Towards Achieving the Maximum Capacity in Large Mobile Wireless Networks Under Delay Constraints

Xiaojun Lin and Ness B. Shroff

Abstract: **In this paper, we study how to achieve the maximum capacity under delay constraints for large mobile wireless networks. We develop a systematic methodology for studying this problem in the asymptotic region when the number of nodes n in the network is large. We first identify a number of key parameters for a large class of scheduling schemes, and investigate the inherent tradeoffs among the capacity, the delay, and these scheduling parameters. Based on these inherent tradeoffs, we are able to compute the upper bound on the maximum per-node capacity of a large mobile wireless network under given delay constraints. Further, in the process of proving the upper bound, we are able to identify the optimal values of the key scheduling parameters. Knowing these optimal values, we can then develop scheduling schemes that achieve the upper bound up to some logarithmic factor, which suggests that our upper bound is fairly tight. We have applied this methodology to both the *i.i.d.* mobility model and the random way-point mobility model. In both cases, our methodology allows us to develop new scheduling schemes that can achieve larger capacity than previous proposals under the same delay constraints. In particular, for the *i.i.d.* mobility model, our scheme can achieve $\Theta(n^{-1/3}/\log^{3/2} n)$ per-node capacity with constant delay. This demonstrates that, under the *i.i.d.* mobility model, mobility increases the capacity even with constant delays. Our methodology can also be extended to incorporate additional scheduling constraints.**

Index Terms: **Capacity-delay tradeoff, large system asymptotics, mobile ad hoc networks, mobile wireless networks.**

I. INTRODUCTION

Since the seminal paper by Gupta and Kumar [1], there has been tremendous interest in the networking research community to understand the fundamental achievable capacity in wireless networks. For a static network (where nodes do not move), Gupta and Kumar show that the per-node capacity decreases as $O(1/\sqrt{n \log n})$[1] as the number of nodes n increases [1]. The capacity of wireless networks can be improved when *mobility* is taken into account. When the nodes are mobile, Grossglauser and Tse show that per-node capacity of $\Theta(1)$ is achievable [2], which is much better than that of static networks. This capacity improvement is achieved at the cost of excessive packet delays. In fact, it has been pointed out in [2] that the packet delay of the proposed scheme could be unbounded.

There have been several recent studies that attempt to address the relationship between the achievable capacity and the packet delay in mobile wireless networks. In the work by Neely and Modiano [3], it was shown that the maximum achievable per-node capacity of a mobile wireless network is bounded by $O(1)$. Under an *i.i.d.* mobility model, the authors of [3] present a scheme that can achieve $\Theta(1)$ per-node capacity and incur $\Theta(n)$ delay, provided that the load is strictly less than the capacity. Further, they show that it is possible to reduce packet delay if one is willing to sacrifice capacity. In [3], the authors formulate and prove a fundamental tradeoff between the capacity and delay. Let the average end-to-end delay be bounded by D. For D between $\Theta(1)$ and $\Theta(n)$, [3] shows that the maximum per-node capacity λ is upper bounded by

$$\lambda \leq O(\frac{D}{n}). \tag{1}$$

The authors of [3] develop schemes that can achieve $\Theta(1)$, $\Theta(1/\sqrt{n})$, and $\Theta(1/(n \log n))$ per-node capacity, when the delay constraint is on the order of $\Theta(n)$, $\Theta(\sqrt{n})$, and $\Theta(\log n)$, respectively.

Inequality (1) leads to the *pessimistic* conclusion that a mobile wireless network can sustain at most $O(1/n)$ per-node capacity with a constant delay bound. This capacity is even worse than that of static networks. It turns out that this pessimistic conclusion is due to certain restrictive assumptions that are implicit in the work in [3] (we will elaborate on these assumptions in Section VII). In fact, Toumpis and Goldsmith [4] present a scheme that can achieve a per-node capacity of $\Theta(n^{(d-1)/2}/\log^{5/2} n)$ when the delay is bounded by $O(n^d)$. The result of [4] has incorporated the effect of fading. If we remove fading, the per-node capacity will be of the order $\Theta(n^{(d-1)/2}/\log^{3/2} n)$. Ignoring the logarithmic term, we find that the following capacity-delay tradeoff is achievable by the scheme of [4]:

$$\lambda^2 = \Theta(\frac{D}{n}). \tag{2}$$

This is better than (1). In particular, the authors of [4] present a scheme that can achieve $\Theta(1/(\sqrt{n} \log^{3/2} n))$ per-node capacity with a constant delay bound. (The capacity will be $\Theta(1/(\sqrt{n \log n}))$ with no fading.) This capacity is now *comparable* to that of the static wireless networks.

An open question then is: *what is the maximum achievable capacity of mobile wireless networks under given delay con-*

Manuscript received xxxxxxxxx xx, 2004; approved for publication by Eytan Modiano, Editor, xxxxxxxx xx, 2004. An earlier version of this paper has appeared in the Third Annual Mediterranean Ad Hoc Networking Workshop [9].

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA (e-mail: linx@ecn.purdue.edu; shroff@ecn.purdue.edu).

[1] We use the following notation throughout:

$$\begin{aligned}
f(n) = o(g(n)) \quad &\leftrightarrow \quad \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0, \\
f(n) = O(g(n)) \quad &\leftrightarrow \quad \limsup_{n\to\infty} \frac{f(n)}{g(n)} < \infty, \\
f(n) = \omega(g(n)) \quad &\leftrightarrow \quad g(n) = o(f(n)), \\
f(n) = \Theta(g(n)) \quad &\leftrightarrow \quad f(n) = O(g(n)) \text{ and } g(n) = O(f(n)).
\end{aligned}$$

straints? Inequality (1) is clearly not optimal. Without a careful study of the various inherent tradeoffs within the system, a constructive methodology such as the one in [4] will only produce a lower bound like (2). In this paper, we attempt to address this question using a systematic methodology as follows. We first identify several key parameters of a general class of scheduling schemes, and investigate the inherent tradeoffs among the capacity, the delay, and these scheduling parameters. Based on these inherent tradeoffs (in the form of inequalities), we are able to compute an upper bound on the maximum per-node capacity of a large mobile wireless network under given delay constraints. In the process of proving the upper bound, we are also able to identify the optimal values of the key scheduling parameters. Knowing these optimal values, we can then develop scheduling schemes that achieve the upper bound up to some logarithmic factor, which suggests that our upper bound is fairly tight. We have applied this methodology to both the *i.i.d.* mobility model and the random way-point mobility model. For the *i.i.d.* mobility model, the inherent tradeoffs that our methodology is based on can be analytically established. For the random way-point mobility model, we use a combination of analytical and numerical techniques to establish these inherent tradeoffs. In both cases, we are able to obtain new insights on the optimal choices of the key scheduling parameters, and develop new scheduling schemes that can achieve larger capacity than previous proposals under the same delay constraints. For example, under the *i.i.d* mobility model, we can achieve a new capacity-delay tradeoff of

$$\lambda^3 \geq \Theta(\frac{\bar{D}}{n}/\log^{9/2} n). \tag{3}$$

In Fig. 1, we draw this new tradeoff (the top line) alongside the capacity-delay tradeoffs achieved by the schemes in [3] and [4] (the bottom line and the middle line, respectively). Our new scheme clearly achieves larger capacity when the delay constraints are small. In particular, when the delay is bounded by a constant, our scheme can achieve $\Theta(n^{-1/3}/\log^{3/2} n)$ per-node capacity. Unlike previous works, this result shows that, even for a constant delay bound, the per-node capacity of mobile wireless networks can be larger than that of the static networks! Finally, our methodology can be extended to incorporate additional constraints of the scheduling schemes.

