Computer Networks 39 (2002) 113–131
www.elsevier.com/locate/comnet

Resource management in energy-limited, bandwidth-limited, transceiver-limited wireless networks for session-based multicasting ☆

Jeffrey E. Wieselthier [a,*], Gam D. Nguyen [a], Anthony Ephremides [b]

[a] *Information Technology Division, Code 5521, Naval Research Laboratory, Washington, DC 20375, USA*
[b] *Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA*

Received 12 September 2001; accepted 12 November 2001

Responsible Editor: I.F. Akyildiz

Abstract

In this paper we consider source-initiated multicast session traffic in an ad hoc wireless network, operating under hard constraints on the available transmission energy as well as on bandwidth and transceiver resources. We describe the similarities and differences between energy-limited and energy-efficient communications, and we illustrate the impact of these overlapping (and sometimes conflicting) considerations on network operation. In energy-limited applications, fundamental objectives include the maximization of a network's useful lifetime and the maximization of traffic that is delivered during this lifetime. We demonstrate how the incorporation of residual energy into the cost metric used for tree construction can provide improved performance based on these criteria. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Wireless multicast; Energy-efficient; Energy-limited; Ad hoc networks

1. Introduction

In applications where nodes in an all-wireless multihop network are equipped with batteries that cannot be recharged during network operation, battery energy is a precious resource that must be carefully managed. In this paper, we consider multicasting in the context of precisely this kind of energy limitation. Specifically, we address the problem of multicasting for session traffic in resource-limited all-wireless (i.e., infrastructureless, peer-to-peer, or ad hoc) multihop networks. At the same time, we assume limited bandwidth and transceiver resources, but our emphasis is on managing the energy resource.

In [1,2] we developed the multicast incremental power (MIP) algorithm for the construction of multicast trees for such networks, and demonstrated the energy efficiency of the trees it

☆ This work was supported by the Office of Naval Research.
* Corresponding author.
E-mail addresses: wieselthier@itd.nrl.navy.mil (J.E. Wieselthier), nguyen@itd.nrl.navy.mil (G.D. Nguyen), tony@eng.umd.edu (A. Ephremides).

1389-1286/02/$ - see front matter
PII: S1389-1286(01)00303-6

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2002**	2. REPORT TYPE	3. DATES COVERED **00-00-2002 to 00-00-2002**
4. TITLE AND SUBTITLE **Resource management in energy-limited, bandwidth-limited, transceiver-limited wireless networks for session-based multicasting**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Code 5221,4555 Overlook Avenue, SW,Washington,DC,20375**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**19**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

produces. In [3,4] we evaluated its performance under realistic conditions involving a wide range of traffic loads and the joint constraints of a finite number of transceivers at each network node and a finite number of available frequencies. In this paper, we introduce the additional hard constraint of a *fixed quantity of energy at each of the network nodes*, and in this setting we compare the performance of MIP to that of a more conventional algorithm, which is based on the use of least-cost paths. We show that the introduction of a hard constraint on energy (in addition to equipment and bandwidth constraints) creates a significantly different networking environment in which even the choice of a performance measure becomes unexpectedly complicated.

A crucial aspect of the performance of multicasting algorithms is the choice of "link" or "node" metric used in tree construction. This metric should reflect the specific performance criteria associated with the problem of interest. Here, we show how the finite available energy at each node can be factored into an appropriate node metric, resulting in significant performance enhancement. Our earlier work on energy-efficient multicasting assumed the availability of unlimited energy, but required its economical use. So, in addition to considering what constitutes an appropriate overall performance measure (which, as we will discuss, is a thorny issue in its own right), we need to consider the "local" metrics that are expected to relate closely to the chosen performance measure.

We first argued in [5] that "node-based" approaches are needed for wireless networking applications, and in [1,2] we demonstrated how the characteristics of the wireless medium can be exploited, while departing from the conventional layered protocol structure. For example, MIP is based on jointly choosing transmitted power levels (and hence determining connectivity) and constructing the multicast tree (a routing function). We demonstrated that MIP provides considerably better performance than "link-based" approaches (which are adaptations of schemes developed for wired networks) over a wide range of system parameters. The scope of [1,2] was limited to tree construction, however, and did not address the operation of a multicast network over an extended period of time and with finite resources.

Examples of the few studies that have addressed multicasting in wireless networks include [6–8], but these have not addressed energy-related issues. Virtually all multicasting studies have been limited to the case of stationary networks that are not wireless (e.g., [9–11]).

To assess the complex trade-offs in wireless multicasting, it is necessary to address them one at a time. For example, in this paper we do not consider mobility. 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. We also do not consider the protocol issues associated with determining connectivity and reserving resources, but rather focus on the basic problem of energy-efficient multicasting, assuming the existence of the underlying protocol that supplies the necessary topological connectivity information.

In Section 2 we present our basic communication model, including a characterization of the resources available at the nodes, and we discuss some of the characteristics of wireless networks that distinguish them from wired networks. In Section 3 we discuss the differences between energy-limited and energy-efficient communications, including their impact on communication problems. In Section 4 we discuss the multicasting problem, including performance measures and the algorithms we have developed. In Section 5 we incorporate the impact of energy limitations into the cost function used in tree construction, and present performance results for the case in which an unlimited supply of transceivers and frequencies is available, thus permitting us to assess the impact of finite energy on performance. In Section 6 we show how the impact of finite transceivers and frequencies is incorporated into our model, and we present performance results. Finally, in Section 7 we present our conclusions.

2. The model

We consider source-initiated, circuit-switched, multicast sessions. [1] The maintenance of a session (using a multicast tree) requires the dedication of a transceiver at each participating node (source node, relay nodes, and destination nodes) throughout the duration of the session. The network consists of N nodes, which are randomly distributed over a specified region. A node can transmit and receive simultaneously, but must do so at different frequencies. Each node has T transceivers, and can thus support up to T multicast sessions simultaneously, each using different transmit and receive frequencies. 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. Congestion (and hence call rejection) may arise when either an insufficient number of transceivers or an insufficient number of frequencies are available.

As a result of these considerations, an intermediate node in the tree requires two frequencies (i.e., one for transmission and one for reception) for each session it participates in. [2] Thus, for a node to make full use of its T transceivers (by supporting T sessions), up to $2T$ frequencies would be needed. Since other sessions may be active in the vicinity of the node of interest, the number of frequencies needed to make full use of the available transceiver resources may be considerably larger than this number.

