Exact Throughput Capacity under Power Control in Mobile Ad Hoc Networks

---

Citation:

Jiajia Liu, Xiaohong jiang, Hiroki Nishiyama, and Nei Kato, "Exact Throughput Capacity under Power Control in Mobile Ad Hoc Networks," 31st IEEE International Conference on Computer Communications (INFOCOM 2012), Orlando, Florida, USA, Mar. 2012.

URL:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6195580

Exact Throughput Capacity under Power Control in Mobile Ad Hoc Networks

Jiajia Liu
Tohoku University
Sendai, Japan 980-8579
Email: liu-jia@it.ecei.tohoku.ac.jp

Xiaohong Jiang
Future University Hakodate
Hokkaido, Japan 041-8655
Email: jiang@fun.ac.jp

Hiroki Nishiyama and Nei Kato
Tohoku University
Sendai, Japan 980-8579
Email: {bigtree,kato}@it.ecei.tohoku.ac.jp

***Abstract*—The lack of a general capacity theory on mobile ad hoc networks (MANETs) is still a challenging roadblock stunting the application of such networks. The available works on this line mainly focus on deriving order sense results, which are helpful for us to explore the general scaling laws of throughput capacity but tell us little about the exact achievable throughput. This paper studies the exact per node throughput capacity of a MANET, where the transmission power of each node can be controlled to adapt to a specified transmission range υ and a generalized two-hop relay with limited packet redundancy f is adopted for packet routing. Based on the concept of automatic feedback control and the Markov chain model, we first develop a general theoretical framework to fully depict the complicated packet delivery process in the challenging MANET environment. With the help of the framework, we are then able to derive the exact per node throughput capacity for a fixed setting of both υ and f. Based on the new throughput result, we further explore the optimal throughput capacity for any f but a fixed υ and also determine the corresponding optimum setting of f to achieve it. This result helps us to understand how such optimal capacity varies with υ (and thus transmission power) and to find the maximum possible throughput capacity of such a network for any f and υ. Surprisingly, our results here indicate that usually such maximum throughput capacity can not be achieved through the local transmission, a fact different from what is generally believed in literature.**

I. Introduction

The mobile ad hoc network (MANET), a very flexible and self-autonomous wireless network architecture, is very promising to find many important applications in the daily information exchange, disaster relief, military troop communication, etc. By now, the lack of a general Shannon limit-like network capacity theory is still a challenging roadblock stunting the development and commercialization of MANETs [1]. It is expected such a theory can help us to understand the basic network throughput capacity limit and thus serves as an instruction guideline for the network design, performance optimization and engineering of future MANETs [2], [3].

Since the seminal work of Grossglauser and Tse (2001) [4], a lot of research efforts have been devoted to a better understanding of the MANET throughput capacity under various mobility models. Grossglauser and Tse [4] showed that under the i.i.d. mobility model, it is possible to achieve a $\Theta(1)$ per node throughput by employing a two-hop relay scheme. Following this line, it was later proved that the $\Theta(1)$ per node throughput can also be achieved under other mobility models, like the random walk model [5], the two-dimensional Brownian motions model [6] and the restricted mobility model [7]. Moraes *et al.* further showed that under uniform mobility model, we can still have the $\Theta(1)$ throughput even with a variant of the two-hop relay scheme, where each packet is only broadcasted once by its source and all nodes that receive the packet will act as its relays [8].

Recently, the trade-off between the throughput capacity and delay performances in MANETs has also been extensively explored. Perevalov *et al.* [9] studied the delay-limited throughput of MANETs and reported that under the i.i.d. mobility model, the achievable throughput is of order $\Theta(n^{-1/3})$ for a fixed delay value d and the throughput increases as $d^{2/3}$ when the delay d is a moderate value. Lin *et al.* [10] considered the Brownian motion model and showed that the two-hop relay scheme proposed by Grossglauser and Tse, while capable of achieving a per node throughput of $\Theta(1)$, incurs an expected packet delay of $\Omega(\log n/\sigma_n^2)$, where σ_n^2 is the variance parameter of the Brownian motion model. Neely *et al.* [11] proved that under the i.i.d. mobility model, it is able to achieve $O(1/\sqrt{n})$ throughput and $O(\sqrt{n})$ delay by introducing exact $\sqrt{n}$ redundancy for each packet. More recently, the per node throughput capacity and delay trade-off has also been studied under the random waypoint model [12] and the Brownian mobility model [13].

It is notable that the above works mainly focus on deriving the order sense results of MANET throughput capacity. Although the order sense results are helpful for us to understand the general scaling law and thus the growth rate of the throughput capacity with network size n, they tell us little about the exact achievable per node throughput capacity. In practice, however, such exact network throughput capacity is of great interest for network designers. Another limit of these works is that the impact of node transmission range on the throughput capacity of MANETs has been largely neglected. Since it is generally believed that the local transmission mode could result in the maximum per node throughput capacity, these work generally adopt the local transmission mode in their analysis, where either each node has a small transmission range of $\Theta(1/\sqrt{n})$ [4], [7]–[10], [12], [13], or it can only transmit to some other node(s) in the same cell [5], [6], [11].

In this paper, we study the exact per node throughput capacity of a MANET, where the transmission power (and thus

transmission range) of each node can be controlled such that the impact of the transmission range on per node throughput capacity can be explored. For packet routing, we consider a generalized two-hop relay with limited packet redundancy (i.e., with a specified limit on the maximum number of distinct relays for each packet), which covers the available two-hop relay schemes [11]–[13] as special cases. The main contributions of this paper are summarized as follows:

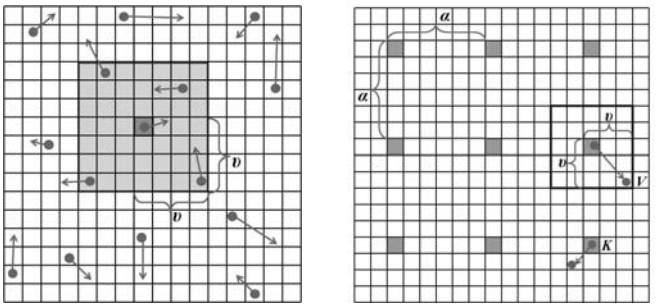
- By modeling the packet dispatching at its source and the packet receiving at its destination as Markov chains and applying the concept of automatic feedback control to characterize the service rate adaption between the source and destination of a flow, we first develop a general theoretical framework to depict the complicated packet delivery process in the challenging MANET environment.
- With the help of the theoretical framework, we then develop the exact per node throughput capacity $\mu(\upsilon, f)$ for any specified setting of transmission range υ and packet redundancy limit f. Simulation results are also provided to validate the throughput capacity result.
- Based on the new throughput result, we further explore the optimal capacity $\max_f\{\mu(\upsilon, f)\}$ for a fixed υ and also determine the corresponding optimum setting of f to achieve it. This result helps us to understand how such optimal capacity varies with υ and to find a suitable υ (and also f) to achieve the maximum possible throughput capacity $\max_{\upsilon,f}\{\mu(\upsilon, f)\}$ of such a network.

