The Fundamental Capacity-Delay Tradeoff in Large Mobile Ad Hoc Networks

Xiaojun Lin and Ness B. Shroff
School of Electrical and Computer Engineering, Purdue University
West Lafayette, IN 47907, U.S.A.
{linx, shroff}@ecn.purdue.edu *

Abstract

There has been recent interest within the networking research community to understand how mobility can improve the capacity of mobile ad hoc networks. Of particular interest is the achievable capacity under delay constraints. In this paper, we establish the following upper bound on the optimal capacity-delay tradeoff in mobile ad hoc networks for an *i.i.d.* mobility model. For a mobile ad hoc network with n nodes, if the per-bit-averaged mean delay is bounded by $\bar{D}$, then the per-node capacity λ is upper bounded by $\lambda^3 \leq O(\frac{\bar{D}}{n}\log^3 n)$. By studying the condition under which the upper bound is tight, we are able to identify the optimal values of several key scheduling parameters. We then develop a new scheme that can achieve a capacity-delay tradeoff close to the upper bound up to a logarithmic factor. Our new scheme achieves a larger per-node capacity than the schemes reported in previous works. In particular, when the delay is bounded by a constant, our scheme achieves a per-node capacity of $\Theta(n^{-1/3}/\log n)$. This indicates that, for the *i.i.d.* mobility model, *mobility*

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

results in a larger capacity than that of static networks even with constant delays. Finally, the insight drawn from the upper bound allows us to identify limiting factors in existing schemes. These results present a relatively complete picture of the achievable capacity-delay tradeoffs under different settings.

1 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 ad hoc networks. For a static network (where nodes do not move), Gupta and Kumar show that the per-node capacity decreases as $O(1/\sqrt{n \log n})$[*] as the number of nodes n increases [1]. The capacity of wireless ad hoc 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 ad hoc networks. In the work by Neely and Modiano [3], it was shown that the maximum achievable per-node capacity of a mobile ad hoc network is bounded by $O(1)$. Under an *i.i.d.* mobility model, the authors of [3] present a

[*] We use the following notation throughout:

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

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 ad hoc 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 6). 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 in [4] the following capacity-delay tradeoff is achievable:

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

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 ad hoc networks.

An open question that still remains is: *what is the optimal capacity-delay tradeoff in mobile ad hoc networks*? Inequality (1) is clearly not optimal. The methodology of [4] is constructive in nature. Hence, inequality (2) is only a lower bound. The search for the optimal capacity-delay tradeoff is important for two reasons. First, it will allow us to see where the fundamental limits

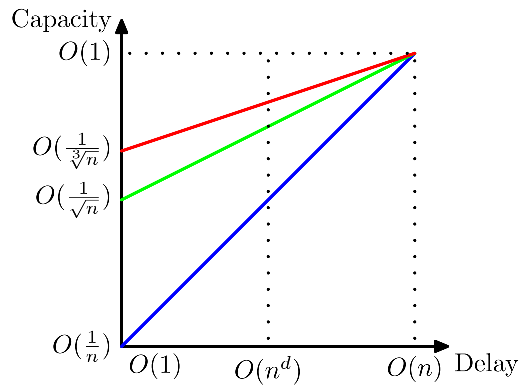

Figure 1: The achievable capacity-delay tradeoffs of existing schemes compared with the upper bound (ignoring the logarithmic terms). The top line corresponds to our upper bound (achievable by the scheme outlined in Section 5 up to a logarithmic factor). The middle line is achieved by the scheme in [4], and the bottom one is achieved by the scheme in [3].

(i.e., upper bounds) are, and how far existing schemes could possibly be improved. Secondly, as has happened in previous works [1, 3], a careful study of the upper bound is usually able to reveal the delicate tradeoffs inherent to the problem. A complete understanding of these tradeoffs will help us identify the possible points of inefficiency in existing schemes and provide directions for further improvement. The ultimate goal is to find a scheme that can achieve the optimal capacity-delay tradeoff.

This paper accomplishes these two goals. Under the *i.i.d.* mobility model studied in [3], we will first establish an *upper bound* on the optimal capacity-delay tradeoff in mobile ad hoc networks. We will show that, if the per-bit-averaged mean delay is bounded by $\bar{D}$, then the per-node capacity λ is upper bounded by

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

In Fig. 1, we draw this upper bound alongside the capacity-delay tradeoffs achieved by the schemes in [3] and [4]. There is obviously a gap between the upper bound and what can be achieved by existing schemes.

Further, in the process of proving the upper bound, we are able to identify the optimal choices for several key parameters of the scheduling policy. We then develop a new scheme that achieves the upper bound on the capacity-delay tradeoff upto a logarithmic factor, which suggests that our upper bound is fairly tight. Our new scheme achieves a larger per-node capacity than the ones in [3] and [4]. In particular, our scheme can achieve $\Theta(n^{-1/3}/\log n)$ per-node capacity with constant delay. Unlike previous works, this result shows that, even for a constant delay bound, the per-node capacity of mobile ad hoc networks can be larger than that of the static networks! Finally, the insight drawn from the upper bound allows us to identify the limiting factors of the schemes in [3] and [4].

The rest of the paper is organized as follows. In Section 2, we outline the network and mobility model. In Section 3, we prove several key properties that capture various tradeoffs inherent in mobile ad hoc networks. We establish the upper bound on the optimal capacity-delay tradeoff in Section 4 and present a scheme in Section 5 that achieves a capacity-delay tradeoff close to the upper bound. In Section 6, we discuss the existing schemes described in [3] and [4]. Then we conclude.

2 Network and Mobility Model

We consider a mobile ad hoc network with n nodes moving within a unit square[†]. We assume that time is divided into slots of unit length. We assume the following *i.i.d. mobility model* proposed in [3]. At each time slot, the positions of each node are *i.i.d.* and uniformly distributed within the unit square. Between time slots, the distributions of the positions of the nodes are independent. Although the assumption on an *i.i.d.* mobility model is somewhat restrictive, its mathematical tractability allows us to gain important insights into the structure of the problem. We will comment on some extensions to the *i.i.d.* mobility model in the conclusion.

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

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 each time slot, 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 further assume the following *separation of time scale*, 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 multiple hops separately within a single time slot.

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 $\bar{D}$.

3 Properties of the Scheduling Policies

In this section, we will prove several key results that capture the various tradeoffs inherent in mobile ad hoc networks. We will first define the class of scheduling policies that we will consider. Because we are interested in the fundamental *achievable* capacity for a given delay, 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. It is important to forward the bit to the destination within a single time slot. Otherwise, since the chosen relay node may move 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.

- *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.

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 is 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.

The following examples are illustrative of the possible scheduling policies within this broad class. The schemes in previous works [3, 4] are all special cases or variants of these examples.

Example A: The number of mobile relays R is fixed and the capture neighborhood is chosen to be a disk with a fixed radius ρ around the destination. Once a bit b enters the system, it is immediately broadcast to the nearest $R-1$ neighboring nodes. When any of the R mobile relays (including the source node) move within distance ρ from the destination, the bit b is then forwarded from the nearest mobile relay to the destination.

Example B: The unit area is divided into a number of cells. Once a bit b enters the system, it is immediately broadcast to all other nodes in the same cell. The number of mobile relays for the bit b then stay unchanged. Note that the actual number of mobile relays depends on the number of nodes that reside in the same cell as the source (at the time slot when the bit b enters the system), and is thus a random variable. When one of the mobile relays moves into the same cell as the destination, the bit b is then forwarded from the nearest mobile relay to the destination.

Example C: In the above two schemes, no duplication for bit b is carried out except at the first time slot when the bit enters the system. A more sophisticated strategy is to use an "opportunistic duplication scheme" such as the example below. The unit area is divided into a number of cells. After a bit b enters the system, at each time slot t, if one of the mobile relays moves into the same cell as the destination, bit b is then forwarded from the nearest mobile relay to the destination. Otherwise, the source node (or, alternatively, the current mobile relays) broadcasts the bit to all other nodes that reside at the same cell. Hence, duplication may occur at each time slot until bit b is delivered to its destination.

In the sequel, we will prove several key inequalities that capture the various tradeoffs inherent in this broad class of scheduling policies. 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. As we will see later, these tradeoffs will determine the fundamental relationship between achievable capacity and delay in mobile ad hoc networks.

3.1 Notations

Let $(\Omega, \mathcal{F}, P)$ be the probability space on which the random mobility of the mobile nodes is defined. Let $X(i,t)$ be the random variable that denotes the position of node i at time slot t. Let b denote a bit that needs to be communicated from a source node $S(b)$ to destination node

$D(b)$. Let $t_0(b)$ be the time slot when bit b first enters the system. Let $I_b(i,t)$ be an indicator function, where $I_b(i,t) = 1$ if node i has a copy of bit b at the beginning of time slot t, $I_b(i,t) = 0$ otherwise. By definition, $I_b(S(b), t_0(b)) = 1$, and $I_b(i,t) = 0$ for all i and $t < t_0(b)$. Let $\mathcal{F}_t$ be the σ-algebra generated by the random variables $X(i,s)$ and $I_b(i,s)$ for all $s \le t$. Hence $\{\mathcal{F}_t, t = 0, 1, ...\}$ is a filtration [5, p231] and $\mathcal{F}_t$ captures all information about the "history" up to time slot t.

Fix any scheduling policy and fix a bit b that enters the system at time slot $t_0(b)$. For any time slot $t \ge t_0(b)$, let $C_b(t) = 1$ if the scheduler decides that a successful capture occurs at this time slot. $C_b(t) = 0$, otherwise. If $C_b(t) = 1$, the scheduler then picks one mobile relay that has a copy of the bit b at the beginning of the time slot to forward the bit towards the destination *within the same time slot* t, using possibly multi-hop transmissions. Let $\tilde{l}_b(t)$ be the distance from the chosen mobile relay to the destination of the bit b. Let $\tilde{l}_b(t) = \infty$ if $C_b(t) = 0$. Finally, let $r_b(t+1)$ denote the number of mobile relays holding the bit b *at the end* of the time slot t, i.e., $r_b(t+1)$ is the cardinality of the set $\{i : I_b(i, t+1) = 1\}$. Since the random variables $C_b(t), \tilde{l}_b(t)$ and $r_b(t+1)$ are all outcomes of the scheduling policy, the causality assumption implies that they are all $\mathcal{F}_t$-measurable[‡].

Let

$$s_b \triangleq \min\{t : t \ge t_0(b) \text{ and } C_b(t) = 1\}$$

be the first time when a successful capture for bit b occurs. Thus s_b is a stopping time [5, p234] with respect to the filtration $\{\mathcal{F}_t, t = 0, 1, ...\}$. Let $R_b \triangleq r_b(s_b)$ denote the number of mobile relays holding the bit b at the time of capture. Let $D_b \triangleq s_b - t_0(b)$ denote the number of time slots from the time bit b enters the system to the time of capture. Let $l_b \triangleq \tilde{l}_b(s_b)$ denote the distance from the chosen last mobile relay node to the destination. The quantities R_b, D_b, and l_b are essential for the tradeoffs that follow. Note that D_b includes possible queueing delay at the

[‡]Here we have excluded *probablistic* scheduling policies. Otherwise, $\mathcal{F}_t$ should be augmented with a σ-algebra that is independent of node mobility in future time slots.

source node or at the relay nodes.

3.2 Tradeoff I : D_b versus R_b and l_b

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

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

where c_1 is a positive constant.