It is also of interest to study systems that use time-division multiple access (TDMA), rather than multiple transceivers, 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 (of perhaps several available) transceiver to a new session. Alternatively, it would be possible to consider a system that uses code-division multiple access (CDMA) [4]. 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. We assume that each node can choose its transmission power level p^{RF}, such that $p_{\min} \leqslant p^{\mathrm{RF}} \leqslant p_{\max}$. 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.

We assume that the received signal power is equal to $p^{\mathrm{RF}} r^{-\alpha}$, where 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. [3] Without loss of generality, we set the threshold constant equal to 1, resulting in:

[1] Unicasting (i.e., single-destination communication), which is a special case of multicasting, also benefits from the techniques presented in this paper.

[2] The source node requires only a single frequency since it is a transmit-only node; similarly, leaf nodes require a single frequency because they do not transmit.

[3] 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.

$$p_{ij}^{\mathrm{RF}} = \text{RF power needed for link between Node } i \text{ and Node } j = \max\{r_{ij}^{\alpha}, p_{\min}\} \quad (1)$$

where r_{ij} is the distance between Node i and Node j. If the maximum permitted transmitter power $p_{\max}$ is sufficiently large, the network is fully connected. 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).

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, although it would be straightforward to incorporate it into our model. The total power expenditure of Node i, when transmitting to Node j, is:

$$p_{ij} = p_{ij}^{\mathrm{RF}} + p^{\mathrm{T}} + p^{\mathrm{R}} 1\ (\text{Node } i \text{ is a receiving node}) \quad (2)$$

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. [4] For example, Node i has energy E_i (0) at time 0. The *residual energy* at Node i at time t is:

$$E_i(t) = E_i(0) - \int_0^t P_i(\tau)\mathrm{d}\tau \quad (3)$$

where $P_i(\tau)$ is the total power expended at Node i at time τ. [5] 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.

Thus far, we have addressed only point-to-point (single destination) communication. However, since we assume the use of omnidirectional antennas, all nodes within communication range of a transmitting node can receive its transmission. In such cases, we can exploit the "wireless multicast advantage," first described in [5]. For example, consider a situation in which Node i transmits directly to its neighbors, Node j and Node k; the power required to reach Node j is p_{ij} and the power required to reach Node k is p_{ik}. A single transmission at power $p_{i,(j,k)} = \max\{p_{ij}, p_{ik}\}$ is sufficient to reach both Node j and Node k (rather than the sum of these powers, as in wired applications).

As a result of the wireless multicast advantage, the omnidirectional wireless communication medium can be viewed as a *node-based* environment. By contrast, in wired models, as long as there is a wire or cable link connecting two nodes, the reception is ensured over that link, and the cost of Node i's transmission to Node j and Node k would be the *sum* of the costs to the individual nodes. Thus, wired networks are *link-based*.

The node-based nature of wireless networks necessitates the development of new networking techniques, because the models developed for wired networks do not adequately capture the characteristics of the wireless medium. 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. 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 incre-

[4] 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 [12]. We also neglect any nonlinear behavior, which may characterize power amplifiers especially at high output levels.

[5] 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.

mental power (BIP) heuristic [1,2]. Related studies of the complexity of tree construction and energy-efficient connectivity establishment, which do not exactly apply to our model, can be found in [13–15].

3. Energy-limited vs. energy-efficient communications

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 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 (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 (the problem 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 [16], 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 any 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 all 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.

4. A multicasting problem

We now address the problem of determining an appropriate multicast tree for each arriving multicast session request, so that a reward function (which incorporates throughput, while reflecting the finite quantity of energy at each node) is maximized. 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 multicast 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 noninterfering 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 (subject to the restriction that the transmitted power does not exceed $p_{\max}$ at any node).

4.1. Performance measures

Qualitatively, our goal is to maximize the amount of communication that is delivered, subject to the constraint of finite energy at each node, while operating under the admit-all admission-control policy discussed above. We consider two alternative (partially overlapping and partially conflicting) views of this objective, namely:

- Maintain operation at (or near) the best possible performance (i.e., reach most, if not all, destinations) for as long as possible.
- Maximize the total quantity of information that is delivered to the destinations.

Before defining our performance measures, we introduce some notation. We assume that, once a session (multicast tree) is established, communication takes place at a constant bit rate of:

$R =$ data rate in bits/s,

which is the same for each session request, and which is independent of λ. The session duration

$d_i =$ duration of session i

is exponentially distributed with expected duration $1/\mu = 1$.

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.

The following performance measures are studied in this paper.

4.1.1. Multicast efficiency

We define the *multicast efficiency* e_i of the ith multicast session to be the fraction of desired destinations of that session request that are actually reached: $e_i = m_i/n_i$. Then, the *cumulative multicast efficiency* over an observation interval of X multicast requests can be defined as:

$$e = \frac{1}{X}\sum_{i=1}^{X} e_i = \frac{1}{X}\sum_{i=1}^{X}\left(\frac{m_i}{n_i}\right). \tag{4}$$

4.1.2. Delivered traffic volume

Multicast efficiency, which defines performance in terms of the fraction of destinations that are reached, does not directly incorporate the duration of the sessions. Thus, the reward that e associates with short sessions is equal to that of long sessions, even though the latter result in the delivery of more information. It is also of interest to evaluate the amount of information (i.e., total number of bits) that is delivered to the desired destinations. This quantity 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:

$b_i = Rd_i$

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:

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

4.2. "Local" cost metrics

It is not feasible to find the multicast trees that guarantee the optimal values of the global performance measures we have studied, e.g., e and B_X^{total}. Therefore, we have focused on the development of "local" strategies that depend on "local" [6] metrics, which find the multicast tree that attempts to minimize an appropriate cost function for each new multicast request. The cost function has been chosen with the goal of providing behavior that is monotonically related to the global performance measure.

Tree formation consists of the choice of transmitting nodes and their transmitting powers. Our BIP algorithm (see Section 5) uses node-based metrics in the construction of the tree, rather than the more-conventional link-based metrics. Link-based metrics assign a cost to each link, e.g., the power needed to maintain the link. The total cost of a multicast tree is then the sum of the costs of the links that form the tree. However, such metrics do not reflect the wireless multicast advantage property, discussed in Section 2. Instead, the total cost of a multicast tree should be evaluated as the sum of the costs of the transmitting nodes that form the tree, as we do with MIP.

[6] "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).

4.3. Algorithms for multicasting

Once a local cost metric has been chosen, the minimum-cost multicast tree problem is well defined. 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. As we noted earlier, we know of no scalable algorithms for the minimum-energy broadcast problem. Thus, heuristics are needed.