The rest of this paper is outlined as follows. Section II provides the system assumptions and definitions, and Section III discusses the issues of transmission scheduling and packet routing. In Section IV, we develop the theoretical model for achievable per node throughput capacity and present numerical results to validate it. We study in Section V the throughput maximization problem and explore the impact of transmission range on the maximum per node throughput capacity, and finally conclude the paper in Section VI.

II. System Assumptions and Definitions

Network Model: Similar to the previous works [6], [11], [14], in this paper we consider a two-dimensional torus network with unit area and n independent mobile nodes. We assume that the time is slotted and the network is evenly divided into $\sqrt{n}\times\sqrt{n}$ cells with $1/n$ area each, as illustrated in Fig. 1a. We consider the limit channel bandwidth scenario such that the total number of bits that can be transmitted per time slot is fixed and normalized to one packet. We further suppose that during each time slot each node has the knowledge about which cell it falls within based on its location information (For node localization, please refer to [15], [16]).

Mobility Model: This paper focuses on the i.i.d. node mobility model [17]–[19], where each node first independently and uniformly chooses a destination cell over all n cells at the beginning of each time slot, and then stays within it for the whole time slot. Under the i.i.d. mobility model, the time a node takes to move from one cell to another cell is neglected, so such model is able to capture the node behaviors in the regime of infinite mobility. The results in [11] indicate that the network capacity derived under the i.i.d. mobility model is actually identical to the one derived under other non-i.i.d. mobility models (like the Markovian random walk model and random waypoint model) if they follow the same steady state channel distribution.



(a) Cell partitioned network. (b) Illustration of transmission-group with $\alpha = 6$.

Fig. 1. Network cell partition and transmission-group.

Communication Model: To account for the interference among simultaneous transmissions, the Protocol Model introduced in [6], [20] is adopted here. For a link i at time slot t, we use $T_i(t)$ and $R_i(t)$ to denote the positions of the corresponding transmitter and receiver, respectively. Based on the Protocol Model, the transmission of the link i can be successful at the time slot t if for any other link j with simultaneous transmission we have

$$|T_j(t) - R_i(t)| \geq (1 + \Delta)|T_i(t) - R_i(t)|$$

here Δ is a protocol specified guardfactor for interference control. In order to explore the impact of power control on network throughput capacity, similar to [21] we assume that each node employs a power level so as to cover a set of cells with horizontal and vertical distances no more than $\upsilon - 1$ cells away from its current cell, where $1 \leq \upsilon \leq \lfloor\frac{\sqrt{n}+1}{2}\rfloor$, and $\lfloor x\rfloor$ is the floor function. With such power control, a node could transmit to any other node in a square area centered at the cell of the node and is of side length $(2\upsilon - 1)/\sqrt{n}$, as that illustrated in Fig. 1a.

Traffic Model: This paper considers the permutation traffic pattern widely adopted in previous studies [4], [11], [14], [17], [22]–[24]. Under such traffic model, there will be in total n distinct flows, where each node is the source of its locally generated traffic flow and at the same time the destination of another flow originated from some other node. For the traffic flow originated at each node, we assume it has an average rate of λ (packets/slot). The packet arrival process at each node is independent of its mobility process and all packets arrive at the beginning of each time slots. For the purpose of throughput capacity analysis, we simply assume that no lifetime is associated with each packet and the buffer size at each node is large enough such that the packet loss due to buffer overflow will never happen.

Throughput Capacity: We call a traffic input rate λ (packets/slot) feasible or achievable if there exists a spatial and temporal scheduling algorithm such that under this input rate the queue length at each node will never increase to infinity as the time goes to infinity. The per node throughput capacity is then defined as the maximum feasible input rate λ. Without incurring any ambiguity, hereafter we call such capacity as throughput capacity for brevity.

III. Transmission Scheduling and Routing

This section introduces the transmission scheduling and routing schemes to be adopted in this paper.

A. Transmission-Group Based Scheduling

According to the protocol interference model, multiple links could simultaneously transmit if they are sufficiently far away from each other. To support as many simultaneous link transmissions as possible while ensuring an acceptable interference among nodes, we consider here a transmission-group based scheduling scheme similar to [14], [22], [25], [26].

Transmission-group: A transmission-group is a subset of cells, where any two of them have a vertical and horizontal distance of some multiple of α cells and all of them could conduct transmission simultaneously.

An example of transmission-group is illustrated in Fig. 1b, where all the shaded cells are of the same transmission-group and each of them can simultaneously support a transmitting node in it without interfering with each other. It is notable that for the transmission-group based scheduling with parameter α, there will be in total α^2 distinct transmission-groups, where each cell belongs to one distinct transmission-group. If all transmission-groups alternatively become active (i.e., get the transmission opportunity), then each transmission-group (and thus each cell) becomes active in every α^2 time slots.

