On Base Station Selection for Anycast Flow Routing in Energy-Constrained Wireless Sensor Networks

Y. Thomas Hou Yi Shi
The Bradley Dept. of Electrical and Computer Engineering
Virginia Tech
Blacksburg, VA, USA
{thou,yshi}@vt.edu

Hanif D. Sherali
The Grado Department of Industrial and Systems Engineering,
Virginia Tech
Blacksburg, VA, USA
hanifs@vt.edu

Abstract

Energy constraints have had a significant impact on the design and operation of wireless sensor networks. In this paper, we investigate base station selection (or anycast) problem in wireless sensor networks. We consider a wireless sensor network having multiple base stations (data sink nodes), where each source node must send all *its locally generated data to only one base station. To maximize the network lifetime, it is essential to optimally match each source node to a particular base station in addition to finding an optimal routing solution. We propose a polynomial time heuristic for optimal base station selection for anycast via a sequential fixing procedure, under the assumption that the bit rate from each source node is constant. Through extensive simulation results, we show that this heuristic has excellent performance behavior and is a tight low bound that is very close to optimal solution for the original optimization problem.*

1 Introduction

Wireless sensor networks consist of battery-powered nodes that are endowed with a multitude of sensing modalities including multi-media (*e.g.,* video, audio) and scalar data (*e.g.,* temperature, pressure, light, magnetometer, infrared). The demand for these networks is spurred by numerous applications that require in-situ, unattended, high-precision, and real-time observations over a vast area. Although there have been significant improvements in processor design and computing, advances in battery technology still lag behind, making energy resource the fundamental constraint in wireless sensor networks.

As a result, there has been active research on exploring optimal flow routing strategies to maximize the lifetime of the network (see, e.g., [4, 6]). Network lifetime refers to the maximum time that nodes in the network remain alive until one or more nodes drain up their energy. Most prior efforts assume that the mapping between a sensor node and its sink node (base station) is given *a priori.* For example, for a sensor network having only a single sink node (e.g., a base station) [3, 11, 14], all the data traffic generated by the sensor nodes will be delivered to this sink node. For a sensor network having multiple sink nodes, data traffic generated by any sensor node may be delivered to multiple base stations simultaneously [4, 6].

However, for the case when multiple base stations are present, there has been little research to-date on addressing optimal base station selection for anycast routing where anycast is defined in the sense that each source node must send *all* its locally generated data to only one base station. There are two aspects about this problem that are interesting. From an application requirement perspective, although many sensing applications allow collected data to be sent to multiple base stations, this is not always the case for certain application. For example, for some real-time multimedia sensing applications (e.g., surveillance video), it is necessary to have all the traffic generated from a source node be routed to the *same* base station (albeit that they may be split into sub-flows traversing different paths) so that decoding and processing can be properly completed. This is because for multimedia traffic such as video, the information contained in different packets from the same source node are highly correlated and dependent (due to compression). A particular base station may not be able to decode the video packets properly if packets generated by a source node are sent to different base stations. From a communication power consumption perspective, how a particular base station (as a destination sink node) is chosen could have

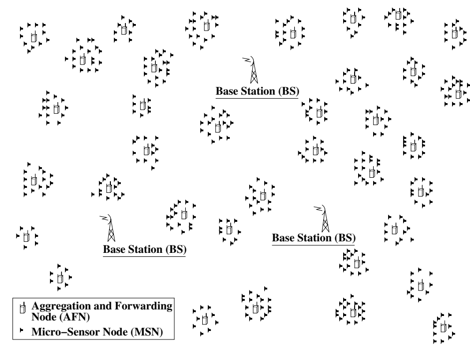
Proceedings of the 2nd Int'l Conf. on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'05)

significant impact on the overall network lifetime performance. This is because communication power consumption is topology dependent; the optimal flow routing strategy (to maximize network lifetime) depends on the particular mapping between a sensor node (source) and a base station (destination). As a result, there appears to be a compelling need to understand how to perform *anycast* in energy-constrained sensor networks.

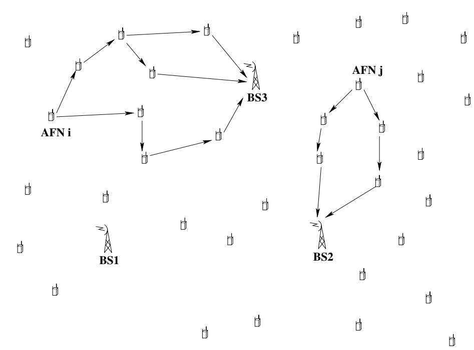
In this paper, we investigate the optimal base station selection problem for anycast with the aim of maximizing network lifetime. We show that the joint base station selection and anycast flow routing problem can be formulated as a *mixed integer nonlinear programming* (MINLP) optimization problem. Since MINLP is NP-hard in general [9] and our base station selection problem is likely to be NP-hard as well, we develop a heuristic algorithm in the hope of providing good solutions.

To provide a measure for the quality of our proposed heuristic, we first explore computing a tight upper bound on the maximization problem by applying a suitable *relaxation* technique. With this upper bound as a performance measure, we move on to develop a heuristic algorithm. Our heuristic, called "ABS", for Anycast Base station Selection, is based on the conjecture that the optimal base station for a node should be closely related to the base station that receives the largest amount of traffic volume *when there is no constraint on the number of destination base stations.* We employ a *sequential fixing procedure* to find the optimal base station for each node. Numerical results show that the ABS algorithm yields a solution that has an objective value very close to the upper bound produced by our relaxation procedure, hence suggesting that the solution offered by our heuristic algorithm must be even closer to the optimal solution.

