Optimal Multicast Capacity and Delay Tradeoffs in MANETs: A Global Perspective

Yun Wang, Xiaoyu Chu, Xinbing Wang
Dept of Electronic Engineering
Shanghai Jiao Tong University, China
Email: {sunshinehk,cxygrace,xwang8}@sjtu.edu.cn

Yu Cheng
Dept. of Elec. and Comp. Engin.
Illinois Institute of Technology, USA
Email: cheng@iit.edu

***Abstract*—In this paper, we give a global perspective of multicast capacity and delay analysis in Mobile Ad-hoc Networks (MANETs). Specifically, we consider four node mobility models: (1) two-dimensional i.i.d. mobility, (2) two dimensional hybrid random walk, (3) one-dimensional i.i.d. mobility, and (4) one-dimensional hybrid random walk. Two mobility time-scales are included in this paper: (i) Fast mobility where node mobility is at the same time-scale as data transmissions; (ii) Slow mobility where node mobility is assumed to occur at a much slower time-scale than data transmissions. In addition, heterogeneous mobile networks with infrastructure support are also discussed in this paper. For each of the eight types of homogeneous networks, given a delay constraint D, we first characterize the multicast capacity upper bound, and then we develop a scheme that can achieve a capacity-delay tradeoff close to the upper bound up to a logarithmic factor. Further, we give the premises of why our one-dimensional model can achieve higher capacity than two-dimensional model, and based on them, we propose our *hybrid dimensional model* for future consideration.**

I. Introduction

Since the seminal paper by Gupta and Kumar [1], where a maximum per-node throughput of $O(1/\sqrt{n})$ was established in a static network with n nodes, there has been tremendous interest in the networking research community to investigate the fundamental achievable capacity in wireless ad-hoc networks. Franceschetti et al. [2] later proved by percolation theory that the same $1/\sqrt{n}$ per-node throughput can be achieved in random ad hoc networks. However, this upper bound decreases to zero when the node number n goes to infinity, which leads to a non-scalable network. Since then, much attention is put on the improving methods. In general, there are mainly two options: mobility and infrastructure support.

Grossglauser and Tse [3] is the first to introduce the mobile network. In their work, a constant capacity is achieved at the cost of excessive packet delay. Then extensive works have been done to explore the optimal capacity-delay tradeoff in mobile ad hoc networks, [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. The tradeoffs in these works differ considerably due to the different mobility models and network settings the authors studied. The mobility models that have been studied include the i.i.d. mobility model [7], [8], [9]; random walk mobility model [4], [10], [11]; random way-point mobility model [12], [13] and Brownian mobility model [5].

The i.i.d. mobility model is widely studied for its modeling simplicity. Neely and Modiano [7] considered a $\Theta(n)$ cell-partitioned network where only one transmission for each cell is allowed in each time-slot. They showed that the capacity-delay tradeoff was $\lambda = O(\frac{D}{n})$, where λ is the throughput per S-D pair, and D is the number of time-slots taken to deliver packets from source to destination. In [7], fast mobility is assumed. A different time-scale of mobility, slow mobility, was considered by Toumpis and Goldsmith in [8], and Lin and Shroff in [9]. For slow mobiles, node mobility is assumed to be much slower than data transmissions. So the packet size can be scaled down as n increases, and multi-hop transmissions are feasible in one time-slot. The delay throughput tradeoff was shown to be $\lambda = O(\sqrt{D/n}\log n)$ in [8]. A better tradeoff was given in [9], where the maximum throughput per S-D pair with mean delay D was shown to be $\lambda = O(\sqrt[3]{D/n}\log n)$, and a scheme was proposed to achieve a capacity-delay tradeoff close to the upper bound up to logarithmic factors.

Most of the above works assume two-dimensional mobility. In [15], one-dimensional motion is considered. Later, Ying et al. [10] gave a broader perspective of optimal delay-throughput tradeoffs in MANETs. They showed the results in eight kinds of networks concerning one-dimension/two dimension, i.i.d. mobility/random walk mobility and fast mobility/slow mobility.

The other capacity improving method, infrastructure support, is studied in [16], [17], [18], [19], [20], [21]. They proposed to connect the more powerful nodes (i.e., base stations) with a wired network, resulting in "hybrid wireless networks". Mostly, transmissions in hybrid networks are classified into two independent modes: ad-hoc mode and infrastructure mode, and the network capacity is the combination of these two. The base stations can make significant contributions to the capacity only when $m = \Omega(n)$, [18]. Note that all these works studied infrastructure support in static networks.

All the above works studied the unicast traffic. As the demand of information sharing increases rapidly, multicast flows are expected to be predominant in many of the emerging applications, such as the order delivery in battlefield networks and wireless video conferences. Related works are [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], including static, mobile and hybrid networks.

Introducing mobility into the multicast traffic pattern, Hu et al. [31] studied a motioncast model. Fast mobility was assumed. Capacity and delay were calculated under two particular algorithms, and the tradeoff derived from them was $\lambda = O(\frac{D}{nk \log k})$, where k was the number of destinations per source. In their network, $\Theta(n)$ cell-partitioning [7] was used, which fixed the transmission range as $O(\frac{1}{n})$. Zhou and Ying [32] also studied the *fast mobility* model and provided an optimal tradeoff under their network assumptions. Specifically, they considered a network that consists of n_s multicast sessions, each of which had one source and p desinations. They showed that given delay constraint D, the capacity per multicast session was $O\Big(\min\Big\{1, (\log p)(\log(n_s p))\sqrt{\frac{D}{n_s}}\Big\}\Big)$. Then a *joint coding/scheduling algorithm* was proposed to achieve a throughput of $O\Big(\min\Big\{1, \sqrt{\frac{D}{n_s}}\Big\}\Big)$. In their network, each multicast session had no intersection with others and the total number of mobile nodes was $n = n_s(p+1)$.

Heterogeneous networks with multicast traffic pattern were studied by Li et al. [34] and Mao et al. [35]. Wired base stations are used and their transmission range can cover the whole network. Li et al. [36] studied a dense network with fixed unit area. The helping nodes in their work are wireless, but have higher power and only act as relays instead of sources or destinations. [34], [35] and [36] all study static networks.

In this paper, we give a general analysis on the optimal multicast capacity-delay tradeoffs in both homogeneous and heterogeneous MANETs. We assume a mobile wireless network that consists of n nodes, among which $n_s = n^s$ nodes are selected as sources and $n_d = n^\alpha$ destined nodes are chosen for each. Thus, n_s multicast sessions are formed. In heterogeneous network, $m = n^\beta$ base stations connected with wires are uniformly placed in the unit square, separating the network into m subregions. Mobile nodes can transmit packets to base stations only when their distance is within the transmission range of mobile nodes L.

We summarize our main results here:

(1) Two-dimensional i.i.d. mobility models:

  (i) Under the fast mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt{\frac{D}{n}n_d}\Big)$ under a delay constraint $D = n^d$. A cell-partitioned scheme is presented to achieve a close capacity when $D = o(\frac{n}{n_d})$ and $n_s n_d \geq n$.

  (ii) Under the slow mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[3]{\frac{D}{n}n_d}\Big)$ under a delay constraint D. A cell-partitioned scheme is presented to achieve a close capacity when $D = o(\frac{n}{n_d})$ and $n_s n_d \geq n$.

(2) Two-dimensional hybrid random walk mobility models:

  (i) Under the fast mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt{\frac{D}{n}n_d}\Big)$ when $B = o(1)$ and $D = \omega(|\log B|/B^2)$.

  (ii) Under the slow mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[3]{\frac{D}{n}n_d}\Big)$ when $B = o(1)$ and $D = \omega(|\log B|/B^2)$.

(3) One-dimensional i.i.d. mobility models:

  (i) Under the fast mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[3]{\frac{D^2}{n}n_d^2}\Big)$ under a delay constraint D. A cell-partitioned scheme is presented to achieve a close capacity when $D = o(\frac{\sqrt{n}}{n_d})$, $n_d = O(\sqrt{n})$ and $n_s n_d \geq n$.

  (ii) Under the slow mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[4]{\frac{D^2}{n}n_d^2}\Big)$ under a delay constraint D. A cell-partitioned scheme is presented to achieve a close capacity when $D = o(\frac{\sqrt{n}}{n_d})$, $n_d = O(\sqrt{n})$ and $n_s n_d \geq n$.

(4) One-dimensional hybrid random walk mobility models:

  (i) Under the fast mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[3]{\frac{D^2}{n}n_d^2}\Big)$ when $B = o(1)$ and $D = \omega(1/B^2)$.

  (ii) Under the slow mobility assumption, it is shown that the maximum throughput per multicast session is $O\Big(\frac{n}{n_s n_d}\sqrt[4]{\frac{D^2}{n}n_d^2}\Big)$ when $B = o(1)$ and $D = \omega(1/B^2)$.

(5) Heterogeneous networks with infrastructure support:

  (i) In infrastructure mode, it is shown that the maximum aggregate input throughput of the whole network is $\max\big\{\Theta(1), \Theta(\frac{m}{n_d})\big\}$.

  (ii) The aggregate input capacity of the heterogeneous networks under the above eight different kinds of mobility models are $T = \max\big\{\Theta(\frac{m}{n_d}), n_s\lambda_a\big\}$, where λ_a is the per session capacity in each of the homogeneous networks presented above. The detail results can be seen in Table VI.

The rest of the paper is organized as follows. In section II, we outline the system models. Eight kinds of mobility models in homogeneous networks are discussed in section III to section VIII respectively. Section IX gives a concise summary of our results and makes comparisons with unicast cases. Section X discusses the impact of mobility dimension on optimal capacity. In Section XI, we study the heterogeneous networks with infrastructure support and establishes the maximum aggregate input throughput. Then we conclude.

II. System Models

In this section, we first present the multicast traffic pattern, wireless interference model and the definition of capacity and delay. Then the mobility models are characterized in both homogeneous and heterogeneous networks.

Multicast Traffic Pattern: We consider a mobile ad-hoc network where n nodes move within a unit square. Among them, n_s nodes are selected as sources, and each has n_d distinct destination nodes. We number the n mobile nodes as node $\{1, 2, \ldots, n\}$ and assume that the first continuous n_s nodes are source nodes, i.e., $\{1, 2, \ldots, n_s\}$. We group each source and its n_d destinations as a multicast session. Note that a particular node may be included by different multicast sessions as either source or destination.

Protocol Model: 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 $i, i = 1, \ldots, n$. Let $|X_i - X_j|$ be the Euclidean distance between nodes i and j. Node i can communicate directly with another node j at W bits per second if and only if the following interference constraint is satisfied for every other node $k \neq i, j$ that is simultaneously transmitting, [1]. Here, Δ is some positive number.

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

Definition of Capacity: We assume the same packet arrival rate per time-slot for each source, say λ. The network is said stable if and only if there exists a certain scheduling scheme which can guarantee the finite length of queue in each node as time goes to infinity. Then the *capacity*, which is short for per-session capacity, is defined as the maximum arrival rate λ that the stable network can support.

Definition of Delay: We define the *survival time* for a certain packet as the time interval counting from the moment it enters the network and ending until one of its copies reaches the last destination. Note that we consider the expectation value over all possible network configurations. Then the *delay*, which is short for per-session delay, is defined as the average *survival time* over all packets during an enough long term. Also note that we do not consider the queueing delays in the network.

Notations: Given non-negative functions $f(n)$ and $g(n)$:

(1) $f(n) = O(g(n))$ means there exist positive constants c and m such that $f(n) \leq cg(n)$ for all $n \geq m$.
(2) $f(n) = o(g(n))$ means that $\lim_{n\to\infty} f(n)/g(n) = 0$.
(3) $f(n) = \Omega(g(n))$ means there exist positive constants c and m such that $f(n) \geq cg(n)$ for all $n \geq m$.
(4) $f(n) = \omega(g(n))$ means that $\lim_{n\to\infty} g(n)/f(n) = 0$.
(5) $f(n) = \Theta(g(n))$ means that both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ hold.

A. Homogeneous Networks

Mobile ad hoc network model: Consider a pure ad hoc network where n wireless mobile nodes are positioned in a unit square. The unit square is assumed to be a torus, where the left and right edges are connected, and top and bottom edges are also connected. We will study the following mobility models in this paper.

(1) **Two-dimensional i.i.d. mobility model:** Our two-dimensional i.i.d. mobility model is defined as follows:
   (i) At each time slot, the nodes are uniformly, randomly positioned in the unit square.
   (ii) The node positions are independent of each other, and independent from time slot to time slot. So the nodes are totally reshuffled at each time slot.

(2) **Two-dimensional hybrid random walk model:** Consider a unit square which is further divided into $1/B^2$ squares of equal size. Each of the smaller square will be called an RW-cell (random walk cell), and indexed by (U_x, U_y) where $U_x, U_y \in \{1, \ldots, 1/B\}$. A node which is in one RW-cell at a time slot moves to one of its eight adjacent RW-cells or stays in the same RW-cell in the next time-slot with each move being equally likely as in Figure 1. Two RW-cells are said to be adjacent if they share a common point. The node position within the RW-cell is randomly uniformly selected.

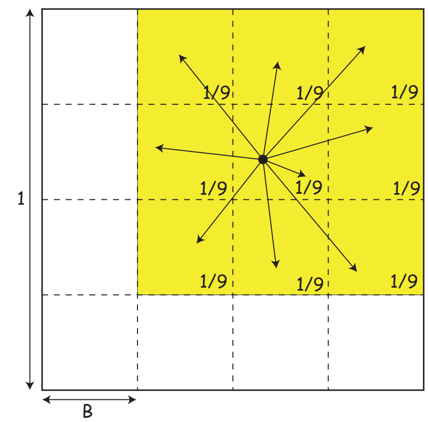

Fig. 1. Two-dimensional hybrid random walk model

(3) **One-dimensional i.i.d. mobility model:** Our one-dimensional i.i.d. mobility model is defined as follows:
   (i) Reasonably, we assume the number of mobile nodes n and source nodes n_s are both even numbers. Among the mobile nodes, $n/2$ nodes (including $n_s/2$ source nodes), named H-nodes, move horizontally; and the other $n/2$ nodes (including the other $n_s/2$ source nodes), named V-nodes, move vertically.
   (ii) Let (x_i, y_i) denote the position of node i. If node i is an H-node, y_i is fixed and x_i is a value randomly uniformly chosen from $[0, 1]$. We also assume that H-nodes are evenly distributed vertically, so y_i takes values $2/n, 4/n, \ldots, 1$. V-nodes have similar properties. See Figure 2.
   (iii) Assume that source and destinations in the same multicast session are the same type of nodes. Also assume that node i is an H-node if i is odd, and a V-node if i is even.
   (iv) The orbit distance of two H(V)-nodes is defined to be the vertical (horizontal) distance of the two nodes.

