Energy-Limited Wireless Networking with Directional Antennas: The Case of Session-Based Multicasting

Jeffrey E. Wieselthier
Code 5521
Information Technology Division
Naval Research Laboratory
Washington, DC 20375
wieselthier@itd.nrl.navy.mil

Gam D. Nguyen
Code 5521
Information Technology Division
Naval Research Laboratory
Washington, DC 20375
nguyen@itd.nrl.navy.mil

Anthony Ephremides
Electrical and Computer Eng. Dept.
and Institute for Systems Research
University of Maryland
College Park, MD 20742
tony@eng.umd.edu

Abstract — **We consider ad hoc wireless networks that use directional antennas and have limited energy resources. The performance objectives of such networks depend largely on the application. However, a robust performance measure is the total traffic volume that the network can deliver when all nodes are equipped with a finite and non-renewable amount of energy. We show that the network's lifetime can be extended significantly by incorporating a simple measure of a node's residual energy into the node's cost function. To explore quantitatively the advantage offered by the use of directional antennas over the case of omnidirectional antennas, we consider the case of connection-oriented multicast traffic. Building upon our prior work on multicasting algorithms, we introduce two protocols that exploit the use of directional antennas and evaluate their performance. We observe significant improvement with respect to the omnidirectional case.**

I. INTRODUCTION

The use of directional antennas can provide energy savings and interference reduction by concentrating RF energy where it is needed. Hence they are especially useful in networks with finite energy resources. In this paper, we develop and evaluate algorithms for multicasting that are suitable for use in networks with directional antennas and limited battery capability, and compare performance to that achieved when antennas are omnidirectional. We focus on the problem of tree construction for source-initiated, session-based traffic in all-wireless (i.e., infrastructureless, peer-to-peer, or ad hoc) multihop networks.

In our earlier studies, we developed energy-aware algorithms for the construction of broadcast and multicast trees for networks with omnidirectional antennas. In this context, we demonstrated the superior performance of "node-based" algorithms, which exploit the "wireless multicast advantage" property associated with omnidirectional antennas, namely the capability for a node to reach several neighbors by using a transmission power level sufficient to reach the most distant one. These algorithms are known as Broadcast Incremental Power (BIP) and Multicast Incremental Power (MIP) [1], [2]. Using the incremental power philosophy as a starting point, we demonstrate the issues that arise when directional antennas are used, and develop algorithms that have varying levels of complexity and performance.

We focus on the case in which the nodes are equipped with batteries that cannot be recharged during network operation. Thus, there is a hard constraint of a *fixed quantity of energy at each of the network nodes*. We address some of the fundamental differences between energy-limited and energy-efficient network operation.

In [3], under the constraint of a fixed quantity of energy at each of the network nodes, we presented preliminary results that compare the performance of MIP to that of a more conventional algorithm, which is based on the use of least-cost paths. These studies have demonstrated the superior performance of MIP over a wide range of system parameter values. Additionally, we demonstrated that the lifetime of the network can be extended significantly by incorporating into the tree-construction process a cost-function that reflects the residual energy at the nodes. The present paper extends the results of [3], not only by considering directional antennas, but also by presenting a more-detailed study of the omnidirectional antenna case as well.

In the spirit of assessing the complex trade-offs in wireless multicasting by addressing them one at a time, we do not consider mobility here. However, its impact can be incorporated later since the choice of transmitter power is adjustable and its magnitude determines the connectivity among the neighboring nodes. Thus, the capability to adjust transmission power provides a degree of "elasticity" to the topological connectivity, particularly when the extent of topological change is small, and hence may reduce the need for immediate hand-offs and accurate tracking. Neither do we consider the protocol issues associated with determining connectivity and reserving resources, but rather focus on the basic problems of energy-efficient (or energy-limited) multicasting, assuming the existence of the underlying protocol that supplies the necessary topological connectivity information.

II. ENERGY-LIMITED VS ENERGY-EFFICIENT COMMUNICATION

When a network of wireless links is deployed and the energy reserves at each node are hard-limited, the first question that arises is "what constitutes desirable performance?". To properly address this question, we must rethink the usual premises of energy efficiency, high throughput, low blocking probability, etc. For session-oriented multicast traffic (the focus of this paper), the following conflicting and overlapping requirements are usually posed:

- Network longevity, i.e., the useful life of the network; several alternative definitions are possible, including the time at which the first (and/or last) node in the network runs out of energy, the time at which performance (as defined below) degrades below an acceptable level, the time until the network becomes disconnected, etc.
- High multicast efficiency (i.e., the ability to reach as many of the intended destinations in each multicast session as possible); this quantity may be measured on an

---

This work was supported by the Office of Naval Research.

0-7803-7476-2/02/$17.00 (c) 2002 IEEE.

instantaneous (per session) basis, averaged over a window of recent sessions, or evaluated on a cumulative basis over the lifetime of the network's operation.

- Low blocking probability (as defined by the percentage of session requests that are entirely blocked at the source, i.e., can reach none of the intended destinations).
- High throughput *volume* rather than *rate* (i.e., high total number of bits delivered, which is a quantity that depends on length of session and number of reached destinations).
- Economical use of available energy (as a means for satisfying the previous requirements).
- A specified quality of service, which results in constraints on one or more of the above requirements.

Clearly, all these requirements are interrelated and have different weight and significance, depending on the applications. For example, in sensor networks (as envisioned in commercial and, especially, military applications) the primary requirement is longevity (although at the same time high throughput volume is desired). In other applications of brief duration, the primary requirement is that of high throughput volume (provided the network does not run out of energy prematurely). Any such performance comparisons should be made on the basis of a given, fixed amount of offered traffic load (i.e., rate of session establishment requests and average session duration).

The introduction of hard constraints on the total amount of energy available at each node results in a problem that is very different from that in which unlimited energy is available (although energy efficiency still may be desired). Under such hard constraints on energy (as studied in this paper), the network is capable of operation for a limited period of time. A node dies (and hence can no longer transmit) when its energy is depleted, and the network dies when it is no longer capable of providing a minimum acceptable level of service. By contrast, when the goal is energy efficiency (e.g., delivering the largest number of bits per unit energy), it is implicitly assumed that ample energy is available; in such cases, the use of energy is essentially treated as a cost function.

Energy-efficient operation does not ensure good performance in *energy-constrained* applications. For example, use of the most energy-efficient routes (or multicast trees) may result in premature depletion of energy at some nodes.

A problem that bears some similarity (although many significant differences) to ours was addressed in [4], where the objective was to choose routes to maximize the lifetime of a network of energy-constrained sensor nodes, which are required to deliver their data to <u>any</u> of several gateway nodes. By contrast, we address the problem of source-initiated multicasting, where all nodes have equal capability, and the goal is to form a tree that reaches <u>all</u> members of the group. Also, their model involved constant-rate data flows, whereas we study randomly generated session arrivals and randomly constituted multicast groups.