We have considered two basic approaches for multicasting:

- Pruning the broadcast tree.
- Superposing the minimum-cost unicast paths to each individual destination.

Examples of these approaches, which are described more thoroughly in [1,2], are discussed below.

(1) *An approach based on pruning:* First, a low-cost broadcast tree is formed, based on the chosen local (i.e., link- or node-based) metric. To obtain the multicast tree, the broadcast tree is pruned by eliminating all transmissions that are not needed to reach the members of the multicast group. More specifically, nodes with no downstream destinations will not transmit, and some nodes will be able to reduce their transmitted power (i.e., if their more-distant downstream neighbors have been pruned from the tree).

In this paper, we focus on the MIP algorithm, which is a pruned version of the BIP algorithm, discussed below. We have also applied the same pruning technique to BLiMST (a link-based MST algorithm studied in [1,2], resulting in the algorithm MLiMST).

BIP is similar in principle to Prim's algorithm for the formation of 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) to reflect the fact that the cost of adding new nodes to a transmitting node's list of neighbors is the *incremental cost*, as defined below. To facilitate the discussion, we assume in this section that the local cost metric is transmission power. 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, the cost of adding it to the tree is set to ∞. [7] 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:

$$p'_{ij} = p_{ij} - p_i, \tag{6}$$

where p_{ij} is the link-based cost (power) of a transmission [8] between Node i and Node j (i.e., it is $r_{ij}^{\alpha} + p^{\mathrm{T}}$), 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.

(2) *An approach based on unicast paths:* A minimum-cost path is established between the source and every desired destination separately (as in the classic routing problem), where the cost of a path is the sum of the costs (powers) of the links on the path. The multicast tree consists of the

[7] 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.

[8] The cost of the link is set to ∞ if p_{ij}^{RF} exceeds $p_{\max}$. We neglect the receiver processing power in this cost measure because it is the same for all possible Node j's.

superposition of the appropriate unicast paths. The three algorithms most often used for finding shortest paths are the Dijkstra, Bellman–Ford, and Floyd–Warshall algorithms [17]. We refer to this as the multicast least-unicast-cost (MLU) algorithm [1,2].

(3) *On the effectiveness of alternative schemes:* In [1,2] we evaluated our multicasting schemes on the basis of the total RF transmitter power of the trees they produce, i.e., the local cost metric used there was simply the RF transmission power. Performance results (for an unlimited number of frequencies) indicate that multicasting schemes based on pruning (MIP and MLiMST) tend to work better than MLU when the number of destinations is a relatively large fraction of the total number of nodes (e.g., 25% or greater), whereas MLU works better than MIP and MLiMST when the fraction of nodes that are destinations is small (e.g., 10% or less). In all cases, MIP provides better performance than MLiMST. We attribute the superiority of MIP to the fact that it exploits the node-based wireless multicast advantage property, whereas MLiMST ignores this property as it forms trees on the basis of link-based costs. In this paper, we present results for MLU and MIP.

(4) *The sweep: removing unnecessary transmissions:* In [1,2] we noted that the performance of our broadcast algorithms can be improved somewhat [9] by using what we call the "sweep" operation, which detects redundant transmissions as well as transmissions that can be reduced in power. The numerical results presented in this paper are based on a version of the sweep in which the entire tree is constructed before searching for opportunities to improve performance. We observed in [1,2] that this approach typically provides better performance than an alternative approach in which a sweep is performed at each step during the tree construction.

[9] The percentage improvement achieved by the sweep is somewhat greater for BLU and BLiMST (typically 5–20%) than for BIP (typically 5–10%), but BIP typically provides better performance than the other schemes (both pre- and post-sweep).

5. Performance results: energy-limited systems with ample transceiver and frequency resources

In this section, we present performance results for systems with limited energy at each node, but an infinite number of transceivers and frequencies. By studying such systems, we are able to isolate the impact of energy limitations. We first discuss how energy limitations are incorporated into the cost function used to construct trees, and then present performance results.

5.1. The incorporation of energy limitations

If the cost metric does not reflect the constraint of finite energy at each node, the greedy nature of MIP and MLU (both of which construct trees without regard to the residual energy available at the nodes) 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, and performance (in the sense of the chosen performance measure) can rapidly deteriorate, as we demonstrate in Section 7.

We can discourage the inclusion of energy-starved nodes in the multicast tree by increasing the cost associated with their use. In Eq. (1) 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

$$C_{ij} = p_{ij}\left(\frac{E_i(0)}{E_i(t)}\right)^{\beta}, \tag{7}$$

where β is a parameter that reflects the importance we assign to the impact of residual energy. [10] Clearly, when $\beta = 0$, the link cost is simply the power needed to maintain the link.

For BIP and MIP, we follow the approach used in the previous section to define C_i to be the nodal cost associated with Node i prior to the addition of Node j. We have

$$C_i = p_i\left(\frac{E_i(0)}{E_i(t)}\right)^{\beta}. \tag{8}$$

[10] Residual energy was incorporated into the cost metric in a similar manner in [16].

We modify BIP for the finite-energy case by adding the Node j that results in the smallest incremental cost:

$$C'_{ij} = C_{ij} - C_i, \tag{9}$$

rather than the smallest incremental power. When β is too small, too much emphasis may be placed on the construction of energy efficient trees, resulting in the 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.

The incorporation of residual energy into the cost metric in this manner is a heuristic approach, and no claim for optimality is made. Its effectiveness is demonstrated in the following subsection.

5.2. Performance results

We first discuss our performance results for examples with unlimited numbers of transceivers and frequencies, but finite energy at each node. Since sufficient transceiver and frequency resources are available, all desired destinations can be reached, provided that the nodes have sufficient energy to support the required trees. In particular, we compare the performance of MIP and MLU, and show that the incorporation of residual energy into the cost metric has a significant impact on increasing the useful operating lifetime of the network. In Section 6 we discuss the incorporation of limited numbers of frequencies and transceivers into the model, and we present performance results for such cases.

We have simulated the performance of MIP and MLU 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. We present results for a propagation constant value of $\alpha = 2$, which results in a required RF power value 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 power (p^R). At one extreme, we neglect both of these quantities by setting them equal to zero; we also consider "moderate" values ($p^T = 0.01$ and $p^R = 0.1$) and "high" values ($p^T = 0.1$ and $p^R = 1$) 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 Section 6.4 we discuss the dependence of performance on the value of p_{max}. In all of our experiments, the initial energy at each node is 200 (arbitrary units, consistent with the units of distance). [11] We demonstrate the impact of incorporating residual energy into the cost metric, and compare performance for several values of β.