The remainder of this paper is organized as follows. In Section 2, we first describe the reference network model, which is based on a two-tier architecture. Our focus is to study optimal base station selection and anycast flow routing for the upper-tier aggregation and forwarding nodes (AFNs). We also describe the power consumption behavior for AFNs and introduce the anycast optimization problem that we plan to investigate. In Section 3, we formulate the anycast problem as an MINLP problem, which is NP-hard in general. Then we develop an upper bound for this MINLP problem as a performance measure for any heuristic algorithm. Since the base station selection problem is likely to be NP-hard as well, in Section 4 we develop a heuristic algorithm (ABS) to the anycast problem. In Section 5, we offer extensive simulation results and show that the ABS algorithm is able to offer near-optimal solution. Section 6 reviews related work and Section 7 concludes this paper.



(a) Physical topology.



(b) Two examples illustrating anycast between an AFN and a BS.

Figure 1. Reference network model. Physical network consisting of BS's, AFNs, and MSNs is shown in (a). In (b), we illustrate with two examples for anycast between an AFN and a BS.

2 Reference Network Model and the Anycast Problem

2.1 Reference Network Model

We consider a two-tiered architecture for wireless sensor networks [7, 18]. Figure 1(a) shows the *physical* network topology for such a network. There are three types of nodes in the network, namely, *micro-sensor nodes* (MSNs), *aggregation and forwarding nodes* (AFNs), and *base stations* (BSs). The MSNs can be application-specific sensor nodes (*e.g.,* temperature sensor nodes (TSNs), pressure sensor nodes (PSNs), and video sensor nodes (VSNs)) and they constitute the lower tier of the network. They are small and low-cost, and are deployed in groups (or clusters) at strategic locations for sensing applications. The objective of an MSN is to collect data and to send it directly to the local AFN.[1]

[1] Due to the small distance between an MSN and its local AFN, multi-hop routing among the MSNs may not be necessary.

Proceedings of the 2nd Int'l Conf. on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'05)
0-7695-2423-0/05 $20.00 © 2005 IEEE

For each cluster of MSNs, there is one AFN, which is different from an MSN in terms of physical properties and functions. The primary functions of an AFN are: (1) *data aggregation* (or "fusion") for information flows coming from the local cluster of MSNs, and (2) *forwarding* (or relaying) the aggregated information to the next hop AFN (toward a base station). For data fusion, an AFN analyzes the content of each data stream (*e.g.,* video) it receives, from which it composes a complete scene by exploiting the correlation among each individual data stream from the MSNs [7]. After data fusion, the aggregated bit rate from an AFN i (denoted as g_i) will be forwarded to toward a base station in either single or multiple hops. Although an AFN is expected to be provisioned with much more energy than an MSN, it also consumes energy at a substantially higher rate (due to wireless communication over large distances). Consequently, an AFN has a limited lifetime. Upon depletion of energy at an AFN, we expect that the sensing *coverage* for the particular area is lost, despite the fact that some of the MSNs within the cluster may still have remaining energy.

The third component in the two-tiered architecture are base stations. Essentially, base stations are the *sink* nodes for all the data collected in the network. In this investigation, we assume that there is sufficient energy resource available at a base station and thus, there is no energy constraint for base stations.

In summary, the main functions of the lower tier MSNs are data acquisition and compression while the upper-tier AFNs are used for data fusion and wireless networking for relaying sensing information to the base stations. Our focus in this paper is on the upper tier wireless multi-hop communications among the AFNs toward a BS via anycast. Table 1 lists the notation used in this paper.

Table 1. Notation.

Symbols	Definitions
N and M	The number of AFNs and base stations
(x_i, y_i)	The location of AFN i
e_i	The initial energy at AFN i
g_i	The locally generated data rate at AFN i
ρ	Power consumption coefficient for receiving data
c_{ik}	Power consumption coefficient for transmitting data from AFN i to node k
α	Distance independent term in power consumption for transmitting data
β	Distance dependent term in power consumption for transmitting data
d_{ik}	Physical distance between AFN i and node k
$f_{A_iA_j}^{A_kB_l}$ (or $f_{A_iB_l}^{A_kB_l}$)	The flow rate from AFN i to AFN j (or base station l) with the source and destination being AFN k and base station l
$\mathcal{F}_{\mathcal{A}A}$ (or $\mathcal{F}_{\mathcal{A}B}$)	The set of flows from an AFN to another AFN (or a base station)
$\mathcal{F}_{\mathcal{A}A_i}$	The set of in-coming flows to AFN i
$\mathcal{F}_{A_i\mathcal{A}}$ (or $\mathcal{F}_{A_i\mathcal{B}}$)	The set of out-going flows from AFN i to another AFN (or a base station)
$\lambda^{A_iB_l}$	If the data generated by AFN i will be transmitted to base station l, then $\lambda^{A_iB_l} = 1$; otherwise $\lambda^{A_iB_l} = 0$
$V_{A_iA_j}^{A_kB_l}$ (or $V_{A_iB_l}^{A_kB_l}$)	The data volume (in bits) transported from AFN i to AFN j (or base station l) with the source and destination being AFN k and base station l
$\mathcal{V}_{\mathcal{A}A}$ (or $\mathcal{V}_{\mathcal{A}B}$)	The set of volumes from an AFN to another AFN (or a base station)
$\mathcal{V}_{\mathcal{A}A_i}$	The set of in-coming volumes to AFN i
$\mathcal{V}_{A_i\mathcal{A}}$ (or $\mathcal{V}_{A_i\mathcal{B}}$)	The set of out-going volumes from AFN i to another AFN (or a base station)
$\mu^{A_iB_l}$	$= \lambda^{A_iB_l}T$ in LP-Relax
T_{UB}	An upper bound for the anycast network lifetime
$d(k)$	The destination for AFN k under each heuristic
T_{ABS}	Network lifetime under the ABS algorithm
T_{nearest}	Network lifetime under the nearest base station selection approach
T_{random}	Network lifetime under the random base station selection approach
L_{ABS}	$= \frac{T_{\text{ABS}}}{T_{\text{UB}}}$, normalized network lifetime under the ABS algorithm
L_{nearest}	$= \frac{T_{\text{nearest}}}{T_{\text{UB}}}$, normalized network lifetime under the nearest base station selection approach
L_{random}	$= \frac{T_{\text{random}}}{T_{\text{UB}}}$, normalized network lifetime under the random base station selection approach