(4) **One-dimensional hybrid random walk model:** Each orbit is divided into $1/B$ RW-intervals (random walk interval). At each time slot, a node moves into one of two adjacent RW-intervals or stays at the current RW-interval (as in Figure 3). The node position in the RW-interval is randomly, uniformly selected.

We further assume that at each time slot, at most W bits can be transmitted in a successful transmission.

Mobility time scales: Two time scales of mobility are considered in this paper:

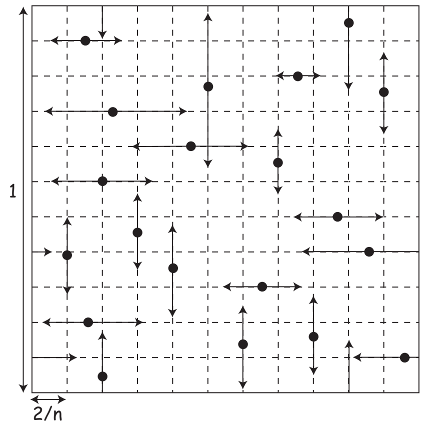

Fig. 2. One-dimensional i.i.d. model

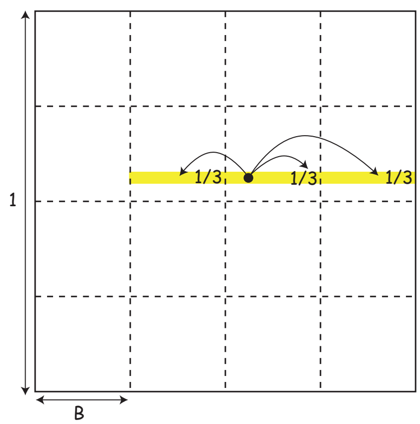

Fig. 3. One-dimensional hybrid random walk model

- **Fast mobility**: The mobility of nodes is at the same time scale as the transmission of packets, i.e., in each time-slot, only one-hop transmission is allowed.
- **Slow mobility**: The mobility of nodes is much slower than the transmission of packets, i.e., multi-hop transmissions may happen within a single time-slot.

Scheduling Policies: We 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, [9]. We say a packet p is successfully multicast if and only if all destinations within the multicast session have received the packet. In each time slot, for each packet p that has not been successfully multicast and each of its unreached destination k, the scheduler needs to perform the following two functions:

- **Capture**: The scheduler needs to decide whether to deliver the packet p to the destination k. If yes, the scheduler then needs to choose one relay node (or source node itself) that has a copy of the packet p at the beginning of the time-slot, and schedule radio transmissions to forward this packet to the destination k 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 k of packet p. We call this chosen relay node the *last mobile relay* for packet p and destination k. And we call the distance between the last mobile relay and the destination a *capture range*.
- **Duplication**: For a packet p that has not been successfully multicast, the scheduler needs to decide whether to *duplicate* packet p to other nodes that do not have the packet at the beginning of the time-slot. The scheduler also needs to decide which nodes to relay from and relay to, and how.

B. Heterogeneous Networks

We introduce m regularly placed base stations (connected with each other via wires) $Z = \{z_1, z_2, \ldots, z_m\}$ into the above pure mobile ad hoc networks and generate a heterogeneous network. Specifically, the base stations are placed at positions $(\frac{1}{2\sqrt{m}} + i\frac{1}{\sqrt{m}}, \frac{1}{2\sqrt{m}} + j\frac{1}{\sqrt{m}})$ with $0 \leq i, j \leq \sqrt{m} - 1$. Clearly, these m regularly distributed base stations divide the original square region into m subregions with side length $\frac{1}{\sqrt{m}}$. Here we assume that m is the square of some integer for simplicity and we use S_i to denote the subregion centered at the base station z_i.

All transmissions can be carried out either in *ad hoc mode* or in *infrastructure mode* (see Figure 4). We assume that the base stations have different transmission bandwidths, denoted by W_i for each, from mobile ad hoc nodes, denoted by W_a for each. Further, we evenly divide the bandwidth W_i into two parts, one for uplink transmissions and the other for downlink transmissions, so that these different kinds of transmissions will not interfere with each other. In each time-slot and for each packet, the scheduler will need to decide whether to use the infrastructure mode or ad hoc mode. A transmission in infrastructure mode is carried out in the following steps:

(i) *Uplink:* The schedular needs to choose a mobile node holding packet p, and transmits this packet to the nearest base station. The uplink transmission can be carried out by using multi-hop transmission under slow mobility models or by using one-hop transmission under fast mobility models.

(ii) *Infrastructure relay:* Once a base station receives a packet from a mobile node, all the other $m - 1$ base stations share this packet immediately, (i.e., the delay is considered to be zero) for they are connected by wires which have a much higher transmission speed.

(iii) *Downlink:* Each base station searches for all the packets needed in its own subregion, and broadcast all of them to their destined mobile nodes. Since base stations have much more powerful antennas, we assume that all base stations can reach all mobile nodes within its subregion directly, and all the downlink transmissions can be done within a single time-slot. So this step suffers no delay either.

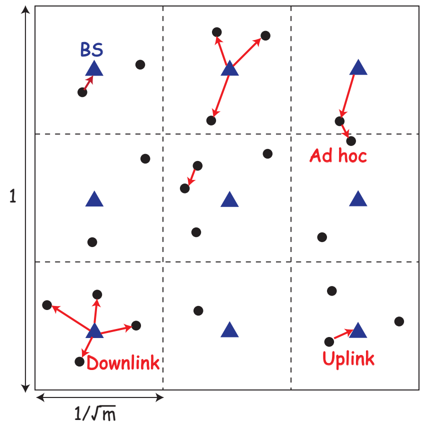

Fig. 4. Heterogeneous network with infrastructure support

III. Two Dimensional I.I.D. Fast Mobility Model

In this section, we present the upper bound on multicast capacity-delay tradeoff under the *two-dimensional i.i.d. fast mobility model*, and then propose a scheme to achieve a capacity close to the upper bound up to logarithmic factors.

A. Upper Bound

Consider packet p and one of its destinations k, let $L_{p,k}$ denote the *capture range* for packet p and destination k, L_p denote the *capture range* for packet p and its last reached destination. Let $D_{p,k}$ denote the number of time slots it takes to reach destination k, D_p denote the number of time slots it takes to reach the last destination of packet p. And let R_p denote the number of mobile relays holding packet p when the packet reaches its last destination. *Note that we use cooperative transmission.* That is, whenever a destination captures packet p, if it is not the last destination, it can still act as a relay node to help deliver the packet to other destinations. Then we have the following lemma.

Lemma 1: Under *two-dimensional i.i.d. mobility model* and concerning successful encounter, the following inequality holds for any causal scheduling policy (c_1 is some positive constant).

$$c_1 \log n \mathbb{E}[D_{p,k}] \geq \frac{1}{\left(\mathbb{E}[L_{p,k}] + \frac{1}{n^2}\right)^2 \mathbb{E}[R_{p,k}]}$$

Proof: See Proposition 1 in [9]. We give some intuitive explanations here. Consider a simpler scenario in which R_p, the number of mobile relays holding packet p, and L_p, the capture range with respect to packet p, are constants in different time slots and for different destinations. Then $1-(1-{L_p}^2)^{R_p}$ is the probability that any one out of the R_p nodes can capture destination k in one time slot. It is easy to show that, the average number of time slots needed to capture destination k, is

$$\mathbb{E}[D_{p,k}] = \frac{1}{1-(1-{L_p}^2)^{R_p}} \geq \frac{1}{{L_p}^2 R_p}$$

■

Deduction of Lemma 1: We let D_p denote the time it takes to reach the last destination of packet p. And let R_p denote the number of mobile relays holding packet p when the packet reaches its last destination. Name the last reached destination of packet p is La_p, it is obvious that $D_p = \max\{D_{p,1}, \dots, D_{p,n_d}\} = D_{p,La_p}$, $R_p = \max\{R_{p,1}, \dots, R_{p,n_d}\} = R_{p,La_p}$. Since the capture range is chosen properly so that there is one and only one relay node within, so as the number of relays increases, the radius of capture range should be no larger than the previous ones. Thus

$$L_{p,La_p} = \min\{L_{p,1}, \dots, L_{p,n_d}\} \triangleq L_p \tag{1}$$

According to *Lemma 1*, these three parameters have the similar inequality.

$$c_1 \log n \mathbb{E}[D_p] \geq \frac{1}{\left(\mathbb{E}[L_p] + \frac{1}{n^2}\right)^2 \mathbb{E}[R_p]} \tag{2}$$

Consider a large enough time interval T. The total number of packets communicated among all sessions is $\lambda n_s T$. Then goes the following lemma,

Lemma 2: Under *fast mobility model* and concerning network radio resources consumption, the following inequality holds for any causal scheduling policy (c_2 is some positive constant).

$$\sum_{p=1}^{\lambda n_s T} \frac{\Delta^2}{4} \frac{\mathbb{E}[R_p] - n_d}{n} + \sum_{p=1}^{\lambda n_s T} \sum_{k=1}^{n_d} \frac{\pi \Delta^2}{4} \mathbb{E}[L_{p,k}^2] \leq c_2 W T \log n \tag{3}$$

Proof: See Appendix A. ■

Theorem 1: Under *two-dimensional i.i.d. fast mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. The following upper bound holds for any causal scheduling policy,

$$\lambda \leq \min\left\{\Theta(1), \Theta\Big(\frac{n}{n_s n_d}\Big), \Theta\Big(\frac{n}{n_s n_d}\sqrt{\frac{n_d D}{n}}\Big)\right\} \tag{4}$$

Proof: Since we assume that no node can transmit and receive over the same frequency at the same time, the following property can be shown as in [1]

$$\frac{WT}{2} n \geq \sum_{p=1}^{\lambda n_s T} \sum_{k=1}^{n_d} 1 = \lambda n_s T n_d$$

Hence, $\lambda \leq \frac{Wn}{2 n_s n_d} = \Theta(\frac{n}{n_s n_d})$. In addition, each source can send out at most W size of packet per time-slot, i.e., $\lambda \leq W = \Theta(1)$. These first two factors are kind of obvious, and we will neglect them in later results.

From the deduction of *Lemma 2*, (2), we have

$$\sum_{p=1}^{\lambda n_s T} \mathbb{E}[R_p] \geq \sum_{p=1}^{\lambda n_s T} \frac{1}{c_1 \log n} \frac{1}{\left(\mathbb{E}[L_p] + \frac{1}{n^2}\right)^2 \mathbb{E}[D_p]} \tag{5}$$

Note that,

$$D = \frac{\sum_{p=1}^{\lambda n_s T} \mathbb{E}[D_p]}{\sum_{p=1}^{\lambda n_s T}} = \frac{\sum_{p=1}^{\lambda n_s T} \mathbb{E}[D_p]}{\lambda n_s T}$$

Using the Cauchy-Schwartz Inequality, we have,

$$\begin{aligned}&\left(\frac{\sum_{p=1}^{\lambda n_s T}\mathbb{E}[D_p]}{\lambda n_s T}\right)\left(\sum_{p=1}^{\lambda n_s T}\frac{1}{\left(\mathbb{E}[L_p]+\frac{1}{n^2}\right)^2\mathbb{E}[D_p]}\right)\\
&\geq \frac{1}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\frac{1}{\mathbb{E}[L_p]+\frac{1}{n^2}}\right)^2\\
&\geq \frac{1}{\lambda n_s T}\left(\frac{\left(\sum_{p=1}^{\lambda n_s T}1\right)^2}{\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n^2}\right)}\right)^2\\
&= \frac{(\lambda n_s T)^3}{\left(\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n^2}\right)\right)^2} \qquad (6)\end{aligned}$$

Equalities hold when $\mathbb{E}[L_p]$ is equal for all p and $\mathbb{E}[D_p]=D$ for all p. Substituting (6) in (5), we have

$$\sum_{p=1}^{\lambda n_s T}\mathbb{E}[R_p]\geq \frac{1}{c_1\log n}\frac{(\lambda n_s T)^3}{D\left(\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n^2}\right)\right)^2} \qquad (7)$$

On the other hand, using Jensens Inequality and Cauchy-Schwartz Inequality, and the definition of L_p in (1) we have,

$$\begin{aligned}&\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\mathbb{E}[L_{p,k}^2]\\
&\geq \sum_{p=1}^{\lambda n_s T} n_d\mathbb{E}[L_p^2]\\
&\geq \sum_{p=1}^{\lambda n_s T} n_d\mathbb{E}^2[L_p]\\
&\geq \frac{n_d\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2}{\sum_{p=1}^{\lambda n_s T}1} \qquad (8)\end{aligned}$$

Equalities hold when $\mathbb{E}[L_{p,k}^2]$ is equal for all k, $k=1,\ldots,n_d$, L_p is almost surely a constant and $\mathbb{E}[L_p]$ is equal for all p. Substituting (7) and (8) into inequality (3), we have,

$$\begin{aligned}\frac{4c_2WT\log n}{\Delta^2} &\geq \sum_{p=1}^{\lambda n_s T}\frac{\mathbb{E}[R_p]-n_d}{n}+\pi\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\mathbb{E}[L_{p,k}^2]\\
&\geq \frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\left(\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n^2}\right)\right)^2}\\
&\quad+\frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2-\lambda T\frac{n_s n_d}{n}\end{aligned}$$

There are two cases we need to consider.

