A Communication Architecture for Large Heterogeneous Wireless Networks

Urs Niesen
Bell Laboratories, Alcatel-Lucent
Murray Hill, NJ 07974
urs.niesen@alcatel-lucent.com

Piyush Gupta
Bell Laboratories, Alcatel-Lucent
Murray Hill, NJ 07974
pgupta@research.bell-labs.com

David Tse
University of California at Berkeley
Berkeley, CA 94720
dtse@eecs.berkeley.edu

***Abstract*—Capacity scaling in wireless networks under homogeneous node placement and traffic demands has been widely studied over the last decade. In general networks, however, both the node placement and traffic demands can be highly nonuniform. In this paper, we present a communication architecture that operates a wireless network under such heterogeneous settings. The proposed scheme includes the existing communication schemes for large wireless networks—multi-hopping, cooperative communication, cooperative multi-hopping—as special cases. We illustrate through a few specific wireless network scenarios that the proposed scheme can perform significantly better, even scaling-wise, than any of these previous schemes.**

I. Introduction

Considerable progress has been made over the last decade in quantifying performance in large wireless networks. Since the capacity region is not known even for small networks (e.g., a three-node relay network, or a four-node interference network), this progress has been made by asking for less: Instead of determining the capacity region exactly, the objective has been to find inner and outer bounds on this region that have the same order of scaling as the network size increases. Even for this simplified question, answers are only known under a number of restrictive homogeneity assumptions on the node placement and traffic demands. This limits the applicability of these results to real networks, which are often highly heterogeneous, and hence do not usually satisfy these assumptions.

In this paper, we do away with all such homogeneity assumptions and allow for arbitrary placement of nodes on a square area of arbitrary size, and with arbitrary unicast or multicast traffic demands between node pairs in the network. We develop a general communication architecture that operates a wireless network under such heterogeneous conditions. The proposed communication scheme incorporates existing communication schemes for large wireless networks, such as multi-hopping and multi-user cooperative communication, in a common unifying framework. The scheme is, however, more general than a mere combination of these schemes, as it determines what specific strategy to deploy in different regions of the network based on the level of heterogeneity in the node placement and/or traffic demands at different scales of the network. We establish that the proposed scheme achieves order-optimal capacity scaling in several scenarios.

This work was supported in part by AFOSR under grant FA9550-09-1-0317.

A. Related Work

Consider an extended wireless network in which n nodes are placed in the square region $[0,\sqrt{n}]^2$ of area n and communicate with each other over Gaussian fading channels with path-loss exponent $\alpha \geq 2$. The objects of interest in this paper are the unicast capacity region $\Lambda^{\mathsf{UC}}(n) \subset \mathbb{R}_+^{n\times n}$ and multicast capacity region $\Lambda^{\mathsf{MC}}(n) \subset \mathbb{R}_+^{n\times 2^n}$, describing achievable rates between all possible source-destination pairs.

The study of capacity scaling laws was initiated by Gupta and Kumar in [1]. Under a *protocol channel model*, in which interference is treated as noise and only point-to-point communication is allowed, and under random source-destination pairing with uniform traffic demands, the largest achievable per-node rate is shown to scale like $n^{-1/2\pm o(1)}$. Hence, under the protocol channel model, and assuming random node placement, [1] provides the scaling behavior of one point of the $n\times n$-dimensional unicast capacity region $\Lambda^{\mathsf{UC}}(n)$.

Subsequent work in the information theory literature focused on removing the protocol channel model assumption made in [1] and instead considered *Gaussian fading channels*. In a series of papers [2]–[9], upper bounds on the achievable rates for random source-destination pairing were derived. In particular, Özgür et al. [8] have shown that for $\alpha \geq 3$ multi-hop communication is indeed optimal, and hence the largest uniformly achievable per-node rate scales like $n^{-1/2\pm o(1)}$. On the other hand, in another stream of work [8], [10]–[13], it is shown that for $2 \leq \alpha < 3$ cooperative communication schemes significantly outperform multi-hop communication. In particular, [8] introduced a hierarchical cooperative communication scheme achieving the order optimal per-node rate scaling of $n^{1-\alpha/2\pm o(1)}$. This provides scaling information, now under the Gaussian fading channel model, but still assuming random node placement and again about one point in $\Lambda^{\mathsf{UC}}(n)$.

The impact of the random node placement assumption on achievable rates was investigated by Niesen et al. in [14], where it is shown that for low path-loss exponent ($\alpha < 3$), the same uniform per-node rate of $n^{1-\alpha/2\pm o(1)}$ is achievable regardless of the node placement. On the other hand, in the high path-loss regime ($\alpha \geq 3$), the regularity of the node placement crucially affects achievable rates as well as the nature of order-optimal communication schemes. In particular, it is shown in [14] that there are node placements (containing "gaps") for which multi-hop communication is not order optimal for any value of α, contrasting with the results for random node placement where multi-hop communication is order optimal for all $\alpha \geq 3$.

General traffic patterns were considered in [15] under random node placement. For $\alpha > 5$, a scaling characterization of the entire $n \times n$-dimensional unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ and the entire $n \times 2^n$-dimensional multicast capacity region $\Lambda^{\mathsf{MC}}(n)$ are provided. For $2 \leq \alpha \leq 5$ the scaling of all but $2n$ dimensions of $\Lambda^{\mathsf{UC}}(n)$ and $\Lambda^{\mathsf{MC}}(n)$ is characterized.

If arbitrary node placement as well as arbitrary traffic patterns are allowed, the problem becomes considerably harder to deal with. To the best of our knowledge, no scaling results for this general wireless network setting are known.