In our simulations, multicast requests arrive with interarrival times that are exponentially distributed with rate λ/N at each node. For the present case of an infinite number of transceivers and frequencies, performance is essentially independent of λ (except for minor "end effects" related to the death of nodes during sessions); we have used $\lambda = 1$ in all simulations, except for those discussed in Section 6.3. 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 $X = 5000$ 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 MLU with BIP and for different values of β.

(1) *Network lifetime and cumulative efficiency:* 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 [16]) because it is no longer possible to reach all of

[11] 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.

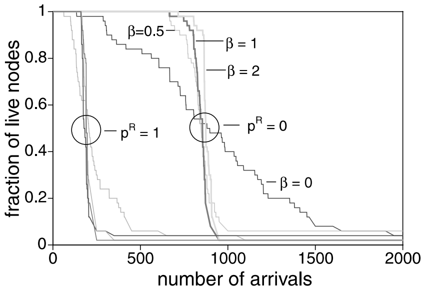


Fig. 1. Evolution of number of live nodes under MIP for 50-node network [$(p^T,\ p^R) = (0, 0)$ and (0.1, 1)].

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. Instead, we examine the time evolution of the number of live nodes.

Fig. 1 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 for the present case of an unlimited number of transceivers and frequencies. 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 5.1, the use of nonzero values of β tends to discourage the use of nodes that have little residual energy. The use of $0.5 \leqslant \beta \leqslant 2$, 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 662, 668, and 716, respectively. [12] The fraction of live nodes decreases to 90% at arrivals 308, 725, 803, and 857 for $\beta = 0$, 0.5, 1, and 2, respectively. Results are qualitatively similar when $(p^T,\ p^R) = (0.1, 1)$; the first node does at arrival 63, 160, 164, and 162 for $\beta = 0$, 0.5, 1, and 2, respectively.

Moreover, for $0.5 \leqslant \beta \leqslant 2$, 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. It can be argued that once about 25% of the nodes have died, the network is no longer providing acceptable performance. 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 seen as an advantage.

Therefore, for $0.5 \leqslant \beta \leqslant 2$ 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 $0.5 \leqslant \beta \leqslant 2$) and on the basis of our observations in additional experiments, we use $\beta = 1$ in most of the examples presented in this paper. No claim of optimality is made.

We note that, in Fig. 1, the fraction of live nodes does not decrease to zero, even though the simulation was continued for 5000 arrivals. This behavior is a consequence of our use of a finite value of p_{max}; thus, 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. (When an infinite value is used for p_{max}, it is typical to be left with a single live node; it does not transmit because none of its potential destination nodes are alive. Of course, if we had included the quiescent energy used by a node when it is not transmitting or receiving, all nodes would eventually die.)

In [16], which studied a sensor network with constant data flow rates, the goal was to maximize system lifetime, which was defined as the time until the first node dies. Thus, performance was considered to be acceptable only as long as all data flows can be supported. It was shown there that the incorporation of residual energy into the link cost metric was able to extend network lifetime. Our results are similar in principle, in that they demonstrate that the incorporation of residual

[12] We have also studied the impact of larger values of β. When $\beta \in \{3, 4, 5, 10, 20\}$ and $(p^T,\ p^R) = (0, 0)$, the first node dies at arrival 481; when $\beta = 50$ the first node dies at arrival 419. The more-rapid death of the first node (as compared with $\beta = 2$) is a consequence of placing too much emphasis in the cost metric on residual energy, rather than on tree power. Also, the total delivered traffic volume (to be discussed shortly) decreases by 8.4% as β is increased from 2 to 20, and by 12.6% as β is increased from 20 to 50.

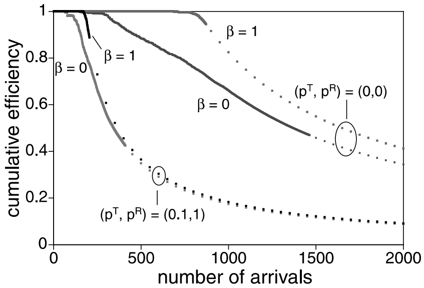


Fig. 2. Evolution of cumulative multicast efficiency under MIP for 50-node network.

energy into the cost metric delays the death of the first node, and keeps the fraction of live nodes at a high level for a longer period of time.

Fig. 2 shows the evolution of cumulative efficiency e for MIP as a function of the number of session arrivals for the same set of parameters. The curves are shown dotted after the point at which 90% of the nodes have died to emphasize the fact that little additional traffic is being supported beyond this point. We see that the use of $\beta = 1$ maintains e at a significantly higher value than that for $\beta = 0$, as long as at least 10% of the nodes are still alive.

(2) *Delivered traffic volume:* We now compare performance on the basis of cumulative volume of delivered traffic, B_X^{total}. Fig. 3(a) shows B_X^{total} as a function of the number of arrivals X for MLU and MIP, and for $\beta = 0$ and 1, for the case in which signal processing power is zero, i.e., $(p^{\text{T}},\ p^{\text{R}}) = (0, 0)$. Recall that B_i (the traffic volume successfully delivered in session i) is proportional to both the number of destinations reached (m_i) and to the duration of the session (d_i). The vertical axis represents the traffic volume in "units" of traffic, where one unit corresponds to the delivery of a message of duration 1 (the mean value of message duration) to a single destination. Again, the curves are shown dotted after the point at which less than 10% of the nodes are alive.

Performance is very similar for all four cases in the early part of the simulation (approximately the first 400 arrivals), when ample energy is available at all nodes. However, the benefits achieved by using $\beta = 1$ with either MIP or MLU are apparent as the simulation progresses past this point. When $\beta = 1$, the delivered traffic volume continues to increase almost linearly, until most of the nodes are dead. When $\beta = 0$, the point at which only 10% of the nodes remain alive is delayed considerably

