Approximation Algorithm for Base Station Placement in Wireless Sensor Networks

Yi Shi and Y. Thomas Hou

The Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA

***Abstract*— Base station location has significant impact on network lifetime performance for a sensor network. For a multi-hop sensor network, this problem is particular challenging as we need to jointly consider base station placement and data routing strategy to maximize network lifetime performance. This paper presents an approximation algorithm that can guarantee $(1-\varepsilon)$ optimal network lifetime performance for base station placement problem with any desired error bound $\varepsilon > 0$. The proposed $(1-\varepsilon)$ optimal approximation algorithm is based on several novel techniques that enable to reduce an infinite search space to a *finite-element* search space for base station location. The first technique used in this reduction is to discretize cost parameters (with performance guarantee) associated with energy consumption. Subsequently, the continuous search space can be broken up into a finite number of subareas. The second technique is to exploit the cost property of each subarea and represent it by a novel notion called "fictitious cost point," each with guaranteed cost bounds. This approximation algorithm offers a simpler and in most cases practically faster algorithm than a state-of-the-art algorithm and represents the best known result to this important problem.**

***Index Terms*— Theory, approximation algorithm, base station placement, network lifetime, sensor network.**

I. Introduction

An important characteristic for wireless sensor networks is the so-called network lifetime performance, which is highly dependent upon the physical topology of the network. This is because energy expenditure at a node to transmit data to another node not only depends on the data bit rate, but also on the physical distance between these two nodes. Consequently, it is important to understand the impact of location related issues on network lifetime performance and to optimize topology during network deployment stage.

This paper considers the important base station placement problem for a given sensor network such that network lifetime can be maximized. Specifically, we consider the following problem. Given a sensor network with each node i producing sensing data at a rate of r_i, where should we place the base station in this sensor network such that all the data can be forwarded to the base station (via multi-hop and multi-path if necessary) such that the network lifetime is maximized?

In Section V, we give a comprehensive review of related work on network lifetime and node placement problems and contrast their differences with this work. The most relevant work on this problem is done by Efrat, Har-Peled, and Mitchell in [7], which represents the state-of-the-art result on this problem. However, the computational complexity of the algorithm in [7] is higher than the one to be presented in this paper for most cases.

The main idea in our approximation algorithm is to exploit a clever way to discretize cost parameters associated with energy consumption with tight upper and lower bounds. As a result, we can divide the continuous search space into a finite number of subareas. By further exploiting the cost property of each subarea, we conceive a novel idea to represent each subarea with a so-called "fictitious cost point," which is an N-tuple cost vector with each component representing the upper bound of cost to a sensor node in the network. Based on these ideas, we can successfully reduce an infinite search space for base station location into finite "points" upon which we can apply a linear programming (LP) to find the corresponding achievable network lifetime and data routing solution for each of these points. By comparing the achievable network lifetime among all the fictitious cost points, we show that the largest is $(1-\varepsilon)$ optimal. We also show that placing the base station at *any point* in the subarea corresponding to the best fictitious cost point is $(1-\varepsilon)$ optimal. We analyze the complexity of our approximation algorithm and show that it is practically faster than the algorithm proposed in [7] for most cases, which was the best known result on this problem. As a result, the algorithm presented in this paper represent the best known result to the base station placement problem.

The rest of this paper is organized as follows. Section II presents the network model used in this study and describes the base station placement problem. Section III presents the main result of this paper, which is a $(1-\varepsilon)$ approximation algorithm for the base station placement problem. In Section IV, we present some numerical results illustrating the efficacy of the proposed algorithm. Section V reviews related work and Section VI concludes this paper.

II. Network Model and Problem Description

A. Network Model

We consider a static sensor network consisting of a set of $\mathcal{N}$ sensor nodes deployed over a two-dimensional area. The location of each sensor node is fixed and the initial energy on sensor node i is denoted as e_i. Each sensor node i generates data at a rate r_i. We assume there is one base station that needs to be deployed to collect sensing data.

This work has been supported in part by NSF Grant CNS-0347390.

1-4244-1268-4/07/$25.00 ©2007 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE SECON 2007 proceedings.

In this paper, we focus on the energy consumption due to communications (i.e., data transmission and reception). Suppose sensor node i transmits data to sensor node j with a rate of f_{ij} b/s. Then we model the transmission power at sensor node i as

$$p_{ij}^t = c_{ij} \cdot f_{ij} \ . \tag{1}$$

c_{ij} is the cost on link (i, j), and can be modeled as

$$c_{ij} = \beta_1 + \beta_2 \cdot d_{ij}^{\alpha} \ , \tag{2}$$

where β_1 and β_2 are constant coefficients, d_{ij} is the physical distance between sensor nodes i and j, α is the path loss index, and $2 \leq \alpha \leq 4$ [8].

The power consumption at the receiving sensor node i can be modeled as [8]:

$$p_i^r = \rho \cdot \sum_{k \in \mathcal{N}}^{k \neq i} f_{ki} \ , \tag{3}$$

where f_{ki} (also in b/s) is the incoming bit-rate received by sensor i from sensor k. It is easy to observe from (1), (2), and (3) that the *location for the base station* and *data routing* in the network both have a profound impact on energy consumption among the nodes.

The above transmission and reception energy model assumes a contention-free MAC protocol, where interference from simultaneous transmission can be effectively avoided. For deterministic rate traffic pattern model in this paper, a contention-free MAC protocol is fairly easy to design (see, e.g., [13]) and its discussion is beyond the scope of this paper. Table I lists all notation used in this paper.

B. Problem Description

The focus of this paper is to investigate how to optimally place a base station to collect data in a sensor network so that the network lifetime can be maximized. The network lifetime is defined as the time until any sensor node uses up its energy. To achieve optimality, the data generated by each sensor node is allowed to be routed to the base station via multi-hop or multi-path. Also, power control at a node is allowed.