2.2 Power Consumption Model

As described, for AFN i, the aggregate bit rate generated locally is $g_i, i = 1, 2, \cdots, N$, which must be routed toward a base station. For an AFN, the energy consumption due to wireless communication (*i.e.,* receiving and transmitting) is considered the dominant source in power consumption [1]. The power dissipation at a radio transmitter can be modeled as:

$$p_t(i,\ k) = c_{ik} \cdot f_{ik} \ , \tag{1}$$

where $p_t(i,\ k)$ is the power dissipated at AFN i when it is transmitting to node k, f_{ik} is the bit rate transmitted from AFN i to node k, and c_{ik} is the power consumption cost of radio link (i, k) and can be modeled as

$$c_{ik} = \alpha + \beta \cdot d_{ik}^m \ , \tag{2}$$

where α is a *distance-independent* constant term, β is a coefficient term associated with the *distance-dependent* term, d_{ik} is the physical distance between node i and node k, and m is the path loss index, with $2 \leq m \leq 4$ [20]. Example values for these parameters are $\alpha = 50$ nJ/b and $\beta = 0.0013$ pJ/b/m^4 (for $m = 4$) [10].[2] Since the power level of an AFN's transmitter can be used to control the distance coverage of an AFN (see, e.g., [19, 21]), different network flow routing topologies can be formed by adjusting the power level of each AFN's transmitter. Therefore, throughout this paper, whenever we have a flow routing topology, we assume that the power level at the underlying physical node is also adjusted accordingly to achieve such inter-nodal communications.

The power dissipation at a receiver can be modeled as [20]:

$$p_r(i) = \rho \cdot \sum_{k \neq i} f_{ki} \ , \tag{3}$$

[2]In this paper, we use $m = 4$ in all of our numerical results.

where $\sum_{k \neq i} f_{ki}$ (in b/s) is the aggregated rate of the received data stream by AFN i. A typical value for the parameter ρ is 50 nJ/b [10].

2.3 Optimal Base Station Selection for Anycast Routing

The anycast problem we investigate in this paper involves an optimal mapping between an AFN and a base station such that the network lifetime can be maximized. There are two components that are deeply coupled in this research. The first component involves the mapping between each AFN and a particular base station. The second component deals with how to perform flow routing for a given mapping such that the network lifetime can be maximized. Many existing papers on optimal flow routing (e.g., [3, 6] only addresses the second component of this problem, i.e., assuming that the mapping between an AFN and one (or more) base station is known *a priori*. However, when the mapping is not given, the joint problem of base selection selection and flow routing (so that the network lifetime can be maximized) becomes more interesting. In addition to its intellectual interest, there are also important application scenarios that motivate us to investigate this problem. In particular, for certain applications (e.g., surveillance video), it is necessary to forward all bit streams generated by an AFN to the same base station (instead of to different base stations). This is because partial data streams from a video source may not be properly decoded and processed at a base station.

It is worth noting that anycast routing is different from single path routing. That is, although we mandate that all bit streams generated by an AFN must be relayed to the same base station, the bit stream can be split into sub-flows and sent to the same base station through different paths (see Fig. 1(b)). Although doing so will result in delay jitter and thus require play-out buffer at the base station, this approach will be much more flexible and energy "wise" than mandating to send the flow along a single path.

3 Problem Formulation and An Upper Bound for Optimal Solution

3.1 Problem Formulation

For the base station selection/anycast routing problem, denote $f_{A_iA_j}^{A_kB_l} \in \mathcal{F}_{\mathcal{A}\mathcal{A}}$ as the flow (in b/s) from AFN i to relay node AFN j with the source and destination of the flow being AFN k and base station l, where $\mathcal{F}_{\mathcal{A}\mathcal{A}} = \{f_{A_iA_j}^{A_kB_l} : 1 \leq i,j,k \leq N, i \neq j, k \neq j, 1 \leq l \leq M\}$. Similarly, denote $f_{A_iB_l}^{A_kB_l} \in \mathcal{F}_{\mathcal{A}\mathcal{B}}$ as the flow from AFN i to base station l with source and destination of the flow being AFN k and base station l, where $\mathcal{F}_{\mathcal{A}\mathcal{B}} = \{f_{A_iB_l}^{A_kB_l} : 1 \leq i,k \leq N, 1 \leq l \leq M\}$.

To formulate the optimization problem for the joint base station selection and anycast flow routing problem, we need to keep track of the incoming and outgoing flows at each AFN. Denote the set of incoming flows to AFN i as $\mathcal{F}_{\mathcal{A}A_i}$, the set of outgoing flows from AFN i to other AFNs as $\mathcal{F}_{A_i\mathcal{A}}$, and the set of outgoing flows from AFN i to base stations as $\mathcal{F}_{A_i\mathcal{B}}$. Then we have $\mathcal{F}_{\mathcal{A}A_i} = \{f_{A_mA_i}^{A_kB_l} : 1 \leq m,k \leq N, m \neq i, k \neq i, 1 \leq l \leq M\}$, $\mathcal{F}_{A_i\mathcal{A}} = \{f_{A_iA_r}^{A_kB_l} : 1 \leq r,k \leq N, r \neq k, r \neq i, 1 \leq l \leq M\}$, and $\mathcal{F}_{A_i\mathcal{B}} = \{f_{A_iB_l}^{A_kB_l} : 1 \leq k \leq N, 1 \leq l \leq M\}$. Denote T as the network lifetime, which is defined as the time until the first AFN drains out of energy. Then the optimization problem for the base station selection (BS) and anycast routing (AR) can be formulated as follows.