B. Organization

The remainder of this paper is organized as follows. We start in Section II with an example to motivate the need for a more general communication architecture to address the combined non-uniformity in node placement and traffic demands. In Section III, we describe the network and the channel models. We present the main results of this paper in Section IV. Section V contains discussions and concluding remarks.

Due to space constraints, proofs are omitted in this paper.

II. A Motivating Example

Consider the node placement in Fig. 1 of area n. V_1, V_2, and V_3 are three sets of $n/3$ nodes, and the nodes V of the wireless network are the union of $\{V_i\}_{i=1}^3$. $\{u_i\}_{i=1}^3$ and $\{w_i\}_{i=1}^3$, as shown in the figure, are nodes in V. Each node $u \in V$ is source for only one destination node w chosen randomly from among the other nodes $V \setminus \{u\}$; this results in n source-destination pairs. We assume that the rate at which different source-destination pairs want to communicate depends only on which of the sets $\{V_i\}$ they are in. For example, if $u, \tilde{u} \in V_1$ and $w, \tilde{w} \in V_3$, and $u, \tilde{u}$ are the sources for $w, \tilde{w}$, respectively, then they both want to communicate at the same rate. The nodes communicate over a Gaussian fading channel with path-loss exponent $\alpha > 2$, i.e., power decays as $r^{-\alpha}$ over distance r.

Let $\{(u_i, w_i)\}_{i=1}^3$ be three source-destination pairs. Furthermore, assume a path-loss exponent $\alpha \in (3, 4)$. For the source-destination pair (u_1, w_1), it can be shown that the order-optimal communication strategy is multi-hop communication, in which each message is encoded, sent to a close-by neighbor, where it is decoded and re-encoded for the next hop. On the other hand, for the source-destination pair (u_2, w_2), the order-optimal communication strategy is cooperative communication, in which the node u_2 establishes a distributed antenna array consisting of order n transmitters (the node u_2 and its $n/18$ neighbors in the upper right corner of V_1) and order n receivers (the node w_2 and its $n/3$ neighbors in V_2). The message is then sent from the source u_2 to its destination w_2 in one direct hop using this virtual multiple-input multiple-output (MIMO) channel. Finally, for the source-destination pair (u_3, w_3), the order-optimal communication strategy is cooperative multi-hop communication, in which a message is routed from u_3 to w_3 using hops of length $n^{1/4}$ and each hop is implemented via cooperative communication of order $n^{1/4}$ nodes. From this, we see that in a efficient communication strategy *different regions in the network need to use different communication schemes*.

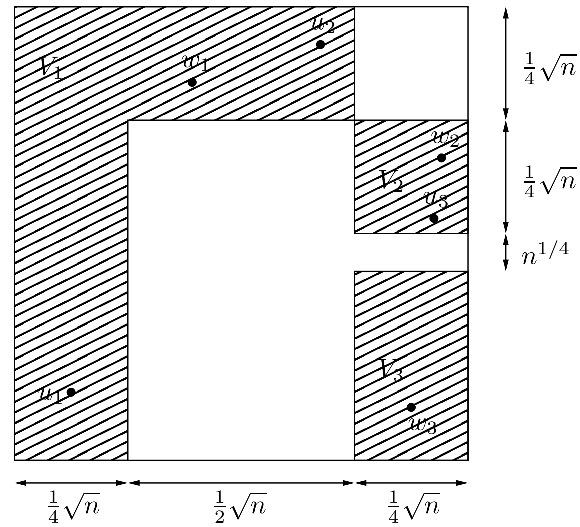

Fig. 1. Node placement illustrating the impact or location and traffic heterogeneity on the order-optimal communication scheme. The sets V_1, V_2, V_3 contain $n/3$ uniformly placed nodes each.

Assume now that u_1 is the source node for destination node w_3, and let $\alpha > 4$. An order optimal communication strategy is to route the message from u_1 through V_1 and V_2 using multi-hop communication, and then use cooperative multi-hop communication to carry the message to the node w_3 in V_3. Thus *various different communication schemes may need to be interconnected to carry traffic over the network*.

Finally, assume again that u_1 is source node for destination node w_3, but consider $\alpha \in (2, 4)$. An order-optimal communication strategy is then to directly communicate between u_1 and w_3, using cooperative communication. In other words, the behavior of the optimal routes depend strongly on whether $\alpha < \alpha^\star$ or $\alpha > \alpha^\star$ with $\alpha^\star = 4$. Moreover, by slightly altering the node placement, the threshold $\alpha^\star$ for abrupt changes of the order-optimal routes can be made to take any value between 3 and 4. Thus *the optimal routes over the network depend strongly and in a rather subtle manner on the node placement and the value of the path-loss exponent* α.

Note that multi-hop communication, cooperative communication, and cooperative multi-hop communication can all be understood as cooperation at various scales in the wireless network. In multi-hop communication, nodes cooperate only at local scale; in cooperative communication, nodes cooperate at global scale; in cooperative multi-hop, nodes cooperate at intermediate scale between these two extremes. Thus, the three observations made above suggest the following communication architecture. Decompose the network into (overlapping) subsets at various scales. In each subset, perform cooperative communication at that scale. To send a message from a source to its destination, route it over such subsets (at possibly different scales), using in each subset the corresponding communication scheme. By optimizing over the routes, we ensure that the proper communication strategy is used in each region of the network.

As we will see in the following, the communication scheme in each subset $\tilde{V} \subset V$ of nodes can be conveniently described as a graph $G_{\tilde{V}}$. These graphs can be connected, resulting in a larger graph G. Routing over subsets can then equivalently be understood as routing over this graph G. Thus, the graph G captures the relevant parts of the heterogeneity of the node placement.

