

Throughput and Delay in Cooperative Wireless Networks With Partial Infrastructure

Xiaoyan Zhu, Pan Li, *Student Member, IEEE*, Yuguang Fang, *Fellow, IEEE*, and Yumin Wang, *Senior Member, IEEE*

Abstract—Hybrid wireless networks have been shown to scale only when the number of base stations m is on the same order of the number of nodes n in the network. Pure wireless ad hoc networks with multiple-input–multiple-output (MIMO) distributed communications have also been shown to scale. A natural question arises: Do hybrid wireless networks always scale no matter how many base stations are placed if MIMO distributed communications are used? In this paper, we address this question and show that hybrid wireless networks cannot always scale with the proposed resource-allocation strategy. Specifically, they can only scale either when $m = \Theta(1)$ and all resources are exclusively allocated to ad hoc mode transmissions or when $m = \Theta(n)$, and all resources are exclusively allocated to infrastructure mode transmissions; moreover, the delay is significantly large in the first case, whereas it is $\Theta(1)$ in the second case, where $\Theta(f(n))$ means on the order of $f(n)$ when n is large.

Index Terms—Delay, hybrid wireless networks, scaling law, throughput capacity, wireless ad hoc networks.

I. INTRODUCTION

WIRELESS networks have attracted intensive attention because they can easily and quickly be deployed at low cost whenever needed. In a wireless network, the throughput and delay are important performance index. In their seminal paper [9], Gupta and Kumar show that the per-node throughput capacity in random wireless ad hoc networks is $\Theta(1/\sqrt{n} \log n)$ bits per second, which means random ad hoc networks cannot scale (the network throughput scales with respect to the number of nodes if it is proportional to the number of nodes in the network).

The work in [9] deals with *dense* networks, where the area is fixed while the density of nodes increases under the assumption that the whole network is connected. Dousse *et al.* [4] study the throughput capacity in *extended* networks, where

the density of nodes is fixed while the area linearly increases with the number of nodes n . They show that, by allowing an arbitrary small fraction of the nodes to be disconnected in 2-D extended networks, the per-node throughput can be $O(1/\sqrt{n})$ bits per second. In addition, Buragohain *et al.* [2] investigate the throughput capacity in extended grid networks, where there are n nodes and the average source–destination distance is d . They show that the $\Omega(n/d)$ throughput can be achieved. Thus, grid networks cannot scale either since $d = \omega(1)$ for most of the cases. Moreover, in [5], Duarte-Melo *et al.* examine the case of *semiextended* networks, where both the node density and the network area increase as the number of nodes n increases. Specifically, they assume that the network area is a disk of radius n^γ , $0 < \gamma < 1/2$. With a $1/(1+d)^\alpha$ propagation model, they show that the per-node throughput capacity is $\Omega(1/n^{1-\gamma})$, i.e., semiextended networks cannot scale.

We can clearly observe that pure wireless ad hoc networks cannot scale using the simple store-and-forward relaying scheme. In other words, pure wireless ad hoc networks cannot provide satisfactory service when the number of network users is large. In this paper, we propose to use hybrid wireless networks, which are also called multihop cellular networks [17], to improve network performance. Hybrid wireless networks can be 1-D, 2-D with a strip area [16], or 2-D with a square area [15]. The traffic pattern in the network can be asymmetric [22] or symmetric [11], [20], [24]. In this paper, we only focus on 2-D square hybrid wireless networks with symmetric traffic.

The capacity of hybrid wireless networks has been investigated in the literature. In [11], Kozat and Tassiulas study the throughput capacity of hybrid wireless networks where both ad hoc nodes and base stations are randomly distributed. They show that the per-node throughput capacity can be $\Theta(1/\log n)$ bits per second if the number of base stations linearly scales with the number of nodes, which means the network cannot scale. Similar results are also reported in [1]. Zemlianov and Veciana investigate in [24] the throughput capacity of hybrid wireless networks where ad hoc nodes are randomly distributed and base stations are arbitrarily placed. They show that the per-node throughput capacity depends on the number of base stations, but the network still cannot scale. Assuming n nodes are randomly distributed and m base stations are regularly placed, Liu *et al.* study the throughput capacity of hybrid wireless networks in [15]. They consider two different routing strategies. Under the k -nearest-cell routing strategy, if m asymptotically grows slower than \sqrt{n} , the maximum per-node throughput capacity is $\Theta(\sqrt{1/(n \log n/m^2)})$, and the benefit of adding base stations is insignificant. However, if m asymptotically grows faster than \sqrt{n} , the maximum per-node

Manuscript received February 15, 2009; revised April 22, 2009. First published May 8, 2009; current version published October 2, 2009. The work of X. Zhu, Y. Fang, and Y. Wang was supported in part by the 111 Project under Grant B08038 with Xidian University, Xi'an, China. The work of P. Li and Y. Fang was supported in part by the U.S. National Science Foundation under Grant CNS-0721744 and Grant DBI-0529012. The work of X. Zhu was supported in part by the National Natural Science Foundation of China under Grant 60772136 and the 863 Project of China under Grant 2007AA01Z435. The review of this paper was coordinated by Dr. L. Cai.

X. Zhu and Y. Wang are with the National Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: xyzhu@mail.xidian.edu.cn; ymwang@xidian.edu.cn).

P. Li is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: lipanleo@ufl.edu).

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA, and also with the National Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: fang@ece.ufl.edu).

Digital Object Identifier 10.1109/TVT.2009.2022533

throughput capacity is $\Theta(m/n)$, which linearly increases with the number of base stations. Under the probabilistic routing strategy, similar results are obtained. The threshold of the number of base stations above which the per-node throughput capacity linearly increases with m is $\sqrt{n/\log n}$. Thus, the network can scale if $m = \Omega(n)$.