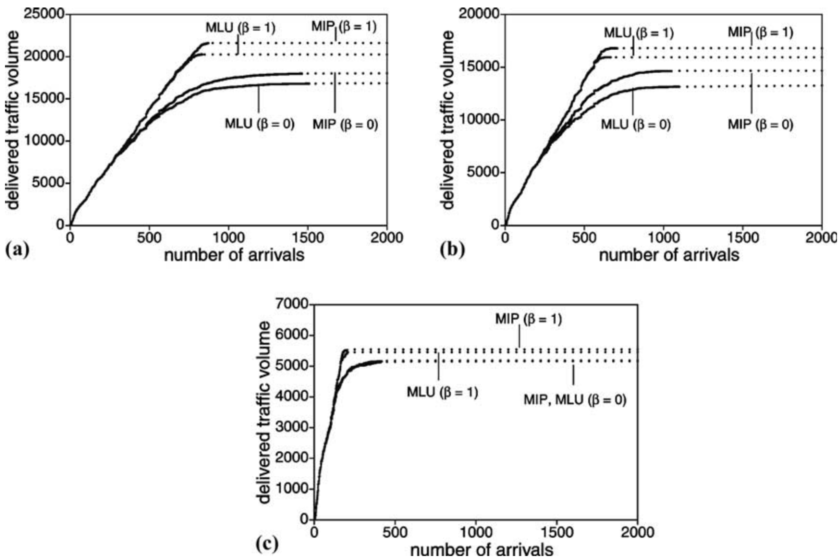


Fig. 3. Evolution of cumulative bit volume under MLU and MIP. (a) $(p^{\text{T}},\ p^{\text{R}}) = (0, 0)$, (b) $(p^{\text{T}},\ p^{\text{R}}) = (0.01, 0.1)$ and (c) $(p^{\text{T}},\ p^{\text{R}}) = (0.1, 1.0)$.

(consistent with the results of Fig. 1), but the total delivered traffic volume is considerably smaller.

For a given value of β, the delivered traffic volume provided by MIP is somewhat greater than that of MLU. We attribute this behavior, at least in part, to the fact that MIP exploits the node-based nature of wireless communications. The superior performance of MIP was pointed out in [1,2] for the limited context of energy-efficient tree construction. The present paper verifies that this advantage is also present for the case of multicast operation over an interval of many randomly generated sessions, in which hard constraints on energy are present. For both MIP and MLU, $\beta = 1$ provides a greater delivered traffic volume than $\beta = 0$ for the present case of zero processing power.

Fig. 3(b) and (c) show similar results for non-zero values of signal processing power, namely $(p^{\mathrm{T}}, p^{\mathrm{R}}) = (0.01, 0.1)$ and $(0.1, 1)$, respectively. The total delivered traffic volume decreases as signal processing power increases. For the "moderate" values of signal-processing power, the RF energy component dominates energy consumption; for the "high" values, the signal-processing component dominates. Again, use of $\beta = 1$ provides better performance than $\beta = 0$. For "high" values of signal-processing power (Fig. 3(c)), there is little difference between the curves for MIP and MLU.

6. Performance results: finite transceiver and frequency resources

The discussions in the previous sections assume the availability of an infinite number of frequencies. However, in realistic situations the number of frequencies is finite, and poses a limitation to overall network throughput. Although, as noted earlier, it is straightforward to incorporate the impact of a finite number of transceivers (i.e., by setting the cost of the node to ∞), the modeling of finite frequency resources is much more complicated. In this section, we first discuss the incorporation of bandwidth limitations into our model, and then discuss performance results in such limited-resource environments.

6.1. 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 the reception of f;
- the use of f by Node m would interfere with ongoing communications at other nodes.

In this section, we discuss the following basic greedy approaches for frequency assignment:

FA1: Assume the availability of an infinite number of frequencies when forming the tree (the approach used in [1,2,5]). 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.

FA2: At each step of the tree-construction, the frequency is chosen along with the transmission power level.

Under FA1, 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. By contrast, 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.

Both FA1 and FA2 are suitable for use with MIP; however, FA2 cannot be used in conjunction with MLU, which requires well-defined link costs that cannot be updated during the construction of a tree.

Let us consider the construction of trees under BIP (which are subsequently pruned to implement MIP) for the case in which the number of frequencies F is finite. Under FA2 the cost of a transmission is set to infinity if no frequency is available. Also, when evaluating the incremental cost of Eq. (5), the multicast advantage applies only when the same frequency can be used by Node i to reach all of its intended neighbors.

Note that FA1 and FA2 actually represent classes of frequency assignment policies. We have used greedy versions, 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, either of these schemes 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.

6.2. Performance results

In this section 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 Section 5. 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, which as in the case for infinite values of T and F, performs better than MLU. We again demonstrate that the incorporation of the residual energy into the cost metric (by using $\beta = 1$) increases significantly the period of time over which the network can support the highest possible multicast efficiency. Our results are based on the use frequency assignment scheme FA1. [13]

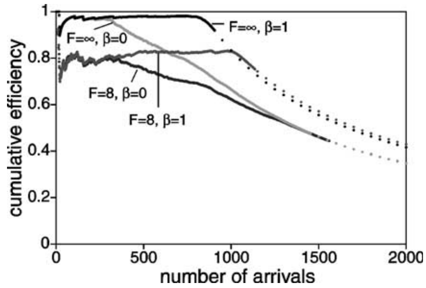


Fig. 4. Evolution of cumulative multicast efficiency under MIP with FA1 [$(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$; $T = 4$; $\lambda = 1$].

(1) *Cumulative multicast efficiency:* Fig. 4 shows the cumulative multicast efficiency for MIP with $T = 4$ and $\lambda = 1$, for the case in which signal-processing power is zero, i.e., $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$. Curves are shown for $F = 8$ and ∞, for both $\beta = 0$ and 1. For $F = \infty$, the primary impact of reducing T from ∞ to 4 is that e cannot be maintained at a value of 1 even when all nodes are alive. We again see that use of $\beta = 1$ maintains e at its maximum possible value for a considerably longer time than the use of $\beta = 0$. For $F = 8$, we observe that the availability of an insufficient number of frequencies results in a considerable lowering of the value of e that can be sustained.

(2) *Delivered traffic volume:* We now consider the delivered traffic volume B_X^{total}. Fig. 5(a) shows the time evolution of B_X^{total} under MIP for several sets of (F,T) pairs for $\beta = 0$, $\lambda = 1$, and $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$. [14] As before, the initial value of energy at each node is $E_i(0) = 200$. Results for nine sets of (F,T) pairs are shown, namely the cases for which $T = 2$, 4, and ∞ and $F = 4$, 8, and ∞. Results for $(F,T) = (4,4)$ are identical to those for $(4,\infty)$; in view of the discussion of Section 2, it is never beneficial to have more transceivers at a node than the total number of frequencies used in the network. The curves for $F = 4$ are significantly lower than the others during the early phase of the simulation (i.e., for approximately the first 1250

[13] Although somewhat better performance is often obtained by using FA2, space does not permit a complete comparison of the two schemes in this paper. We present results only for FA1 because FA2 cannot be used in conjunction with MLU. Results for FA2 are qualitatively similar to those for FA1.