III. Models and Notation

Consider a wireless network consisting of n nodes $V(n) \subset A(n)$ placed in an arbitrary fashion on the square region $A(n)$ of area

$$|A(n)| \triangleq n^{\nu}.$$

The parameter ν couples the growth of the network area with the growth of the number of nodes n. Two special cases are worth mentioning: For $\nu = 1$ this results in an *extended wireless network* for which the network area scales linearly with the number of nodes; for $\nu = 0$ this results in a *dense network* for which the network area is constant regardless of the number of nodes. In general, ν can be any nonnegative real number. Note that we make no probabilistic assumption for the node placement, but rather allow for any arbitrary (deterministic) node placement. Denote by $r_{u,v}$ the Euclidean distance between nodes u and v in $V(n)$. Moreover, for subset $W \subset V(n)$ and node $u \in V(n)$, define

$$r_W \triangleq \max_{u,w \in W} r_{u,w}.$$

The nodes in the wireless network communicate with each other over the wireless channel modeled as follows. The baseband-equivalent *received signal* $y_v[t]$ at node v at time t is given by

$$y_v[t] \triangleq \sum_{u \in V(n) \setminus \{v\}} h_{u,v}[t] x_u[t] + z_v[t]. \tag{1}$$

Here, $x_u[t]$ is the *transmitted signal* by node u at time t, and we impose an average unit power constraint on $x_u[t]$. $h_{u,v}[t]$ is the *channel gain* between the nodes u and v and is assumed to be of the form

$$h_{u,v}[t] \triangleq r_{u,v}^{-\alpha/2} \exp(\sqrt{-1}\theta_{u,v}[t]), \tag{2}$$

where the constant $\alpha \geq 2$ is the *path-loss exponent*, and $\theta_{u,v}[t]$ is the *phase shift* between nodes u and v at time t. As a function of the nodes (u, v), the phase shifts $\{\theta_{u,v}[t]\}_{u,v}$ are assumed to be independently and identically distributed (i.i.d.) uniform random variables on $[0, 2\pi)$. As a function of time t, the phase shifts $\{\theta_{u,v}[t]\}_t$ are assumed to be stationary and ergodic. All phase shifts $\{\theta_{u,v}[t]\}_{u,v}$ are assumed to be known causally at all nodes in the network, i.e., every node u knows all n^2 channel gains $\{\theta_{u,v}[t]\}_{u,v}$ at time t. Since the network is static (i.e., $r_{u,v}$ does not vary as a function of t), this implies that full causal channel state information (CSI) is available at all the nodes in the network. Finally, $z_v[t]$ in (1) is additive noise at node v at time t. We assume that $\{z_v[t]\}_{v,t}$ is i.i.d. circularly symmetric Gaussian with mean zero and variance one, independent of the transmitted signals and the channel gains.

Making the assumption of i.i.d. phase fading in (2), i.e., the independence of $\{\theta_{u,v}[t]\}_{u,v}$ as a function of the nodes (u, v), requires some care. In particular, it is shown in [9], [16], [17] that the i.i.d. phase fading assumption is only valid if the wavelength λ_c of the carrier frequency of communication, which is not explicitly captured in the baseband-equivalent channel model (1), satisfies $\lambda_c \leq |A(n)|^{1/2}/n$. In the following, we assume that the network operates in the regime in which this condition is satisfied, so that the channel model (1)–(2) is valid.

IV. Main Results

This section contains the main results of the paper. We first define a directed capacitated (noiseless) graph $G = (V_G, E_G)$ such that $V(n) \subset V_G$. This construction is described in Section IV-A. We then argue that G is implementable in the wireless network. More precisely, if messages can be routed at rates λ over G, then these messages can also be reliably transmitted over the wireless network. The resulting inner bounds for the unicast and multicast capacity regions $\Lambda^{\mathsf{UC}}(n)$ and $\Lambda^{\mathsf{MC}}(n)$ of the wireless network are presented in Sections IV-B and IV-C, respectively.

A. Construction of G

For each $\ell \in \mathbb{N}$, partition the area $A(n)$ into 4^{ℓ} subsquares $\{A_{\ell,i}^1\}_{i=1}^{4^{\ell}}$ of equal size. In the following, we will be interested in values of ℓ between 0 and[1]

$$L(n) \triangleq \frac{1}{2}\nu \log(n),$$

i.e., for subsquares varying in size from n^{ν} [for $\ell = 0$] to 1 [for $\ell = L(n)$]. The sidelength of the subsquare $A_{\ell,i}^1$ is $n^{\nu/2}2^{-\ell}$. Assume we "shift" the way the subsquares at level ℓ are defined by $\frac{1}{2}n^{\nu/2}2^{-\ell}$ to the right. Call $\{A_{\ell,i}^2\}_2$ the resulting partition of $A(n)$. Note that the subsquares $A_{\ell,i}^2$ at the boundary of $A(n)$ will now have size that is only half of the ones in the center, and hence the number of subsquares is now $4^{\ell} + 2^{\ell}$. For $j \in \{3, 4\}$, define $\{A_{\ell,i}^2\}_j$ similarly by "shifts" of length $\frac{1}{2}n^{\nu/2}2^{-\ell}$ to either the top or both to the right and the top, respectively. The number of subsquares is $4^{\ell} + 2^{\ell}$ and $4^{\ell} + 2^{\ell} + 1$. Denote by

$$V_{\ell,i}^j \triangleq V(n) \cap A_{\ell,i}^j$$

the nodes in the subsquare $A_{\ell,i}^j$.

[1] All logarithms are to base 2.