Setting of Parameter α: To support as many simultaneous transmissions as possible, we need to properly set the parameter α of transmission-group based on the parameter υ for power control and the parameter Δ for interference control. As illustrated in Fig. 1b, suppose that the node V is scheduled to receive from some transmitting node, while the node K in another active cell of the same transmission-group is transmitting to some other node. Notice that in this paper we consider a network scenario where each node employs a power level so as to cover a set of cells which have a horizontal and vertical distance of no greater than $\upsilon - 1$ cells away from its current cell, $1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$. Thus, we assume that the node V is at a distance of (x, y) ($x, y \in [-\upsilon+1, \upsilon-1]$) cells away from its transmitting node, where the x and y denote the horizontal distance and vertical distance, respectively. It is trivial to see that we only need to consider the cases that $x \in [0, \upsilon-1], y = \upsilon-1$. We can easily see that the distance from node V to its transmitting node is at most $\frac{1}{\sqrt{n}}\sqrt{\upsilon^2+(x+1)^2}$, while another simultaneous transmitting node (say the node K in Fig. 1b) is at least $\frac{1}{\sqrt{n}}\sqrt{(\alpha-\upsilon)^2+x^2}$ away from the node V. According to the interference model, the condition that K will not interfere with the reception at the V is that for any $x \in [0, \upsilon-1]$,

$$\frac{1}{\sqrt{n}}\sqrt{(\alpha-\upsilon)^2+x^2} \geq (1+\Delta)\frac{1}{\sqrt{n}}\sqrt{\upsilon^2+(x+1)^2} \quad (1)$$

To ensure above inequality for each $x \in [0, \upsilon-1]$, we have

$$\alpha \geq \upsilon + \sqrt{2(\Delta+1)^2\upsilon^2 - (\upsilon-1)^2} \quad (2)$$

Since α is an integer and $\alpha \leq \sqrt{n}$, we can set α as follow to support as many simultaneous transmissions as possible.

$$\alpha = \min\{\upsilon + \lceil\sqrt{2(\Delta+1)^2\upsilon^2 - (\upsilon-1)^2}\rceil, \lfloor\sqrt{n}\rfloor\} \quad (3)$$

Selection of Transmitting Node: The results in [14] indicate that at any time slot, the event that there are at least two nodes falling within an an active cell happens with a non-negligible probability (approaches $1-2e^{-1}$ as n goes to infinity). When there are more than one node in an active cell, the selection of transmitting node from the cell can be implemented by a mechanism similar to the DCF. At the beginning of each time slot, each node independently judges whether it is inside an active cell or not. If not, it remains silent and will not contend for the transmitting opportunity. Otherwise, it starts its back-off counter with a seed randomly selected from $[0, CW]$ (CW represents the contention window), and then overhears the channel until its back-off counter becomes 0 or it hears a broadcasting message from a transmitter. If no broadcasting message is heard during the back-off counting process, it broadcasts out a message denoting itself as the transmitter. Based on the back-off counting mechanism, each node of an active cell has the same probability to win the transmission opportunity and thus becomes the transmitting node.

B. 2HR-f Routing Scheme

In this paper, we consider a generalization of the classic two-hop routing scheme with f-cast (2HR-f) [11], [14], [27], $f \in [1, n-2]$, where each packet waiting at the source is delivered to at most f distinct relay nodes (i.e., each packet has a limited redundancy f) and should be received in order at its destination. For the permutation traffic pattern considered in this paper, there are in total n distinct traffic flows. Without loss of generality, we focus on a tagged flow in our discussion and use the S and D to denote the source node and the destination node, respectively. The source S labels each packet P of the tagged flow with a *sequence number* $SN(P)$, while the destination D maintains a *request number* $RN(D)$ to indicate the sequence number of the packet it is currently requesting. Notice that the sequence number mechanism ensures that every packet is received in order at the destination and also helps to remove the remnant copies of those packets already received [11]. The overall 2HR-f scheme is summarized as follows.

2HR-f Routing Scheme: When the source S wins the transmission opportunity at the current time slot, S first overhears the channel for a specified interval of time to check whether the node D is inside the one-hop transmission range.

1) If S hears the reply from D within the specified time interval, it initiates a handshake with D and then transmits a packet directly to D ("Source-to-Destination" transmission);
2) If no broadcasting reply is overheard during the specified time interval, a receiving node (say R) is randomly selected among the nodes within the one-hop transmission range of S based on a mechanism similar to the selection of transmitting node. With probability $1/2$, the S and R then perform either the "Source-to-Relay" or "Relay-to-Destination" transmission:
   - Source-to-Relay: Suppose that the packet P is the packet locally generated for which S is currently delivering copies, S first initiates a handshake with R to check whether R has already received a copy of P before. If not, S delivers out a new copy of P to R if only less than f copies of P have already been delivered out from S by now; otherwise, S remains idle for this time slot.
   - Relay-to-Destination: S initiates a handshake with R to check if S carries a packet P^* destined for node R with $SN(P^*) = RN(R)$. If so, S delivers the packet P^* to node R; otherwise, S remains idle for this time slot.

IV. Throughput Capacity

In this section, we first introduce some basic probabilities and use them to analyze the service time at the source node S and at the destination node D, and then provide the analysis and also related validation for throughput capacity.

A. Some Basic Probabilities

Lemma 1: For the tagged flow and a given time slot, we use p_1 and p_2 to denote the probability that S conducts a source-to-destination transmission and the probability that S conducts a source-to-relay or relay-to-destination transmission, respectively. By setting $m = (2\upsilon - 1)^2$, we have

$$p_1 = \frac{1}{\alpha^2}\left\{\frac{m-1}{n-1}\left(1 - \left(\frac{n-1}{n}\right)^{n-1}\right) + \frac{1}{n}\left(\frac{n-1}{n}\right)^{n-1}\right\} \quad (4)$$

$$p_2 = \frac{1}{\alpha^2}\left\{\frac{n-m}{n-1}\left(1 - \left(\frac{n-1}{n}\right)^{n-1}\right) - \left(\frac{n-m}{n}\right)^{n-1}\right\} \quad (5)$$

Lemma 2: For the tagged flow, suppose that the source S is delivering copies for some packet P in the current time slot, the destination D is also requesting for P, i.e., $SN(P) = RN(D)$, and there are already j $(1 \le j \le f+1)$ copies of P in the network (including the original one at S). For the next time slot, we use $p_r(j)$ to denote the probability that D will receive P, use $p_d(j)$ to denote the probability that S will successfully deliver out a copy of P to some new relay (if $j \le f$). Then we have

$$p_r(j) = p_1 + \frac{j-1}{2(n-2)}p_2 \quad (6)$$

$$p_d(j) = \frac{n-j-1}{2(n-2)}p_2 \quad (7)$$

The proofs of Lemmas 1 and 2 can be found in [28].