[14] Unlike our earlier figures, we do not indicate here the point at which 90% of the nodes die, simply because doing so might make it difficult to distinguish between adjacent and overlapping curves.

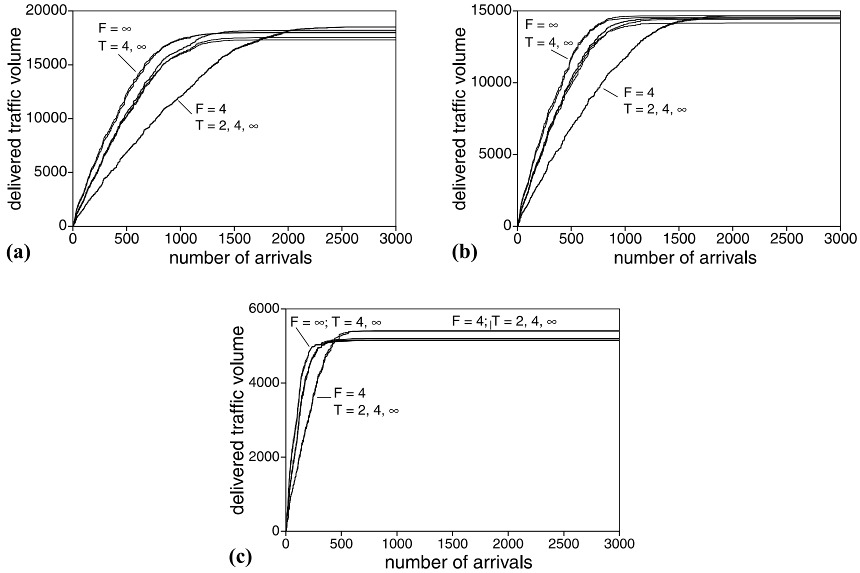


Fig. 5. Evolution of cumulative bit volume under MIP with FA1 for several sets of (F, T) pairs $(\lambda = 1, \beta = 0)$. (a) $(p^T, p^R) = (0, 0)$, (b) $(p^T, p^R) = (0.01, 0.1)$ and (c) $(p^T, p^R) = (0.1, 1.0)$.

arrivals) because the small number of frequencies results in a significant level of blocking. There is virtually no difference among these curves for different values of T, because additional transceivers cannot increase data volume when only four frequencies are available. When an unlimited number of frequencies is available (i.e., $F = \infty$), there is little blocking when $T \geqslant 4$, and $T = 4$ transceivers provides performance that is similar to that for $T = \infty$. The remaining curves are not labeled because there is relatively little difference among them.

Among the sets of (F, T) pairs, the highest final value is achieved for $F = 4$ (the value in this case is nearly independent of the value of T). Despite the strong dependence of B_X^{total} on T and F during the first approximately 1250 arrivals, the highest final value is only 6.5% greater than the lowest final value, which occurs for $(F, T) = (\infty, 2)$. Thus, although there is significant variation in the "rate" of delivered traffic volume as a function of T and F in the early stages of the simulation (when all nodes are still alive), there is relatively little difference in the final value. We postpone our discussion of the final value of B_X^{total} to Section 6.3, where we address the impact of higher arrival rates (and hence higher levels of blocking), which result in a greater dependence of total bit volume on the values of T and F.

Curves for moderate and high values of processing power, shown in Fig. 5(b) and (c) demonstrate qualitatively similar behavior. As signal-processing power increases, nodes die sooner and the final value of B_X^{total} decreases. It is clear from these figures that when $(p^T, p^R) = (0.01, 0.1)$ most of the energy is expended for RF transmission, whereas when $(p^T, p^R) = (0.1, 1.0)$ most of the energy is expended for signal processing.

Fig. 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 appears to be approximately linear 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. 1. Thus, there are two distinct regions of operation. When all (or most)

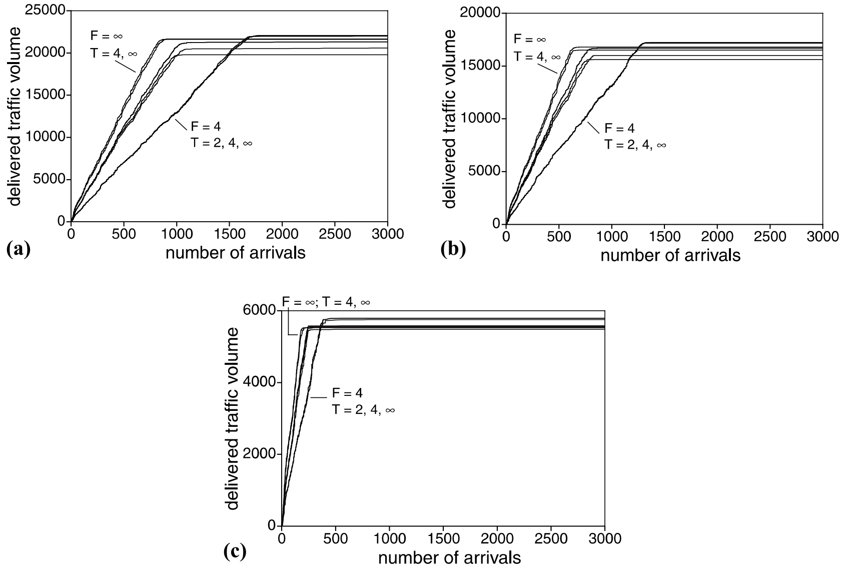


Fig. 6. Evolution of cumulative bit volume under MIP with FA1 for several sets of (F,T) pairs ($\lambda = 1, \beta = 1$). (a) $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$, (b) $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0.01, 0.1)$ and (c) $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0.1, 1.0)$.

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.

The value of multicast efficiency e while all nodes are alive can be inferred from the behavior of these curves in their linear region. Consider the case of $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$, shown in Fig. 6(a). At $X = 500$ arrivals, the value of B_X^{total} is 13,796 for $(F,T) = (\infty,\infty)$ and 7,026 for $(4,\infty)$. Since $e = 1$ when $(F,T) = (\infty,\infty)$, we can therefore infer that $e = 0.509$ when $(F,T) = (4,\infty)$.

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 $\beta = 0$.