Case 1: If $\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\leq\frac{\lambda n_s T}{n^2}$ then,

$$\begin{aligned}\frac{4c_2WT\log n}{\Delta^2} &\geq \frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\left(\frac{2\lambda n_s T}{n^2}\right)^2}-\lambda T\frac{n_s n_d}{n}\\
&= \frac{1}{4c_1\log n}\frac{\lambda T n_s n^3}{D}-\lambda T\frac{n_s n_d}{n}\end{aligned}$$

When $D=o(\frac{n}{n_d})$, the first term dominates when n is large. Hence, for n large enough,

$$\begin{aligned}\frac{4c_2WT\log n}{\Delta^2} &\geq \frac{1}{8c_1\log n}\frac{\lambda T n_s n^3}{D}\\
\lambda &\leq \frac{32c_1c_2W}{\Delta^2}\frac{D\log^2 n}{n_s n^3} \qquad (9)\end{aligned}$$

Case 2: If $\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\geq\frac{\lambda n_s T}{n^2}$ then,

$$\begin{aligned}\frac{4c_2WT\log n}{\Delta^2} &\geq \frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\left(2\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2}\\
&\quad+\frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2-\lambda T\frac{n_s n_d}{n}\\
&\geq 2\sqrt{\frac{1}{c_1 n\log n}\frac{\pi n_d}{\lambda n_s T}\frac{(\lambda n_s T)^3}{4D}}-\lambda T\frac{n_s n_d}{n}\\
&= \lambda T\frac{n_s n_d}{n}\sqrt{\frac{\pi n}{c_1 n_d D\log n}}-\lambda T\frac{n_s n_d}{n} \qquad (10)\end{aligned}$$

Since $D=o\left(\frac{n}{n_d\log n}\right)$, the first term dominates when n is large, i.e.,

$$\begin{aligned}\frac{4c_2WT\log n}{\Delta^2} &\geq \lambda T\frac{n_s n_d}{2n}\sqrt{\frac{\pi n}{c_1 n_d D\log n}}\\
\lambda &\leq \sqrt{\frac{64c_1c_2^2W^2}{\pi\Delta^4}}\frac{n}{n_s n_d}\sqrt{\frac{D\log^3 n}{n}n_d} \qquad (11)\end{aligned}$$

Finally, we compare the two inequalities we have obtained, i.e., (8) and (11). Since $D=o\left(\frac{n}{n_d}\right)$ inequality (11) will eventually be the loosest for large n. Hence, the optimal capacity-delay tradeoff is upper bounded by

$$\lambda\leq\Theta\left(\frac{n}{n_s n_d}\sqrt{\frac{D\log^3 n}{n}n_d}\right)$$

■

B. Achievable Lower Bound

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

Choosing Optimal Values of Key Parameters:

From *Theorem 1*, we have

$$\lambda=\Theta\left(\frac{n}{n_s n_d}\sqrt{\frac{n_d D\log^3 n}{n}}\right)=\Theta(n^{-\frac{2s+\alpha-1-d}{2}}\log^{\frac{3}{2}}n)$$

In order to achieve the maximum capacity on the right hand side, all inequalities in the proofs of Theorem 1 should hold with equality. By studying the conditions under which these inequalities are tight, we will be able to identify that the optimal choices of various key parameters of the scheduling policy. We can infer that the parameters (such as $\mathbb{E}[D_{p,k}]$,

$\mathbb{E}[L_{p,k}]$) of each packet p and each destination node k should be the same and concentrate on their respective average values. This implies that the scheduling policy should use the same parameters for all packets and all destinations. We further assume that $n_s = n^s, 0 \le s \le 1$; $n_d = n^\alpha, 0 \le \alpha \le 1$ and $D = n^d, 0 \le d < 1 - \alpha$. In addition, we limit the mobile nodes $n \le n_s n_d$.

From inequality (10), we get the *capture range* L when it becomes equality,

$$\begin{aligned} \frac{1}{c_1 n \log n} \frac{(\lambda n_s T)^3}{D\left(2\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2} &= \frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2 \\ \frac{1}{4c_1 n \log n}\frac{(\lambda n_s T)^3}{D(\lambda n_s T L)^2} &= \frac{\pi n_d}{\lambda n_s T}(\lambda n_s T L)^2 \\ L &= \Theta(n^{-\frac{1+\alpha+d}{4}}/\log^{\frac{1}{4}} n) \end{aligned}$$

From (1), we have

$$R = \Theta(\frac{1}{DL^2 \log n}) = \Theta(n^{\frac{1+\alpha-d}{2}}/\log^{\frac{1}{2}} n)$$

The optimal values are summarized in Table I. Note that the optimal values are independent of the number of sources n_s.

TABLE I
THE ORDER OF THE OPTIMAL VALUES OF THE PARAMETERS IN TWO-DIMENSIONAL FAST I.I.D. MOBILITY MODEL. ($n_s = n^s$ MULTICAST SESSIONS ARE INCLUDED, EACH OF WHICH HAS $n_d = n^\alpha$ DESTINATIONS. DELAY IS BOUNDED BY $D = n^d$.)

L: Capture Range	$\Theta(n^{-\frac{1+\alpha+d}{4}}/\log^{\frac{1}{4}} n)$
R: # of Duplicates	$\Theta(n^{\frac{1+\alpha-d}{2}}/\log^{\frac{1}{2}} n)$

Capacity Achieving Scheme I:

We propose a more flexible cell-partitioning scheme, [9], to achieve a capacity that is close to the upper bound, using broadcasting and time division. Cell-partitioning schemes, like [7] and [9], divide the network into several non-overlapping and independent cells and only allow transmissions within the same cell. As *Lemma 2* in [36] shows, each cell in the network can transmit at a rate of $c_3 W$, where c_3 is a deterministic positive constant.

We group every D time-slots into a super-slot.

(1) At each odd super-slot, we schedule transmissions from the sources to the relays in every time-slot. We divide the unit square into $\mathcal{C}_d = \Theta\left(\frac{n^{(1-\alpha+d)/2}}{\log n}\right)$ cells. Each cell is a square of area $1/\mathcal{C}_d$. We refer to each cell in the odd super-slot as a *duplication cell*. By *Lemma 6* in [9], each cell can be active for $1/c_4$ amount of time, where c_4 is some constant. When a cell is scheduled to be active, each source node in the cell broadcasts a new packet to all other nodes in the same cell for $\Theta\left(\frac{n^{-(2s+\alpha-1-d)/2}}{\log^2 n}\right)$ amount of time. These other nodes then serve as mobile relays for the packet. The nodes within the same *duplication cell* coordinate themselves to broadcast sequentially.

(2) At each even super-slot, we schedule transmissions from the mobile relays to the destination nodes in every time-slot. Note that the positions of the mobile relays have changed and are now independent of their positions in the previous time-slots. We divide the unit square into $\mathcal{C}_c = \Theta\left(n^{(1+\alpha+d)/2}\right)$ cells. Each cell is a square of area $1/\mathcal{C}_c$. We refer to each cell in the even super-slot as the *capture cell*. In each time-slot, for each destination node $\mathcal{D}$ and each of its source node $\mathcal{S}$, pick a node $Y_{\mathcal{SD}}$ that is in the same *capture cell* with node $\mathcal{D}$ in current time-slot and in the same *duplication cell* with node $\mathcal{S}$ some time-slot in previous super-slot and hold a copy of the packet source node $\mathcal{S}$ generated in that very time-slot. If there are multiple relay nodes, just pick one, which we call a *representative relay*, and transmit the destined packet to $\mathcal{D}$. At the end of each even super-slot, clear all the buffers of mobile nodes, and prepare for a new turn of duplication and capture.

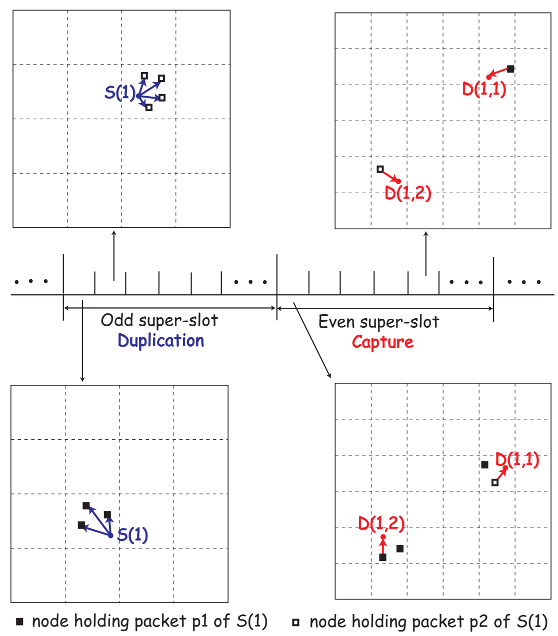

Fig. 5. Capacity Achieving Scheme I

See Figure 5 for better understanding of our scheme. We can show that as $n \to \infty$, with high probability (w.h.p.), all packets generated in odd duplication super-slot will finish its n_d destined transmission within the following even capture super-slot.

Proposition 1: With probability approaching one, as $n \to \infty$, the above scheme allows each source to send D packets of size $\lambda = \Theta\left(\frac{n^{-(2s+\alpha-1-d)/2}}{\log^2 n}\right)$ to their respective destinations within $2D$ time-slots.

Proof: Similar but simpler proof of *Proposition 2*. ■

Remarks: Our scheme uses different cell-partitioning in the odd super-slot that that in the even super-slot. The size of the duplication cell is chosen such that the average number of nodes in each cell, $n/\mathcal{C}_d$, is close to the optimal value of R.

The size of the capture cell is chosen such that its area, $1/\mathcal{C}_c$ is close to the optimal value of L^2.

IV. Two Dimensional I.I.D. Slow Mobility Model

In this section, we present the upper bound on multicast capacity-delay tradeoff under the *two-dimensional i.i.d. slow mobility model*, and then propose a scheme to achieve a capacity close to the upper bound up to logarithmic factors.

A. Upper Bound

Recall that under slow mobility model, the mobility of nodes is much slower than the transmission of packets, i.e., multi-hop transmissions may happen within a single time-slot. For packet p and one of its destinations k, $1 \le k \le n_d$, let $h_{p,k}$ denote the number of hops packet p takes during the *capture* to node k. And let $S^h_{p,k}$, $h = 1, 2, \ldots, h_{p,k}$ denote the length of each hop. Hence, similar to *Lemma 2*, the following lemma holds,

Lemma 3: Under *slow mobility model* and concerning network radio resources consumption, the following inequality holds for any causal scheduling policy (c_4 is some positive constant).

$$\sum_{p=1}^{\lambda n_s T}\frac{\Delta^2}{4}\frac{\mathbb{E}[R_p]-n_d}{n}+\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}\frac{\pi\Delta^2}{4}\mathbb{E}\big[(S^h_{p,k})^2\big] \le c_5 WT\log n. \quad (12)$$

Where the sum of each hop's length of the $h_{p,k}$ hops must be no smaller than the straight-line distance, i.e., the capture radius:

$$\sum_{h=1}^{h_{p,k}} S^h_{p,k} \ge L_{p,k} \quad (13)$$

Theorem 2: Under *two-dimensional i.i.d. slow mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. The following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d D}{n}}\right) \quad (14)$$

Proof: Since a node can either transmit or receive at one time, it is easy to see that,

$$\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}} 1 \le \frac{WT}{2}n \quad (15)$$

Using the Cauchy-Schwartz inequality and (15), we have

$$\begin{aligned}&\left(\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}} S^h_{p,k}\right)^2\\ \le\ &\left(\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}} 1\right)\left(\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}(S^h_{p,k})^2\right)\\ \le\ &\frac{WTn}{2}\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}(S^h_{p,k})^2\end{aligned}$$

Equalities hold when (15) becomes an equality and $S^h_{p,k}$ is equal for all p, k and h.

Further, using Jensen's Inequality and (13), we have,

$$\begin{aligned}&\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}\mathbb{E}\big[(S^h_{p,k})^2\big]\\ =\ &\mathbb{E}\Big[\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}(S^h_{p,k})^2\Big]\\ \ge\ &\frac{2}{WTn}\mathbb{E}\Bigg[\Big(\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}S^h_{p,k}\Big)^2\Bigg]\\ \ge\ &\frac{2}{WTn}\left(\mathbb{E}\Big[\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}S^h_{p,k}\Big]\right)^2\\ \ge\ &\frac{2}{WTn}\left(\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\mathbb{E}[L_{p,k}]\right)^2\\ \ge\ &\frac{2n_d^2}{WTn}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2 \quad (16)\end{aligned}$$

Equalities hold when (16) is tight, $L_{p,k}$ is equal for all k, $\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}} S^h_{p,k}$ is almost surely a constant and (13) becomes an equality.

