

MIT Open Access Articles

The multicast capacity region of large wireless networks

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Niesen, U., P. Gupta, and D. Shah. "The Multicast Capacity Region of Large Wireless Networks." INFOCOM 2009, IEEE. 2009. 1881-1889. © 2010 Institute of Electrical and Electronics Engineers.

As Published: <http://dx.doi.org/10.1109/INFCOM.2009.5062109>

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: <http://hdl.handle.net/1721.1/58799>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



The Multicast Capacity Region of Large Wireless Networks

Urs Niesen
Massachusetts Institute of Technology
Email: uniesen@mit.edu

Piyush Gupta
Bell Labs, Alcatel-Lucent
Email: pgupta@research.bell-labs.com

Devavrat Shah
Massachusetts Institute of Technology
Email: devavrat@mit.edu

Abstract—We study the problem of determining the multicast capacity region of a wireless network of n nodes randomly located in an extended area and communicating with each other over Gaussian fading channels. We obtain an explicit information-theoretic characterization of the scaling of the multicast capacity region for n nodes in terms of $2n$ weighted cuts. These cuts only depend on the geometry of the locations of the source nodes and their destination nodes and the traffic demands between them, and thus can be readily evaluated. The results are constructive and provide a two-layer architecture for achieving nearly the entire multicast capacity region in the scaling sense: The top layer routes traffic from each of the source nodes to its set of destination nodes, and the bottom layer physically distributes/concentrates traffic among appropriate nodes through one of the two cooperative communication schemes – hierarchical relaying and multi-hopping – depending on the wireless-channel characteristics.

I. INTRODUCTION

In recent years, large wireless networks have become the architecture of choice in many emerging scenarios, such as mesh networks for providing infrastructure in metro areas, communication infrastructure under extreme conditions in military applications, cost-effective sensor networks for monitoring and surveillance, peer-to-peer networks between handheld devices in a social setup, etc.

Motivated by such applications, numerous protocols have been proposed to perform multicast in wireless networks (e.g., [1]–[4]). An important issue is to determine the performance of these protocols, or more generally, the best performance that can be achieved by any protocol. Some recent progress has been made for quantifying the performance in the *homogeneous* traffic case (discussed more in detail below). However, in many of the aforementioned applications, different users may wish to send varying number of messages to different groups of users. For example, a local news channel may wish to broadcast the news over the metro network to all users almost all the time, but an individual user may wish to send messages (e.g., voice, email, or SMS) to only a small group of his/her friends occasionally. The key issue of quantifying the performance of wireless networks under such *heterogeneous* demands of various users remains wide open.

Thus, in this paper, we consider the problem of determining the complete multicast capacity region of wireless

networks. For a n node wireless network, this corresponds to characterizing a $n \times 2^n$ -dimensional region. As the main result of this paper, we obtain an information-theoretic scaling characterization of this region under random node placement.

A. Related Work

Before we describe our contributions in detail, we summarize related work. Given the popularity of this topic, the complete list is very large, and we will only sample a small subset of them here.

The result by Gupta and Kumar [5] started the inquiry of the scaling of the (unicast) capacity region for large wireless networks. Under random node placement and a protocol interference model, they showed that the maximal uniformly achievable per-node rate essentially scales as $\Theta(1/\sqrt{n})$ for a network of n nodes when traffic demands are generated by n random source-destination pairs. More recently, a complete information-theoretic (i.e., without making the protocol interference assumption) analogue of these results were obtained by Özgür, Lévêque, and Tse [6]. This information-theoretic characterization was further generalized by the present authors in a recent result [7] with nodes of the wireless network placed arbitrarily (as opposed to randomly) with a minimum-distance requirement. An excellent survey by Xue and Kumar [8] summarizes the results for this kind of traffic demand, as well as results for the transport capacity, up to early 2005.

It can be shown that analyzing traffic demands of this type (i.e., n random source-destination pairs, each with the same rate requirement) yields the scaling of the unicast capacity region only along one particular direction [9]. In contrast, the complete *unicast capacity region* for a network of n nodes is a n^2 -dimensional region. Under the protocol model and arbitrary node placement, its scaling characterization follows from certain results on multi-commodity flows in wireline networks in terms of appropriately weighted cuts (cf. [9] using [10]) – the number of cuts that need to be considered can be up to 2^n . An alternate approach based on transport-capacity bounds is discussed in [11]. In a very recent result [12], the present authors have obtained a complete information-theoretic scaling characterization of this n^2 -dimensional unicast capacity region under random node placement. Somewhat surprisingly, as is shown in [12], the scaling of the n^2 -dimensional unicast capacity region is obtained in terms of only $2n$ weighted cuts.

The work of U. Niesen and D. Shah was supported in parts by DARPA grant (ITMANET) 18870740-37362-C; the work of P. Gupta was supported in part by NSF Grants CCR-0325673 and CNS-0519535.

The *multicast capacity region* of a network of n nodes is a $n \times 2^n$ -dimensional region. Again, under the protocol model and arbitrary node placement, characterizing this region is equivalent to characterizing the multicast capacity region of a wireline network. However, unlike in the unicast case, no scaling characterization of the multicast region is known even for wireline networks! This has led various authors to consider restricted setups to obtain an explicit characterization of such scaling (see [13]–[18], among others). For example, in [13], Li, Tang, and Frieder obtained a scaling characterization under the protocol model and random node placement for multicast traffic when each node chooses a certain number of its destinations uniformly at random. Independently, in [14], Shakkottai, Liu, and Srikant considered a similar setup and also obtained the precise scaling when sources and their multicast destinations are chosen at random. Both of these results are non information-theoretic (in that they assume a specific protocol model). Furthermore, they correspond to the scaling characterization of the $n \times 2^n$ -dimensional multicast capacity region only along one particular dimension.

In the context of wireline networks with only one source and an arbitrary subset of multicast destinations, the min-cut among the various single source-destination cuts characterizes the precise achievable multicast rate. For example, the popular network coding schemes achieve this rate (cf. [19]–[21]). In a recent result, Koetter [22] has established the equivalence of the multicast capacity region of a wireline network with noisy links (modeled as independent discrete memoryless channels) and the capacity region of a wireline network with noiseless links having the same capacity as in the noisy case. However, this equivalence relation still does not provide any hints for a characterization of the multicast capacity region even for wireline networks.

B. Our Contributions

The main contribution of this paper is an explicit and information-theoretic scaling characterization of the $n \times 2^n$ -dimensional multicast capacity region of a wireless network of n nodes placed uniformly at random in a square of area n . Our characterization is in terms of $2n$ weighted cuts, which only depend on the geometry of the locations of the source nodes and their destination nodes and the traffic demands between them, and thus can be easily evaluated.

To establish our results, we develop two novel equivalence (up to scaling) relations of the multicast capacity region. The first relation is conceptual, stating that the scaling of the multicast capacity region is equivalent to the scaling of the intersection of a set of *induced* unicast capacity regions. For a precise statement, see Theorem 3. This can be thought of as a generalization of the min-cut characterization of a single-source multicast in wireline networks (cf. I-A). The second relation is algorithmic, suggesting an optimal (in the scaling sense) two-layer architecture for multicast traffic. The top or routing layer sees the network as a rooted tree over which it routes messages from each of the source nodes to its set of destination nodes. The bottom or physical layer provides this

tree abstraction by performing cooperative communication at various scales through one of the two schemes – hierarchical relaying and multi-hop communication – depending on the power path loss exponent of the wireless channel. See Theorem 1 and the communication schemes in Section III for details. These results also implicitly establish that a separation based approach, where the routing layer works essentially independently of the physical layer, can achieve nearly the entire multicast capacity region in the scaling sense.

