Control Reduces Delay in Mobile Wireless Networks with Maximal Throughput

Holly A. Waisanen
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: waisanh@mit.edu

Devavrat Shah
Department of Electrical Engineering
and Computer Science
& Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: devavrat@mit.edu

Munther A. Dahleh
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: dahleh@mit.edu

Abstract

Previous research in the tradeoff between throughput and delay in wireless networks has focused on networks with fixed nodes or nodes with random mobility. In this paper, motivated to study fundamental limits of delay in a wireless network with maximal throughput, we examine delay scaling in networks in which the mobile nodes may control their own motion and schedule the pickup and delivery of messages so as to achieve maximal throughput while minimizing delay. We find that by letting nodes control their mobility, even with very minimal information, results in significant delay reduction.

I. Introduction

We consider a communication network in which messages orginating in a geographic region must be delivered to their destinations elsewhere in the region. This service is carried out by a number of (possibly) mobile nodes or vehicles. Each node is capable of transmitting messages wirelessly to other nodes, however a very distinct characteristic of wireless transmission is *interference*: a wireless transmission by one node may adversely affect other transmissions occuring simultaneously. Interference generally implies that wireless transmissions should only take place over short distances to minimize interference and allow the greatest number of nodes to transmit simulataneously. Since messages may be destined to locations very far from their origin, several wireless transmissions may be required for delivery unless there is another method of transport.

If the nodes are mobile, messages may also be physically carried over distances in the region. Physical transport of messages has the advantage that it does not create interference and many messages may be carried at one time by a single node. However, node velocity is typically much less than the speed of electromagnetic propagation of wireless transmissions, and so this is a much slower method of delivering messages.

Building on the seminal results in throughput analysis of [1], [2], El Gamal, Mammen, Prabhakar and Shah [3] studied the question of achievable throughput and delay tradeoff. They obtained the following optimal tradeoff: (a) for fixed random networks, the throughput per node, $T(n)$ and delay per packet, $D(n)$ are related as $T(n) = \Theta(nD(n))$ for $T(n) = O(1/\sqrt{n \log n})$; and (b) for mobile networks with each node performing independent random walk, for most of the throughput[1], the delay scales $D(n) = \Theta(n \log n)$. The result of [3] for mobile networks provides a very pessimistic conclusion: even at the loss of significant throughput, the delay can not be reduced under a random walk based mobility model.

In search of better delay scaling, various authors [4]–[6] have suggested different mobility models. While some of these models provide significant delay reduction at the loss of

[1]Precisely, for throughput $T(n) \in [n^{-0.5+\delta}, \Theta(1)]$, for any $\delta > 0$.

throughput, they are far from being realistic. Most assume that node motion is completely random, regardless of the current message delivery requirements. Because they consider specific models, they are not able to make statements on optimal performance achievable under any mobility model.

In practice, one expects nodes to have control of their movement. Motivated by this consideration, in this paper, we do not assume any particular node mobility model. Instead, we consider a situation in which each node can move within a geographic area in any manner it chooses with the physical constraint of bounded velocity. In fact, in contrast to previous results on random mobility, we assume that the primary task of each node is to provide network infrastructure. Given this, we wish to answer the following questions: (a) Is it possible to reduce delay while retaining maximal throughput by controlling node mobility? (b) More generally, what is the optimal delay scaling achievable for maximal throughput? Note that maximal throughput requires that much of the message transport is achieved by physical means. This is most directly related to the results of Grossglauser and Tse [1]. Future work will consider extensions to the case where throughput may be decreased in favor of also decreasing delay.

In this paper, we answer the above questions. We find that by controlling node mobility, the delay improves drastically over the case of the random mobility models previously considered. Before we state our results, we present the details of the model, the problem statement and some notation. We also make a connection to models and results in vehicle routing, a field that seems far removed from the traditional throughput-delay analysis of wireless networks.

A. Model and Problem

Let the n nodes be in a geographic area $\mathcal{A} \subset \mathbb{R}^2$, which is a convex, compact set with volume A. For simplicity, we consider $\mathcal{A} = [0, \sqrt{A}]^2$, with the understanding that these results may be extended to other convex environments with the same area. Each node may move in any direction at any time with a velocity of magnitude $\leq v$. For simplicity, assume $v = 1$. We will call each node a *vehicle*. Messages are generated according to a Poisson process with time intensity λ. We assume that $\lambda/n \to \infty$. We note that the Poisson assumption is not necessary, however we consider it for simplicity. Associated with each message j are source and destination locations denoted by $s(j) \in \mathcal{A}$ and $d(j) \in \mathcal{A}$ respectively. Source and destinations locations are IID and uniformly distributed in the region $\mathcal{A}$.

