grid and service computing.