C. Organization

The remainder of this paper is organized as follows. In Section II, we formally present our main results. Section III contains a description of the communication schemes used to prove achievability. In Section IV, we present the proof of the main result. Finally, Section V contains concluding remarks.

II. MAIN RESULT

A. Model

Consider the square $A(n) \triangleq [0, \sqrt{n}]^2$ of area n . The n wireless nodes $V(n) \subset A(n)$ (with $|V(n)| = n$) are placed independently and uniformly at random on $A(n)$. These nodes share a common wireless medium modeled as Gaussian fading channels. Specifically, let $\{X_u[t]\}_{u,t}$ be the (sampled) signals sent by the nodes in $V(n)$. Then the (sampled) received signal at node v and time t is

$$Y_v[t] = \sum_{u \in V(n) \setminus \{v\}} H_{u,v}[t] X_u[t] + Z_v[t] \quad (1)$$

for all $v \in V(n), t \in \mathbb{N}$. Here $\{Z_v[t]\}_{v,t}$ are i.i.d. circularly symmetric complex Gaussian random variables with mean 0 and variance 1, and

$$H_{u,v}[t] = r_{u,v}^{-\alpha/2} \exp(\sqrt{-1} \Theta_{u,v}[t]),$$

where α is the *path loss exponent* with $\alpha > 2$, $r_{u,v}$ is the Euclidean distance between u and v , and the $\{\Theta_{u,v}[t]\}_{u,v}$ are assumed to be i.i.d. with uniform distribution on $[0, 2\pi)$. We assume either *fast fading* or *slow fading*. That is, $\{\Theta_{u,v}[t]\}_t$ is either stationary and ergodic as a function of t (fast fading), or $\{\Theta_{u,v}[t]\}_t$ is constant as a function of t (slow fading). In either case, we assume full channel state information (CSI) is available at all nodes¹, i.e., each node knows all $\{H_{u,v}[t]\}_{u,v}$ at time t . We also impose an average power constraint of P on the signal $\{X_u[t]\}_t$ for every node $u \in V(n)$.

B. Problem Statement

A *multicast traffic matrix* $\tilde{\lambda} = [\tilde{\lambda}_{u,W}] \in \mathbb{R}_+^{n \times 2^n}$ represents that source $u \in V(n)$ wishes to send common data to multicast set $W \subset V(n)$ at rate $\tilde{\lambda}_{u,W} \geq 0$. The *multicast capacity region* of a wireless network, denoted by $\tilde{\Lambda}(n) \subset \mathbb{R}_+^{n \times 2^n}$, is

¹We make the full CSI assumption in all the converse results in this paper. Achievability can be shown to hold under weaker assumptions on the availability of CSI. In particular, for $\alpha \geq 3$, no CSI is necessary, and for $\alpha \in (2, 3)$, a 2 bit quantization of the channel state $\{\Theta_{u,v}\}_{u,v}$ available at all nodes is sufficient.

the collection of multicast traffic matrices that are achievable. Define for any multicast traffic matrix $\tilde{\lambda}$

$$\rho_{\tilde{\lambda}}^*(n) = \sup\{b \geq 0 : b\tilde{\lambda} \in \tilde{\Lambda}(n)\}.$$

Since $\tilde{\Lambda}(n)$ is a convex region, it can be completely characterized by the knowledge of $\rho_{\tilde{\lambda}}^*(n)$ for all traffic matrices $\tilde{\lambda} \in \mathbb{R}_+^{n \times 2^n}$. Our goal in this paper is to obtain the precise scaling (i.e., asymptotic) characterization of $\rho_{\tilde{\lambda}}^*(n)$ for any $\tilde{\lambda} \in \mathbb{R}_+^{n \times 2^n}$.

C. Main Theorem

Before stating the main result, we introduce some notation. Partition $A(n)$ into $L(n)$ different square-grids with

$$L(n) \triangleq \frac{1}{2} \log(n)(1 - \log^{-1/2}(n)).$$

The ℓ -th square-grid, $0 \leq \ell \leq L(n)$, divides $A(n)$ into 4^ℓ squares, each of sidelength $2^{-\ell}\sqrt{n}$, denoted by $\{A_{\ell,i}(n)\}_{i=1}^{4^\ell}$. Let $V_{\ell,i}(n) \subset V(n)$ be the nodes in $A_{\ell,i}(n)$ (see Figure 1). The choice of $L(n)$ is such that with high probability $|V_{L(n),i}| \rightarrow \infty$ as $n \rightarrow \infty$, but not too fast (i.e., $|V_{L(n),i}| = n^{o(1)}$).

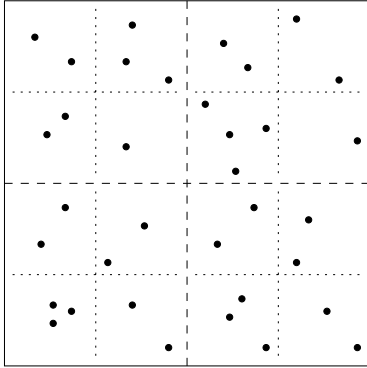


Fig. 1. Square-grids with $0 \leq \ell \leq 2$, i.e., with $L(n) = 2$. The grid at level $\ell = 0$ is the area $A(n)$ itself. The grid at level $\ell = 1$ is indicated by the dashed lines. The grid at level $\ell = 2$ by the dashed and the dotted lines. Assume for the sake of example that the subsquares are numbered from left to right and then from bottom to top (the precise order of numbering is immaterial). Then $V_{0,1}(n)$ are all the nodes $V(n)$, $V_{1,1}(n)$ are the nine nodes in the lower left corner (separated by dashed lines), and $V_{2,1}(n)$ are the three nodes in the lower left corner (separated by dotted lines).

Define the function $g_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$,

$$g_\alpha(r) \triangleq \begin{cases} r^{2-\min\{3,\alpha\}/2} & \text{if } r \geq 1, \\ 1 & \text{else.} \end{cases}$$

As we shall see, $g_\alpha(4^{-\ell}n)$ corresponds approximately to the total rate crossing the boundary of $V_{\ell,i}(n)$ that is achievable. Finally, for a given multicast traffic matrix $\tilde{\lambda} \in \mathbb{R}_+^{n \times 2^n}$, set

$$\phi_{\tilde{\lambda}}^*(n) \triangleq \min_{\ell \in \{1, \dots, L(n)\} \cup \{\log(n)\}} \min_{i \in \{1, \dots, 4^\ell\}} \frac{g_\alpha(4^{-\ell}n)}{D_{\tilde{\lambda}}(V_{\ell,i}(n))}, \quad (2)$$

where, for any $U \subset V(n)$,

$$D_{\tilde{\lambda}}(U) \triangleq \sum_{u \in U} \sum_{\substack{W \subset V(n): \\ W \cap U^c \neq \emptyset}} \tilde{\lambda}_{u,W} + \sum_{u' \in U^c} \sum_{\substack{W' \subset V(n): \\ W' \cap U \neq \emptyset}} \tilde{\lambda}_{u',W'}, \quad (3)$$

is the total demand across the boundary of U .

We now state the main result of this paper characterizing the scaling of the complete multicast capacity region.