Assume that base station B is located at a point p. Denote (x_B, y_B) the position of point p and T the network lifetime. Then a feasible routing solution achieving this network lifetime T should satisfy both flow balance and energy constraints at each sensor node. These constraints can be formally stated as follows. Denote f_{ij} and f_{iB} the data rates from sensor node i to sensor node j and base station B, respectively (since we allow multi-path). Then the flow balance for each sensor node i is

$$\sum_{k \in \mathcal{N}}^{k \neq i} f_{ki} + r_i = \sum_{j \in \mathcal{N}}^{j \neq i} f_{ij} + f_{iB} \ , \tag{4}$$

i.e., the sum of total incoming flow rates plus self-generated data rate is equal to the sum of total outgoing flow rates. The energy constraint for each sensor node i is

$$\sum_{k \in \mathcal{N}}^{k \neq i} \rho \cdot f_{ki} T + \sum_{j \in \mathcal{N}}^{j \neq i} c_{ij} \cdot f_{ij} T + c_{iB}(p) \cdot f_{iB} T \leq e_i \ , \tag{5}$$

TABLE I
NOTATION

Symbols	Definitions
$\mathcal{A}$	The search space for the base station, which can be the smallest enclosing disk to cover all sensor nodes
$\mathcal{A}_m$	The m-th subarea in the search space
B	The base station
c_{ij} (or $c_{iB}(p)$)	Power consumption coefficient for transmitting data from sensor i to sensor j (or base station B at point p)
$C_{iB}^{\min}, C_{iB}^{\max}$	Lower and upper bounds of $c_{iB}(p)$
$C[h]$	$=\beta_1(1+\varepsilon)^h$, the transmission cost for the h-th circle
d_{ij} (or d_{iB})	Distance between sensor i and sensor j (or base station B)
e_i	Initial energy at sensor i
f_{ij} (or f_{iB})	Data rate from sensor i to sensor j (or base station B)
H_i	Number of circles at sensor node i for a given ε
K	Number of total circles for a given ε
M	Number of subareas for a given ε
N	Number of sensor nodes in the network
$\mathcal{N}$	Set of sensor nodes in the network
$O_{\mathcal{A}}$	The center of $\mathcal{A}$
p_m	Fictitious cost point for the m-th subarea
p_{opt}	The best location among all points in $\mathcal{A}$
p^*	The best location among M fictitious cost points
p_ε	A point in the subarea corresponding to p^*
r_i	Sensing data rate produced at sensor i
$R_{\mathcal{A}}$	The radius of $\mathcal{A}$
T_m	Maximum achievable network lifetime by p_m
T_{opt}	Optimal network lifetime achieved by p_{opt}
T^*	$= \max\{T_m : m = 1, 2, \cdots, M\}$
T_ε	$(1-\varepsilon)$ optimal network lifetime achieved by p_ε
V_{ij} (or V_{iB})	Total data volume from sensor i to sensor j (or base station B)
α	Path loss index
β_1, β_2	Constant terms in transmission power modeling
ε	Desired small approximation error, $\varepsilon > 0$
ρ	Power consumption coefficient for receiving data
ψ_{opt}	The best routing solution for p_{opt}
ψ^*	Routing solution for p^* obtained via LP

i.e., total consumed energy due to receiving and transmission over time T cannot exceed its initial energy e_i. By (2), we have

$$c_{iB}(p) = \beta_1 + \beta_2 \left[\sqrt{(x_B - x_i)^2 + (y_B - y_i)^2} \right]^{\alpha} \ ,$$

which is a non-linear function of base station location (x_B, y_B).

It is not hard to see that the above formulation will lead to an optimization problem in the form of *non-convex programming*.

III. Algorithm Design

A. Optimal Routing for A Given Base Station Location

As discussed earlier, the maximum network lifetime depends on both base station location and data routing. To start with, we show that for a *given* base station location, we can find the corresponding maximum achievable network lifetime and optimal routing via a single linear programming (LP) [1].

The objective function is network lifetime T and the constraints are given in (4) and (5). Multiply both sides of (4) by T and denote $V_{ij} = f_{ij}T$ and $V_{iB} = f_{iB}T$, where V_{ij} (or V_{iB}) can be interpreted as the total data volume from sensor node

i to sensor node j (or base station B). We have

$$\begin{aligned}
\text{Max} \quad & T \\
\text{s.t.} \quad & \sum_{k\in\mathcal{N}}^{k\neq i} V_{ki} + r_i T - \sum_{j\in\mathcal{N}}^{j\neq i} V_{ij} - V_{iB} = 0 \qquad (i \in \mathcal{N}) \\
& \sum_{k\in\mathcal{N}}^{k\neq i} \rho V_{ki} + \sum_{j\in\mathcal{N}}^{j\neq i} c_{ij} V_{ij} + c_{iB}(p) V_{iB} \leq e_i \qquad (i \in \mathcal{N}) \quad (6) \\
& T, V_{ij}, V_{iB} \geq 0 \qquad (i, j \in \mathcal{N}, i \neq j) \; .
\end{aligned}$$

It should be note that for a given base station location, $c_{iB}(p)$'s are constants. Once we solve the above LP, we can obtain optimal routing solution for f_{ij} and f_{iB} by $f_{ij} = \frac{V_{ij}}{T}$ and $f_{iB} = \frac{V_{iB}}{T}$.

B. Our Approach

Although for a given base station location, we can find the corresponding maximum achievable network lifetime via a single LP, it is not possible to examine all points (infinite) in the two-dimensional plane and select the point with the maximum network lifetime among all the points. Our approach to this search problem is to narrow down the search space to a *finite* set of points, among which there exist at least one point with $(1-\varepsilon)$ optimal network lifetime.