There are numerous control parameters that can be adjusted to satisfy the requirements listed above. An important one that we do not consider here is admission control. To address it prematurely would open a Pandora's box of difficulties, and we choose to assume that the network tries its best to greedily accept all session requests it can, i.e., a session is rejected or a destination is not reached only if it cannot be reached because of insufficient resources (i.e., transceivers, frequencies, or energy). Another potential control parameter is the transmission rate or other transmission parameter (which can affect session duration, energy usage, quality of service, etc.). We also choose to assume that the channel bandwidth and signal design parameters are set so that the bit rate is fixed.

What remains, and which we do concentrate on here, is the choice of multicast tree for each session. That is, we focus on the selection of multicast routes, which in the wireless environment translate to choosing transmission power and set of receiving neighbor nodes at each level in the multicast tree.

An important feature of our approach, which is enabled by the energy limitations and by the nature of the wireless environment, is the possibility of assigning a "local" metric to each node (and, indirectly, to each potential link) in the network. In this fashion, the session routing problem is amenable to solution methods that are normally applicable to data routing only (e.g., use of "shortest" path trees, distributed algorithms, etc.). This, in its own right, is an innovative feature of our approach.

III. THE MODEL

We consider source-initiated, circuit-switched, multicast sessions. The maintenance of a session requires the dedication of a transceiver at each participating node (source node, relay nodes, and destination nodes), as well as the needed amount of interference-free bandwidth, throughout the duration of the session. The network consists of N nodes, which are randomly distributed over a specified region. Each node has T transceivers, and can thus support up to T multicast sessions simultaneously. We assume that there is a total of F frequencies available to the network. Frequencies can be reused, provided that doing so does not create interference. Thus, congestion (and hence the inability to reach one or more destinations) may arise when either an insufficient number of transceivers or an insufficient number of frequencies are available at one or more nodes along the path. Alternatively, energy-inefficient paths may have to be used when the best paths are not available.

It is also of interest to study systems that use time-division multiple access (TDMA), rather than multiple transceivers or multiple channels, to support multiple sessions simultaneously. In TDMA-based systems, the need to assign specific time slots creates a much more difficult problem than that of simply assigning any transceiver (of perhaps several available) to a new session. Alternatively, it would be possible to consider code-division multiple access (CDMA) [5]. The study of TDMA- and CDMA-based systems is not pursued here, since we want to place emphasis on the energy constraint with as little complication from the MAC layer as possible.

Any node is permitted to initiate multicast sessions. Multicast requests and session durations are generated randomly at the network nodes. Each multicast group consists of the source node plus at least one destination node. Additional nodes may be used as relays either to provide connectivity to all members of the multicast group or to reduce overall energy consumption. The set of nodes that support a multicast session (the source node, all destination nodes, and all relay nodes) is referred to as a *multicast tree.* Notice the difference between this definition and the conventional one that is based on links (or edges); here the links are incidental and their existence depends on the transmission power of each node. Thus it is the nodes (rather than the links) that are the fundamental units in constructing the tree.

The connectivity of the network depends on the transmission power and antenna pattern. We assume that each node can choose its RF power level p^{RF}, such that $p_{\min} \le p^{RF} \le p_{\max}$. The nodes in any particular multicast tree do not necessarily have to use the same power levels; moreover, a node may use different power levels for the various multicast trees in which it participates.

A. Propagation Model

When considering omnidirectional antennas and uniform propagation conditions, we assume that the received signal power is equal to $pr^{-\alpha}$, where p is the transmission power, r is the distance and α is a parameter that typically takes on a value between 2 and 4, depending on the characteristics of the communication medium. Based on this model, the transmitted power required to support a link between two nodes separated by distance r is proportional to r^{α}, since the received power must exceed some threshold.[1] Without loss of generality, we set the threshold constant equal to 1, resulting in:

$$\begin{aligned} p_{ij}^{RF} &= \text{RF power needed for link between Nodes } i \text{ and } j \\ &= \max\{r_{ij}^{\alpha}, p_{\min}\} \qquad (1) \end{aligned}$$

where r_{ij} is the distance between Node i and Node j. The use of a nonzero value of $p_{\min}$ is a way to account for the fact that the $r^{-\alpha}$ dependence applies only in the far-field region (i.e., even when two nodes are arbitrarily close to each other, a nonzero power level $p_{\min}$ is required to support communication between them).

The use of directional antennas can permit energy savings by concentrating transmission energy where it is needed. On the other hand, only the nodes located within the transmitting node's antenna beam can receive the signal, thus possibly diminishing the effect of the wireless multicast advantage. We use an idealized model in which we assume that all of the transmitted energy is concentrated uniformly in a beam of width θ (thus we ignore the possibility of sidelobe interference). Then, the RF power needed by a node to transmit to a distance r using beamwidth θ is

$$p^{RF}(r,\theta) = \max\left\{ r^{\alpha}\frac{\theta}{360},\ p_{\min} \right\}. \qquad (2)$$

Consequently, the use of narrow beams permits energy saving (for a given communication range) or range extension (for a given transmitter power level), as compared to the use of omnidirectional antennas. Specifically, for a given value of $p_{\max}$, the maximum range is increased by a factor of $(360/\theta)^{1/\alpha}$, compared with the case of omnidirectional antennas.

We assume that the beamwidth θ can be chosen so that $\theta_{\min} \le \theta \le \theta_{\max}$. Furthermore, we assume that each node knows the precise locations of its potential neighbors, and that each antenna beam can be pointed in any desired direction to provide connectivity to a subset of the nodes that are within communication range. (In practice, the number of antenna elements needed tends to increase as $\theta_{\min}$ decreases.)

We assume that one antenna beam can be supported for each session in which a node participates; thus the use of directional antennas does not have an impact on the number of sessions that a node can support simultaneously (as compared to an implementation with omnidirectional antennas). Additionally, both θ and the direction in which the beam points are chosen independently for each session in which a node participates. Although setting $\theta = \theta_{\min}$ is appropriate for point-to-point applications, it is often appropriate to use larger values of θ in multicast applications, since a node may have several downstream neighbors, all of which must be included in a single beam (based on the assumption just made above). We discuss the choice of θ in our discussion of multicast algorithms in Section V.

Although we do consider energy expenditures associated with processing at each node (in addition to that for RF transmission), we do not explicitly connect the amount of processing energy with the beamwidth of the antenna. This coupling is deferred for future investigation.

The use of directional receiving antennas would also have a beneficial impact, since background noise and other-user interference would be troublesome only when located within the antenna beamwidth rather than the entire omnidirectional region. Thus, lower signal levels would be needed to provide the required performance. However, we assume the use of omnidirectional receiving antennas to simplify the model.

It is also possible to consider alternative models, which may incorporate one or more of the following:

- fixed beamwidth (i.e., $\theta_{\min} = \theta_{\max}$);
- a single beam per node;
- multiple beams per session;
- constraint on number of beams per node (possibly > T);
- directional receiving antennas.

However, these are not addressed in this paper.

B. Energy Expenditure

In addition to RF propagation, energy is also expended for transmission (encoding, modulation, etc.) and reception (demodulation, decoding, etc.). We define:

p^T = transmission processing power
p^R = reception processing power.

We assume that these quantities are the same at all nodes, and we neglect any energy consumption occurring when the node is simply "on" without transmitting or receiving. The total power expenditure of Node i, when transmitting to Node j, is

$$p_{ij} = p_{ij}^{RF} + p^T + p^R\, 1(\text{Node } i \text{ is a receiving node}) \qquad (3)$$

where the indicator function is included because the p^R term is not needed for the source node. A leaf node, since it does not transmit but only receives, has a total power expenditure of p^R.