Theorem 1. *Under either fast or slow fading, for any $\alpha > 2$ and any $\varepsilon > 0$, there exists $b(n) = O(n^\varepsilon)$ such that with probability $1 - o(1)$ as $n \rightarrow \infty$ for any multicast traffic matrix $\tilde{\lambda} \in \mathbb{R}_+^{n \times 2^n}$*

$$b(n)^{-1} \phi_{\tilde{\lambda}}^*(n) \leq \rho_{\tilde{\lambda}}^*(n) \leq b(n) \phi_{\tilde{\lambda}}^*(n).$$

Theorem 1 implies that the quantity $\phi_{\tilde{\lambda}}^*(n)$ determines the scaling of $\rho_{\tilde{\lambda}}^*(n)$. By definition, this can be computed by evaluating weighted cuts as in (2). For $1 \leq \ell \leq L(n)$, we need to evaluate at most n cuts (for n large enough). For $\ell = \log(n)$, we only need to take those cuts into account that are non-empty. Since there are at most n such cuts with $\ell = \log(n)$, the number of cuts in (2) that need to be evaluated is at most $2n$. By (3), evaluating each such cut takes at most $|\{(u, W) : \tilde{\lambda}_{u,W} > 0\}|$ operations. Thus computing (2) takes at most $\Theta(n)$ times more operations than required to just read the problem parameters.

D. Some Example Scenarios

Example 1. (Broadcast from one source)

Assume we have only one source (say $u_0 \in V(n)$) that wants to broadcast the same message to all other nodes. In other words, we consider the multicast traffic matrix

$$\tilde{\lambda}_{u,W} = \begin{cases} \rho(n) & \text{if } u = u_0 \text{ and } W = V(n), \\ 0 & \text{else,} \end{cases}$$

for some $\rho(n) > 0$. Applying Theorem 1 yields that $\rho^*(n)$, the largest achievable $\rho(n)$, satisfies

$$\Omega(n^{-\varepsilon}) = \rho^*(n) = O(n^\varepsilon)$$

with probability $1 - o(1)$ as $n \rightarrow \infty$. This result says that the source can broadcast its information at essentially a constant rate independent of n to all nodes in the network. Hence, transmitting information from one source to its destination is (at least asymptotically) as difficult as broadcasting information from one source to all nodes in the network.

Two comments are in order here. First, this example shows that in large wireless networks, network coding can provide at most a factor $O(n^\varepsilon)$ gain for the multicasting problem with one source. Second, note that this result is independent of α . This is in contrast to the unicast situation with random source-destination pairing, where the resulting capacity depends strongly on the value of α (see [6], [7]). \diamond

Example 2. (Broadcast from many sources)

Consider a scenario with n^β sources, $\{u_1, \dots, u_{n^\beta}\}$, for some $0 \leq \beta \leq 1$, each broadcasting an independent message to all other nodes at the same rate. In other words, we have a multicast traffic matrix of the form

$$\tilde{\lambda}_{u,W} = \begin{cases} \rho(n) & \text{if } u = u_i \text{ for some } i \text{ and } W = V(n), \\ 0 & \text{else,} \end{cases}$$

for some $\rho(n) > 0$. Applying Theorem 1 yields that $\rho^*(n)$, the largest achievable $\rho(n)$, satisfies

$$\Omega(n^{-\beta-\varepsilon}) = \rho^*(n) = O(n^{-\beta+\varepsilon}),$$

with probability $1 - o(1)$ as $n \rightarrow \infty$. \diamond

Example 3. (Multicast from many sources)

Consider n^{β_1} sources, each generating independent multicast traffic for a set of n^{β_2} destinations. All these sources and their destinations are chosen independently and uniformly at random from $V(n)$. Let $\{u_1, \dots, u_{n^{\beta_1}}\}$ be the source nodes and $\{W_1, \dots, W_{n^{\beta_2}}\}$, with $|W_i| = n^{\beta_2}$, be the corresponding destination nodes. Then the multicast traffic matrix $\tilde{\lambda}$ is of the form

$$\tilde{\lambda}_{u,W} = \begin{cases} \rho(n) & \text{if } u = u_i, W = W_i \text{ for some } i, \\ 0 & \text{else,} \end{cases} \quad (4)$$

for some $\rho(n) > 0$. Applying Theorem 1 yields (after a somewhat lengthy computation) that $\rho^*(n)$ satisfies

$$\begin{aligned} & \Omega\left(\min\{n^{-\varepsilon}, n^{(1-\beta_2)\tilde{\alpha}-\beta_1-\varepsilon}\}\right) \\ &= \rho^*(n) \\ &= O\left(\min\{n^{\varepsilon}, n^{(1-\beta_2)\tilde{\alpha}-\beta_1+\varepsilon}\}\right) \end{aligned} \quad (5)$$

with probability $1 - o(1)$ as $n \rightarrow \infty$, and where

$$\tilde{\alpha} \triangleq 2 - \min\{3, \alpha\}/2.$$

Example 1 is a special case of the setup here with $\beta_1 = 0$ and $\beta_2 = 1$. Plugging this into (5), we recover the result from Example 1. Similarly, Example 2 is a special case of the setup here with $\beta_1 = \beta$ and $\beta_2 = 1$, which combined with (5) yields the result from Example 2. \diamond

Example 4. (Localized multicast from many sources)

Consider the setup of Example 3 above, except now each source picks its n^{β_2} destinations uniformly at random from among nodes within a distance of $n^{\frac{\beta_3}{2}}$, where $\beta_3 > \beta_2$. In other words, each source node does localized multicast. Again, let $\{u_1, \dots, u_{n^{\beta_1}}\}$ denote the source nodes and $\{W_1, \dots, W_{n^{\beta_2}}\}$, with $|W_i| = n^{\beta_2}$, denote the corresponding destination nodes, where now $r_{u_i,v} \leq n^{\frac{\beta_3}{2}}$, for each $v \in W_i$. The multicast traffic matrix $\tilde{\lambda}$ is of the form (4). Then, an application of Theorem 1 shows that $\rho^*(n)$ satisfies

$$\begin{aligned} & \Omega\left(\min\{n^{-\varepsilon}, n^{(\beta_3-\beta_2)\tilde{\alpha}-\max\{0,\beta_1+\beta_3-1\}-\varepsilon}\}\right) \\ &= \rho^*(n) \\ &= O\left(\min\{n^{\varepsilon}, n^{(\beta_3-\beta_2)\tilde{\alpha}-\max\{0,\beta_1+\beta_3-1\}+\varepsilon}\}\right) \end{aligned}$$

with probability $1 - o(1)$ as $n \rightarrow \infty$. Note that setting $\beta_3 = 1$, i.e., each source picks its destinations uniformly over the entire region $A(n)$, yields the same scaling of $\rho^*(n)$ as in Example 3. \diamond

Example 5. (Multiple classes of localized multicast)

All the examples so far have homogeneous traffic requirements from different sources. We next relax that. There are