6.3. The impact of arrival rate λ

The performance results presented thus far are based on an arrival rate of $\lambda = 1$ (in conjunction with an average session duration of $1/\mu = 1$). As λ increases (when F and/or T are finite), the level of blocking increases. Fig. 7(a) and (b) show the evolution of B_X^{total} for $\beta = 0$ and 1, respectively, for

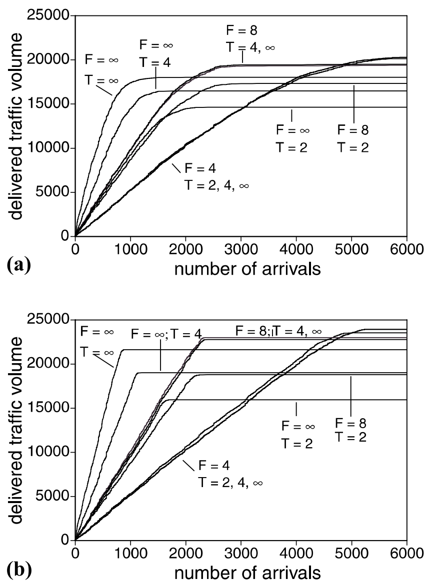


Fig. 7. Evolution of cumulative bit volume under MIP with FA1 for several sets of (F,T) pairs [$\lambda = 5$, $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$]. (a) $\beta = 0$, and (b) $\beta = 1$.

the case of $\lambda = 5$ and $(p^{\mathrm{T}},\ p^{\mathrm{R}}) = (0,0)$. The curves for $T = F = \infty$ are identical to those for $\lambda = 1$; as before, there is no blocking until destinations start to die. As F and/or T decrease, blocking increases, resulting in a decreased rate of delivered traffic volume. Blocking is considerably greater for $\lambda = 5$ than for $\lambda = 1$, resulting in a significantly lower rate of traffic delivery for a given (F, T) pair (again referring to the early part of the simulation). Overall, the rate of traffic delivery is more sensitive to the values of F and T when the value of λ is high. Also, the nodes stay alive for a greater number of arrivals (although not a greater time interval) when $\lambda = 5$ because a greater number of destinations are blocked (thus no energy is expended to reach them).

The sensitivity of the final value of B_X^{total} to F and T is also greater for $\lambda = 5$ than for $\lambda = 1$. For $\beta = 0$, the largest value (which is obtained for $F = 4$, $T = 4$ or ∞) is 38.7% greater than the smallest; for $\beta = 1$, the largest value (which is obtained for the same values of F and T) is 50% greater than the smallest. It is also interesting that the largest values obtained for $\lambda = 5$ are greater than those obtained for $\lambda = 1$, despite the increased level of blocking that is associated with larger arrival rates. This behavior is not difficult to explain. When F is small, it is difficult to reach destinations that are more than one or two hops away. Thus, those that are in fact reached tend to be low-cost destinations, thus resulting in a higher final value of B_X^{total}. (This is the reason that the highest values of B_X^{total} were observed for $F = 4$ in Figs. 5 and 6, as well.) When T is small, but F is moderate to large, the unavailability of transceivers not only results in blocked destinations, but also results in the inability to use the best relay nodes, thus resulting in a lower final value of B_X^{total}.

When $\lambda = 5$ and $F = 4$, the value of multicast efficiency e in the linear region of the curve is 0.189, which is likely too low for practical applications. Nevertheless, our observations on operation for this set of parameters provides insight into the dependence of performance on arrival rates and system resources.

6.4. The impact of the value of $p_{\max}$

The value of $p_{\max}$ determines connectivity. It is difficult to determine a priori the value of $p_{\max}$ that would provide optimal performance, based on one of the criteria we have studied (e.g., total bit volume). For example, setting $p_{\max} = \infty$ provides the possibility of direct connectivity between every pair of nodes, as well as the possibility of reaching several neighbors with a single transmission. However, use of $p_{\max} = \infty$ may also permit the use of excessively long links that result in rapid depletion of energy (especially since we have assumed that trees are created to reach all reachable destinations, regardless of the cost required to do so).

Table 1 shows the impact of $p_{\max}$ and β on performance. The upper entry in each cell shows the total delivered traffic volume for a simulation of 5000 arrivals; the lower entry shows the fraction of nodes that remain alive at the end of the simulation. Results are provided for $T = \infty$, $F = \infty$, $\beta = 0$, 0.5, 1, and 2, and for $p_{\max} = 10$, 25, and ∞. We note first that use of $0.5 \leqslant \beta \leqslant 2$ provides a considerable increase in delivered traffic volume (as compared with $\beta = 0$) for all values of $p_{\max}$ (which is consistent with our results for $p_{\max} = 25$ presented earlier). For $p_{\max} \geqslant 25$, the highest value of total bit volume is achieved for $\beta = 2$; for $p_{\max} = 10$, it is achieved for $\beta = 0.5$.

Next we examine the impact of the value of $p_{\max}$ on total delivered traffic volume. We observe that for $\beta = 0$ and 0.5, use of $p_{\max} = 10$ provides better results, whereas for $\beta = 1$ or 2 (which are preferable because they provide increased total traffic volume), it is better to use $p_{\max} = 25$ or ∞. We have already noted that use of $\beta = 1$ results in a good degree of load balancing, which keeps the vast majority of nodes alive considerably longer than

Table 1
Total delivered traffic volume and fraction of live nodes after a simulation of 5000 arrivals ($\lambda = 1$)

	$p_{\max} = 10$	$p_{\max} = 25$	$p_{\max} = \infty$
$\beta = 0$	18630 0.16	18022 0.04	18023 0.02
$\beta = 0.5$	21813 0.22	21374 0.06	21468 0.06
$\beta = 1$	21338 0.14	21608 0.02	21825 0.02
$\beta = 2$	21723 0.08	22206 0.02	22043 0.02

use of $\beta = 0$. This load balancing makes it "safe" to use larger values of p_{max} because high levels of transmitter power are used only when they are beneficial to overall performance.

Summarizing the above, we conclude that the use of $\beta \geqslant 0.5$ has a much greater impact on performance than the value of p_{max}.

It is also interesting to look at the fraction of nodes that remain alive at the end of the simulation, which is shown as the lower entry in each of the cells. When $p_{max} = 10$, a significant fraction of nodes remain alive. Typically, these nodes have few (if any) live neighbors (i.e., live nodes within a range of 3.16). This fraction is considerably smaller for $p_{max} \geqslant 25$, in which case at most three of the 50 nodes remain alive. When $p_{max} = \infty$, if the simulation is run "forever," one node will typically remain alive (corresponding to a fraction of 0.02 for our example with 50 nodes) because it has no potential destination nodes. Of course, if the quiescent energy required to maintain a node in the "on" state were incorporated into the model (which would be straightforward to do), all nodes would eventually die regardless of the value of p_{max}.