We assume that each node starts with a finite quantity of battery energy.[2] For example, Node i has energy $E_i(0)$ at time 0. The *residual energy* at Node i at time t is

$$E_i(t) = E_i(0) - \int_0^t P_i(\tau)\,d\tau \qquad (4)$$

---

[1] This threshold depends on factors such as signal parameters, detector structure, and noise levels (including other-user interference). In this paper, we assume that these characteristics are fixed; thus, the required level of received power is the same at all nodes. Thus, we neglect fading effects that arise in wireless channels.

[2] We assume that the battery has a fixed capacity, i.e., we neglect the fact that the total energy that can be supplied by a battery depends in part on the discharge rate and duty cycle [6]. We also neglect any nonlinear behavior, which may characterize power amplifiers especially at high output levels.

where $P_i(\tau)$ is the total power expended at Node i at time τ.[3] We say that a node is "alive" as long as its residual energy is positive, and that it dies when its residual energy decreases to zero. Based on our assumptions, a "dead" node cannot participate, even as a receive-only leaf node.

IV. THE MULTICASTING PROBLEM

The establishment of a multicast tree requires the specification of the transmitted power levels, the frequencies used by each node, and the commitment of the needed transceiver resources throughout the duration of the session.

We assume that multicast session requests arrive to each of the N nodes at rate λ/N arrivals per unit time. The set of desired destinations is chosen randomly for each arrival. We say that a destination can be *reached* if the following conditions are satisfied:

- there exists a path from the source to it (i.e., the transmitted power required to support the path does not exceed $p_{\max}$ at any node);
- a transceiver is available (i.e., not already supporting another session) at each node along the path;
- a suitable frequency assignment can be found to support the path (i.e., a non-interfering frequency is available to support the link between each node pair in the network along the path; these frequency assignments must not interfere with, or suffer interference from, currently ongoing sessions).

As noted earlier, all multicast requests are accepted as long as one or more of the intended destinations can be reached, and paths are established to all reachable destinations, regardless of the cost required to do so.

A. Performance Measures

In this paper, we focus on one particular performance measure, which is especially well suited for energy-limited applications, namely the total delivered traffic volume during the lifetime of the network. We also consider the related quantity of traffic volume per unit energy.

We first introduce some notation. We assume that, once a session (multicast tree) is established, communication takes place at a constant rate of R bits/s, which is the same for each session request, and which is independent of λ. The duration of session i (d_i) is exponentially distributed with mean $1/\mu = 1$.

Since partial multicast sessions may take place (because some nodes may be unreachable), the performance metric should provide a reward that reflects the number of destinations that are actually reached. We define

n_i = # of intended destinations in session i

m_i = # of destinations reached in session i

P_i = sum of the transmitter powers used by all nodes in session i.

Delivered traffic volume

The delivered traffic volume is directly proportional to both the number of destinations that are reached and to the duration of each session. Specifically, each destination node participating in multicast session i receives $b_i = R\, d_i$ bits during the course of the session. The total quantity of data delivered during session i is then

B_i = total number of bits delivered to all reached destinations in session i

$= m_i\, b_i$.

Then, the total quantity of information delivered to all destinations over an observation interval of X multicast requests is:

$$B_X^{total} = \sum_{i=1}^{X} B_i = R \sum_{i=1}^{X} m_i d_i . \tag{5}$$

Delivered traffic volume per unit energy

The energy expenditure in session i is $P_i\, d_i$. Thus, the total energy expenditure over the observation interval is

$$E_X = \sum_{i=1}^{X} P_i d_i . \tag{6}$$

Therefore, the delivered traffic volume per unit energy over an interval of X arrivals is

$$B_{X,E} = \frac{B_X^{total}}{E_X} = \frac{R \sum_{i=1}^{X} m_i d_i}{\sum_{i=1}^{X} P_i d_i} . \tag{7}$$

B. "Local" Cost Metrics

Tree formation consists of the specification of transmitting nodes and their downstream neighbors. When omnidirectional antennas are used, it is sufficient to specify the set of transmitting nodes and their RF transmission power levels; when directional antennas are used, the antenna pattern must also be specified. It is not feasible to find the multicast trees that guarantee the optimal values of global performance measures such as multicast efficiency, B^{total}, etc. Therefore, we have focused on the development of "local" strategies that depend on "local"[4] metrics, which find the multicast tree that attempts to minimize an appropriate cost function for each new multicast request.

In particular, the basic approach taken in [1] and [2] is to minimize the power needed to maintain the tree associated with each newly arriving session.[5] This power includes the RF transmission power of all transmitting nodes as well as the signal processing power expended at transmitting and receiving nodes. We recognize that local optimization does not guarantee global optimization, e.g., minimizing tree power does not guarantee the minimization of energy over an observation interval of many arrivals. Moreover, even if it were possible to do so, this would certainly not guarantee the optimization of the desired global performance measures. Nevertheless, it has been our experience that this approach works reasonably well.

The problem of finding minimum-power trees in wireless networks is a difficult one. For example, let us consider the broadcasting problem, in which a minimum-cost tree must be found from the source node to all other nodes in the network.

---

[3] Since Node i may be transmitting as a member of several trees simultaneously, $P_i(\tau)$ is the sum of the powers for all such trees at time τ.

[4] "Global" is used here to refer to optimization over a long observation interval. "Local" is used here both in the sense of time-local (i.e., for each arrival of a multicast session request), as well as in the topological sense (i.e., pertaining to an individual link or node).

[5] In Section VI, we introduce a cost metric that also involves the residual energy at each node.

In wired networks, the broadcasting problem can be formulated as the well-known, and easily solved, minimum-cost spanning tree (MST) problem. However, we do not know of any scalable solutions to the node-based version of this problem, for which we developed the Broadcast Incremental Power (BIP) heuristic [1], and we suspect and conjecture that this problem is NP-complete. Related studies of complexity of tree construction and energy-efficient connectivity establishment, which do not exactly apply to our model, can be found in [8], [9], [10].

The multicasting problem is similar to the broadcasting problem, except that only a specific subset of the nodes are required to be in the tree. It is well known that the determination of a minimum-cost multicast tree in wired networks is a difficult problem, which can be modeled as the NP-complete Steiner tree problem, even though the broadcasting problem is easily formulated as the MST problem, which has low complexity. The multicasting problem appears to be at least as hard in wireless networks as it is in wired networks. Thus, heuristics are needed for both broadcasting and multicasting. The two basic approaches we have used for multicasting are the "pruning" of broadcast trees and the superposition of unicast paths [1], [2].

V. ALGORITHMS FOR BROADCASTING AND MULTICASTING WITH DIRECTIONAL ANTENNAS

We have considered two basic approaches for broadcasting and multicasting with directional antennas:

- Construct the tree by using an algorithm designed for omnidirectional antennas; then reduce each antenna beam to the minimum possible width that can support the tree;
- Incorporate directional antenna properties into the tree-construction process.