K classes of multicast sources, for some fixed K . Each source node in class k generates multicast traffic for $n^{\beta_{2,k}}$ destinations, which are randomly and uniformly chosen among all nodes within a distance of $n^{\frac{\beta_{3,k}}{2}}$ from it, where $\beta_{3,k} > \beta_{2,k}$. There are $n^{\beta_{1,k}}$ sources of class k , each chosen independently and uniformly over $A(n)$, each generating multicast traffic at rate $\rho_k(n)$. Then, Theorem 1 yields that $\{\rho_k^*(n)\}_{k=1}^K$, the largest achievable $\{\rho_k(n)\}_{k=1}^K$, satisfy

$$\begin{aligned} & \Omega\left(\min\{n^{-\varepsilon}, n^{(\beta_{3,k}-\beta_{2,k})\tilde{\alpha}-\max\{0,\beta_{1,k}+\beta_{3,k}-1\}-\varepsilon}\}\right) \\ &= \rho_k^*(n) \\ &= O\left(\min\{n^{\varepsilon}, n^{(\beta_{3,k}-\beta_{2,k})\tilde{\alpha}-\max\{0,\beta_{1,k}+\beta_{3,k}-1\}+\varepsilon}\}\right) \end{aligned}$$

for all $k \in \{1, \dots, K\}$ with probability $1 - o(1)$ as $n \rightarrow \infty$. In other words, time sharing between all the K classes is order optimal. \diamond

III. COMMUNICATION SCHEMES

In this section, we provide a high-level description of the communication schemes used to prove achievability in Theorem 1 (see Section III-B below). We start off in Section III-A by recalling results for unicast traffic from prior work that will be used as building blocks in the following.

A. Unicast Traffic

In unicast traffic, there is distinct data that needs to be communicated between each source-destination pair. A *unicast traffic matrix* $\lambda = [\lambda_{u,v}] \in \mathbb{R}_+^{n \times n}$ represents that source $u \in V(n)$ wishes to send (distinct) data to destination $v \in V(n)$ at rate $\lambda_{u,v} \geq 0$.

A unicast traffic matrix λ is called a *permutation traffic* if for every $u \in V(n)$ there exists unique $v, \tilde{v} \in V(n) \setminus \{u\}$ such that $\lambda_{u,v} > 0$ and $\lambda_{\tilde{v},u} > 0$. Here we discuss (asymptotically) optimal communication schemes for such permutation traffic with uniform rate for all source-destination pairs. We shall use these communication schemes as building blocks in the following.

The type of optimal communication scheme depends drastically on the path loss exponent α . For $\alpha \in (2, 3]$, i.e., the path loss exponent is small, cooperative communication on a global scale is necessary to achieve optimal performance. For $\alpha > 3$, i.e., the path loss exponent is large, only local communication between neighboring nodes is necessary, and traffic is routed in a multi-hop fashion from the source to the destination. We will refer to the optimal scheme for $\alpha \in (2, 3]$ as *hierarchical relaying scheme*, and to the optimal scheme for $\alpha > 3$ as *multi-hop scheme*.

Given a permutation traffic on $V(n)$. For $\alpha \in (2, 3]$, hierarchical relaying achieves a per-node rate of $n^{1-\alpha/2-o(1)}$. For $\alpha > 3$, multi-hop communication achieves a per-node rate of $n^{-1/2-o(1)}$. By choosing the appropriate scheme, we can thus achieve a per-node rate of $n^{1-\min\{3,\alpha\}/2-o(1)}$. We provide a short description of the hierarchical relaying scheme in the following. The details can be found in [7].

Consider n nodes placed independently and uniformly at random on $A(n)$. Divide $A(n)$ into $n^{\frac{2}{\alpha} \log^{-1/3}(n)}$ squarelets

of equal size. Call a squarelet *dense*, if it contains a number of nodes proportional to its area. For each source-destination pair, choose such a dense squarelet as a *relay*, over which it will transmit information (see Figure 2).

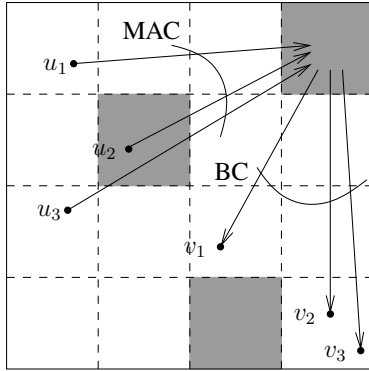


Fig. 2. Sketch of one level of the hierarchical relaying scheme. Here $\{(u_i, v_i)\}_{i=1}^3$ are three source-destination pairs. Groups of source-destination pairs relay their traffic over relay squarelets, which contain a number of nodes proportional to their area (shaded). We time share between the different relay squarelets. Within all relay squarelets the scheme is used recursively to enable joint decoding and encoding at each relay.

Consider now one such relay squarelet and the nodes that are transmitting information over it. If we assume for the moment that the nodes within the relay squarelets could cooperate then between the source nodes and the relay squarelet we would have a multiple access channel, where each of the source nodes has one transmit antenna, and the relay squarelet (acting as one node) has many receive antennas. Between the relay squarelet and the destination nodes, we would have a broadcast channel, where each destination node has one receive antenna, and the relay squarelet (acting again as one node) has many transmit antennas. The cooperation gain from using this kind of scheme arises from the use of multiple antennas for this multiple access and broadcast channel.

To actually enable this kind of cooperation at the relay squarelet, local communication within the relay squarelets is necessary. It can be shown that this local communication problem is actually the same as the original problem, but at a smaller scale. Indeed, we are now considering a square of size $n^{1-\frac{2}{\alpha}} \log^{-1/3}(n)$ with equal number of nodes (at least order wise). Hence we can use the same scheme recursively to solve this subproblem. We terminate the recursion after $\log^{1/3}(n)$ iterations, at which point we use simple TDMA to bootstrap the scheme.

Observe that at the final level of the scheme, we have divided $A(n)$ into

$$\left(n^{\frac{2}{\alpha}} \log^{-1/3}(n)\right)^{\log^{1/3}(n)} = n^{2/\alpha}$$

squarelets. A sufficient condition for the scheme to succeed is that all these squarelets are dense (i.e., contain a number of nodes proportional to their area). However much weaker conditions are sufficient as well (see [7]). The per-node rate achievable with this scheme is at least $n^{1-\alpha/2-o(1)}$.

We note in passing that for traffic matrices where a constant fraction of source-destination pairs are at distance $\Theta(\sqrt{n})$ (as is the case with probability $1 - o(1)$ as $n \rightarrow \infty$ if the permutation traffic is chosen uniformly at random), this is asymptotically the best uniformly achievable per-node rate.

B. Multicast Traffic

In this section, we present a scheme to transmit general multicast traffic. As will soon become clear, the special tree structure of this scheme is crucial in proving Theorem 1.