The rest of the paper is structured as follows. In Section II, we outline the network and mobility models. We will first focus on the *i.i.d.* mobility model in Sections III-VII. In Section III, we outline the general class of scheduling policies that we will consider. We then identify a number of key scheduling parameters and study their inherent tradeoffs in Section IV. We establish the upper bound on the optimal capacity-delay tradeoff in Section V and present a scheme in Section VI that achieves a capacity-delay tradeoff close to the upper bound. In Section VII, we discuss how to treat additional scheduling constraints such as those that appear in previous works [3], [4]. The extension to the random way-point mobility model is studied in Section VIII. Then we conclude.

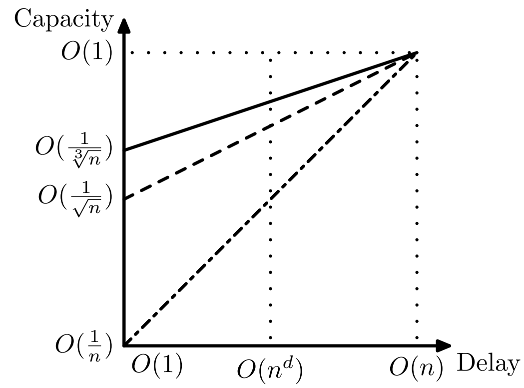

Fig. 1. Comparison of the achievable capacity-delay tradeoffs (ignoring the logarithmic terms).

II. NETWORK AND MOBILITY MODELS

We consider a mobile wireless network with n nodes moving within a unit square[2]. For simplicity, we assume the following traffic model similar to the models in [3], [4]. We assume that the number of nodes n is even and the nodes can be labeled in such a way that node $2i-1$ communicates with node $2i$, and node $2i$ communicates with node $2i-1$, $i = 1, 2, ..., n/2$. The communication between any source-destination pairs can go through multiple other nodes as *relays*. That is, the *source* can either send a message directly to the *destination*; or, it can send the message to one or more *relay* nodes; the relay nodes can further forward the message to other relay nodes (possibly after moving to another position); and finally some relay node forwards the message to the destination.

We assume the following Protocol Model from [1] that governs *direct radio transmissions* between nodes. Let W be the bandwidth of the system. Let X_i denote the position of node i, $i = 1, ..., n$. Let $|X_i - X_j|$ be the Euclidean distance between nodes i and j. At any time, node i can communicate *directly* with another node j at W bits per second if and only if the following interference constraint is satisfied [1]:

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

for every other node $k \neq i, j$ that is simultaneously transmitting. Here, Δ is some positive number. Note that an alternative model for direct radio transmission is the Physical Model [1], [4]. In the Physical Model, a node can communicate with another node if the signal-to-interference ratio is above a given threshold. It has been shown that, under certain conditions, the Physical Model can be reduced to the Protocol Model with an appropriate choice of Δ [1]. Hence, we will not consider the Physical Model any further in this paper. We also assume that no nodes can transmit and receive over the same frequency at the same time.

We will study two types of mobility models.

1) **The *i.i.d.* Mobility Model:** In the *i.i.d* mobility model [3], the time is divided into slots of unit length. At each time slot, the positions of each node are *i.i.d.* and uniformly distributed

[2]Note that changing the shape of the area from a square to a circle or other topologies will not affect our main results.

within the unit square. Between time slots, the distributions of the positions of the nodes are independent. Although the *i.i.d.* mobility model is somewhat restrictive in assuming the distribution of the node positions to be independent across time slots, its mathematical tractability allows us to gain important insights into the structure of the problem, which can then be extended to other more realistic mobility models.

2) **The Random Way-point (*RWP*) Mobility Model:** In the random way-point (*RWP*) mobility model, we assume that the unit square is a *torus*, i.e., a node can move out of the unit square from an edge and immediately move into the unit square from the opposite edge.[3] The initial positions of the nodes at time $t = 0$ are *i.i.d.* and uniformly distributed within the unit square. Each node then moves independently in *trips*: for each *trip*, the node picks a target position uniformly distributed within the unit square, and moves towards the target position along the shortest path at a constant speed v. (Note that since the unit square is a torus, the shortest path may not always be the straight line.) When the node reaches the target position, it immediately starts another *trip* by picking a new target position randomly. Unlike [5], we assume that, when a node is picked as the relay node for a message, the information about the future motion of the relay node is not available to the scheduler. (On the other hand, the scheduling scheme in [5] has exploited this knowledge to obtain a different capacity-delay tradeoff than ours under a somewhat similar *uniform mobility model*.) Following the convention in related studies [6], we assume that the speed v scales as $v(n) = \Theta(1/\sqrt{n})$ when the number of nodes n increases.[4]

Under both mobility models, we assume the following *separation of time scales*, i.e., radio transmission can be scheduled at a time scale much faster than that of node mobility. This is usually a reasonable assumption in real networks. Hence, a message may be divided into multiple bits and each bit can be forwarded "instantaneously" across multiple hops as if the positions of all nodes are frozen.

We assume a uniform traffic pattern, that is, all source nodes communicate with their destination nodes at the same rate λ. Let $\bar{D}$ be the mean delay averaged over all messages and all source-destination pairs. Both λ and $\bar{D}$ will depend on how the transmissions between mobile nodes are scheduled. We are interested in capturing the fundamental tradeoff between the achievable capacity λ and the delay $\bar{D}$. That is, over all possible ways of scheduling the radio transmissions, what is the maximum per-node capacity λ given certain constraint on the delay $\bar{D}$.

III. THE CLASS OF SCHEDULING POLICIES

In Sections III-VII, we will focus on the *i.i.d* mobility model, and we will defer the study of the random way-point mobility model until Section VIII. In this section, we define the class of scheduling policies that we will consider for the *i.i.d.* mobility model. Because we are interested in the fundamental *achievable* capacity for given delay constraints, we will assume that there exists a scheduler that has all the information about the current and past status of the network, and can schedule any radio transmission in the current and future time slots. At each time slot t, for each bit b that has not been delivered to its destination yet, the scheduler needs to perform the following two functions:

- *Capture:* The scheduler needs to decide whether to deliver the bit b to the destination within the current time slot. If yes, the scheduler then needs to choose one relay node (possibly the source) that has a copy of the bit b at the beginning of the time slot t, and schedule radio transmissions to forward this bit to the destination *within the same time slot*, using possibly multi-hop transmissions. When this happens successfully, we say that the chosen relay node has successfully *captured* the destination of bit b, or a successful *capture* has occurred for bit b.
- *Duplication:* If capture does not occur for bit b, the scheduler needs to decide whether to *duplicate* bit b to other nodes that do not have the bit at the beginning of the time slot t. The scheduler also needs to decide which nodes to relay from and relay to, and how to schedule radio transmissions to forward the bit to these new relay nodes.

The *capture* function and the *duplication* function are mutually exclusive for a given bit b. Once a successful *capture* occurs, the bit b will be delivered to its destination within the same time slot, and hence can exit the system. Note that once a successful capture occurs, it is important to forward the bit b to the destination within a single time slot. Otherwise, since the chosen relay node may move arbitrarily far away from the destination in the next time slot, the nodes that received the bit b in the current time slot will only count as new relay nodes for the bit b, and they have to capture again in the next time slot.