Problem BS-AR: Max T

$$\text{s.t.} \sum_{r \neq i} f_{A_iA_r}^{A_iB_l} + f_{A_iB_l}^{A_iB_l} - g_i \lambda^{A_iB_l} = 0 \quad (1 \leq i \leq N, 1 \leq l \leq M) \quad (4)$$

$$\sum_{r \neq i,k} f_{A_iA_r}^{A_kB_l} + f_{A_iB_l}^{A_kB_l} - \sum_{m \neq i} f_{A_mA_i}^{A_kB_l} = 0 \quad (1 \leq i \leq N, 1 \leq l \leq M, 1 \leq k \leq N, k \neq i) \quad (5)$$

$$\left[\sum_{f_{A_iA_r}^{A_kB_l} \in \mathcal{F}_{A_i\mathcal{A}}} c_{A_iA_r} f_{A_iA_r}^{A_kB_l} + \sum_{f_{A_iB_l}^{A_kB_l} \in \mathcal{F}_{A_i\mathcal{B}}} c_{A_iB_l} f_{A_iB_l}^{A_kB_l} + \sum_{f_{A_mA_i}^{A_kB_l} \in \mathcal{F}_{\mathcal{A}A_i}} \rho f_{A_mA_i}^{A_kB_l} \right] T \leq e_i \quad (1 \leq i \leq N) \quad (6)$$

$$\sum_{1 \leq l \leq M} \lambda^{A_iB_l} = 1 \quad (1 \leq i \leq N) \quad (7)$$

$$T, f_{A_iA_j}^{A_kB_l}, f_{A_iB_l}^{A_kB_l} \geq 0, \lambda^{A_iB_l} = 0 \text{ or } 1 \quad (f_{A_iA_j}^{A_kB_l} \in \mathcal{F}_{\mathcal{A}\mathcal{A}}, f_{A_iB_l}^{A_kB_l} \in \mathcal{F}_{\mathcal{A}\mathcal{B}}, 1 \leq i,j,k \leq N, i \neq j, k \neq j, 1 \leq l \leq M).$$

Note that $\lambda^{A_iB_l}$ is a binary variable used for base station selection: if the data stream generated by AFN i will be transmitted to base station l, then $\lambda^{A_iB_l} = 1$; otherwise $\lambda^{A_iB_l} = 0$. The set of constraints in (4) to (7) can be interpreted as follows. The first set of constraints in (4) focuses on traffic flow generated *locally* at each AFN i. They state that, for each AFN i, if base station l is the destination, then the locally generated bit rate (i.e., g_i) will be equal to the outgoing data flows from AFN i toward base station l via a single hop (i.e., $f_{A_iB_l}^{A_iB_l}$) or multi-hop (i.e., $f_{A_iA_r}^{A_iB_l}$); otherwise, all flows corresponding to the source-destination pair (A_iB_l) must be zero. The second set of constraints in (5) focus on the traffic that use AFN i as a relay node. They state that at each relay node i, the total amount of incoming traffic (i.e., $\sum_{m \neq i} f_{A_mA_i}^{A_kB_l}$) should be the same as the total amount of outgoing traffic (i.e., $\sum_{r \neq i,k} f_{A_iA_r}^{A_kB_l} + f_{A_iB_l}^{A_kB_l}$), for each source-destination pair (A_iB_l). The third set of constraints in (6) concerns energy consumption at AFN i. They state that, for each AFN i, the energy consumption due to transmission and receiving (see Eqs. (1) and (3)) over the course of the network lifetime should not exceed the initial energy provision e_i. Note that in (6), both flows generated locally at AFN i and those flows that use AFN i as relay node are included. Finally, the remaining two sets of constraints enforce that AFN i can only transmit all of its data to one base station under our anycast requirement, along with the logical restrictions on the optimization variables $\lambda^{A_iB_l}$, $f^{A_kB_l}_{A_iA_j}$, and $f^{A_kB_l}_{A_iB_l}$. Note that ρ, g_i, e_i, $c_{A_iA_r}$, and $c_{A_iB_l}$ are all constants in this optimization problem.

Problem BS-AR is a *mixed-integer non-linear programming* (MINLP) problem, which is, unfortunately, NP-hard in general [9]. Although there exist software (e.g. BARON [2]) to solve such problems, the solutions are obtainable only for small networks. Although we do not have a formal proof in this paper, we conjecture that our BS-AR problem is also NP-hard. As a result, we pursue a heuristic algorithm to address this problem.

In addition to designing a heuristic that offers a lower bounding solution, we also develop an upper bound to this problem, which can be used as a measure for the quality of the heuristic solution obtained. In particular, if our heuristic produces a solution close to this upper bound, then the solution offered by the heuristic must be *even closer* to the actual optimal solution, hence demonstrating its performance.

3.2 An Upper Bound for Optimal Solution

In this section, we develop an upper bound for the BS-AR problem (see Section 3.1) by studying a closely related problem that can be formulated and solved via linear programming (LP). This process involves two steps. As the first step, we relax the binary requirement on $\lambda^{A_iB_l}$ by letting $\lambda^{A_iB_l}$ be a *real* number with $\lambda^{A_iB_l} \in [0,\ 1]$. Consequently, the integer component in the MINLP problem disappears and we now have a *non-linear programming* (NLP) formulation. Apparently, the solution to this NLP formulation gives an upper bound to the BS-AR problem since the continuous relaxation of $\lambda^{A_iB_l}$ only increases the solution space to the original BS-AR problem. There is also an intuitive physical interpretation to this NLP problem. That is, under this NLP problem, we allow the data from AFN i to be sent to multiple base stations instead of to just one base station. The fraction is determined by $\lambda^{A_iB_l}$, i.e., AFN i sends $\lambda^{A_iB_l}$ of its data to base station l.