The communication scheme consists of two layers: A top or routing layer, and a bottom or physical layer. Seen from the routing layer, the network consists of a capacitated communication (or an equivalent wireline) graph G . This graph is a tree, whose leaf nodes represent the nodes $V(n)$ in the wireless network. The intermediate nodes of G represent larger clusters of nodes (i.e., subsets of $V(n)$) in the wireless network (see Figure 3). In the routing layer, messages are sent from

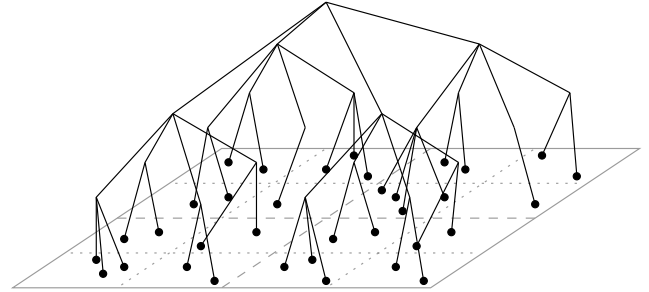


Fig. 3. Construction of the tree graph G . We consider the same nodes as in Figure 1 with $L(n) = 2$. The leaves of G are the nodes $V(n)$ of the wireless network. They are always at level $\ell = L(n) + 1$ (i.e., 3 in this example). At level $0 \leq \ell \leq L(n)$ in G , there are 4^ℓ nodes. The tree structure is the one induced by the grid decomposition $\{V_{\ell,i}(n)\}_{\ell,i}$ as shown in the figure. Level 0 contains the root node of G .

each source to its set of destinations by routing them over G . To send information along an edge of G , the physical layer is used. More precisely, to send information from a child node to its parent in G (i.e., towards the root node of G), the message at the cluster in $V(n)$ represented by the child node is distributed (over the wireless medium) evenly among all nodes in the bigger cluster in $V(n)$ represented by the parent node. To send information from a parent node to a child node in G (i.e., away from the root node of G), the message at the cluster in $V(n)$ represented by the parent node is concentrated on the cluster in $V(n)$ represented by the child node. This distribution and concentration of messages in the wireless network is performed by either using hierarchical relaying (for $\alpha \in (2, 3]$) or multi-hop communication (for $\alpha > 3$). It is this operation of each edge in the physical layer that determines the edge capacity of the graph G as seen from the routing layer.

Consider now a multicast message that needs to be transmitted from a source node $u \in V(n)$ to its set of intended destinations $W \subset V(n)$. As noted earlier, the scheme works by routing the message over the tree G (where sending a

message over an edge of G in the routing layer corresponds to the physical operation of the wireless network in the physical layer described in the last paragraph). Since G is a tree, the routing part is simple. In fact, between u and every $v \in W$ there exists only one (simple) path in G connecting u and v . Consider the union of all those paths. It is easy to see that this union is a subtree of G . Indeed, it is the smallest subtree of G that covers $\{u\} \cup W$. Traffic is optimally routed over G from u to W by sending it along the edges of this subtree. See Figure 4 for an example of how this two layer communication scheme works.

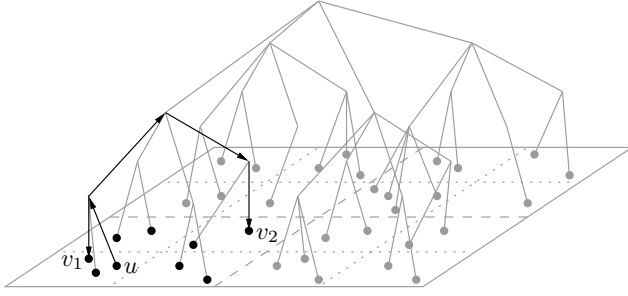


Fig. 4. We consider the same nodes as in Figure 1, and we use the same numbering of the $\{V_{\ell,i}(n)\}_{\ell,i}$ as there (i.e., from left to right and then bottom to top). Assume source node u wants to multicast a message to $\{v_1, v_2\}$. In the routing layer, we find the subgraph $G(\{u, v_1, v_2\})$ (indicated by black lines in the figure). Messages are sent from the source to its destinations by routing them along this subgraph (as indicated by the arrows). To send a message along an edge, the physical layer is called upon. Consider the path from u to v_2 . In the physical layer, communication along the first edge results in distributing the message from u to the remaining two nodes in $V_{2,1}(n)$. Communication along the second edge results in distributing the message further from $V_{2,1}(n)$ to the remaining six nodes in $V_{1,1}(n)$. Communication along the third edge results in concentrating the message from $V_{1,1}(n)$ on $V_{2,6}(n)$. Finally, communication along the fourth edge results in concentrating the message from $V_{2,6}(n)$ on v_2 .

IV. PROOF OF THEOREM 1

The proof of Theorem 1 relies on showing that in the case of wireless networks, knowledge of the unicast capacity region is sufficient to determine the multicast capacity region. We start by recalling a result from [12] characterizing the unicast capacity region.

The *unicast capacity region* of a wireless network, denoted by $\Lambda(n) \subset \mathbb{R}_+^{n \times n}$, is the collection of unicast traffic matrices that are achievable. Define for any unicast traffic matrix λ

$$\rho_\lambda^*(n) = \sup\{b \geq 0 : b\lambda \in \Lambda(n)\}.$$

Since $\Lambda(n)$ is a convex region, it can be completely characterized by the knowledge of $\rho_\lambda^*(n)$ for all traffic matrices $\lambda \in \mathbb{R}_+^{n \times n}$. In a previous work [12], we obtained the asymptotic (scaling) characterization of $\rho_\lambda^*(n)$ for any $\lambda \in \mathbb{R}^{n \times n}$, yielding an asymptotic characterization of $\Lambda(n)$.

For a unicast traffic matrix λ , define

$$\phi_\lambda^*(n) \triangleq \min_{\ell \in \{1, \dots, L(n)\} \cup \{\log(n)\}} \min_{i \in \{1, \dots, 4^\ell\}} \frac{g_\alpha(4^{-\ell}n)}{D_\lambda(V_{\ell,i}(n))},$$

where for any $U \subset V(n)$

$$D_\lambda(U) \triangleq \sum_{u \in U, v \in U^c} (\lambda_{u,v} + \lambda_{v,u}).$$

Then we have the following result.

Theorem 2 (Theorem 4 in [12]). *Under either fast or slow fading, for any $\alpha > 2$, $\varepsilon > 0$, there exist $b_1 = n^{-o(1)}$, $b_2 = O(n^\varepsilon)$ such that with probability $1 - o(1)$ as $n \rightarrow \infty$ for any unicast traffic matrix $\lambda \in \mathbb{R}_+^{n \times n}$*

$$b_1(n)\phi_\lambda^*(n) \leq \rho_\lambda^*(n) \leq b_2(n)\phi_\lambda^*(n).$$

We say that a unicast traffic matrix λ is *compatible* with a multicast traffic matrix $\tilde{\lambda}$ if there exists a mapping $f : V(n) \times 2^{V(n)} \rightarrow V(n)$ such that $f(u, W) \in W$, for all (u, W) , and

$$\lambda_{u,v} = \sum_{\substack{W \subset V(n): \\ f(u, W) = v}} \tilde{\lambda}_{u,W}$$

for all (u, v) . In words, $\tilde{\lambda}$ is compatible with λ if we can create the unicast traffic matrix λ from $\tilde{\lambda}$ by simply discarding the traffic for the pair (u, W) at all the nodes $W \setminus \{f(u, W)\}$. Let $\Gamma(\tilde{\lambda})$ be the set of all unicast traffic matrices compatible with the multicast traffic matrix $\tilde{\lambda}$.

Theorem 3. *Under either fast or slow fading, for any $\alpha > 2$, $\varepsilon > 0$, there exists $b(n) = O(n^\varepsilon)$ such that with probability $1 - o(1)$ as $n \rightarrow \infty$ for any multicast traffic matrices $\tilde{\lambda} \in \mathbb{R}_+^{n \times 2^n}$*

$$\begin{aligned} \tilde{\lambda} \in \tilde{\Lambda}(n) &\implies \Gamma(\tilde{\lambda}) \subset \Lambda(n), \\ \tilde{\lambda} \in \tilde{\Lambda}(n) &\iff b(n)\Gamma(\tilde{\lambda}) \subset \Lambda(n). \end{aligned}$$

Combining Theorems 2 and 3 proves Theorem 1.