As a first step, we show that it is only necessary to consider points inside the so-called *smallest enclosing disk* (SED),[1] which is a unique disk with the smallest radius that contains all the N sensor nodes in the network and can be found in $O(N)$ time [5]. This is formally stated in the following lemma, which can be easily proved via contradiction [12].

Lemma 1: To maximize network lifetime, the base station location must be within the smallest enclosing disk $\mathcal{A}$ that covers all the N sensor nodes in the network.

Now we have narrowed down the search space for base station B from a two-dimensional plane to SED $\mathcal{A}$. However, the number of points in $\mathcal{A}$ remains infinite. It is tempting to divide $\mathcal{A}$ into small subareas (e.g., a grid-like structure), $\mathcal{A}_1, \mathcal{A}_2, \cdots$, up to say $\mathcal{A}_M$, i.e.,

$$\mathcal{A} = \bigcup_{m=1}^{M} \mathcal{A}_m \; .$$

When each subarea is sufficiently small (i.e., M is sufficiently large), we can use some point $q_m \in \mathcal{A}_m$ to represent $\mathcal{A}_m$, $m = 1, 2, \cdots, M$. By applying an LP on each of the M points, we can select the best location among all points and obtain a good solution for base station placement. However, such approach is *heuristic* at best and does not provide any *theoretical guarantee* on performance.

The key to provide a theoretical guarantee on performance is to divide the subarea in such a way that tight bounds can be guaranteed on any point in the subarea. If this is possible, then we may be able to exploit such properties and develop an approximation algorithm that yields *provably* $(1-\varepsilon)$ optimal network lifetime performance. In the following section, we show a novel technique to divide SED $\mathcal{A}$ into subareas where each subarea can be represented by a point with a set of tight bounds. Consequently, a $(1-\varepsilon)$ optimal approximation algorithm can be developed.

[1] In fact, we can consider points in an even smaller area, i.e., the convex hull of all sensor nodes. However, using convex hull cannot reduce the order of complexity of our algorithm. On the other hand, the use of SED can simplify the discussion.

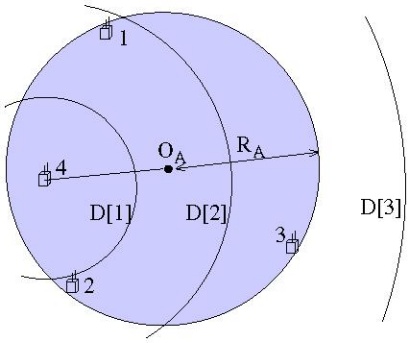

Fig. 1. A sequence of circles with increasing costs with center at node 4.

C. Subarea Division and Fictitious Cost Points

1) Subarea Division: The proposed subarea division (with guaranteed performance bounds) hinges upon discretization of the cost parameters. A close look at the energy constraint in (6) suggests that the location of the base station is *embedded* in the cost parameter $c_{iB}(p)$'s. In other words, if we can discretize these cost parameters, we may also discretize the location for the base station.

Since the search space is narrowed down to the SED $\mathcal{A}$, we can limit the range for the distance between a sensor node i to the possible location for the base station. Denote $O_{\mathcal{A}}$ and $R_{\mathcal{A}}$ the origin and radius of the SED $\mathcal{A}$. For each sensor node $i \in \mathcal{N}$, denote $D_{i,O_{\mathcal{A}}}$ the distance from sensor node i to the origin of disk $\mathcal{A}$ (see node 4 in Fig. 1 as an example). Denote $D_{iB}^{\min}$ and $D_{iB}^{\max}$ the minimum and maximum distances between sensor node i and possible location for the base station B, respectively. Then we have

$$\begin{aligned}
D_{iB}^{\min} &= 0 \; , \\
D_{iB}^{\max} &= D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}} \; .
\end{aligned}$$

Corresponding to $D_{iB}^{\min}$ and $D_{iB}^{\max}$, denote $C_{iB}^{\min}$ and $C_{iB}^{\max}$ the minimum and maximum cost between sensor node i and base station B, respectively. Then by (2), we have

$$C_{iB}^{\min} = \beta_1 \; , \qquad (7)$$

$$C_{iB}^{\max} = \beta_1 + \beta_2 (D_{iB}^{\max})^{\alpha} = \beta_1 + \beta_2 (D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha} \; . \qquad (8)$$

Given the range of $d_{iB} \in [D_{iB}^{\min}, D_{iB}^{\max}] = [0, D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}}]$ for each sensor node i, we now show how to divide disk $\mathcal{A}$ into a finite number of subareas with the distance of each subarea to sensor node $i \in \mathcal{N}$ meeting some tight bounds. Specifically, from a sensor node i, we draw a sequence of circles centered at this sensor node, each with increasing radius $D[1], D[2], \cdots, D[H_i]$ corresponding to costs

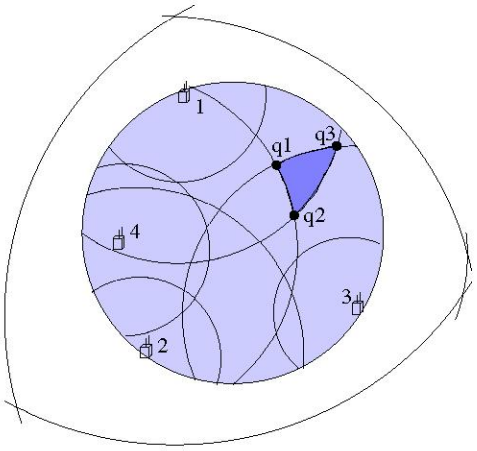

Fig. 2. An example of subareas within disk $\mathcal{A}$ that are obtained by intersecting arcs from different circles.