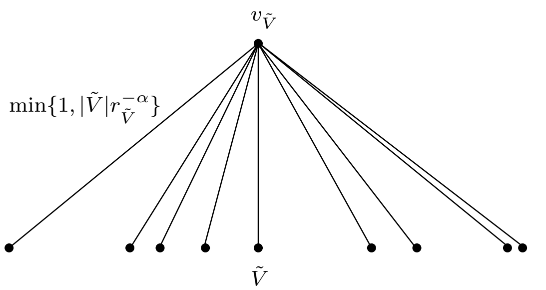

Fig. 2. Construction of the graph $G_{\tilde{V}}$. $\tilde{V}$ is a subset of the nodes $V(n)$ of the wireless network; $v_{\tilde{V}}$ is an additional node not in $V(n)$. Each edge in $G_{\tilde{V}}$ is undirected with capacity $\min\{1, |\tilde{V}| r_{\tilde{V}}^{-\alpha}\}$.

Let $\ell \in \{0, \ldots, L(n)\}$, and consider the nodes $V_{\ell,i}^{j}$. For ease of notation, we will denote the nodes $V_{\ell,i}^{j}$ by $\tilde{V}$. Construct the graph $G_{\tilde{V}}$ as follows. The nodes $G_{\tilde{V}}$ are the union of $\tilde{V}$ and an additional node $v_{\tilde{V}}$ that is not part of $V(n)$. For each $u \in U$, add an undirected edge e between u and $v_{\tilde{V}}$ with edge capacity

$$c(e) \triangleq c(u, v_{\tilde{V}}) \triangleq \min\{1, |\tilde{V}| r_{\tilde{V}}^{-\alpha}\}.$$

This construction is illustrated in Fig. 2.

The construction of $G_{\tilde{V}}$ can intuitively be understood as follows. The node $v_{\tilde{V}}$ represents the collection of nodes $\tilde{V}$. Routing along the edge from $u \in \tilde{V}$ to $v_{\tilde{V}}$ in $G_{\tilde{V}}$ is equivalent to distributing a message from u to all other nodes in $\tilde{V}$ in the wireless network. We will argue that this is possible at a per-node rate of essentially $\min\{1, |\tilde{V}| r_{\tilde{V}}^{-\alpha}\}$, where the $|\tilde{V}|$ term accounts for the multi-user gain and where $r_{\tilde{V}}^{-\alpha}$ accounts for the power loss of communicating over distance $r_{\tilde{V}}$. Routing along the edge from $v_{\tilde{V}}$ to $w \in \tilde{V}$ is equivalent to concentrating a message distributed over the nodes $\tilde{V}$ onto the node w. We will argue that this is again possible at a per-node rate up to essentially $\min\{1, |\tilde{V}| r_{\tilde{V}}^{-\alpha}\}$.

The graph G is then constructed as the union of the graphs $\{G_{\tilde{V}}\}$ for every $\tilde{V} \triangleq V_{\ell,i}^{j}$, $\ell \in \{0, \ldots, L(n)\}$, $i \in \{1, \ldots, 4^{\ell} + 2^{\ell} + 1\}$, $j \in \{1, \ldots, 4\}$. This is illustrated in Fig. 3. Note that, unlike what the appearance of Fig. 3 may suggest, the graph G is actually quite small as well as sparse. First, there are n nodes in $V(n)$ and an additional at most $8n^{\nu}$ nodes of the form $v_{\tilde{V}}$. Thus G has at most

$$|V_G| \leq n + 8n^{\nu} \tag{3}$$

nodes, i.e., V_G has polynomial size in n. Second, each node $u \in V(n)$ is connected in G to at most $4(1 + L(n)) = 4 + 2\nu \log(n)$ nodes. Hence the total number of edges in G is at most

$$|E_G| \leq 4n + 2\nu n \log(n);$$

in other words, G is quite sparse.

B. Unicast Capacity Region

In this section, we consider general unicast traffic, modeled as follows. A *unicast traffic matrix* $\lambda^{\mathsf{UC}} \in \mathbb{R}_{+}^{n \times n}$ associates

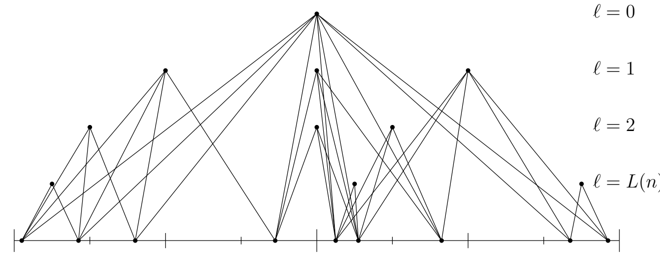

Fig. 3. Construction of the graph G as union of subgraphs $\{G_{\tilde{V}}\}$. For simplicity a one-dimensional network is shown.

with each pair of nodes $(u, w) \in V(n) \times V(n)$ the non-negative rate $\lambda_{u,w}^{\mathsf{UC}}$ at which the source node u wants to transmit a message to the destination node w. There are n ways to choose the source node u and n ways to choose the destination w, and hence the traffic matrix λ^{UC} is a n^2-dimensional matrix. We assume that the messages for distinct (u, w) pairs are independent. However, the same node u can be source for several (independent) messages to be sent to distinct destination nodes $w_1, w_2, \ldots$ at possibly different rates. Similarly, the same node w can be destination for several (independent) messages sent from distinct source nodes $u_1, u_2, \ldots$ at possibly different rates. Furthermore, each node can be both source as well as destination for several (again independent) messages.

The following example illustrates this definition of a unicast traffic matrix.