The messages need to be picked up from their source locations and delivered to their destination locations by vehicles. Here, the aim is to study delay scaling for maximal throughput. This essentially means that each message is not transmitted more than a constant number of times. Inspired by the scheme of Grossglauser and Tse [1], we require that each message is picked up and delivered by the same vehicle without any hopping (or transmission of the message to other vehicles).

Further, we require that the vehicle can pick up (deliver) a message only by reaching the source (destination) location and receiving (transmitting) it. This effectively eliminates any wireless interference that may be created by longer distance transmissions. Without delving into further details, note that we will not directly consider the effect of longer range transmissions on throughput and delay in this paper. When a vehicle is at the source (destination) location of a message, it needs to spend time δ to receive a message of δ size. Under our model, we assume that message size is $\hat{p} = \frac{n}{K\lambda}$ for large enough constant K. Thus, each vehicle spends $\hat{p}$ amount of time at each pickup (delivery) location. This packet scaling will be required to ensure the stability of the network for large message arrival rates λ.

Under the above setup, if the vehicles manage to pickup and deliver all messages arriving at rate λ, then the corresponding total throughput, $T(n)$, that is, the rate at which data is transfered through the network is

$$T(n) \;=\; \lambda \times \hat{p} \;=\; \frac{n}{K} \;=\; \Theta(n). \tag{1}$$

Thus, maximal possible network throughput of $\Theta(n)$ or $\Theta(1)$ per vehicle is achieved.

A control policy is a set of decision making rules for each vehicle that decides the pickup and delivery schedule of arriving messages, based on a set of constraints on information available to the vehicle. In particular, we consider two types of information structure: (a) *Source only information*: vehicles do not know the destination of messages until they pick them up; and (b) *Source-destination information*: vehicles know both the source and destination information before they pick up. Under both structures, we assume that vehicles have information about messages within a local region and that message information is transmitted instantaneously over a separate wireless control channel which does not create interference with the transmission of messages.

Therefore, the objective of minimizing delay for maximal throughput corresponds to the following problem.

Problem statement. Design a valid control policy for each vehicle that decides the pickup and delivery schedule of arriving messages such that: (a) The messages are delivered at rate λ by the n vehicles collectively, and (b) the average message delay is minimized, where delay of a message is the time it takes to reach destination location from time of its arrival (not pickup). We will call the above defined control problem as the Source-Destination-Pickup-Delivery (SDPD) problem.

B. Related work

The SDPD problem is naturally related to a very well-studied problem of vehicle routing (VR) in the Operations Research community. Bertsimas and van Ryzin [7]–[9] formulated the VR problem as the Dynamic Traveling Repair-person Problem (DTRP), which is as follows: requests arrive in a convex environment $\mathcal{A}$ of area A according to a Poisson process of time intensity λ with locations of requests being IID distributed uniformly in $\mathcal{A}$. There are n repair-persons to serve them with bounded velocity. Each repair-person spends a random amount time to serve a request after reaching its location. [7]–[9] obtain several policies which achieve order-optimal average delay for this problem and showed that under the optimal policy the average delay per request scales as $\Theta(\lambda A/n^2)$, where delay of a request is the time taken to be serviced (including service time) since its arrival. Intuitively, this seems similar to our problem as both pickup and delivery of message in SDPD can be treated as separate requests in the DTRP setup. However, there is a strong causality relation between pickup and delivery of each message. Hence, both problems are very distinct.

Frazzoli and Bullo [10] have recently presented a set of distributed algorithms for the n-vehicle DTRP that achieve the same delay performance as the optimal lower bounds in [8] using only local communication between vehicles and locally sensed information. We will seek similarly distributed control policies in this research.

As the results for the DTRP are not directly applicable to the SDPD problem studied here, we mention a more relevant problem, the Online Dial-a-Ride Problem (OLDARP) [11], [12]. As in our problem, messages need to be picked up by vehicles from a random arrival location and dropped off at random destination location, but he goal of the OLDARP problem is usually to minimize the service completion time of a given finite collection of messages.