Proof of Theorem 3. We start with some definitions concerning capacitated tree graphs. Let $G = (V_G, E_G)$ be a tree graph with leaf nodes² $V \subset V_G$, and with edge capacities $c(e)$ in both directions for $e \in E_G$. For a non-terminal node u in G , let $\mathcal{D}(u)$ be the descendants of u in G (including u itself, i.e., $u \in \mathcal{D}(u)$), and let $\mathcal{L}(u)$ denote the leaf nodes in $\mathcal{D}(u)$. Denote by $\Lambda_G \subset \mathbb{R}_+^{n \times n}$ the set of feasible unicast traffic matrices between leaf nodes of G (i.e., the set of flow rates with both source and destination in $V \subset V_G$ that can be routed through G). Similarly, denote by $\tilde{\Lambda}_G \subset \mathbb{R}_+^{n \times 2^n}$ the set of feasible multicast traffic matrices between leaf nodes of G .

Lemma 4. *There exists a undirected capacitated tree $G = (V_G, E_G)$ with leaf nodes $V \subset V_G$ such that for any $\varepsilon > 0$*

$$\begin{aligned} n^{-o(1)}\tilde{\Lambda}_G &\subset \tilde{\Lambda}, \\ \Omega(n^{-\varepsilon})\Lambda &\subset \Lambda_G, \end{aligned}$$

with probability $1 - o(1)$ as $n \rightarrow \infty$.

Proof. Construct a graph $G = (V_G, E_G)$ as follows. G is a full tree (i.e., all its leaf nodes are on the same level). G has n

²To simplify notation, we suppress dependence on n within proofs whenever this dependence is clear from the context.

leaves, each of them representing an element of V . To simplify notation, we assume that $V \subset V_G$, so that the leaves of G are exactly the elements of $V \subset V_G$. Whenever the distinction is relevant, we use u, v for nodes in $V \subset V_G$ and μ, ν for nodes in $V_G \setminus V$ in the following. The non-terminal nodes of G correspond to $V_{\ell,i}$ for all $\ell \in \{0, \dots, L(n)\}$, $i \in \{1, \dots, 4^\ell\}$, with hierarchy induced by the one on $A(n)$ (see Figure 5 and Figure 3 in Section III-B).

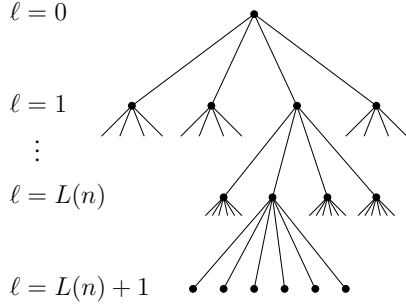


Fig. 5. Communication graph G constructed in the proof of Theorem 1. Nodes on levels $\ell \in \{0, \dots, L(n) - 1\}$ have each four children, nodes on level $\ell = L(n)$ have each $\Theta(n^{\log^{-1/2}(n)})$ children. The total number of terminal nodes is n , one representing each node in the wireless network $V(n)$. A non-terminal node in G at level $\ell \in \{0, \dots, L(n)\}$ represents the collection of nodes in $V_{\ell,i}(n)$ for some i .

Thus, nodes in V_G at level $\ell < L(n)$ have each 4 children, nodes in V_G at level $\ell = L(n)$ have between $4^{-L(n)-1}n$ and $4^{-L(n)+1}n$ children with high probability, and nodes in V_G at level $\ell = L(n) + 1$ are the leaves of the tree. To understand the relation between V_G and V , we define the *representative* $\mathcal{R} : V_G \rightarrow 2^V$ of μ as follows. For a leaf node $u \in V \subset V_G$ of G , let

$$\mathcal{R}(u) \triangleq \{u\}.$$

For $\mu \in V_G$ at level $L(n)$, choose $\mathcal{R}(\mu) \subset \mathcal{L}(\mu) \subset V$ such that

$$|\mathcal{R}(\mu)| = 4^{-L(n)-1}n.$$

This is possible with probability $1 - o(1)$ as $n \rightarrow \infty$ for a random node placement. Finally, for $\mu \in V_G$ at level $\ell < L(n)$, and with children $\{\nu_i\}_{i=1}^4$, let

$$\mathcal{R}(\mu) \triangleq \bigcup_{j=1}^4 \mathcal{R}(\nu_j).$$

We now define an edge capacity $c(\mu, \nu)$ for each edge $(\mu, \nu) \in E_G$. If μ is a leaf of G and ν its parent, set

$$c(\mu, \nu) = c(\nu, \mu) \triangleq 1. \quad (6)$$

If μ is a non-terminal node at level ℓ in G and ν its parent, then set

$$c(\mu, \nu) = c(\nu, \mu) \triangleq (4^{-\ell}n)^{2-\min\{3, \alpha\}/2} = g_\alpha(4^{-\ell}n). \quad (7)$$

We argue now that if a multicast traffic $\tilde{\lambda}$ in which only the leaf nodes of G are sources or destinations can be routed through G , then $n^{o(-1)}\tilde{\lambda} \in \tilde{\Lambda}$ (i.e., almost the same flow

can be reliably transmitted over the wireless network). The idea (that will be made precise) is the following. To transmit information from a non-terminal node $\mu \in V_G$ to its parent node ν , we split the message at each node in $\mathcal{R}(\mu)$ into four parts and send one part to each node in $\mathcal{R}(\nu)$. In other words, we distribute the message by a factor four over the wireless network. To transmit information from a node $\mu \in V_G$ with non-terminal children $\{\nu_j\}_{j=1}^4$ to one of them, say ν_1 , we send the message parts from each $\{\mathcal{R}(\nu_j)\}_{j=2}^4$ to a corresponding node in $\mathcal{R}(\nu_1)$ and combine them there. In other words, we concentrate the message by a factor four over the wireless network. The scheme is bootstrapped at the leaves $V \subset V_G$ of G (where, by our definition of $\tilde{\lambda}$, all traffic originates and ends) as follows. To send a message from a leaf node $u \in V \subset V_G$ to its parent ν in G , the message is split at u into $|\mathcal{R}(\nu)|$ equal pieces, and one piece is sent to each node in $\mathcal{R}(\nu)$ over the wireless network. In other words, we distribute the message again over the wireless network, but this time by a factor of $|\mathcal{R}(\nu)|$. To send a message to a leaf node $u \in V \subset V_G$ from its parent ν in G , each node in $\mathcal{R}(\nu)$ sends its piece of the message to u over the wireless network. Thus, again we concentrate the message over the network, but this time by a factor of $|\mathcal{R}(\nu)|$.

We now analyze this scheme in more detail. Note first that by time sharing between the $L(n) + 1$ non-terminal levels of the tree, and by appropriate spatial reuse within each level, we only loose a factor of at most

$$K \frac{1}{4^{L(n)+1}} \leq n^{-o(1)}$$

for some constant $K > 0$ in rate. Hence it is sufficient to consider communication between a non-terminal node of G and its children.