Substituting (16) and previous result (7) into inequality (12), we have,

$$\begin{aligned}&\frac{4c_5 WT\log n}{\Delta^2}\\ \ge\ &\sum_{p=1}^{\lambda n_s T}\frac{\mathbb{E}[R_p]-n_d}{n}+\pi\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\sum_{h=1}^{h_{p,k}}\mathbb{E}\big[(S^h_{p,k})^2\big]\\ \ge\ &\frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\Big(\sum_{p=1}^{\lambda n_s T}\big(\mathbb{E}[L_p]+\frac{1}{n^2}\big)\Big)^2}\\ &+\frac{2\pi n_d^2}{WTn}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2-\lambda T\frac{n_s n_d}{n}\end{aligned}$$

Case 1: If $\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p] \le \frac{\lambda n_s T}{n^2}$ then,

$$\begin{aligned}&\frac{4c_5 WT\log n}{\Delta^2}\\ \ge\ &\frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\Big(\frac{2\lambda n_s T}{n^2}\Big)^2}-\lambda T\frac{n_s n_d}{n}\\ =\ &\frac{1}{4c_1\log n}\frac{\lambda T n_s n^3}{D}-\lambda T\frac{n_s n_d}{n}\end{aligned}$$

When $D = o(\frac{n}{n_d})$, the first term dominates when n is large. Hence, for n large enough,

$$\begin{aligned}\frac{4c_5 WT\log n}{\Delta^2} &\ge \frac{1}{8c_1\log n}\frac{\lambda T n_s n^3}{D}\\ \lambda &\le \frac{32c_1c_5 W}{\Delta^2}\frac{D\log^2 n}{n_s n^3} \quad (17)\end{aligned}$$

Case 2: If $\sum_{p=1}^{\lambda n_s T} \mathbb{E}[L_p] \geq \frac{\lambda n_s T}{n^2}$ then,

$$\begin{aligned} &\frac{4c_5 WT\log n}{\Delta^2} \\ \geq\ & \frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\Big(2\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\Big)^2} \\ &+\frac{2\pi n_d^2}{WTn}\Big(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\Big)^2 - \lambda T\frac{n_s n_d}{n} \\ \geq\ & 2\sqrt{\frac{1}{c_1 n\log n}\frac{2\pi n_d^2}{WTn}\frac{(\lambda n_s T)^3}{4D}} - \lambda T\frac{n_s n_d}{n} \quad (18) \\ =\ & \lambda T\frac{n_s n_d}{n}\sqrt{\frac{2\pi\lambda n_s}{c_1 WD\log n}} - \lambda T\frac{n_s n_d}{n} \end{aligned}$$

Therefore, since $D = o(\frac{n}{n_d})$, either

$$\lambda \leq \frac{c_1 W}{2\pi}\frac{D\log n}{n_s} \quad (19)$$

Or if $\lambda = \omega(\frac{D\log n}{n_s})$, then the first term in (18) dominates when n is large. In this case,

$$\begin{aligned} \frac{4c_5 WT\log n}{\Delta^2} &\geq \lambda T\frac{n_s n_d}{n}\sqrt{\frac{2\pi\lambda n_s}{c_1 WD\log n}} \\ \lambda &\leq \sqrt[3]{\frac{32c_1c_5^2W^3}{\Delta^4}}\frac{n}{n_s n_d}\sqrt[3]{\frac{D\log^3 n}{n}n_d} \quad (20) \end{aligned}$$

Finally, we compare the three inequalities we have obtained, i.e., (17), (19) and (20). Since $D = o\big(\frac{n}{n_d}\big)$ inequality (20) will eventually be the loosest for large n. Hence, the optimal capacity-delay tradeoff is upper bounded by

$$\lambda \leq \Theta\Bigg(\frac{n}{n_s n_d}\sqrt[3]{\frac{D\log^3 n}{n}n_d}\Bigg)$$

■

B. Achievable Lower Bound

We will first calculate the critical values of key parameters in our two-dimensional i.i.d. slow mobility model, and then propose our cell-partitioned scheme II.

Choosing Optimal Values of Key Parameters:

From *Theorem 2*, we have

$$\lambda = \Theta\Bigg(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d D\log^3 n}{n}}\Bigg) = \Theta(n^{-\frac{3s+2\alpha-2-d}{3}}\log n)$$

From inequality (18), we get the *capture range* L^2 when it becomes equality,

$$\begin{aligned} \frac{1}{c_1 n\log n}\frac{(\lambda n_s T)^3}{D\Big(2\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\Big)^2} &= \frac{2\pi n_d^2}{WTn}\Big(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\Big)^2 \\ \frac{1}{4c_1 n\log n}\frac{(\lambda n_s T)^3}{D(\lambda n_s TL)^2} &= \frac{2\pi n_d^2}{WTn}(\lambda n_s TL)^2 \\ L &= \Theta(n^{-\frac{1+2\alpha+2d}{6}}/\log^{\frac{1}{2}} n) \end{aligned}$$

By setting (15) to equality, we have

$$\begin{aligned} \lambda n_s T n_d H &= \frac{WT}{2}n \\ H &= \Theta(n^{\frac{1-\alpha-d}{3}}/\log n) \end{aligned}$$

By setting (13) to equality and recall that all the values are uniformed, we have

$$S = \frac{L}{H} = \Theta\Big(\sqrt{\frac{\log n}{n}}\Big)$$

From (2), we can get

$$R = \Theta(\frac{1}{DL^2\log n}) = \Theta(n^{\frac{1+2\alpha-d}{3}})$$

We summarize the optimal values in Table II.

TABLE II
THE ORDER OF THE OPTIMAL VALUES OF THE PARAMETERS IN TWO-DIMENSIONAL SLOW I.I.D. MOBILITY MODEL. ($n_s = n^s$ MULTICAST SESSIONS ARE INCLUDED, EACH OF WHICH HAS $n_d = n^\alpha$ DESTINATIONS. DELAY IS BOUNDED BY $D = n^d$.)

L: Capture Range	$\Theta(n^{-\frac{1+2\alpha+2d}{6}}/\log^{\frac{1}{2}} n)$
R: # of Duplicates	$\Theta(n^{\frac{1+2\alpha-d}{3}})$
H: # of Hops	$\Theta(n^{\frac{1-\alpha-d}{3}}/\log n)$
S: Hop Length	$\Theta(\sqrt{\log n/n})$

Capacity Achieving Scheme II: We group every D time-slots into a super-slot.

(1) At each odd super-slot, we schedule transmissions from the sources to the relays in every time-slot. We divide the unit square into $\mathcal{C}_d = \Theta\big(\frac{n^{(2-2\alpha+d)/3}}{\log n}\big)$ cells. Each cell is a square of area $1/\mathcal{C}_d$. We refer to each cell in the odd super-slot as a *duplication cell*. By *Lemma 6* in [9], each cell can be active for $1/c_4$ amount of time, where c_4 is some constant. When a cell is scheduled to be active, each node in the cell broadcasts a new packet to all other nodes in the same cell for $\Theta\big(\frac{n^{-(3s+2\alpha-2-d)/3}}{\log^2 n}\big)$ amount of time. These other nodes then serve as mobile relays for the packet. The nodes within the same *duplication cell* coordinate themselves to broadcast sequentially.

(2) At each even super-slot, we schedule transmissions from the mobile relays to the destination nodes in every time-slot. Note that the positions of the mobile relays have changed and are now independent of their positions in the previous time-slots. We divide the unit square into $\mathcal{C}_c = \Theta\big(n^{(1+2\alpha+2d)/3}\big)$ cells. cell is a square of area $1/\mathcal{C}_c$. We refer to each cell in the even time slot as the *capture cell*. In each time-slot, for each destination node $\mathcal{D}$ and each of its source node $\mathcal{S}$, pick a node $Y_{\mathcal{SD}}$ that is in the same *capture cell* with node $\mathcal{D}$ in current time-slot and in the same *duplication cell* with node $\mathcal{S}$ some time-slot in previous super-slot and hold a copy of the packet source node $\mathcal{S}$ generated in that very time-slot. If there are multiple relay nodes, just pick one, which we call it *representative relay*. We then schedule multi-hop

transmissions in the following fashion to forward each packet from the *representative relay* to its destination in the same *capture cell*. We further divide each *capture cell* into $\mathcal{C}_h = \Theta\left(\frac{n^{(2-2\alpha-2d)/3}}{\log n}\right)$ *hop-cells* (in $\sqrt{\mathcal{C}_h}$ rows and $\sqrt{\mathcal{C}_h}$ columns, see Figure 6). Each hop-cell is a square of area $1/(\mathcal{C}_c\mathcal{C}_h)$. By *Lemma 6* in [9], there exists a scheduling scheme where each hop-cell can be active for $1/c_4$ amount of time. When each hop-cell is active, it forwards a packet to another node in the neighboring hop-cell. If the destination of the packet is in the neighboring cell, the packet is forwarded directly to the destination node. The packets from each *representative relay* are first forwarded towards neighboring cells along the X-axis, then to their destination nodes along the Y-axis. At the end of each even super-slot, clear all the buffers of mobile nodes, and prepare for a new turn of duplication and capture.

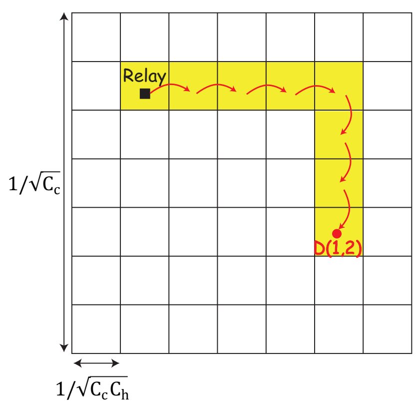

Fig. 6. Multi-hop scheme in capture cell.

Proposition 2: With probability approaching one, as $n \to \infty$, the above scheme allows each source to send D packets of size $\lambda = \Theta\left(\frac{n^{-(3s+2\alpha-2-d)/3}}{\log^2 n}\right)$ to their respective destinations within $2D$ time-slots.

Proof: See Appendix B. ∎

<u>*Remarks:*</u> In our scheme, the size of the hop-cell is chosen such that each hop to the neighboring hop-cell is of length $1/\sqrt{\mathcal{C}_c\mathcal{C}_h}$, which is close to the optimal value of S.

V. Two Dimensional Hybrid R.W. Mobility Model

In this section, we study the two-dimensional hybrid random walk mobility model with both fast and slow mobiles. We will obtain the maximum throughput for $D = \omega(|\log B|/B^2)$.

From Appendix G in [10], we can have the following lemma.

Lemma 4: Under *two-dimensional hybrid random walk mobility model*, with given delay constraint $D = \omega(|\log B|/B^2)$, for any $L \in [0, B/\sqrt{\pi})$, we have

$$Pr(\widetilde{L}_p \leq L) \leq 36L^2D \tag{21}$$

Where $\widetilde{L}_p$ is the minimum distance between the *representative relay* node to the destination, which is introduced in the proof of *Lemma 1*.

Compared with two-dimensional i.i.d. mobility model, where $Pr(\widetilde{L}_p \leq L) \leq \pi L^2 D$, Lemma 4 is only different in the coefficient, which does not influence the orders of the final result. So following the same proof procedure of Theorem 1 and Theorem 2, we have the following two results in two-dimensional hybrid random walk model with fast and slow mobiles respectively.

Theorem 3: Under *two-dimensional hybrid random walk fast mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $B = o(1)$ and $D = \omega(|\log B|/B^2)$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt{\frac{n_d D}{n}}\right) \tag{22}$$

Theorem 4: Under *two-dimensional hybrid random walk slow mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $B = o(1)$ and $D = \omega(|\log B|/B^2)$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d D}{n}}\right) \tag{23}$$

As the two-dimensional random walk mobility model has the same capacity upper bound as two-dimensional i.i.d. mobility model, capacity achieving scheme I and scheme II still apply to R.W. mobility model with fast and slow mobiles respectively, with some extra limitations on delay constraint. We do not extend this part for space limitations.

VI. One Dimensional I.I.D. Fast Mobility Model

In this section, we study the one-dimensional i.i.d. fast mobility model.

A. Upper Bound

As under the two-dimensional i.i.d. fast mobility model, *Lemma 2* holds for one-dimensional i.i.d. fast mobility model. Besides, we have the following lemma.

Lemma 5: Under *one-dimensional i.i.d. mobility model* and concerning successful encounter, the following inequality holds for any causal scheduling policy (c_6 is some positive constant).

$$c_6 \log n \mathbb{E}[D_{p,k}] \geq \frac{1}{(\mathbb{E}[L_{p,k}] + \frac{1}{n})\mathbb{E}[R_{p,k}]}$$

Proof: Let ρ_p denote the distance from any one mobile node of packet p to one of its destinations in a particular time-slot. Under one-dimensional i.i.d. mobility model, when two nodes are vertical to each other, $\widetilde{L}_p \leq L$ holds only if at some time-slot within D, they are in the square with side length $2L$ as in Figure 7. In this scenario, we have

$$Pr(\rho_p \leq L) \leq 4L^2$$

When the orbits of these two nodes are parallel to each other, then

$$Pr(\rho_p \leq L) \leq 2L$$

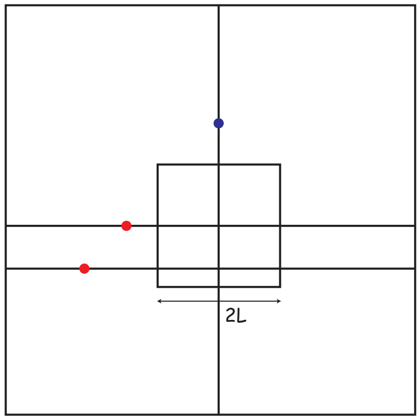

Fig. 7. Vertical and parallel orbits.