The closest to SDPD problem is the Dynamic Pick-up Deliveryperson Problem (DPDP) proposed in [13]. In this setup, a single service vehicle is responsible for picking up and delivering all messages that arrive. The goal is to minimize the average delay experienced by all messages. In contrast to DPDP, we are interested in the multi-vehicle case.

C. *Main results*

Our model is different from that of [1]–[3]. Hence, results of delay scaling of [3] do not apply directly. To this end, we consider a policy (described in detail in Section II) based on random walk similar to the mobility model of [3]. This policy, denoted by RW, has delay lower bounded by as stated below, which is similar to the scaling in [3].

Theorem 1: For any λ such that $n \ll \lambda$, the RW policy achieves maximal throughput and the average delay scales as $\Omega\left(\frac{\lambda A}{n}\log\frac{A\lambda^2}{n^2}\right)$.

We note that the quantitative difference between the delay scaling of [3] (which was $\Theta(n\log n)$) and the statement of Theorem 1 is due to the difference in normalization of velocity, area and packet-size.

We then analyze two policies which exploit the two different information structures described in Section I-A and demonstrate that with controlled mobility, the average delay experienced by messages in the system can be reduced. In particular, we state the following two theorems:

Theorem 2: Under the Source only information structure, the minimum average delay per message scales as $O(\lambda A/n)$.

Theorem 3: Under Source-Destination information, the minimum average delay per message scales as $O(\lambda A/n^{3/2})$.

While the policies examined in the proofs of Theorems 2 and 3 provide an upper bound on optimal performance by their existence, we conjecture that these policies are in fact of optimal order.

D. *Organization*

The rest of the paper is organized as follows. In Section II, we provide the proof of Theorem 1. Section III describes and analyzes policies that achieves the claimed performance. Finally, in Section IV we present discussion and directions for future work.

II. Proof of Theorem 1: The RW policy

In this section, we prove the Theorem 1. We first describe the Random Walk (RW) policy as follows: divide the area $\mathcal{A}$ into a $\frac{\lambda\sqrt{A}}{c_1 n} \times \frac{\lambda\sqrt{A}}{c_1 n}$ grid of square cells, each of area $\frac{c_1^2 n^2}{\lambda^2}$, where $c_1 < 1$ is an arbitary cell scaling constant. For ease of reference, let $n_A = \frac{\lambda^2 A}{c_1^2 n^2}$ be the number of cells in the grid. Each vehicle performs a symmetric random walk along the centers of these squares. That is, each time step the vehicle moves to the center of one of the four neighboring squares of the square in which it is currently residing, each transition being equally likely. Each step of the random walk occurs every $c_2\frac{n}{\lambda}$ time, where c_2 is a time scaling constant to be determined. c_2 will in general depend on the cell area constant c_1.

The pickup and delivery policies are described as follows:

(a) *Source pickup.* When a vehicle enters a square, say square $\mathbf{s}$, containing unserviced and unassigned messages, the vehicle selects one message to service, uniformly at random.

This assignment is communicated to any other vehicles in the same square, preventing multiple assignment of the message. When a message is assigned to a vehicle, the vehicle moves from the center of s to the source location, spends time $\frac{n}{K\lambda}$ for transmission, and returns to the center before any further operation.

(b) *Destination delivery.* Once any message pickups are performed, the vehicle performs message deliveries for at most one message currently being carried by the vehicle that has its destination in square s. Again, the vehicle starts from the center of the square and goes to the destination location, spends time $\frac{n}{K\lambda}$ and returns to center of square before delivering or picking up any other message in the same square. When multiple messages have destinations in the current square, the vehicle delivers them in the first come first serve order.

At most one assignment/pickup and at most one delivery may occur each time step before the vehicle moves to the next square on its random walk.

Although this model is far from realistic, in comparison to the fully controlled policies we will soon examine, this example serves to show that even a small amount of forced randomness significantly increases delay.

Now we state the proof of Theorem 1.

Proof: [Theorem 1] First note that, under the RW policy, the queue formed at any of the vehicles has arrival and service processes dependent on the external Poisson arrival process, uniform distribution of source and destination locations, and the Markov chain describing the independent random walks of n vehicles. These are jointly stationary and ergodic processes. Hence, the queue-size process has jointly stationary and ergodic arrival and service processes.

Given this, for stability, we need to show that the average interarrival time is strictly greater than the average service time. Now, due to the independent random walks of the vehicles and the uniform source location distribution, under the RW policy, the arrival rate to each node is λ/n. Therefore, for stability, we require that the average service time is less than n/λ.