The first approach can be used with any tree-construction algorithm. The "beam-reduction" phase is performed after the tree is constructed by using an additional "post-processing" algorithm, which is appended to the tree-construction algorithm. The second approach, which requires decisions on beamwidth to be made at each step of the tree construction process, can be used only with algorithms that construct trees by adding one node at a time, such as BIP (and its multicasting counterpart MIP). In this section, we describe these approaches in detail.

A. An Approach based on Beamwidth-Reduction: Reduced Beam BIP (RB-BIP) and MIP (RB-MIP)

First, a low-cost broadcast or multicast tree is formed, using any tree-construction algorithm (e.g., BIP or MIP), under the assumption that the transmitting antennas are omnidirectional. Then, after the tree is constructed in this manner, each transmitting node's antenna beamwidth is reduced to the smallest possible value that provides coverage of the node's downstream neighbors, subject to the constraint $\theta_{\min} \leq \theta \leq 360$. Thus the tree structure is independent of $\theta_{\min}$. We assume perfect antenna patterns that provide uniform gain throughout the cone of beamwidth θ (with no sidelobes), so it is not necessary to extend θ beyond the direction of the nodes at the edges of the cone. When applied to BIP, the resulting scheme is called Reduced-Beam BIP (RB-BIP); when applied to MIP, the resulting scheme is called RB-MIP.

B. An Approach based on Incremental Power: Directional BIP (D-BIP) and MIP (D-MIP)

In [1] and [2] we proposed the Broadcast Incremental Power (BIP) algorithm, a centralized heuristic for the determination of low-power broadcast trees in networks with omnidirectional antennas. BIP is the basis for the Multicast Incremental Power (MIP) algorithm, under which the tree produced by BIP is pruned by eliminating all transmissions that are not needed to reach the members of the multicast group. More specifically, under MIP, nodes with no downstream destinations do not transmit, and some nodes may be able to reduce their transmitted power (i.e., if their more-distant downstream neighbors have been pruned from the tree).

BIP is similar in principle to Prim's algorithm for the formation of minimum-cost spanning trees (MSTs), in the sense that new nodes are added to the tree one at a time (on a minimum-cost basis) until all nodes are included in the tree. In fact, the implementation of this algorithm is based on the standard Prim algorithm, with one fundamental difference. Whereas the inputs to Prim's algorithm are the link costs p_{ij} (which remain unchanged throughout the execution of the algorithm), BIP must dynamically update the costs at each step (i.e., whenever a new node is added to the tree). This updating is done to reflect the fact that the cost of adding a new node to a transmitting node's list of neighbors is the *incremental cost*, i.e., the additional cost associated with adding a new downstream neighbor, given that the node is already transmitting at some particular power level. Consider an example in which Node i is already in the tree (it may be either a transmitting node or a leaf node), and Node j is not yet in the tree. If Node j is already participating in T sessions (hence no transceivers are available for an additional session), the cost of adding it to the tree is set to ∞.[6] Otherwise, for all such Nodes i (i.e., all nodes already in the tree), and Nodes j (i.e., nodes not yet in the tree), the following is evaluated:

$$p'_{ij} = p_{ij} - p_i \tag{8}$$

where p_{ij} is the link-based cost (power) of a transmission[7] between Node i and Node j (i.e., it is $r_{ij}^{\alpha} + p^T$), and p_i is Node i's transmission cost prior to the addition of Node j; (which includes p^T if node i is already transmitting; if Node i is currently a leaf node, $p_i = 0$). The quantity p'_{ij} represents the *incremental cost* associated with adding Node j to the set of nodes to which Node i already transmits. The pair $\{i,j\}$ that results in the minimum value of p'_{ij} is selected, i.e., Node i transmits at a power level sufficient to reach Node j. Thus, one node is added to the tree at every step of the algorithm.

Unlike Prim's algorithm, which guarantees the formation of minimum-cost spanning trees for link-based costs (as in wired networks), BIP does not necessarily provide minimum-cost trees for wireless networks. However, neither do any other scalable algorithms that we are aware of.

The incremental power philosophy, originally developed for use with omnidirectional antennas, can be applied to broadcast tree construction in networks with directional antennas as well. At each step of the tree-construction

---

[6] It is also possible to associate a higher cost with nodes that have low "residual capacity" (i.e., few available transceivers); however, we do not do so in this paper.

[7] We neglect p^R in this cost measure because it is the same for all possible Node j's. However, p^R is included when energy consumption is evaluated.

process, a single node is added, as above. However, whereas the only variable involved in computing the cost (and incremental cost) in the omnidirectional case was the transmitter power, the directional-antenna case involves the choice of beamwidth θ as well. Based on the propagation model of (2), the required RF power increases in proportion to the α power of the distance to the farthest downstream neighbor, and linearly with θ.

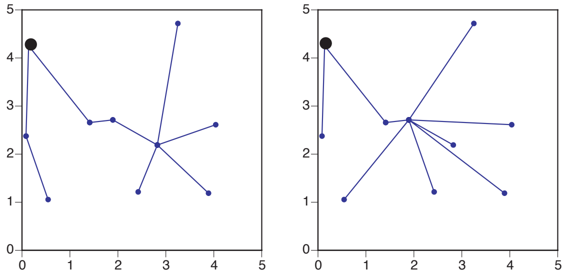
Consider a situation in which Node i is already transmitting to several other nodes. The incremental cost of adding Node j to Node i's set of downlink neighbors depends on the relative location of Node j with respect to the region already included in Node i's antenna's cone of coverage. For example, if Node j is located within the angle of Node i's beam, it suffices to increase Node i's communication range, without changing the width or direction of the beam.[8] On the other hand, if Node j is not located within the angle of Node i's beam, then the beam must be adjusted; this is usually done by increasing θ, although it is sometimes possible to simply shift the beam if all of a node's downstream neighbors are located within a cone not greater than θ_{min}. Thus, to add a new node, it is sometimes sufficient to simply increase transmission range, it is sometimes sufficient to simply shift the beam, sometimes the beam has to be made wider, and sometimes a combination of increased communication range and beam characterization must be done. Note that there is no incremental cost associated with shifting a beam (while maintaining the same angle of coverage).

When applied to the broadcasting problem, the resulting scheme is called Directional BIP (D-BIP). When applied to the multicasting problem, a broadcast tree is formed using D-BIP. To implement Directional MIP (D-MIP), the broadcast tree produced by D-BIP is pruned, as discussed at the beginning of this subsection. Note that when $\theta_{min} = 360$, D-BIP, RB-BIP, and BIP are identical.

C. *Example Broadcast Trees*

Figure 1(a) shows the broadcast tree produced by BIP for a ten-node network, where the source node is shown larger than the other nodes. As noted in Section III.A, RB-BIP uses the same tree as BIP (which is based on omnidirectional antennas); the only difference is that the antenna beamwidth is reduced. Figure 1(b) shows the optimal tree for omnidirectional antennas, which was obtained by exhaustive search. The tree structure, as well as the resulting value of total tree power P, depend on the value of the propagation constant α; our results are based on $\alpha = 2$. Tree power P is listed in the figure caption for $\theta_{min} = 360$ (the omnidirectional case), as well as $\theta_{min} = 30$ and 1. There is relatively little power savings when θ_{min} is reduced below 30 because the two highest-range transmissions require the use of $\theta > 30$ to reach all of their downstream neighbors.