In [15], the authors assume that $k = 0$, i.e., a node transmits to its destination in the ad hoc mode only if it is in the same cell. However, the 0-nearest-cell routing strategy cannot efficiently make use of the wireless channel. For example, if a source node and its destination are within one-hop distance of each other but are not in the same cell, then they cannot directly communicate in the ad hoc mode according to the routing strategy. Instead, the transmissions between these two nodes can only be carried out through base stations. Moreover, the case $k = 0$ is too specific. Li *et al.* [14] analyze the throughput by employing a more efficient resource-allocation strategy and show that the result in [15] is just a special case in their analysis. They can provide higher throughput using the same number of base stations when $m = o(n)$. In addition, they show that hybrid wireless networks can scale only when $m = \Omega(n)$. Li and Fang [12] investigate the throughput capacity in hybrid wireless networks and show that both the network topology and the traffic pattern have significant impacts on the network capacity. When a directional antenna is deployed, Li *et al.* [13] have also given some characterization on asymptotic connectivity, which is a closely related property to the scaling law.

Other than improving the throughput by placing some infrastructure, Ozgur *et al.* [18] study the throughput capacity of a connected ad hoc network with intelligent node cooperation and distributed multiple-input–multiple-output (MIMO) communication. Their results show that the dense networks can linearly scale with the number of nodes n and that the extended networks can also scale if there is no power limit. Thus, a natural question arises: How can hybrid wireless networks scale if distributed MIMO cooperative communication is used?

In this paper, we investigate the throughput and delay in hybrid wireless networks with distributed cooperative MIMO communication. We allow a source–destination pair to use the ad hoc mode transmission only when these two nodes are in the same cell. Using this scheme, we show that, even with distributed MIMO communication, hybrid wireless networks can only scale either when $m = \Theta(1)$ and all resources are exclusively allocated to ad hoc mode transmissions or when $m = \Theta(n)$ and all resources are exclusively allocated to infrastructure mode transmissions; moreover, the delay is very large in the first case, whereas it is $\Theta(1)$ in the second case.

The rest of this paper is organized as follows: In Section II, we introduce some definitions and models used in this paper. Section III gives an overview of the hierarchical cooperative MIMO scheme proposed in [19]. In Section IV, we derive the network throughput and the average delay in hybrid wireless networks, respectively. We conclude this paper in Section V.

II. DEFINITIONS AND MODELS

In this section, we give the definitions and models used in this paper.

Notations: We use the following classical notations [10].

- 1) $f(n) = o(g(n))$ means $f(n)$ is asymptotically negligible with respect to $g(n)$, i.e., $\lim_{n \rightarrow \infty} |f(n)/g(n)| = 0$.
- 2) $f(n) = O(g(n))$ means $f(n)$ is asymptotically upper bounded by $g(n)$, i.e., $\limsup_{n \rightarrow \infty} |f(n)/g(n)| < \infty$.
- 3) $f(n) = \Omega(g(n))$ means $f(n)$ is asymptotically lower bounded by $g(n)$, i.e., $\liminf_{n \rightarrow \infty} |f(n)/g(n)| > 0$.
- 4) $f(n) = \Theta(g(n))$ means $f(n)$ is asymptotically tight bounded by $g(n)$, i.e., $0 < \liminf_{n \rightarrow \infty} |f(n)/g(n)| \leq \limsup_{n \rightarrow \infty} |f(n)/g(n)| < \infty$.
- 5) $f(n) = \omega(g(n))$ means $f(n)$ is asymptotically dominant with respect to $g(n)$, i.e., $\lim_{n \rightarrow \infty} |f(n)/g(n)| = \infty$.

1) *Throughput:* As defined in the usual way, throughput is the time average of the number of bits per second that can be transmitted by each node to its destination, which is called the *per-node throughput*. The sum of per-node throughput over all the nodes in a network is called the *throughput of the network* or *network throughput*.

2) *Average Delay:* As defined in [7], the *delay* in a network is the time it takes a bit or a packet of constant size to reach the destination after it leaves the source, where averaging is taken over all the bits or packets. Note that the delay here is different from that defined in [6] and [14], where the packet size is assumed to scale as the per-node throughput. We also note that this definition does not consider the queuing delay at the source node, which largely depends on the packet arrival process.

3) *Network Model:* Similar to [14], we consider a two-tier hybrid wireless network on the surface of a torus of unit area. The low tier is composed of n normal nodes operating in the ad hoc mode, and the high tier consists of m base stations operating in the infrastructure mode. We assume that n nodes are uniformly and independently distributed in the unit area. They have the same transmission power and, hence, the same transmission range, which is denoted by $r(n)$. We follow the process in [8] to choose random sender–receiver pairs so that each node is a source node for one flow and a destination node for at most $O(1)$ flows. We also assume that the m base stations are regularly placed in the network, dividing the area into a hexagonal tessellation, which is exactly the classical seven-cell reuse model, as described in [21]. Each hexagon is called a cell, and there is one base station in the center of each cell. Base stations do not serve as data sources or data destinations. Instead, they only help relay the packets for the normal nodes. Furthermore, we also assume that base stations are interconnected by a wired network in which the link bandwidth is abundant. Thus, the wired network has no bandwidth constraints.

4) *Interference Model:* We employ the protocol model in [9] as the interference model. Suppose node X_i transmits to another node X_j . X_i and X_j also denote the positions of these two nodes. Then, the transmission is successful if the positions of other transmitters X_k simultaneously transmitting over the same channel satisfy

$$|X_k - X_j| \geq (1 + \Delta)|X_i - X_j|.$$

The quantity $\Delta > 0$ models situations where a guard zone is specified by the protocol to prevent a neighboring node from transmitting on the same channel at the same time. It also allows for imprecision in the achieved range of transmissions.