Each message is completely served in two stages, pickup and delivery, during two possibly different random walk time steps. We may therefore upper bound the service time by twice the random walk step time. In each step time, the vehicle must pickup at most one message, traveling back and forth between center of the cell and the service location. This distance is at most $\sqrt{2}\frac{c_1 n}{\lambda}$. The vehicle may also deliver at most one message, again traveling a distance of at most $\sqrt{2}\frac{c_1 n}{\lambda}$. Each of these services also requires a transmission time of $\frac{n}{K\lambda}$. Finally, at the end of the time step, the vehicle must travel $\frac{c_1 n}{\lambda}$ to travel to the center of the next cell on its random walk. Adding these together, the random walk time step must be

$$c_2\frac{n}{\lambda} \geq \frac{2\sqrt{2}}{c_1}\frac{n}{\lambda} + \frac{2}{K}\frac{n}{\lambda} + c_1\frac{n}{\lambda} = (\frac{2\sqrt{2}}{c_1} + \frac{2}{K} + c_1)\frac{n}{\lambda}. \tag{2}$$

Bounding the service time by twice the random walk step size, we see that stability requires $\frac{n}{\lambda} > 2c_2\frac{n}{\lambda}$. By appropriate choice of packet size and cell area scaling constants, this service time can be made strictly smaller than the interarrival time and hence the queue at each vehicle is stable.

Next, we lower bound delay. For this, consider a message, say 0, arriving at a vehicle. Let the squares of the $\sqrt{n_A} \times \sqrt{n_A}$ grid be numbered from $\{0, \ldots, n_A\}$. Without loss of generality, let the message arrive in the square numbered 0. Let T_k be the random number of time-steps required by random walk to reach square k starting from square 0. Now, the destination of this message is distributed uniformly at random. Hence, the destination of message is in square

$k \in \{0, \ldots, n_A\}$ with probability $1/n_A$. It is well-known (see equation (22), pp. 17 [3]) that

$$\mathbb{E}\left[\frac{1}{n_A}\sum_{k=0}^{n_A-1} T_k\right] = \Theta(n_A \log n_A). \tag{3}$$

Let T denote the random time it takes to reach the destination square after the arrival of the message 0. Then from (3) and recalling that each time-step of random walk is $\Theta(n/\lambda)$ time, we obtain

$$\mathbb{E}[T] = \Theta\left(\frac{n}{\lambda} n_A \log n_A\right) \tag{4}$$

$$= \Theta\left(\frac{\lambda A}{n} \log \frac{\lambda^2 A}{n^2}\right). \tag{5}$$

For purpose of lower bound, we can use $\mathbb{E}[T]$ as a lower bound on average delay of message 0. Since message 0 was arbitrary, the bound of (5) applies to the average delay experienced by messages in the RW policy. Therefore, delay scales as $\Omega(\frac{\lambda A}{n} \log \frac{\lambda^2 A}{n^2})$.

This completes the proof of Theorem 1. ■

III. Policies

In this section, we describe policies that achieve the delay performance claimed in Theorems 2 and 3 for Source Only and Source Destination information respectively.

A. Source Only Policy

Recall that in the Source Only information structure, vehicles do not know the destination of messages before they are picked up, thus this information may not be used by vehicles in deciding which messages to pick up. Unlike the Random Walk policy above, messages are not scheduled for pickup immediately upon arrival but are instead assigned to vehicles which then schedule the pickups and deliveries according to a batching procedure.

(a) *Message Assignment.* Upon arrival, each message is assigned to the vehicle closest to its source location at the time of the message arrival. The message is not immediately picked up, but the vehicle is notified of the message assignment and retains the source location information of this message. According to the results of [1], this nearest neighbor communication is possible with throughput $O(1)$, thus the message assignment notification may be performed in a distributed way. Since the message source distribution is uniform and independent of the vehicle locations, the assignment of messages to each vehicle is Poisson with an expected arrival rate of λ/n. All messages assigned to a single vehicle that arrive in the interval $[kT, (k+1)T)$ form a batch, where T, the batch time interval, is a parameter to be determined. Each batch is deposited into a queue for its assigned vehicle upon formation at time $(k+1)T$ for appropriate k.

(b) *Message pickup and delivery.* Batches for each vehicle are served in First Come, First Serve order from the vehicle's batch queue. Pickups are performed along a shortest path through the source locations which is computed at the beginning of the interval. Once pickups are complete, a shortest path through the delivery locations is computed and the deliveries are performed accordingly. To perform each service, the vehicle stops at the source (destination) location for $\frac{n}{\lambda K}$ time to pickup (deliver) the associated message.

Theorem 2 The average delay per message for the Source only policy described scales as:

$$W_{SO} = O(\frac{\lambda A}{n}) \tag{6}$$

Therefore, the optimal average delay per message over all policies with Source Only information is $O(\frac{\lambda A}{n})$.

Proof: [Theorem 2] Since the batch interarrival time is fixed at T, the batches form a D/G/1 queue. This batching protocol is stable if and only if the expected time to service each batch of messages is less than T, the expected time between batch arrivals. The batch service time, T_B, has two components: 1) the travel time between service locations and 2) the onsite transmission time required to pickup (deliver) messages. Let N_T be the number of messages arriving in $[kT, (k+1)T)$ that are assigned to vehicle i. Each corresponds to two locations, both source and destination. To bound the travel time through the N_T source locations, note that the length of the shortest path is no more than the length of the optimal TSP tour. The length of the TSP tour may be bounded with the following asymptotic result originally due to Beardwood, Halton, and Hammersley [14]:

Theorem 4: Given N points uniformly distributed over a region of area A, there exists a constant $0 < \beta < \infty$ such that as $N \to \infty$, the expected length of the TSP tour through these points is:

$$E[L_N|N] = \beta\sqrt{N}\sqrt{A} \tag{7}$$

Furthermore, note that the expected length of a TSP tour through N uniformly distributed points is an upper bound on the expected length of a TSP tour through N points with an arbitrary distribution over the same support since clustering of locations can only decrease the average interpoint travel. Therefore, using Theorem 4 to bound the travel time required for each of the shortest paths (pickup and delivery), the total expected service time required to service the messages accumulated in $[kT, (k+1)T)$ is:

$$\begin{aligned} E[T_B] &\leq 2E[E[L_{N_T}|N_T] + N_T\frac{n}{\lambda K}] && (8)\\ &= 2E[\beta\sqrt{A}\sqrt{N_T} + N_T\frac{n}{\lambda K}] && (9)\\ &\leq 2\beta\sqrt{A}\sqrt{E[N_T]} + 2E[N_T]\frac{n}{\lambda K} && (10)\\ &= 2\beta\sqrt{A}\sqrt{\frac{\lambda T}{n}} + 2\frac{\lambda T}{n}\frac{n}{\lambda K} && (11) \end{aligned}$$

where (10) is by concavity of $\sqrt{\cdot}$.

Therefore the following bound on T is sufficient for stability:

$$\begin{aligned} T &\geq 2\beta\sqrt{\frac{\lambda A}{n}}\sqrt{T} + 2\frac{T}{K} \geq E[T_B] && (12)\\ \Longrightarrow T &\geq \frac{4\beta^2\lambda A}{n}(1 - \frac{2}{K})^{-2} && (13) \end{aligned}$$

Message delay has four components: 1) time waiting for batch to form, 2) time batch spends in queue, 3) time waiting for service of other vehicles in batch, and 4) time of own service. Since batch interarrival time $= T$, each message waits at most T for its batch to form, bounding 1). The expected time waiting for service of messages in the batch may be upper bounded by the expected batch service time, thus bounding 3)+4). Therefore the average message delay may be bounded as

$$W_{SO} \leq T + T_Q + T \tag{14}$$

where T_Q is the amount of time the batch spends in queue.

T_Q may be bounded by Kingman's bound:

$$T_Q \leq \frac{\lambda(\sigma_a^2 + \sigma_s^2)}{2(1-\rho)} \tag{15}$$

In this context, $\lambda = 1/T$, the arrival rate of batches. Since this is a D/G/1 queue, the variance of interarrival times is $\sigma_a^2 = 0$. Fixing T to be $\kappa\frac{4\beta^2\lambda A}{m}(1-\frac{2}{K})^{-2}$ for some $\kappa > 1$, we have $\rho \leq 1/\kappa$ and

$$T_Q = O\left(\frac{\sigma_s^2}{(1-1/\kappa)T}\right). \tag{16}$$

The variance of the batch service time, σ_s^2, may be bounded as follows:

Lemma 1: $\sigma_s^2 = O(\frac{\lambda A}{n}T) + O(\sqrt{\frac{\lambda A}{n}}T^{3/2})$

The proof of lemma 1 may be found in [15].