$C[1], C[2], \cdots, C[H_i]$ that are defined as follows.

$$C[h] = C_{iB}^{\min}(1+\varepsilon)^h = \beta_1(1+\varepsilon)^h \quad (1 \le h \le H_i) \qquad (9)$$

The geometric series $C[h]$ (with a factor of $(1+\varepsilon)$) is carefully chosen and will offer tight performance bounds for any point in a subarea (more on this later). The number of required circles H_i can be determined by having the last circle in the sequence (with radius $D[H_i]$) to completely contain disk $\mathcal{A}$, i.e. $D[H_i] \ge D_{iB}^{\max}$, or equivalently,

$$C[H_i] \ge C_{iB}^{\max} .$$

This will set the limit H_i as

$$\begin{aligned} H_i &= \left\lceil \frac{\ln(C_{iB}^{\max}/C_{iB}^{\min})}{\ln(1+\varepsilon)} \right\rceil \\ &= \left\lceil \frac{\ln\left(1+\frac{\beta_2}{\beta_1}(D_{i,O_{\mathcal{A}}}+R_{\mathcal{A}})^{\alpha}\right)}{\ln(1+\varepsilon)} \right\rceil . \end{aligned} \qquad (10)$$

For example, for node 4 in Fig. 1, we have $H_4 = 3$, i.e., $D[3]$ is the circle centered at node 4 that will completely contain the disk. As a result, with sensor node i as center, we have a total of H_i circles, each with cost $C[h]$, $h = 1, 2, \cdots, H_i$.

The above partitioning of SED $\mathcal{A}$ is with respect to a specific node i. We now perform the above process for *all* sensor nodes. These intersecting circles will cut disk $\mathcal{A}$ into a finite number of *non-uniform* subareas, with the boundaries of each subarea being either an arc (with a center at some sensor node i and some cost $C[h]$, $1 \le h < H_i$) or an arc from SED $\mathcal{A}$. As an example, the SED $\mathcal{A}$ in Fig. 2 is now divided into 28 subareas.

We now show that under our subarea partitioning technique, for any point in a given subarea, its cost to each sensor node in the network can be *tightly* bounded.

Note that in regard to every sensor node i, a subarea $\mathcal{A}_m$ must be within some ring (with a center at sensor node i). Denote the index of this ring (w.r.t. sensor node i) as $h_i(\mathcal{A}_m)$. So when the base station B is at any point $p \in \mathcal{A}_m$, we have

$$C[h_i(\mathcal{A}_m)-1] \le c_{iB}(p) \le C[h_i(\mathcal{A}_m)] , \qquad (11)$$

where we define $C[0] = C_{iB}^{\min} = \beta_1$. Since $\frac{C[h_i(\mathcal{A}_m)]}{C[h_i(\mathcal{A}_m)-1]} = 1+\varepsilon$ by (9), we have a very tight upper and lower bounds for $c_{iB}(p)$ (now we see the benefit of our discretization technique for cost and distance).

2) Fictitious Cost Point: To represent each subarea $\mathcal{A}_m$, $m = 1, 2, \cdots, M$, with a point, we introduce a novel concept called *fictitious cost point*.

Definition 1: Denote the fictitious cost point for subarea $\mathcal{A}_m$ ($m = 1, 2, \cdots, M$) as p_m, which is represented by an N-tuple vector with its i-th element ($i = 1, 2, \cdots, N$) being upper cost bound for any point in subarea $\mathcal{A}_m$ to the i-th sensor node in the network.

That is, the N-tuple cost vector for fictitious cost point p_m is $[c_{1B}(p_m), c_{2B}(p_m), \cdots, c_{NB}(p_m)]$, with the i-th component $c_{iB}(p_m)$ being

$$c_{iB}(p_m) = C[h_i(\mathcal{A}_m)] , \qquad (12)$$

where $h_i(\mathcal{A}_m)$ is determined by (11).

As an example, the fictitious cost point for subarea with corner points (q_1, q_2, q_3) in Fig. 2 can be represented by 4-tuple cost vector $[c_{1B}(p_m), c_{2B}(p_m), c_{3B}(p_m), c_{4B}(p_m)] = [C[2], C[3], C[2], C[3]]$, where the first component $C[2]$ represents an upper bound of cost for any point in this subarea to sensor node 1, the second component $C[3]$ represents an upper bound of cost (which is loose here) for any point in this subarea to sensor node 2, and so forth.

We emphasize that the reason we use the word "fictitious" is that a fictitious cost point p_m may not be mapped to a *physical* point within the corresponding subarea $\mathcal{A}_m$. This happens when there does not exist a physical point in subarea $\mathcal{A}_m$ that has its costs to all the N sensor nodes *equal* (one by one) to the respective N-tuple cost vector embodied by p_m *simultaneously*. As an example, any point within the dark subarea bounded by corner points q_1, q_2, and q_3 cannot have its costs to the four sensor nodes in the network equal to the respective element in $[C[2], C[3], C[2], C[3]]$ simultaneously, where $[C[2], C[3], C[2], C[3]]$ is the cost vector of the fictitious cost point for this subarea.

Using "fictitious" points to represent subareas and to construct a finite search space is a key step in design our low complexity approximation algorithm. As a contrast, in [7], the authors use physical points to construct a finite search space. Thus, they cannot discretize cost directly. Instead, they consider how to discretize transmission energy, flow rate, and network lifetime such that cost can be discretized. The number of discretized costs is the product of the numbers of discretized transmission energies, flow rates, and network lifetimes. We will show that the complexity associated with the algorithm in [7] is higher than ours for most cases.

The following important property for fictitious cost point p_m will be used in the proof of $(1-\varepsilon)$ optimal of the approximation algorithm.