III. HIERARCHICAL COOPERATION SCHEME

In this section, we present a brief overview of the one-layer hierarchical cooperation scheme introduced in [19], followed by an extension to the multilayer hierarchical cooperation scheme, and then analyze the delay under this cooperation scheme. Notice that it has been shown in [9] that the protocol model is in fact equivalent to the physical model when the path-loss exponent is greater than 2. Thus, the following result can also be obtained when we employ the protocol model.

Assume that there are n nodes randomly and uniformly distributed in a torus of unit area. A single channel is used, and the channel capacity is normalized to 1 bit/s. The network is first evenly divided into squares of M nodes, the value of which will be determined later. Distributed MIMO transmissions are carried out according to a three-phase scheme, as follows.

1) *Phase 1—Setting Up Transmit Cooperation*: Assume that there are M nodes in each cluster (corresponding to a square here) and that each node has M bits to transmit. When applying the *network coding theory* [23], these data can be represented by an M -dimensional column vector $m \in F^M$, where $F = GF(2)$ is a base field for a bit. Considering one source node, we can denote the M bits data by an $M \times 1$ matrix, i.e.,

$$m = [m_1, m_2, m_3, \dots, m_M]^T$$

where T indicates the matrix transpose operation. Assume that the M nodes in the same cluster are numbered from 1 to M . The source node formulates a linear network code for the j th node ($1 \leq j \leq M$) by applying the following mapping matrix, which is denoted by $f_j \in [\cdot]_{1 \times M}$, to m :

$$f_j(p) = \begin{cases} 1, & \text{if } p = j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Then, the source node transmits the message $m^j = f_j \times m$ to the j th node, one after another, respectively. Thus, at the end of the phase, each node has 1 bit from each of the source nodes in the same cluster.

Since there are M source nodes in each cluster, this gives a traffic demand of exchanging $M(M-1) \sim M^2$ bits. Under a 9-TDMA scheme (in which each node transmits once in every nine time slots) [19], clusters can work in parallel, and the M^2 bits can be exchanged in at most M^2 time slots.

2) *Phase 2—MIMO Transmissions*: For any source node, the M nodes in the same cluster transmit, at the same time, the 1 bit they received from the same source node in Phase 1 to the M nodes in the cluster in which the destination node is located, forming a long-distance distributed MIMO transmission. In this MIMO transmission, the M nodes in the source square act as M transmitting antennas, whereas the M nodes in the destination square act as the M receiving antennas. Thus, the transmission of M bits can be completed in one time slot.

Since MIMO transmissions are successively performed for each source node, we need n time slots to finish Phase 2.

3) *Phase 3—Cooperate to Decode*: Since there are M nodes inside each destination cluster, each cluster receives M MIMO transmissions in Phase 2: one intended for each of the destination nodes. Thus, each node in the cluster has received M observations, one from each of the MIMO transmissions, and each observation is to be conveyed to a different node in its cluster. Nodes quantize each observation into fixed Q bits; thus, there are now a total of QM^2 bits to exchange inside each cluster. Using exactly the same scheme as in Phase 1, we can complete Phase 3 in at most QM^2 time slots.

This can be considered as a one-layer hierarchical cooperation scheme. The network throughput under this scheme, which is denoted by $T_{\text{mimo}}^1(n)$, can be calculated as follows [19]:

$$T_{\text{mimo}}^1(n) = \frac{nM}{M^2 + n + QM^2}. \quad (2)$$

By choosing $M = \sqrt{n}$, we can obtain $T_{\text{mimo}}^1(n) = \Theta(\sqrt{n})$. In addition, the average delay, which is denoted by $D_{\text{mimo}}^1(n)$, can easily be derived as

$$D_{\text{mimo}}^1(n) = \Theta(M^2 + n + QM^2) = \Theta(M^2) = \Theta(n).$$

Notice that, in Phase 1 and Phase 3, the bits are transmitted as usual directly from one node to another. By further dividing each square into smaller squares and carrying out the transmissions in Phase 1 and Phase 3 according to the previously described three-phase scheme, we can have a two-layer hierarchical cooperation scheme. Let M_1 and M_2 be the number of nodes in the first- and second-layer squares, respectively. Similar to that in (2), the network throughput with the two-layer cooperation scheme, which is denoted by $T_{\text{mimo}}^2(n)$, can be obtained as follows:

$$\frac{nM_1M_2}{(M_2^2 + M_1 + QM_2^2)M_1 + nM_2 + (M_2^2 + M_1 + QM_2^2)QM_1}.$$

By choosing $M_1 = n^{2/3}$ and $M_2 = n^{1/3}$, we have $T_{\text{mimo}}^2(n) = \Theta(n^{2/3})$. The average delay in this case, which is denoted by $D_{\text{mimo}}^2(n)$, is given by

$$\begin{aligned} D_{\text{mimo}}^2(n) &= \Theta((M_2^2 + M_1 + QM_2^2)M_1 + nM_2 \\ &\quad + (M_2^2 + M_1 + QM_2^2)QM_1) \\ &= \Theta(M_1M_2^2) \\ &= \Theta(n^{4/3}). \end{aligned}$$