Example 1. Consider a network with $n = 4$ nodes, $V(n) = \{v_j\}_{j=1}^{4}$. Assume node v_1 has to send a message $m_{1,3}$ to node v_3 at rate 1 bits per channel use. Node v_2 has to send a message $m_{2,1}$ to node v_1 at rate 2 bits per channel use and a message $m_{2,3}$ to node v_3 at rate 3 bits per channel use. Assume all messages are independent. This traffic requirement can be described by the traffic matrix $\lambda^{\mathsf{UC}} \in \mathbb{R}_{+}^{4 \times 4}$ with $\lambda_{v_1,v_3}^{\mathsf{UC}} = 1$, $\lambda_{v_2,v_1}^{\mathsf{UC}} = 2$, $\lambda_{v_2,v_3}^{\mathsf{UC}} = 3$, and $\lambda_{u,w}^{\mathsf{UC}} = 0$ for all other (u, w). Note that node v_2 is source for two messages, node v_3 is destination for two messages, and node v_1 is both source and destination. Node v_4 acts as neither source nor destination in this example, and can be understood as a helper node. $\lozenge$

The *unicast capacity region* $\Lambda^{\mathsf{UC}}(n) \subset \mathbb{R}_{+}^{n \times n}$ of the wireless network $V(n)$ is the closure of the collection of all achievable unicast traffic matrices $\lambda^{\mathsf{UC}} \in \mathbb{R}_{+}^{n \times n}$. In other words, $\lambda^{\mathsf{UC}} \in \Lambda^{\mathsf{UC}}(n)$ if all the n^2 source-destination pairs $(u, w) \in V(n) \times V(n)$ can simultaneously and reliably transmit an independent message from u to w at rate $\lambda_{u,w}^{\mathsf{UC}}$.

For the graph G constructed in Section IV-A, define

$$\hat{\Lambda}_G^{\mathsf{UC}}(n) \triangleq \Bigg\{ \lambda^{\mathsf{UC}} \in \mathbb{R}_{+}^{n \times n} : \sum_{u \in S \cap V(n)} \sum_{w \in V(n) \setminus S} \lambda_{u,w}^{\mathsf{UC}} \leq \sum_{u \in S} \sum_{v \in V_G \setminus S} c(u, v), \quad \forall S \subset V_G \Bigg\}. \tag{4}$$

In words, $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ is the collection of all unicast traffic matrices λ^{UC} on $V(n) \subset V_G$ such that, for every cut S in the graph G, the total traffic

$$\sum_{u \in S \cap V(n)} \sum_{w \in V(n) \setminus S} \lambda_{u,w}^{\mathsf{UC}}$$

across the cut S is not more than the sum of the capacities

$$\sum_{u \in S} \sum_{v \in V_G \setminus S} c(u,v)$$

of edges in G crossing S. Or, put differently, $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ is the collection of unicast traffic matrices that satisfy all the cut-set bounds on the graph G.

The next theorem states that (appropriately scaled) $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ is an inner bound for the unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ of the wireless network.

Theorem 1. *For all $\alpha > 2$, $\nu \in \mathbb{R}_+$, there exists $\hat{b}(n) \geq n^{-o(1)}$ such that for any $n \in \mathbb{N}$, and any node placement $V(n)$ on $A(n)$ with area n^ν,*

$$\hat{b}(n)\hat{\Lambda}_G^{\mathsf{UC}}(n) \subset \Lambda^{\mathsf{UC}}(n).$$

Theorem 1 provides an analytic inner bound to the unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ of the wireless network. Whereas $\Lambda^{\mathsf{UC}}(n)$ is not computable, the region $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ is explicitly defined in (4). However, the set $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ has the disadvantage of being difficult to handle from a computational point of view. In fact, (4) defines $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ through an exponential (in the size of the network n) number of linear inequalities. We next argue that $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ can be efficiently approximated through a linear program. Together with Theorem 1, this yields a less explicit, but computationally more efficient, inner bound on the unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ of the wireless network.

Denote by $\Lambda_G^{\mathsf{UC}}(n) \subset \mathbb{R}_+^{n\times n}$ the unicast capacity region of G for traffic between nodes in $V(n) \subset V_G$ under routing. In other words, $\lambda^{\mathsf{UC}} \in \Lambda_G^{\mathsf{UC}}(n)$ if simultaneously for every node pair $(u,w) \in V(n) \times V(n)$ we can route data from u to w over G at rate $\lambda_{u,w}^{\mathsf{UC}}$. The capacity region $\Lambda_G^{\mathsf{UC}}(n)$ of G under routing can be described by a linear program of polynomial size in $|V_G|$ and therefore is readily computed, as is described next.

For a unicast traffic matrix $\lambda^{\mathsf{UC}} \in \mathbb{R}_+^{n\times n}$, define

$$\phi_{\lambda^{\mathsf{UC}}}^\star(n) \triangleq \max\{\phi : \phi\lambda^{\mathsf{UC}} \in \Lambda_G^{\mathsf{UC}}(n)\}.$$

In words, $\phi_{\lambda^{\mathsf{UC}}}^\star(n)$ is the largest multiplier ϕ such that the scaled traffic matrix $\phi\lambda^{\mathsf{UC}}$ can be routed over G. Clearly, $\lambda^{\mathsf{UC}} \in \Lambda_G^{\mathsf{UC}}(n)$ if and only if $\phi_{\lambda^{\mathsf{UC}}}^\star(n) \geq 1$. Moreover, since the capacity region $\hat{\Lambda}_G^{\mathsf{UC}}(n)$ is convex, knowledge of $\phi_{\lambda^{\mathsf{UC}}}^\star(n)$ for all λ^{UC} completely characterizes $\Lambda_G^{\mathsf{UC}}(n)$.

