Noname manuscript No.
(will be inserted by the editor)

Controlled Mobility in Stochastic and Dynamic Wireless Networks

Güner D. Çelik · **Eytan H. Modiano**

Invited Paper

Abstract We consider the use of controlled mobility in wireless networks where messages arriving randomly in time and space are collected by mobile receivers (collectors). The collectors are responsible for receiving these messages via wireless communication by dynamically adjusting their position in the network. Our goal is to utilize a combination of *wireless transmission* and *controlled mobility* to improve the throughput and delay performance in such networks. In the first part of the paper we consider a system with a single collector. We show that the necessary and sufficient stability condition for such a system is given by $\rho < 1$ where ρ is the average system load. We derive lower bounds for the average message waiting time in the system and develop policies that are stable for all loads $\rho < 1$ and have asymptotically optimal delay scaling. We show that the combination of mobility and wireless transmission results in a delay scaling of $\Theta(\frac{1}{1-\rho})$ with the system load ρ in contrast to the $\Theta(\frac{1}{(1-\rho)^2})$ delay scaling in the corresponding system where the collector visits each message location. In the second part of the paper we consider the system with multiple collectors. In the case where simultaneous transmissions to different collectors do not interfere with each other, we show that the stability condition is given by $\rho < 1$, where ρ is the system load on multiple collectors. We develop lower bounds on delay and generalize policies established for the single collector case to multiple collectors case. We show that the delay scaling of $\Theta(\frac{1}{1-\rho})$ extends to the case of multiple collectors, in contrast to the $\Theta(\frac{1}{(1-\rho)^2})$ delay scaling in the corresponding multi-collector system without wireless transmission. We also consider the case where simultaneous transmissions to different collectors interfere with each other. We characterize the stability region of the system in terms of interference constraints. We show that a frame-based version of the well-known Max-Weight policy is throughput-optimal asymptotically in the frame length and derive an upper bound on average delay under this policy.

Keywords Spatial Queueing Models · Controlled Mobility in Wireless Networks · Dynamic Vehicle Routing · DTRP · Delay Tolerant Networks

Güner D. Çelik
Massachusetts Institute Of Technology
77 Massachusetts Avenue, 31-011, Cambridge, MA, 02139
Tel.: +1 (617) 230-1558
E-mail: gcelik@mit.edu

Eytan H. Modiano
Massachusetts Institute Of Technology
77 Massachusetts Avenue, 33-412, Cambridge, MA, 02139
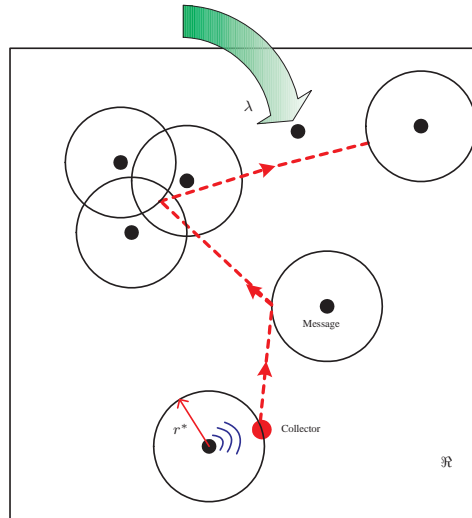Tel.: +1 (617) 452-3414
E-mail: modiano@mit.edu

Fig. 1: The system model for the case of a single collector. The collector adjusts its position in order to receive randomly arriving messages via wireless communication. The circles with radius r^* represent the communication range and the dashed lines represent the collector's path.

1 Introduction

There has been a significant amount of interest in performance analysis of mobility assisted wireless networks in the last decade (e.g., [21], [23], [33], [32], [41], [43] [44], [49], [50], [52]). Typically, throughput and delay performance of networks were analyzed where nodes moving according to a random mobility model were utilized for relaying data (e.g., [20, 21, 23, 32, 39]). More recently, networks deploying nodes with controlled mobility have been considered focusing primarily on route design and ignoring the communication aspect of the problem (e.g., [13], [21], [33], [44], [49], [26], [50], [53]). In this paper we explore the use of *controlled mobility* and *wireless transmission* in order to improve the throughput and delay performance of wireless networks. We consider a dynamic vehicle routing problem where a vehicle (collector) uses a combination of physical movement and wireless reception to receive randomly arriving data messages.