Following the foregoing process, and iterating the scheme k times, we can finally arrive at a k -layer hierarchical cooperation scheme, as shown in Fig. 1. Let M_1, M_2, \dots, M_k denote the numbers of nodes in layer-1, layer-2, \dots , layer- k squares. The resulting network throughput, which is denoted by $T_{\text{mimo}}^k(n)$, is

$$T_{\text{mimo}}^k(n) = \Theta\left(n^{\frac{k}{k+1}}\right) \quad (3)$$

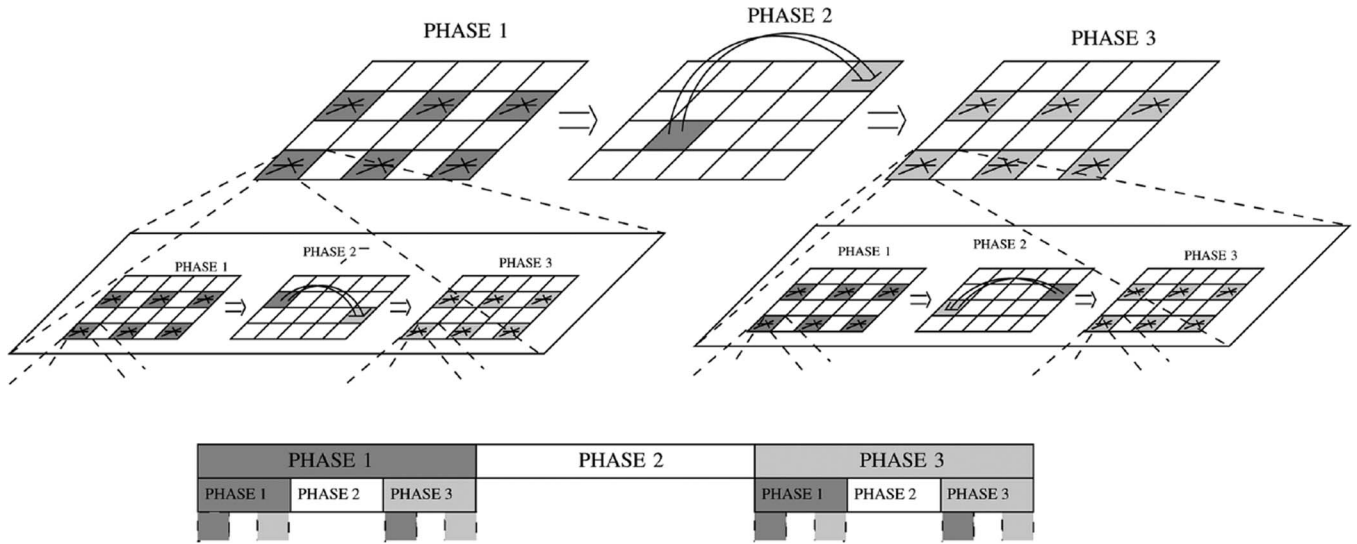


Fig. 1. Hierarchical cooperation scheme proposed in [19]. This figure is also taken from [19].

by choosing $M_j = n^{k-j+1/k+1}$, where $1 \leq j \leq k$. The average delay, which is denoted by $D_{\text{mimo}}^k(n)$, is

$$\begin{aligned}
 D_{\text{mimo}}^k(n) &= \Theta(M_1 M_2, \dots, M_k^2) \\
 &= \Theta\left(n^{\frac{k}{k+1}} n^{\frac{k-1}{k+1}}, \dots, n^{\frac{2}{k+1}} n^{\frac{1}{k+1}} n^{\frac{1}{k+1}}\right) \\
 &= \Theta\left(n^{\frac{k^2+k+2}{2(k+1)}}\right). \tag{4}
 \end{aligned}$$

IV. THROUGHPUT AND DELAY IN HYBRID WIRELESS NETWORKS

We assume a total bandwidth of W bits per second, which is split into three bands, i.e., W_1 for the ad hoc mode, W_2 for uplink for the infrastructure mode, and W_3 for downlink for the infrastructure mode, respectively. Since the uplink has the same amount of traffic as the downlink, we have $W_2 = W_3$. Thus, we have $W = W_1 + 2W_2$. We also assume that there is no interference between these three types of traffics. Thus, the throughput of the network with n nodes and m base stations, which is denoted by $T(n, m)$, can be represented as

$$T(n, m) = T_a(n, m) + T_i(n, m) \tag{5}$$

where $T_a(n, m)$ and $T_i(n, m)$ denote the throughput contributed by the ad hoc and infrastructure mode transmissions, respectively.

Moreover, let n_1 denote the number of nodes in the ad hoc mode. Let $D_a(n, m)$ and $D_i(n, m)$ denote the average delay in the ad hoc and infrastructure modes, respectively. Thus, the average delay of the network, which is denoted by $D(n, m)$, can be calculated as

$$D(n, m) = \frac{n_1 D_a(n, m) + (n - n_1) D_i(n, m)}{n}. \tag{6}$$

In this paper, we allocate the resources as follows: If a source node and its destination node are located in the same cell, the

transmission between them is carried out in the ad hoc mode. Otherwise, the transmission is performed in the infrastructure mode. In this paper, we investigate two cases for our study: 1) when the number of base stations is significantly smaller than the number of nodes and 2) when the number of base stations is significantly more than the first case.