Although the resulting bilinear problem is still NP-hard in general [9], the particular structure of problem BS-AR permits it to be equivalently transformed into a linear programming. To see this, let us multiply (4), (5), and (7) by T and then use the linearizing substitutes $V^{A_kB_l}_{A_iA_j} = T \cdot f^{A_kB_l}_{A_iA_j}$, $V^{A_kB_l}_{A_iB_l} = T \cdot f^{A_kB_l}_{A_iB_l}$, and $\mu^{A_iB_l} = T \cdot \lambda^{A_iB_l}$. Also denote $\mathcal{V}_{\mathcal{AA}}$ as the set of traffic volumes being transported among the AFNs (i.e., the $V^{A_kB_l}_{A_iA_j}$ variables) and $\mathcal{V}_{\mathcal{AB}}$ as the set of traffic volumes being transported between AFNs and base stations (i.e., the $V^{A_kB_l}_{A_iB_l}$ variables). Furthermore, for each AFN i, denote $\mathcal{V}_{\mathcal{A}A_i}$ as the set of in-coming traffic volumes (i.e., the $V^{A_kB_l}_{A_mA_i}$ variables), $\mathcal{V}_{A_i\mathcal{A}}$ as the set of out-going traffic volumes to other AFNs (i.e., the $V^{A_kB_l}_{A_iA_r}$ variables), and $\mathcal{V}_{A_i\mathcal{B}}$ as the set of out-going volumes to base stations (i.e., the $V^{A_kB_l}_{A_iB_l}$ variables). Then, the NLP problem can be reformulated into the following equivalent LP problem.

<u>LP-Relax:</u> Max T

$$\text{s.t.} \sum_{r\neq i} V^{A_iB_l}_{A_iA_r} + V^{A_iB_l}_{A_iB_l} - g_i\mu^{A_iB_l} = 0 \quad (1 \le i \le N, 1 \le l \le M) \quad (8)$$

$$\sum_{r\neq i,k} V^{A_kB_l}_{A_iA_r} + V^{A_kB_l}_{A_iB_l} - \sum_{m\neq i} V^{A_kB_l}_{A_mA_i} = 0 \quad (1 \le i \le N, 1 \le l \le M, 1 \le k \le N, k \neq i) \quad (9)$$

$$\sum_{V^{A_kB_l}_{A_mA_i} \in \mathcal{V}_{\mathcal{A}A_i}} \rho V^{A_kB_l}_{A_mA_i} + \sum_{V^{A_kB_l}_{A_iA_r} \in \mathcal{V}_{A_i\mathcal{S}}} c_{A_iA_r} V^{A_kB_l}_{A_iA_r} + \sum_{V^{A_kB_l}_{A_iB_l} \in \mathcal{V}_{A_i\mathcal{B}}} c_{A_iB_l} V^{A_kB_l}_{A_iB_l} \le e_i \quad (1 \le i \le N) \quad (10)$$

$$\sum_{1 \le l \le M} \mu^{A_iB_l} - T = 0 \quad (1 \le i \le N) \quad (11)$$

$$T, V^{A_kB_l}_{A_iA_j}, V^{A_kB_l}_{A_iB_l}, \mu^{A_iB_l} \ge 0 \quad (V^{A_kB_l}_{A_iA_j} \in \mathcal{V}_{\mathcal{AA}}, V^{A_kB_l}_{A_iB_l} \in \mathcal{V}_{\mathcal{AB}}, 1 \le i,j,k \le N, i \neq j, k \neq j, 1 \le l \le M) .$$

where Eqs. (8) and (9) follow from the flow balance equations (4) and (5), Eqs. (10) follow from the energy constraints in (6), and Eqs. (11) follow from the energy constraints in (7). Note that T, $V^{A_kB_l}_{A_iA_j}$, $V^{A_kB_l}_{A_iB_l}$, and $\mu^{A_iB_l}$ are variables, and ρ, g_i, e_i, $c_{A_iA_r}$, and $c_{A_iB_l}$ are all constants.

We now have a standard LP formulation, which was transformed directly from the NLP problem. By their equivalence, the solution to this LP problem yields an upper bound to problem BS-AR. We will use this solution as a performance measure for heuristics. Our numerical results show that this upper bound is extremely tight to the optimal solution to the MINLP problem, as our would expect from the convex hull results presented by Sherali *et al.* in [22].

4 ABS: A Heuristic Algorithm

Suppose that we know an optimal mapping between each AFN i and a base station. Then we can find an optimal flow routing using an LP formulation similar to that in [6]. Since such an optimal mapping is not available, we develop our heuristic solution in two steps:[3] 1) find a good mapping

[3] Clearly, such a problem decomposition will yield a sub-optimal solution to the BS-AR problem. That is, the solution produced by this heuristic algorithm is a lower bound for the BS-AR problem.

between each AFN and a base station; 2) find an optimal flow routing for this mapping.