$\phi_{\lambda^{\mathsf{UC}}}^\star(n)$ can be computed as the solution to the following linear program:

$$\begin{array}{lrcll}
\max & \phi & & & \\
\text{s.t.} & \displaystyle\sum_{p \in P_{u,w}} f_p & \geq & \phi\lambda_{u,w}^{\mathsf{UC}}, & \forall\, u,w \\
& \displaystyle\sum_{u \in V(n)} \sum_{w \in V(n)} \sum_{\substack{p \in P_{u,w}: \\ e \in p}} f_p & \leq & c(e), & \forall\, e \in E_G \\
& f_p & \geq & 0, & \forall\, u,w,p \in P_{u,w}
\end{array} \tag{5}$$

where $P_{u,w}$ is the collection of all paths from u to w in G, and where the maximization is over ϕ and the flow variables $\{f_p\}$ between nodes u and w over path $p \in P_{u,w}$. The linear program (5) solves for the maximal ϕ such that the total flow for each source-destination pair (u,w) is at least a multiple ϕ of the traffic demand $\lambda_{u,w}^{\mathsf{UC}}$, and such that the total flow over each edge e of G is at most the capacity $c(e)$ of that edge.

While the number of paths $P_{u,w}$ from u to w in the graph G could be exponential in the number of nodes $|V_G|$ of G, (5) can nevertheless be solved in polynomial time in $|V_G|$. This can be seen by reformulating it in terms of per-edge flow variables. By (3), G has only polynomially many nodes in n, and hence the linear program (5) can also be solved in polynomial time in the size of the wireless network n. Therefore, the region $\Lambda_G^{\mathsf{UC}}(n)$ can be efficiently characterized through its equivalent description $\phi_{\lambda^{\mathsf{UC}}}^\star(n)$.

The next theorem states that the unicast capacity region $\Lambda_G^{\mathsf{UC}}(n)$ of G under routing yields an inner bound to the unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ of the wireless network.

Theorem 2. *For all $\alpha > 2, \nu \in \mathbb{R}_+$, there exists $b(n) \geq n^{-o(1)}$ such that for any $n \in \mathbb{N}$, and any node placement $V(n)$ on $A(n)$ with area n^ν,*

$$b(n)\Lambda_G^{\mathsf{UC}}(n) \subset \Lambda^{\mathsf{UC}}(n).$$

Theorem 2 shows that if traffic can be routed over the graph G, then approximately the same traffic can be transmitted reliably over the wireless network. Given the description of $\Lambda_G^{\mathsf{UC}}(n)$ through the linear program (5), this provides an efficiently computable inner bound to the unicast capacity region $\Lambda^{\mathsf{UC}}(n)$ of the wireless network.

Moreover, Theorem 2 suggests a communication architecture for heterogeneous wireless networks. Given a wireless network, construct its graph G. This construction of G handles the heterogeneity of the node placement. Given a traffic matrix λ^{UC} for the wireless network, find the optimal routing over G by solving the linear program (5). By Theorem 2 and the construction of G as described in Section IV-A, the optimal routes found for G can be used to transmit data over the wireless network. This process of optimal routing handles the heterogeneity of the traffic pattern.

Using an approximate max-flow min-cut result for unicast traffic on undirected capacitated graphs [18], it can be shown that

$$\Lambda_G^{\mathsf{UC}}(n) \subset \hat{\Lambda}_G^{\mathsf{UC}}(n) \subset O(\log(n))\Lambda_G^{\mathsf{UC}}(n).$$

Hence, the inner bounds in Theorems 1 and 2 are equivalent for scaling purposes.

While Theorems 1 and 2 provide only inner bounds on the capacity region for general heterogeneous wireless networks, matching (in the scaling sense) outer bounds can be found for several important special cases. In particular, this is the case for dense wireless networks with arbitrary node placement and traffic pattern, for random node placement with random source-destination pairing on a square of arbitrary area, and random node placement with arbitrary "symmetric" traffic among others. Prior order-optimal communication schemes proposed for large wireless networks under certain homogeneity assumptions (as opposed to the general heterogeneous wireless networks considered in this paper) can be understood as a specific routing strategy over G. The proposed communication architecture thus unifies these various communication strategies. Moreover, in many cases with heterogeneous node placement or traffic demands, the additional degrees of freedom offered by allowing arbitrary routing over G is crucial for optimal operation of the wireless network. In fact, it is easy to construct scenarios with heterogeneous node placements or traffic patterns, similar to the Example in Section II, where the proposed scheme is optimal in the scaling sense and neither of the previously known schemes achieves the correct scaling exponent.

C. Multicast Capacity Region

In this section, we consider general multicast traffic. A *multicast traffic matrix* $\lambda^{\mathsf{MC}} \in \mathbb{R}_+^{n\times 2^n}$ associates with each pair of source node $u \in V(n)$ and destination group $W \subset V(n)$ the nonnegative rate $\lambda^{\mathsf{MC}}_{u,W}$ at which the source node u wants to transmit the *same* message to *all* the destination nodes $w \in W$. There are n ways to choose the source node u and 2^n ways to choose the destination group W, and hence λ^{MC} is a $n \times 2^n$-dimensional matrix. We assume that the messages for distinct (u, W) pairs are independent. However, the same node u can be source for several (independent) messages to be sent to distinct destination groups $W_1, W_2, \dots$ at possibly different rates. Similarly, the same subset of nodes W can be destination group for several (independent) messages sent from distinct source nodes $u_1, u_2, \dots$ at possibly different rates.

The following example illustrates this definition of a multicast traffic matrix.