7. 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. When the energy available for transmission at each node is limited, even the definition of performance measures becomes a difficult question to address, and the relative performance of multicasting schemes depends strongly on the criteria used to evaluate them.

We have shown that the incorporation of residual energy into local cost metrics used for tree construction, which results in spreading the burden of energy use among more of the nodes, has a considerable impact on network performance. Most significantly, we have shown that multicast efficiency can be maintained at high levels significantly longer when residual energy is taken into account, and that the overall volume of delivered traffic can be increased.

We have demonstrated that our MIP algorithm (which exploits the node-based nature of wireless communications) performs better than MLU (which is an adaptation of link-based unicast routing) in the context of limited-energy networks. Our conclusions reaffirm the results obtained in our earlier work on energy-efficient multicasting.

References

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

[2] J.E. Wieselthier, G.D. Nguyen, 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, A. Ephremides, Algorithms for bandwidth-limited energy-efficient wireless broadcasting and multicasting, Proceedings of MILCOM'2000, Paper No. 15.1, 2000.

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

[5] J.E. Wieselthier, G.D. Nguyen, A. Ephremides, Algorithms for energy-efficient multicasting in static ad hoc wireless networks, Mobile Networks and Applications (MONET), 6–3, 2001, pp. 251–263.

[6] M. Nagy, S. Singh, Multicast scheduling algorithms in mobile networks, Cluster Computing 1–2 (1998) 177–185.

[7] C.-C. Chiang, M. Gerla, L. Zhang, Forwarding group multicast protocol (FGMP) for multihop mobile wireless networks, Cluster Computing 1–2 (1998) 187–196.

[8] S.-J. Lee, W. Su, J. Hsu, M. Gerla, R. Bagrodia, A performance comparison study of ad hoc wireless multicast protocols, Proceedings of IEEE INFOCOM 2000, 2000, pp. 565–574.

[9] C. Diot, W. Dabbous, J. Crowcroft, Multipoint communication: survey of protocols, functions and mechanisms, IEEE Journal on Selected Areas in Communications 15-3 (1997) 277–290.

[10] M.H. Ammar, G.C. Polyzos, S.K. Tripathi, editors, Special issue on network support for multipoint communication, IEEE Journal on Selected Areas in Communications 15-3 (1997).

[11] L. Zhang, S. Deering, D. Estrin, S. Shenker, D. Zappala, RSVP: a new resource ReSerVation Protocol, IEEE Network Magazine (1993) 8–18.

[12] M. Zorzi, R.R. Rao, Error control and energy consumption in communications for nomadic computing, IEEE Transactions on Computers 46-3 (1997) 279–289.

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

[14] W.-T. Chen, N.-F. Huang, The strongly connecting problem on multihop packet radio networks, IEEE Transactions on Communications 37-3 (1989) 293–295.

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

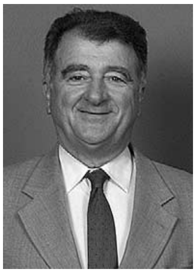
[16] J.-H. Chang, L. Tassiulas, Energy conserving routing in wireless adhoc networks, Proceedings of IEEE INFOCOM 2000, 2000, pp. 22-31.

[17] D. Bertsekas, R. Gallager, Data Networks, Prentice Hall, Englewood Cliffs, 1992.

Jeffrey E. Wieselthier was born in Brooklyn, NY, on March 14, 1949. He received the S.B. degree from the Massachusetts Institute of Technology, Cambridge, in 1969, the M.S. degree from the Johns Hopkins University, Baltimore, MD, in 1971, and the Ph.D. degree from the University of Maryland, College Park, in 1979, all in Electrical Engineering. He was employed at the Naval Surface Warfare Center, White Oak, Silver Spring, MD, from 1969 to 1979. Since 1979 he has been with the Information Technology Division of the Naval Research Laboratory, Washington, DC. He was Technical Program Co-Chair of the Third IEEE Symposium on Computers and Communications in Athens, Greece, in 1998 and Treasurer of the 1991 IEEE International Symposium on Information Theory in Budapest, Hungary. He won the IEEE Fred W. Ellersick Award for the best unclassified paper at MILCOM 2000. He has studied a variety of communication networking problems, including multiple access and routing in spread-spectrum networks and the use of neural networks and other approaches for network performance evaluation, optimization and control. His current interests include wireless communication networks, with an emphasis on issues relating to energy-efficient operation.

Dr. Wieselthier is a member of Eta Kappa Nu and Sigma Xi, and is a Senior Member of IEEE.

Gam D. Nguyen received the Ph.D. in Electrical Engineering from the University of Maryland, College Park, MD in 1990. He has been at the Naval Research Laboratory, Washington, DC, since 1991. He won the IEEE Fred W. Ellersick Award for the best unclassified paper at MILCOM 2000. His research interests include communication systems, computer communication networks, and information processing.



Anthony Ephremides received his B.S. degree from the National Technical University of Athens (1967), and M.S. (1969) and Ph.D. (1971) degrees from Princeton University, all in Electrical Engineering. He has been at the University of Maryland since 1971, and currently holds a joint appointment as Professor in the Electrical and Computer Engineering Department and the Institute for Systems Research (ISR). He is co-founder of the NASA Center for Commercial Development of Space on Hybrid and Satellite Communications Networks established in 1991 at Maryland as an off-shoot of the ISR.

He was a Visiting Professor in 1978 at the National Technical University in Athens, Greece, and in 1979 at the EECS Department of the University of California, Berkeley, and at INRIA, France. During 1985–1986 he was on leave at MIT and ETH in Zurich, Switzerland. He was the General Chairman of the 1986 IEEE Conference on Decision and Control in Athens, Greece and of the 1991 IEEE International Symposium on Information Theory in Budapest, Hungary. He was the Technical Program Co-Chair of the IEEE INFOCOM in New York City in 1999 and of the IEEE International Symposium on Information theory in Sorrento, Italy in 2000. He has also been the Director of the Fairchild Scholars and Doctoral Fellows Program, an academic and research partnership program in Satellite Communications between Fairchild Industries and the University of Maryland. He has received several awards, including the IEEE Donald E. Fink Prize Paper Award (1992) and the Sigmobile Award of the ACM for contributions to wireless communications (1997). He has been the President of the Information Theory Society of the IEEE (1987), and has served on the Board of the IEEE (1989 and 1990). He is a Fellow of IEEE.

Dr. Ephremides' interests are in the areas of communication theory, communication systems and networks, queueing systems, signal processing, and satellite communications.