Our heuristic algorithm in the first step is called ABS (abbreviation for Anycast Base station Selection) and is motivated by the solution to the LP-Relax problem discussed in the Section 3.2. Under LP-Relax, each AFN is allowed to send its traffic to multiple base stations. The basic idea is to assign a source AFN, say i, to the base station that receives the *largest* amount of traffic volume (in bits) from AFN i among all the base stations in the solutions to LP-Relax. In particular, we use the *sequential fixing* procedure with a tunable parameter θ to find the destinations for all AFNs.[4]

Algorithm 1 (ABS)

1. *Solve the LP-Relax problem.*
2. *Now, fix some AFNs' base stations via the solution to LP-Relax problem. There are two cases.*
   (a) *If there are AFNs that send at least θ of its data to one base station. For such an AFN i, i.e., $\lambda^{A_iB_l}(=\frac{\mu^{A_iB_l}}{T}) \geq \theta$,[5] fix this base station as its destination.*
   (b) *If there is no AFN sends at least θ of its data to one base station. Assume that $\mu^{A_iB_l}$ is the largest among all μ values, fix B_l as AFN i's destination.*
3. *If all AFNs' destinations are fixed, stop; otherwise, reformulate the LP-Relax problem. In this LP-Relax, if AFN i's destination is fixed as B_l, then $\mu^{A_iB_l} = T$ (i.e., $\lambda^{A_iB_l} = 1$) and all other μ (or λ) variables for AFN i are zero.*[6]
4. *Go to Step 1.*

There is one small subtle detail that deserves further consideration under this approach. Suppose that in Step 2(b), the largest traffic volume sent by AFN i to a base station is comparable to the second largest traffic volume sent by AFN i to a different base station. Which base station should we then choose as the optimal base station for anycast? Clearly, the distance factor should be taken into consideration since, doing so, would help reduce energy consumption and help increase the network lifetime. Under ABS, we choose the base station that is closer to source AFN i whenever the difference between the largest and second largest traffic volumes (generated by AFN i) destined to two difference base stations is within a certain range. More formally, we introduce a threshold parameter (ϵ) to quantify the gap between the largest traffic volume and other traffic volumes destined to different base stations. For source AFN i, if the largest traffic volume to a base station l is comparable to the second largest traffic volume to base station m under LP-Relax, i.e., $\lambda^{A_iB_l} - \lambda^{A_iB_m} < \epsilon$, where $\lambda^{A_iB_l} = \frac{\mu^{A_iB_l}}{T}$ and $\lambda^{A_iB_m} = \frac{\mu^{A_iB_m}}{T}$, respectively, in the solution to LP-Relax, *and* the base station m is closer to AFN i than base station l, then we choose base station m as AFN i's anycast destination. The parameter ϵ is an engineering parameter that is tunable.[7]

We emphasize that the amount of traffic volumes to different base stations under LP-Relax is the dominant reason to map the AFN to a base station in our ABS algorithm. The proximity of a base station to the AFN is considered only if the largest and second largest traffic volumes to two different base stations are comparable. In Section 5, we will show that choosing the nearest base station to AFN i is not a good approach.

Denote $d(i)$ as the resulting destination for AFN i using the above mapping. Then we have $\mu^{A_id(i)} = T$ and $\mu^{A_iB_l} = 0$ for $B_l \neq d(i)$. In the second step, we can find the routing solution via an LP formulation as follows.

LP-Routing: Max T

$$\text{s.t.} \quad \sum_{r\neq i} V_{A_iA_r}^{A_id(i)} + V_{A_id(i)}^{A_id(i)} = g_iT \qquad (1 \leq i \leq N)$$

$$\sum_{r\neq i,k} V_{A_iA_r}^{A_kd(k)} + V_{A_id(k)}^{A_kd(k)} - \sum_{m\neq i} V_{A_mA_i}^{A_kd(k)} = 0 \qquad (1 \leq i \leq N,\ 1 \leq k \leq N, k \neq i)$$

$$\sum_{V_{A_mA_i}^{A_kd(k)} \in \mathcal{V}_{\mathcal{A}A_i}} \rho V_{A_mA_i}^{A_kd(k)} + \sum_{V_{A_iA_r}^{A_kd(k)} \in \mathcal{V}_{A_i\mathcal{A}}} c_{A_iA_r} V_{A_iA_r}^{A_kd(k)}$$

$$+ \sum_{V_{A_id(k)}^{A_kd(k)} \in \mathcal{V}_{A_i\mathcal{B}}} c_{A_id(k)} V_{A_id(k)}^{A_kd(k)} \leq e_i \qquad (1 \leq i \leq N)$$

$$T, V_{A_iA_j}^{A_kd(k)}, V_{A_id(k)}^{A_kd(k)} \geq 0 \qquad (V_{A_iA_j}^{A_kd(k)} \in \mathcal{V}_{\mathcal{A}\mathcal{A}},\ V_{A_id(k)}^{A_kd(k)} \in \mathcal{V}_{\mathcal{A}\mathcal{B}}, 1 \leq i,j,k \leq N, i \neq j, k \neq j)\ .$$

5 Performance Evaluation

In this section, we present numerical results demonstrating the performance of the ABS algorithm. In our experiments, we consider different network sizes and numbers of base stations under various topologies. In particular, we consider networks consisting of $N = 10$, 20, and 30 AFNs along with $M = 4$, 5, and 6 base stations. That is, we have a total of 9 possible N and M combinations. For each combination, we run 10 experiments (each under a randomly generated network topology for the AFNs), thus obtaining 90 sets of data.

[4] We use $\theta = 0.85$ in our numerical results.

[5] Recall that AFN i sends a fraction of $\lambda^{A_iB_l}$ of its data to base station B_l.

[6] Corresponding variables and constraints can be removed.

[7] We use 0.1 in our numerical results.

For each topology, an AFN i is placed randomly with uniform distribution along both x and y dimensions within the following range: $x_i, y_i \in [0,\ 1000]$ (m). The base stations, B_1, B_2, B_3, and B_4 are located at $(0,0)$, $(0,1000)$, $(1000,0)$, and $(1000,1000)$ (all in meters), respectively. When there are five base stations present, B_5 is located at $(500,500)$; when there are six base stations present, B_5 and B_6 are located at $(0,500)$ and $(1000,500)$, respectively. The initial energy at AFN i is also randomly generated following a uniform distribution with $e_i \in [250,\ 500]$ (kJ). The data rate generated by AFN i, g_i, is also uniformly distributed within $[2,\ 10]$ (kb/s).