Example 2. Consider again a network with $n = 4$ nodes $V(n) = \{v_j\}_{j=1}^4$. Assume node v_1 has to send a message $m_{1,\{2,3\}}$ to nodes $\{v_2, v_3\}$ at rate 1 bit per channel use and a message $m_{1,\{2,3,4\}}$ to nodes $\{v_2, v_3, v_4\}$ at rate 2 bits per channel use. Node v_2 has to send a message $m_{2,\{1\}}$ to just node v_1 at rate 3 bits per channel use. And node v_4 has to send a message $m_{4,\{2,3\}}$ to nodes $\{v_2, v_3\}$ at rate 4 bits per channel use. Assume all messages are independent. This traffic requirement can be described by the multicast traffic matrix $\lambda^{\mathsf{MC}} \in \mathbb{R}_+^{4\times 2^4}$ with $\lambda^{\mathsf{MC}}_{v_1,\{v_2,v_3\}} = 1$, $\lambda^{\mathsf{MC}}_{v_1,\{v_2,v_3,v_4\}} = 2$, $\lambda^{\mathsf{MC}}_{v_2,\{v_1\}} = 3$, $\lambda^{\mathsf{MC}}_{v_4,\{v_2,v_3\}} = 4$, and $\lambda^{\mathsf{MC}}_{u,W} = 0$ for all other (u, W). Note that node v_1 is sources for two messages, subset $\{v_2, v_3\}$ is destination for two messages. The message $m_{1,\{2,3,4\}}$ is to be sent from node v_1 to all other nodes and can hence be understood as a broadcast message. Moreover, the message $m_{2,\{1\}}$ can be understood as a unicast message; thus unicast traffic can be recovered as a special case of multicast traffic. ◊

Similar to the unicast case, the *multicast capacity region* $\Lambda^{\mathsf{MC}}(n) \subset \mathbb{R}_+^{n\times 2^n}$ of the wireless network $V(n)$ is defined as the closure of the collection of all achievable multicast traffic matrices $\lambda^{\mathsf{MC}} \in \mathbb{R}_+^{n\times 2^n}$. In other words, $\lambda^{\mathsf{MC}} \in \Lambda^{\mathsf{MC}}(n)$ if all the $n \times 2^n$ source–destination-group pairs (u, W) can simultaneously reliably transmit an independent message from u to W at rate $\lambda^{\mathsf{MC}}_{u,w}$.

In analogy to the unicast case, define

$$\hat{\Lambda}^{\mathsf{MC}}_G(n) \triangleq \Big\{\lambda^{\mathsf{MC}} \in \mathbb{R}_+^{n\times 2^n} : \sum_{u\in S\cap V(n)} \sum_{\substack{W\subset V(n):\\ W\setminus S\neq\emptyset}} \lambda^{\mathsf{MC}}_{u,W} \leq \sum_{u\in S}\sum_{v\in V_G\setminus S} c(u,v), \quad \forall S \subset V_G\Big\}. \tag{6}$$

In words, $\hat{\Lambda}^{\mathsf{MC}}_G(n)$ is the collection of all multicast traffic matrices λ^{MC} on $V(n) \subset V_G$ such that for every cut S in the the graph G the total traffic

$$\sum_{u\in S\cap V(n)} \sum_{\substack{W\subset V(n):\\ W\setminus S\neq\emptyset}} \lambda^{\mathsf{MC}}_{u,W}$$

across the cut S is not more than the sum of capacities of the edges across S. Note that, unlike the unicast case, we count traffic $\lambda^{\mathsf{MC}}_{u,W}$ to be crossing the cut S if the source u is in S and there exists at least one destination $w \in W$ that is not in S. The next theorem states that (appropriately scaled) $\hat{\Lambda}^{\mathsf{MC}}_G(n)$ is an inner bound for the multicast capacity region $\Lambda^{\mathsf{MC}}(n)$ of the wireless network.

Theorem 3. *For all $\alpha > 2$, $\nu \in \mathbb{R}_+$, there exists $\hat{b}(n) \geq n^{-o(1)}$ such that for any $n \in \mathbb{N}$, and any node placement $V(n)$ on $A(n)$ with area n^ν,*

$$\hat{b}(n)\hat{\Lambda}^{\mathsf{MC}}_G(n) \subset \Lambda^{\mathsf{MC}}(n).$$

Theorem 3 provides an analytic inner bound to the multicast capacity region $\Lambda^{\mathsf{MC}}(n)$ of the wireless network, with the inner bound $\hat{\Lambda}^{\mathsf{MC}}_G(n)$ explicitly defined in (6). However, as in the unicast case, $\hat{\Lambda}^{\mathsf{MC}}_G(n)$ is difficult to evaluate computationally. We next provide an alternative inner bound that is computationally more manageable.

Similar to the unicast case, denote by $\Lambda^{\mathsf{MC}}_G(n) \subset \mathbb{R}_+^{n\times 2^n}$ the multicast capacity region of G for traffic between nodes in $V(n) \subset V_G$ under routing. In other words, $\lambda^{\mathsf{MC}} \in \Lambda^{\mathsf{MC}}_G(n)$ if simultaneously for every source $u \in V(n)$ and destination-group $W \subset V(n)$ we can route data from u to every node $w \in W$ over G at rate $\lambda^{\mathsf{MC}}_{u,W}$.

The next theorem states that the multicast capacity region $\Lambda^{\mathsf{MC}}_G(n)$ of G under routing yields an inner bound to the multicast capacity region $\Lambda^{\mathsf{MC}}(n)$ of the wireless network.