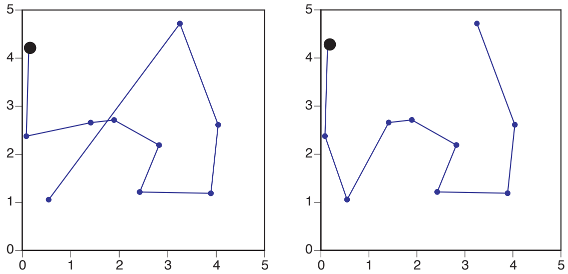
Under D-BIP (unlike RB-BIP), the tree structure depends on the value of θ_{min}. Figure 2(a) shows the tree for the same network for D-BIP with $\theta_{min} = 1$. In this example, D-BIP produces a tree in which each node has only a single downstream neighbor (thus $\theta = \theta_{min}$ at each node) resulting in a zigzag path with no branching. The value of P is greatly reduced by using highly directional antennas. However, this value is 84% greater than that of the optimal tree for $\theta_{min} = 1$, as shown in Fig. 2(b).



(a) BIP, RB-BIP*
$\theta_{min} = 360$: $P = 14.06$
$\theta_{min} = 30$: $P = 4.26$
$\theta_{min} = 1$: $P = 3.99$

(b) optimal (based on $\theta_{min} = 360$)
$\theta_{min} = 360$: $P = 10.71$
$\theta_{min} = 30$: $P = 3.728$
$\theta_{min} = 1$: $P = 3.709$

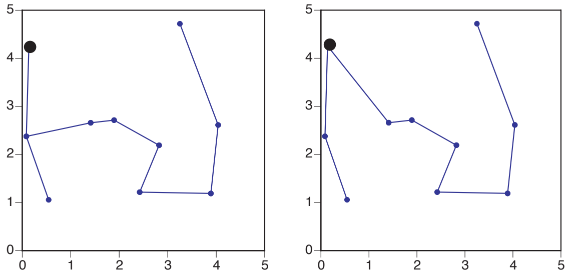
Fig. 1 — Example ten-node broadcast trees based on use of omnidirectional antennas
(*the same tree is used for RB-BIP, independent of the value of θ_{min}).



(a) D-BIP ($P = 0.1051$) (b) optimal ($P = 0.05707$)

Fig. 2 — Example ten-node broadcast trees for $\theta_{min} = 1$.

Figure 3(a) shows the D-BIP tree for the same network, but with $\theta_{min} = 30$. Here, P is 7.2% greater than that of the optimal tree, as shown in Fig. 3(b).



(a) D-BIP ($P = 1.722$) (b) optimal ($P = 1.607$)

Fig. 3 — Example ten-node broadcast trees for $\theta_{min} = 30$.

These results demonstrate that the use of directional antennas can facilitate considerable energy saving through the use of algorithms such as RB-BIP and D-BIP. Moreover, D-BIP provides lower-power trees than RB-BIP for a given value of θ_{min}, and this advantage increases as θ_{min} decreases. However, when θ_{min} is very small, even the tree produced by D-BIP is likely to have a significantly higher value of RF transmission power than the optimal tree (on a percentage basis).

We attribute the relatively good performance of BIP when $\theta_{min} \geq 30$ (as measured by the closeness of tree power to its optimal value, on a percentage basis) to the wireless multicast advantage (see Section III.A). However, this property no longer applies when highly directional antennas are used because power is directly proportional to beamwidth θ; thus, it is costly to expand a beam to accommodate additional nodes.

---

[8] It is also necessary to examine whether Node j could be added to the tree at lower cost by using a different node (e.g., one of Node i's downstream neighbors) as its upstream neighbor.

Therefore, the greedy nature of our incremental power approach suffers when used with extremely narrow beams, and alternative approaches may be desirable.

VI. THE INCORPORATION OF RESOURCE LIMITATIONS

The discussions in the previous sections implicitly assume that sufficient resources are available to implement the trees created by the algorithms. These resources include transceivers, frequencies, and battery energy. In this section we discuss how limitations on these resources are incorporated into our model, and how our algorithms can be modified to cope with limited energy.

It is straightforward to incorporate the impact of a finite number of transceivers. When constructing a tree for a new arrival, the cost of a node is set to ∞ if all of its transceivers are currently supporting other sessions. However, the modeling of finite frequency resources is much more complicated.

A. The Incorporation of Bandwidth Limitations

Let us consider the case in which Node m wants to transmit to Node n. Any particular frequency f may be unusable for one of the following reasons:

- f is already in use (for either transmission or reception) at either Node m or Node n;
- f is being used by one or more nodes that create interference at Node n, thereby preventing reception at f;
- the use of f by Node m would interfere with ongoing communications at other nodes.

In this paper we use the following basic greedy approach for frequency assignment, which we referred to as FA1 in [5]:

> Assume the availability of an infinite number of frequencies when forming the tree (the approach used in [1] and [7]). Then attempt to assign the available frequencies to the tree. The assignment process is complete when either frequencies have been assigned to all transmissions, or when no additional frequencies are available to support portions of the tree.

Under this scheme, the tree construction process ignores the possibility that frequencies may not be available to provide the required connectivity. Thus, if appropriate frequencies cannot be found along the paths to all desired destinations, then some destinations will not be reached. We have used a greedy version, in which frequencies are assigned using an orderly procedure, without the possibility of backtracking to change assignments and without the use of exhaustive search (or other scheme) to determine whether a consistent frequency assignment is possible. Specifically, we simply assign the lowest-numbered available non-interfering frequency to each node. Thus, this scheme can result in unreached destinations, even though they might be reachable through a better frequency assignment. But this is a common characteristic of all heuristic procedures.

In [5] we also considered an alternative scheme (FA2) under which, at each step of the tree-construction, the frequency is chosen along with the transmission power level. Under FA2 the tree is formed using only nodes that do, in fact, have frequencies available. Again, there is no guarantee that all destinations will be reached. However, FA2 provides a richer search space than FA1.

In this paper, we focus exclusively on FA1 because it is simple to use and is applicable to any tree-construction algorithm. FA2 can be used with BIP (and similar schemes in which one node is added to the tree at each step), but not with some of the other algorithms discussed in [1] and [2].

B. The Incorporation of Energy Limitations

Use of a cost metric that involves only the total power required to maintain the tree can result in rapid energy depletion at some nodes. When nodes "die" in this manner, it may be no longer possible to create energy-efficient trees.

We can discourage the inclusion of energy-poor nodes in the multicast tree by increasing the cost associated with their use. In (4) we defined the residual energy at Node i at time t to be $E_i(t)$. We now define the cost of a link between Node i and Node j to be

$$C_{ij} = p_{ij}\left(\frac{E_i(0)}{E_i(t)}\right)^{\beta} \qquad (9)$$

where β is a parameter that reflects the importance we assign to the impact of residual energy.[9] Clearly, when $\beta = 0$, the link cost is simply the power needed to maintain the link.