For each run (90 total), we can obtain the upper bound for the network lifetime (denoted as T_{UB}) through LP-Relax as discussed in Section 3.2. Denote T_{ABS} as the network lifetime obtained via our ABS algorithm. For comparison against the performance of ABS, we also consider the network lifetime obtained under two other approaches. One approach is that each AFN i simply chooses the nearest base station as its anycast base station. We denote the network lifetime performance under this approach as T_{nearest}. The other approach is that each AFN i chooses a random base station as its anycast base station. We denote the network lifetime under this approach as T_{random}. For the ease of comparison among T_{UB}, T_{ABS}, T_{nearest}, and T_{random} across all 90 sets of data, we present the normalized network lifetime for T_{ABS}, T_{nearest}, and T_{random} with respect to T_{UB} for each experiment and denote these normalized network lifetimes as $L_{\text{ABS}} = \frac{T_{\text{ABS}}}{T_{\text{UB}}}$, $L_{\text{nearest}} = \frac{T_{\text{nearest}}}{T_{\text{UB}}}$, and $L_{\text{random}} = \frac{T_{\text{random}}}{T_{\text{UB}}}$, respectively. The normalized network lifetimes, L_{ABS}, L_{nearest}, and L_{random}, are plotted in Fig. 2. Evidently, L_{ABS} is very close to the upper bound of 1 and exhibits a very stable performance. Since the optimal normalized lifetime for the original BS-AR problem lies between L_{ABS} and 1, we conclude that *this upper bound is extremely tight and that the network lifetime performance under ABS is even closer to the optimal solution.*

From Fig. 2, we can see that the heuristic ABS is significantly superior to the nearest approach (in most cases), which not only yields a worse performance than ABS in most cases, but also, exhibits very wide oscillations in network lifetime performance (L_{Nearest}), which is undesirable. Furthermore, the random base station selection approach offers a very poor performance (in most cases) compared to the ABS heuristic. Although in rare cases, the random selection approach solution may coincide with that for the ABS heuristic, in most cases, its performance falls far below that of the ABS algorithm.

Table 2 summarizes the statistical behavior of all the results from these 90 runs, which reveals some quantitative comparison among the approaches. First, in the worst case (among the 90 runs), the ABS algorithm stays within 20.0% of upper bound (even closer to the true optimum). On the

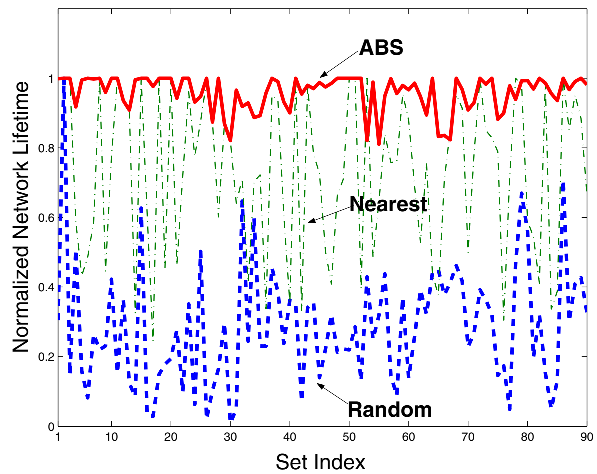


Figure 2. The normalized network lifetimes under the ABS, Nearest, and Random base station selection for 90 data sets.

Table 2. Statistical comparison of the normalized network lifetime for ABS, nearest, and random base station selection approaches.

	Worst Case	Average	95% Confidence Interval
ABS	0.8041	0.9585	[0.9455, 1]
Nearest	0.2445	0.7251	[0.6845, 1]
Random	0.0085	0.2794	[0.2483, 1]

other hand, the worst case performance for the nearest and random base station selection approaches are 75.6% and 99.2% away from the upper bound. Second, on average, the ABS algorithm is within 4.2% of the upper bound, and is 23.3% better than the nearest base station selection approach, and 67.9% better than the random base station selection approach, respectively. The 95% confidence interval for the ABS algorithm is also much narrower than that for the nearest and random base station selection approaches.

To get a sense of what the real (instead of normalized) network lifetimes look like, we list the network lifetimes (all in days) for the last 10 sets of data from the 90 sets of numerical results (with 30 AFNs and 6 base stations) in Table 3. Clearly, the ABS algorithm is much more superior that the nearest and random base station selection approaches in most cases. For set 86 in the table, we find that the nearest approach happens to coincide with ABS and the upper bound. This indicates that for this particular network topology and initial parameters, ABS and the nearest base station selection approach both yield the optimal solution. But in general, the nearest base station selection approach

Table 3. The actual network lifetime performance (in days) for the last 10 sets of data.

Set Number	T_{UB}	T_{ABS}	T_{nearest}	T_{random}
81	129.68	125.68	70.59	25.71
82	105.09	105.09	43.15	33.64
83	172.16	169.81	169.81	26.44
84	193.63	185.15	65.15	9.51
85	185.28	173.41	71.84	24.20
86	30.98	30.98	30.98	21.93
87	106.83	103.35	91.06	32.22
88	47.84	47.84	45.52	19.57
89	131.23	131.22	116.13	56.03
90	92.62	90.90	61.30	30.09

cannot offer good performance as ABS algorithm.

6 Related Work

For the Internet environment, anycast has been addressed extensively (see, e.g., [17]). But the Internet environment is radically different from wireless sensor networks (e.g., severe energy constraint) and thus results on anycast for the Internet may not be directly carried over to wireless sensor networks.