Therefore, substituting the result of Lemma 1 into (16), and recalling our definition of T as $O(\frac{\lambda A}{n})$, we have $T_Q = O(\frac{\lambda A}{n})$. Substituting T_Q into (14), we have the following bound on W_{SO}:

$$W_{SO} \leq T + O(\frac{\lambda A}{n}) + T \tag{17}$$

Again recalling that T was defined as $T = \kappa\frac{4\beta^2\lambda A}{m}(1-\frac{2}{K})^{-2} = O(\frac{\lambda A}{n})$, this becomes

$$W_{SO} = O(\frac{\lambda A}{n}) \tag{18}$$

■

B. Source-Destination Policy

In the Source-Destination information structure, destination information may be used by vehicles in deciding which messages to pick up. By exploiting this information, vehicles need not traverse the entire geographical region when servicing messages, but may instead only pick up messages that have both source and destination locations in a limited area. For this, a spatially based message assignment policy is used.

(a) *Message Assignment.* Divide the geographical region into an $\sqrt{\frac{A}{\sqrt{n}}} \times \sqrt{\frac{A}{\sqrt{n}}}$ grid of subregions, each of area $\frac{A}{\sqrt{n}}$. To each of the n ordered pairs of subregions, assign exactly one vehicle to service that pair. Each vehicle is assigned to pickup all messages that originate in the first subregion of its assigned ordered pair that have a destination location in second assigned subregion. As before, all messages assigned to a single vehicle that arrive in the interval $[kT, (k+1)T)$ form a batch, where T, the batch time interval, is a parameter to be determined. Each batch is deposited into a queue for its assigned vehicle upon formation at time $(k+1)T$ for appropriate k. Since message assignments are spatially based, this policy is easily implementable in a distributed way, assuming that region pairs are assigned upon initialization. It is assumed that each vehicle can keep track of its own batch queue despite possibly traveling out of communication range of its message source region while delivering messages.

(b) *Message Pickup and Delivery.* As before, batches for each vehicle are served in First Come, First Serve order from the vehicle's batch queue. Batch pickups and deliveries are performed in the same way as in the policy with Source only information with the notable addition of possible interregion travel time between source region and destination region.

Theorem 3 The average delay per message for the Source-Destination policy described above scales as:

$$W_{SD} = O(\frac{\lambda A}{n^{3/2}}) \tag{19}$$

Therefore, the optimal average delay per message over all policies with Source and Destination information is $O(\frac{\lambda A}{n^{3/2}})$.

Proof: [Theorem 3] Service of assigned messages is the same as in the Source only policy described above except that the TSP tours are performed over possibly distinct subregions of the environment. Each TSP tour now ranges over a subset of the geographical region $\mathcal{A}$ with area $A/\sqrt{n}$. Travel time between subregions must also be included in the batch service time analysis. Since the total geographical region is a square of area A, this interregion travel time may be upper bounded by $2\sqrt{A}$.

Therefore, as before, the total expected service time required to service the messages accumulated in $[kT, (k+1)T]$ is:

$$\begin{aligned} E[T_B] &\leq 2E[E[L_{N_T}|N_T] + N_T\frac{n}{\lambda K} + 2\sqrt{A}] \qquad (20) \\ &= 2\beta\sqrt{\frac{A}{\sqrt{n}}}\sqrt{\frac{\lambda T}{n}} + 2\frac{\lambda T}{n}\frac{n}{\lambda K} + 4\sqrt{A} \qquad (21) \end{aligned}$$

Therefore the following bound on T is sufficient for stability:

$$T \geq 2\beta\sqrt{\frac{\lambda A}{n^{3/2}}}\sqrt{T} + 2\frac{T}{K} + 4\sqrt{A} \geq E[T_B] \tag{22}$$

If $\lambda T/n^{3/2} >> 1$ (equivalently $\lambda >> n^{3/2}$), that is, the TSP service time dominates the interregion travel, the third term is dominated by the first two and by similar analysis as before:

$$T = O(\frac{\lambda A}{n^{3/2}}) \tag{23}$$

As before, the delay in the FCFS batch queue may be computed to bound message delay using Kingman's Bound and Lemma 1 as $\leq T + O(\frac{\lambda A}{n^{3/2}}) + T$. Lemma 1 was applied with $A = A/\sqrt{n}$ and $T = O(\frac{\lambda A}{n^{3/2}})$. Therefore, similar to the Source Only case with the altered scaling of the area A,

$$W_{SD} = O(\frac{\lambda A}{n^{3/2}}) \tag{24}$$

■

IV. Discussion and Conclusions

Previous analysis in throughput scaling as a function of n, the number of nodes in a wireless network, has focused on networks with fixed nodes or nodes with random mobility. In this paper, we consider using controlled mobility models in which vehicles (nodes) decide how to service the arriving messages. These decisions are carried out in a distributed manner with varying levels of information available to the vehicles at the time of message pickup. Taking a cue from [1], we focused on delay scaling for maximal $O(1)$ throughput by allowing messages to be transmitted to only a single mobile vehicle. First, we determined that the delay scaling for a network with a Random Walk mobility model with maximal throughput scales as $\Omega(\frac{\lambda A}{n}\log\frac{A\lambda^2}{n^2})$. Policies were then presented which retained maximal throughput but achieved delay scaling of smaller order than in the case of random mobility. In the case that Source Only information is available, we showed that the average delay scales as $O(\frac{\lambda A}{n})$ which is an $O(\log\frac{A\lambda^2}{n^2})$ improvement over the model with random mobility. In the case that both Source and Destination information is available, the average delay scales as $O(\frac{\lambda A}{n^{3/2}})$ which is an additional $O(\sqrt{n})$ improvement over the case where only source information is available. From a system design standpoint, these scalings quantify the perfomance improvements achievable by adding additional control and information gathering capabilities to the mobile nodes.

In future work, we intend to demonstrate that the policies presented for the two information structures are indeed order optimal. That is, we must show that these policies achieve the optimal delay scaling for maximal throughput. Another future challenge is an extension of this work to networks with non-maximal throughput, using transmissions between mobile nodes to achieve smaller delay.

REFERENCES

[1] M. Grossglauser and D. Tse, "Mobility increases the capacity of wireless networks," in *Proc. IEEE INFOCOM*, Anchorage, Alaska, 2001, pp. 1360–1369.

[2] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. on Info Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[3] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," IEEE IT-ToN, submitted.

[4] M. Neely and E. Modiano, "Capacity and delay tradeoffs in mobile ad hoc networks," in *IEEE BroadNets*, San Jose, CA, 2004.

[5] X. Lin and N. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Mediterranean Ad Hoc Networking Workshop*, Bodrum, Turkey, 2004.

[6] S. Toumpis and A. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Tran. Wireless Comm.*, vol. 2, no. 4, pp. 736–748, 2003.

[7] D. Bertsimas and G. van Ryzin, "A stochastic and dynamic vehicle routing problem in the euclidean place," *Oper Res*, vol. 39, no. 4, pp. 601–615, 1991.

[8] ——, "Stochastic and dynamic vehicle routing in the euclidean plane: the multiple-server, capacitated vehicle case," *Oper Res*, vol. 41, pp. 60–76, 1993.

[9] ——, "Stochastic and dynamic vehicle routing with general demand and interarrival time distributions," *Adv App Prob*, vol. 20, no. 4, pp. 947–978, 1993.

[10] E. Frazzoli and F. Bullo, "Decentralized algorithms for vehicle routing in a stochastic time-varying environment," in *IEEE Conf. on Decision and Control*, Paradise Island, Bahamas, 2004, pp. 3357–3363.

[11] E. Feuerstein and L. Stougie, "On-line single-server dial-a-ride problems," *Theo Comp Sci*, vol. 268, pp. 91–105, 2001.

[12] M. Lippman, X. Lu, W. de Paepe, R. Sitters, and L. Stougie, "On-line dial-a-ride problems under a restricted information model," *Algorithmica*, vol. 40, pp. 319–329, 2004.

[13] M. Swihart and J. Papastavrou, "A stochastic and dynamic model for the single-vehicle pick-up and delivery problem," *Eur J Oper Res*, vol. 114, pp. 447–464, 1999.

[14] J. Beardwood, J. Halton, and J. Hammersley, "The shortest path through many points," *Proc Cambridge Phil Soc*, vol. 55, pp. 299–327, 1959.

[15] H. Waisanen, D. Shah, and M. A. Dahleh, "Effects of controlled mobility on throughput and delay," preprint.