The proof is available in Appendix A. This novel result is one of the cornerstones for deriving the optimal capacity-delay tradeoff in mobile ad hoc networks. It captures the following tradeoff: the smaller the number R_b of mobile relays the 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. This seemingly odd relationship is actually motivated by some simple examples. Consider Example A at the beginning of Section 3. 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,

$$\mathbf{E}[D_b] = \frac{1}{1 - (1 - A_b)^{R_b}} \geq \frac{1}{A_b R_b}.$$

If, as in Example B, R_b and possibly A_b are random but *fixed after the first time slot* $t_0(b)$, then

$$\mathbf{E}[D_b | R_b, A_b] \geq \frac{1}{A_b R_b}.$$

By Hőlder's Inequality [5, p15],

$$\mathbf{E}^2[\frac{1}{\sqrt{A_b}}] \leq \mathbf{E}[R_b] \mathbf{E}[\frac{1}{A_b R_b}].$$

Hence,

$$\begin{aligned} \mathbf{E}[D_b] &\geq \mathbf{E}[\frac{1}{A_b R_b}] \geq \mathbf{E}^2[\frac{1}{\sqrt{A_b}}] \frac{1}{\mathbf{E}[R_b]} \\ &\geq \frac{1}{\mathbf{E}^2[\sqrt{A_b}] \mathbf{E}[R_b]}, \end{aligned}$$

where in the last step we have applied Jensen's Inequality [5, p14]. Note that on average l_b is on the order of $\sqrt{A_b}$. Hence,

$$\mathbf{E}[D_b] \geq \frac{c_1'}{\mathbf{E}^2[l_b]\mathbf{E}[R_b]} \text{ for all bits } b, \tag{5}$$

where c_1' is a positive constant. It may appear that, when an "opportunistic duplication scheme" such as the one in Example C is employed, such a scheme might achieve a better tradeoff than (5) by starting off with fewer mobile relays and a smaller capture neighborhood, if the node positions at the early time slots after the bit's arrival turns out to be favorable. However, Proposition 1 shows that no scheduling policy can improve the tradeoff by more than a $\log n$ factor. For details, please refer to Appendix A.

3.3 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.,*

$$\sum_{h=1}^{h_b} S_b^h \geq l_b. \tag{6}$$

3.4 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 $D_b \le c_2 n^2$ for $n \ge N_0$. If the positions of 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 $n \ge N_1$,*

$$\sum_{b=1}^{\lambda n T} \frac{\Delta^2}{4} \frac{\mathbf{E}[R_b] - 1}{n} + \mathbf{E}[\sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} \frac{\pi \Delta^2}{4} (S_b^h)^2] \le c_3 W T \log n. \tag{7}$$

The assumption that $D_b \le c_2 n^2$ for large n is not as restrictive as it appears. It has been shown in [3] that the maximal achievable per-node capacity is $\Theta(1)$ and this capacity can be achieved with $\Theta(n)$ delay. 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.

We briefly outline the motivation behind the inequality (7). The details of the proof are quite technical and available in Appendix B. Consider nodes i, j that directly transmit to nodes k and l, respectively, at the same time. Then, according to the interference constraint:

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

Hence,

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

Similarly,

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

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[§]. 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 $\frac{\pi s^2}{n}$ 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

$$\sum_{b=1}^{\lambda nT} \frac{\Delta^2}{4}\frac{\mathbf{E}[R_b]-1}{n} + \mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}\frac{\pi\Delta^2}{4}(S_b^h)^2] \leq c_3' WT, \tag{8}$$

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

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

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 Appendix B.

3.5 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,*

$$\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} 1 \leq \frac{WT}{2} n. \tag{9}$$

4 The Upper Bound on the Capacity-Delay Tradeoff

Our first main result is to derive, from the above four tradeoffs, the upper bound on the optimal capacity-delay tradeoff of mobile ad hoc 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,*

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

Proof: Using the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \left(\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} S_b^h \right)^2 &\leq \left(\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} 1 \right) \left(\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} (S_b^h)^2 \right) \\ &\leq \frac{WTn}{2} \sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} (S_b^h)^2, \end{aligned} \tag{10}$$

where in the last step we have used Tradeoff IV (9). Equality holds in (10) when inequality (9) is tight and when S_b^h is equal for all b and h. We thus have,

$$\begin{aligned}\mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}(S_b^h)^2] &\geq \frac{2}{WTn}\mathbf{E}[\left(\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}S_b^h\right)^2] \\ &\geq \frac{2}{WTn}\left(\mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}S_b^h]\right)^2 \qquad (11)\\ &\geq \frac{2}{WTn}\left(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]\right)^2, \qquad (12)\end{aligned}$$

where in the last two steps we have used Jensen's Inequality and the Tradeoff II (6), respectively. Inequality (11) is tight when $\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}S_b^h$ is almost surely a constant, and (12) is tight when (6) is tight.

From Tradeoff I (4), we have

$$\sum_{b=1}^{\lambda nT}\mathbf{E}[R_b] \geq \sum_{b=1}^{\lambda nT}\frac{1}{c_1\log n}\frac{1}{(\mathbf{E}[l_b]+\frac{1}{n^2})^2\mathbf{E}[D_b]}. \qquad (13)$$

Let

$$\bar{D} = \frac{\sum_{b=1}^{\lambda nT}\mathbf{E}[D_b]}{\sum_{b=1}^{\lambda nT}1} = \frac{\sum_{b=1}^{\lambda nT}\mathbf{E}[D_b]}{\lambda nT}.$$

Using Jensen's Inequality and Hőlder's Inequality, we have,

$$\begin{aligned}\frac{1}{\left(\frac{\sum_{b=1}^{\lambda nT}(\mathbf{E}[l_b]+\frac{1}{n^2})}{\sum_{b=1}^{\lambda nT}1}\right)^2} &\leq \left(\frac{\sum_{b=1}^{\lambda nT}\frac{1}{(\mathbf{E}[l_b]+\frac{1}{n^2})}}{\sum_{b=1}^{\lambda nT}1}\right)^2 \\ &\leq \frac{\sum_{b=1}^{\lambda nT}\frac{1}{(\mathbf{E}[l_b]+\frac{1}{n^2})^2\mathbf{E}[D_b]}}{\sum_{b=1}^{\lambda nT}1}\frac{\sum_{b=1}^{\lambda nT}\mathbf{E}[D_b]}{\sum_{b=1}^{\lambda nT}1}. \qquad (14)\end{aligned}$$

Equality holds when $\mathbf{E}[l_b]$ is the same for all b and $\mathbf{E}[D_b] = \bar{D}$ for all b. Substituting (14) in (13), we have

$$\sum_{b=1}^{\lambda nT} \mathbf{E}[R_b] \geq \frac{1}{c_1 \log n} \frac{\left(\sum_{b=1}^{\lambda nT} 1\right)^3}{\bar{D}\left(\sum_{b=1}^{\lambda nT} (\mathbf{E}[l_b] + \frac{1}{n^2})\right)^2}. \tag{15}$$

Substituting (12) and (15) into Inequality (7), we have

$$\begin{aligned}
\frac{4c_3 WT \log n}{\Delta^2} &\geq \sum_{b=1}^{\lambda nT} \frac{\mathbf{E}[R_b] - 1}{n} + \pi \mathbf{E}[\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} (S_b^h)^2] \\
&\geq \frac{1}{c_1 n \log n} \frac{(\lambda nT)^3}{\bar{D}\left(\sum_{b=1}^{\lambda nT} (\mathbf{E}[l_b] + \frac{1}{n^2})\right)^2} \\
&\quad + \frac{2\pi}{WTn} \left(\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]\right)^2 - \lambda T.
\end{aligned}$$

There are two cases that we need to consider.

Case 1: If $\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b] \leq \frac{\lambda T}{n}$, then

$$\begin{aligned}
\frac{4c_3 WT \log n}{\Delta^2} &\geq \frac{1}{c_1 n \log n} \frac{(\lambda nT)^3}{\bar{D}\left(\frac{2\lambda T}{n}\right)^2} - \lambda T \\
&= \frac{1}{4c_1 \log n} \frac{\lambda T n^4}{\bar{D}} - \lambda T.
\end{aligned}$$

When $\bar{D} = O(n^d), d < 1$, the first term dominates when n is large. Hence, for n large enough,

$$\begin{aligned}
\frac{4c_3 WT \log n}{\Delta^2} &\geq \frac{1}{8c_1 \log n} \frac{\lambda T n^4}{\bar{D}} \\
\lambda &\leq \frac{32 c_1 c_3 W}{\Delta^2} \frac{\bar{D} \log^2 n}{n^4}.
\end{aligned} \tag{16}$$

Case 2: If $\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b] \geq \frac{\lambda T}{n}$, then

$$\begin{aligned}
\frac{4c_3 WT \log n}{\Delta^2} &\geq \frac{1}{c_1 n \log n} \frac{(\lambda nT)^3}{\bar{D}\left(2 \sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]\right)^2} \\
&\quad + \frac{2\pi}{WTn} \left(\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]\right)^2 - \lambda T
\end{aligned} \tag{17}$$

$$\geq \quad 2\sqrt{\frac{1}{c_1 \log n}\frac{2\pi}{WTn^2}\frac{(\lambda nT)^3}{4\bar{D}}} - \lambda T \tag{18}$$

$$= \quad 2\sqrt{\frac{\pi}{2c_1 \log n}\frac{\lambda^3 nT^2}{\bar{D}W}} - \lambda T. \tag{19}$$

Therefore, either

$$\lambda \leq O(\frac{\bar{D}\log n}{n}), \tag{20}$$

or, if $\lambda = \omega(\frac{\bar{D}\log n}{n})$, then the first term in (19) dominates when n is large. In the latter case, for n large enough,

$$\begin{aligned}\frac{4c_3 WT\log n}{\Delta^2} \quad &\geq \quad \sqrt{\frac{\pi}{2c_1 \log n}\frac{\lambda^3 nT^2}{\bar{D}W}} \\ \lambda^3 \quad &\leq \quad \frac{32c_1c_3^2W^3}{\pi\Delta^4}\frac{\bar{D}\log^3 n}{n}.\end{aligned} \tag{21}$$

Finally, we compare the three inequalities we have obtained, i.e., (16), (20) and (21). Since $\bar{D} = o(n^d), d < 1$, inequality (21) will eventually be the loosest for large n. Hence, the optimal capacity-delay tradeoff is upper bounded by

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

Q.E.D.

5 An Achievable Lower Bound on the Capacity-Delay Tradeoff

The capacity-delay tradeoff in Proposition 5 is better than those reported in [3] and [4]. Assuming that the delay bound is $\Theta(n^d)$, $0 \leq d < 1$, 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 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. Precisely, we met several inequalities (10)-(18) during the derivation of the upper bound. By studying the conditions under which these inequalities are tight, we will be able to identify the optimal choices of various key parameters of the scheduling policy. In the end, the knowledge of the optimal choices of the parameters will help us develop a new scheme that is superior to existing ones.

5.1 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,

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

In order to achieve the maximum capacity on the right hand side, all inequalities (10)-(18) should hold with equality. By checking the conditions when (10)-(14) are tight, we can infer that the 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. From now on, we will assume that all key parameters (such as R_b, l_b, etc.) are indeed the same for all bits.

The inequality (18) is essential for deriving the optimal values of these parameters. Note that equality holds in (18) if and only if

$$\frac{1}{4c_1 n\log n}\frac{(\lambda nT)^3}{\bar{D}(\sum\limits_{b=1}^{\lambda nT}\mathbf{E}[l_b])^2} = \frac{2\pi}{WTn}(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b])^2.$$

Substituting $\sum\limits_{b=1}^{\lambda nT}\mathbf{E}[l_b] = \lambda nTl_b$, we can solve for l_b,

$$\begin{aligned}\frac{1}{4c_1 n\log n}\frac{\lambda nT}{\bar{D}l_b^2} &= \frac{2\pi}{WTn}(\lambda nT)^2 l_b^2 \\ l_b^4 &= \frac{1}{8\pi c_1}\frac{W}{\bar{D}\lambda n\log n}.\end{aligned}$$

Substituting $\lambda = \Theta(n^{(d-1)/3} \log n)$ and $\bar{D} = n^d$, we obtain the optimal value of l_b,

$$l_b = \Theta(n^{-\frac{1+2d}{6}} \log^{-\frac{1}{2}} n)$$

A reasonable choice for the area of capture neighborhood, A_b, is then,

$$A_b = l_b^2 = \Theta(n^{-\frac{1+2d}{3}} / \log n).$$

By setting (9) of Tradeoff IV to equality, we have

$$\begin{aligned} \lambda n T h_b &= \frac{WTn}{2} \\ h_b &= \frac{W}{2\lambda} = \Theta(n^{\frac{1-d}{3}} / \log n). \end{aligned}$$

By setting (6) of Tradeoff II to equality, we have

$$S_b^h = \frac{l_b}{h_b} = \Theta(\sqrt{\frac{\log n}{n}}).$$

From (4) of Tradeoff I , equality is attained when

$$R_b = \Theta(\frac{1}{c_1 \log n} \frac{1}{l_b^2 \bar{D}}) = \Theta(n^{\frac{1-d}{3}}).$$

The optimal values of these parameters are summarized in Table 1.