Property 1: For any point $p \in \mathcal{A}_m$ and the corresponding fictitious cost point p_m, we have

$$c_{iB}(p_m) \leq (1+\varepsilon)c_{iB}(p) \ .$$

Proof: By (11) and definition of fictitious cost point p_m (see (12)), we have

$$\begin{aligned} c_{iB}(p_m) &= C[h_i(\mathcal{A}_m)] = (1+\varepsilon)\cdot C[h_i(\mathcal{A}_m)-1] \\ &\leq (1+\varepsilon)\cdot c_{iB}(p) \ , \end{aligned}$$

where the inequality follows from (11). This completes the proof. ■

D. Summary of Algorithm and Example

By discretizing the cost parameters and the corresponding distances, we have partitioned the search space (SED $\mathcal{A}$) into a finite number (M) subareas. By introducing the concept of fictitious cost points, we can represent each subarea with a point. As a result, we can now readily apply the LP approach discussed in Section III-A to examine each point and choose the point that offers the maximum network lifetime. The complete approximation algorithm is outlined below.

Algorithm 1:

1) *Find the smallest enclosing disk $\mathcal{A}$ that covers all the N nodes.*
2) *Within $\mathcal{A}$, compute the lower and upper cost bounds C_{iB}^{min} and C_{iB}^{max} for each node $i \in \mathcal{N}$ by (7) and (8).*
3) *For a given $\varepsilon > 0$, define a sequence of costs $C[1]$, $C[2], \cdots, C[H_i]$ by (9), where H_i can by calculated by (10).*
4) *For each node i, draw a sequence of $H_i - 1$ circles corresponding to cost $C[h]$ centered at node i, $1 \leq h < H_i$. The intersection of these circles within disk $\mathcal{A}$ will divide $\mathcal{A}$ into M subareas $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_M$.*
5) *For each subarea $\mathcal{A}_m$, $1 \leq m \leq M$, define a fictitious cost point p_m by an N-tuple cost vector $[c_{1B}(p_m), c_{2B}(p_m), \cdots, c_{NB}(p_m)]$, where $c_{iB}(p_m)$ is defined in (12).*
6) *For each fictitious cost point p_m, $1 \leq m \leq M$, apply the LP in Section III-A and obtain the achievable network lifetime T_m.*
7) *Choose the fictitious cost point p^* that offers the maximum network lifetime among these M fictitious cost points. The base station can be placed at any point p_ε within the subarea corresponding to p^*.*
8) *For point p_ε, apply the LP in Section III-A and obtain $(1-\varepsilon)$ optimal network lifetime T_ε.*

Remark. Note that if there is additional area constraint on base station placement, then the above algorithm can be easily extended to accommodate this constraint. That is, we can define $\mathcal{A}$ in Step 1 to be the intersection between SED and the allowed area for base station placement.

Example. We use a small 3-node network to illustrate the steps of the approximation algorithm. The location, data rate, and initial energy for each sensor are shown in Table II, where the units of distance, rate, and energy are all normalized. Also,

TABLE II
SENSOR LOCATIONS, DATA RATE, AND INITIAL ENERGY OF THE EXAMPLE SENSOR NETWORK

Node Index	(x_i, y_i)	r_i	e_i
1	$(0.1, 0.5)$	0.8	390
2	$(1.1, 0.7)$	1.0	400
3	$(0.4, 0.1)$	0.5	130

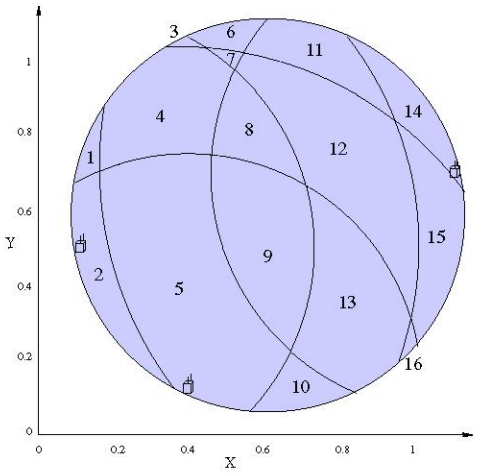

Fig. 3. The SED is divided into 16 subareas in the example sensor network.

we set $\alpha = 2$, $\beta_1 = 1$, $\beta_2 = 0.5$, and $\rho = 1$ under the normalized units. For illustration, we set the error bound to $\varepsilon = 0.2$.[2]

Step 1. We identify SED $\mathcal{A}$ with origin $O_{\mathcal{A}} = (0.61, 0.57)$ and radius $R_{\mathcal{A}} = 0.51$ (see Fig. 3).

Step 2. We first have $D_{i,O_{\mathcal{A}}} = R_{\mathcal{A}} = 0.51$ for each node i, $1 \leq i \leq 3$. We then find the lower and upper bounds of c_{iB} for each node i, $1 \leq i \leq 3$, as follows.

$$\begin{aligned} C_{iB}^{\min} &= \beta_1 = 1 \\ C_{iB}^{\max} &= \beta_1 + \beta_2 (D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^\alpha \\ &= 1 + 0.5 \cdot (0.51 + 0.51)^2 = 1.52 \end{aligned}$$

Step 3. For each node i, $1 \leq i \leq 3$, we find

$$\begin{aligned} H_i &= \left\lceil \frac{\ln\left(1 + \frac{\beta_2}{\beta_1}(D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^\alpha\right)}{\ln(1+\varepsilon)} \right\rceil \\ &= \left\lceil \frac{\ln\left(1 + \frac{0.5}{1}(0.51+0.51)^2\right)}{\ln(1+0.2)} \right\rceil = 3 \end{aligned}$$

and

$$\begin{aligned} C[1] &= \beta_1(1+\varepsilon) = 1 \cdot (1+0.2) = 1.20 \ , \\ C[2] &= \beta_1(1+\varepsilon)^2 = 1 \cdot (1+0.2)^2 = 1.44 \ , \\ C[3] &= \beta_1(1+\varepsilon)^3 = 1 \cdot (1+0.2)^3 = 1.73 \ . \end{aligned}$$

[2]This ε is used for illustration here only. In our numerical results in Section IV, we use $\varepsilon = 0.05$. for all computations.