The incremental cost associated with adding Node j to the set of Node i's downstream neighbors, given that Node i is already transmitting at power level p_i (hence at cost C_i) is:

$$C'_{ij} = C_{ij} - C_i \ . \qquad (10)$$

When β is too small, too much emphasis may be placed on the construction of energy efficient trees, resulting in the rapid depletion of energy at some of the nodes. By contrast, when β is too large, too much emphasis may be placed on balancing energy use throughout the network, while under-emphasizing the need for energy efficiency.

Performance results in Section VII show the beneficial effects of using β in the range [0.5, 2]. It would be possible to develop alternative cost functions to (9) that also discourage the use of energy-poor nodes; we make no claim of optimality. Our objective is to demonstrate that load balancing based on residual energy can extend a network's useful lifetime.

VII. PERFORMANCE RESULTS

Important performance measures for energy-constrained networks include network lifetime and delivered traffic volume. In this section we present our performance results for the two schemes we have developed for directional antennas, namely Reduced-Beamwidth MIP (RB-MIP) and Directional MIP (D-MIP).

We have simulated the performance of RB-MIP and D-MIP for a network of $N = 50$ nodes that are randomly located in a region with dimensions 5×5 (arbitrary units of distance); the same node locations are used in all examples presented in this paper. In extensive performance evaluation, we have observed that these results are representative of other random node distributions as well. We present results for a propagation constant value of $\alpha = 2$, which results in required RF power values of r^2 to support a link between two nodes that are separated by distance r. We set arbitrary values for transmission processing power (p^T) and reception processing

---

[9] Residual energy was incorporated into the cost metric in a similar manner in [4].

power (p^R). In particular, we consider $(p^T, p^R) = (0, 0)$ as well as "moderate" (0.01 , 0.1) and "high" (0.1 , 1) values of these quantities. RF transmission power levels are bounded by $p_{\min} = 0$ and $p_{\max} = 25$ (corresponding to a maximum communication range of 5). In most of our experiments, the initial energy at each node is 200 (arbitrary units, consistent with the units of distance).[10] We demonstrate the impact of incorporating residual energy into the cost metric, and compare performance for $\beta = 0, 0.5, 1$, and 2.

In our simulations, multicast requests arrive with interarrival times that are exponentially distributed with rate λ/N at each node; we have used $\lambda = 1$ in our simulations. Session durations are exponentially distributed with mean 1. Multicast groups are chosen randomly for each session request; the number of destinations is uniformly distributed between 1 and N–1.

Each simulation run consists of $X = 10{,}000$ multicast sessions, some of which may be blocked because of lack of resources (which in general include transceivers, frequencies, and energy). The same random number sequence is used to drive each of our experiments, thereby facilitating a meaningful comparison of results for different values of β.

A. *Network Lifetime*

A fundamental issue in limited-energy applications is network lifetime, i.e., the interval over which the network can provide acceptable levels of service. Clearly, a suitable definition of network lifetime depends on the specific application. For example, in some applications one may view network death as the time at which the first node dies (e.g., see [4]) because it is no longer possible to reach all of the nodes. Alternatively, network death may be defined as the death of a specified fraction of the nodes. In this paper, we don't specify a particular definition of network death, although we do feel that a reasonable definition of acceptable performance would require that at least 50% of the nodes remain alive. Instead, we examine the time evolution of the number of live nodes.

In this subsection, we consider the case of unlimited numbers of transceivers and frequencies, but finite energy at each node. We present results for omnidirectional antennas (although results are qualitatively similar for directional antennas). Thus, we are able to focus on the impact of energy constraints, without addressing other system parameters. In such cases, all desired destinations can be reached, provided that live nodes are available to support the required trees.

Figure 4 shows the evolution of the number of live nodes as a function of the number of session arrivals for $\beta = 0, 0.5, 1$, and 2. Results are shown for the cases of zero and "high" processing power, i.e., $(p^T, p^R) = (0, 0)$ and (0.1, 1), respectively. As noted in Section VI, the use of nonzero values of β tends to discourage the use of nodes that have little residual energy. The use of $0.5 \le \beta \le 2$, rather than 0, results in a significant delaying of the first node's death, and keeps a large fraction (e.g., 80% or 90%) of the nodes alive for a considerably greater number of sessions. Specifically, for zero processing power, when $\beta = 0$, the first node dies at arrival 136; for $\beta = 0.5$, 1, and 2, the first node dies at arrival 563, 668, and 599, respectively. Results are qualitatively similar when $(p^T, p^R) = (0.1, 1)$, except that nodes die much faster because of the energy consumed by signal processing.

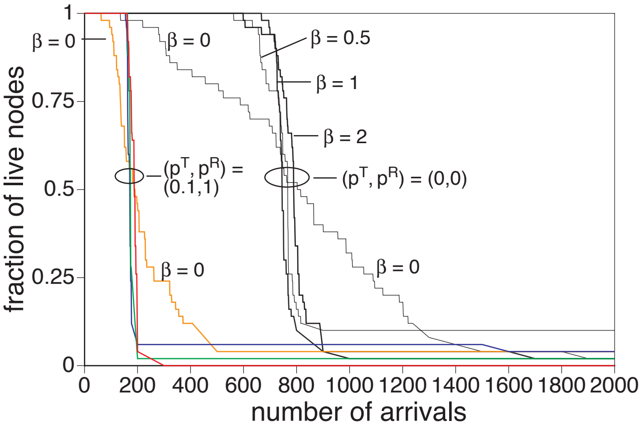


Fig. 4 — Evolution of number of live nodes under MIP with omnidirectional antennas for 50-node network.

Moreover, for $0.5 \le \beta \le 2$, once about 10% of the nodes have died, the fraction of live nodes decreases to below 10% shortly thereafter. The rapid death of nodes in this manner is not a harmful effect. Once about 50% of the nodes are dead, a significant number of the remaining live nodes are typically unreachable. Thus, the fact that use of $\beta = 0$ maintains a certain fraction (say 25%) of the nodes alive considerably longer than use of larger values of β is not beneficial.

Thus, for $0.5 \le \beta \le 2$ we have achieved a high degree of load balancing that keeps almost all of the nodes alive for a relatively long time, thereby maintaining network connectivity and high levels of throughput much longer than for the case in which $\beta = 0$. In view of the relative insensitivity of node lifetime to the value of β (in the region $0.5 \le \beta \le 2$), we use $\beta = 1$ in the examples presented in this paper. No claim for optimality is made.

Since we use a finite value of $p_{\max}$, it is typical to achieve a final state in which a number of nodes still have energy, but further communication is impossible because of a lack of connectivity among the live nodes.

B. *Delivered Traffic Volume*

We now consider the delivered traffic volume B^{total}. In doing so, we address the impact of realistic constraints on the number of transceivers (T) available at each node and on the number of frequencies (F) available for communication. Our modeling assumptions are the same as those of the previous subsection. Unlike the case of infinite transceiver and frequency resources, performance depends strongly on the arrival rate λ because high traffic loads require a large number of transceivers and frequencies to support them. We present results for MIP, first for omnidirectional and then for directional antennas. Our results are based on the use of frequency assignment scheme FA1.