Thus, for $L \le 1/2$, we can conclude that

$$Pr(\rho_p \le L) \le 2L \tag{24}$$

For $L \ge 1/2$, we have

$$Pr(\rho_p \le L) \le 4L^2 \tag{25}$$

The rest of the proof is available in Appendix C. ■

Obviously, the deduction of this lemma holds too

$$c_6 \log n \mathbb{E}[D_p] \ge \frac{1}{(\mathbb{E}[L_p] + \frac{1}{n})\mathbb{E}[R_p]} \tag{26}$$

Theorem 5: Under *one-dimensional i.i.d. fast mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $D = o(\frac{\sqrt{n}}{n_d})$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d^2 D^2}{n}}\right) \tag{27}$$

Proof: From the deduction of *Lemma 5*, (2), we have

$$\sum_{p=1}^{\lambda n_s T} \mathbb{E}[R_p] \ge \sum_{p=1}^{\lambda n_s T} \frac{1}{c_6 \log n} \frac{1}{(\mathbb{E}[L_p] + \frac{1}{n})\mathbb{E}[D_p]} \tag{28}$$

Using the Cauchy-Schwartz Inequality, we have,

$$\begin{aligned} &\left(\frac{\sum_{p=1}^{\lambda n_s T}\mathbb{E}[D_p]}{\lambda n_s T}\right)\left(\sum_{p=1}^{\lambda n_s T}\frac{1}{(\mathbb{E}[L_p]+\frac{1}{n})\mathbb{E}[D_p]}\right) \\ \ge\ & \frac{1}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\frac{1}{\sqrt{\mathbb{E}[L_p]+\frac{1}{n}}}\right)^2 \\ \ge\ & \frac{1}{\lambda n_s T}\left(\frac{\left(\sum_{p=1}^{\lambda n_s T}1\right)^2}{\sum_{p=1}^{\lambda n_s T}\sqrt{\mathbb{E}[L_p]+\frac{1}{n}}}\right)^2 \\ \ge\ & \frac{(\lambda n_s T)^3}{(\lambda n_s T)\left(\sum_{p=1}^{\lambda n_s T}(\mathbb{E}[L_p]+\frac{1}{n})\right)} \\ =\ & \frac{(\lambda n_s T)^2}{\sum_{p=1}^{\lambda n_s T}(\mathbb{E}[L_p]+\frac{1}{n})} \end{aligned} \tag{29}$$

Substituting (29) in (28), we have

$$\sum_{p=1}^{\lambda n_s T}\mathbb{E}[R_p] \ge \frac{1}{c_6 \log n}\frac{(\lambda n_s T)^2}{D\left(\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n}\right)\right)} \tag{30}$$

Substituting (30) and (8) into inequality (3), we have,

$$\begin{aligned} \frac{4c_2 W T \log n}{\Delta^2} &\ge \sum_{p=1}^{\lambda n_s T}\frac{\mathbb{E}[R_p]-n_d}{n} + \pi\sum_{p=1}^{\lambda n_s T}\sum_{k=1}^{n_d}\mathbb{E}[L_{p,k}^2] \\ &\ge \frac{1}{c_6 n \log n}\frac{(\lambda n_s T)^2}{D\left(\sum_{p=1}^{\lambda n_s T}\left(\mathbb{E}[L_p]+\frac{1}{n}\right)\right)} \\ &\quad + \frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2 - \lambda T\frac{n_s n_d}{n} \end{aligned}$$

Case 1: If $\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p] \le \frac{\lambda n_s T}{n}$ then,

$$\begin{aligned} \frac{4c_2 W T \log n}{\Delta^2} &\ge \frac{1}{c_6 n \log n}\frac{(\lambda n_s T)^2}{D\frac{2\lambda n_s T}{n}} - \lambda T\frac{n_s n_d}{n} \\ &= \frac{1}{2c_6 \log n}\frac{\lambda T n_s}{D} - \lambda T\frac{n_s n_d}{n} \end{aligned}$$

When $D = o(\frac{n}{n_d \log n})$, the first term dominates when n is large. Hence, for n large enough,

$$\begin{aligned} \frac{4c_2 W T \log n}{\Delta^2} &\ge \frac{1}{4c_6 \log n}\frac{\lambda T n_s}{D} \\ \lambda &\le \frac{16 c_6 c_2 W}{\Delta^2}\frac{D \log^2 n}{n_s} \end{aligned} \tag{31}$$

Case 2: If $\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p] \ge \frac{\lambda n_s T}{n}$ then,

$$\begin{aligned} &\frac{4c_2 W T \log n}{\Delta^2} \\ \ge\ & \frac{1}{c_6 n \log n}\frac{(\lambda n_s T)^2}{D\left(2\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)} \\ &+ \frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2 - \lambda T\frac{n_s n_d}{n} \\ \ge\ & 3\sqrt[3]{\left(\frac{1}{c_6 n \log n}\frac{(\lambda n_s T)^2}{4D}\right)^2\frac{\pi n_d}{\lambda n_s T}} - \lambda T\frac{n_s n_d}{n} \\ =\ & \lambda T\frac{n_s n_d}{n}\sqrt[3]{\frac{27\pi n}{16 c_6^2 n_d^2 D^2 \log^2 n}} - \lambda T\frac{n_s n_d}{n} \end{aligned} \tag{32}$$

Since $D = o(\frac{\sqrt{n}}{n_d})$, the first term dominates when n is large, i.e.,

$$\begin{aligned} \frac{4c_2 W T \log n}{\Delta^2} &\ge \lambda T\frac{n_s n_d}{n}\sqrt[3]{\frac{27\pi n}{16 c_6^2 n_d^2 D^2 \log^2 n}} \\ \lambda &\le \sqrt[3]{\frac{16 c_2^3 c_2^6 W^3}{\pi \Delta^6}}\frac{4n}{3 n_s n_d}\sqrt[3]{\frac{D^2 \log^5 n}{n}n_d^2} \end{aligned} \tag{33}$$

Finally, we compare the two inequalities we have obtained, i.e., (31) and (33). Since $D = o\left(\frac{\sqrt{n}}{n_d}\right)$ inequality (33) will eventually be the loosest for large n. Hence, the optimal capacity-delay tradeoff is upper bounded by

$$\lambda \leq \Theta\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{D^2 \log^5 n}{n}}n_d^2\right)$$

■

B. Achievable Lower Bound

Choosing Optimal Values of Key Parameters:

From *Theorem 5*, we have the optimal multicast tradeoff as

$$\lambda = \Theta\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d^2 D^2 \log^5 n}{n}}\right) = \Theta(n^{-\frac{3s+\alpha-2-2d}{3}} \log^{\frac{5}{3}} n)$$

From inequality (32), we get the *capture range* L when it becomes equality,

$$\begin{aligned}
\frac{1}{c_6 n \log n}\frac{(\lambda n_s T)^2}{D\left(4\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)} &= \frac{\pi n_d}{\lambda n_s T}\left(\sum_{p=1}^{\lambda n_s T}\mathbb{E}[L_p]\right)^2 \\
\frac{1}{4c_6 n \log n}\frac{(\lambda n_s T)^2}{D\lambda n_s TL} &= \frac{\pi n_d}{\lambda n_s T}(\lambda n_s TL)^2 \\
L &= \Theta(n^{-\frac{1+\alpha+d}{3}}/\log^{\frac{1}{3}} n)
\end{aligned}$$

From (26), we have

$$R = \Theta\left(\frac{1}{DL\log n}\right) = \Theta(n^{\frac{1+\alpha-2d}{3}}/\log^{\frac{2}{3}} n)$$

We summarize our optimal values in Table III.

TABLE III
THE ORDER OF THE OPTIMAL VALUES OF THE PARAMETERS IN ONE-DIMENSIONAL FAST I.I.D. MOBILITY MODEL. ($n_s = n^s$ MULTICAST SESSIONS ARE INCLUDED, EACH OF WHICH HAS $n_d = n^\alpha$ DESTINATIONS. DELAY IS BOUNDED BY $D = n^d$.)

L: Capture Range	$\Theta(n^{-\frac{1+\alpha+d}{3}}/\log^{\frac{1}{3}} n)$
R: # of Duplicates	$\Theta(n^{\frac{1+\alpha-2d}{3}}/\log^{\frac{2}{3}} n)$

Capacity Achieving Scheme III:

Carefully study the fundamental difference between two-dimensional models and one-dimensional models and we find that, with the same capture range and delay constraint, one-dimensional mobile can reach a higher probability for successful capture when the orbits of the two nodes in the capture procedure are parallel. Thus, to fully take advantage of this benefit, we assume capture only happens within two parallel nodes, defined as *H(V) capture*. However, what if the orbit of the source node is far away from those of destination nodes? To offset this flaw, we introduce those V(H) nodes whose orbits are vertical with them and let these nodes relay the packet to some orbits close to the destinations.

The transmission of a packet in the H(V) multicast session will go through H(V)-V(H) duplication, V(H)-H(V) duplication and H(V)-H(V) capture, three procedures, sequentially (see Figure 8).

We propose a flexible *rectangle-partition* scheme, similar to [10], to achieve a capacity-delay tradeoff that is close to the upper bound. Rectangle-partition model divides the unit square into multiple horizontal rectangles, named as H-rectangles; and multiple vertical rectangles, named as V-rectangles as in Figure 8. A packet is said to be destined to a rectangle if the orbit of one of its destinations is contained in the rectangle. Each H-rectangle and V-rectangle cross to form a cell, and transmissions only happen in the same *crossing cell*.

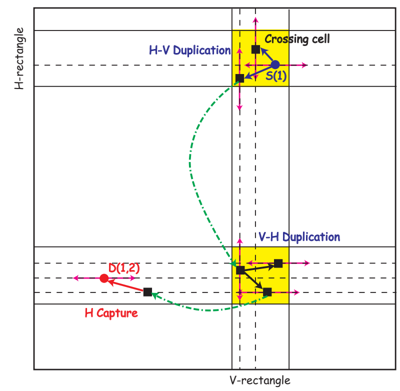

Fig. 8. One-dimensional transmissions in scheme III.

We group every D time-slots into a super-slot, and let z denote any non-negative integer.

(1) At each $3z+1$ super-slot, we schedule transmissions from the H(V)-sources to the V(H)-relays in every time-slot. We divide the unit square into $\mathcal{R}_d$ H-rectangles and $\mathcal{R}_d$ V-rectangles, i.e., $\mathcal{R}_d^2 = \Theta\left(\frac{n^{(2-\alpha+2d)/3}}{\log n}\right)$ crossing cells. Each cell is a square of area $1/\mathcal{R}_d^2$. We refer to each cell in the $3z+1$ super-slot as a *duplication cell*. Note that we use the same partition in *V-H duplication* and *H-V duplication*, so we name the crossing cell as the same. By *Lemma 6* in [9], each cell can be active for $1/c_4$ amount of time, where c_4 is some constant. When a cell is scheduled to be active, each H(V)-source node in the cell broadcasts a new packet to all other V(H)-nodes in the same cell for $\Theta\left(\frac{n^{-(3s+\alpha-2-2d)/3}}{\log^2 n}\right)$ amount of time. These other V(H)-nodes then serve as mobile V(H)-relays for the packet to complete the *V(H)-H(V) duplications* in the next super-slot. The source nodes within the same *duplication cell* coordinate themselves to broadcast sequentially.

(2) At each $3z+2$ super-slot, we schedule transmissions from the V(H)-relay nodes to the H(V)-relay nodes in every time-slot. Note that the positions of the mobile relays have changed and are now independent of their positions in the previous time-slots. We use the same partition method as the one used in $3z+1$ super-slot. When a cell is scheduled to be active, search for V(H)-relay nodes holding the packet, which is destined to the H(V)-rectangle containing this crossing cell and has not been *V(H)-H(V) duplicated* yet. If there are multiple satisfied V(H)-nodes for one packet, randomly choose one and broadcast the packet to all other H(V)-nodes in the same cell. We can easily prove

that with R V(H)-relay nodes for each packet p, which are generated in *H(V)-V(H) duplication* of former $3z+1$ super-slot, w.h.p., there must be a time-slot within this $3z+2$ super-slot that at least one of them reaches the destined H(V)-rectangle of packet p. And under proper scheduling, the throughput in this period cannot be smaller than that in $3z+1$ super-slot.

(3) At each $3z+3$ super-slot, we schedule transmissions from the mobile H(V)-relays to the H(V)-destination nodes in every time-slot. We divide the unit square into $\mathcal{R}_c = \Theta\big(n^{(1+\alpha+d)/3}\big)$ H-rectangles and $\mathcal{R}_c$ V-rectangles, i.e., $\mathcal{R}_c^2$ crossing cells. Each cell is a square of area $1/\mathcal{R}_c^2$. We refer to each cell in the $3z+3$ super-slot as the *capture cell*. In each time-slot, for each H(V)-destination node $\mathcal{D}$ and each of its destined packet p, search for H(V)-relay nodes in the same *capture cell* holding packet p. If there are multiple ones, randomly pick one, which we call a *representative H(V)-relay*, and transmit the destined packet p to $\mathcal{D}$. In the end of each $3z+3$ super-slot, clear all the buffers of mobile nodes, and prepare for a new turn of duplications and capture.

Following the proof of Proposition 2, we have

Proposition 3: With probability approaching one, as $n \to \infty$, the above scheme allows each source to send D packets of size $\lambda = \Theta\big(\frac{n^{-(3s+\alpha-2-2d)/3}}{\log^2 n}\big)$ to their respective destinations within $3D$ time-slots.

<u>*Remarks:*</u> The size of the duplication cell is chosen such that the average number of nodes in each cell, $n/\mathcal{R}_d^2$, is close to the optimal value of R. The size of the capture cell is chosen such that its area, $1/\mathcal{R}_c^2$ is close to the optimal value of L^2.

VII. One Dimensional I.I.D. Slow Mobility Model

In this section, we study the one-dimensional i.i.d. slow mobility model.

A. Upper Bound

Lemma 3 and *Lemma 5* hold for one dimensional i.i.d. slow mobility model. By these two lemmas, and following the proof of *Theorem 5*, we have the following theorem.

Theorem 6: Under *one-dimensional i.i.d. slow mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $D = o\big(\frac{\sqrt{n}}{n_d}\big)$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[4]{\frac{n_d^2 D^2}{n}}\right) \tag{34}$$

B. Achievable Lower Bound

The optimal values of key parameters of one-dimensional i.i.d. slow mobility model are given in Table IV.

TABLE IV
THE ORDER OF THE OPTIMAL VALUES OF THE PARAMETERS IN ONE-DIMENSIONAL SLOW I.I.D. MOBILITY MODEL. ($n_s = n^s$ MULTICAST SESSIONS ARE INCLUDED, EACH OF WHICH HAS $n_d = n^\alpha$ DESTINATIONS. DELAY IS BOUNDED BY $D = n^d$.)

L: Capture Range	$\Theta(n^{-\frac{1+2\alpha+2d}{4}} / \log^{\frac{3}{4}} n)$
R: # of Duplicates	$\Theta(n^{\frac{1+2\alpha-2d}{4}} / \log^{\frac{1}{4}} n)$
H: # of Hops	$\Theta(n^{\frac{1-2\alpha-2d}{4}} / \log^{\frac{5}{4}} n)$
S: Hop Length	$\Theta(\sqrt{\log n/n})$

Capacity achieving scheme IV: We group every D time-slots into a super-slot, and let z denote any non-negative integer.