Step 4. We draw circles centered at each node i, $1 \le i \le 3$, and with cost $C[h]$, $1 \le h < H_i = 3$, to divide the whole disk $\mathcal{A}$ into 16 subareas $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{16}$.

Step 5. We define a fictitious cost point p_m for each subarea $\mathcal{A}_m$, $1 \le m \le 16$. For example, for fictitious cost point p_1, we define the 3-tuple cost vector as $[c_{1B}(p_1), c_{2B}(p_1), c_{3B}(p_1)] = [C[1], C[3], C[2]] = [1.20, 1.73, 1.44]$.

Step 6. We apply LP in Section III-A on these 16 fictitious cost points and obtain the achievable network lifetime.

Step 7. Since the fictitious cost point $p^* = p_9$ has the maximum achievable network lifetime 226.47 among all 16 fictitious cost points, we can place the base station at any point in subarea $\mathcal{A}_9$, e.g. $p_\varepsilon = (0.6, 0.6)$.

Step 8. We apply LP in Section III-A on p_ε and obtain a $(1-\varepsilon)$ optimal network lifetime $T_\varepsilon = 227.07$. This completes the algorithm.

E. Proof Sketch and Complexity Analysis

Denote p_{opt} the optimal location for base station placement (unknown) and T_{opt} and ψ_{opt} the corresponding maximum network lifetime and data routing solution. For T_ε obtained in Algorithm 1, we have the following theorem.

Theorem 1: The achieved network lifetime T_ε is $(1-\varepsilon)$ optimal, i.e., $T_\varepsilon \ge (1-\varepsilon)T_{opt}$.

A complete proof of the above theorem can be found in [12]. We give a sketch of the proof here. Suppose that p_{opt} is in a subarea $\mathcal{A}_m$. By Property 1, for the corresponding fictitious cost point p_m, we have $c_{iB}(p_m) \le (1+\varepsilon)c_{iB}(p_{\text{opt}})$. It can be shown that under p_m and ψ_{opt}, the total consumed energy at each node i is no more than e_i at time $(1-\varepsilon)T_{\text{opt}}$, i.e., the network lifetime under p_m and ψ_{opt} is at least $(1-\varepsilon)T_{\text{opt}}$. Then for the achievable network lifetime T_m by p_m under its optimal data routing, we have

$$T_m \ge (1-\varepsilon)T_{\text{opt}} \,. \tag{13}$$

Note that the achievable network lifetime T^* by p^* is the largest among the achievable network lifetimes by all fictitious cost points (see Step 7 in Algorithm 1). In particular, $T^* \ge T_m$. By (13), we have

$$T^* \ge T_m \ge (1-\varepsilon)T_{\text{opt}} \,. \tag{14}$$

By Definition 1, for a point p_ε in the corresponding subarea to fictitious cost point p^*, we have $c_{iB}(p_\varepsilon) \le c_{iB}(p^*)$. Denote ψ^* the optimal data routing solution for p^*. It can be shown that under p_ε and ψ^*, the total consumed energy at each node i is no more than e_i at time T^*, i.e., the network lifetime under p_ε and ψ^* is at least T^*. Then for the achievable network lifetime T_ε by p_ε under its optimal data routing, we have $T_\varepsilon \ge T^*$. Thus, by (14), we have

$$T_\varepsilon \ge T^* \ge (1-\varepsilon)T_{\text{opt}} \,,$$

which is the result in Theorem 1.

The complexity of Algorithm 1 can be measured by the number of LPs that need to be solved, which is equal to the number of subareas M. The boundaries of each subarea being either an arc centered at some sensor node i (with some cost $C[h]$, $1 \le h < H_i$, where H_i is defined in (10)) or an arc of disk $\mathcal{A}$. Since there are $H_i - 1$ circles radiating from each sensor node i and one circle for disk $\mathcal{A}$, the number of circles is $K = 1 + \sum_{i \in \mathcal{N}}(H_i - 1)$. The maximum number of subareas that can be obtained by K circles is upper bounded by [5]

$$M \le K^2 - K + 2 \,. \tag{15}$$

We have $M = O(K^2) = O((\sum_{i \in \mathcal{N}} H_i)^2) = O((\frac{N}{\varepsilon})^2)$.

As for comparison, the complexity of the approximation algorithm proposed in [7] is given in Section IV. Numerical comparison on complexity for some network topologies are also given there.

TABLE III

EACH NODE'S CARTESIAN COORDINATES, DATA GENERATION RATE AND INITIAL ENERGY FOR A SMALL 10-NODE NETWORK.

(x_i, y_i)	r_i	e_i	(x_i, y_i)	r_i	e_i
$(0.81, 0.86)$	0.7	390	$(0.25, 0.71)$	0.4	400
$(0.47, 0.44)$	1.0	440	$(0.28, 0.03)$	0.6	330
$(0.25, 0.36)$	0.2	440	$(0.48, 0.22)$	0.1	300
$(0.53, 0.16)$	0.8	410	$(0.66, 0.52)$	0.2	210
$(0.91, 0.86)$	0.1	320	$(0.44, 0.21)$	0.9	330

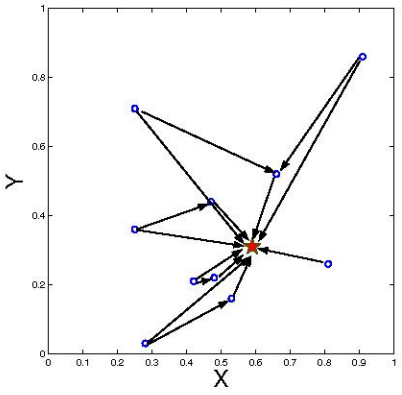

Fig. 4. A schematic showing the routing solution for the 10-node network with base station being placed at $(0.59, 0.31)$.

IV. NUMERICAL RESULTS

In this section, we apply the approximation algorithm on various network topology and use numerical results to demonstrate its efficacy. The units of distance, rate, and energy are all normalized appropriately. The normalized parameters in energy consumption model are $\beta_1 = \beta_2 = \rho = 1$ and we set path loss index $\alpha = 2$.