1) *Case I: $m = o(\sqrt{n})$* : Recall that m base stations are regularly distributed in the network. Let N_i ($1 \leq j \leq n$) be a random variable, which is defined as follows:

$$N_i = \begin{cases} 1, & \text{source node } i \text{ transmits to its destination} \\ & \text{node in the ad hoc mode} \\ 0, & \text{otherwise.} \end{cases}$$

Let N_T denote the total number of source nodes transmitting in the ad hoc mode, i.e., $N_T = \sum_{i=1}^n N_i$. Thus, the expected number of source nodes in the ad hoc mode is

$$E(N_T) = E\left(\sum_{i=1}^n N_i\right) = \sum_{i=1}^n E(N_i).$$

Since $P(N_i = 1) = 1/m^2$, we have $E(N_i) = 1 \cdot (1/m^2) = 1/m^2$. Thus

$$E(N_T) = n \cdot \frac{1}{m^2} = \frac{n}{m^2}.$$

Recalling the Chernoff bounds [3], we can obtain the following:

- For any $\delta > 0$

$$\begin{aligned}
 P\left[N_T > (1 + \delta) \frac{n}{m^2}\right] &< \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^{n/m^2} \\
 &= e^{-[(1+\delta) \log(1+\delta) - \delta] \frac{n}{m^2}}.
 \end{aligned}$$

- For any $0 < \delta < 1$

$$P\left[N_T < (1 - \delta) \frac{n}{m^2}\right] < e^{-\frac{\delta^2 n}{2m^2}}.$$

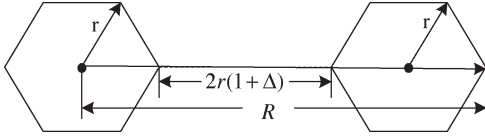


Fig. 2. Calculating the number of interfering cells.

From the foregoing discussion, for any $0 < \delta < 1$, since $(1 + \delta) \log(1 + \delta) - \delta \geq \delta^2/2$, we can obtain

$$P\left[\left|N_T - \frac{n}{m^2}\right| > \delta \frac{n}{m^2}\right] < e^{-\theta \frac{n}{m^2}}, \quad \theta = (1 + \delta) \log(1 + \delta) - \delta.$$

Thus, when $m = o(\sqrt{n})$ as $n \rightarrow \infty$, in an arbitrary cell, the total number of source nodes transmitting in the ad hoc mode is on the order of n/m^2 with probability 1. Similarly, we can prove that, in an arbitrary cell, the number of source nodes transmitting in the infrastructure mode is on the order of $(n/m) - (n/m^2)$, i.e., on the order of n/m , with probability 1.

Moreover, the ad hoc mode transmissions in different cells may interfere with each other. We have the following result.

Lemma 1: A cell has at most c interfering cells, where $c > 0$ is a constant, irrespective of the number of nodes or the number of base stations.

Proof: As indicated in Fig. 2, the maximal distance for the ad hoc mode transmissions within a cell i is $2r$, where r is the cell radius. Thus, those cells where ad hoc mode transmissions may interrupt with the transmissions in cell i must be included in a disk centered at the center of cell i with a radius of $R = r + 2r(1 + \Delta) + 2r$. Thus, the number of cells within this area, which is denoted by c' , is

$$c' = \frac{\pi R^2}{\frac{3\sqrt{3}}{2}r^2} = \frac{2\sqrt{3}(5 + 2\delta)^2\pi}{9}.$$

Thus, $c = c' - 1$ is a constant, irrespective of n or m . ■

By Lemma 1, the bandwidth used by the ad hoc mode transmissions in a cell is lower bounded by $W_1/(1 + c)$, which is obviously upper bounded by W_1 . Thus, the bandwidth for ad hoc mode transmissions in a cell is on the order of W_1 . Moreover, since there are m cells, according to the result in (3), we can obtain the throughput contributed by ad hoc transmissions utilizing the distributed MIMO technique, i.e., $T_a(n, m)$, as follows:

$$T_a(n, m) = \Theta\left(m \left(\frac{n}{m^2}\right)^{\frac{k}{k+1}} W_1\right). \quad (7)$$

We then derive the throughput contributed by transmissions in the infrastructure mode, i.e., $T_i(n, m)$. Notice that each packet transmitted from a source to its destination in the infrastructure mode will use one uplink and one downlink, and hence, it should be counted only once for the throughput.

Since the bandwidth for uplink is W_2 bits per second, the per-cell throughput in the infrastructure mode, which is denoted by $T_i^c(n, m)$, is upper bounded by W_2 . As we have mentioned previously, the base stations divide the area into a hexagonal tessellation, i.e., a seven-cell frequency reuse pattern that we

assume in this paper. Thus, the per-cell throughput in the infrastructure mode is lower bounded by $(1/7)W_2$. Thus, we can have

$$T_i(n, m) = mT_i^c(n, m) = \Theta(mW_2). \quad (8)$$

Substituting (7) and (8) into (5), we can obtain the throughput of the hybrid wireless networks, i.e.,

$$T(n, m) = \Theta\left(m \left(\frac{n}{m^2}\right)^{\frac{k}{k+1}} W_1\right) + \Theta(mW_2). \quad (9)$$

Moreover, we derive the average delay as follows: Based on the analysis in Section III, in $D_{\text{mimo}}^k(n/m^2)/W_1$ time slots, there are in total $W_1 \cdot T_{\text{mimo}}^k(n/m^2) \cdot D_{\text{mimo}}^k(n/m^2)/W_1$, i.e., $T_{\text{mimo}}^k(n/m^2)D_{\text{mimo}}^k(n/m^2)$, bits transmitted in the ad hoc mode, with the delay of each bit being $D_{\text{mimo}}^k(n/m^2)$. Since we assume that each packet has constant size and the number of nodes in infrastructure mode in each cell is on the order of n/m , in the infrastructure mode, there are $W_2 \cdot D_{\text{mimo}}^k(n/m^2)/W_1$ bits transmitted, and the delay of each bit is $\Theta(n/m)$. As a result, according to (6), the average delay of hybrid wireless networks, which is denoted by $D(n, m)$, is on the order of

$$\frac{T_{\text{mimo}}^k(n/m^2)D_{\text{mimo}}^k(n/m^2)D_{\text{mimo}}^k(n/m^2) + \left(W_2 \frac{D_{\text{mimo}}^k(n/m^2)}{W_1}\right) \frac{n}{m}}{T_{\text{mimo}}^k(n/m^2)D_{\text{mimo}}^k(n/m^2) + W_2 \frac{D_{\text{mimo}}^k(n/m^2)}{W_1}}.$$