Our model consists of collectors that are responsible for gathering messages that arrive randomly in time at uniformly distributed geographical locations. The messages are transmitted when a collector is within their communication distance and depart the system upon successful transmission. Collectors adjusts their positions in order to successfully receive these messages in the least amount of time as shown in Fig. 1 for the case of one collector. This setup is particularly applicable to networks deployed in a large area so that mobile elements are necessary to provide connectivity between spatially separated entities in the network [13], [26], [33], [52]. For instance, this model is applicable to a densely deployed sensor network where mobile base stations collect data from a large number of sensors densely deployed inside the network, [27], [33], [44], [52], [53]. Another application is utilizing Unmanned Aerial Vehicles (UAVs) as data harvesting devices or as communication relays on a battlefield environment [18], [41], [26], [53]. This model also applies to networks in which data rate is relatively low so that data transmission time is comparable to the collector's travel time, for instance in underwater sensor networks [1], [45].

Vehicle Routing Problems (VRPs) have been extensively studied in the past (e.g., [2], [6], [8], [9], [10], [17], [18], [34], [49], [50], [51]). The common example of a VRP is the Euclidean Traveling

---

A preliminary version of this paper was presented in IEEE CDC'10, Dec. 2001 [14].

Salesman Problem (TSP) in which a single server is to visit each member of a fixed set of locations on the plane such that the total travel cost is minimized. Several extensions of TSP have been considered in the past such as stochastic demand arrivals and the use of multiple servers [2], [8], [9], [18]. In particular, in the TSP with neighborhoods (TSPN) problem the vehicle is to visit a neighborhood of each demand location [6], [17], [34], which can model a mobile collector receiving messages from a communication distance. A more detailed review of the literature in this field can be found in [9], [34] and [51].

Of particular relevance to us among the VRPs is the Dynamic Traveling Repairman Problem (DTRP) due to Bertsimas and van Ryzin [8], [9], [10]. DTRP is a stochastic and dynamic VRP in which a vehicle is to serve demands that arrive randomly in time and space. Fundamental lower bounds on delay were established and several vehicle routing policies were analyzed for DTRP for a single server in [8], for multiple servers in [9], and for general demand and interarrival time distributions in [10]. Altman and Levy [2] considered a similar problem termed *queuing in space* and and proposed stabilizing algorithms. Later, [49], [50] generalized the DTRP model to analyze Dynamic Pickup and Delivery Problem (DPDP) where fundamental bounds on delay were established. We apply the DTRP model to wireless networks where the demands are data messages to be transmitted to a collector which is capable of wireless communication[1]. In our system the problem has considerably different characteristics since in this case the collector does not have to visit message locations but rather can receive the messages from a distance using wireless communication. The objective in our system is to effectively utilize this combination of wireless transmission and controlled mobility in order to minimize the time average message waiting time.

In a closely related problem where multiple mobile nodes with controlled mobility and communication capability relay the messages of static nodes, [43] derived a lower bound on node travel times. Message sources and destinations are modeled as static nodes in [43] and these nodes have saturated arrivals hence queuing aspects were not considered. In an independent work, [27] considered utilizing mobile wireless servers as data relays on periodic routes and applied various delay relations from Polling models to this setup. A mobile server harvesting data from spatial queues in a wireless network was considered in [41] where the stability region of the system was characterized using a fluid model approximation. In [15] we analyzed a one-collector model similar to the current paper but for which the arriving messages were transmitted to the collector using a random access scheme, creating interference among neighboring transmissions. In this paper, the message transmissions are scheduled, i.e., there is only one transmission in the system at a given time, and the collector decides on the message to be transmitted next. The two systems have considerably different characteristics as will be explained in the following sections.

Another related body of literature lies in the area of utilizing mobile elements that can control their mobility to collect sensor data in Delay Tolerant Networks (DTN) (e.g., [13, 33, 44, 45, 52, 53]). Route selection (e.g., [33], [44], [53]), scheduling or dynamic mobility control (e.g., [13], [45], [52]) algorithms were proposed to maximize network lifetime, to provide connectivity or to minimize delay. More detailed surveys of the related work in the area of utilizing mobility in DTN and Sensor Networks can be found in [45] and [53]. These works focus primarily on mobility and usually consider particular policies for the mobile element. To the best of our knowledge, this is the first attempt to develop fundamental bounds on delay in a system where a collector is to gather data messages randomly arriving in time and space using *wireless communication* and *controlled mobility*.