We consider two randomly generated networks consisting of 10 and 50 nodes deployed over an 1×1 square. In all cases, the targeted accuracy for approximation algorithm is 0.95 of optimal, i.e., $\varepsilon = 0.05$.

The network setting (location, data rate, and initial energy for each node) for the 10-node network is given in Table III. By applying Algorithm 1, we find that fictitious cost point with cost vector $(1.05, 1.28, 1.05, 1.22, 1.16, 1.05, 1.05, 1.05, 1.41,$

TABLE IV

EACH NODE'S CARTESIAN COORDINATES, DATA GENERATION RATE AND INITIAL ENERGY FOR A 50-NODE NETWORK.

(x_i, y_i)	r_i	e_i	(x_i, y_i)	r_i	e_i
$(0.13, 0.15)$	0.1	260	$(0.08, 0.69)$	0.4	130
$(0.19, 0.06)$	1.0	390	$(0.59, 0.16)$	1.0	290
$(0.10, 0.60)$	0.9	240	$(0.89, 0.78)$	1.0	300
$(0.19, 0.10)$	0.2	60	$(0.89, 0.54)$	0.3	140
$(0.00, 0.02)$	0.9	270	$(0.21, 0.74)$	0.1	80
$(0.42, 0.24)$	0.9	400	$(0.43, 0.24)$	0.2	360
$(0.19, 0.79)$	0.5	380	$(0.11, 0.64)$	1.0	380
$(0.75, 0.23)$	0.3	440	$(0.33, 0.54)$	1.0	310
$(0.47, 0.82)$	0.4	60	$(0.69, 0.33)$	0.3	190
$(0.53, 0.04)$	0.1	500	$(0.22, 0.84)$	0.8	150
$(0.87, 0.27)$	0.4	290	$(0.47, 0.00)$	0.9	290
$(0.67, 0.02)$	0.5	480	$(0.94, 0.24)$	0.3	150
$(0.48, 0.20)$	0.8	420	$(0.32, 0.26)$	0.6	460
$(0.68, 0.80)$	0.4	90	$(0.37, 0.96)$	0.3	60
$(0.52, 0.35)$	0.3	130	$(0.81, 0.86)$	0.7	240
$(0.97, 0.30)$	0.2	400	$(0.37, 0.50)$	0.8	270
$(0.23, 0.42)$	0.3	150	$(0.65, 0.48)$	0.9	220
$(0.56, 0.40)$	0.5	220	$(0.91, 0.23)$	0.4	470
$(0.17, 0.00)$	0.5	460	$(0.44, 0.47)$	0.7	150
$(0.41, 0.10)$	0.1	360	$(0.81, 0.98)$	0.8	110
$(0.03, 0.47)$	0.8	150	$(0.30, 0.39)$	0.1	310
$(0.19, 0.17)$	0.3	170	$(0.78, 0.58)$	0.6	110
$(0.56, 0.16)$	0.1	290	$(0.73, 0.52)$	0.7	290
$(1.00, 0.88)$	0.7	430	$(0.91, 0.29)$	0.9	340
$(0.86, 0.14)$	0.6	130	$(0.96, 0.98)$	0.2	190

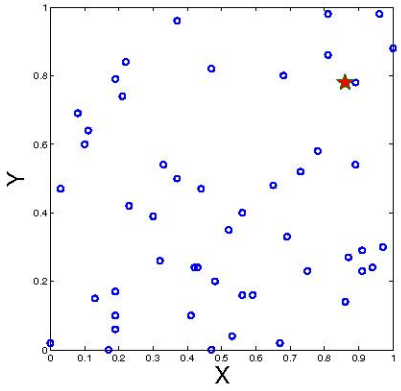

Fig. 5. A 50-node network.

1.05) having the maximum network lifetime $T^* = 357.49$, which is at least 95% of the optimal. By placing the base station at a point in the corresponding subarea, e.g., at point $(0.59, 0.31)$, the network lifetime is $T_\varepsilon = 359.17 > T^*$. Thus, the network lifetime is also at least 95% of the optimal. The flow routing solution is shown in Fig. 4, where a circle represents a sensor node and a star represents the location of the base station $(0.59, 0.31)$.

The network setting (location, data rate, and initial energy for each node) for the 50-node network is given in Table IV. By applying Algorithm 1, we obtain a $(1 - \varepsilon)$ optimal solution with $T_\varepsilon = 135.17$ when the base station is placed at $(0.51, 0.68)$ in Fig. 5.

A. Complexity Comparison

We now compare the complexity of our algorithm (Section III-D) with the approximation algorithm proposed in [7]. Similarly, the complexity of the approximation algorithm in [7] can also be measured by the number of LPs that need to be solved, which is

$$\left\lfloor \frac{4}{\varepsilon} \right\rfloor \left\lceil \frac{\alpha \ln 2}{\ln(1+\varepsilon/8)} \right\rceil \left\lceil \frac{8\alpha\pi}{\varepsilon} \right\rceil \cdot \sum_{i \in \mathcal{N}} \left\lceil \frac{\ln\left(8N^3 \sum_{j \in \mathcal{N}} r_j/(\varepsilon r_i)\right)}{\ln(1+\varepsilon/8)} \right\rceil . \quad (16)$$

To have a sense of quantitative comparison of complexity between our algorithm and the one in [7], we use (15) and (16) on the 10 and 50-node network considered in this section. Corresponding to each network topology, we find that the complexity of the approximation algorithm in [7] is $3.7\text{x}10^7$ and $5.2\text{x}10^6$ times of the complexity of our proposed approximation algorithm.