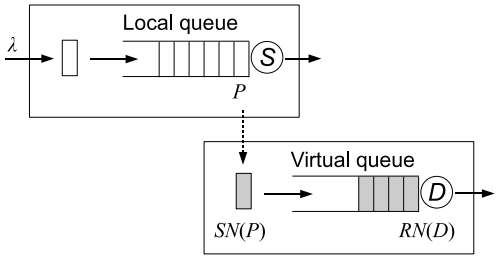


Fig. 2. Illustration of the local queue at the source S and the virtual queue at the destination D.

B. Service Time at Source S and Destination D

Before deriving the service time at S and D, we first define the following two queues. As shown in Fig. 2 that the first queue is a local queue at the source S, which stores the locally generated packets and operates as follows: every time a local packet arrives at S, it is put to the end of the queue; every time S finishes the copy dispatching for the head-of-line packet, S takes it out of the queue and moves ahead the remaining packets behind it. Thus, the head-of-line packet of the queue is the one for which S is currently distributing copies.

The second queue is a virtual queue defined at the destination D. As shown in Fig. 2, the virtual queue stores only the sequence numbers of those packets not received yet by D and operates as follows: every time a packet P is moved to the head-of-line of the local queue at S, the corresponding packet sequence number $SN(P)$ is put to the end of the virtual queue; every time D receives a packet whose sequence number equals to the head-of-line entry, D moves the head-of-line entry out of the virtual queue and moves ahead the remaining entries. Thus, the head-of-line entry of the virtual queue is the sequence number of the packet that D is currently requesting for, i.e., the $RN(D)$.

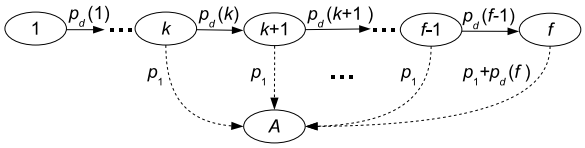
The service time at S and D can then be defined as follows:

Definition 1: For a packet P of the tagged flow, its service time at the source S is defined as the time elapsed between the time slot when S starts to deliver copies for P and the time slot when S stops distributing copies for P.

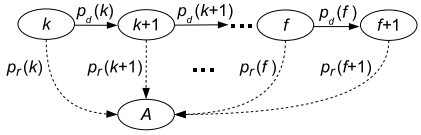
Definition 2: For a packet P of the tagged flow, the service time at the destination D is defined as the time elapsed between the time slot when D starts to request for P and the time slot when D receives P.

For a packet P of the tagged flow, suppose that there are k copies of P in the network when its destination D starts to request for P, $1 \le k \le f+1$. If we denote by state A the absorbing state (i.e., the termination of the service process) for P, then the service processes for the packet P at its source S and at its destination D can be defined by two finite-state absorbing Markov chains shown in Fig. 3a and Fig. 3b, respectively.

Suppose that there are k copies of P in the network when D starts to request for the packet, we denote by $X_S(k)$ and $X_D(k)$ the corresponding service time of the packet at S and



(a) Absorbing Markov chain for the packet distribution process at the source node S.



(b) Absorbing Markov chain for the packet reception process at the destination node D.

Fig. 3. Absorbing Markov chains for a packet P of the tagged flow, given that D starts to request for P when there are already k copies of P in the network. For each transient state, the transition back to itself is not shown for simplicity.

D, respectively [1]. From the theory of Markov chain [29] we know that the $X_S(k)$ is the time the Markov chain in Fig. 3a takes to become absorbed given that the chain starts from the state 1, and the $X_D(k)$ is the time the Markov chain in Fig. 3b takes to become absorbed given that the chain starts from the state k.

Lemma 3: For a packet P of the tagged flow, suppose that there are k copies of P in the network when the destination D starts to request for P, $1 \le k \le f+1$, then we have

$$\mathbb{E}\{X_S(k)\} = \begin{cases} \sum_{i=1}^{k-1} \frac{1}{p_d(i)} + \frac{1}{p_1+p_d(k)} \\ \quad \cdot\big(1+\sum_{j=1}^{f-k}\phi_1(k,j)\big) & \text{if } 1 \le k \le f, \\ \sum_{i=1}^{f} \frac{1}{p_d(i)} & \text{if } k = f+1. \end{cases} \tag{8}$$

$$\mathbb{E}\{X_D(k)\} = \begin{cases} \frac{1}{p_1+p_2/2}\big(1+\sum_{j=1}^{f-k}\phi_2(k,j) \\ +\frac{p_d(f)}{p_r(f+1)}\phi_2(k,f-k)\big) & \text{if } 1 \le k \le f-1, \\ \frac{1}{p_1+p_2/2}\big(1+\frac{p_d(f)}{p_r(f+1)}\big) & \text{if } k = f, \\ \frac{1}{p_r(f+1)} & \text{if } k = f+1. \end{cases} \tag{9}$$

where

$$\phi_1(k,j) = \prod_{t=1}^{j} \frac{p_d(k+t-1)}{p_1+p_d(k+t)}$$

$$\phi_2(k,j) = \prod_{t=1}^{j} \frac{p_d(k+t-1)}{p_1+p_2/2}$$

Lemma 4: For any $1 \le k \le f$, we have

$$\mathbb{E}\{X_S(k)\} < \mathbb{E}\{X_S(k+1)\} \tag{10}$$

$$\mathbb{E}\{X_D(k)\} > \mathbb{E}\{X_D(k+1)\} \tag{11}$$

For the tagged flow, if we further denote by $\overline{X}_S$ the average service time of all packets locally generated at the source S, and denote by $\overline{X}_D$ the average service time of all packets

[1]The $X_S(f+1)$ corresponds to the case that D starts to request for the packet P from the state that there are $f+1$ copies in the network, i.e., the f copies of P have already been distributed.

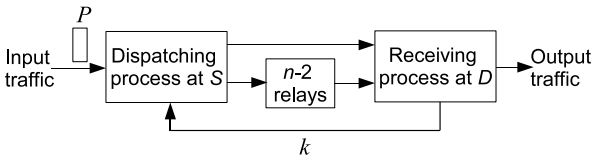


Fig. 4. Illustration of the automatic feedback control system defined for the packet delivery process of the tagged flow, where the parameter k is automatically updated to adjust to the service rates at the S and D.

received at the destination D, then we can establish the following result based on Lemma 4.