Figure 5 shows the time evolution of B^{total} under MIP, with omnidirectional antennas, for several sets of (F,T) pairs for $\beta = 0$, $\lambda = 1$, and $(p^T, p^R) = (0, 0)$. One unit on the vertical axis corresponds to the delivery of a message of average length (one time unit) to a single destination (see definition in (5)). The initial value of energy at each node is $E_i(0) = 200$.

---

[10] We assume that if a node is alive at the beginning of a session, it will be able to complete the session (regardless of whether it is a transmitting or a receive/only node). Thus, we neglect the minor "end effects" associated with a node's death during a session.

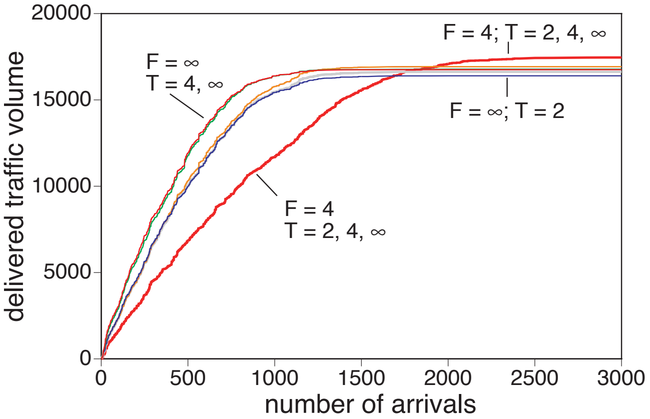


Fig. 5 — Evolution of cumulative bit volume under MIP for several sets of (F, T) pairs ($\beta = 0$; $(p^T, p^R) = (0, 0)$).

Results for nine sets of (F, T) pairs are shown, namely the cases for which $F = 4, 8$, and ∞ and $T = 2, 4$, and ∞. Three of the curves are significantly lower than the others during the early phase of the simulation (i.e., for approximately the first 1250 arrivals); these are the curves for $F = 4$. Among the sets of (F, T) pairs, the highest final value is achieved for $F = 4$ (the precise value in this case is nearly independent of the value of T). This value is 6.5% greater than the lowest final value, which occurs for $(F, T) = (\infty, 2)$.

Figure 6 shows similar results for $\beta = 1$. Qualitatively, performance is similar to that for $\beta = 0$ in some ways. In particular, the three curves for $F = 4$ are again significantly lower than the others in the early part of the simulation, and somewhat higher at the end. However there are significant differences as well. For each (F, T) pair, the curve can be approximated well by a linear increase until the final value is reached, a departure from the asymptotic performance observed for $\beta = 0$. This behavior can be explained by the fact that the use of $\beta = 1$ results in the rapid transition from a state in which most nodes are alive to one in which most are dead, as shown in Fig. 4. Thus, there are two distinct regions of operation. When all (or most) nodes are alive, the rate of traffic delivery is maintained at (or near) its maximum value. When most nodes are dead, the rate of traffic delivery is close to (or equal to) zero. We also observe that the highest final value, which occurs for $(F, T) = (4, 4)$ and $(4, \infty)$ is 14.5% greater than the lowest value, which occurs for $(F, T) = (\infty, 2)$. This percentage difference is more than twice that observed for $\beta = 0$.

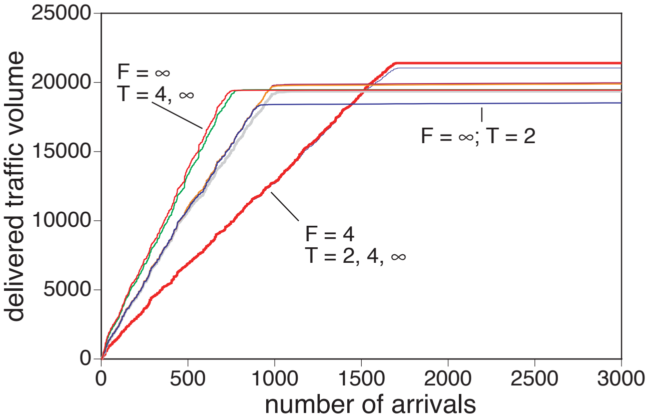


Fig. 6 — Evolution of cumulative bit volume under MIP for several sets of (F, T) pairs ($\beta = 1$; $(p^T, p^R) = (0, 0)$).

We now consider the case of directional antennas. Figure 7 shows the time evolution of B^{total} for RB-MIP and D-MIP for several values of $\theta_{\min}$. Results are shown for $\beta = 1$, zero processing power, and $T = F = \infty$. The case of $\theta_{\min} = 360$ corresponds to the use of omnidirectional antennas. Our first observation is that the use of RB-MIP and D-MIP provide significantly increased values of delivered traffic volume, and that this volume increases as $\theta_{\min}$ decreases. The increase is less than linear in $1/\theta_{\min}$ because some beamwidths may be greater than $\theta_{\min}$.

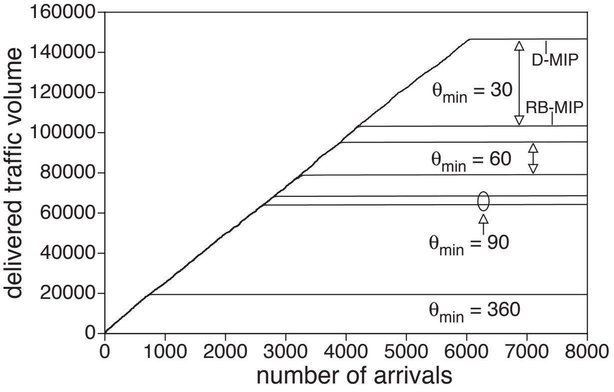


Fig. 7 — Evolution of cumulative bit volume under MIP with directional antennas for D-MIP and RB-MIP; $T = \infty$, $F = \infty$ ($\beta = 1$; $(p^T, p^R) = (0, 0)$).

For $\theta_{\min} = 30$, 60, and 90, two curves are shown for each value; the lower curve is for RB-MIP and the upper curve is for D-MIP. In all cases, D-MIP provides better performance than RB-MIP, and its advantage increases as $\theta_{\min}$ decreases. Like Fig. 6, the curves can be closely approximated by straight lines until the final value is reached. The slope of the curve is independent of $\theta_{\min}$.

Figure 8 shows similar results for finite transceiver and frequency resources, namely $T = 4$ and $F = 8$. The same observations made for infinite resources apply here as well, although there are slight differences in total traffic volume, the point at which the curves reach their final values, and the slope for $\theta_{\min} = 360$.