V. RELATED WORK

Due to energy constraint, network lifetime for a wireless sensor network is limited. As a result, there is a flourish of research activities on how to prolong network lifetime. Many of these efforts (see, e.g., [2], [4], [10], [17]) studied lifetime problem under given network topology and without explicit consideration on the impact of node placement on network performance.

Among the body of research on node placement, researchers have studied sensor node placement [6], [14], [15], [18], relay node placement [9], [16], and base station placement [3], [7], [11]. The main focus of sensor node placement has been on coverage in order to have either better geographical coverage of the area or better connectivity in the network. Relay node placement deals with how to place special auxiliary nodes within a sensor network so that network performance (e.g., connectivity, lifetime) can be improved. Related work in relay node placement (e.g., [9], [16]) have been limited to heuristic algorithm instead of providing performance guarantee.

Related work on base station placement include [3], [7], [11]. In [3], Bogdanov et al. studied how to place base station so that the network flow is proportionally maximized subject to link capacity. The authors show that although it is possible to find optimal solutions for special network topology (e.g., grid), the base station placement problem for an arbitrary network is NP-complete. The authors also pointed out that an approximation algorithm with any guarantee was not known at the time of their paper and subsequently proposed two heuristic algorithms. In [11], Pan et al. studied base station placement problem to maximize network lifetime. The optimal location is only determined for the simple case where only single-hop routing is allowed. The more difficult problem involving multi-hop routing was not addressed.

In [7], Efrat et al. proposed the first $(1 - \varepsilon)$ optimal approximation algorithm to the base station placement problem. However, since they constructed a finite search space of physical points, the computational complexity of their

algorithm is higher than the one proposed in this paper for most cases.

VI. Conclusions

In this paper, we investigated the base station placement problem for a multi-hop sensor network. The main result is an approximation algorithm that can guarantee $(1-\varepsilon)$ optimal network lifetime performance for base station placement problem with any desired error bound $\varepsilon > 0$. The proposed $(1-\varepsilon)$ approximation algorithm was based on several novel techniques such as discretization of cost parameters (and distances), division of search space into finite number of subareas, and representation of subareas with fictitious points (with nice bounding properties on costs). The proposed approximation algorithm offers significant complexity reduction when compared to a state-of-the-art algorithm and represents the best known result to the described base station placement problem.

References

[1] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali, *Linear Programming and Network Flows*, second edition, John Wiley & Sons, 1990.

[2] M. Bhardwaj and A.P. Chandrakasae, "Bounding the lifetime of sensor networks via optimal role assignments," in *Proc. IEEE Infocom*, pp. 1587–1596, New York, NY, June 23-27, 2002.

[3] A. Bogdanov, E. Maneva, and S. Riesenfeld, "Power-aware base station positioning for sensor networks," in *Proc. IEEE Infocom*, pp. 575–585, Hong Kong, China, March 7–11, 2004.

[4] T.X. Brown, H.N. Gabow, and Q. Zhang, "Maximum flow-life curve for a wireless ad hoc network," in *Proc. ACM MobiHoc*, pp. 128–136, Long Beach, CA, Oct. 4–5, 2001.

[5] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, second edition, Chapter 4, Springer–Verlag, New York, NY, 1998.

[6] S.S. Dhillon and K. Chakrabarty, "Sensor plocement for effective coverage and strveillance in distributed sensor networks," in *Proc. IEEE WCNC*, pp. 1609–1614, New Orleans, LA, March 16–20, 2003.

[7] A. Efrat, S. Har-Peled, and J. Mitchell, "Approximation algorithms for location problems in sensor networks," in *Proc. IEEE Broadband Communications, Networks, and Systems*, pp. 767–776, Boston, MA, Oct. 3–7, 2005.

[8] W. Heinzelman, *Application-specific Protocol Architectures for Wireless Networks*, Ph.D. thesis, MIT, 2000.

[9] Y.T. Hou, Y. Shi, H.D. Sherali, and S.F. Midkiff, "Prolonging sensor network lifetime with energy provisioning and relay node placement," in *Proc. IEEE SECON*, pp. 295–304, Santa Clara, CA, Sep. 26–29, 2005.

[10] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Maximum lifetime data gathering and aggregation in wireless sensor setworks," in *Proc. IEEE International Conference on Networking*, pp. 685–696, Atlanta, GA, Aug. 26–29, 2002.

[11] J. Pan, Y.T. Hou, L. Cai, Y. Shi, and S.X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM Mobicom*, pp. 286–299, San Diego, CA, Sep. 14–19, 2003.

[12] Y. Shi and Y.T. Hou, "Approximation algorithm for base station placement in wireless sensor networks," Technical Report, the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, Nov. 2006. Available at http://www.ece.vt.edu/thou/Research.html.

[13] K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie, "Protocols for self-organizing of a wireless sensor network," *IEEE Personal Communications Magazine*, vol. 7, pp. 16–27, Oct. 2000.

[14] G. Wang, G. Cao, and T. La Porta, "Movement-assisted sensor deployment," in *Proc. IEEE Infocom*, pp. 2469–2479, Hong Kong, China, March 7–11, 2004.

[15] J. Wu and S. Yang, "SMART: A scan-based movement-assisted sensor deployment method in wireless sensor networks," in *Proc. IEEE Infocom*, pp. 2313–2324, Miami, FL, March 13–17, 2005.

[16] K. Xu, H. Hassanein, and G. Takahara, "Relay node deployment strategies in heterogeneous wireless sensor networks: multiple-hop communication case," in *Proc. IEEE SECON*, pp. 575-585, Santa Clara, CA, Sep. 26–29, 2005.

[17] H. Zhang and J. Hou, "On deriving the upper bound of α-lifetime for large sensor networks," in *Proc. ACM MobiHoc*, pp. 121–132, Tokyo, Japan, May 24–26, 2004.

[18] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization based on virtual forces," in *Proc. IEEE Infocom*, pp. 1293–1303, San Francisco, CA, April 1–3, 2003.