Lemma 5: For any given transmission range parameter υ and packet redundancy limit f, $1 \le \upsilon \le \lfloor \frac{\sqrt{n}+1}{2} \rfloor$, $1 \le f \le n-2$, we have

$$\mathbb{E}\{X_S(1)\} \le \overline{X}_S \le \mathbb{E}\{X_S(f+1)\} \tag{12}$$

$$\mathbb{E}\{X_D(f+1)\} \le \overline{X}_D \le \mathbb{E}\{X_D(1)\} \tag{13}$$

For the proofs of the above lemmas, please refer to [28].

C. Per Node Throughput Capacity

For the tagged flow, suppose that currently packet P is the head-of-line packet at the local queue of S, D just starts to request for P and there are k ($1 \le k \le f$) copies of P now in the network. We further assume that the packet waiting right behind P in the local queue is packet P', and D will start to request for P' when there are k' copies of P' in the network. Then we have the following two cases:

- If $\mathbb{E}\{X_S(k)\} \le \mathbb{E}\{X_D(k)\}$, then in the average case we have $k' \ge k$. According to (10) and (11), it follows that $\mathbb{E}\{X_S(k')\} \ge \mathbb{E}\{X_S(k)\}$ and $\mathbb{E}\{X_D(k')\} \le \mathbb{E}\{X_D(k)\}$. Then we have
$$\mathbb{E}\{X_D(k') - X_S(k')\} \le \mathbb{E}\{X_D(k) - X_S(k)\} \tag{14}$$
The above condition indicates that statistically the expected gap between the service time at the destination and the service time at the source tends to reduce. Since (14) also holds for the packets (if any) waiting behind P', we can see that the $\overline{X}_S$ and $\overline{X}_D$ (and thus the average service rates at S and D) will gradually approach each other until a balance is achieved [2].
- If $\mathbb{E}\{X_S(k)\} > \mathbb{E}\{X_D(k)\}$, on average we then have $k' < k$. Similar to the above case, it follows that
$$\mathbb{E}\{X_S(k') - X_D(k')\} < \mathbb{E}\{X_S(k) - X_D(k)\} \tag{15}$$
This condition indicates that statistically the service time at S tends to decrease while the service time at D tends to increase, thus the network system will gradually evolve towards a stable state.

[2]As we will show later that since the server for the local queue at S and the server for the virtual queue at D may have vacancy time, even in the case that the two average service rates $1/\overline{X}_S$ and $1/\overline{X}_D$ can not achieve a balance, the network system may still be stabilized as long as the input rate λ is feasible.

The above analysis indicates that under the 2HR-f routing scheme, the parameter k is automatically updated from packet to packet to adjust to the service rates at the S and D. Based on this intrinsic feature of automatic updating for parameter k, we can model the packet delivery process of the tagged flow as an automatic feedback control system shown in Fig. 4, where the packet dispatching process at S and the packet receiving process at D can be defined by the two absorbing Markov chains in Fig. 3a and Fig. 3b, respectively[3].

Now we are ready to derive the throughput capacity for the tagged flow. We first denote by V_S the long-term average packet dispatching rate at S and denote by V_D the long-term average packet receiving rate at D, where

$$V_S = \lim_{t\to\infty} \frac{\text{the number of dispatched packets at } S}{t} \quad (16)$$

$$V_D = \lim_{t\to\infty} \frac{\text{the number of received packets at } D}{t} \quad (17)$$

We then have the following result.

Lemma 6: For the tagged flow and any given parameters of υ and f, we have

$$V_S \le \frac{1}{\mathbb{E}\{X_S(1)\}} \quad (18)$$

$$V_D \le \frac{1}{\mathbb{E}\{X_D(f+1)\}} \quad (19)$$

Proof: We first prove (18). For the local queue at the source S, suppose that during some time interval t node S has successfully served a total of $N_S(t)$ locally generated packets (i.e., S has distributed copies for $N_S(t)$ local packets). According to the definition in (16), we have

$$V_S = \lim_{t\to\infty} \frac{N_S(t)}{t} \quad (20)$$

Notice that during the time interval t, the local queue may be empty and thus the queue server at S may become idle. Denote by $I_S(t)$ the accumulated vacancy time at S during the time interval t, then we have

$$\overline{X}_S = \lim_{t\to\infty} \frac{t - I_S(t)}{N_S(t)} \quad (21)$$

Since $I_S(t) \ge 0$, combining the (20) and (21), we have

$$V_S \le \frac{1}{\overline{X}_S} \quad (22)$$

Substituting (12) into (22), (18) then follows. After a similar derivation, (19) also follows. ■

We now can establish the following main result on per node throughput capacity.

[3]Notice that for a packet P of the tagged flow, its corresponding parameter k depends only on the delivery process of the last packet received just before the P. Thus, for the automatic feedback control system in Fig. 4, the parameter k will be automatically updated for each packet and there is no need for any extra transmissions to deliver the parameter k between S and D.

Theorem 1: Consider a cell partitioned MANET, where nodes move according to the i.i.d. mobility model, each node could transmit to the cells with horizontal and vertical distance no more than $\upsilon - 1$ cells away from its current cell, $1 \le \upsilon \le \lfloor \frac{\sqrt{n}+1}{2} \rfloor$, and each packet follows the 2HR-f routing scheme, $1 \le f \le n-2$. If we denote by $\mu(\upsilon, f)$ the per node (flow) throughput capacity, i.e., the network can stably support any input rate $\lambda \le \mu(\upsilon, f)$, then we have

$$\mu(\upsilon, f) = \min\left\{ p_1 + \frac{f}{2(n-2)} p_2, \frac{p_1 + p_2/2}{1 + \sum_{j=1}^{f-1} \prod_{t=1}^{j} \frac{(n-t-1)p_2}{2(n-2)p_1 + (n-t-2)p_2}} \right\} \quad (23)$$

Proof: For the tagged flow, as its packet delivery process can be defined by the automatic feedback control system in Fig. 4, we can see that if the network is stable (i.e., the queue length at each node will not go to infinity) under the input rate λ, then we have

$$\lambda = V_S = V_D \quad (24)$$

This is due to the fact that in a stable control system, the long-term average rate of the input traffic is equal to that of the output one.

Based on (18) and (19), we then have

$$\lambda \le \min\left\{ \frac{1}{\mathbb{E}\{X_S(1)\}}, \frac{1}{\mathbb{E}\{X_D(f+1)\}} \right\} \quad (25)$$

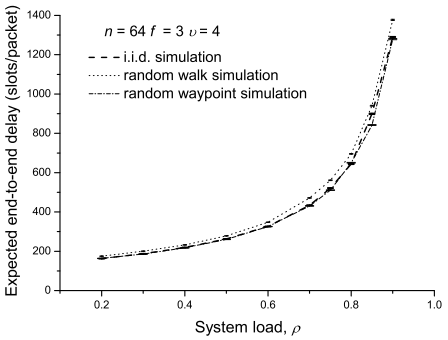
After substituting (8) and (9) into (25), (23) follows. ■

D. Model Validation

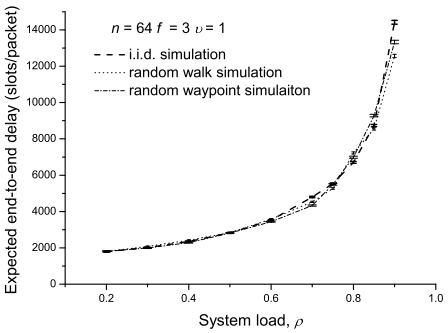
To validate our theoretical per node throughput capacity, a dedicated simulator in C++ was developed, which is now available at [30]. In our simulation, the traffic flow originated from each node is assumed to be a Poisson stream, and, similar to the settings in [31], the guard factor Δ is fixed as $\Delta = 1$ here. In addition to the i.i.d. mobility model considered in this paper, we also implemented the simulations for the popular random walk and random waypoint models:

- **Random Walk Model [5]:** At the beginning of each time slot, each node independently makes a decision regarding its mobility action, either staying inside its current cell or moving to one of its eight adjacent cells. Each action happens with the same probability of $1/9$.
- **Random Waypoint Model [18]:** At the beginning of each time slot, each node independently and randomly generates a two-dimensional vector $\boldsymbol{\nu} = [\nu_x, \nu_y]$, where the values of ν_x and ν_y are uniformly drawn from $[1/\sqrt{n}, 3/\sqrt{n}]$. The node then moves a distance of ν_x along the horizontal direction and a distance of ν_y along the vertical direction.

In our simulation, we consider both a smaller network with $n = 64$, $f = 3$ and $\upsilon = \{1, 4\}$ and a larger network with $n = 256$, $f = 6$ and $\upsilon = \{1, 6\}$. The corresponding simulation results are summarized in Fig. 5 and Fig. 6, where all the

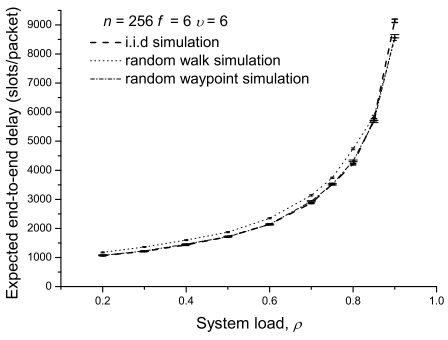


(a) Network scenario ($n = 64$, $f = 3$, $\upsilon = 4$) with $\mu(4,3) = 7.60 \times 10^{-3}$ (packets/slot).

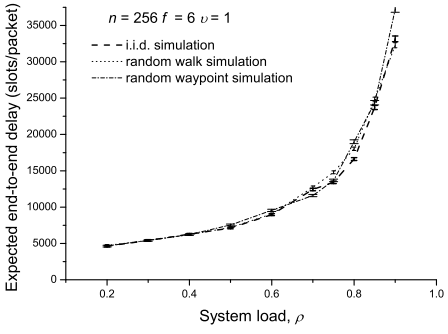


(b) Network scenario ($n = 64$, $f = 3$, $\upsilon = 1$) with $\mu(1,3) = 7.53 \times 10^{-4}$ (packets/slot).

Fig. 5. The expected end-to-end packet delay for $n = 64$.



(a) Network scenario ($n = 256$, $f = 6$, $\upsilon = 6$) with $\mu(6,6) = 1.17 \times 10^{-3}$ (packets/slot).



(b) Network scenario ($n = 256$, $f = 6$, $\upsilon = 1$) with $\mu(1,6) = 2.84 \times 10^{-4}$ (packets/slot).

Fig. 6. The expected end-to-end packet delay for $n = 256$.

results are reported with the 95% confidence interval. Figs. 5 and 6 indicate clearly that our theoretical models can precisely depict the throughput capacity of a MANET under the packet redundancy control and the transmission range control. As can be observed from these two figures that when the system load $\rho = \lambda/\mu$ approaches to 1 (i.e., when the traffic input rate λ approaches the throughput capacity μ), the packet delay rises up sharply and becomes extremely sensitive to the variations of ρ. Such skyrocketing behavior of packet delay as ρ approaches 1 serves as an intuitive validation for the throughput capacity determined by our theoretical model.

It is also interesting to notice from Figs. 5 and 6 that the networks we consider here actually exhibit very similar behaviors under either the i.i.d. model, the random walk model or the random waypoint model. In this sense, our theoretical throughput capacity model, although was developed under the i.i.d. mobility model, can also be used to nicely capture the network behaviors under the random walk and the random waypoint models as well.

V. Throughput Optimization

Based on the new theoretical capacity model, we further explore the following throughput optimization problem.

Throughput Optimization Problem: For a 2HR-f-based MANET with a fixed transmission range υ for each node, calculate its maximum per node throughput capacity for any value of f.

For a fixed transmission range υ, if we denote by μ^* the corresponding maximum per node throughput capacity, then the throughput optimization problem can be formulated as

$$\begin{aligned}\mu^* &= \max_f\{\mu(\upsilon, f)\} \\ &= \max \ \min\left\{\frac{1}{\mathbb{E}\{X_S(1)\}}, \frac{1}{\mathbb{E}\{X_D(f+1)\}}\right\} \quad (26)\end{aligned}$$

subject to:

$$1 \le f \le n-2, 1 \le \upsilon \le \lfloor \frac{\sqrt{n}+1}{2} \rfloor$$

where the $\mathbb{E}\{X_S(1)\}$ and $\mathbb{E}\{X_D(f+1)\}$ are defined in (8) and (9), respectively.

Regarding the solution of this optimization problem, we have the following result.

Lemma 7: For any given $\upsilon \in [1, \lfloor \frac{\sqrt{n}+1}{2} \rfloor]$, we have

$$\mu^* = max\left\{\frac{1}{\mathbb{E}\{X_D(f+1)\}\big|_{f=f_0}}, \frac{1}{\mathbb{E}\{X_S(1)\}\big|_{f=f_1}}\right\} \quad (27)$$

where

$$f_0 = \max\Big\{f \mid \mathbb{E}\{X_S(1)\} \le \mathbb{E}\{X_D(f+1)\}\Big\} \quad (28)$$

$$f_1 = \min\Big\{f \mid \mathbb{E}\{X_D(f+1)\} \le \mathbb{E}\{X_S(1)\}\Big\} \quad (29)$$

Proof: We first prove that f_0 and f_1 defined above do exist. According to (8) and (9), we have

$$\mathbb{E}\{X_S(1)\}|_{f=1} = \frac{1}{p_1 + p_d(1)} \le \frac{1}{p_1 + \frac{p_2}{2(n-2)}} = \mathbb{E}\{X_D(f+1)\}|_{f=1} \quad (30)$$

Notice that

$$\begin{aligned}&\mathbb{E}\{X_S(1)\}|_{f=n-2}\\ &= \frac{1}{p_1 + p_2/2}\Big(1 + \sum_{j=1}^{n-3}\prod_{t=1}^{j}\frac{(n-t-1)p_2}{2(n-2)p_1 + (n-t-2)p_2}\Big)\\ &\ge \frac{1}{p_1 + p_2/2} = \mathbb{E}\{X_D(f+1)\}|_{f=n-2}\end{aligned} \quad (31)$$

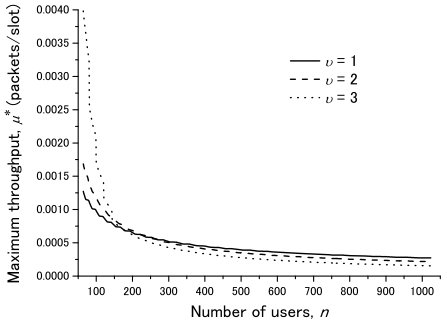
It is easy to see from (8) and (9) that as f increases, the $\mathbb{E}\{X_S(1)\}$ monotonically increases while the $\mathbb{E}\{X_D(f+1)\}$ monotonically decreases. Combining with the results in (30) and (31), we can see that the f_0 and f_1 defined above do exist.

From (25) we know that

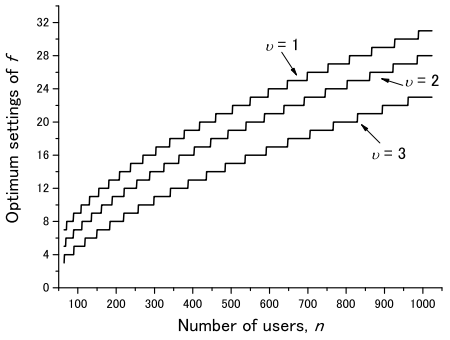
$$\mu(\upsilon, f) = \begin{cases} 1/\mathbb{E}\{X_D(f+1)\} & \text{if } 1 \le f \le f_0, \\ 1/\mathbb{E}\{X_S(1)\} & \text{if } f_1 \le f \le n-2. \end{cases} \quad (32)$$

Thus, (27) can be derived directly based on (32). The above results indicate that for a MANET with a fixed υ, there exists an optimum setting of f (f_0 or f_1) to achieve the optimal per node throughput capacity μ^*. ■

To illustrate the optimal throughput capacity μ^*, we show in Fig. 7a and Fig. 7b that for $\upsilon = \{1, 2, 3\}$, how μ^* and the corresponding optimum setting of f vary with network size n. Fig. 7a shows clearly that for all the three settings of υ here, although the corresponding μ^* all decrease quickly as the network size increases, their varying tendencies with n are actually different. A careful observation of Fig. 7a indicates that among the three settings of υ here, the μ^* of the case $\upsilon = 3$ decreases most dramatically with n while the one of the case $\upsilon = 1$ decreases least significantly with n. It is also interesting to notice that when $n \le 143$, the μ^* of the case $\upsilon = 3$ is always the greatest one among that of all three cases, while the μ^* of the case $\upsilon = 1$ becomes the greatest one when n is larger than 270. The results in Fig. 7b show that for a given υ, the corresponding optimum setting of f is actually a piecewise function of n. We can also see from the figure that for each network size n, the optimum setting of f of the case $\upsilon = 3$ is the smallest one among that of all three cases. This can be intuitively interpreted as follows. For one given network, if a bigger υ (and thus a larger transmission range) is adopted, a node will have higher probability to meet its destination or relay nodes and thus can deliver packets more fast, resulting in a fewer number of redundant copies for each packet before it arrives at its destination.



(a) The maximum throughput capacity μ^* vs. n



(b) The optimum setting of f vs. n

Fig. 7. The maximum throughput capacity μ^* and the corresponding optimum setting of f for networks with n varying from 64 to 1024.

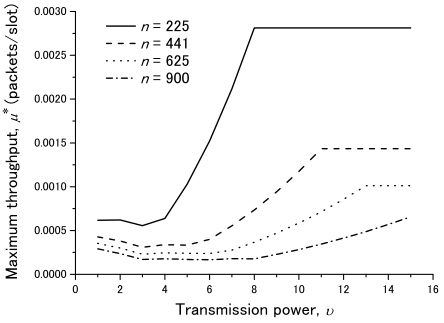


Fig. 8. The throughput capacity μ^* vs. node transmission region υ.

To further explore how υ affects μ^*, we summarize in Fig. 8 how μ^* varies with υ for networks of $n = \{225, 441, 625, 900\}$. It is interesting to see that, for each network scenario here, as υ increases μ^* always first decreases and then increases. This is because that the effect of increasing υ is two-fold: on one hand, it increases the probability that a node meets the destination or relay nodes, but on the other hand it decreases the number of simultaneous transmissions. As illustrated in Fig. 8 that when υ is small, the latter negative

effect dominates, while as υ gradually increases, the former positive effect becomes the dominant one. It is further noticed that for the cases $n = 225$, $n = 441$ and $n = 625$, μ^* does not increase any more when υ increases beyond some threshold ($\upsilon = 8, 11$ and 13, respectively), where a node is able to cover the whole network region. It is notable that the results in both Fig. 8 and Fig. 7a actually imply a fact that for the network scenarios considered in this paper, we may have a very significant per node throughput capacity improvement and may achieve its maximum possible value through adopting a bigger υ (and thus a larger transmission range) for each node, which is different from what is generally believed in literature that a smaller υ usually results in a higher throughput capacity.

VI. Conclusion

Distinguished from the available works which mainly focus on deriving the order sense results and exploring the scaling laws of the throughput capacity in MANETs, this paper addressed another basic problem: for a MANET with general node transmission range control and packet redundancy control, what is the exact achievable per node throughput capacity. We found that for the network scenarios considered in this paper, it may not be always true that adopting local transmission can achieve the maximum per node throughput capacity, as what is generally believed in literature. This finding indicates that further deliberate studies are necessary to reveal the real achievable network throughput of MANETs. Another interesting finding of our work is that the MANETs considered in this paper actually exhibit very similar behaviors in terms of packet delay and per node throughput under different node mobility models, like the i.i.d., random walk and random waypoint.

Acknowledgment

Part of this work was supported through the A3 Foresight Program by the Japan Society for the Promotion of Science (JSPS), the National Natural Science Foundation of China (NSFC), and the National Research Foundation of Korea (NRF).

References

[1] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, December 2008.

[2] A. Goldsmith, M. Effros, R. Koetter, M. Medard, A. Ozdaglar, and L. Zheng, "Beyond shannon: The quest for fundamental performance limits of wireless ad hoc networks," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 195–205, May 2011.

[3] L. X. Cai, X. S. Shen, J. W. Mark, and L. Cai, "Capacity analysis and mac enhancement for uwb broadband wireless access networks," *Elsevier Computer Networks*, vol. 51, no. 11, pp. 3265–3277, August 2007.

[4] M. Grossglauser and D. N. Tse, "Mobility increases the capacity of ad hoc wireless networks," in *INFOCOM*, 2001.

[5] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks-part i: The fluid model," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2568–2592, June 2006.

[6] ——, "Throughput-delay trade-off in wireless networks," in *INFOCOM*, 2004.

[7] J. Mammen and D. Shah, "Throughput and delay in random wireless networks with restricted mobility," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1108–1116, 2007.

[8] R. M. de Moraes, H. R. Sadjadpour, and J. Garcia-Luna-Aceves, "Throughput-delay analysis of mobile ad-hoc networks with a multi-copy relaying strategy," in *SECON*, 2004.

[9] E. Perevalov and R. S. Blum, "Delay-limited throughput of ad hoc networks," *IEEE Transactions on Communications*, vol. 52, no. 11, pp. 1957–1968, November 2004.

[10] X. Lin, G. Sharma, R. R. Mazumdar, and N. B. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with brownian mobility," *IEEE/ACM Transactions on Networking, Special Issue on Networking and Information Theory*, vol. 52, no. 6, pp. 2777–2784, June 2006.

[11] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917–1936, June 2005.

[12] G. Sharma and R. Mazumdar, "Delay and capacity trade-off in wireless ad hoc networks with random way-point mobility," in *Dept. Elect. Comput. Eng., Purdue Univ.,West Lafayette, IN*, 2005. [Online]. Available: http://ece.purdue.edu/∼gsharma/

[13] ——, "On achievable delay/capacity trade-offs in mobile ad hoc networks," in *Wiopt*, 2004.

[14] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Delay and capacity in ad hoc mobile networks with f-cast relay algorithms," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2738–2751, August 2011.

[15] M. Li and Y. Liu, "Rendered path: Range-free localization in anisotropic sensor networks with holes," in *MobiCom*, 2007.

[16] Z. Zhu, A. M.-C. So, and Y. Ye, "Universal rigidity: Towards accurate and efficient localization of wireless networks," in *INFOCOM*, 2010.

[17] L. Ying, S. Yang, and R. Srikant, "Optimal delay-throughput trade-offs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4119–4143, September 2008.

[18] S. Zhou and L. Ying, "On delay constrained multicast capacity of large-scale mobile ad-hoc networks," in *INFOCOM*, 2010.

[19] Y. Wang, X. Chu, X. Wang, and Y. Cheng, "Optimal multicast capacity and delay tradeoffs in manets: A global perspective," in *INFOCOM*, 2011.

[20] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[21] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming, W. M. McEneaney, G. Yin, and Q. Zhang, pp. 547–566, 1998.

[22] P. Li, Y. Fang, and J. Li, "Throughput, delay, and mobility in wireless ad-hoc networks," in *INFOCOM*, 2010.

[23] M. Garetto, P. Giaccone, and E. Leonardi, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: The subcritical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1888–1901, December 2009.

[24] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad-hoc networks," in *INFOCOM*, 2010.

[25] S. R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.

[26] C. Zhang, Y. Fang, and X. Zhu, "Throughput-delay tradeoffs in large-scale manets with network coding," in *INFOCOM*, 2010.

[27] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77–90, February 2008.

[28] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Impact of power control on throughput and delay in mobile ad hoc networks," Graduate school of information sciences, Tohoku university, 2011, technical report 201107.01. [Online]. Available: http://distplat.blogspot.com

[29] C. M. Grinstead and J. L. Snell, *Introduction to Probability: Second Revised Edition*. American Mathematical Society, 1997.

[30] C++ simulator for the 2hr-f manet with power control. [Online]. Available: http://distplat.blogspot.com

[31] The network simulator ns-2. [Online]. Available: http://www.isi.edu/nsnam/ns/