We first consider communication up the tree (i.e., towards the root). Let $u \in V \subset V_G$ be a leaf node of G and ν be its parent. To send traffic at rate $c(u, \nu)$ from u to ν , node u splits its traffic into $|\mathcal{R}(\nu)| = 4^{-L(n)-1}n$ equal parts and sends each part to one node in $\mathcal{R}(\nu)$. Recall $\mathcal{R}(\nu) \subset V_{L(n),i}$ for some i . Since

$$r_{u,v} \leq 2(4^{-L(n)}n)^{1/2}$$

for any $u, v \in V_{L(n),i}$, communicating between u and v , we incur a power loss of

$$r_{u,v}^{-\alpha} \geq 2^{-\alpha}(4^{-L(n)}n)^{-\alpha/2} \geq 2^{-\alpha}n^{-\log^{-1/2}(n)\alpha/2},$$

and hence we can communicate between u and v at a rate of at least

$$\log(1 + P2^{-\alpha}n^{-\log^{-1/2}(n)\alpha/2}) \geq n^{-o(1)}.$$

By time sharing between all the destination nodes in $\mathcal{R}(\nu)$, and since all message parts are only $|\mathcal{R}(\nu)|^{-1}$ of the size of the original message, all message parts of u can be transmitted from u to $\mathcal{R}(\nu)$ at this rate. It can be shown that for random node placement

$$|\mathcal{L}(\nu)| \leq 4^{-L(n)+1}n = 4n^{\log^{-1/2}(n)},$$

with high probability, and hence further time sharing between all source nodes in $\mathcal{L}(\nu)$, we can communicate simultaneously from all leaf nodes $u \in \mathcal{L}(\nu)$ to $\mathcal{R}(\nu)$ at a rate at least

$$4^{-1}n^{-\log^{-1/2}(n)}n^{-o(1)} \geq n^{-o(1)}c(u, \nu).$$

Let now $\nu \in V_G$ be a node in level $\ell < L(n)$ in G and let $\{\mu_j\}_{j=1}^4$ be its children. Since

$$|\mathcal{R}(\mu_j)| = 4^{-\ell(n)-2}n$$

for all $j \in \{1, \dots, 4\}$, we can find a one-to-one correspondence between $\mathcal{R}(\mu_j)$ and $\mathcal{R}(\mu_k)$. Choose an arbitrary such correspondence for each $j, k \in \{1, \dots, 4\}, j \neq k$. Now, we know that the traffic to be sent from μ_j to ν originates at one or several nodes in $\mathcal{L}(\mu_j)$. Thus by construction of the previous stages, all nodes in $\mathcal{R}(\mu_j)$ possess an equal part of the total message to be transmitted from μ_j to ν . Split each such message part further into four equal parts and consider one particular node $u \in \mathcal{R}(\mu_k)$. The first part of the message stays at u . The other three parts are to be transmitted to the corresponding nodes in $\{\mathcal{R}(\mu_j)\}_{j \neq k}$. Time sharing between all 12 possible (j, k) pairs, we only incur a constant loss. Hence we can focus on communication between a particular (j, k) pair. We are now in a situation with $4^{-\ell(n)-2}n$ nodes in $\mathcal{R}(\mu_j)$, each with a message of equal size for the corresponding node in $\mathcal{R}(\mu_k)$. Note that this is a permutation traffic (as defined in Section III-A). Hence using existing results on unicast permutation traffic (see again Section III-A), we can therefore use hierarchical relaying (for $\alpha \in (2, 3]$) or multi-hop for ($\alpha > 3$) to transmit at a total rate up to

$$n^{-o(1)}(4^{-\ell}n)^{2-\min\{3, \alpha\}/2} = n^{-o(1)}c(\mu, \nu).$$

Consider now communication down the tree (i.e., away from the root). Communication between ν and μ works in the same fashion by concentrating the messages. The same arguments as in the previous two paragraphs show that any rate up to (6) or (7) are achievable up to a factor $n^{-o(1)}$. Time sharing between the two directions, yields an additional rate loss of a factor $1/2$. Together, this shows that

$$n^{-o(1)}\tilde{\Lambda}_G \subset \tilde{\Lambda},$$

proving the first half of the lemma.

We now turn to the proof of the second half of the lemma. Let $e = (u, v) \in E_G$, and assume that v is the parent of u in G ; with slight abuse of notation, define $\mathcal{D}(e) \triangleq \mathcal{D}(u)$. For a unicast traffic matrix $\lambda \in \mathbb{R}_+^{n \times n}$, define

$$\gamma_\lambda^* \triangleq \sup\{b \geq 0 : b\lambda \in \Lambda_G\},$$

and for $e \in E_G$, let

$$d_\lambda(e) \triangleq \sum_{\substack{u \in \mathcal{D}(e) \\ v \notin \mathcal{D}(e)}} (\lambda_{u,v} + \lambda_{v,u}).$$

We now show that

$$\gamma_\lambda^* \geq \min_{e \in E_G} \frac{c(e)}{d_\lambda(e)}. \quad (8)$$

Indeed, assume (8) does not hold. Then the load over every edge $e \in E_G$ is strictly less than $c(e)$, and hence it is possible to increase the flow for each (u, v) pair by a strictly positive amount. This contradicts the definition of γ_λ^* , and hence shows (8).

Consider now an edge $e = (\mu, \nu) \in E_G$, and assume that ν is the node closer to the root of G . Let ℓ be the level of μ in the tree. Then, by construction, $c(e)$ is only a function of ℓ and given by either (6) or (7), which, in turn, is equal to $g_\alpha(4^{-\ell}n)$. Moreover, $d_\lambda(e)$ is either equal to $D_\lambda(\{u\})$ for some $u \in V$ if $\ell = L(n) + 1$, or equal to $D_\lambda(V_{\ell,i})$ for some i if $\ell \leq L(n)$. It can be shown that for all $u \in V$, we have $\{u\} = V_{\log(n),i}$ for some i with probability $1 - o(1)$, and thus

$$\begin{aligned} \gamma_\lambda^* &\geq \min_{e \in E_G} \frac{c(e)}{d_\lambda(e)} \\ &= \min_{\ell \in \{1, \dots, L(n)\} \cup \{\log(n)\}} g_\alpha(4^{-\ell}n) \min_{i \in \{1, \dots, 4^\ell\}} \frac{1}{D_\lambda(V_{\ell,i}(n))} \\ &= \phi_\lambda^*(n). \end{aligned}$$

Comparing this to the upper bound in Theorem 2 shows that

$$\rho_\lambda^* \leq O(n^\varepsilon)\gamma_\lambda^*,$$

where the $O(n^\varepsilon)$ term is uniform in λ . By convexity of Λ and Λ_G , this proves the second half of the lemma. \square

Lemma 5. Let $G = (V_G, E_G)$ be an undirected capacitated tree and let Λ_G and $\tilde{\Lambda}_G$ be the unicast and multicast capacity region of G , respectively. Then

$$\tilde{\lambda} \in \tilde{\Lambda}_G \iff \Gamma(\tilde{\lambda}) \subset \Lambda_G.$$

Proof. Clearly,

$$\tilde{\lambda} \in \tilde{\Lambda}_G \implies \Gamma(\tilde{\lambda}) \subset \Lambda_G$$

(this is, in fact, true for all graphs G , not only trees).