Several remarks are in order. Since it is sufficient to control all parameters around these optimal values, simple cell-based schemes such as the one in Example B of Section 3 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 sketched a blueprint of the optimal scheduling scheme for us. We next present schemes that can achieve capacity-delay tradeoffs that are close to the upper bound up to a logarithmic factor.

Table 1: The order of the optimal values of the 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	$\Theta(\sqrt{\frac{\log n}{n}})$

5.2 Achievable Capacity with $\Theta(1)$ Delay

We first focus on the case when the mean delay is bounded by a constant, i.e., the exponent $d = 0$. By Proposition 5, the per-node throughput is bounded by $O(n^{-1/3} \log n)$. We now present a scheme that can achieve $\Theta(n^{-1/3}/\log n)$ capacity with $\Theta(1)$ delay for large n. 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, mobility increases the capacity even with constant delay.

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, see Fig. 2). 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 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 Appendix C.

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.*

The capacity achieving scheme is as follows.

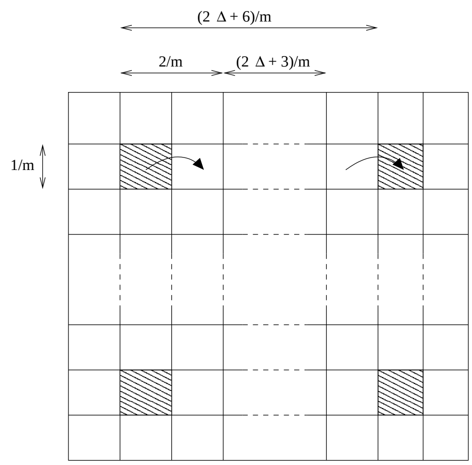

Figure 2: Cells that are $\lfloor 2\Delta + 6 \rfloor / m$ apart (i.e., the shaded cells in the figure) can be active together.

Capacity Achieving Scheme:

1) At each *odd* time slot, we schedule transmissions from the sources to the relays. We divide the unit square into $g_1(n) = \lfloor \left(\frac{n^{2/3}}{8 \log n} \right)^{\frac{1}{2}} \rfloor^2$ cells. Each cell is a square of area $1/g_1(n)$. We refer to each cell in the odd time slot as a *sending cell.* By Lemma 6, each cell can be active for $\frac{1}{c_4}$ amount of time. When a cell is scheduled to be active, each node in the cell broadcasts a new message to all other nodes in the same cell for $\frac{1}{32 c_4 n^{1/3} \log n}$ amount of time (Fig. 3). These other nodes then serve as mobile relays for the message. The nodes within the same sending cell coordinate themselves to broadcast sequentially. If any sending cell has more than $32 n^{1/3} \log n$ nodes, we refer to it as a Type-I error [4]. Unless a Type-I error occurs, each source can broadcast a message of length $\frac{W}{32 c_4 n^{1/3} \log n}$ to all other nodes in the same sending cell.

2) At each *even* time slot, we schedule transmissions from the mobile relays to the destination nodes. Note that the positions of the mobile relays have changed and are now independent of their positions in the previous time slot. We divide the unit square into $g_2(n) = \lfloor \left(n^{1/3} \right)^{\frac{1}{2}} \rfloor^2$ cells. Each cell is a square of area $1/g_2(n)$. We refer to each cell in the even time slot as the *receiving cell.* For any *receiving cell* $i = 1, ..., g_2(n)$ and any *sending cell* $j = 1, ..., g_1(n)$, pick a node Y_{ij} that is in the *receiving cell* i in the current time slot and that was in the *sending cell* j in the

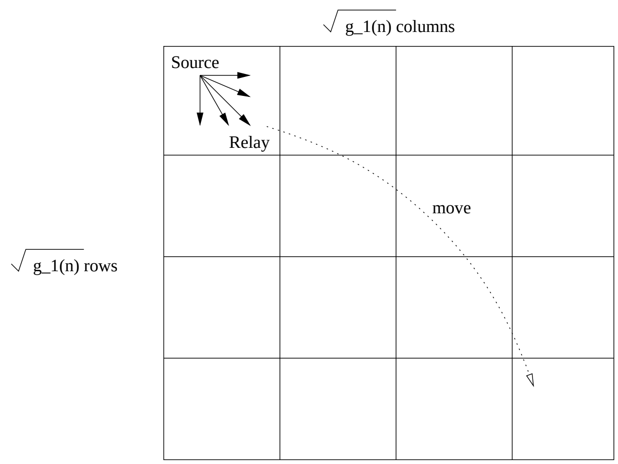

Figure 3: Transmission schedule in the odd time slot. $g_1(n) = \lfloor \left(\frac{n^{2/3}}{\log n} \right)^{\frac{1}{2}} \rfloor^2$.

previous time slot. We refer to this node Y_{ij} as the *designated mobile relay in* receiving cell i and *for* sending cell j. If there is no such node Y_{ij} for any i or j, we refer to it as a Type-II error. There may be multiple nodes that can serve as the designated mobile relay for some i, j. In this case we only pick one. Unless a Type-II error occurs, each receiving cell will contain one designated mobile relay from every sending cell. Therefore, each destination node can now find a designated mobile relay that holds the message intended for the destination node and that resides in the same receiving cell (see Fig. 4). We then schedule multi-hop transmissions in the following fashion to forward each message from the designated mobile relay to its destination in the same receiving cell. We further divide each receiving cell i into $g_3(n) = \lfloor \left(\frac{n^{2/3}}{4 \log n} \right)^{\frac{1}{2}} \rfloor^2$ *mini-cells* (in $\sqrt{g_3(n)}$ rows and $\sqrt{g_3(n)}$ columns, see Fig. 5). Each mini-cell is a square of area $1/(g_2(n)g_3(n))$. By Lemma 6, there exists a scheduling scheme where each mini-cell can be active for $\frac{1}{c_4}$ amount of time. When each mini-cell is active, it forwards a message (or a part of a message) to one other node in the neighboring mini-cell. If the destination of the message is in the neighboring cell, the message is forwarded directly to the destination node. The messages from each designated mobile relay are first forwarded towards neighboring cells along the X-axis, then to their destination nodes along the Y-axis (see Fig. 5). In this fashion, a successful schedule

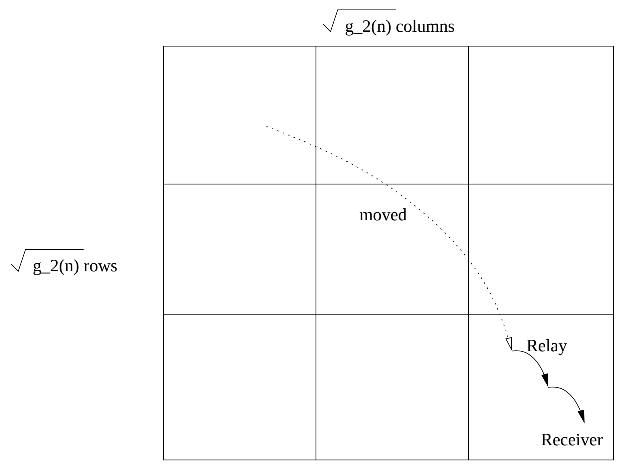

Figure 4: Transmission schedule in the even time slot. $g_2(n) = \lfloor \left(n^{1/3}\right)^{\frac{1}{2}} \rfloor^2$.

will allow each destination node to receive a message of length $\frac{W}{32c_4 n^{1/3}\log n}$ from its respective designated mobile relay residing in the same receiving cell. For details on constructing such a schedule, see Appendix D. If no such schedule exists, we refer to it as a Type-III error. At the end of each even time slot, if there are any packets that cannot be delivered to the destination nodes due to Type-II or Type-III errors, they are dropped.

We can show that, as $n \to \infty$, the probabilities of errors of all types will go to zero. The proof is available in Appendix D.

Proposition 7 *With probability approaching one, as $n \to \infty$, the above scheme allows each source to send a message of length $\frac{W}{32c_4 n^{1/3}\log n}$ to its respective destination node within two time slots.*

Remark: Our scheme uses different cell-partitioning in the odd time slots than that in the even time slots. 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 our tighter lower bound for the capacity. In particular, the size of the sending cell is chosen such that the average number of nodes in each cell, $n/g_1(n) = \Theta(n^{1/3}\log n)$, is close to the optimal value of R_b in Section 5.1 (with $d = 0$). The size of the receiving cell is chosen such that its area, $1/g_2(n) = \Theta(n^{1/3})$, is

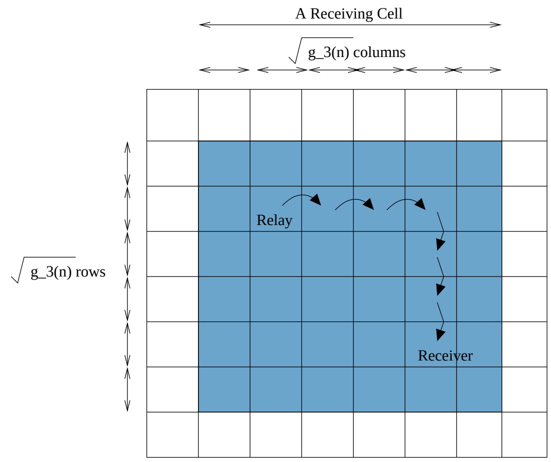

Figure 5: Multi-hop transmissions within a receiving cell.

close to the optimal value of l_b^2. Finally, the size of the mini-cell is chosen such that each hop to the neighboring cell is of length $1/\sqrt{g_2(n)g_3(n)} = \Theta(\sqrt{\log n/n})$, which is close to the optimal value of S_b^h.