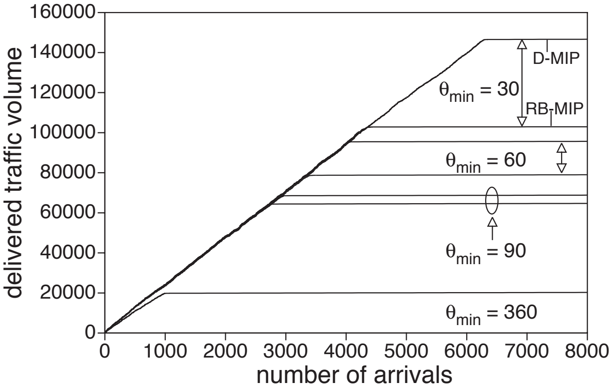
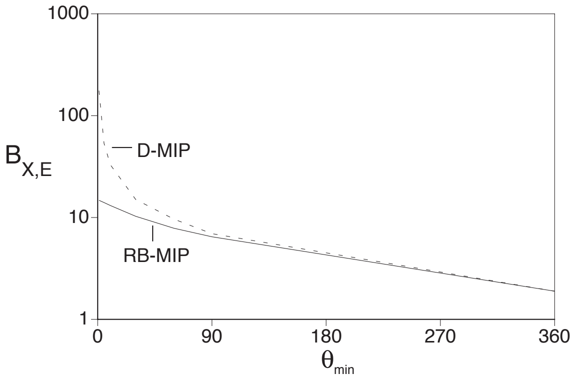
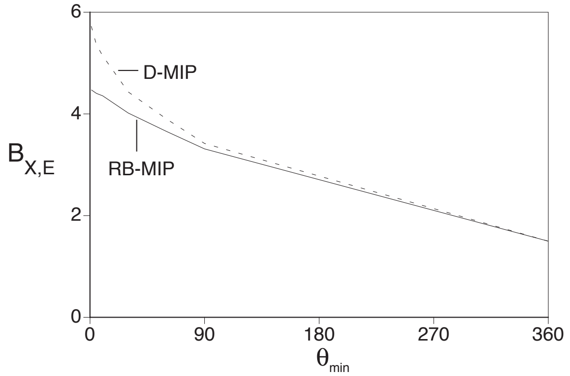


Fig. 8 — Evolution of cumulative bit volume under MIP with directional antennas for D-MIP and RB-MIP; $T = 4$, $F = 8$ ($\beta = 1$; $(p^T, p^R) = (0, 0)$).

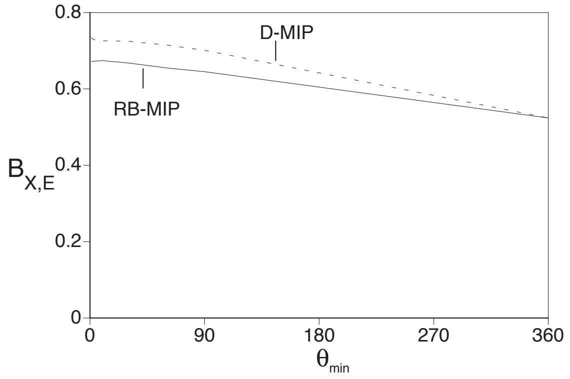
It is also of interest to study the dependence of traffic volume on $\theta_{\min}$. Figure 9 shows $B_{X,E}$, the total number of bits delivered per unit energy over the entire lifetime of the network (in this case until no pair of live nodes is within communication range), as a function of $\theta_{\min}$, for both RB-MIP and D-MIP.



(a) $(p^T, p^R) = (0, 0)$



(b) $(p^T, p^R) = (0.01, 0.1)$



(c) $(p^T, p^R) = (0.1, 1)$

Fig. 9 — Bit volume per unit energy vs $\theta_{\min}$ for D-MIP and RB-MIP ($\beta = 1$).

Figure 9(a) shows $B_{X,E}$ for $(p^T, p^R) = (0, 0)$ and $\beta = 1$. Consistent with the results presented above, D-BIP provides better performance than RB-BIP, and this difference increases as $\theta_{\min}$ decreases. There is little difference in performance for $\theta_{\min} > 90$. However, there is approximately an order of magnitude difference for $\theta_{\min} = 1$ (the smallest value for which results were obtained).

Figures 9(b) and 9(c) show the impact of processing power, for the cases of $(p^T, p^R) = (0.01, 0.1)$ and $(0.1, 1)$, respectively. The most obvious impact of processing power is the reduced value of $B_{X,E}$. Since energy is now expended for signal processing, less is available for RF transmission. Therefore, the overall delivered traffic volume is reduced greatly (note that the vertical scale is logarithmic in Fig. 9(a) and linear in the others). Moreover, the advantage of using D-BIP decreases as processing power increases, again because a smaller fraction of energy is available for RF transmission.

VIII. CONCLUSIONS

In this paper, we have identified the fundamental issues that arise in all-wireless networks that are subject to hard constraints on energy, and we have addressed the similarities and differences between energy-limited and energy-efficient operation. We have studied the problem of source-initiated, session-based multicasting, and have developed algorithms that are suitable for use with directional antennas.

One of these algorithms, Reduced-Beamwidth MIP (RB-MIP), uses the trees formed by MIP under the assumption of omnidirectional antennas, and then reduces the beamwidth to concentrate the RF energy in the cone where it is needed. The other, Directional-MIP (D-MIP), exploits the directionality of the antennas throughout the tree-construction process.

We have shown that the incorporation of residual energy into local cost metrics, which results in load balancing that spreads the burden of energy use among more of the nodes, has a considerable impact on network performance. Most importantly, we have shown that the time of the first node's death can be delayed significantly, thus permitting operation at maximum throughput rates much longer than is possible when a criterion of minimum-power trees is used. Additionally, the overall volume of data that is delivered is increased. System operation is highly robust with respect to the residual-energy parameter β; values between 0.5 and 2 have been shown to work well.

Both RB-MIP and D-MIP provide significant improvement in terms of network lifetime and total delivered traffic volume, as compared to MIP with omnidirectional antennas (except when signal-processing power dominates energy expenditure, in which case the improvement is small). The improvement is greatest for small values of $\theta_{\min}$. Moreover, D-MIP provides significantly better performance than RB-MIP, especially for small values of $\theta_{\min}$ and small values of processing power.

REFERENCES

[1] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks," *Proc. IEEE INFOCOM 2000*, pp. 585-594, March 2000.

[2] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy-efficient broadcast and multicast trees in wireless networks," *Mobile Networks and Applications (MONET)*, in press.

[3] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy efficiency in energy-limited wireless networks for session-based multicasting," *Proc. 2001 Spring Vehicular Technology Conference*, May 2001.

[4] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," *Proc. IEEE INFOCOM 2000*, pp. 22-31, March 2000.

[5] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy-efficient wireless multicast of session traffic," *Proc. Hawaii International Conference on System Sciences (HICSS-34)*, January 2001.

[6] M. Zorzi and R. R. Rao, "Error control and energy consumption in communications for nomadic computing," *IEEE Trans. Computers*, **46**-3, pp. 279-289, March 1997.

[7] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Algorithms for energy-efficient multicasting in static ad hoc wireless networks," *Mobile Networks and Applications (MONET)*, **6**, pp. 251-263, June 2001.

[8] L. Viennot, "Complexity results on election of multipoint relays in wireless networks," Report RR-3584, INRIA, December 1998.

[9] W.-T. Chen and N.-F. Huang, "The strongly connecting problem on multihop packet radio networks," *IEEE Trans. Communications*, **37**-3, pp. 293-295, March 1989.

[10] A. E. F. Clementi, P. Penna, and R. Silvestri, "Hardness results for the power range assignment problem in packet radio networks," *Proc. Third International Workshop on Randomization and Approximation Techniques in Computer Science*, pp. 197-208, 1999.