Thus, we can have

$$\begin{aligned} D(n, m) &= \Theta\left(\frac{\left(\frac{n}{m^2}\right)^{\frac{2k^2+4k+4}{2k+2}} W_1 + \left(\frac{n}{m}\right) \left(\frac{n}{m^2}\right)^{\frac{k^2+k+2}{2k+2}} W_2}{\left(\frac{n}{m^2}\right)^{\frac{k^2+3k+2}{2k+2}} W_1 + \left(\frac{n}{m}\right)^{\frac{k^2+k+2}{2k+2}} W_2}\right) \\ &= \Theta\left(\frac{\left(\frac{n}{m^2}\right)^{\frac{k^2+k+2}{2k+2}} W_1 + \left(\frac{n}{m}\right) \left(\frac{n}{m^2}\right)^{-\frac{2k}{2k+2}} W_2}{W_1 + \left(\frac{n}{m^2}\right)^{-\frac{2k}{2k+2}} W_2}\right). \end{aligned} \quad (10)$$

Moreover, in (9), since $m = o(\sqrt{n})$, we have that

$$m \left(\frac{n}{m^2}\right)^{\frac{k}{k+1}} = \omega(m).$$

Thus, $T(n, m)$ in (9) is dominated by the first term. Hence, by allocating all resources to the ad hoc mode transmissions, i.e., $W_1 = W$ and $W_2 = W_3 = 0$, we can maximize the throughput. Denoting this optimal throughput by $T^o(n, m)$, we have

$$T^o(n, m) = \Theta\left(m \left(\frac{n}{m^2}\right)^{\frac{k}{k+1}} W\right). \quad (11)$$

In this case, there are only ad hoc mode transmissions in the network, and the delay, which is denoted by $D_T(n, m)$, is given by

$$D_T(n, m) = \Theta\left(\left(\frac{n}{m^2}\right)^{\frac{k^2+k+2}{2(k+1)}}\right).$$

However, as shown in (10), when we allocate all resources to the ad hoc mode transmissions, i.e., $W_1 = W$ and $W_2 = W_3 = 0$, the delay would be

$$D^o(n, m) = \Theta \left(\left(\frac{n}{m^2} \right)^{\frac{k^2+k+2}{2k+2}} \right)$$

and when we exclusively allocate all resources to the infrastructure mode transmissions, i.e., $W_1 = 0$ and $W_2 = W_3 = W/2$, the delay is

$$D^o(n, m) = \Theta \left(\frac{n}{m} \right).$$

Assume that $m = n^x$, where $0 \leq x < 1/2$. By letting

$$\left(\frac{n}{m^2} \right)^{\frac{k^2+k+2}{2k+2}} > \frac{n}{m}$$

we have $k > ((1 + \sqrt{1 + 8(1 - 2x)}) / 2(1 - 2x))$. In other words, when $m = n^x$, and $k > ((1 + \sqrt{1 + 8(1 - 2x)}) / 2(1 - 2x))$, the delay when we exclusively allocate all resources to the infrastructure mode transmissions will be at a lower order than that when we exclusively allocate all resources to the ad hoc mode transmissions. Moreover, the throughput under the former condition is

$$T_D(n, m) = \Theta \left(m \frac{W}{2} \right) = \Theta(mW).$$

Thus, we can clearly observe that there is a tradeoff between throughput and delay. Furthermore, the optimal throughput shown in (11) becomes $\Theta(n/m)$ as k is relatively large, which means we can have linear scaling throughput only when $m = \Theta(1)$.

2) *Case II: $m = \Omega(\sqrt{n})$:* Similar to that in Case I, according to the Chernoff bound, we have for any $\delta > 0$

$$P \left[N_T > (1 + \delta) \frac{n}{m^2} \right] < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{n/m^2}.$$

Let $\theta = (1 + \delta) \log(1 + \delta) - \delta$. Then, we can easily show that $\theta > 0$ and have

$$P \left[N_T > (1 + \delta) \frac{n}{m^2} \right] < e^{-\frac{\theta n}{m^2}}. \tag{12}$$

By choosing $1 + \delta = (m^2/n) \log n$, we can obtain

$$\begin{aligned} \theta &= \left(\frac{m^2}{n} \log n \right) \log \left(\frac{m^2}{n} \log n \right) - \left(\frac{m^2}{n} \log n - 1 \right) \\ &= \Theta \left(\left(\frac{m^2}{n} \log n \right) \log \left(\frac{m^2}{n} \log n \right) \right). \end{aligned}$$

Substituting θ into (12), we can have

$$P[N_T > \log n] < e^{-\frac{\theta n}{m^2}} \rightarrow 0$$

which gives us $N_T = O(\log n)$.