A recent survey on wireless sensor network research is given in [1]. Although there has been active research on energy efficient *unicast* [11, 15] and *multicast* (including broadcast) [5, 8, 16, 23, 24, 25] for wireless sensor networks, there is very limited research on how to perform anycast in such networks.

To the best of our knowledge, the first anycast routing protocol for ad hoc wireless sensor networks was proposed in [13]. Under this protocol, packets are delivered to the nearest sink node. However, energy constraints and lifetime performance were not considered in this effort. As we have shown in Section 5, the nearest sink node approach does not offer good performance for anycast flow routing.

A recent work on anycast routing was presented in [12]. In this effort, Hu et al. studied anycast routing by building a source-based tree. This approach is somewhat similar to the nearest-sink node approach in [13] in the sense that both routings consider minimum energy path. But adapting minimum energy paths does not guarantee a good performance with respect to the network lifetime.

In contrast, our work in this paper focuses on maximizing the network lifetime under energy constraints, and provides different results from the minimum energy path approach. Furthermore, the joint problem of base station selection and anycast flow routing differs from a minimum energy path routing.

7 Conclusions

This paper considers a wireless sensor network having multiple base stations as data sink nodes. Since many real time multimedia applications require to have each source node send all its collected data to one base station for data processing (e.g., video decoding), it is necessary to optimally map each source node to a base station. We have investigated the joint problem of base station selection and anycast flow routing with the aim of maximizing the network lifetime. We propose a heuristic algorithm having polynomial time complexity that is shown to be near-optimal. This result contributes to the basic understanding of anycast routing problems in sensor networks under energy constraint.

Acknowledgements

The work of Y.T. Hou and Y. Shi has been supported in part by NSF under Grants ANI-0312655 and CNS-0347390 and ONR under Grant No. N00014-03-1-0521. The work of H.D. Sherali was supported in part by the NSF under Grant No. DMI-0094462.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks (Elsevier),* vol. 38, no. 4, pp. 393–422, 2002.

[2] BARON Global Optimization Software, `http://archimedes.scs.uiuc.edu/baron/baron.html`.

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

[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] J. Cartigny, D. Simplot, and I. Stojmenovic, "Localized minimum-energy broadcasting in ad-hoc networks," in *Proc. IEEE Infocom,* pp. 2210–2217, San Francisco, CA, March 30–April 3, 2003.

[6] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *Proc. IEEE Infocom,* pp. 22–31, Tel Aviv, Israel, March 26–30, 2000.

[7] J. Chou, D. Petrovis, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. IEEE Infocom,* pp. 1054–1062, San Francisco, CA, March 30–April 3, 2003.

[8] A.K. Das, R.J. Marks, M. El-Sharkawi, P. Arabshabi, and A. Gray, "Minimum power broadcast trees for wireless networks: integer programming formulations," in *Proc. IEEE Infocom,* pp. 1001–1010, San Francisco, CA, March 30–April 3, 2003.

[9] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness, W.H. Freeman and Company,* pp. 245–248, New York, NY, 1979.

[10] W. Heinzelman, *Application-specific Protocol Architectures for Wireless Networks,* Ph.D. thesis, Massachusetts Institute of Technology, June 2000.

[11] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. on Wireless Commun.,* vol. 1, no. 4, pp. 660–670, October 2002.

[12] W. Hu, N. Bulusu, and S. Jha, "An anycast service for hybrid sensor/actuator networks," in *Proc. the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC),* Barcelona, Spain, Sept. 5-8, 2004.

[13] C. Intanagonwiwat and D. De Lucia, "The sink-based anycast routing protocol for ad hoc wireless sensor networks," USC Computer Science Department, Technical Report 95-698, available at `http://www.usc.edu/dept/cs/technical_reports.html`.

[14] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks," *Computer Networks.* vol. 42, issue 6, pp. 697–716, August 2003.

[15] Q. Li, J. Aslam, and D. Rus, "Online power-aware routing in wireless Ad-hoc networks," in *Proc. ACM Mobicom,* pp. 97–107, Rome, Italy, July 16-21, 2001.

[16] W. Liang, "Constructing minimum-energy broadcast trees in wireless ad hoc networks," in *Proc. ACM MobiHoc,* pp. 112-122, Lausanne, Switzerland, June 9-11, 2002.

[17] C. Partridge, T. Mendez, and W. Milliken, "Host anycasting service," RFC 1546, Internet Engineering Task Force, Nov. 1993.

[18] S.S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense sensor network," *IEEE Signal Processing Magazine,* vol. 19, no. 2, pp. 51–60, March 2002.

[19] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment," in *Proc. IEEE Infocom,* pp.404–413, Tel-Aviv, Israel, March 26-30, 2000.

[20] T.S. Rappaport, *Wireless Communications: Principles and Practice,* Prentice Hall, Upper Saddle River, NJ, 1996.

[21] V. Rodoplu and T.H. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, pp. 1333–1344, August 1999.

[22] H.D. Sherali, W.P. Adams, and P.J. Driscoll, "Exploiting special structures in constructing a hierarchy of relaxations for 0-1 mixed integer problems," *Operations Research,* vol. 46, no. 3, pp. 396–405, 1998.

[23] P.-J. Wan, G. Galinescu, X.-Y. Li, and O. Frieder, "Minimum-energy broadcast routing in static ad hoc wireless networks," in *Proc. IEEE Infocom,* pp. 1162–1171, Anchorage, AK, April 22-26, 2001.

[24] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides, "Resource management in energy-limited, bandwidth limited, transceiver-limited wireless networks for session-based multicasting," *Computer Networks,* vol. 39, no. 2, pp. 113–131, June 2002.

[25] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides, "Energy-limited wireless networking with directional antennas: The case of session-based multicasting," in *Proc. IEEE Infocom*, pp. 190–199, New York, NY, June 23-27, 2002.