In the first part of the paper we consider a system with a single collector and extend the results of [8] for the DTRP problem to the communication setting. In particular, we show that $\rho < 1$ is the necessary and sufficient condition for the stability of the system where ρ is the system load. We derive lower bounds on delay and develop algorithms that are asymptotically within a constant factor of the lower bounds. We show that the combination of mobility and wireless transmission results in a delay scaling of $\Theta(1/(1 - \rho))$ in contrast to the $\Theta(1/(1 - \rho)^2)$ delay scaling in the system where

---

[1] In previous works such as [2], [8], [9], the collector needs to be at the message location in order to be able to serve it, therefore, we will refer to the DTRP model as the system without wireless transmission.

the collector visits each message location analyzed in [2], [8]. In the second part of the paper we consider the system with multiple collectors under the assumption that simultaneous transmissions to different collectors do not interfere with each other. We show that the necessary and sufficient stability condition is still given by $\rho < 1$, where ρ is the load on multiple collectors. We develop fundamental lower bounds on delay in the system and generalize the single-collector policies analyzed in the first part to the multiple collectors case. Finally we consider a multiple-collector system under interference constraints on simultaneous transmissions to different collectors. We formulate a scheduling problem and characterize the stability region of the system in terms of interference constraints on simultaneous transmissions. We show that a frame-based version of the seminal Max-Weight scheduling policy can stabilize the system whenever it is stabilizable and we derive an upper bound on average delay under this policy.

This paper is organized as follows. In Section 2 we consider the single collector case. We present the model in Section 2.1, characterize the necessary and sufficient stability condition in Section 2.2, derive the delay lower bound in Section 2.3, and analyze single-collector policies in Section 2.4. In Section 3 and the subsections therein we extend the results for a single collector to systems with multiple collectors whose transmissions do not interfere with each other. In Section 4 we consider the system with interference constraints on simultaneous transmissions to collectors. We first present the model and characterize the stability region, and then analyze the frame based Max-Weight policy in Section 4.1 and propose an upper bound on the delay performance of this policy in Section 4.2.

2 Single Collector

In this section we consider the case of a single collector and develop fundamental insights into the problem. We extend the stability and the delay results in [2] and [8], established for the system where the collector visits each message location, to systems with wireless transmission capability. We show that the combination of mobility and wireless transmission results in a delay scaling of $\Theta(\frac{1}{1-\rho})$ with the system load ρ in contrast to the $\Theta(\frac{1}{(1-\rho)^2})$ delay scaling in the corresponding system without wireless transmission in [2] and [8].

2.1 Model

Consider a square region $\mathcal{R}$ of area A and messages arriving into $\mathcal{R}$ according to a Poisson process (in time) of intensity λ. Upon arrival the messages are distributed independently and uniformly in $\mathcal{R}$ and they are to be gathered by a collector via wireless reception. An arriving message is transmitted to the collector when the collector comes within the *reception distance* of the message location and grants access for the message's transmission. Therefore, there is no interference power from the neighboring nodes during message receptions.

We assume a Disk Model (or communication range model) [16], [24] for determining successful message receptions. Let r^* be the *reception distance* of the collector. Under the disk model, a transmission can be received only if it is within a disk of radius r^* around the collector. Note that the Disk Model is similar to the Signal to Noise Ratio (SNR) packet reception model [16], [23], [24], termed the SNR Model, under which a transmission is successfully decoded at the collector if it's received SNR is above a threshold β. To see this, if P_T is the constant transmit power level of a transmission at distance r away from the collector, due to distance-attenuation, the received power satisfies $P_R = P_T r^{-\alpha}$ [16], [23], [24], where α is the power loss exponent. Therefore, under the SNR Model, a transmission at distance r to the collector is successful if $r \leq r^* \doteq (P_T/(P_N \beta))^{1/\alpha}$, showing the equivalence to the Disk Model. Under the Disk Model, if the location of the next message to be received is within r^*, the collector stops and attempts to receive the message. Otherwise, the collector travels towards the message location until it is within a distance r^* away from the message. Under the

disk model, transmissions are assumed to be at a constant rate taking a fixed amount of time denoted by s.

The collector travels from the current message reception point to the next message reception point at a constant speed v. We assume that at a given time the collector knows the locations and the arrival times of the messages that arrived before this time. The knowledge of the service locations is a standard assumption in vehicle routing literature [2], [6], [8], [17], [18], [29], [34], [49].

Let $N(t)$ denote the total number of messages in the system at time t.