(1) At each $3z+1$ super-slot, we schedule transmissions from the H(V)-sources to the V(H)-relays in every time-slot. We divide the unit square into $\mathcal{R}_d$ H-rectangles and $\mathcal{R}_d$ V-rectangles, i.e., $\mathcal{R}_d^2 = \Theta\big(\frac{n^{(3-2\alpha+2d)/4}}{\log n}\big)$ crossing cells. Each cell is a square of area $1/\mathcal{R}_d^2$. We refer to each cell in the $3z+1$ super-slot as a *duplication cell*. Note that we use the same partition in *V-H duplication* and *H-V duplication*, so we name the crossing cell as the same. By *Lemma 6* in [9], each cell can be active for $1/c_4$ amount of time, where c_4 is some constant. When a cell is scheduled to be active, each H(V)-source node in the cell broadcasts a new packet to all other V(H)-nodes in the same cell for $\Theta\big(\frac{n^{-(4s+2\alpha-3-2d)/4}}{\log^2 n}\big)$ amount of time. These other V(H)-nodes then serve as mobile V(H)-relays for the packet to complete the *V(H)-H(V) duplications* in the next super-slot. The source nodes within the same *duplication cell* coordinate themselves to broadcast sequentially.

(2) At each $3z+2$ super-slot, we schedule transmissions from the V(H)-relay nodes to the H(V)-relay nodes in every time-slot. Note that the positions of the mobile relays have changed and are now independent of their positions in the previous time-slots. We use the same partition method as the one used in $3z+1$ super-slot. When a cell is scheduled to be active, search for V(H)-relay nodes holding the packet, which is destined to the H(V)-rectangle containing this crossing cell and has not been *V(H)-H(V) duplicated* yet. If there are multiple satisfied V(H)-nodes for one packet, randomly choose one and broadcast the packet to all other H(V)-nodes in the same cell. We can easily prove that with R V(H)-relay nodes for each packet p, generated in *H(V)-V(H) duplication* of former $3z+1$ super-slot, w.h.p., there must be a time-slot within this $3z+2$ super-slot that at least one of them reaches the destined H(V)-rectangle of packet p. And under proper scheduling, the throughput in this period cannot be smaller than that in $3z+1$ super-slot.

(3) At each $3z+3$ super-slot, we schedule transmissions from the mobile H(V)-relays to the H(V)-destination nodes in every time-slot. We divide the unit square into $\mathcal{R}_c = \Theta\big(n^{(1+2\alpha+2d)/4}\big)$ H-rectangles and $\mathcal{R}_c$ V-rectangles, i.e., $\mathcal{R}_c^2$ crossing cells. Each cell is a square of area $1/\mathcal{R}_c^2$. We refer to each cell in the $3z+3$ super-slot as the *capture cell*. In each time-slot, for each H(V)-destination node $\mathcal{D}$ and each of its destined packet p, search for H(V)-relay nodes in the same *capture cell* holding packet p. If there are multiple ones, randomly pick one, which we call a *representative H(V)-relay*. We then schedule multi-

hop transmissions in the following fashion to forward this destined packet p from the *representative H(V)-relay* to $\mathcal{D}$. We further divide each *capture cell* into $\mathcal{R}_h = \Theta\big(\frac{n^{(1-2\alpha-2d)/4}}{\sqrt{\log n}}\big)$ H-rectangles and $\mathcal{R}_h$ V-rectangles, i.e., $\mathcal{R}_h^2$ crossing *hop-cells*. Each *hop-cell* is a square of side length $1/(\mathcal{R}_c\mathcal{R}_h)$. By *Lemma 6* in [9], there exists a scheduling scheme where each hop-cell can be active for $1/c_4$ amount of time. When each hop-cell is active, it forwards a packet to another H(V)-node in the neighboring *hop-cell*. If the H(V)-destination node of the packet is in the neighboring cell, the packet is forwarded directly to the H(V)-destination node. The packets from each *representative H(V)-relay* are first forwarded towards neighboring cells along the X-axis, then to their destination nodes along the Y-axis. At the end of each $3z+3$ super-slot, clear all the buffers of mobile nodes, and prepare for a new turn of duplications and capture.

Proposition 4: With probability approaching one, as $n \to \infty$, the above scheme allows each source to send D packets of size $\lambda = \Theta\big(\frac{n^{-(4s+2\alpha-3-2d)/4}}{\log^2 n}\big)$ to their respective destinations within $3D$ time-slots.

Remarks: In our scheme, the size of the hop-cell is chosen such that each hop to the neighboring hop-cell is of length $1/\sqrt{\mathcal{R}_c\mathcal{R}_h}$, which is close to the optimal value of S.

VIII. One Dimensional Hybrid R.W. Mobility Model

In this section, we present the optimal multicast capacity-delay tradeoffs of the one-dimensional hybrid random walk mobility models with both fast and slow mobiles. The results can be established by following similar analysis as one-dimensional i.i.d. mobility models.

Theorem 7: Under *one-dimensional hybrid random walk fast mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $B = o(1)$ and $D = \omega(1/B^2)$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[3]{\frac{n_d^2 D^2}{n}}\right) \tag{35}$$

Theorem 8: Under *one-dimensional hybrid random walk slow mobility model*, let D be the mean delay averaged over all packets, and let λ be the capacity per multicast session. When $B = o(1)$ and $D = \omega(1/B^2)$, the following upper bound holds for any causal scheduling policy,

$$\lambda = O\left(\frac{n}{n_s n_d}\sqrt[4]{\frac{n_d^2 D^2}{n}}\right) \tag{36}$$

IX. Results Discussions

Our results of optimal multicast capacity-delay tradeoffs in mobile ad-hoc networks give a global perspective for the following reasons:

- It generalizes the optimal delay-throughput tradeoffs in unicast traffic pattern in [10], when taking $n_s = n$ and $n_d = 1$.
- It generalizes the multicast capacity result $O(\sqrt{D/n_s})$ under delay constraint in [32], which is better than [31], when considering the two-dimensional i.i.d. fast mobility model and taking $n_s n_d = n$.

We summarize our results in Table V. Setting $n_s = n$ and $n_d = 1$, our results are shown in the second column. Setting $n_s = n$ and $n_d = k$, our results are shown in the third column.

TABLE V
Optimal multicast capacity and delay tradeoffs in MANETs: a global perspective

λ (i.i.d./hybrid r.w.)	unicast	multicast
2D fast mobility	$O\left(\sqrt{\frac{D}{n}}\right)$	$O\left(\frac{1}{k}\sqrt{\frac{D}{n}k}\right)$
2D slow mobility	$O\left(\sqrt[3]{\frac{D}{n}}\right)$	$O\left(\frac{1}{k}\sqrt[3]{\frac{D}{n}k}\right)$
1D fast mobility	$O\left(\sqrt[3]{\frac{D^2}{n}}\right)$	$O\left(\frac{1}{k}\sqrt[3]{\frac{D^2}{n}k^2}\right)$
1D slow mobility	$O\left(\sqrt[4]{\frac{D^2}{n}}\right)$	$O\left(\frac{1}{k}\sqrt[4]{\frac{D^2}{n}k^2}\right)$

X. Mobility Dimension and Capacity

Since M. Grossglauser and D. Tse [3] proved that mobility can increase capacity, lots of works have been done to study the improvement of capacity by introducing different kinds of mobility. Results do verify the capacity improvement. Recently, Pan Li et al. [37] studied a restricted mobility model. Specifically they assumed a network of unit area with n nodes. The network is evenly divided into $n^{2\alpha}$ cells, each of which is further evenly divided into squares with an area of $n^{-2\beta}$. All nodes can only move inside the cell where they are initially distributed, and at the beginning of each time slot, every node moves from its current square to a uniformly chosen point in a uniformly chosen adjacent square. The capacity is established in such networks with lower bound $n^{2\beta-\alpha-1}$ and upper bound $n^{\beta-\alpha-\frac{1}{2}}$. They showed that *restriction on mobility range decreases capacity*.

However, our one-dimensional mobility models, where nodes can move either horizontally or vertically, can even achieve larger capacities than two-dimensional models. This intuition-contradictory result can be explained by the following reasons:

- *One-dimensional mobility models assume lower dimensional mobility in higher dimensional space.* Note that if the space is also one-dimensional, the results will be the same as those in our two-dimensional mobility models.
- *Mobilities are not restricted within limited range.* Though nodes are limited to only moving horizontally or vertically, yet the mobility range on their orbit lines is not restricted.
- *It uses compensating mobile nodes to help relay.* For H(V) multicast sessions, the V(H)-relay nodes are used to compensate for the lack of vertical(horizontal) mobility.

Therefore, the one-dimensional mobility model in our paper is actually a *hybrid dimensional model*, where one-dimensional mobile nodes transmit packets in two-dimensional space. Lower dimensional mobility has its advantage in that the mobility pattern is simple and easy to predict, thus increasing the inter contact rate.

To better understand the idea of hybrid dimensional model, we take three-dimensional space transmission with lower dimensional mobiles as an example. It is our future work to study the capacity improvement of these networks.

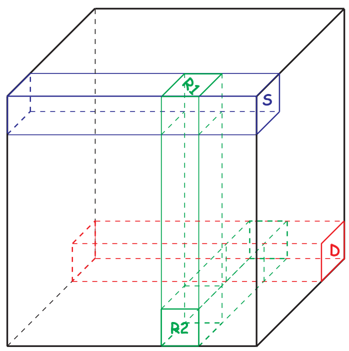

Fig. 9. Transmission in three-dimensional space with one-dimensional mobile multicast sessions and two kinds of one-dimensional mobile relays.

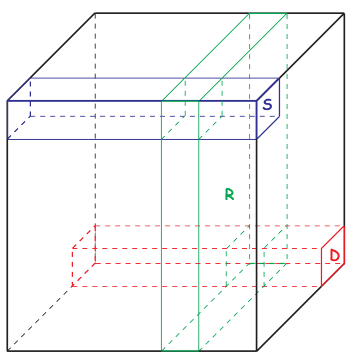

Fig. 10. Transmission in three-dimensional space with one-dimensional mobile multicast sessions and two-dimensional mobile relays.

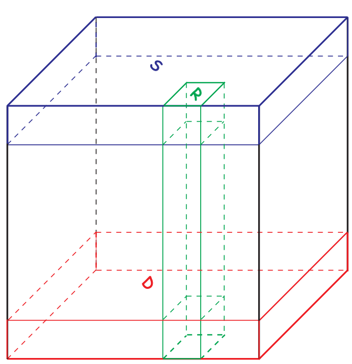

Fig. 11. Transmission in three-dimensional space with two-dimensional mobile multicast sessions and one-dimensional mobile relays.

XI. Heterogeneous Networks with Infrastructure Support

As we know, in heterogeneous networks, transmissions can be carried out either in infrastructure mode or in ad hoc mode (see Figure 12). Let T_i and T_a denote an achievable *aggregate input throughput* of the whole network when all the transmissions are carried out in infrastructure mode and in ad hoc mode, respectively. Then, the maximum aggregate throughput achievable in heterogeneous wireless networks, denoted by T, can be calculated as follows:

$$T = \max\{T_i, T_a\} \tag{37}$$

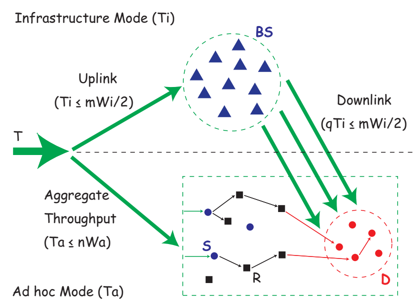

Fig. 12. Two modes in heterogeneous networks.

T_a has been studied in former sections under different mobility models. We only discuss T_i here.

Since all base stations share information simultaneously, they can be regarded as an integrated relay. We define this specific huge relay as *BS relay*, with the same maximum input and output throughput $mW_i/2$.

Under multicast traffic pattern, multiple deliveries are needed to reach all the n_d destinations. Hence, for each input flow of the *BS relay*, there are at least q output flows. Note that $q \neq n_d$, for that there may be multiple destination nodes within the same subregion and only one broadcast is enough for all of them. Therefore, T_i can be calculated as

$$T_i = \Theta(\frac{m}{q}) \tag{38}$$

Lemma 6: For each packet, and in each subregion, w.h.p., there are at most $\Theta(\frac{n_d}{m})$ destination nodes when $m = o(n_d)$, and at most $\Theta(1)$ destination nodes when $m = \omega(n_d)$.

Proof: Consider subregion i. Let X_i be a random variable denoted as the number of destination nodes in subregion i, and $\mathbb{E}[X_i]$ the expectation of X_i. Then, we have $\mathbb{E}[X_i] = \frac{n_d}{m}$.

Recall the Chernoff Bound in [23]:

For any $\delta > 0$,

$$Pr\big(X_i > (1+\delta)\mathbb{E}[X_i]\big) < e^{-\mathbb{E}[X_i]f(\delta)} \tag{39}$$

Wherein $f(\delta) = (1+\delta)\log(1+\delta) - \delta$.

1) $m = o(n_d)$, i.e., $0 \leq \beta < \alpha \leq 1$.

According to the Chernoff bound (39), we can obtain that

$$Pr(X_i > 2\frac{n_d}{m}) < e^{-\frac{n_d}{m}f(1)}$$

Wherein $f(1) = 2\log 2 - 1 > 0$. Since $0 \le \beta < \alpha$, when n is large enough, we have

$$\begin{aligned} Pr(X_i \le 2\frac{n_d}{m} \text{ for any i }) &\ge 1 - mPr(X_i > 2\frac{n_d}{m}) \\ &> 1 - n^{\beta} e^{-n^{\alpha-\beta} f(1)} \\ &\to 1 \end{aligned}$$

2) $m = \omega(n_d)$, i.e., $0 \le \alpha < \beta \le 1$.
Again, according to the Chernoff bound (39), we can obtain that

$$\begin{aligned} Pr(X_i > (1+\delta)\frac{n_d}{m}) &< e^{-\frac{n_d}{m}\left[(1+\delta)\log(1+\delta)-\delta\right]} \\ &= \frac{e^{\delta\frac{n_d}{m}}}{(1+\delta)^{(1+\delta)\frac{n_d}{m}}} \end{aligned}$$

Let $(1+\delta)\frac{n_d}{m} = c_7$. where c_7 is a constant that will be determined later. Then we have

$$\begin{aligned} Pr(X_i > c_7) &< \frac{e^{c_7 - n^{\alpha-\beta}}}{c_7 n^{c_7(\beta-\alpha)}} \\ &= \frac{e^{2-n^{\alpha-\beta}}}{c_7 n^{c_7(\beta-\alpha)}} \end{aligned}$$

Hence

$$\begin{aligned} Pr(X_i \le c_7 \text{ for any i }) &\ge 1 - mPr(X_i > c_7) \\ &> 1 - \frac{e^{2-n^{\alpha-\beta}}}{c_7 n^{(c_7-1)\beta - c_7\alpha}} \end{aligned}$$

When we choose $c_7 > \frac{\beta}{\beta-\alpha}$, which makes $(c_7-1)\beta - c_7\alpha > 0$, recall that $\alpha < \beta$, this probablitiy goes to 1 when n is large enough.