Theorem 4. *For all $\alpha > 2, \nu \in \mathbb{R}_+$, there exists $b(n) \geq n^{-o(1)}$ such that for any $n \in \mathbb{N}$, and any node placement $V(n)$ on $A(n)$ with area n^ν,*

$$b(n)\Lambda_G^{\mathsf{MC}}(n) \subset \Lambda^{\mathsf{MC}}(n).$$

While the unicast capacity region $\Lambda_G^{\mathsf{UC}}(n)$ of G can be conveniently described by a linear program, the same is not true for the multicast capacity region $\Lambda_G^{\mathsf{MC}}(n)$. However, using a result by Räcke [19], $\Lambda_G^{\mathsf{MC}}(n)$ can be efficiently approximated. In fact, [19] shows that any graph G can be approximated by a collection of tree graphs $\{T_i\}$, called *decomposition trees*. Each leaf node in T_i corresponds to a node in G, and each internal node in T_i corresponds to a subset S of nodes in G. Each edge e in T_i corresponds to a cut in the graph G, and the capacity of e in T_i is equal to the capacity of the cut in G.

This decomposition of G into trees $\{T_i\}$ can be found in polynomial time in $|V_G|$ and hence, by (3), also in n. This algorithm also produces nonnegative weights $\{\omega_i\}$ summing to one and corresponding to the decomposition trees $\{T_i\}$. Approximately optimal routes for multicast traffic in G can then be computed as follows. Find the optimal routes for the multicast traffic λ^{MC} for each tree T_i. Since T_i is a tree, this is trivial. The routes in T_i for λ^{MC} induce routes in G for λ^{MC}. We time share between these routes computed according to T_i with weight ω_i, i.e., for a fraction ω_i of time we route according to T_i. [19] shows that this procedure is close to optimal for the graph G up to a factor $O(\log(n))$.

Since this algorithm yields approximately optimal routes for multicast traffic in G, it can be used to approximately evaluate the region $\Lambda_G^{\mathsf{MC}}(n)$. In other words, by loosing an additional factor of order $\log(n)$, the inner bound $\Lambda_G^{\mathsf{MC}}(n)$ in Theorem 4 can be efficiently evaluated. Moreover, as in the unicast case, this procedure suggests a communication architecture for multicast traffic in heterogeneous wireless networks. As in the unicast case, the construction of G handles heterogeneity in the node placement, and optimal routing handles heterogeneity in the traffic demands. In addition to the procedure in the unicast case, the decomposition of the graph G into trees $\{T_i\}$ is used to handle the "heterogeneity" of having to deal with multicast traffic.

The result by Räcke also shows that

$$\Lambda_G^{\mathsf{MC}}(n) \subset \hat{\Lambda}_G^{\mathsf{MC}}(n) \subset O(\log(n))\Lambda_G^{\mathsf{MC}}(n).$$

Thus, for multicast traffic over undirected graphs, an approximate max-flow min-cut result holds. This implies that the inner bounds in Theorems 3 and 4 are equivalent for scaling purposes.

As in the unicast case, the inner bounds given by Theorems 3 and 4 can be shown to be tight in the scaling sense for several important special cases. As before, the flexibility of choosing various levels of cooperation in different areas of the network can be shown to be necessary under heterogeneous node placements and traffic patterns.

V. Conclusions and Future Work

In this paper, we have studied the problem of quantifying performance in heterogeneous wireless networks, where both the node placements and user traffic demands can be highly non-uniform. We have developed a general communication architecture to overcome such heterogeneities. The proposed scheme deploys a suitable combination of multihopping and cooperative communication at different scales to efficiently operate a wireless network under non-uniform node placement and traffic demands. We have established that the proposed scheme performs as well as any existing communication scheme for large wireless networks—multihopping, multi-user cooperative communication, cooperative multi-hopping, among others. Further, for scenarios with heterogeneous node placements or traffic patterns, the proposed scheme can perform strictly better (even in terms of the scaling exponent) than the known schemes.

While the inner bound on the capacity region derived in this paper can be shown to be tight in the scaling sense in several important special cases, it is not clear whether this is the case in general. Deriving tight outer bounds on the scaling of the capacity region under both node placement and traffic demand non-uniformities is hence of interest. The currently known techniques, such as those used in [8] and [15], work well under homogeneous node placement, but have limited applicability to the arbitrary setting.

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[2] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 748–767, May 2004.

[3] A. Jovičić, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2555–2565, Nov. 2004.

[4] F. Xue, L.-L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 834–847, Mar. 2005.

[5] O. Lévêque and İ. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 858–865, Mar. 2005.

[6] S. H. A. Ahmad, A. Jovičić, and P. Viswanath, "On outer bounds to the capacity region of wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2770–2776, June 2006.

[7] L.-L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of the transport capacity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2313–2328, June 2006.

[8] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3549–3572, Oct. 2007.

[9] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: Information theoretic and physical limits," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3413–3424, Aug. 2009.

[10] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

[11] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Sept. 2005.

[12] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1348–1358, Apr. 2005.

[13] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2044–2059, June 2007.

[14] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Trans. Inf. Theory*, vol. 56, pp. 3959–3982, Sept. 2009.

[15] U. Niesen, P. Gupta, and D. Shah, "The balanced unicast and multicast capacity regions of large wireless networks," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2249–2271, May 2010.
[16] S.-H. Lee and S.-Y. Chung, "On the capacity scaling of wireless networks: Effect of finite wavelength," in *Proc. IEEE ISIT*, June 2010.
[17] A. Özgür, O. Lévêque, and D. N. C. Tse, "Linear capacity scaling in wireless networks: Beyond physical limits?," in *Proc. ITA*, Feb. 2010.
[18] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, pp. 215–245, June 1995.
[19] H. Räcke, "Optimal hierarchical decompositions for congestion minimization in networks," *Proc. ACM STOC*, pp. 255–263, May 2008.