Thus, according to (3) and (4), we can obtain the throughput and the delay of the ad hoc mode transmissions in hybrid

wireless networks, respectively, i.e.,

$$\begin{aligned} T_a(n, m) &= O \left(m (\log n)^{\frac{k}{k+1}} W_1 \right) \\ D_a(n, m) &= O \left((\log n)^{\frac{k^2+k+2}{2(k+1)}} \right). \end{aligned}$$

As in Case I, we know that the throughput and the delay of the infrastructure mode transmissions, respectively, are

$$\begin{aligned} T_i(n, m) &= \Theta(mW_2) \\ D_i(n, m) &= \Theta \left(\frac{n}{m} \right). \end{aligned}$$

As a result, according to (5), we finally obtain

$$T(n, m) = O \left(m (\log n)^{\frac{k}{k+1}} W_1 \right) + \Theta(mW_2). \tag{13}$$

Similar to that in (10), we can also obtain the average delay, i.e.,

$$D(n, m) = O \left(\frac{(\log n)^{\frac{2k^2+4k+4}{2k+2}} W_1 + \left(\frac{n}{m} \right) (\log n)^{\frac{k^2+k+2}{2k+2}} W_2}{(\log n)^{\frac{k^2+3k+2}{2k+2}} W_1 + (\log n)^{\frac{k^2+k+2}{2k+2}} W_2} \right). \tag{14}$$

Assume that m is on a certain exponential order of n , i.e., $m = n^x$, where $(1/2) < x < 1$. Then, we find in (13) that $m(\log n)^{k/k+1}$ and m have the same exponential order of n . Thus, we choose to allocate all resources to infrastructure mode transmissions, i.e., $W_1 = 0$, and $W_2 = W_3 = W/2$. Thus, the optimal throughput and delay under this condition, respectively, are

$$\begin{aligned} T^o(n, m) &= \Theta(mW) \\ D_T(n, m) &= \Theta \left(\frac{n}{m} \right). \end{aligned}$$

As shown in (10), when we exclusively allocate all resources to the ad hoc mode transmissions, the delay would be

$$D^o(n, m) = O \left((\log n)^{\frac{k^2+k+2}{2k+2}} \right)$$

and when we exclusively allocate all resources to the infrastructure mode transmissions, the delay is

$$D^o(n, m) = \Theta \left(\frac{n}{m} \right).$$

Thus, when $m = o(n)$, the optimal delay of hybrid wireless networks achieved when all resources are exclusively allocated to the ad hoc mode transmissions, i.e., $W_1 = W$, and $W_2 = W_3 = 0$, and the throughput under this condition, respectively, are

$$\begin{aligned} D^o(n, m) &= O \left((\log n)^{\frac{k^2+k+2}{2k+2}} \right) \\ T_D(n, m) &= O \left(m (\log n)^{\frac{k}{k+1}} W \right). \end{aligned}$$

Moreover, when $m = \Theta(n)$, the optimal delay is obtained when all resources are exclusively allocated to the infrastructure

mode transmissions. The throughput and the delay under this condition are the same as $T^o(n, m)$ and $D_T(n, m)$ shown above.

Obviously, we also find that there is a tradeoff between the throughput and the delay in Case II. Furthermore, when we allocate all resources to the ad hoc mode transmissions, the network throughput linearly increases with the number of base stations, and the average delay decreases inversely proportionally to the number of base stations. When $m = \Theta(n)$, we can have $T^o(n, m) = \Theta(nW)$ and $D_T(n, m) = \Theta(1)$.

As a result, from both Case I and Case II, we find that, in the hybrid wireless networks using the proposed resource-allocation strategy, even with distributed MIMO technology, the network can achieve linear scaling throughput either when $m = \Theta(1)$ and $W_1 = W$ or when $m = \Theta(n)$ and $W_2 = W_3 = W/2$. In addition, the average delay is very large in the first case and is $\Theta(1)$ in the second case.

V. CONCLUSION

In this paper, we have proposed to use distributed MIMO technology in hybrid wireless networks. We first employed a resource allocation strategy such that a source and a destination communicate in the ad hoc mode only when they are in the same cell. Then, we showed that, under this strategy, the network can achieve linear scaling throughput either when $m = \Theta(1)$ and all resources are allocated to the ad hoc mode transmissions or when $m = \Theta(n)$ and all resources are allocated to the infrastructure mode transmissions. The delay can be $\Theta(1)$ in the second case.

However, we also noticed that this resource-allocation strategy may not be the most efficient strategy. In our future work, we will further investigate the throughput and delay tradeoff in hybrid wireless networks using more efficient strategies. Moreover, the transmission power issue in MIMO transmissions is very important but was not investigated in this paper. We also leave this issue to our future research.

REFERENCES

- [1] A. Agarwal and P. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 3, pp. 71–81, Jul. 2004.
- [2] C. Buragohain, S. Suri, C. Toth, and Y. Zhou, "Improved throughput bounds for interference-aware routing in wireless networks," in *Proc. COCOON*, Banff, AB, Canada, Jul. 2007, pp. 210–221.
- [3] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.
- [4] O. Dousse, M. Franceschetti, and P. Thiran, "On the throughput scaling of wireless relay networks," *IEEE Trans. Inf. Theory—Joint Special Issue IEEE/ACM Trans. Netw.*, vol. 52, no. 6, pp. 2756–2761, Jun. 2006.
- [5] E. Duarte-Melo, A. Josan, M. Liu, D. Neuhoff, and S. Pradhan, "The effect of node density and propagation model on throughput scaling of wireless networks," in *Proc. IEEE ISIT*, Jul. 2006, pp. 1693–1697.
- [6] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks—Part I: The fluid model," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2568–2592, Jun. 2006.
- [7] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks—Part II: Constant-size packets," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5111–5116, Nov. 2006.
- [8] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
- [9] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