Conversely, assume that $\tilde{\lambda} \notin \tilde{\Lambda}_G$. Since G is a tree, there is only one way to route multicast traffic from u to W , namely among the subtree induced by $\{u\} \cup W$. For a set $U \subset V_G$ of nodes, denote by $G(U)$ the smallest subtree of G that covers U . Then traffic from u to W is routed along $G(\{u\} \cup W)$. Hence for any edge $e \in E_G$, the traffic that needs to be routed over e is equal to

$$d_{\tilde{\lambda}}(e) = \sum_{\substack{u \in V_G, W \subset V_G: \\ e \in E_G(\{u\} \cup W)}} \tilde{\lambda}_{u,W}.$$

Now, since $\tilde{\lambda} \notin \tilde{\Lambda}_G$, there exists $e \in E_G$ such that

$$d_{\tilde{\lambda}}(e) > c(e).$$

But then, by definition of $d_{\tilde{\lambda}}(e)$, there exists a function $f : V(n) \times 2^{V(n)} \rightarrow V(n)$ with $f(u, W) \in W$ for all (u, W) , and such that

$$d_{\lambda_f(\tilde{\lambda})}(e) = d_{\tilde{\lambda}}(e) > c(e),$$

where $\lambda_f(\tilde{\lambda})$ is the unicast traffic matrix produced by applying f to the multicast traffic matrix $\tilde{\lambda}$. Thus $\lambda_f(\tilde{\lambda}) \notin \Lambda_G$, and therefore $\Gamma(\tilde{\lambda}) \setminus \Lambda_G \neq \emptyset$. This implies

$$\tilde{\lambda} \in \tilde{\Lambda}_G \iff \Gamma(\tilde{\lambda}) \subset \Lambda_G.$$

□

We are now ready to prove Theorem 3. Consider first the upper bound. It is clear that if $\tilde{\lambda} \in \tilde{\Lambda}$ then $\lambda \in \Lambda$ for any $\lambda \in \Gamma(\tilde{\lambda})$. Thus

$$\tilde{\lambda} \in \tilde{\Lambda} \implies \Gamma(\tilde{\lambda}) \subset \Lambda,$$

On the other hand, assume $\Gamma(\tilde{\lambda}) \subset \Lambda$. With G as in Lemma 4, we then have

$$\Gamma(\tilde{\lambda}) \subset \Lambda \subset O(n^{\tilde{\varepsilon}})\Lambda_G,$$

or, equivalently,

$$\Gamma(\Omega(n^{-\tilde{\varepsilon}})\tilde{\lambda}) \subset \Lambda_G.$$

Therefore by Lemma 5,

$$\Omega(n^{-\tilde{\varepsilon}})\tilde{\lambda} \in \tilde{\Lambda}_G,$$

and applying Lemma 4

$$\Omega(n^{-2\tilde{\varepsilon}})\tilde{\lambda} \in \tilde{\Lambda}.$$

Setting $\varepsilon = 2\tilde{\varepsilon}$, we have hence shown that

$$\Gamma(\tilde{\lambda}) \subset \Lambda \implies \Omega(n^{-\varepsilon})\tilde{\lambda} \in \tilde{\Lambda},$$

where the $\Omega(n^{-\varepsilon})$ is uniform in $\tilde{\lambda}$. Since $\tilde{\varepsilon} > 0$ and thus $\varepsilon > 0$ is arbitrary, this concludes the proof of Theorem 3. □

V. CONCLUSIONS

In this paper, we have obtained an explicit characterization of the scaling of the multicast capacity region of wireless networks with n nodes located uniformly at random in an extended region. The characterization is in terms of $2n$ weighted cuts, which are based on the geometry of the locations of the source nodes and their destination nodes and the traffic demands between them, and thus can be readily evaluated. This characterization is obtained by establishing two equivalence relations of the multicast capacity region. The first equivalence is of the scaling of the multicast capacity region to that of the intersection of a set of induced unicast capacity regions. The second equivalence establishes that the multicast capacity region of a logical rooted tree, derived from the wireless network, has essentially the same scaling as that of the original network. The latter also suggests a two-layer architecture for achieving the optimal scaling of the multicast capacity region. The top layer, which we refer to as the routing layer, establishes paths from each of the source nodes to its set of destinations over the rooted tree, whose nodes correspond to hierarchically growing sets of nodes. The bottom layer, which we refer to as the physical layer, distributes the traffic among the corresponding set of nodes as a path travels up the rooted tree and concentrates the traffic on to the corresponding subset of nodes when the path travels down the tree. How

this distribution/concentration of traffic is physically achieved depends on the path loss exponent: In the low path loss case, $\alpha \in (2, 3]$, it is achieved through hierarchical relaying, while in the high path loss case ($\alpha > 3$), the same is achieved through multi-hopping. This scheme also implicitly establishes that a separation based approach, where the routing layer works essentially independently of the physical layer, can achieve nearly the entire multicast capacity region in the scaling sense. Thus, such techniques as network coding can provide at most a small increase in the scaling.

REFERENCES

- [1] E. M. Royer and C. E. Perkins, "Multicast operation of the ad hoc on-demand distance vector routing protocol," *Procs. ACM MobiCom*, Seattle, WA, August 1999.
- [2] P. Sinha, R. Sivakumar, and V. Bharghavan, "MCEDAR: Multicast Core-Extraction Distributed Ad Hoc Routing," *Procs. Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, September 1999.
- [3] S.-J. Lee, M. Gerla, and C.-C. Chiang, "On-Demand Multicast Routing Protocol," *Procs. Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, September 1999.
- [4] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks," *Procs. IEEE INFOCOM*, Tel-Aviv, Israel, March 2000.
- [5] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, 46(2):388–404, March 2000.
- [6] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inform. Theory*, 53(10):3549–3572, October 2007.
- [7] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *submitted to IEEE Trans. Inform. Theory*, November 2007. Available online at <http://arxiv.org/abs/0711.2745>.
- [8] F. Xue and P. R. Kumar, "Scaling laws for ad-hoc wireless networks: An information theoretic approach," *Foundation and Trends in Networking*, 1(2), 2006.
- [9] R. Madan, D. Shah, and O. Lévêque, "Product multicommodity flow in wireless networks," *IEEE Trans. Inform. Theory*, 54(4):1460–1476, April 2008.
- [10] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, 15(2):215–245, June 1995.
- [11] S. Subramanian, S. Shakkottai, and P. Gupta, "Optimal geographic routing for wireless networks with near-arbitrary holes and traffic," *Procs. IEEE INFOCOM*, Phoenix, AZ, Apr. 13–18, 2008.
- [12] U. Niesen, P. Gupta, and D. Shah, "The capacity region of large wireless networks," *submitted to IEEE Trans. Inform. Theory*, July 2008. Available online at <http://arxiv.org/abs/0809.1344>.
- [13] X.-Y. Li, S.-J. Tang, and O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," *Procs. ACM MobiCom*, pages 263–277, 2007.
- [14] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," *Procs. ACM MobiHoc*, 2007.
- [15] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," *Procs. ACM MobiCom*, 2006.
- [16] P. Chaporkar and S. Sarkar, "Wireless Multicast: Theory and Approaches," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1954–1972, June 2005.
- [17] P. Jacquet and G. Rodolakis, "Multicast scaling properties in massively dense ad hoc networks," *Procs. 11th International Conference on Parallel and Distributed Systems* Washington, DC, 2005.
- [18] P. Jacquet, "Geometry of information propagation in massively dense ad hoc networks," *Procs. ACM MobiHoc*, Roppongi Hills, Japan, May 2004.
- [19] R. Alshwede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, 46(4):1204–1216, 2000.
- [20] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, 49(2):371–381, February 2003.
- [21] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, 11(5):782–795, 2003.
- [22] R. Koetter, "Separation for network coding," *Preprint*, 2007.