■

By *Lemma 6* and (38), we have

Theorem 9: In heterogeneous networks with wire connected base stations as infrastructure support, the maximum aggregate capacity throughput provided by infrastructure mode is

$$T_i = \max\left\{\Theta(1), \Theta(\frac{m}{n_d})\right\} \quad (40)$$

Therefore, $T = \max\left\{\Theta(1), \Theta(\frac{m}{n_d}), T_a\right\}$. Note that as $n_d \le n$ and $D = \Omega(1)$, T_a is always larger than $\Theta(1)$ regardless of the node mobility models. Thus,

$$T = \max\left\{\Theta(\frac{m}{n_d}), T_a\right\} = \max\left\{\Theta(\frac{m}{n_d}), n_s\lambda_a\right\} \quad (41)$$

By comparing the T_i and T_a under all the eight mobility models presented above, we summarize our maximum aggregate capacity throughput in all the heterogeneous networks in Table VI. The aggregate capacity throughput of heterogeneous networks is closely related with the number of base stations, m. To achieve a larger capacity than its homogeneous counterpart, the number of base stations in the heterogeneous network must be larger than m_c, as is given by Table VI.

TABLE VI
CRITICAL VALUES OF m, AND CORRESPONDING T. WHEN $m < m_c$, THE AD HOC MODE IS PREFERRED AND $T = T_a$. OTHERWISE, INFRASTRUCTURE MODE IS PREFERRED AND $T = T_i = \Theta(\frac{m}{n_d})$.

	m_c	$T = T_a$
2D fast mobility	$\Theta\left(\sqrt{nn_dD}\right)$	$\Theta\left(\sqrt{\frac{nD}{n_d}}\right)$
2D slow mobility	$\Theta\left(\sqrt[3]{n^2n_dD}\right)$	$\Theta\left(\sqrt[3]{\frac{n^2D}{n_d^2}}\right)$
1D fast mobility	$\Theta\left(\sqrt[3]{n^2n_d^2D^2}\right)$	$\Theta\left(\sqrt[3]{\frac{n^2D^2}{n_d}}\right)$
1D slow mobility	$\Theta\left(\sqrt[4]{n^3n_d^2D^2}\right)$	$\Theta\left(\sqrt[4]{\frac{n^3D^2}{n_d^2}}\right)$

XII. CONCLUSION

In this paper, we have studied the multicast capacity-delay tradeoffs in both homogeneous and heterogeneous mobile networks. Specifically, in homogeneous networks, we established the upper bound on the optimal multicast capacity-delay tradeoffs under two-dimensional/one-dimensional i.i.d./hybrid random walk fast/slow mobility models and proposed capacity achieving schemes to achieve capacity close to the upper bound. In addition, we find that though the one-dimensional mobility models constrain the direction of nodes' mobility, it achieves larger capacity than two-dimensional models. Then we give a discussion on the relationship between mobility dimension and capacity, and propose our *hybrid dimensional model*. In heterogeneous networks, with base stations as infrastructure support, we give the aggregate input capacity of the whole network under different mobility models.

REFERENCES

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," in *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, March 2000.

[2] M. Franceschetti, O. Dousse, D. N. Tse and P. Thiran, "Closing the Gap in the Capacity of Wireless Networks via Percolation Theory," in *IEEE Transactions on Information Theory*, 53(3):1009-1018, 2007.

[3] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, August 2002.

[4] A. E. Gamal, J. Mammen, B. Prabhakar and D. Shah, "Throughput-delay trade-off in wireless networks," in *Proc. IEEE Infocom*, 2004, pp. 475-485.

[5] X. Lin, G. Sharma, R. R. Mazumdar and N. B. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with Brownian mobility," in *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 277-2784, June 2006.

[6] G. Sharma, R. Mazumdar and N. Shroff, "Delay and capacity tradeoffs in mobile ad hoc networks: A global perspective," in *Proc. IEEE Infocom*, April 2006, pp. 1-12.

[7] M. Neely and E. Modiano, "Capacity and Delay Tradeoffs for Ad-hoc Mobile Networks," in *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1917-1937, 2005.

[8] S. Toumpis and A. Goldsmith, "Large Wireless Networks under Fading, Mobility, and Delay Constraints," in *Proc. IEEE INFOCOM*, 2004, pp. 609-619.

[9] X. Lin and N. B. Shroff, "The Fundamental Capacity-Delay Tradeoff in Large Mobile Ad Hoc Networks" in *Proc. Third Annu. Mediterranean Ad Hoc Netw. Workshop*, 2004.

[10] L. Ying, S. Yang and R. Srikant, "Optimal delay-throughput trade-offs in mobile ad-hoc networks," in *IEEE Transactions on Information Theory*, vol. 9, no. 54, pp. 4119-4143, September 2008.
[11] J. Mammen and D. Shah, "Throughput and delay in random wireless networks with restricted mobility," in *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1108-1116, March 2007.
[12] G. Sharma and R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," in *Proc. IEEE ICC*, 2004, pp. 3869-3873.
[13] G. Sharma and R. Mazumdar, "Delay and Capacity Trade-Off in Wireless Ad Hoc Networks with Random Way-Point Mobility," School of ECE, Purdue Univ., W. Lafayette, IN, 2005. Proprint. Available: [On-line] at http://ece.purdue.edu/ gsharma.
[14] P. Li, Y. Fang and J. Li, "Throughput, Delay, and Mobility in Wireless Ad Hoc Networks," IEEE International Conference on Computer Communications (INFOCOM10), San Diego, CA, March 15-19, 2010.
[15] S. N. Diggavi, M. Grossglauser and D. Tse., "Even one-dimensional mobility increases ad hoc wireless capacity," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2002, p. 352.
[16] A. Agarwal and P. Kumar, "Capacity Bounds for Ad Hoc and Hybrid Wireless Networks." in *ACM SIGCOMM Computer Communication Review*, 34(3):71-81, July 2004.
[17] U. Kozat and L. Tassiulas, "Throughput Capacity of Random Ad Hoc Networks with Infrastructure Support." in *Proceedings of ACM MobiCom*, San Diego, CA, USA, June 2003.
[18] P. Li, C. Zhang and Y. Fang, "Capacity and Delay of Hybrid Wireless Broadband Access Networks." in *IEEE Journal on Selected Areas in Communications (JSAC)* - Special Issue on Broadband Access Networks, 27(2):117-125, February 2009.
[19] B. Liu, Z. Liu and D. Towsley, "On the Capacity of Hybrid Wireless Networks." in *Proceeding of IEEE INFOCOM*, San Francisco, California, USA, March 2003.
[20] A. Zemlianov and G. Veciana, "Capacity of Ad Hoc Wireless Networks with Infrastructure Support." in *IEEE Journal on Selected Areas in Communications*, 23(3), March 2005.
[21] S. Zhao and D. Raychaudhuri, "On the Scalability of Hierarchical Hybrid Wireless Networks." in *Proceeding of IEEE CISS*, Princeton, NJ, USA, March 2006.
[22] R. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Duxbury Press, 1996.
[23] T. Cormen, C. Leiserson, R. Rivest and C. Stein, *Introduction to Algorithms (2nd ed.)*, MIT Press, 2001.
[24] P. Jacquet and G. Rodolakis, "Multicast Scaling Properties in Massively Dense Ad Hoc Networks," in *Proc. ICPADS*, July 2005.
[25] X. Li, S. Tang and O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," in *ACM MobiCom*, Sept. 2007.
[26] S. Shakkottai, X. Liu and R. Srikant, "The multicast capacity of large multihop wireless networks," in *Proc. ACM MobiHoc*, Sept. 2007.
[27] A. Keshavarz-Haddad and R. Riedi, "Multicast capacity of large homogeneous multihop wireless networks," in *Proc. WiOPT*, Apr. 2008.
[28] Z. Wang, S. Karande, H. R. Sadjadpour and J. J. Garcia-Luna-Aceves, "On the capacity improvement of multicast traffic with network coding," in *Proc. MILICOM*, Sept. 2008.
[29] U. Niesen, P. Gupta and D. Shah, "The multicast capacity region of large wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2009.
[30] U. Lee, S. Y. Oh, K. W. Lee and M. Gerla, "Relaycast: Scalable multicast routing in delay tolerant networks."in *ICNP'08*, Orlando, Florida, Oct. 2008.
[31] C. Hu, X. Wang and F. Wu, "MotionCast: On the Capacity and Delay Tradeoffs," in *ACM MobiHoc 09*, New Orleans, May 2009.
[32] S. Zhou and Lei Ying. "On Delay Constrained Multicast Capacity of Large-Scale Mobile Ad-Hoc Networks." In *Proc. INFOCOM 2010 mini-conference*, San Diego, CA, 2010.
[33] A. Keshavarz-Haddad, V. Ribeiro and R. Riedi, "Broadcast capacity in multihop wireless networks." in *ACM MobiCom* 2006, pp. 239-250.
[34] P. Li and Y. Fang, "Impacts of Topology and Traffic Pattern on Capacity of Hybrid Wireless Networks." in *IEEE Transactions on Mobile Computing*, 8(12):1585-1595, December 2009.
[35] X. Mao, X. Li and S. Tang, "Multicast Capacity for Hybrid Wireless Networks." in *ACM MobiHoc 2008*.
[36] P. Li and Y. Fang, The Capacity of Heterogeneous Wireless Networks, in IEEE International Conference on Computer Communications (INFOCOM 2010), San Diego, CA, March 15-19, 2010.
[37] P Li, Y Fang and J Li, "Throughput, Delay, and Mobility in Wireless Ad Hoc Networks", in *Proc. INFOCOM 2010*.

APPENDIX A – PROOF OF *Lemma 2*

Here are some intuitive explanations for *Lemma 2*. Proofs are similar to and can be easily inferred from Appendix B in [9].

Consider the Protocol Model, [1]. By the interference constraint, if nodes i and j directly transmit to nodes k and l, respectively, at the same time, we have

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

That is, disks of radius $\frac{\Delta}{2}$ times the transmission range centered at the transmitter are disjoint from each other. This property motivates us to measure the radio resources each transmission consumes by the areas of these disjoint disks, [1]. Next we will calculate the radio resources consumption during the *Duplication* and *Capture* procedures, respectively.

- *Capture*: For each packet p and each of its destination k, the *one-hop capture*[1] consumes area of $\frac{\pi\Delta^2}{4}(L_{p,k})^2$. Hence, the lower bound on the expected area consumed by all n_d successful captures of packet p is $\sum_{k=1}^{n_d} \frac{\pi\Delta^2}{4}\mathbb{E}[L_{p,k}^2]$.
- *Duplication*: If the radius of transmission range is s, then, w.h.p., there are $\pi s^2 n$ nodes which can receive the broadcast packets, and a disk of area $\frac{\pi\Delta^2}{4}s^2$ centered at the transmitter will be disjoint from others. Therefore, we can use $\frac{\Delta^2}{4}\frac{\mathbb{E}[R_p]-n_d}{n}$ as a lower bound on the expected area consumed by producing $R_p - n_d$ copies of the packet to other nodes before any of them or the source itself successfully forwards the packet to the last destination and finally finishes its trip. Note that since we use cooperative mode [31], where destinations can also act as relays, the copies produced in *Duplication* should not only exclude the source node but also exclude the $n_d - 1$ destinations which receive the copies in *Capture* procedure.

APPENDIX B – PROOF OF *Proposition 2*

We will focus on the case when the mean delay is bounded by a constant, i.e., $D = 1$. Let $\lfloor x \rfloor$ be the largest integer smaller than or equal to x. We use the following values[2]: $\mathcal{C}_d = \lfloor (\frac{n^{(2-2\alpha)/3}}{8\log n})^{\frac{1}{2}} \rfloor^2$, $\mathcal{C}_c = \lfloor (n^{(1+2\alpha)/3})^{\frac{1}{2}} \rfloor^2$, $\mathcal{C}_h = \lfloor (\frac{n^{(2-2\alpha)/3}}{4\log n})^{\frac{1}{2}} \rfloor^2$. Inspired by the proof of Proposition 7 in [9], which studied the *unicast* scenario, we show that our scheme can obtain a capacity of $\frac{W}{32n^{(3s+2\alpha-2)/3}\log n}$ w.h.p. when concerning *multicast* pattern. First, we present a lemma which will be used frequently in later proof. It has already been proven in [9] (see Lemma 11).

Lemma 7: Consider an experiment where we randomly throw n balls into $m \leq n$ independent urns. The success probability for each ball to enter any one of the urns is $p \leq 1$.

[1]When concerning the *multi-hop capture*, consumption area is summed up by each hop transmission.

[2]To ensure the positive values, we assume $\alpha < 1$.