- [10] D. Knuth, *The Art of Computer Programming*. Reading, MA: Addison-Wesley, 1998.
- [11] U. Kozat and L. Tassiulas, "Throughput capacity of random ad hoc networks with infrastructure support," in *Proc. ACM MobiCom*, San Diego, CA, Jun. 2003, pp. 55–65.
- [12] P. Li and Y. Fang, "Impacts of topology and traffic pattern on capacity of hybrid wireless networks," *IEEE Trans. Mobile Comput.*, to be published. [Online]. Available: <http://www2.computer.org/portal/web/csdl/doi/10.1109/TMC.2009.79>
- [13] P. Li, C. Zhang, and Y. Fang, "Asymptotic connectivity in wireless ad hoc networks using directional antenna," *IEEE/ACM Trans. Netw.*, 2009, to be published. [Online]. Available: <http://ieeexplore.ieee.org/xpl/preabsprintf.jsp?arnumber=4731842>
- [14] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE J. Sel. Areas Commun.—Special Issue Broadband Access Netw.*, vol. 27, no. 2, pp. 117–125, Feb. 2009.
- [15] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar. 2003, pp. 1543–1552.
- [16] B. Liu, P. Thiran, and D. Towsley, "Capacity of a wireless ad hoc network with infrastructure," in *Proc. ACM MobiHoc*, Montreal, QC, Canada, Sep. 2007, pp. 239–246.
- [17] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: A unified cellular and ad-hoc network architecture," in *Proc. ACM MobiCom*, San Diego, CA, Sep. 2003, pp. 353–367.
- [18] A. Ozgur, O. Leveque, and D. Tse, "How does the information capacity of ad hoc networks scale?" in *Proc. 44th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, Sep. 2006.
- [19] A. Ozgur, O. Leveque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.
- [20] Y. Pei, J. Modestino, and X. Wang, "On the throughput capacity of hybrid wireless networks using an L -maximum-hop routing strategy," in *Proc. IEEE Veh. Technol. Conf.*, 2003, pp. 2173–2176.
- [21] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall PTR, 2002.
- [22] S. Toumpis, "Capacity bounds for three classes of wireless networks," in *Proc. ACM MobiHoc*, Tokyo, Japan, May 2004, pp. 133–144.
- [23] R. W. Yeung, S.-Y. Li, N. Cai, and Z. Zhang, "Network coding theory," *Foundation Trends Commun. Inf. Theory*, vol. 2, no. 4/5, pp. 241–381, Mar. 2005.
- [24] A. Zemplianov and G. Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 657–667, Mar. 2005.



Xiaoyan Zhu received the B.E. degree in information engineering and the M.E. degree in information and communications engineering in 2000 and 2004, respectively, from Xidian University, Xian, China, where she is currently working toward the Ph.D. degree.

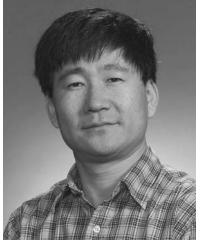
Her research interests include wireless security and network coding.



Pan Li (S'06) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2005. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Florida, Gainesville.

His research interests include capacity and connectivity analysis, medium-access control, routing algorithms, and cross-layer design in wireless networks.

Mr. Li is a Student Member of the ACM.



Yuguang "Michael" Fang (S'92–M'97–SM'99–F'08) received the Ph.D. degree in systems engineering from Case Western Reserve University, Cleveland, OH, in 1994 and the Ph.D. degree in electrical engineering from Boston University, Boston, MA, in 1997.

He was an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, from 1998 to 2000. He then joined the Department of Electrical and Computer Engineering, University of Florida, Gainesville, in May 2000 as an Assistant Professor, got an early promotion to Associate Professor with tenure in August 2003 and another promotion to Full Professor in August 2005. He holds a University of Florida Research Foundation Professorship from 2006 to 2009, a Changjiang Scholar Chair Professorship with Xidian University, Xi'an, China, from 2008 to 2011, and a Guest Chair Professorship with Tsinghua University, Beijing, China, from 2009 to 2012. He has authored more than 250 papers published in refereed professional journals and conference proceedings. He is also active in professional activities.

Dr. Fang is a member of the Association for Computing Machinery. He currently serves as the Editor-in-Chief of IEEE WIRELESS COMMUNICATIONS and has served on several Editorial Boards of technical journals, including the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, *IEEE Wireless Communications Magazine*, and *ACM Wireless Networks*. He was an Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING and currently serves on its Steering Committee. He has been actively participating in professional conference organizations, such as serving as the Steering Committee Cochair for the International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, the Technical Program Vice-Chair for the 2005 IEEE Conference on Computer Communications (INFOCOM), the Technical Program Symposium Cochair for the 2004 IEEE Global Telecommunications Conference, and a Technical Program Committee Member for the INFOCOM in 1998, 2000, and 2003–2010. He was the recipient of the National Science Foundation Faculty Early Career Award in 2001; the Office of Naval Research Young Investigator Award in 2002; the Best Paper Award at the IEEE International Conference on Network Protocols in 2006; and the IEEE TCGN Best Paper Award at the IEEE High-Speed Networks Symposium, IEEE GLOBECOM, in 2002.



Yumin Wang (SM'98) received the B.S. degree from Xidian University, Xi'an, China, in 1959.

He was a Lecturer from 1959 to 1979. From 1979 to 1981, he was a Visiting Research Scholar with the Department of Electrical Engineering, University of Hawaii, Honolulu. He then went back to Xidian University as an Associate Professor from 1981 to 1986 and has been a Professor with the School of Telecommunications since 1986.

Mr. Wang is a Fellow of the Chinese Institute of Electronics and a Fellow of the Chinese Institute of Communications. He is also an Honorary Member of the Board of Directors for the Chinese Institute of Cryptography. He was the recipient of the Scientific Achievement Award from the China Ministry of Electronics in 1983, the Scientific Advancement Award from the China Ministry of Electronics in 1992 and 1997, and the China National Technological Innovation Award in 2005.