Definition 1 **(Stability [5], [35], [37])** *The system is stable under a given control policy π if*

$$\limsup_{t \to \infty} \mathbb{E}[N(t)] < \infty,$$

namely, the long term expected number of messages in the system is finite. Let $\rho = \lambda s$ denote the load arriving into the system per unit time. For stable systems, ρ denotes the fraction of time the collector spends receiving messages.

Definition 2 **(Stability Region [37], [38], [46])** *The stability region $\mathbf{\Lambda}$ is the set of all loads ρ such that there exists a control algorithm that stabilizes the system.*

A policy is said to be throughput-optimal if it stabilizes the system for all loads strictly inside $\mathbf{\Lambda}$.

We define T_i as the time between the arrival of message i and its successful reception. T_i has three components: $W_{d,i}$, the waiting time due to collector's travel distance from the time message i arrives until it gets served, $W_{s,i}$, the waiting time due to the reception times of messages received from the time message i arrives until it gets served, and s, reception time of the message. The total waiting time of message i is denoted by $W_i = W_{d,i} + W_{s,i}$, hence $W_i = T_i - s$. We let d_i be the collector travel distance from the collector's reception location for the message served prior to message i to collector's reception location for message i. The time average per-message travel distance of the collector, denoted by d, is defined by an expectation in the steady state given by $d = \lim_{i \to \infty} E[d_i]$. The time average delays T, W, W_d and W_s are defined similarly to have $T = W_d + W_s + s$ whenever the limits exist. T^* is defined to be the optimal system time which is given by the policy that minimizes T.

2.2 Stability

In this section we show that $\rho < 1$ is a necessary and sufficient condition for the stability of the system. Note that this condition is also necessary and sufficient for stability of the corresponding system without wireless transmission, as shown in [2], as well as for a G/G/1 queue [28]. Here we prove this result using simpler techniques than [2]. The analysis in this section will be essential for generalizing the stability condition and some delay results to the case of multiple collectors.

2.2.1 Necessary Condition for Stability

We lower bound the number of messages in the system by that in the equivalent system in which travel times are zero (i.e., $v = \infty$). This technique was used in [2] to establish a necessary stability condition for the corresponding system without wireless transmission. Here we give a simpler proof of this fact in Appendix A for completeness.

Theorem 1 *A necessary condition for stability is $\rho < 1$. Furthermore, we have*

$$W \geq \frac{\lambda s^2}{2(1 - \rho)}. \tag{1}$$

The proof in Appendix A first establishes that the steady state time average delay in the system under any policy π is at least as big as the delay of any work-conserving[2] policy in the equivalent system in which travel times are zero (i.e., $v = \infty$). This is based on an induction argument that the total number of messages in the system is always greater than that in the infinite velocity system. This is because the service time per message is greater than that in the infinite speed system. Since the latter system behaves as an M/D/1 queue (a queue with Poisson arrivals, constant service times and 1 server), its average waiting time is given by the Pollaczek-Khinchin (P-K) formula for M/G/1 queues [7, p. 189], given in (1). A direct consequence of this lemma is that a necessary condition for stability in the infinite speed system is also necessary for our system. The necessary and sufficient condition for stability in an M/G/1 queue is given by $\rho < 1$ (see e.g., [7] or [22]).

2.2.2 Sufficient Condition for Stability

Here we prove that $\rho < 1$ is a sufficient condition for stability of the system under a policy based on Euclidean TSP with neighborhoods (TSPN). TSPN is a generalization of TSP in which the server is to visit a neighborhood of each demand location via the shortest path [6], [17], [34]. In our case the neighborhoods are disks of radius r^* around each message location. TSPN is an NP-Hard problem such as TSP. Recently, [34] proved that a Polynomial Time Approximation Scheme (PTAS) exists for TSPN among fat regions in the plane. A region is said to be *fat* if it contains a disk whose size is within a constant factor of the diameter of the region, e.g., a disk, and a PTAS belongs to a family of $(1 + \epsilon)$-approximation algorithms parameterized by $\epsilon > 0$.

Under the TSPN policy, the collector performs a cyclic service of the messages present in the system starting and ending the cycle at the center of the network region. Let time t_k be the time that the collector returns to the center for the kth time, where $t_0 \doteq 0$. Assume the system is initially empty at time t_0. The TSPN Policy is described in detail in Algorithm 1.

Algorithm 1 *TSPN Policy*

1: Initially at $t = t_0$, the collector waits at the center of $\mathcal{R}$ until the first message arrival, moves to serve this message and returns to the center.
2: If the system is empty at time $t_k, k = 1, 2, ...$, the collector repeats the above process.
3: If there are messages waiting for service at time $t_k, k = 1, 2, ...$, the collector computes the TSPN tour (e.g., using the PTAS in [34]) through all the messages that are present in the system at time t_k, receives these messages in that tour and returns to the center.