Let $B_i, i = 1, \ldots, m$ be the number of balls in urn i after n balls are thrown. Then the expectation of B_i is $\mathbb{E}[B_i] = \frac{np}{m}$. And as $n \to \infty$, we have

(a) If $\frac{np}{m} \geq c \log n$, and $c \geq 8$, then

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

(b) If $\frac{np}{m} \geq cn^{\alpha}$, where $c > 0$ and $\alpha > 0$, then

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

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

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

Analysis of *Duplication*

We consider the experiment in which we throw n_s balls into $\mathcal{C}_d$ urns with $p = 1$. As

$$16n^{(3s+2\alpha-2)/3} \log n \geq \frac{n_s}{\mathcal{C}_d} \geq 8n^{(3s+2\alpha-2)/3} \log n$$

Let $N_d(i)$ denote the number of source nodes in *duplication cell* i. Since $n \leq n_s n_d$, i.e., $s + \alpha \geq 1$, by *Lemma 7* (a), we have

$$\begin{aligned} & \mathbf{P}[N_d(i) \geq 32n^{(3s+2\alpha-2)/3} \log n \text{ for any } i] \\ \leq \quad & \mathbf{P}[N_d(i) \geq 2\frac{n_s}{\mathcal{C}_d} \text{ for any } i] \leq \frac{1}{n} \end{aligned}$$

Hence, w.h.p., there are no more than $32n^{(3s+2\alpha-2)/3} \log n$ source nodes within the same *duplication cell*. Then using time division, we can let each source broadcast a packet for $\frac{1}{32n^{(3s+2\alpha-2)/3} \log n}$ amount of time in sequence.

Analysis of *Capture*

We consider the experiment in which we throw n balls into $\mathcal{C}_c\mathcal{C}_h$ urns with $p = 1$. As

$$\frac{n}{\mathcal{C}_c\mathcal{C}_h} \geq 4 \log n$$

Let $N_h(i)$ denote the number of nodes in *hopping cell* i, by *Lemma 7* (c), we have

$$\mathbf{P}[N_h(i) = 0 \text{ for any } i] = O(\frac{1}{n})$$

Hence, w.h.p., there is always a node in each *hopping cell* helping the multi-hop transmission.

Then we consider the experiment in which we throw n balls into $\mathcal{C}_d\mathcal{C}_c$ urns with $p = 1$. As

$$16 \log n \geq \frac{n}{\mathcal{C}_d\mathcal{C}_c} \geq 8 \log n$$

Let $N_{dc}(i, j)$ denote the number of nodes that are in *duplication cell* i in the previous time-slot and now in *capture cell* j in the current time-slot. By *Lemma 7* (c), we have

$$\mathbf{P}[N_{dc}(i, j) = 0 \text{ for any } i, j] = O(\frac{1}{n})$$

Hence, w.h.p., in each *capture cell* j, there is always a node which used to be in *duplication cell* i, and it has all the packets broadcasted in that duplication cell. If there are multiple satisfied nodes, we only pick one for each i as *representative relay*, so there are $\mathcal{C}_d$ *representative relays* in *capture cell* j. On the other hand, all packets can be found in each *capture cell*, and each destination node can find all the destined packets it desires from the *representative relays*. Hence, we just need to calculate the maximum transmissions passing through each *hopping cell* if all desired transmissions are allowed.

We consider the possible *transmission pairs* instead of the actual mobile nodes. We classify the destinations based on the sessions they belong to, i.e., one destination may be calculated multiple times when it belongs to different sessions. Thus, there are $n_s n_d$ number of pairs either with different transmission nodes or requiring different packets.

For the transmissions horizontally passing through the *hopping cell*. We consider the experiment in which we throw $n_s n_d$ balls into $\mathcal{C}_d\mathcal{C}_c$ urns with $p = 1$. As

$$16n^{s+\alpha-1} \log n \geq \frac{n_s n_d}{\mathcal{C}_d\mathcal{C}_c} \geq 8n^{s+\alpha-1} \log n$$

Let $N_s(i, j)$ denote the number of *transmission pairs* whose source nodes are located in *duplication cell* i and destination nodes are located in *capture cell* j. By *Lemma 7* (a), we have

$$\begin{aligned} & \mathbf{P}[N_s(i, j) \geq 32n^{s+\alpha-1} \log n \text{ for any } i, j] \\ \leq \quad & \mathbf{P}[N_s(i, j) \geq 2\frac{n_s n_d}{\mathcal{C}_d\mathcal{C}_c} \text{ for any } i, j] \leq \frac{1}{n_s n_d} \end{aligned}$$

Hence, w.h.p., in each *capture cell* j, each *representative relay* will serve no more than $32n^{s+\alpha-1} \log n$ *transmission pairs*.

Since $\mathcal{C}_d$ *representative relays* are chosen in each *capture cell*, we consider the experiment where we throw $\mathcal{C}_d$ balls into $\sqrt{\mathcal{C}_h}$ urns with $p = 1$. As

$$\frac{\sqrt{2}n^{(1-\alpha)/3}}{4\sqrt{\log n}} \geq \frac{\mathcal{C}_d}{\sqrt{\mathcal{C}_h}} \geq \frac{n^{(1-\alpha)/3}}{8\sqrt{\log n}}$$

Let $N_r(l)$ denote the number of *representative relays* in row l. Since $\alpha < 1$, by *Lemma 7* (b), we have

$$\begin{aligned} & \mathbf{P}[N_r(j, l) \geq \frac{\sqrt{2}n^{(1-\alpha)/3}}{2\sqrt{\log n}} \text{ for any } j, l] \\ \leq \quad & \mathbf{P}[N_r(j, l) \geq 2\frac{\mathcal{C}_d}{\sqrt{\mathcal{C}_h}} \text{ for any } j, l] = O(\frac{1}{\mathcal{C}_d}) \to 0 \end{aligned}$$

Hence, w.h.p., the number of horizontal transmissions $T_x = N_s(i, j)N_r(j, l) \leq 16n^{(3s+2\alpha-2)/3}\sqrt{2 \log n}$.

For the transmissions vertically passing through the *hopping cell*. We consider the experiment where we throw $n_s n_d$ balls into $\mathcal{C}_c\sqrt{\mathcal{C}_h}$ urns with $p = 1$. As

$$4n^{(3s+2\alpha-2)/3}\sqrt{2 \log n} \geq \frac{n_s n_d}{\mathcal{C}_c\sqrt{\mathcal{C}_h}} \geq 2n^{(3s+2\alpha-2)/3}\sqrt{2 \log n}$$

Let $N_{ch}(j, l)$ denote the number of *transmission pairs* whose destinations are located in *capture cell* j and column l. By *Lemma 7* (b), we have

$$\begin{aligned} & \mathbf{P}[N_{ch}(j, l) \geq 8n^{(3s+2\alpha-2)/3}\sqrt{2 \log n} \text{ for any } j, l] \\ \leq \quad & \mathbf{P}[N_{ch}(j, l) \geq 2\frac{n_s n_d}{\mathcal{C}_c\sqrt{\mathcal{C}_h}} \text{ for any } l] = O(\frac{1}{n_s n_d}) \end{aligned}$$

Hence, w.h.p., the number of vertical transmissions is $T_y = N_{ch}(j,l) \leq 8n^{(3s+2\alpha-2)/3}\sqrt{2\log n}$. And the total transmissions passing through a single *hopping cell* are $T_x + T_y \leq 32n^{(3s+2\alpha-2)/3}\log n$.

APPENDIX C – PROOF OF *Lemma 5*

We will need the following lemma on the minimum distance between the mobile relays and the destination at any time slot. Fix a packet p that enters into the system at time slot $t_0(p)$, and one of its destinations k. At each time slot $t \geq t_0(p)$, let $r_p(t)$ denote the number of mobile relays holding the packet p at the beginning of the time slot t. Among these $r_p(t)$ mobile relays, there is one mobile relay whose distance to the destination k of packet p is the smallest. Let $\widetilde{L}_{p,k}(t)$ denote this minimum distance, and let

$$L_{p,k}(t) = \max\{\frac{1}{n}, \widetilde{L}_{p,k}(t)\}$$

It is easy to verify that

$$\widetilde{l}_{p,k}(t) \geq \widetilde{L}_{p,k}(t) \geq L_{p,k}(t) - \frac{1}{n}$$

Where $\widetilde{l}_{p,k}(t)$ is the distance between a chosen relay node holding packet p and the destination k.

Lemma 8: Under the one-dimensional i.i.d. mobility model, if $n \geq 3$, then

$$\mathbb{E}\left[\frac{1}{L_{p,k}(t)r_p(t)}|\mathcal{F}_{t-1}{}^3\right] \leq 8\log n \text{ for all } t \leq t_0(p)$$

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

$$\mathbb{E}\left[\frac{1}{L_{p,k}(t)}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[n\mathbf{I}_{\{\widetilde{L}_{p,k}(t)\leq\frac{1}{n}\}}|\mathcal{F}_{t-1}\right] + \mathbb{E}\left[\frac{1}{\widetilde{L}_{p,k}(t)}\mathbf{I}_{\{\widetilde{L}_{p,k}(t)>\frac{1}{n}\}}|\mathcal{F}_{t-1}\right]$$

Since the nodes move on a unit square, $\widetilde{L}_{p,k}(t) \leq \sqrt{2}$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\widetilde{L}_{p,k}(t)}\mathbf{I}_{\{\widetilde{L}_{p,k}(t)>\frac{1}{n}\}}|\mathcal{F}_{t-1}\right] \\
&= \int_{\frac{1}{n}}^{\sqrt{2}} \frac{1}{u} d\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}] \\
&= \frac{1}{u}\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}]\Big|_{\frac{1}{n}}^{\sqrt{2}} - \int_{\frac{1}{n}}^{\sqrt{2}} \mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}]d\frac{1}{u} \\
&= \frac{1}{\sqrt{2}} - n\mathbf{P}[\widetilde{L}_{p,k}(t) \leq \frac{1}{n}|\mathcal{F}_{t-1}] + \int_{\frac{1}{n}}^{\sqrt{2}} \frac{1}{u^2}\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}]du
\end{aligned}$$

[3] $\mathcal{F}_t$ be the σ-algebra captures all information about the "history" up to time-slot t, including the nodes' positions and the packets they have.

Hence,

$$\mathbb{E}\left[\frac{1}{L_{p,k}(t)}|\mathcal{F}_{t-1}\right] = \frac{1}{\sqrt{2}} + \int_{\frac{1}{n}}^{\sqrt{2}} \frac{1}{u^2}\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}]du$$

Under the one-dimensional i.i.d. mobility model, we have,

$$\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}] \leq 1 - (1-2u)^{r_p(t)} \leq 2r_p(t)u$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{L_{p,k}(t)}|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{\sqrt{2}} + \int_{\frac{1}{n}}^{\sqrt{2}} \frac{1}{u^2}\mathbf{P}[\widetilde{L}_{p,k}(t) \leq u|\mathcal{F}_{t-1}]du \\
&\leq \frac{1}{\sqrt{2}} + \int_{\frac{1}{n}}^{\sqrt{2}} 2r_p(t)\frac{1}{u}du \\
&= \frac{1}{\sqrt{2}} + 2r_p(t)\log u\Big|_{\frac{1}{n}}^{\sqrt{2}} \\
&= \frac{1}{\sqrt{2}} + 2r_p(t)(\log\sqrt{2} + 2\log n) \\
&\leq 8r_p(t)\log n
\end{aligned}$$

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

$$\mathbb{E}\left[\frac{1}{L_{p,k}(t)r_p(t)}|\mathcal{F}_{t-1}\right] = \frac{1}{r_p(t)}\mathbb{E}\left[\frac{1}{L_{p,k}(t)}|\mathcal{F}_{t-1}\right] \leq 8\log n$$

■

Proof of *Lemma 5*: Let

$$V_t = 8\log n[t - t_0(p)] - \sum_{s=t_0(p)+1}^{t} \frac{1}{L_{p,k}(t)r_p(t)}\mathbf{I}_{\{C_{p,k}(t)=1\}}$$

Where $C_{p,k}(t) = 1$ denotes that the scheduler decides that a successful capture of packet p to destination k occurs at time-slot t. Then for all $t \geq t_0(p)$, V_t is also $\mathcal{F}_t$-measurable and $V_{t_0(p)} = 0$. By *Lemma 8*, we have

$$\begin{aligned}
&\mathbb{E}[V_t - V_{t-1}|\mathcal{F}_{t-1}] \\
&= 8\log n - \mathbb{E}\left[\frac{1}{L_{p,k}(t,k)r_p(t)}\mathbf{I}_{\{C_{p,k}(t)=1\}}\right] \\
&\geq 8\log n - \mathbb{E}\left[\frac{1}{L_{p,k}(t)r_{p,k}(t)}|\mathcal{F}_{t-1}\right] \\
&\geq 0
\end{aligned}$$

Hence,

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

i.e., V_t is a sub-martingale. We denote that $s_{p,k} = \min\{t : t \geq t_0(p) \text{ and } C_{p,k}(t) = 1\}$. Since $s_{p,k}$ is a stopping time, by appropriately invoking the Optional Stopping Theorem (Theorem 4.1 in [22]), we have,

$$\mathbb{E}[V_{s_{p,k}}] \geq 0$$

Hence,

$$8\log n\mathbb{E}[D_{p,k}] \geq \mathbb{E}\left[\frac{1}{L_{p,k}(s_{p,k})R_{p,k}}\right]$$

Using Hölder's Inequality

$$\mathbb{E}\left[\frac{1}{L_{p,k}(s_{p,k})}\right] \leq \mathbb{E}[R_{p,k}]\mathbb{E}\left[\frac{1}{L_{p,k}(s_{p,k})R_{p,k}}\right]$$

Then we have

$$8\log n\mathbb{E}[D_{p,k}] \geq \mathbb{E}\left[\frac{1}{L_{p,k}(s_{p,k})}\right]\frac{1}{\mathbb{E}[R_{p,k}]} \geq \frac{1}{\mathbb{E}[L_{p,k}(s_{p,k})]\mathbb{E}[R_{p,k}]}$$

Finally, by definition,

$$l_{p,k} = \widetilde{l}_{p,k}(s_{p,k}) \geq L_{p,k}(s_{p,k}) - \frac{1}{n}$$

Therefore,

$$8\log n\mathbb{E}[D_{p,k}] \geq \frac{1}{(\mathbb{E}[l_{p,k}] + \frac{1}{n})\mathbb{E}[R_{p,k}]}$$