5.3 The Effect of Queueing

When we defined the delay D_b of each bit b in Section 3, it includes the possible queueing delay 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. We now show how to analyze the queueing delay of the capacity-achieving scheme in Section 5.2. This scheme attempts to deliver one message of length $\frac{W}{32c_4 n^{1/3}\log n}$ for each source-destination pair every two time slots. Let p_{e} be the probability that a message is successfully delivered to the destination at the end of the even time slot. (Note that p_{e} is the same for all source-destination pairs due to symmetry, and by Proposition 7, $p_{\mathrm{e}} \to 1$ as $n \to \infty$.) Assume that, if such delivery is unsuccessful, messages that have not been delivered to the destinations at the end of each even time slot are discarded and have to be retransmitted at the source nodes. Further, assume that packets of length $\frac{W}{32c_4 n^{1/3}\log n}$ arrive at each source according to certain stochastic process. Then packets may get enqueued at the source nodes. If we observe the system at the end of each even time slot, the number of packets queued for each source-destination pair will evolve as that of a discrete-time queue with geometric service time distributions [6], and the queues for each source-destination pair can be studied independently. If we know the packet arrival process, we can then compute the queueing delay. For example, if the arrival process is Bernoulli, i.e., one new packet for each source-destination pair arrives at the source every two time slots with probability Λ, then using standard results for discrete time $M/M/1$ queues [6, p82], we can compute the queueing delay as,

$$\mathcal{D} = 2\frac{1-\Lambda}{p_{\mathrm{e}}-\Lambda}.$$

As $n \to \infty$, $p_{\mathrm{e}} \to 1$. Hence,

$$\mathcal{D} \to 2, \text{ as } n \to \infty.$$

On the other hand, if the arrival process is Poisson with rate Λ, then the number of packets arriving at a source-destination pair every two time slots is a Poisson random variable with mean 2Λ. Hence, using results for discrete time $M^{a_n}/M/1$ queues [6, p89], we can compute the queueing delay as

$$\mathcal{D} = 2\frac{1-\Lambda}{p_{\mathrm{e}}-2\Lambda}.$$

Assume $2\Lambda \leq 1-\epsilon$, where $0 < \epsilon < 1$. As $n \to \infty$, $p_{\mathrm{e}} \to 1$. Hence,

$$\mathcal{D} \to 2\frac{1-\Lambda}{\epsilon}, \text{ as } n \to \infty.$$

Note that in both cases, the queueing delay is at most a constant multiple of 2 (time slots) provided that ϵ (i.e., the difference between the arrival rate and the capacity) is positive and bounded away from zero as $n \to \infty$. Hence, the capacity-achieving scheme in Section 5.2 can sustain $\Theta(n^{-1/3}/\log n)$ throughput (in bits per time slot) with $O(1)$ *queueing delay.*

5.4 The Capacity Achieving Scheme for Arbitrary Delay Bound

The above scheme can be generalized to arbitrary delay bounds. Let the mean delay be bounded by $\bar{D} = \Theta(n^d)$, $0 \leq d < 1$. We can group every $\lfloor n^d \rfloor + 1$ time slots into a *super-frame*. In each *odd* super-frame, we schedule transmissions from the sources to the relays. We divide the unit square into $g_1(n) = \Theta(n^{(2+d)/3}/\log n)$ sending cells of equal area. Within each sending cell, each source broadcasts a new message to all other nodes within the same cell for a duration of $\Theta(\frac{1}{n^{(1-d)/3}\log^2 n})$ every time slot.

In each *even* super-frame, we schedule transmissions from the relays to the destination nodes. We divide the unit square into $g_2(n) = \Theta(n^{(1+2d)/3})$ receiving cells of equal area. In every time slot, some mobile relays will have messages intended for some other destination nodes in the same receiving cell. We then schedule multi-hop transmissions to deliver the messages from the mobile relays to the destination nodes in the same receiving cell.

Using similar techniques as the one in [4] and the one in Appendix D, we can show that, with probability approaching one as $n \to \infty$, each source can send $\lfloor n^d \rfloor + 1$ messages of length $\Theta(n^{-(1-d)/3}/\log^2 n)$ to its destination within $2(\lfloor n^d \rfloor + 1)$ time slots. The queueing delay can also be studied in a similar fashion as in Section 5.3. The details are omitted because of space constraints.

6 The Limiting Factors in Existing Schemes

In Section 5, we have shown that choosing the optimal values of the scheduling parameters is the key to achieve the optimal capacity-delay tradeoff. In this section, we will show that deviating from these optimal values will lead to suboptimal capacity-delay tradeoffs. In particular, we will identify the limiting factors in the existing schemes in [3] and [4] by comparing the optimal values of scheduling parameters in Section 5.1 with those used by the existing schemes. Our model in Section 4 can be extended to study the upper bounds on the capacity-delay tradeoff when one imposes additional restrictive assumptions that correspond to these limiting factors. We will see that these new upper bounds are inferior to the capacity-delay tradeoff reported in Sections 4 and 5. The existing schemes of [3] and [4] in fact achieve capacity-delay tradeoffs that are close to the respective upper bounds. These results will give us new insights on which schemes to use under different conditions.

6.1 The Limiting Factor in the Scheme of [3]

The scheme by Neely and Modiano [3] divides the unit square into n cells each of area $1/n$. A mobile relay will forward messages to the destination only when they both reside in the same cell. Hence, the distance from the last mobile relay to the destination, l_b, is on average on the order of $O(1/\sqrt{n})$, regardless of the delay constraints. However, we have shown in Section 5.1 that the optimal choice for l_b should be on the order of $\Theta(n^{-(1+2d)/6}\log^{-1/2} n)$, when the mean delay is bounded by $\Theta(n^d)$. The next Proposition shows that the restrictive choice of l_b is indeed the limiting factor of the scheme in [3]. The proof is available in Appendix E

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 $\bar{D} = O(n^d), 0 \leq d < 1$ and $\mathbf{E}[l_b] = O(1/\sqrt{n})$, then for any causal scheduling policy,*

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

Remark: The scheme of [3] achieves the above upper bound up to a logarithmic factor.

6.2 The Limiting Factor in the Scheme of [4]

In the scheme by Toumpis and Goldsmith [4], a mobile relay will always use single-hop transmission to forward the messages directly to the destination. That is, the number of hops from the last mobile relay to the destination node, h_b, is always 1. However, we have shown in Section 5.1 that the optimal value of h_b is $\Theta(n^{(1-d)/3}/\log n)$ when the mean delay is bounded by $\Theta(n^d)$. The next Proposition shows that the restriction on h_b is indeed the limiting factor of the scheme in [4]. The proof is available in Appendix F.

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,*

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

Remark: The scheme of [4] achieves the above upper bound up to a logarithmic factor.

Propositions 5, 8 and 9 present three different upper bounds on the capacity-delay tradeoff of mobile ad hoc networks under different assumptions. Assume that the mean delay is bounded by $n^d, 0 \le d < 1$. When the capacity is the main concern, Proposition 5 shows that the per-node throughput is at most $O(n^{-(1-d)/3} \log n)$. The capacity-achieving scheme reported in Section 5 can achieve close to this upper bound up to a logarithmic factor. However, this capacity-achieving scheme requires sophisticated coordination among the mobile nodes. Hence, it may not be suitable *when simplicity is the main concern.* On the other hand, the scheme of [3] only requires coordination among nodes that are within a cell of area $1/n$. Note that the average number of nodes in such a cell is $\Theta(1)$. Proposition 8 then shows that, when coordination among a large number of nodes is prohibited, the scheme of [3] is close to optimal. Similarly, the scheme of [4] only requires single-hop transmissions from the mobile relays to the destinations. Proposition 9 shows that, when multi-hop transmissions are undesirable, the scheme of [4] is close to optimal. Therefore, the results reported in this paper present a relatively complete picture of the achievable capacity-delay tradeoffs under different conditions.

An interesting open problem for future work is to investigate whether these insights apply to the capacity-delay tradeoff under mobility models other than the *i.i.d.* model. For example, [7] and [8] have studied the capacity-delay tradeoff under the Brownian Motion mobility model. In these works, the authors also have implicit restrictions on the scheduling policy. In particular, the scheme in [7] uses at most one mobile relay at any time (i.e., $R_b = 1$), and the scheme in [8] schedules a transmission from the mobile relay to the destination only when they are at a distance of $O(1/\sqrt{n})$ away (i.e., $l_b = O(1/\sqrt{n})$). As we have shown in this paper, under the *i.i.d.* mobility model, the optimal capacity-delay tradeoff can only be achieved when R_b, l_b and h_b all vary as functions of the delay exponent d. Putting restrictions on any one of these variables will lead to suboptimal capacity for a given delay constraint. For our future work, we plan to study whether these kind of restrictions will also limit the capacity-delay tradeoff obtained in existing works under other mobility models.

7 Conclusion and Future Work

In this paper, we have studied the fundamental capacity-delay tradeoff in mobile ad hoc networks under the *i.i.d.* mobility model. Our contributions are three-fold. We have established the upper bound on the optimal capacity-delay tradeoff over all causal scheduling policies. The upper bound not only provides the fundamental limits of capacity and delay, but also helps to identify the optimal values of the key scheduling parameters in order to achieve the optimal capacity-delay tradeoff. Our second contribution is to develop a new scheduling scheme that can achieve a capacity-delay tradeoff that differs from the upper bound only by a logarithmic factor, which also implies that our upper bound is fairly tight. The capacity achievable by our new scheme is larger than that of the existing schemes in [3] and [4]. In particular, when the delay is bounded by a constant, our scheme achieves a per-node capacity of $\Theta(n^{-1/3}/\log n)$. This indicates that, under the *i.i.d.* mobility model, mobility increases the capacity even with constant delays. Our third contribution is to use the insight drawn from the upper bound to identify the limiting factors in the existing schemes. These results present a relatively complete picture of the achievable capacity-delay tradeoffs under different considerations.

In this paper, we have assumed an *i.i.d.* mobility model. For future work, we plan to study the optimal capacity-delay tradeoff for mobile ad hoc networks under other mobility models.

Among the properties that we proved in Section 3, we expect that the Tradeoffs II to IV will be relatively invariant to the choice of mobility models, while Tradeoff I is likely to depend on a specific model. Hence, future work will concentrate on how to tailor Tradeoff I for other mobility models. Some immediate extensions to the *i.i.d.* mobility model are possible. For example, at each time slot, each node may independently choose to stay in its old position with probability p, and to move to a new random position with probability $1-p$. This model may approximate scenarios where nodes move at a fast speed and then stay for a relatively long period of time. Tradeoff I will hold for this extension of the *i.i.d.* mobility model, and hence our main results will hold as well. An extension of our methodology to the random way-point model [8, 9] is included in Appendix G. Other mobility models that we plan to investigate are, the Brownian motion mobility model [7, 8], and the linear mobility model [10], etc.

Other aspects to consider are how the upper bound will be impacted by the use of diversity coding [11], effect of fading [4], and the use of information-theoretic approaches [12, 13].

Appendix

A Proof of Proposition 1

We will need the following lemma on the minimum distance from the mobile relays to the destination at any time slot. Fix a bit b that enters into the system at time slot $t_0(b)$. At each time slot $t \geq t_0(b)$, recall that $r_b(t)$ is the number of mobile relays holding the bit b at the beginning of the time slot. Among these $r_b(t)$ mobile relays, there is one mobile relay whose distance to the destination of bit b is the smallest. Let $\tilde{L}_b(t)$ denote this minimum distance, and let

$$L_b(t) = \max\{\frac{1}{n^2}, \tilde{L}_b(t)\}.$$

It is easy to verify that

$$\tilde{l}_b(t) \geq \tilde{L}_b(t) \geq L_b(t) - \frac{1}{n^2}.$$

Lemma 10 *Under the* i.i.d. *mobility model, if* $n \geq 3$*, then*

$$\mathbf{E}\left[\frac{1}{L_b^2(t) r_b(t)} | \mathcal{F}_{t-1}\right] \leq 8\pi \log n \textit{ for all } t \geq t_0(b).$$

Proof: Let $\mathbf{I}_A$ be the indicator function on the set A. By the definition of $L_b(t)$, we have,

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{L_b^2(t)} | \mathcal{F}_{t-1}\right] &= \mathbf{E}\left[n^4 \mathbf{I}_{\{\tilde{L}_b(t) \leq \frac{1}{n^2}\}} | \mathcal{F}_{t-1}\right] \\
&\quad + \mathbf{E}\left[\frac{1}{\tilde{L}_b^2(t)} \mathbf{I}_{\{\tilde{L}_b(t) > \frac{1}{n^2}\}} | \mathcal{F}_{t-1}\right].
\end{aligned}$$

Since the nodes move on a unit square, $\tilde{L}_b(t) \leq \sqrt{2}$. Hence,

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{\tilde{L}_b^2(t)} \mathbf{I}_{\{\tilde{L}_b(t) > \frac{1}{n^2}\}} | \mathcal{F}_{t-1}\right] \\
= &\int_{\frac{1}{n^2}}^{\sqrt{2}} \frac{1}{u^2} d\mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] \\
= &\frac{1}{u^2} \mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] |_{\frac{1}{n^2}}^{\sqrt{2}} - \int_{\frac{1}{n^2}}^{\sqrt{2}} \mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] d\frac{1}{u^2} \\
= &\frac{1}{2} - n^4 \mathbf{P}[\tilde{L}_b(t) \leq \frac{1}{n^2} | \mathcal{F}_{t-1}] + \int_{\frac{1}{n^2}}^{\sqrt{2}} \frac{2}{u^3} \mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] du.
\end{aligned}$$

Hence,

$$\mathbf{E}\left[\frac{1}{L_b^2(t)} | \mathcal{F}_{t-1}\right] = \frac{1}{2} + \int_{\frac{1}{n^2}}^{\sqrt{2}} \frac{2}{u^3} \mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] du.$$

Let ρ_b be the distance from any one mobile node to the destination of the bit b. Then, due to the *i.i.d.* mobility model, we have,

$$\mathbf{P}[\rho_b \leq u | \mathcal{F}_{t-1}] \leq \pi u^2,$$

and,

$$\begin{aligned}
\mathbf{P}[\tilde{L}_b(t) \leq u | \mathcal{F}_{t-1}] &\leq 1 - (1 - \pi u^2)^{r_b(t)} \\
&\leq \pi r_b(t) u^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{L_b^2(t)}|\mathcal{F}_{t-1}\right] &= \frac{1}{2} + \int_{\frac{1}{n^2}}^{\sqrt{2}} \frac{2}{u^3}\mathbf{P}[\tilde{L}_b(t) \leq u|\mathcal{F}_{t-1}]du \\
&\leq \frac{1}{2} + \int_{\frac{1}{n^2}}^{\sqrt{2}} \pi r_b(t)\frac{2}{u}du \\
&= \frac{1}{2} + 2\pi r_b(t)\log u|_{\frac{1}{n^2}}^{\sqrt{2}} \\
&= \frac{1}{2} + 2\pi r_b(t)(\log\sqrt{2} + 2\log n) \\
&\leq 8\pi r_b(t)\log n,
\end{aligned}$$

when $n \geq 3$. Finally, since $r_b(t)$ is $\mathcal{F}_{t-1}$-measurable, we have

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{L_b^2(t)r_b(t)}|\mathcal{F}_{t-1}\right] &= \frac{1}{r_b(t)}\mathbf{E}\left[\frac{1}{L_b^2(t)}|\mathcal{F}_{t-1}\right] \\
&\leq 8\pi\log n.
\end{aligned}$$

Q.E.D.

Proof of Proposition 1 : Let

$$V_t = 8\pi\log n\,[t - t_0(b)] - \sum_{s=t_0(b)+1}^{t} \frac{1}{L_b^2(t)r_b(t)}\mathbf{I}_{\{C_b(t)=1\}},$$

Then for all $t \geq t_0(b)$, V_t is also $\mathcal{F}_t$-measurable and $V_{t_0(b)} = 0$. By Lemma 10, we have

$$\begin{aligned}
&\mathbf{E}[V_t - V_{t-1}|\mathcal{F}_{t-1}] \\
&= 8\pi\log n - \mathbf{E}\left[\frac{1}{L_b^2(t)r_b(t)}\mathbf{I}_{\{C_b(t)=1\}}|\mathcal{F}_{t-1}\right] \\
&\geq 8\pi\log n - \mathbf{E}\left[\frac{1}{L_b^2(t)r_b(t)}|\mathcal{F}_{t-1}\right] \\
&\geq 0.
\end{aligned}$$

Hence,

$$\mathbf{E}[V_t|\mathcal{F}_{t-1}] \geq V_{t-1},$$

i.e., V_t is a sub-martingale. Recall that $s_b = \min\{t : t \geq t_0(b) \text{ and } C_b(t) = 1\}$. Since s_b is a stopping time, by appropriately invoking the Optional Stopping Theorem [5, p249, Theorem 4.1], we have,

$$\mathbf{E}[V_{s_b}] \;\geq\; 0.$$

Hence,

$$8\pi \log n \mathbf{E}[D_b] \geq \mathbf{E}\left[\frac{1}{L_b^2(s_b)R_b}\right].$$

Using Hőlder's Inequality [5, p15]

$$\mathbf{E}^2[\frac{1}{L_b(s_b)}] \leq \mathbf{E}[R_b]\mathbf{E}[\frac{1}{L_b^2(s_b)R_b}],$$

we have,

$$\begin{aligned} 8\pi \log n \mathbf{E}[D_b] &\geq \mathbf{E}^2[\frac{1}{L_b(s_b)}]\frac{1}{\mathbf{E}[R_b]} \\ &\geq \frac{1}{\mathbf{E}^2[L_b(s_b)]\mathbf{E}[R_b]}. \end{aligned}$$

Finally, by definition,

$$l_b = \tilde{l}_b(s_b) \geq L_b(s_b) - \frac{1}{n^2},$$

therefore,

$$8\pi \log n \mathbf{E}[D_b] \geq \frac{1}{(\mathbf{E}[l_b] + \frac{1}{n^2})^2 \mathbf{E}[R_b]}.$$

Q.E.D.

B Proof of Proposition 3

The next Lemma will be used frequently in the proof of Propositions 3 and 7. Consider an experiment where we randomly throw n balls into $m \leq n$ urns. The probability that each ball j enters urn i is $\frac{p}{m}$ and is independent of the position of other balls. Thus, $p \leq 1$ is the *success* *probability* that the ball is thrown into *any* one of the urns. Let $B_i, i = 1, ...m$ be the number of balls in urn i after n balls are thrown. It is obvious that $\mathbf{E}[B_i] = \frac{np}{m}$. The following Lemma shows that, when n is large, with high probability all B_i will concentrate on its mean.

Lemma 11 *As $n \to \infty$,*

1. *If $\frac{np}{m} = c$, where c is a positive constant, then*

$$\mathbf{P}[B_i \geq \frac{np}{m} \log n \textit{ for any } i] = O(\frac{1}{n}).$$

2. *If $\frac{np}{m} \geq c \log n$ and $c \geq 8$, then*

$$\mathbf{P}[B_i \geq 2\frac{np}{m} \textit{ for any } i] \leq \frac{1}{n}.$$

3. *If $\frac{np}{m} \geq cn^{\alpha}$, where $c > 0$ and $\alpha > 0$, then*

$$\mathbf{P}[B_i \geq 2\frac{np}{m} \textit{ for any } i] = O(\frac{1}{n}).$$

4. *If $\frac{np}{m} \geq c \log n$ and $c \geq 4$, then*

$$\mathbf{P}[B_i = 0 \textit{ for any } i] = O(\frac{1}{n}).$$

5. *If $\frac{np}{m} \geq c \log n$ and $c \geq 16$, then*

$$\mathbf{P}[B_i \geq 2\frac{np}{m} \textit{ for any } i] \leq \frac{1}{n^3}.$$

Proof: By known results on the characteristic function of Bernoulli random variables, we have, for any $\theta > 0$,

$$\begin{aligned}
\mathbf{E}[e^{\theta B_i}] &= \left[e^{\theta}\frac{p}{m} + (1 - \frac{p}{m})\right]^n \\
&= \left[1 + (e^{\theta} - 1)\frac{p}{m}\right]^n \\
&\leq \exp\left[\frac{np}{m}(e^{\theta} - 1)\right] \text{ for all urn } i,
\end{aligned}$$

where in the last step we have used the inequality that

$$(1+x)^{\frac{1}{x}} \le e \text{ for } x > 0.$$

Using the Markov Inequality [5, p15], for any $y > 0$,

$$\begin{aligned}
\mathbf{P}[B_i \ge y] &\le \frac{\mathbf{E}[e^{\theta B_i}]}{e^{\theta y}} \\
&\le \exp\left[\frac{np}{m}(e^\theta - 1) - \theta y\right].
\end{aligned}$$

Hence, by the union bound

$$\mathbf{P}[B_i \ge y \text{ for any } i] \le n \exp\left[\frac{np}{m}(e^\theta - 1) - \theta y\right].$$

To prove part 1, let $\frac{np}{m} = c$ and $y = \frac{np}{m}\log n$, then

$$\begin{aligned}
\mathbf{P}[B_i \ge y \text{ for any } i] &\le n \exp\left[\frac{np}{m}(e^\theta - 1) - \theta y\right] \\
&= n \exp\left[c(e^\theta - 1) - \theta c \log n\right].
\end{aligned}$$

Let $\theta = 2/c$, then

$$\begin{aligned}
\mathbf{P}[B_i \ge y \text{ for any } i] &\le n\frac{1}{n^2}\exp\left[c(e^{2/c} - 1)\right] \\
&= O(\frac{1}{n}).
\end{aligned}$$

To prove part 2, let $y = 2\frac{np}{m}$, hence

$$\begin{aligned}
\mathbf{P}[B_i \ge y \text{ for any } i] &\le n \exp\left[\frac{np}{m}(e^\theta - 1) - \theta y\right] \\
&= n \exp\left[\frac{np}{m}(e^\theta - 1 - 2\theta)\right]. \qquad (22)
\end{aligned}$$

Let $\theta = \log 2$, then

$$e^\theta - 1 - 2\theta = -(2\log 2 - 1) = -0.386 \le -\frac{1}{4}.$$

Hence, when $\frac{np}{m} \ge c\log n$ and $c \ge 8$, we have

$$\begin{aligned}
\mathbf{P}[B_i \ge y \text{ for any } i] &\le n \exp\left[-\frac{c}{4}\log n\right] \\
&\le n\frac{1}{n^2} = O(\frac{1}{n}).
\end{aligned}$$

To prove part 3, substituting $\theta = \log 2$ and $\frac{np}{m} \geq cn^\alpha$ into (22), we have,

$$\begin{aligned} \mathbf{P}[B_i \geq y \text{ for any } i] &\leq n \exp\left[-\frac{c}{4}n^\alpha\right] \\ &\leq O(\frac{1}{n}). \end{aligned}$$

To prove part 4, note that for any i,

$$\begin{aligned} \mathbf{P}[B_i = 0] &= \left[1 - \frac{p}{m}\right]^n \\ &\leq \left[1 - \frac{4\log n}{n}\right]^n \\ &= \left[1 - \frac{4\log n}{n}\right]^{\frac{n}{4\log n}4\log n}. \end{aligned}$$

Since $\lim_{x\to 0}(1-x)^{1/x} = 1/e$, we have

$$\left[1 - \frac{4\log n}{n}\right]^{\frac{n}{4\log n}} \leq \frac{1}{\sqrt{e}} \text{ for large } n.$$

Hence, for large n,

$$\begin{aligned} \mathbf{P}[B_i = 0] &\leq \left[\frac{1}{\sqrt{e}}\right]^{4\log n} \\ &= \frac{1}{n^2}. \end{aligned}$$

Therefore,

$$\mathbf{P}[B_i = 0 \text{ for any } i\ ] \leq n\frac{1}{n^2} = O(\frac{1}{n}).$$

Part 5 can be shown analogously as part 2. *Q.E.D.*

Proof of Proposition 3 : At each time slot t, an opportunistic broadcast scheme has to determine how to duplicate each bit b to a larger number of mobile nodes. Some of the mobile nodes that already have the bit b have to be selected to transmit the bit b, and some of the other mobile nodes have to be selected to receive the bit.

Let $v_b(t,i)$ be the distance from a node i that is chosen to transmit the bit b at time slot t, to the furthest node that is chosen to receive the bit ($v_b(t,i) = 0$ if node i is not chosen to transmit the bit or if the bit b has cleared the system). Let $u_b(t,i)$ be the number of mobile nodes that are chosen to receive the bit b from node i and that do not have the bit b prior to time slot t ($u_b(t,i) = 0$ if $v_b(t,i) = 0$). Then $u_b(t,i)$ is bounded from above by the number of nodes covered by a disk of radius $v_b(t,i)$ centered at node i. It is easy to verify that

$$R_b - 1 = \sum_{t=1}^{T} \sum_{i=1}^{n} u_b(t,i).$$

Fix a time slot t. We next bound the number of nodes that is covered by each disk of radius $v_b(t,i)$ centered at node i. We divide the unit square into $g_4(n) = \lfloor \left(\frac{n}{16 \log n}\right)^{\frac{1}{2}} \rfloor^2$ cells (in $\sqrt{g_4(n)}$ rows and $\sqrt{g_4(n)}$ columns). Each cell is a square of area $1/g_4(n)$. Let B_i be the number of nodes in cell i, $i = 1, ..., g_4(n)$. Then $\mathbf{E}[B_i] = \frac{n}{g_4(n)}$. When n is large, we have

$$16 \log n \leq \frac{n}{g_4(n)} \leq 32 \log n.$$

Let $\mathcal{A}$ be the event that

$$B_i \leq \frac{2n}{g_4(n)} \text{ for all } i = 1, ..., g_4(n).$$

By part 5 of Lemma 11, $\mathbf{P}[\mathcal{A}^c] \leq 1/n^3$, Now consider each disk of radius $v_b(t,i)$. We need at most

$$\left[2v_b(t,i)\sqrt{g_4(n)} + 2\right]^2$$

cells to completely cover the disk. Hence, if event $\mathcal{A}$ occurs, the number of nodes in the disk of radius $v_b(t,i)$ will be bounded from above by

$$\begin{aligned} & \left[2v_b(t,i)\sqrt{g_4(n)} + 2\right]^2 \frac{2n}{g_4(n)} \\ \leq \; & 16nv_b^2(t,i) + \frac{16n}{g_4(n)} \\ \leq \; & 16nv_b^2(t,i) + 512 \log n \end{aligned}$$

Note that the above relationship holds for all b and i. Let $c_7 = 16/\pi$, and $c_8 = 512$. Since $u_b(t,i)$ is no greater than the number of nodes covered by the disk of radius $v_b(t,i)$, we have,

$$\mathbf{P}\left[\frac{u_b(t,i)}{n} > c_7 \pi v_b^2(t,i) + c_8 \frac{\log n}{n} \text{ for any } b, i\right] \leq \mathbf{P}[\mathcal{A}^c] \leq \frac{1}{n^3}.$$

Fix a bit b. Let $\mathcal{B}$ be the event that

$$\frac{u_b(t,i)}{n} \le c_7 \pi v_b^2(t,i) + c_8 \frac{\log n}{n} \text{ for all } i \text{ and } t = t_0(b), ..., t_0(b) + c_2 n^2.$$

Then,

$$\mathbf{P}[\mathcal{B}^c] \le \frac{1}{n^3} c_2 n^2 = \frac{c_2}{n}.$$

Since $u_b(t,i) \le n$, we have

$$\begin{aligned}
\mathbf{E}\left[\frac{u_b(t,i)}{n}\right] &= \mathbf{E}\left[\frac{u_b(t,i)}{n}\mathbf{I}_{\{\mathcal{B}\}}\right] + \mathbf{E}\left[\frac{u_b(t,i)}{n}\mathbf{I}_{\{\mathcal{B}^c\}}\right] \\
&\le \mathbf{E}\left[c_7 \pi v_b^2(t,i) + c_8 \frac{\log n}{n}\right] + \mathbf{P}[\mathcal{B}^c] \\
&\le c_7 \pi \mathbf{E}[v_b^2(t,i)] + c_8 \frac{\log n}{n} + \frac{c_2}{n} \\
&\le c_7 \pi \mathbf{E}[v_b^2(t,i)] + (c_8 + 1) \frac{\log n}{n}
\end{aligned}$$

when $n \ge \max\{N_0, \exp(c_2)\}$,

We now use the idea in Section 3 that disks of radius $\frac{\Delta}{2}$ times the transmission range centered at the transmitter are disjoint from each other. For each unicast transmission (i.e., the transmission over each hop S_b^h), the transmission range is just S_b^h. For broadcast, the transmission range is the distance from the transmitter to the furthest node that can successfully receive the bit, i.e. $v_b(t,i)$. By counting the area covered by all the disks, we have

$$\sum_{b=1}^{\lambda n T} \sum_{t=1}^{T} \sum_{i=1}^{n} \pi \frac{\Delta^2}{4} v_b^2(t,i) + \sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} \frac{\pi \Delta^2}{4} (S_b^h)^2 \le WT. \tag{23}$$

Since there are at most n nodes that can serve as transmitters at any time, we have

$$\sum_{b=1}^{\lambda n T} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathbf{I}_{\{v_b(t,i) > 0\}} \le WTn.$$

Hence,

$$\begin{aligned}
&\sum_{b=1}^{\lambda n T} \frac{\mathbf{E}[R_b] - 1}{n} \\
= \; &\mathbf{E}\left[\sum_{b=1}^{\lambda n T} \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{u_b(t,i)}{n}\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq c_7\pi\mathbf{E}\left[\sum_{b=1}^{\lambda nT}\sum_{t=1}^{T}\sum_{i=1}^{n} v_b^2(t,i)\right] \\
&\qquad +(c_8+1)\mathbf{E}\left[\sum_{b=1}^{\lambda nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\log n}{n}\mathbf{I}_{\{v_b(t,i)>0\}}\right] \\
&\leq c_7\pi\mathbf{E}\left[\sum_{b=1}^{\lambda nT}\sum_{t=1}^{T}\sum_{i=1}^{n} v_b^2(t,i)\right] + (c_8+1)WT\log n. \qquad (24)
\end{aligned}
$$

Substituting (24) into (23), we have

$$
\begin{aligned}
&\sum_{b=1}^{\lambda nT}\frac{\Delta^2}{4}\frac{\mathbf{E}[R_b]-1}{n} + \sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}\frac{\pi\Delta^2}{4}\mathbf{E}[(S_b^h)^2] \\
&\leq \left\{c_7\mathbf{E}\left[\sum_{b=1}^{\lambda nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\pi\frac{\Delta^2}{4}v_b^2(t,i)\right] + \sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}\frac{\pi\Delta^2}{4}\mathbf{E}[(S_b^h)^2]\right\} \\
&\qquad +\frac{(c_8+1)\Delta^2}{4}WT\log n \\
&\leq c_7WT + \frac{(c_8+1)\Delta^2}{4}WT\log n \\
&\leq \frac{(c_8+2)\Delta^2}{4}WT\log n
\end{aligned}
$$

when $n \geq \max\{N_0, \exp(c_2), \exp(\frac{4c_7}{\Delta^2})\}$. *Q.E.D.*

C Proof of Lemma 6

We can group all cells into $c_4 = \lfloor 2\Delta + 6\rfloor^2$ lattices. Each lattice consists of nodes that are $\lfloor 2\Delta + 6\rfloor/m$ apart along the X-axis or the Y-axis (see Fig.2). The cells of each lattice can be active at the same time since

- the transmission range from a node to a neighboring node is at most $2/m$, and
- any interfering transmitters are at least $(2\Delta + 2)/m$ distance away from the receiver.

We can schedule the c_4 lattices in a round-robin fashion and each lattice is active for $1/c_4$ amount of time.

D Proof of Proposition 7

We only need to show that the probabilities of errors of all types will go to zero as $n \to \infty$. Let x_j be the number of nodes in sending cell $j = 1, ..., g_1(n)$. Let p_{I} be the probability of the Type-I error, i.e., $x_j \geq 32n^{1/3}\log n$ for any j. Equivalently, we can consider the experiment that we throw n balls into $g_1(n)$ urns with success probability $p = 1$. It is easy to show that

$$\frac{n^{2/3}}{16\log n} \leq g_1(n) \leq \frac{n^{2/3}}{8\log n} \text{ when } n \text{ is large.}$$

The average number of nodes in each sending cell is then

$$\frac{n}{g_1(n)} = \Theta(n^{1/3}\log n).$$

Hence, by part 3 of Lemma 11,

$$\begin{aligned}
p_{\mathrm{I}} &= \mathbf{P}[x_j \geq 32n^{1/3}\log n \text{ for any } j] \\
&\leq \mathbf{P}[x_j \geq \frac{2n}{g_1(n)} \text{ for any } j] \\
&= O(1/n).
\end{aligned}$$

Let p_{II} be the probability of the Type-II error. For each receiving cell i and sending cell j, let y_{ij} be the number of nodes that are in the receiving cell i in the even time slot and in the sending cell j in the previous odd time slot. Equivalently, we can consider the experiment that we throw n balls to $g_1(n)g_2(n)$ urns with success probability $p = 1$. Since

$$\mathbf{E}[y_{ij}] = \frac{n}{g_1(n)g_2(n)} \geq 8\log n,$$

by part 4 of Lemma 11,

$$p_{\mathrm{II}} = \mathbf{P}[y_{ij} = 0 \text{ for any } i, j] = O(\frac{1}{n}).$$

Hence, with probability $(1 - p_{\mathrm{I}} - p_{\mathrm{II}})$ approaching one as $n \to \infty$, each destination node can now find a designated mobile relay that holds the message intended for the destination node and that resides in the same receiving cell. Next we study p_{III}, the probability of the Type-III error. We need to specify how to schedule the hop-by-hop transmissions from the designated mobile relays to the destination nodes within each receiving cell. We divide each receiving cell i into $g_3(n) = \lfloor \left(\frac{n^{2/3}}{4\log n}\right)^{\frac{1}{2}} \rfloor^2$ *mini-cells* (in $\sqrt{g_3(n)}$ rows and $\sqrt{g_3(n)}$ columns, see Fig. 5). Each mini-cell is a square of area $1/(g_2(n)g_3(n))$. By Lemma 6, there exists a scheduling scheme where each mini-cell can be active for $\frac{1}{c_4}$ amount of time. When each mini-cell is active, it forwards a message (or a part of a message) to one other node in the neighboring mini-cell. If the destination of the message is in the neighboring cell, the message is forwarded directly to the destination node. The messages from each designated mobile relay are first forwarded towards neighboring cells along the X-axis, then to their destination nodes along the Y-axis (see Fig. 5).

Note that there are totally n messages of length $\frac{W}{32c_4 n^{1/3}\log n}$ that need to be scheduled. The above scheduling scheme can successfully forward all messages from the designated mobile relays to the destinations provided that:

- Each mini-cell contains at least one node. Hence, each node can always find some node in the neighboring cell to serve as static relays.
- The number of messages that go through any mini-cell is bounded by $32n^{1/3}\log n$. Because each message is of length $\frac{W}{32c_4 n^{1/3}\log n}$, each mini-cell then only needs to be active for at most $\frac{1}{c_4}$ amount of time, which is always possible by Lemma 6¶.

In order to show that p_{III} goes to zero as $n \to \infty$, we only need to show that both of the above conditions will hold with probability approaching one. First note that the average number of nodes in a mini-cell is

$$\frac{n}{g_2(n)g_3(n)} \geq 4\log n.$$

¶An assumption we have used here is the *separation of time scale*, i.e., we assume that radio transmissions can be scheduled at a time scale much faster than that of node mobility. Hence, each message can be divided into many smaller pieces and the transmissions of different pieces can be pipelined to achieve maximum throughput [1]. We also assume that the overhead of dividing a message into many smaller pieces is negligible.

Let $p^{\mathrm{a}}_{\mathrm{III}}$ be the probability that any of the $g_2(n)g_3(n)$ mini-cells are empty. Equivalently, we can consider the experiment that we throw n balls into $g_2(n)g_3(n)$ urns with success probability $p = 1$. Then, by part 4 of Lemma 11,

$$p^{\mathrm{a}}_{\mathrm{III}} = O(1/n).$$

Next we group the nodes in each receiving cell by the positions of their corresponding source nodes in the previous time slot. Let $\mathcal{Y}_{ij}$ be the set of nodes in the receiving cell i that are the destination nodes for some source nodes in the sending cell j (in the previous time slot). Let $p^{\mathrm{b}}_{\mathrm{III}}$ be the probability that any set $\mathcal{Y}_{ij}, i = 1, ..., g_2(n), j = 1, ..., g_1(n)$, has more than $32 \log n$ nodes. Equivalently, we can consider the experiment that we throw n balls into $m = g_1(n)g_2(n)$ urns with success probability $p = 1$. Since

$$16 \log n \geq \frac{np}{m} = \frac{n}{g_1(n)g_2(n)} \geq 8 \log n \text{ for large } n,$$

by part 2 of Lemma 11,

$$\begin{aligned} p^{\mathrm{b}}_{\mathrm{III}} &= \mathbf{P}[|\mathcal{Y}_{ij}| \geq 32 \log n, \text{ for any } i, j] \\ &\leq \mathbf{P}[|\mathcal{Y}_{ij}| \geq 2\frac{np}{m} \text{ for any } i, j] \\ &= O(\frac{1}{n}). \end{aligned}$$

Hence, with high probability, each designated mobile relay will serve no more than $32 \log n$ destination nodes in the same receiving cell. As presented earlier, the message will first be forwarded along the X-axis, then along the Y-axis. We next bound the number of messages that go through any mini-cell along the X-axis. Within a given receiving cell i, fix any mini-cell $k = 1, ..., g_3(n)$. Let $Z^x_{i,k}$ be the number of designated mobile relays in receiving cell i that reside at the same *row* with the mini-cell k. Note that there are at most $g_1(n)$ designated mobile relays and $\sqrt{g_3(n)}$ rows of mini-cells in a given receiving cell i. Let $p^{\mathrm{c}}_{\mathrm{III}}(i)$ be the probability that $Z^x_{i,k} \geq \frac{1}{2}\sqrt{\frac{2n^{2/3}}{\log n}}$ for any mini-cell k in a given receiving cell i. Equivalently, we can consider the experiment that we throw $g_1(n)$ balls into $\sqrt{g_3(n)}$ urns with success probability $p = 1$. The average number of balls per urn is $g_1(n)/\sqrt{g_3(n)}$, and

$$\frac{1}{4}\sqrt{\frac{2n^{2/3}}{\log n}} \geq \frac{g_1(n)}{\sqrt{g_3(n)}} \geq \frac{1}{8}\sqrt{\frac{n^{2/3}}{\log n}} \text{ for large } n.$$

By part 3 of Lemma 11,

$$\begin{aligned}
p^{\mathrm{c}}_{\mathrm{III}}(i) &= \mathbf{P}[Z^x_{i,k} \geq \frac{1}{2}\sqrt{\frac{2n^{2/3}}{\log n}} \text{ for any } k] \\
&\leq \mathbf{P}[Z^x_{i,k} \geq 2\frac{g_1(n)}{\sqrt{g_3(n)}} \text{ for any } k] \\
&= O(\frac{1}{g_1(n)}) = O(n^{-\frac{2}{3}} \log n).
\end{aligned}$$

Let $p^{\mathrm{c}}_{\mathrm{III}}$ be the probability that $Z^x_{i,k} \geq \frac{1}{2}\sqrt{\frac{2n^{2/3}}{\log n}}$ for any mini-cell in *any* receiving cell. Since there are $\Theta(n^{1/3})$ receiving cells, by the union bound,

$$\begin{aligned}
p^{\mathrm{c}}_{\mathrm{III}} &\leq \sum_{i=1}^{g_2(n)} p^{\mathrm{c}}_{\mathrm{III}}(i) \\
&\leq n^{\frac{1}{3}} O(n^{-\frac{2}{3}} \log n) \\
&= O(n^{-\frac{1}{3}} \log n).
\end{aligned}$$

Therefore, with high probability, there will be at most $\frac{1}{2}\sqrt{\frac{2n^{2/3}}{\log n}}$ designated mobile relays that reside at the same row as any mini-cell, and each of them is the origin of at most $32 \log n$ messages. Hence, with probability approaching one as $n \to \infty$, the number of messages that have to go through any mini-cell along the X-axis is less than

$$16\sqrt{2n^{2/3}\log n}.$$

Similarly, let $Z^y_{i,k}$ be the number of nodes in the receiving cell i that reside at the same *column* as the mini-cell k. Let $p^{\mathrm{d}}_{\mathrm{III}}(i)$ be the probability that $Z^y_{i,k} \geq 8n^{1/3}\sqrt{2\log n}$ for any mini-cell k in a given receiving cell i. Equivalently, we can consider the experiment that we throw n balls into $\sqrt{g_3(n)}$ urns with success probability $p = 1/g_2(n)$. This experiment is independent of the previous one, because the X-coordinates of the nodes are independent of their Y-coordinates.

The average number of balls per urn is $n/(g_2(n)\sqrt{g_3(n)})$, and

$$4n^{1/3}\sqrt{2\log n} \geq \frac{n}{g_2(n)\sqrt{g_3(n)}} \geq 2n^{1/3}\sqrt{\log n} \text{ for large } n.$$

Hence, by part 3 of Lemma 11,

$$\begin{aligned} p^{\mathrm{d}}_{\mathrm{III}}(i) &= \mathbf{P}[Z^y_{i,k} \geq 8n^{1/3}\sqrt{2\log n} \text{ for any } k\ ] \\ &\leq \mathbf{P}[Z^y_{i,k} \geq 2\frac{n}{g_2(n)\sqrt{g_3(n)}} \text{ for any } k\ ] \\ &= O(\frac{1}{n}). \end{aligned}$$

Let $p^{\mathrm{d}}_{\mathrm{III}}$ be the probability that $Z^y_{i,k} \geq 8n^{1/3}\sqrt{2\log n}$ for any mini-cell in *any* receiving cell. By the union bound,

$$\begin{aligned} p^{\mathrm{d}}_{\mathrm{III}} &\leq \sum_{i=1}^{g_2(n)} p^{\mathrm{d}}_{\mathrm{III}}(i) \\ &= n^{1/3}O(\frac{1}{n}) = O(n^{-2/3}). \end{aligned}$$

Therefore, with probability approaching one as $n \to \infty$, the number of messages that have to go through any mini-cell along the Y-axis is less than

$$8n^{1/3}\sqrt{2\log n}.$$

Combining all of the above results, with probability no greater than

$$p_{\mathrm{III}} = p^{\mathrm{a}}_{\mathrm{III}} + p^{\mathrm{b}}_{\mathrm{III}} + p^{\mathrm{c}}_{\mathrm{III}} + p^{\mathrm{d}}_{\mathrm{III}} = O(n^{-1/3}\log n),$$

the number of messages that have to go through any mini-cell $k = 1, ...g_3(n)$ in any receiving cell i is less than

$$24n^{1/3}\sqrt{2\log n} \leq 32n^{1/3}\log n \text{ for large } n.$$

Proposition 7 then follows. *Q.E.D.*

E Proof of Proposition 8

We start from inequality (17). Since $\mathbf{E}[l_b] \leq \sqrt{c_5} n^{-1/2}$ for some positive constant c_t, we have,

$$\begin{aligned} \frac{1}{c_1 n \log n} \frac{(\lambda n T)^3}{\bar{D}(2 \sum_{b=1}^{\lambda n T} \mathbf{E}[l_b])^2} &\geq \frac{1}{4 c_1 n \log n} \frac{(\lambda n T)^3}{\bar{D}(\lambda n T)^2 c_5 n^{-1}} \\ &= \frac{1}{4 c_1 c_5 \log n} \frac{\lambda n T}{\bar{D}}, \end{aligned}$$

and,

$$\begin{aligned} \frac{2\pi}{WTn} (\sum_{b=1}^{\lambda n T} \mathbf{E}[l_b])^2 &\leq \frac{2\pi}{WTn} (\lambda n T)^2 c_5 n^{-1} \\ &= \frac{2\pi c_5}{W} \lambda^2 T. \end{aligned}$$

Substitute the above two inequalities into (17). Note that when $\bar{D} = o(n)$, the first term of (17) dominates the rest for large n. Hence

$$\begin{aligned} \frac{4c_3}{\Delta^2} WT \log n &\geq \frac{1}{2} \frac{1}{c_1 n \log n} \frac{(\lambda n T)^3}{\bar{D}(2 \sum_{b=1}^{\lambda n T} \mathbf{E}[l_b])^2} \\ &\geq \frac{1}{8 c_1 c_5 \log n} \frac{\lambda n T}{\bar{D}}. \end{aligned}$$

We can then solve for λ,

$$\lambda \leq \frac{\bar{D} \log^2 n}{n} \frac{32 c_1 c_3 c_5 W}{\Delta^2}.$$

Q.E.D.

F Proof of Proposition 9

Since $h_b \leq c_6$ for some positive number c_6, we have,

$$\begin{aligned} \left(\sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} S_b^h \right)^2 &\leq \left(\sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} 1 \right) \left(\sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} (S_b^h)^2 \right) \\ &\leq \lambda n T c_6 \sum_{b=1}^{\lambda n T} \sum_{h=1}^{h_b} (S_b^h)^2. \end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}(S_b^h)^2] &\geq \frac{1}{c_6\lambda nT}\mathbf{E}[\left(\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}S_b^h\right)^2] \\
&\geq \frac{1}{c_6\lambda nT}\left(\mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}S_b^h]\right)^2 \\
&\geq \frac{1}{c_6\lambda nT}\left(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]\right)^2 . \qquad (25)
\end{aligned}$$

Substitute (15) and (25) into (7), we have,

$$\begin{aligned}
\frac{4c_3}{\Delta^2}WT\log n &\geq \sum_{b=1}^{\lambda nT}\frac{\mathbf{E}[R_b]-1}{n}+\pi\mathbf{E}[\sum_{b=1}^{\lambda nT}\sum_{h=1}^{h_b}(S_b^h)^2] \\
&\geq \frac{1}{c_1 n\log n}\frac{(\sum_{b=1}^{\lambda nT}1)^3}{\bar{D}\left(\sum_{b=1}^{\lambda nT}(\mathbf{E}[l_b]+\frac{1}{n^2})\right)^2} \\
&\quad +\frac{\pi}{c_6\lambda nT}(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b])^2-\lambda T
\end{aligned}$$

As in the proof of Proposition 5, the case with $\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]\geq\lambda T/n$ will again prevail. Hence,

$$\begin{aligned}
\frac{4c_3}{\Delta^2}WT\log n &\geq \frac{1}{c_1 n\log n}\frac{(\sum_{b=1}^{\lambda nT}1)^3}{\bar{D}\left(2\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]\right)^2} \\
&\quad + \frac{\pi}{c_6\lambda nT}(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b])^2-\lambda T \\
&\geq 2\left[\frac{\pi}{4c_1c_6 n\log n\bar{D}}(\lambda nT)^2\right]^{1/2}-\lambda T \\
&= 2\sqrt{\frac{\pi}{4c_1c_6}\frac{\lambda^2 nT^2}{\bar{D}\log n}}-\lambda T.
\end{aligned}$$

When $\bar{D}=o(n)$, the first term dominates for large n. Hence,

$$\begin{aligned}
\frac{4c_3}{\Delta^2}WT\log n &\geq \sqrt{\frac{\pi}{4c_1c_6}\frac{\lambda^2 nT^2}{\bar{D}\log n}} \\
\lambda^2 &\leq \frac{\bar{D}\log^3 n}{n}\frac{64c_1c_3^2c_6W^2}{\pi\Delta^4}.
\end{aligned}$$

Q.E.D.

G 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.

G.1 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.[‖] 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 [9], 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 [9] 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 [8], we assume that the speed v scales as $v(n) = \Theta(1/\sqrt{n})$ when the number of nodes n increases.[**]

[‖]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.

[**]It is also possible to extend out 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.

G.2 Inherent Tradeoffs under the Random Way-point Mobility Model

We will refer to the capacity-achieving schemes in Section 5 as *cell-based schemes.* 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 a cell-based scheme. Indeed, by choosing the appropriate cell-partitioning (i.e., $g_1(n)$ sending cells and $g_2(n)$ receiving cells), the cell-based scheme in Section 5 can 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. 6, 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. 7). Note that each line in Fig. 7 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. 6 can be expressed as

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

which corresponds to

$$\mathbf{E}[D_b] \approx \Theta(\frac{n^{\frac{1}{2}}}{\mathbf{E}[l_b]\mathbf{E}[R_b]}). \tag{26}$$

This relationship between D_b, l_b and R_b under the *RWP* mobility model is consistent with the findings in [8]. When l_b and R_b are fixed, it has been shown in [8] 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 (26). However,

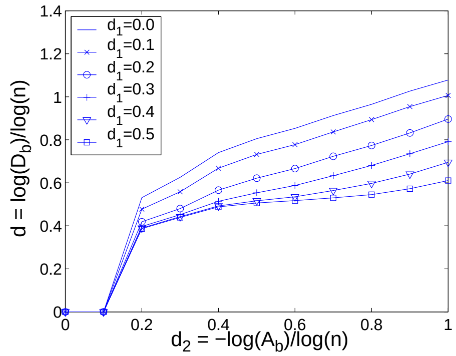

Figure 6: Delay exponent d versus d_1 and d_2 for the random way-point mobility model.

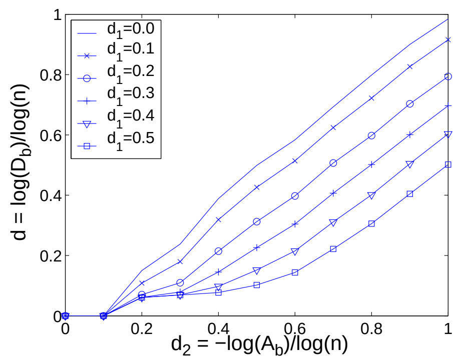

Figure 7: Delay exponent d versus d_1 and d_2 for the *i.i.d.* mobility model.

this relationship only holds when $d \geq 0.5$. In fact, it has been shown in [8] that, in order to take advantage of mobility, the minimum amount of delay under the *RWP* mobility model is $\Theta(\sqrt{n})$. Hence, schemes for constant delay constraints (such as the one in Section 5.2) will not work under the *RWP* mobility model.

G.3 Upper Bound on the Capacity-Delay Tradeoff

Combining relationship (26) with Tradeoffs II, III and IV, we can compute the upper bound on the maximum capacity under given delay constraints as in Proposition 5. We can in fact assume a more general version of the estimate (26):

$$\mathbf{E}[D_b] \geq \frac{c_1 n^{\frac{1}{2}-\epsilon}}{\mathbf{E}[l_b]\mathbf{E}[R_b]}, \tag{27}$$

where $c_1 > 0$ and ϵ is a positive number close to 0. We then have

$$\sum_{b=1}^{\lambda nT} \mathbf{E}[R_b] \geq c_1 n^{\frac{1}{2}-\epsilon} \sum_{b=1}^{\lambda nT} \frac{1}{\mathbf{E}[l_b]\mathbf{E}[D_b]}. \tag{28}$$

Let

$$\bar{D} = \frac{\sum_{b=1}^{\lambda nT} \mathbf{E}[D_b]}{\sum_{b=1}^{\lambda nT} 1} = \frac{\sum_{b=1}^{\lambda nT} \mathbf{E}[D_b]}{\lambda nT}.$$

Using Jensen's Inequality and Hőlder's Inequality, we have,

$$\begin{aligned} \frac{1}{\left(\frac{\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]}{\sum_{b=1}^{\lambda nT} 1}\right)} &\leq \left(\frac{\sum_{b=1}^{\lambda nT} \frac{1}{\sqrt{\mathbf{E}[l_b]}}}{\sum_{b=1}^{\lambda nT} 1}\right)^2 \\ &\leq \frac{\sum_{b=1}^{\lambda nT} \frac{1}{\mathbf{E}[l_b]\mathbf{E}[D_b]}}{\sum_{b=1}^{\lambda nT} 1} \frac{\sum_{b=1}^{\lambda nT} \mathbf{E}[D_b]}{\sum_{b=1}^{\lambda nT} 1}. \end{aligned} \tag{29}$$

Equality holds when $\mathbf{E}[l_b]$ is the same for all b and $\mathbf{E}[D_b] = \bar{D}$ for all b. Substituting (29) in (28), we have

$$\sum_{b=1}^{\lambda nT} \mathbf{E}[R_b] \quad \geq \quad c_1 n^{\frac{1}{2}-\epsilon} \frac{\left(\sum_{b=1}^{\lambda nT} 1\right)^2}{\bar{D} \sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]}. \tag{30}$$

Substituting (12) and (30) into Tradeoff III (7), we have

$$\begin{aligned} \frac{4c_3 WT \log n}{\Delta^2} &\geq \sum_{b=1}^{\lambda nT} \frac{\mathbf{E}[R_b] - 1}{n} + \pi \mathbf{E}[\sum_{b=1}^{\lambda nT} \sum_{h=1}^{h_b} (S_b^h)^2] \\ &\geq c_1 n^{-\frac{1}{2}-\epsilon} \frac{(\lambda nT)^2}{\bar{D} \sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]} \\ &\quad + \frac{2\pi}{WTn} \left(\sum_{b=1}^{\lambda nT} \mathbf{E}[l_b]\right)^2 - \lambda T. \end{aligned}$$

Using the inequality

$$a + b + c \geq 3\sqrt[3]{abc},$$

we have,

$$\frac{4c_3WT\log n}{\Delta^2} \geq 3\sqrt[3]{\left(\frac{c_1}{2}n^{-\frac{1}{2}-\epsilon}\frac{(\lambda nT)^2}{\bar{D}\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]}\right)^2\frac{2\pi}{WTn}\left(\sum_{b=1}^{\lambda nT}\mathbf{E}[l_b]\right)^2} - \lambda T \tag{31}$$

$$= 3\sqrt[3]{\frac{c_1^2\pi}{2}\frac{\lambda^4 n^{2-2\epsilon}T^3}{\bar{D}^2W}} - \lambda T. \tag{32}$$

Therefore, either

$$\lambda \leq O(\frac{\bar{D}^2}{n^{2-2\epsilon}}), \tag{33}$$

or, if $\lambda = \omega(\frac{\bar{D}^2}{n^{2-2\epsilon}})$, then the first term in (32) dominates when n is large. In the latter case, for n large enough,

$$\begin{aligned}
\frac{4c_3WT\log n}{\Delta^2} &\geq \frac{3}{2}\sqrt[3]{\frac{c_1^2\pi}{2}\frac{\lambda^4 n^{2-2\epsilon}T^3}{\bar{D}^2W}} \\
\frac{64c_3^3W^3T^3\log^3 n}{\Delta^6} &\geq \frac{27}{8}\frac{c_1^2\pi}{2}\frac{\lambda^4 n^{2-2\epsilon}T^3}{\bar{D}^2W} \\
\lambda^2 &\leq \sqrt{\frac{1024c_3^3W^4}{27\pi c_1^2\Delta^6}}\frac{\bar{D}}{n^{1-\epsilon}}\log^{\frac{3}{2}} n.
\end{aligned} \tag{34}$$

Compare (33) and (34). Since $\bar{D} = o(n^d), d < 1$, inequality (34) will eventually be the loosest for large n. Hence, the optimal capacity-delay tradeoff is upper bounded by

$$\lambda^2 \leq O(\frac{\bar{D}}{n^{1-\epsilon}}\log^{\frac{3}{2}} n).$$

We can now take $\epsilon \to 0$ and obtain

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

G.4 Optimal Values of the Key Scheduling Parameters

To obtain the optimal value of the key scheduling parameters, we can infer as in Section 5.1 again that the scheduling policy should use the same parameters for all bits. Hence, we will assume that all key parameters (such as R_b, l_b, etc.) are indeed the same for all bits. Assume that the mean delay is bounded by $n^d, d < 1$. By earlier derivations, we have,

$$\lambda \le \Theta(n^{\frac{d-1+\epsilon}{2}} \log^{\frac{3}{4}} n).$$

Note that equality holds in (31) if and only if

$$\frac{c_1}{2} n^{-\frac{1}{2}-\epsilon} \frac{(\lambda n T)^2}{\bar{D} \sum_{b=1}^{\lambda n T} \mathbf{E}[l_b]} = \frac{2\pi}{WTn} (\sum_{b=1}^{\lambda n T} \mathbf{E}[l_b])^2.$$

Substituting $\sum_{b=1}^{\lambda n T} \mathbf{E}[l_b] = \lambda n T l_b$, we can solve for l_b,

$$\begin{aligned} \frac{c_1}{2} n^{-\frac{1}{2}-\epsilon} \frac{\lambda n T}{\bar{D} l_b} &= \frac{2\pi}{WTn} (\lambda n T)^2 l_b^2 \\ l_b^3 &= \frac{c_1}{4\pi} \frac{W}{\bar{D} \lambda n^{\frac{1}{2}+\epsilon}}. \end{aligned}$$

Substituting $\lambda = \Theta(n^{(d-1+\epsilon)/2} \log^{3/4} n)$ and $\bar{D} = n^d$, we obtain the optimal value of l_b,

$$l_b = \Theta(n^{-\frac{d+\epsilon}{2}} \log^{-\frac{1}{4}} n).$$

A reasonable choice for the area of capture neighborhood, A_b, is then,

$$A_b = l_b^2 = \Theta(n^{-d-\epsilon}/\sqrt{\log n}).$$

By setting (9) of Tradeoff IV to equality, we have

$$\begin{aligned} \lambda n T h_b &= \frac{WTn}{2} \\ h_b &= \frac{W}{2\lambda} = \Theta(n^{\frac{1-d-\epsilon}{2}} / \log^{\frac{3}{4}} n). \end{aligned}$$

By setting (6) of Tradeoff II to equality, we have

$$S_b^h = \frac{l_b}{h_b} = \Theta(\sqrt{\frac{\log n}{n}}).$$

From (27), equality is attained when

$$R_b = \Theta(\frac{c_1 n^{\frac{1}{2}-\epsilon}}{l_b \bar{D}}) = \Theta(n^{\frac{1-d-\epsilon}{2}} \log^{\frac{1}{4}} n).$$

Taking $\epsilon \to 0$, we summarize the optimal values of the scheduling parameters below when the delay constraint is $D_b = O(n^d)$:

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

Hence, to use the cell-based capacity-achieving scheme as in Section 5, 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 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 [8]. Analogous to Section 6, we can show that a restrictive choice of the receiving cell size is again the performance limiting factor of the scheme in [8].

References

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, 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, no. 4, 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] R. Durrett, *Probability : Theory and Examples*, 2nd ed. Belmont, CA: Duxbury Press, 1996.

[6] M. E. Woodward, *Communication and Computer Networks: Modelling with Discrete-Time Queues*. Los Alamitos, CA: IEEE Computer Society Press, 1994.

[7] 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.

[8] 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.

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

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

[11] E. Perevalov and R. Blum, "Delay Limited Capacity of Ad hoc Networks: Asymptotically Optimal Transmission and Relaying Strategy," in *Proceedings of IEEE INFOCOM*, San Francisco, CA, April 2003.

[12] M. Gastpar and M. Vetterli, "On the Capacity of Wireless Networks: The Relay Case," in *Proceedings of IEEE INFOCOM*, New York, June 2002.

[13] P. Gupta and P. R. Kumar, "Towards an Information Theory of Large Networks: An Achievable Rate Region," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1877–1894, August 2003.