Let the total number of messages waiting for service at time t_i, $N(t_i)$, be the system state at time t_i. Note that $N(t_i)$ is an irreducible Markov chain on countable state space **N**. We show the stability of the TSPN policy through the ergodicity of this Markov chain.

Theorem 2 *The system is stable under the TSPN policy for all loads $\rho < 1$.*

Proof Given the system state $N(t_i)$ at time t_i, we apply the algorithm in [34] to find a TSPN tour of length L_i through the $N(t_i)$ neighborhoods that is at most $(1 + \epsilon)$ away from the optimal TSPN tour length L_i^*. Note that L_i^* can be upper bounded by a constant L for all $N(t_i)$. This is because the collector does not have to move for messages within its communication range and a finite number of such disks of radius r^* can cover the network region for any $r^* > 0$. The collector then can serve the messages in each disk from its center incurring a tour of constant length L (an example of such a tour is shown in Fig. 2). We will use the Foster-Lyapunov criterion to show that the Markov chain described by the states $N(t_i)$ is positive recurrent [5]. We use $V(N_i) = sN(t_i)$, the total load served

---

[2] A work-conserving policy is such that the server does not idle when the queue is not empty.

during i^{th} cycle, as the Lyapunov function. Note that $V(0) = 0$, $S_k = \{x : V(x) \leq K\}$ is a bounded set for all finite K and $V(.)$ is a non-decreasing function. Since the arrival process is Poisson, the expected number of arrivals during a cycle can be upper-bounded as follows:

$$\mathbb{E}[N(t_{i+1})|N(t_i)] \leq \lambda(L/v + sN(t_i)). \tag{2}$$

Hence we obtain the following drift expression for the load during a cycle.

$$\mathbb{E}[sN(t_{i+1}) - sN(t_i)|N(t_i)] \leq \rho L/v - (1-\rho)sN(t_i). \tag{3}$$

Since $\rho < 1$, there exist a $\delta > 0$ such that $\rho + \delta < 1$:

$$\mathbb{E}[sN(t_{i+1}) - sN(t_i)|N(t_i)] \leq \rho L/v - \delta s N_i$$
$$\leq -\delta s + \frac{\rho L}{v}.\mathbf{1}_{\{N(t_i)\in S\}}, \tag{4}$$

where $\mathbf{1}_{\{N\in S\}}$ is equal to 1 if $N \in S$ and zero otherwise and $S = \{N \in \mathbf{N} : N \leq K\}$ is a bounded set with $K = \lceil \frac{\rho L}{v\delta s} + 1 \rceil$. Hence the drift is negative as long as $N(t_i)$ is outside a bounded set. Therefore, by the standard Foster-Lyapunov criterion [3], [5], the Markov chain $(N(t_i))$ is positive recurrent and it has a unique stationary distribution [5]. Furthermore, using (3) we can bound the steady state average $N(t_i)$ as [35]

$$\limsup_{t_i \to \infty} \mathbb{E}[N(t_i)] \leq \frac{\lambda L}{v(1-\rho)}.$$

Moreover, given some $t \in [t_i, t_{i+1}]$, we have

$$\limsup_{t \to \infty} \mathbb{E}[N(t)] \leq \limsup_{t_i \to \infty} \mathbb{E}[N(t_i) + N(t_{i+1})]$$
$$\leq 2\frac{\lambda L}{v(1-\rho)} < \infty. \tag{5}$$

This establishes the stability of the TSPN policy for any load $\rho < 1$. $\qquad\square$

This establishes that $\rho < 1$ is a necessary and sufficient condition for stability. The travel time of the collector does not affect the stability region of the system. In other words, we have the same stability region as a G/G/1 queue. The intuition behind this result is that as the system load is increased, under stable policies, the fraction of time spent on travel goes to zero. This observation is also true for the corresponding system without wireless transmission since $\rho < 1$ is also necessary and sufficient for stability of such systems [2].

The communication capability does not enlarge the stability region, however, it fundamentally affects the delay scaling in the system. The delay scaling of the TSPN policy with load ρ is $O(\frac{1}{1-\rho})$ as shown in (5), the same delay scaling as in a G/G/1 queue. This is a fundamental improvement in delay due to the wireless transmission capability as the delay scaling for the corresponding system without wireless transmission is $\Theta(\frac{1}{(1-\rho)^2})$ [8]. This delay scaling can be easily obtained as follows. The system without wireless transmission corresponds to having $r^* = 0$ in our system. In this case, one utilizes a $(1 + \epsilon)$ PTAS for the optimal TSP tour through the message locations instead of the TSPN tour. An upper bound on the TSP tour for any $N(t_i)$ points arbitrarily distributed in a square of area A is given by $\sqrt{2AN(t_i)} + 1.75\sqrt{A}$ [29]. Similar arguments as in the proof of Theorem 2 leads to the drift condition

$$\mathbb{E}[sN(t_{i+1}) - sN(t_i)|N(t_i)] \leq \rho(\kappa_1\sqrt{AN(t_i)} + \kappa_2) - (1-\rho)sN(t_i), \tag{6}$$

for some constants κ_1 and κ_2, where the drift is again negative as long as $N(t_i)$ is outside a bounded set S. The difference in this case is that the travel time per cycle scales with the number of messages $N(t_i)$ as $\sqrt{N(t_i)}$ and using (6) we can show that the delay scaling with the load ρ is $O(\frac{1}{(1-\rho)^2})$.

2.3 Lower Bound On Delay

For wireless networks with a small area and/or very good channel quality such that $r^* \geq \sqrt{A/2}$, the collector does not need to move as every message will be in its reception range if it just stays at the center of the network region. In that case the system can be modeled as an M/D/1 queue with service time s and the associated queuing delay is given by the P-K formula for M/G/1 queues, i.e., $W = \lambda s^2/(2(1-\rho))$. However, when $r^* < \sqrt{A/2}$, the collector has to move in order to receive some of the messages. In this case the reception time s is still a constant, however, the travel time per message is now a random variable which is not independent over messages (for example, observing small travel times for the previous messages implies a dense network, and hence the future travel times per message are also expected to be small). Next we provide a lower bound similar to a lower bound in [8] with the added complexity of communication capability in our system.

Theorem 3 *The optimal steady state time average delay T^* is lower bounded by*

$$T^* \geq \frac{\mathbb{E}[(||U|| - r^*)^+]}{v(1-\rho)} + \frac{\lambda s^2}{2(1-\rho)} + s. \tag{7}$$

where $(||U|| - r^*)^+$ represents $\max(0, ||U|| - r^*)$, U is a uniformly distributed random variable over the network region $\mathcal{R}$, and $\rho = \lambda s$ is the system load.

Note that the $\mathbb{E}[(||U|| - r^*)^+]$ term can be further lower bounded by $\mathbb{E}[||U||] - r^*$, where $\mathbb{E}[||U||] = 0.383\sqrt{A}$ [8].

Proof As outlined in Section 2.1, the delay of message i, T_i has three components: $T_i = W_{d,i} + W_{s,i} + s_i$. Taking expectations and the limit as $i \to \infty$ yields

$$T = W_d + W_s + s. \tag{8}$$

A lower bound on W_d is found as follows: Note that $W_{d,i}.v$ is the average *distance* the collector moves during the waiting time of message i. This distance is at least as large as the average distance between the location of message i and the collector's location at the time of message i's arrival less the reception distance r^*. The location of an arrival is determined according to the uniform distribution over the network region, while the collector's location distribution is in general unknown as it depends on the collector's policy. We can lower bound W_d by characterizing the expected distance between a uniform arrival and the best a priori location in the network that minimizes the expected distance to a uniform arrival. Namely we are after the location ν that minimizes $\mathbb{E}[||U - \nu||]$ where U is a uniformly distributed random variable. The location ν that solves this optimization is called the *median* of the region and in our case the median is the center of the square shaped network region. Because the travel distance is nonnegative, we obtain the following bound on W_d:

$$W_d \geq \frac{\mathbb{E}[(||U|| - r^*)^+]}{v}. \tag{9}$$

Let N be the average number of messages received in a waiting time and let R be the average residual reception (service) time. Due to the PASTA property of Poisson arrivals (Poisson Arrivals See Time Averages) (see for example [7, p. 171]) a given arrival in steady state observes the time average steady state occupancy distribution. Therefore, the average residual time observed by an arrival is also R and is given by $\lambda s^2/2$ [7, p. 188] and we have

$$W_s = sN + R. \tag{10}$$

Since in a stable system in steady state the average number of messages received in a waiting time is equal to the average number of arrivals in a waiting time (a variation of Little's law [8], [49]) we have $N = \lambda W = \lambda(W_d + W_s)$. