Delay and Capacity in Ad Hoc Mobile Networks with f-cast Relay Algorithms

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 2-hop relay algorithm and its variants have been attractive for ad hoc mobile networks, because they are simple yet efficient, and more importantly, they enable the capacity and delay to be studied analytically. This paper considers the 2-hop relay with f-cast (2HR-f) under i.i.d. mobility model, a general 2-hop relay algorithm that allows one packet to be delivered to at most f distinct relay nodes. The 2HR-f algorithm covers the available 2-hop relay algorithms ($f = 1, \sqrt{n}$) as special cases. Closed-form analytical models rather than order sense ones are developed for the 2HR-f algorithm with a careful consideration of important medium contention and queuing delay issues, which enable an accurate delay and capacity analysis to be performed for ad hoc mobile networks employing 2HR-f. Based on our models and some typical settings of f (say, $f = 1, \sqrt{n}$), one can easily derive the corresponding order sense results.**

I. Introduction

Since the seminal work of Grossglauser and Tse (2001) [1], the 2-hop relay algorithm and its variants have become a class of attractive routing algorithms for ad hoc mobile networks, because they are simple yet efficient, and more importantly, they enable the capacity and delay to be studied analytically. The 2-hop relay algorithm defines two phases for packet transmission, where in phase 1 a packet is transmitted from its source node to an intermediate node (relay node), and then in phase 2 the packet is transmitted from the relay node to its destination node. Since the source node can directly transmit a packet to its destination node every time such transmission opportunity arises, every packet goes through at most 2 hops to reach its destination in a 2-hop relay network.

By now, extensive order sense results of delay and capacity have been reported to illustrate the scaling laws of 2-hop relay ad hoc mobile networks under various mobility models. Grossglauser and Tse (2001) [1] showed that it is possible to achieve a $\Theta(1)$ throughput per node under i.i.d. mobility model. Later, Gamal *et al.* [2] showed that the $\Theta(1)$ throughput is also achievable under a random walk model, but which comes at the price of a $\Theta(n \log n)$ delay. Mammen *et al.* [3] proved that the same throughput and delay scaling are also achievable even with a variant of the Grossglauser-Tse 2-hop relay and a restricted mobility model. The delay and throughput trade-off has been further widely studied under different mobility models, like the i.i.d. mobility model [4], hybrid random walk and discrete random direction models [5], Brownian motion model [6], [7], and correlated mobility model [8]. These order sense results are helpful for us to understand the general scaling laws of delay and capacity in a 2-hop relay ad hoc mobile network, but they tell us a little about the real end-to-end delay and capacity of such networks. In practice, however, the real delay and capacity results are of great interest for network designers.

It is notable that the relay algorithms discussed above can be regarded as the basic 2-hop relay schemes without packet redundancy (i.e., allowing redundant copies of each packet). As throughput and delay can be traded with each other in a multi-hop way, this trade-off can also be achieved to somewhat extent via delivering redundant copies for each packet. Neely and Modiano [9] considered a modified version of the Grossglauser-Tse 2-hop algorithm for ad hoc mobile networks, and proved that under i.i.d. mobility model it achieve $O(1/\sqrt{n})$ throughput and $O(\sqrt{n})$ delay with exact $\sqrt{n}$ redundancy for each packet. Sharma and Mazumdar [10] explored the order sense delay and capacity trade-off in ad hoc mobile networks under random way-point mobility model and multiple redundancy for each packet. The idea of using packet redundancy has also been adopted to reduce average packet deliver delay [11]–[13] in intermittently connected mobile networks (ICMNs).

In this paper we consider the 2-hop relay with f-cast (2HR-f) under i.i.d. mobility model, a general 2-hop relay algorithm that allows one packet to be delivered to at most f distinct relay nodes, and develop closed form analytical models rather than order sense ones for an accurate delay and capacity analysis for 2HR-f-based ad hoc mobile networks. With such closed form models and some typical settings of f (say, $f = 1, \sqrt{n}$), one can easily derive the corresponding order sense results for delay and capacity.

The rest of the paper is organized as follows. In Section II, we provide the network model, interference model and mobility model considered in our analysis. Section III introduces the the 2-hop relay algorithm f-*cast* (2HR-f) and the corresponding scheduling scheme. We develop the closed-form models in Section IV to analyze the expected end-to-end delay and capacity, and finally we conclude the paper in Section V.

II. System Models

A. Network Model

We assume the network is a square region of unit area, with n mobile nodes which are initially independently and uniformly distributed inside the network region. The unit square is evenly divided into $\sqrt{n} \times \sqrt{n}$ cells each of which has an area of $1/n$. All the n mobile nodes are independently roaming from cell to cell, and time is assumed to be slotted so that each node remains in its current cell for one time slot.

We consider a bi-dimensional i.i.d. mobility model, or so-called reshuffling model in this paper. At the beginning of each time slot, every node independently selects a destination cell and stays in it for the whole time slot, thus the position of each node is updated every time slot. The destination cell is chosen uniformly and randomly among all n cells, so each cell would be chosen with same probability $1/n$.

We assume a time-scale of fast mobility [8] in this paper, i.e., the mobility of nodes is at the same time-scale as the data transmission. Thus, only one-hop transmissions are feasible during any single slot, and the total number of bits that can be transmitted in a time slot is a fixed constant independent of n. We normalize this constant to 1 here.

B. Interference Model

Similar to [1], we assume a uniform communication range $r = \Theta(1/\sqrt{n})$ for all nodes, and adopt the protocol model introduced in [14] to account for interference among simultaneous transmissions. Suppose node i is transmitting to node j at some time slot t, and their Euclidean distance is $d_{ij}(t)$. According to the protocol model, this transmission can be successful if and only if the following two conditions hold:

(1) $d_{ij}(t) \leq r$;
(2) $d_{kj}(t) \geq (1+\Delta)r$ for every other node k which transmits simultaneously, where Δ is a protocol specified guard-factor to prevent interference.

C. Traffic Model

Similar to previous works we consider permutation traffic patterns, in which each node is a source and at the same time a destination of some other node. Hence there are n source-destination pairs in the network. For convenience of expression, we use $S(K)$ and $D(K)$ to denote the source node and destination node of node K, respectively. We further assume a homogeneous scenario in which the traffic originating at each node is a Poisson stream with rate λ (packets/slot). We also assume that the packet arrives at the beginning of time slots, and the arrival process at each node is independent of its mobility process.

III. 2HR-f Algorithm and Scheduling Scheme

A. 2HR-f Algorithm

We consider a generalization of the 2-hop relay algorithm [9] with *f-cast* (2HR-f), which allows up to f ($1 \leq f \leq \sqrt{n}$) relay nodes for a single packet. As we keep one copy at the source node, so there are at most $f+1$ copies of a single packet to coexist in the network.

As a node can be a potential relay for any of other $n-2$ flows (except the two flows originating at and destined for itself), we assume that every node maintains n individual queues at its buffer, one local-queue for storing its locally generated packets waiting for copy-distribution, one already-sent-queue for storing packets whose f replicas have already been distributed but reception status are not confirmed yet (from destination node), and $n-2$ parallel relay-queues for storing packets of other flows (one queue per flow). We further assume all packets of one flow are labeled with sequence numbers, so that a packet can be efficiently retrieved from the queue buffers of its source node or relay node(s) according to its sequence number and destination information. During each time slot, a TCP-style handshake is proceeded before packet transmission to indicate which packet the receiver current needs. Now we are ready to formally define the 2HR-f algorithm.

2HR-f Algorithm: Every time a node (say K) gets a transmission opportunity, it operates as follows:

Step 1: (Source-to-Destination) Check if node $D(K)$ is among its one-hop neighbors. If so, a handshake goes as follows: $D(K)$ first sends its current *request number* RN to K, then K compares RN with the *send number* $SN(P_h)$ of the packet P_h at the head of its local-queue.

- If $SN(P_h) > RN$, K retrieves from its already-sent-queue the packet with $SN = RN$, and deletes all packets with $SN \leq RN$ inside the already-sent-queue;
- If $SN(P_h) = RN$, K sends P_h directly to $D(K)$, moves ahead remaining packets waiting at its local-queue and deletes all packets with $SN < RN$ in its already-sent-queue;
- If $SN(P_h) < RN$ (then $RN = SN(P_h) + 1$), K sends the packet behind P_h with the *send number* equal to RN directly to $D(K)$, steps ahead remaining packets inside its local-queue (by two packets) and empties its already-sent-queue.

Step 2: Otherwise, K flips a unbiased coin, and choose either one from the following:

- (Source-to-Relay) K randomly selects one node as relay from its current one-hop neighbors, and a similar handshake between them proceeds to indicate whether the selected node, say R, has received one copy of P_h, i.e., the packet for which node K is distributing copies. If so, K remains idle for this time slot. Otherwise, K sends a copy of P_h to R, and checks whether f copies has already been delivered out for P_h, if yes, K move ahead its local-queue and put P_h to the end of its already-sent-queue. At the relay node, R adds P_h to the end of its relay-queue specified for node $D(K)$.
- (Relay-to-Destination) K acts as a relay and randomly selects one node as receiver from its one-hop neighbors. The selected receiver, say V, sends its *request number* $RN(V)$ to K, and K checks whether a packet with

$SN = RN(V)$ exists inside its relay-queue destined for V. If found, K sends it directly to V, and deletes all packets with $SN \leq RN(V)$ from its relay-queue for V. Otherwise, K remains idle for this time slot.

Note that in both of the above two steps, every time a node moves ahead its local-queue by one packet (or receives a packet destined for itself), it increases its *send number* (or *request number*) by one.

Remark 1: Although we allow multiple copies of a single packet to coexist in the network, each copy travels at most two hops, from source to a relay node, and from relay node to the destination. Actually only the copy which reaches the destination at the first time will travel two hops, the other f copies will be flushed out from their queue buffers after receiving confirmation from the destination node. Thus the queue sizes of the already-sent-queue and $n-2$ relay-queues will not grow indefinitely. And according to the *request number* sequence, all packets are received in order by their destinations.

Remark 2: It takes up at most $f+1$ transmission opportunities for each packet to reach its destination, and nearly every packet received at its destination will consume exact $f+1$ transmission opportunities. There are only two ways in which a packet will take less than $f+1$ transmissions, (1) node $D(K)$ receives this packet directly from node K, (2) $D(K)$ first receives this packet from one of its relay nodes and then crashes into K (to notify K to stop delivering out remaining copies). Note that either of these two events should happen before K finishes the distribution of all f copies, which happens with a vanishingly small probability (as we show later). So in the actual case, as n scales up every packet takes up $f+1$ transmissions to reach its destination with high probability.

B. A Scheduling Scheme

As we assume a cell-partitioned network and a time slotted system, we make the following restrictions on the transmissions during each time slot:

- For each time slot, we allow at most one transmitter inside each cell, if there are more than one nodes inside a cell then a transmitter is chosen randomly.
- If some node, say K, wins the transmission opportunity in its cell for the current time slot, it can send packets only to nodes in the same cell or its eight adjacent cells. Two cells are said to be adjacent if they share a common points. Thus, the maximum distance between a transmitter and receiver is $\sqrt{8/n}$, then we set the communication range as $r = \sqrt{8/n}$.
- Every time a node gets a transmission opportunity, it follows 2HR-f.

As the wireless transmissions interfere with each other, only cells that are sufficiently far away could simultaneously transmit without interfering each other. In our scheme we allow a receiver to be selected among any of the eight adjacent cells, thus the maximum number of cells that support a transmitting node during every time slot is finite. Toward this end, we define "transmission-group" of cells such that one node in each cell of the transmission-group can transmit simultaneously without interfering with one another.

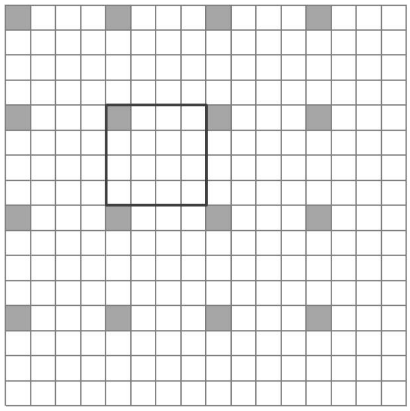

Fig. 1. An example of a transmission-group of cells with $\alpha = 4$. The shaded cells all belong to the same transmission-group.

Transmission-group: A transmission-group is defined to be a subset of cells, which keeps a neighboring distance of some constant number of cells, say α (an integer), in both vertical and horizontal directions. The shaded cells in Fig. 1 represent one such transmission-group.

Now we are able to determine the value of α for our scheduling scheme. Suppose during some time slot, node V is scheduled to receive a packet. Then, according to the definition of "transmission-group", the closest another simultaneous transmitting node (other than V's transmitter), say K, is at a distance of at least $(\alpha-2)/\sqrt{n}$ away from V. The condition that K will not interfere the reception at V is that,

$$(\alpha-2)/\sqrt{n} \geq (1+\Delta) \cdot r$$

substituting $r = \sqrt{8/n}$, we obtain that

$$\alpha \geq (1+\Delta)\sqrt{8} + 2.$$

As α is an integer, we take $\alpha = \lceil (1+\Delta)\sqrt{8} \rceil + 2$, where $\lceil x \rceil$ returns the smallest integer not smaller than x.

Note that there are only a finite number of transmission-groups, i.e., α^2, and each cell belongs to an individual transmission-group. If we let each transmission-group become active (have transmission opportunity) alternatively, then each transmission-group will be active in every α^2 time slots. In other words, each cell is activated in every α^2 time slots.

C. Probability of Medium Contention

We show here explicitly the probability of contending for a transmitting or receiving opportunity.

Lemma 1: During any time slot, consider some active cell, as n approaches infinity, there exists contention for a transmitting opportunity with probability not smaller than $1-2\mathrm{e}^{-1}$; and there exists contention for a receiving opportunity with probability not smaller than $1-\mathrm{e}^{-1}-\frac{19}{2}\mathrm{e}^{-9}$.

Proof: For a given active cell, it has a contention for a transmitter, i.e., at least two nodes appear inside simultaneously. It happens with

$$1-(1-\frac{1}{n})^n-\binom{n}{1}\frac{1}{n}(1-\frac{1}{n})^{n-1}$$
$$=1-(1-\frac{1}{n})^{n-1}(2-\frac{1}{n})\to 1-2\mathrm{e}^{-1}.$$

Similarly, the probability of having a contention for a receiver can be expressed as

$$\begin{aligned}&1-(1-\frac{1}{n})^n-\sum_{k=1}^{2}\binom{n}{k}(\frac{1}{n})^k(1-\frac{9}{n})^{n-k}\\&-\binom{n}{2}\binom{2}{1}\frac{1}{n}\cdot\frac{8}{n}(1-\frac{9}{n})^{n-2}\\&=1-(1-\frac{1}{n})^n-(1-\frac{9}{n})^{n-2}(\frac{19}{2}-\frac{35}{2n})\\&\to 1-\mathrm{e}^{-1}-\frac{19}{2}\mathrm{e}^{-9}.\end{aligned}$$

■

IV. Expected End-to-End Delay and Throughput Capacity

Before we move on with our main results, we first present some basic probability results under 2HR-f and the scheduling scheme introduced in Section III.

Lemma 2: For any time slot, consider some node, say K, we denote by p_1 the probability of K conducting a packet transmission, by p_2 the probability of K conducting a source-to-destination transmission, and by p_3 the probability of K conducting a source-to-relay or relay-to-destination transmission. If we assume a saturate traffic and the local-queue of K always has waiting packets, then we have the following expressions for p_1, p_2, p_3 under 2HR-f and our scheduling scheme,

$$p_1=\frac{1}{\alpha^2}\Big(1-(1-\frac{1}{n})^n-(1-\frac{9}{n})^{n-1}\Big)\tag{1}$$
$$p_2=\frac{1}{\alpha^2}\Big(\frac{8}{n-1}-(1-\frac{1}{n})^{n-2}(\frac{7}{n}+\frac{1}{n^2})\Big)\tag{2}$$
$$p_3=\frac{1}{\alpha^2}\Big(\frac{n-9}{n-1}-(1-\frac{9}{n})(1-\frac{1}{n})^{n-2}-(1-\frac{9}{n})^{n-1}\Big)\tag{3}$$

Proof: Under 2HR-f and our scheduling scheme, K conducts a packet transmission iff the following three events happen simultaneously, K appears in some active cell, K is selected as the transmitter, and there is at least one other node inside the corresponding nine cells. Thus, we have

$$\begin{aligned}p_1=\frac{1}{\alpha^2}\Bigg(&\sum_{k=1}^{n-1}\binom{n-1}{k}(\frac{1}{n})^k(1-\frac{1}{n})^{n-1-k}\frac{1}{k+1}\\&+\sum_{k=1}^{n-1}\binom{n-1}{k}(\frac{8}{n})^k(1-\frac{9}{n})^{n-1-k}\Bigg).\end{aligned}\tag{4}$$

Similarly, K conducts a source-to-destination transmission iff the following three events happen simultaneously, K appears in some active cell, K is selected as the transmitter, and node $D(K)$ appears inside the corresponding nine cells. Thus, we have

$$\begin{aligned}p_2=\frac{1}{\alpha^2}\Bigg(&\sum_{k=0}^{n-2}\binom{n-2}{k}(\frac{1}{n})^k(1-\frac{1}{n})^{n-2-k}\frac{1}{k+2}\frac{1}{n}\\&+\sum_{k=0}^{n-2}\binom{n-2}{k}(\frac{1}{n})^k(1-\frac{1}{n})^{n-2-k}\frac{1}{k+1}\frac{8}{n}\Bigg).\end{aligned}\tag{5}$$

K conducts a source-to-relay or relay-to-destination transmission iff the following four events happen simultaneously, K appears in some active cell, K is selected as the transmitter, there is at least one other node (except K and $D(K)$) inside the corresponding nine cells, and node $D(K)$ appears in one of the other $n-9$ cells. Thus, we have

$$\begin{aligned}p_3=\frac{1}{\alpha^2}(1-\frac{9}{n})\Bigg(&\sum_{k=1}^{n-2}\binom{n-2}{k}(\frac{1}{n})^k(1-\frac{1}{n})^{n-2-k}\frac{1}{k+1}\\&+\sum_{k=1}^{n-2}\binom{n-2}{k}(\frac{8}{n})^k(1-\frac{9}{n})^{n-2-k}\Bigg).\end{aligned}\tag{6}$$

After several basic algebraic operations, (1), (2) and (3) can be derived with (4), (5) and (6) respectively. ■

Remark 3: As a verification of the expressions of p_1, p_2, and p_3, one can easily prove that $p_1=p_2+p_3$. As discussed in Remark 2, we can verify that p_2 vanishes quickly as n scales up.

Lemma 3: For some node, say K, consider the packet at the head of its local-queue, i.e., packet P_h. Given that there are already m ($m\leq f+1$) copies of P_h inside the network and $SN(P_h)=RN(D(K))$, we assume in next time slot, node $D(K)$ will receive P_h with probability $p_a(m)$, and K will successfully deliver out a new copy of P_h (if $m\leq f$) with probability $p_c(m)$. Then we have the following results for $p_a(m)$ and $p_c(m)$,

$$p_a(m)=p_2+\frac{m-1}{2(n-2)}\cdot p_3\tag{7}$$
$$p_c(m)=\frac{n-m-1}{2(n-2)}\cdot p_3.\tag{8}$$

Proof: Given that there are already m copies of packet P_h inside the network, i.e., $m-1$ replicas have been distributed by node K to distinct relay nodes. We further assume that in next time slot, node $D(K)$ will directly receive P_h from K with probability $p_{s\to t}(m)$, and from some relay node with probability $p_{r\to t}(m)$. Then we have

$$p_{s\to t}(m)=p_2\tag{9}$$

$$\begin{aligned}&p_{r\to t}(m)\\&=\phi_1\sum_{t=0}^{n-3}\binom{n-3}{t}\sum_{k=0}^{t}\binom{t}{k}(\frac{1}{n})^{k+1}(\frac{8}{n})^{t-k}(1-\frac{9}{n})^{n-3-t}\cdot\\&\quad\cdot\frac{1}{t+1}(\frac{1}{k+2}+\frac{8}{k+1})\end{aligned}\tag{10}$$
$$\begin{aligned}&=\phi_1\frac{1}{n-2}(\frac{n}{n-1}-(1-\frac{1}{n})^{n-2}-(1-\frac{9}{n})^{n-2})\\&=\frac{p_3}{2(n-2)}\end{aligned}\tag{11}$$

where $\phi_1 = \frac{1}{2\alpha^2}(1-\frac{9}{n})$ and (10) follows directly from the definition of 2HR-f and our scheduling scheme in Section III.

As in the next time slot, $D(K)$ either receives packet P_h from K directly or from one of other $m-1$ relays. Note that these are mutually exclusive events, thus we have

$$p_a(m) = p_{s\to t}(m) + \sum_{i=1}^{m-1} p_{r\to t}(m). \tag{12}$$

After substituting (9) and (11) into (12), it follows (7).

According to Step 2 in 2HR-f, a relay node is chosen randomly from the one-hop neighbors of node K. Thus K will successfully deliver out a new copy of packet P_h, equals that a node other than the $m-1$ relay nodes (who already owns a copy of P_h) will be chosen as receiver in Step 2. Hence we have

$$\begin{aligned} &p_c(m) \\ &= \phi_2 \sum_{k=1}^{n-2} \binom{n-2}{k} \sum_{j=0}^{k} \binom{k}{j} (\frac{1}{n})^j (\frac{8}{n})^{k-j} (1-\frac{9}{n})^{n-2-k} \frac{1}{j+1} \\ &= \frac{n-m-1}{2(n-2)} \cdot p_3 \end{aligned} \tag{13}$$

where $\phi_2 = \frac{1}{2\alpha^2}(1-\frac{9}{n})\frac{n-m-1}{n-2}$. ■

As indicated in former section, some packet at the head of its local-queue, say P_h, might successfully reach its destination with less than $f+1$ transmission opportunities consumed, although with vanishingly small probability. However, even this case happens, it does not necessarily mean the packet waiting just behind P_h in the local-queue, will move ahead into the head of line and get service immediately. Actually if the destination node receives a copy of P_h from one of its relay nodes rather than the source node, the source node will continue to deliver out the remaining copies for P_h, unless the destination node crashes into it and request for the next packet before remaining copies are delivered. Obviously, this happens with a much smaller probability.

The above analysis allows us to assume a fixed service rate for the local-queue at each node, i.e., exact f copies for each packet waiting in the local-queue are delivered to distinct relay nodes, and the destination starts to receive some packet after $f+1$ copies already exist in network. Obviously, the mean end-to-end delay derived under this assumption naturally becomes an upper bound for the actual expected end-to-end delay of 2HR-f, and the whole system can further be treated as two single-server queues in tandem. Hence we have the following result.

Theorem 1: If we denote by S_1 the time it takes some node to deliver out f copies of the head-of-line packet at the local-queue, and S_2 the time it takes the destination node to receive one of the $f+1$ copies, $\mathbb{E}\{T_e\}$ the actual expected end-to-end delay per packet of 2HR-f, and μ the per-node throughput capacity of 2HR-f, then the network can stably support rates $\lambda < \mu$, and we have

$$\mu = \min\{\frac{1}{\mathbb{E}\{S_1\}}, \frac{1}{\mathbb{E}\{S_2\}}\} \tag{14}$$

$$\mathbb{E}\{T_e\} \leq \frac{\mathbb{E}\{S_1\}}{1-\rho_1} + \frac{\mathbb{E}\{S_2\}}{1-\rho_2} - \frac{\rho_2}{2(1-\rho_2)} \tag{15}$$

$$\mathbb{E}\{S_1\} = \frac{2(n-2)}{p_3} \sum_{m=1}^{f} \frac{1}{n-m-1} \tag{16}$$

$$\mathbb{E}\{S_2\} = \frac{1}{p_2 + \frac{f}{2(n-2)} \cdot p_3} \tag{17}$$

where $\rho_1 = \lambda\mathbb{E}\{S_1\}$ and $\rho_2 = \lambda\mathbb{E}\{S_2\}$.

Proof: We first derive the expected time it takes some node to deliver out f copies of the head-of-line packet at its local-queue, i.e., the expected service time $\mathbb{E}\{S_1\}$.

$$\mathbb{E}\{S_1\} = \sum_{m=1}^{f} \frac{1}{p_c(m)} \tag{18}$$

and (16) follows after substituting (8) into (18).

Note that given there are $f+1$ copies of some packet (with a *send number* equals *request number* of its destination node) inside the network, then this packet will arrive at its destination with a probability $p_a(f+1)$ during next time slot. Hence we have

$$\mathbb{E}\{S_2\} = \frac{1}{p_a(f+1)} \tag{19}$$

and (17) follows after substituting (7) into (19).

The proof of (15) is similar to the derivation of the standard *Pollaczek-Khinchin* formula for mean waiting time in an $M/G/1$ queue. Consider some tagged packet, arriving to the local-queue of some node, say K, at the beginning of some time slot. First it has to wait (if the queue is not empty) in the local-queue for service, i.e., to be replicated and delivered to f distinct relays. Let W_1 denote the time this packet spends waiting in the local-queue before getting service, hence we have

$$W_1 = \sum_{i=1}^{L_q} X_i + R \tag{20}$$

where variable R denote the residual service time, L_q represent the number of packets waiting in the queue, and X_i denote the service time of the i_{th} packet. As the service times $\{X_i\}$ are independently in expectation bounded by $\mathbb{E}\{S_1\}$, thus we have $\mathbb{E}\{X_i\} \leq \mathbb{E}\{S_1\}$. If we let ρ_{actual} represent the probability that the node is busy with delivering copies of some packet, and let $\mathbb{E}\{X\}$ represent the actual mean time it takes the node to service a generic packet, since $R = \rho_{actual}\mathbb{E}\{X\}$ and by Little's law (applied to the server) $\rho_{actual} = \lambda\mathbb{E}\{X\}$, it follows that $\rho_{actual} \leq \rho_1$ and $R \leq \rho_1\mathbb{E}\{S_1\}$. It yields by taking expectations of the both sides of (20):

$$\begin{aligned} \mathbb{E}\{W_1\} &\leq \mathbb{E}\{L_q\}\mathbb{E}\{S_1\} + \rho_1\mathbb{E}\{S_1\} \\ &= \lambda\mathbb{E}\{W_1\}\mathbb{E}\{S_1\} + \rho_1\mathbb{E}\{S_1\} \\ &= \rho_1\mathbb{E}\{W_1\} + \rho_1\mathbb{E}\{S_1\}. \end{aligned} \tag{21}$$

We thus have

$$\mathbb{E}\{W_1\} \leq \frac{\rho_1 \mathbb{E}\{S_1\}}{1-\rho_1} \tag{22}$$

where $\lambda < 1/\mathbb{E}\{S_1\}$ and $\rho_1 = \lambda\mathbb{E}\{S_1\}$.

Note that we say the packet enters the second queuing process as soon as f copies of this packet have been delivered (i.e., as soon as this packet finishes its service in its local-queue and departs its source node). As the time required to deliver out the $(m+1)_{th}$ copies of this packet has a geometric distribution with mean $1/p_c(m)$, S_1 can be interpreted as a sum of f mutually independent, identically distributed geometric random variables. According to theory of departure processes in [15], the input to the second queuing process can be approximated as a Poisson stream with arrival rate λ. Recall that in 2HR-f destination node receives packets in *request number* order, thus in the second queuing process, a packet has to wait until its *send number* equals the *request number* of its destination node (i.e., the destination node has received all the preceding packets). We denote by W_2 this waiting time, using the *Pollaczek-Khinchin* formula, we have

$$\mathbb{E}\{W_2\} = \frac{\rho_2 \mathbb{E}\{S_2\}}{1-\rho_2} - \frac{\rho_2}{2(1-\rho_2)} \tag{23}$$

where $\lambda < 1/\mathbb{E}\{S_2\}$ and $\rho_2 = \lambda\mathbb{E}\{S_2\}$.

When a new packet reaches the head-of-line at its local-queue, and its SN equals the RN of its destination node, the time required for the packet to reach its destination is at most $S_1 + S_2$, thus we have

$$\mathbb{E}\{T_e\} \leq \mathbb{E}\{W_1\} + \mathbb{E}\{S_1\} + \mathbb{E}\{W_2\} + \mathbb{E}\{S_2\} \tag{24}$$

where $\mathbb{E}\{T_e\}$ denotes the actual expected end-to-end delay per packet of 2HR-f. After substituting (22) and (23) into (24), it follows (15). ∎

Remark 4: Theorem 1 provides a closed-form (rather than order sense) results for the achievable throughput per node and the expected end-to-end delay per packet in 2HR-f-based ad hoc mobile networks. Based on Theorem 1 and some typical settings of f, one can easily derive the corresponding order sense results. For example, by setting $f = 1$ one can obtain a $\Theta(1/n)$ throughput and $O(n)$ delay; by setting $f = \sqrt{n}$, one can easily recover the order sense results ($O(1/\sqrt{n})$ throughput and $O(\sqrt{n})$ delay) reported in Theorem 6 of Neely and Modiano [9].

Remark 5: One may also notice that when setting $f = 1$, Theorem 1 results in a $O(n)$ delay and a $\Theta(1/n)$ throughput, which is different from the throughput result $\Theta(1)$ reported in [1]. This reduced throughput is due to the rule of "reception in order" employed in 2HR-f. The restriction of receiving packets according to *request number* ensures that all packets arrive at the destination in order, but it wastes a lot of opportunities to receive "out of order but fresh" packets (i.e., packets with *send number* larger than the current *request number* of destination node). Thus the benefits of receiving all packets in order comes at the price of a reduced throughput per node.

V. Conclusion

We extend the classic 2-hop relay algorithm under a general setting, i.e., f-*cast*, and propose a 2-hop relay f-cast algorithm (2HR-f) and a corresponding scheduling scheme. We develop closed form upper bounds for the expected end-to-end delay and obtainable throughput in a 2HR-f-based ad hoc mobile network. Furthermore, our closed form bound covers a range of models, from which one can easily derive the order sense results.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and capacity trade-offs for mobile ad hoc networks: A global perspective," *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. 15, no. 5, pp. 981–992, October 2007.

[6] 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.

[7] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *INFOCOM*, 2004.

[8] 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.

[9] 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.

[10] 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/

[11] 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.

[12] R. Groenevelt, G. Koole, and P. Nain, "Message delay in manet (extended abstract)," in *ACM Sigmetrics*, 2005.

[13] T. Small and Z. Hass, "Resource and performance tradeoffs in delay-tolerant wireless networks," in *ACM SIGCOMM Workshop on Delay Tolerant Networks (WDTN)*, 2005.

[14] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 46, no. 2, pp. 388–404, March 2000.

[15] P. Burke, "The output of queuing systems," *Operations Research*, vol. 4, no. 6, pp. 699–704, December 1956.