In this paper, we will consider the class of *causal* scheduling policies that perform the above two functions at each time slot. The causality assumption essentially requires that, when the scheduler makes the capture decision and the duplication decision, it can only use information about the current and the past status of the network. In particular, at any time slot t, the scheduler cannot use information about the *future* positions of the nodes at any time slot $s > t$.

This class of scheduling policies is clearly very general, and encompasses nearly any practical scheduling scheme we can think of. (Note that even *predictive* scheduling schemes have to rely on current and past information only.) Some remarks on the *capture* process are in order. Although we do allow for other less intuitive alternatives, in a typical scheduling policy a successful *capture* usually occurs when some relay nodes are within an area close to the destination node, so that fewer resources will be needed to forward the information to the destination. For example, a relay node could enter a disk of a certain radius around the destination, or a relay node could enter the same cell as the destination. We call such an area a *capture neighborhood*. The relay nodes that has the bit b at the beginning of the time slot t are called *mobile relays* for bit b. The mobile relay that is chosen to forward the bit b to the destination is called the *last mobile relay* for bit b.

Within this large class of general scheduling policies, the following *cell-based scheme* is of particular interest as we will use it to construct the optimal capacity-achieving scheme later on.

The Cell-Based Scheduling Scheme: The unit square is divided into $g_1(n)$ *sending cells* with equal size. Once a bit b

[3]The assumption of a torus could be removed. It is included here for mathematical convenience so that we do not need to deal with the edge effects.

[4]It is also possible to extend our methodology to the case when the speed is randomly distributed between $[v(n), cv(n)]$ for some $c > 1$, and to the case when nodes pause between trips.

enters the system, it is immediately broadcast to all other nodes in the same sending cell. The number of mobile relays for the bit b then stays unchanged. To decide a successful *capture*, we divide the unit square into $g_2(n)$ *receiving cells* with equal size. When one of the mobile relays for bit b moves into the same receiving cell as the destination, the bit b is then forwarded from the last mobile relay to the destination in a multi-hop fashion.

Compared with the most general scheduling policies, the cell-based scheme clearly lacks in flexibility. Hence, it is not *a priori* obvious why the choice of setting up a rigid cell-structure beforehand is a good one for achieving the maximum capacity. This choice will however become evident after we establish the upper bound on the maximum capacity under delay constraints.

IV. INHERENT TRADEOFFS AMONG THE KEY SCHEDULING PARAMETERS

We find three key parameters in the cell-based scheme: $g_1(n)$ will determine the number of mobile relays (denoted by R_b with mean $n/g_1(n)$), $g_2(n)$ will determine the area of the capture neighborhood (denoted by $A_b = 1/g_2(n)$), and the network also needs to determine the number of hops h_b that each bit b takes from the last mobile relay to its destination. It turns out that these three key parameters can also be defined for general scheduling policies that are *not* cell-based, and these parameters combined will determine the performance of any scheduling policy. Indeed, fixing any scheduling policy, we can define the following random variables for each bit b that needs to be communicated from its source node to its destination node. Let $t_0(b)$ denote the time slot when bit b first enters the system. Let $s_b \geq t_0(b)$ be the time slot when the scheduler decides that a successful capture for bit b occurs. Define the delay $D_b \triangleq s_b - t_0(b)$. The scheduler also needs to choose one mobile relay that has a copy of bit b at the beginning of the time slot s_b to forward bit b towards its destination *within the same time slot* s_b. Let R_b be the number of mobile nodes holding the bit b at the time of capture, and let l_b denote the distance from the chosen last mobile relay to the destination of bit b. (Note that for the cell-based scheme, l_b is on the order of $\sqrt{A_b}$.) Let h_b be the number of hops that bit b takes from the last mobile relay to its destination. The parameters R_b, l_b, and h_b are essential for the subsequent tradeoffs that determine the relationship between the achievable capacity and the delay D_b[5]. Note that D_b includes possible queueing delays at the source node or at the relay nodes.

We are now ready to state the inherent tradeoffs among capacity, delay and these key scheduling parameters. Intuitively, the larger the number of mobile relays and the larger the capture neighborhood, the smaller the delay. On the other hand, in order to improve capacity, we need to consume fewer radio resources, which implies a smaller number of mobile relays and a shorter distance from the last mobile relay to the destination. In this section, we will state these tradeoffs precisely for the *i.i.d.* mobility model, and defer the discussion on the random way-point mobility model until Section VIII.

A. Tradeoff I : D_b versus R_b and l_b

Proposition 1 captures the following tradeoff: the smaller the number R_b of mobile relays that bit b is duplicated to, and the shorter the targeted distance l_b from the last mobile relay to the destination, the longer it takes to capture the destination.

Proposition 1: Under the *i.i.d.* mobility model, the following inequality holds for any causal scheduling policy when $n \geq 3$,

$$c_1 \log n \mathbf{E}[D_b] \geq \frac{1}{(\mathbf{E}[l_b] + \frac{1}{n^2})^2 \mathbf{E}[R_b]} \text{ for all bits } b, \quad (4)$$

where c_1 is a positive constant.

The proof is available in [8]. This seemingly odd relationship is actually motivated by simple examples. When R_b and the area of the capture neighborhood A_b are constants, then $1 - (1 - A_b)^{R_b}$ is the probability that any one out of the R_b nodes can capture the destination in one time slot. It is easy to show that, the average number of time slots needed before a successful capture occurs is,

$$\mathbf{E}[D_b] = \frac{1}{1 - (1 - A_b)^{R_b}} \geq \frac{1}{A_b R_b} \geq \frac{c_1'}{\mathbf{E}^2[l_b] R_b},$$

where c_1' is a positive constant. Proposition 1 shows that, even when R_b and A_b are random variables, no scheduling policy can improve the tradeoff by more than a $\log n$ factor. For details, please refer to [8].

B. Tradeoff II : Multihop

Once a successful capture occurs, the chosen mobile relay (i.e., the *last mobile relay*) will start transmitting the bit to the destination *within a single time slot*, using possibly other nodes as relays. We will refer to these latter relay nodes as *static relays*. The static relays are only used for forwarding the bit to the destination *after a successful capture occurs*. Let h_b be the number of hops it takes from the last mobile relay to the destination. Let S_b^h denote the transmission range of each hop $h = 1, .., h_b$. The following relationship is trivial.

Proposition 2: The sum of the transmission ranges of the h_b hops must be no smaller than the straight-line distance from the last mobile relay to the destination, i.e.,

$$\sum_{h=1}^{h_b} S_b^h \geq l_b. \quad (5)$$

C. Tradeoff III : Radio Resources

It consumes radio resources to duplicate each bit to mobile relays and to forward the bit to the destination. Proposition 3 below captures the following tradeoff: the larger the number of mobile relays R_b and the further the multi-hop transmissions towards the destination have to traverse, the smaller the achievable capacity. Consider a large enough time interval T. The total number of bits communicated end-to-end between all source-destination pairs is $\lambda n T$.

Proposition 3: Assume that there exist positive numbers c_2 and N_0 such that $D_b \leq c_2 n^2$ for $n \geq N_0$. If the positions of

[5]Using the notion of *filtrations* and *stopping times* [7, p231], we can rigorously define these quantities as random variables on the same probability space where the random mobility of the mobile nodes is defined. The expectation in the following tradeoffs are then taken with respect to this probability space. See [8] for details.

the nodes within a time slot are *i.i.d.* and uniformly distributed within the unit square, then there exist positive numbers N_1 and c_3 that only depend on c_2, N_0 and Δ, such that the following inequality holds for any causal scheduling policy when $n \geq N_1$,

$$\sum_{b=1}^{\lambda nT} \frac{\Delta^2}{4} \frac{\mathbf{E}[R_b]-1}{n} + \mathbf{E}[\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} \frac{\pi\Delta^2}{4} (S_b^h)^2] \leq c_3 WT \log n. \tag{6}$$

The assumption that $D_b \leq c_2 n^2$ for large n is not as restrictive as it appears. It has been shown in [3] and [6] that the maximal achievable per-node capacity is $\Theta(1)$ and this capacity can be achieved with $\Theta(n)$ delay under both the *i.i.d.* mobility model and the random way-point mobility model. Hence, we are most interested in the case when the delay is not much larger than the order $O(n)$. Further, Proposition 3 only requires that the stationary distribution of the positions of the nodes within a time slot is *i.i.d.* It does not require the distribution between time slots to be independent, and hence can be extended to the random way-point mobility model as well.

We briefly outline the motivation behind the inequality (6). The details of the proof are quite technical and available in [8]. Consider nodes i, j that directly transmit to nodes k and l, respectively, at the same time. Then, according to the interference constraint:

$$\begin{aligned} |X_j - X_k| &\geq (1+\Delta)|X_i - X_k] \\ |X_i - X_l| &\geq (1+\Delta)|X_j - X_l]. \end{aligned}$$

Hence,

$$\begin{aligned} |X_j - X_i| &\geq |X_j - X_k| - |X_i - X_k| \\ &\geq \Delta|X_i - X_k|. \end{aligned}$$

Similarly,

$$|X_i - X_j| \geq \Delta|X_j - X_l|.$$

Therefore,

$$|X_i - X_j| \geq \frac{\Delta}{2}(|X_i - X_k| + |X_j - X_l|).$$

That is, disks of radius $\frac{\Delta}{2}$ times the transmission range centered at the transmitter are disjoint from each other[6]. This property can be generalized to *broadcast* as well. We only need to define the transmission range of a broadcast as the distance from the transmitter to the furthest node that can successfully receive the bit. The above property motivates us to measure the radio resources each transmission consumes by the areas of these disjoint disks [1]. For unicast transmissions from the last mobile relay to the destination, the area consumed by each hop is $\frac{\pi\Delta^2}{4}(S_b^h)^2$. For duplication to other nodes, broadcast is more beneficial since it consumes fewer resources. Assume that each transmitter chooses the transmission range of the broadcast independently of the positions of its neighboring nodes. If the transmission range is s, then on average no greater than $n\pi s^2$ nodes can receive the broadcast, and a disk of radius $\frac{\Delta}{2}s$ (i.e., area $\frac{\pi\Delta^2}{4}s^2$) centered at the transmitter will be disjoint from other disks. Therefore, we can use $\frac{\Delta^2}{4}\frac{\mathbf{E}[R_b]-1}{n}$ as a lower bound on the expected area consumed by duplicating the bit to $R_b - 1$ mobile relays (excluding the source node). This lower bound will hold even if the duplication process is carried out over multiple time slots, because the average number of *new* mobile relays each broadcast can cover is at most proportional to the area consumed by the broadcast. Therefore, inspired by [1], the amount of radio resources consumed must satisfy

$$\sum_{b=1}^{\lambda nT} \frac{\Delta^2}{4} \frac{\mathbf{E}[R_b]-1}{n} + \mathbf{E}[\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} \frac{\pi\Delta^2}{4} (S_b^h)^2] \leq c_3' WT,$$

where c_3' is a positive constant.

However, $\frac{\Delta^2}{4}\frac{\mathbf{E}[R_b]-1}{n}$ may fail to be a lower bound on the expected area consumed by duplicating to $R_b - 1$ mobile relays if the following *opportunistic broadcast scheme* is used. The source may choose to broadcast *only when there are a larger number of nodes close by*. If the source can afford to wait for these "good opportunities", an *opportunistic broadcast scheme* may consume less radio resources than a non-opportunistic scheme to duplicate the bit to the same number of mobile relays. Nonetheless, Proposition 3 shows that no scheduling policies can improve the tradeoff by more than a $\log n$ factor. For details, please refer to [8].

D. Tradeoff IV : Half Duplex

Finally, since we assume that no node can transmit and receive over the same frequency at the same time (a practically necessary assumption for most wireless devices), the following property can be shown as in [1].

Proposition 4: The following inequality holds,

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} 1 \leq \frac{WT}{2} n. \tag{7}$$

V. THE UPPER BOUND ON THE MAXIMUM ACHIEVABLE CAPACITY UNDER DELAY CONSTRAINTS

Our first main result is to derive from the above four tradeoffs the upper bound on the optimal capacity-delay tradeoff of mobile wireless networks under the *i.i.d.* mobility model. Since the maximal achievable per-node capacity is $\Theta(1)$ and this capacity can be achieved with $\Theta(n)$ delay by the scheme of [3], we are only interested in the case when the mean delay is $o(n)$.

Proposition 5: Let $\bar{D}$ be the mean delay averaged over all bits and all source-destination pairs, and let λ be the throughput of each source-destination pair. If $\bar{D} = O(n^d), 0 \leq d < 1$, the following upper bound holds for any causal scheduling policy under the *i.i.d.* mobility model,

$$\lambda^3 \leq O(\frac{\bar{D}}{n} \log^3 n). \tag{8}$$

Proof: Using Jensen's Inequality and Hőlder's Inequality [7, p14-15], we can reduce the four inequalities (4-7) to one inequality and eliminate all variables except $\bar{D}$ and λ. We can then obtain (8). For details, please refer to [9]. □

The capacity-delay tradeoff in Proposition 5 is better than those reported in [3] and [4]. Assuming that the delay bound

[6]A similar observation is used in [1] except that they take a receiver point of view.

is $\Theta(n^d)$, $0 \le d < 1$, the achievable per-node capacity is $O(n^{-(1-d)})$ by the scheme in [3], and $O(n^{-(1-d)/2})$ by the scheme in [4]. Our upper bound, however, implies a per-node capacity of $O(n^{-(1-d)/3})$ (we have ignored all $\log n$ factors). Since $d < 1$, there is clearly room to substantially improve existing schemes (see Fig. 1). In the next section, we will present a new scheme that can achieve the upper bound (8) up to a logrithmic factor.

VI. A LOWER BOUND ON THE ACHIEVABLE CAPACITY UNDER DELAY CONSTRAINTS

In this section, we will show how the study of the upper bound also helps us to develop a new scheme that can achieve a capacity-delay tradeoff that is close to the upper bound.

A. Choosing the Optimal Values of the Key Parameters

Assume that the mean delay is bounded by $n^d, d < 1$. By Proposition 5, we have,

$$\lambda \le \Theta(\sqrt[3]{\frac{\bar{D}}{n}\log^3 n}) = \Theta(n^{\frac{d-1}{3}}\log n). \tag{9}$$

In the process of proving the upper bound in Proposition 5, we obtain several inequalities (see [9]). In order to achieve the maximum capacity on the right hand side of (9), all of these inequalities need to hold with equality. By studying the conditions under which these inequalities are tight, we are able to identify the optimal choices of the key scheduling parameters (see [8] for the detail). In particular, we find that the scheduling parameters (such as $S_b^h, \mathbf{E}[l_b], \mathbf{E}[D_b]$) of each bit b should be about the same and should concentrate on their respective average values. This implies that the scheduling policy should use the same parameters for all bits. We can also obtain the optimal values of these parameters. The results are summarized in Table 1. The details of the derivation are available in our on-line technical report [8].

Several remarks are in order. Since it is sufficient to control all parameters around these optimal values, *simple cell-based schemes in Section III suffice*. Secondly, the optimal values for R_b and l_b can provide guidelines on how to choose the cell partitioning. Thirdly, the optimal value for S_b^h is roughly the average distance between neighboring nodes when n nodes are uniformly distributed in a unit square. Hence, it is desirable to use multi-hop transmission over neighboring nodes to forward the information from the last mobile relay to the destination. These guidelines have allowed us to sketch a blueprint of the optimal scheduling scheme. We next present schemes that can achieve capacity-delay tradeoffs that are close to the upper bound up to a logarithmic factor.

B. Achievable Capacity with $\Theta(n^d)$ Delay

We now present a scheme that can achieve the maximum per-node capacity in (9) up to a logarithmic factor. Our scheme can achieve $\Theta(n^{\frac{d-1}{3}}/\log^{3/2} n)$ per-node capacity with $\Theta(n^d)$ delay, $0 \le d < 1$. When $d = 0$, i.e., when the delay is bounded by a constant, our scheme can achieve $\Theta(n^{-1/3}/\log^{3/2} n)$ per-node capacity with $\Theta(1)$ delay[7]. This is an encouraging result for mobile networks because we know that the per-node capacity of static networks is $O(1/\sqrt{n\log n})$ [1]. Hence, for the *i.i.d.* mobility model, mobility increases the capacity even with constant delay. This is the first such result of its kind in the literature.

We will need the following Lemma before stating the main scheduling scheme. We will repeatedly use the following type of cell-partitioning. Let m be a positive integer. Divide the unit square into $m \times m$ cells (in m rows and m columns). Each cell is a square of area $1/m^2$. As in [4], we call two cells *neighbors* if they share a common boundary, and we call two nodes *neighbors* if they lie in the same or neighboring cells. We say that a group of cells can be *active* at the same time when one node in each active cell can successfully transmit to or receive from a neighboring node, subject to the interference from other cells that are active at the same time. Let $\lfloor x \rfloor$ be the largest integer smaller than or equal to x. The proof of the following Lemma is available in [8].

Lemma 6: There exists a scheduling policy such that each cell can be active for at least $1/c_4$ amount of time, where c_4 is a constant independent of m.

Group every $\lfloor n^d \rfloor$ time slots into a *super-frame*. The capacity-achieving scheme is as follows.

Capacity-Achieving Scheme:

1) In every *odd* super-frame (i.e., the 1st, 3rd, 5th, ... super-frame), we schedule transmissions from the sources to the relays. We will refer to the $\lfloor n^d \rfloor$ time slots in each odd super-frame as the *sending* time slots. At each sending time slot:

- We divide the unit square into $g_1(n) = \lfloor \left(\frac{n^{\frac{2+d}{3}}}{16\log n}\right)^{\frac{1}{2}} \rfloor^2$ *sending cells*. By Lemma 6, each cell can be active for $\frac{1}{c_4}$ amount of time in each sending time slot.
- When a cell is scheduled to be active, each node in the cell broadcasts a new packet to all other nodes in the same cell for $\frac{1}{4096c_4 n^{(1-d)/3}\log^{3/2} n}$ amount of time (Fig. 2). The length of the packet is $\frac{W}{4096c_4 n^{(1-d)/3}\log^{3/2} n}$. These other nodes then serve as mobile relays for the packet. The nodes within the same sending cell coordinate themselves to broadcast sequentially in each sending time slot. Note that on average there are $\mathbf{E}[R_b] = n/g_1(n) = \Theta(n^{(1-d)/3}\log n)$ mobile relays for each packet. Compared with the value in Table 1, the additional $\log n$ term for R_b is to improve the capture probability later in the even super-frame.
- If, in any of the $\lfloor n^d \rfloor$ sending time slots, there exists a sending cell that has more than $64n^{(1-d)/3}\log n$ nodes, we refer to it as a Type-I error [4]. If no Type-I errors occur, each source can broadcast a total of $\lfloor n^d \rfloor$ distinct packets, each of length $\frac{W}{4096c_4 n^{(1-d)/3}\log^{3/2} n}$, during the entire odd super-frame.

2) In every *even* super-frame (i.e., the 2nd, 4th, 6th, ... super-frame), we schedule transmissions from the mobile relays to the destination nodes. We will refer to the $\lfloor n^d \rfloor$ time slots in each even super-frame as the *receiving* time slots. At each receiving time slot:

- We divide the unit square into $g_2(n) = \lfloor \left(n^{\frac{1+2d}{3}}\right)^{\frac{1}{2}} \rfloor^2$ *receiving cells*. *Capture* occurs for each packet k when one of its mobile relays moves within the same receiving cell as the destination node (see Fig. 3). Note that the average distance from the last mobile relay to the destination is $\mathbf{E}[l_b] = \Theta(\sqrt{1/g_2(n)}) =$

[7] The scheme for $d = 0$ can be further refined to achieve $\Theta(n^{-1/3}/\log n)$ per-node capacity with $\Theta(1)$ delay [9].

Table 1. The order of the optimal values of the key scheduling parameters when the mean delay is bounded by n^d.

R_b: # of Duplicates	$\Theta(n^{(1-d)/3})$
l_b: Distance to Destination	$\Theta(n^{-(1+2d)/6}/\log^{1/2} n)$
h_b: # of Hops	$\Theta(n^{(1-d)/3}/\log n)$
S_b^h: Transmission Range of Each Hop	$\Theta(\sqrt{\frac{\log n}{n}})$

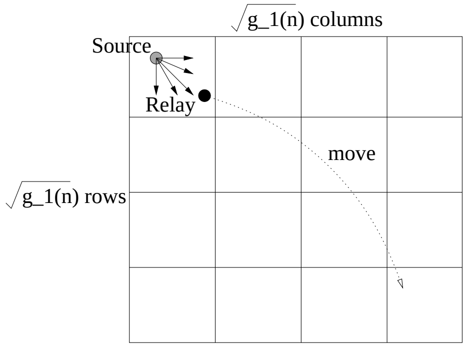

Fig. 2. Transmission schedule in each time slot of the odd super-frame. $g_1(n) = \lfloor \left(\frac{n^{\frac{2+d}{3}}}{16 \log n} \right)^{\frac{1}{2}} \rfloor^2$.

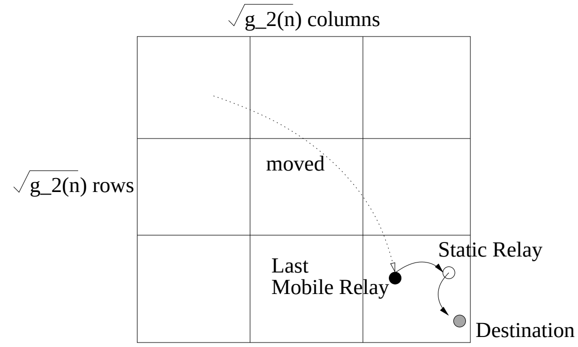

Fig. 3. Transmission schedule in each time slot of the even super-frame. $g_2(n) = \lfloor \left(n^{\frac{1+2d}{3}} \right)^{\frac{1}{2}} \rfloor^2$.

$\Theta(n^{-(1+2d)/6})$. Again, compared with the value in Table 1, the removal of the $\log n$ terms for l_b is to improve the capture probability. Let $\mathcal{Y}_j(t)$ denote the set of packets that meet the criteria for *capture* in receiving cell j at receiving time slot t. We will refer to these packets as the *active packets*. If a packet k does not meet the criteria for *capture* in any of the $\lfloor n^d \rfloor$ time slots, i.e., packet k does not belong to $\mathcal{Y}_j(t)$ for any $j = 1, ..., g_2(n)$ and $t = 1, ..., \lfloor n^d \rfloor$, we will refer to it as a Type-II error.

- The active packets in each set $\mathcal{Y}_j(t)$ are then forwarded to their destination nodes within the same time slot t in the following multi-hop fashion. We further divide the receiving cell j into $g_3(n) = \lfloor \left(\frac{n^{\frac{2-2d}{3}}}{8 \log n} \right)^{\frac{1}{2}} \rfloor^2$ *mini-cells*. By Lemma 6, there exists a scheduling scheme where each mini-cell can be active for $\frac{1}{c_4}$ amount of time in time slot t. When each mini-cell is active, it forwards an active packet (or a part of the packet) to one other node in the neighboring mini-cell. If the destination of the active packet is in the neighboring mini-cell, the packet is forwarded directly to the destination node. Note that the distance traveled by each hop is $S_b^h = \Theta(\sqrt{1/(g_2(n) g_3(n))}\ ) = \Theta(\sqrt{\log n / n})$, which is consistent with the value in Table 1. The active packets from each mobile relay are first forwarded towards neighboring mini-cells along the X-axis, then to their destination nodes along the Y-axis. In this fashion, a successful schedule will be able to deliver all active packets in $\mathcal{Y}_j(t)$ to their respective destination nodes (in the same receiving cell j) within the time slot t. For details on constructing such a schedule, see [8]. If no such schedule exists, we refer to it as a Type-III error.
- At the end of each even super-frame, any packets that remain in the buffer of the mobile relays (i.e., that have not been delivered to the destination nodes) are dropped.

If the packets that have already been delivered to the destination nodes are not immediately removed from the buffer of other mobile relays, it is possible that the above scheme delivers the same packet for more than once to the same destination node. In this case, we will assume that each packet has a sequence number so that the destination node can detect duplicate packets and only keep the new packets.

We can show that, as $n \to \infty$, the probabilities of errors of Type-I, II, and III, will all go to zero. The following proposition then holds, which shows that our scheme can achieve $\Theta(n^{(d-1)/3}/\log^{3/2} n)$ per-node capacity with $\Theta(n^d)$ delay. The proof is available in [8].

Proposition 7: Assume the *i.i.d.* mobility model. With probability approaching one, as $n \to \infty$, the above scheme allows each source to send $\lfloor n^d \rfloor$ packets of length $\frac{W}{4096 c_4 n^{(1-d)/3} \log^{3/2} n}$ to its respective destination node within $2\lfloor n^d \rfloor$ time slots.

Remark: Our scheme uses different cell-partitioning in the odd super-frames than that in the even super-frames. Note that in previous works [3], [4], the cell structure remains the same over all time slots. Our judicious choice of the cell-structures is the key to the development of a tighter lower bound for the capacity. As is discussed when we present the capacity-achieving scheme, we have chosen $g_1(n), g_2(n)$ and $g_3(n)$ such that the key scheduling parameters R_b, l_b and h_b are all close to their respective optimal values in Table 1.

Remark: We can incorporate the queueing delays into the above analysis. In fact, when we defined the delay D_b of each bit b in Section IV, it includes possible queueing delays at the source node and at the relay nodes. The upper bound on the capacity-delay tradeoff (Proposition 5) thus holds regardless of the queueing discipline used in the system, and $\bar{D}$ also includes the queueing delay. For the lower bound, provided that the offered load is at a fixed percentage below the capacity, we have shown in [8] that the queueing delay is again a constant multiple of n^d. Hence, our capacity-achieving scheme can sus-

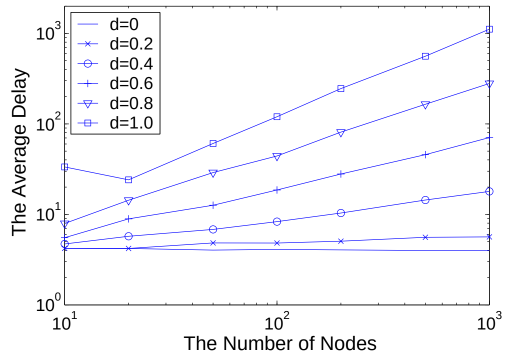

Fig. 4. Mean queueing delay (in number of time slots) versus the number of nodes.

tain $\Theta(n^{(d-1)/3}/\log^{3/2} n)$ per-node throughput (in bits per time slot) with $O(n^d)$ *queueing delay*. For details, please refer to [8].

C. Simulation Results

We have simulated the capacity-achieving scheme in Section VI-B for different values of the delay exponent d and the number of nodes n. In our simulation, we use the following Bernoulli packet arrival processes, i.e., one packet arrives at each source-destination pair every time-slot with probability p. The packet size scales as $\frac{W}{4096c_4 n^{(1-d)/3}\log^{3/2} n}$. Hence, for a fixed p, the offered load for each source-destination pair scales as $\Theta(n^{(d-1)/3}/\log^{3/2} n)$ (in bits per time slot). We use $p = 0.4$ in the simulation. Fig. 4 shows the mean queueing delay (in number of time slots) versus the number of nodes n at different values of d. We observe that, for large n, the mean queueing delay evolves as $\Theta(n^d)$ for all values of d. This figure demonstrates that our scheme can indeed sustain $\Theta(n^{(d-1)/3}/\log^{3/2} n)$ per-node throughput with $\Theta(n^d)$ delay.

VII. INCORPORATING ADDITIONAL SCHEDULING CONSTRAINTS

Our methodology can be extended to incorporate additional constraints in the scheduling policy. For example, if coordination among a large number of nodes is prohibited, i.e., each node is only allow to obtain the information about a constant number of other nodes in its close neighborhood, then the distance l_b from the last mobile relay to the destination node should be on the same order as the average distance between neighboring nodes, i.e., $\mathbf{E}[l_b] = O(1/\sqrt{n})$. Combining this new constraint with the four tradeoffs in Section IV, and using a similar argument as that of Proposition 5, we can establish the following new upper bound on the maximum achievable capacity under delay constraints.

Proposition 8: Let $\bar{D}$ be the mean delay averaged over all bits and all source-destination pairs, and let λ be the throughput of each source-destination pair. If $\bar{D} = O(n^d), 0 \le d < 1$ and $\mathbf{E}[l_b] = O(1/\sqrt{n})$, then for any causal scheduling policy under the *i.i.d.* mobility model,

$$\lambda \le O(\frac{\bar{D}}{n}\log^2 n).$$

The proof is available in [8]. The scheme of [3] achieves the above upper bound up to a logarithmic factor. Note that their scheme only allows coordination among nodes that are within a cell of area $1/n$. Hence, Proposition 8 shows that the restrictive choice of l_b is indeed the performance limiting factor of the scheme in [3].

Similarly, if coordination among more than $\Theta(1)$ nodes is allowed, however multi-hop transmissions are undesirable, then the number of hops h_b from the last mobile relay to the destination is of order $O(1)$. We can then obtain the following new capacity-delay tradeoff.

Proposition 9: Let $\bar{D}$ be the mean delay averaged over all bits and all source-destination pairs, and let λ be the throughput of each source-destination pair. If $\bar{D} = O(n^d), 0 \le d < 1$ and $h_b = O(1)$, then for any causal scheduling policy under the *i.i.d.* mobility model,

$$\lambda^2 \le O(\frac{\bar{D}}{n}\log^3 n).$$

The proof is available in [8]. The scheme of [4] achieves the above upper bound up to a logarithmic factor. Note that their scheme uses direct single-hop transmission from the last mobile relay to the destination (i.e., $h_b = 1$). Hence, Proposition 9 shows that the restriction on h_b is indeed the performance limiting factor of the scheme in [4].

VIII. THE RANDOM WAY-POINT MOBILITY MODEL

The analyses in the previous sections have focused on the *i.i.d.* mobility model. In this section, we will extend our methodology to the random way-point (*RWP*) mobility model. In previous sections, we have shown that, at least for the *i.i.d.* mobility model, there is not a significant loss of generality by using cell-based schemes. Indeed, by choosing the appropriate cell-partitioning, we have been able to develop cell-based schemes that asymptotically achieve the maximum capacity under given delay constraints. Hence, in this section, we will restrict our attention to cell-based schemes only, and our focus will be to find the optimal cell-partitioning (i.e., the values of $g_1(n)$ and $g_2(n)$) for the *RWP* mobility model. However, in the *RWP* mobility model, the nodes move continuously instead of in time slots. Hence, we need to modify the cell-based scheme as follows. We still divide the time into slots of unit length. After a bit b enters the system, it is broadcast to all other nodes in the same sending cell by the end of the next time slot. Let R_b be the number of mobile relays that receive the bit b. After certain delay D_b, one of the mobile relays (i.e., the *last mobile relay*) moves into the same receiving cell (of area A_b) as the destination node of bit b. The bit b is then forwarded from the last mobile relay to the destination by the end of the next time slot. Since the velocity of the nodes is $v(n) = \Theta(1/\sqrt{n})$, the distance any node can move within one time slot is of order $\Theta(1/\sqrt{n})$, which is small compared to the sizes of the sending cells and the receiving cells that we will choose later. Hence, the mobility of the nodes will not interfere much with both the duplication of the bit at the very beginning and the multi-hop forwarding after capture. Let h_b be the number of hops that bit b takes from the last mobile relay to the destination.

With the above modification of the cell-based scheme, the Tradeoffs II, III, and IV can be readily extended to the *RWP* mobility model. However, the exact counterpart to Tradeoff I is quite difficult to obtain analytically. We instead use numerical methods to study the likely form of Tradeoff I under the *RWP* mobility model. In the cell-based scheme, the number of mobile relays for each bit b is determined at the beginning of the duplication process, and all of these mobile relays were close to the source node of bit b when they received bit b. Hence, we can use the following simple simulation model to study the tradeoff between D_b, R_b and A_b. At time $t = 0$, we put one (destination) node at a random position uniformly distributed within the unit square. We put R_b mobile relays at another random position. Let the size of the receiving cell be $A_b = 1/g_2(n)$. We then let these nodes move according to the *RWP* mobility model and record the mean delay $\mathbf{E}[D_b]$ (averaged over simulation runs) before any one of the R_b mobile relays moves within the same receiving cell as the destination node. Varying R_b and A_b, we can thus obtain the relationship between $\mathbf{E}[D_b]$, R_b and A_b. However, note that we are only interested in the relationship when n is large. In order to extract the most useful information, we let $R_b = n^{d_1}, 0 < d_1 < 1$, and $A_b = n^{-d_2}, 0 < d_2 < 1$. With any fixed d_1 and d_2, we observe from our simulations that, when n is large, $\frac{\log \mathbf{E}[D_b]}{\log n}$ will converge to a number d, i.e., the delay is approximately n^d. In Fig. 5, we plot the relationship between d, d_1, and d_2 for the *RWP* mobility model. It is instructive to compare with the same plot obtained under the *i.i.d.* mobility model (Fig. 6). Note that each line in Fig. 6 can be expressed as

$$d = d_2 - d_1, \text{ for } d \geq 0,$$

which is consistent with (4) noting that $l_b = \Theta(\sqrt{A_b})$. On the other hand, each line in Fig. 5 can be expressed as

$$d = \frac{1 + d_2}{2} - d_1, \text{ for } 0.5 < d < 1,$$

which corresponds to

$$\mathbf{E}[D_b] \approx \Theta(\frac{n^{\frac{1}{2}}}{\mathbf{E}[l_b]\mathbf{E}[R_b]}). \quad (10)$$

This relationship between D_b, l_b and R_b under the *RWP* mobility model is consistent with the findings in [6]. When l_b and R_b are fixed, it has been shown in [6] that a given mobile relay can move within a distance l_b from the destination node during a single trip with probability $\Theta(l_b)$. Since odd trips are independent from each other, the expected number of trips for any of the R_b mobile relays to move within distance l_b from the destination node is $\Theta(\frac{1}{l_b R_b})$. Finally, as $v(n) = \Theta(1/\sqrt{n})$, each trip will take $\Theta(\sqrt{n})$ amount of time. We then obtain (10). However, this relationship only holds when $d \geq 0.5$. In fact, it has been shown in [6] that, in order to take advantage of mobility, the minimum amount of delay under the *RWP* mobility model is $\Theta(\sqrt{n})$.

Combining relationship (10) with Tradeoffs II, III and IV, we can compute the upper bound on the maximum capacity under given delay constraints (as in Proposition 5), and we can also identify the optimal values of the parameters R_b, A_b and h_b for achieving the upper bound. Due to space constraints, we state the results directly (for details, please refer to [8]). We obtain the following upper bound on the capacity-delay tradeoff under the *RWP* mobility model:

$$\lambda^2 \leq \Theta(\frac{\bar{D}}{n} \log^{\frac{3}{2}} n).$$

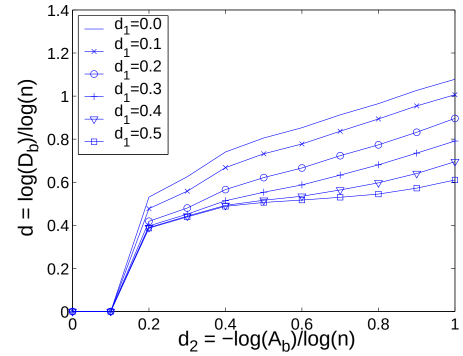

Fig. 5. Delay exponent d versus d_1 and d_2 for the random way-point mobility model.

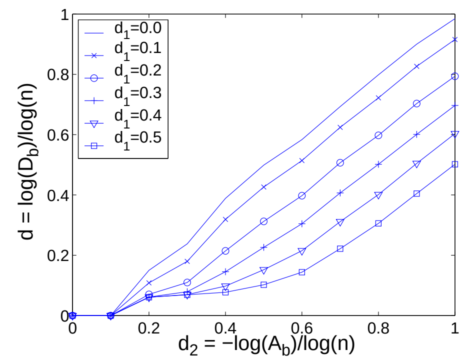

Fig. 6. Delay exponent d versus d_1 and d_2 for the *i.i.d.* mobility model.

Further, when the delay constraint is $D_b = O(n^d)$, the optimal values of the scheduling parameters are

$$R_b = \Theta(n^{\frac{1-d}{2}} \log^{\frac{1}{4}} n), \; l_b = \Theta(n^{-\frac{d}{2}} / \log^{\frac{1}{4}} n),$$
$$h_b = \Theta(n^{\frac{1-d}{2}} / \log^{\frac{3}{4}} n), \text{ and } S_b^h = \Theta(\sqrt{\frac{\log n}{n}}).$$

Hence, to use the cell-based capacity-achieving scheme as in Section VI, the number of sending cells and receiving cells should be $g_1(n) = \Theta(\frac{n^{\frac{1+d}{2}}}{\log n})$ and $g_2(n) = \Theta(n^d)$, respectively. We have simulated this cell-based scheme under the *RWP* mobility model and find it to achieve the following capacity-delay tradeoff

$$\lambda^2 \geq \Theta(\frac{\bar{D}}{n} / \log^2 n) \text{ when } 0.5 < d < 1.$$

Note that this capacity-delay tradeoff is better than the tradeoff reported in earlier studies [6]. Analogous to Section VII, we can show that a restrictive choice of the receiving cell size is again the performance limiting factor of the scheme in [6].

IX. CONCLUSION AND FUTURE WORK

In this paper, we have presented a systematic methodology for studying the maximum achievable capacity under given delay constraints for large mobile wireless networks. We first identify a number of key parameters of a large class of scheduling schemes, and investigate the inherent tradeoffs among the capacity, the delay, and these scheduling parameters. Based on these inherent tradeoffs, we are able to compute the upper bound on the maximum per-node capacity of a large mobile wireless network under given delay constraints, and develop scheduling schemes that achieve the upper bound up to some logarithmic factor. We have applied this methodology to both the *i.i.d.* mobility model and the random way-point mobility model. In both cases, our methodology sheds important insights on the optimal values of the key scheduling parameters, and allows us to develop new schemes that can achieve larger capacity than previous proposals under the same delay constraints. In particular, for the *i.i.d.* mobility model, our scheme can achieve $\Theta(n^{-1/3}/\log^{3/2} n)$ per-node capacity with constant delay. This demonstrates that, under the *i.i.d.* mobility model, mobility increases the capacity even with constant delays. Our methodology can also be extended to incorporate additional scheduling constraints.

In this paper, our treatment of the *i.i.d.* mobility model is purely analytical and rigorous. On the other hand, the treatment of the random way-point mobility model uses a combination of analytical and numerical techniques. For future work, we plan to investigate how to analytically establish the Tradeoff I for the random way-point mobility model as well. We also plan to investigate other mobility models, such as the Brownian motion mobility model [10], [11], and the linear mobility model [12].

ACKNOWLEDGMENTS

This work has been partially supported by the NSF grants ANI-0207728, EIA-0130599 and the Indiana 21st Century Fund for Wireless Networks.

REFERENCES

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, March 2000.

[2] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, August 2002.

[3] M. J. Neely and E. Modiano, "Capacity and Delay Tradeoffs for Ad-Hoc Mobile Networks," *submitted to IEEE Transactions on Information Theory, available at http://www-rcf.usc.edu/∼mjneely/*, 2003.

[4] S. Toumpis and A. J. Goldsmith, "Large Wireless Networks under Fading, Mobility, and Delay Constraints," in *Proceedings of IEEE INFOCOM*, (Hong Kong, China), March 2004.

[5] N. Bansal and Z. Liu, "Capacity, Delay and Mobility in Wireless Ad-Hoc Networks," in *Proceedings of IEEE INFOCOM*, (San Francisco, CA), April 2003.

[6] G. Sharma and R. R. Mazumdar, "Delay and Capacity Tradeoffs for Wireless Ad Hoc Networks with Random Mobility," *preprint available at http://www.ece.purdue.edu/∼mazum/*, October 2003.

[7] R. Durrett, *Probability : Theory and Examples*. Belmont, CA: Duxbury Press, second ed., 1996.

[8] X. Lin and N. B. Shroff, " The Fundamental Capacity-Delay Tradeoff in Large Mobile Ad Hoc Networks," *Technical Report, Purdue University, http://min.ecn.purdue.edu/∼linx/papers.html*, 2004.

[9] X. Lin and N. B. Shroff, "The Fundamental Capacity-Delay Tradeoff in Large Mobile Ad Hoc Networks," in *Third Annual Mediterranean Ad Hoc Networking Workshop*, (Bodrum, Turkey), June 2004.

[10] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-off in Wireless Networks," in *Proceedings of IEEE INFOCOM*, (Hong Kong, China), March 2004.

[11] G. Sharma and R. R. Mazumdar, "On Achievable Delay/Capacity Tradeoffs in Mobile Ad Hoc Networks ," in *Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WIOPT)*, (Cambridge, UK), March 2004.

[12] S. Diggavi, M. Grossglauser, and D. Tse, "Even One-Dimensional Mobility Increases Ad Hoc Wireless Capacity," in *ISIT 02*, (Lausanne, Switzerland), June 2002.

Xiaojun Lin received his B.S. in Electronics and Information Systems from Zhongshan University, Guangzhou, China, in 1994, and his M.S. in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, in 2000.
From 1995 to 1998, he worked as an I/T Specialist in IBM China. In 2002, he spent three months in IBM T.J. Watson Research Labs. Since 2000, he has been working towards the Ph.D. degree in Electrical and Computer Engineering at Purdue University, West Lafayette, IN.
His research interests are resource allocation, optimization, network pricing, routing, congestion control, network as a large system, cross-layer design in wireless networks, mobile ad hoc and sensor networks.

Ness B. Shroff received his Ph.D. degree from Columbia University, NY in 1994. He has been at Purdue University since 1994, and is currently a Professor of Electrical and Computer Engineering at Purdue.
His research interests span the areas of wireless and wireline communication networks. He is especially interested in fundamental problems in the design, performance, control, security, and pricing of these networks. His research has been funded by various companies such as Intel, Hewlett Packard, Nortel, AT&T, BAE systems, and L.G. Electronics; and government agencies such as the National Science Foundation, DARPA, Indiana Dept. of Transportation, and the Indiana 21st Century fund.
Dr. Shroff is an editor for the IEEE/ACM Trans. on Networking and the Computer Networks Journal, and past editor of IEEE Communications Letters. He was the Technical Program co-chair for IEEE INFOCOM'03 (San Francisco, CA), the panel co-chair for ACM Mobicom'02 (Atlanta, GA), program co-chair for the symposium on high-speed networks, Globecom 2001 (San Francisco, CA), and conference chair for the 14th Annual IEEE Computer Communications Workshop (Estes Park, CO). He was the co-organizer of the NSF Workshop on "Fundamental Research in Networking" in April 2003. He received the NSF Career award in 1996 and the Computer Network Journal's best paper award of 2003.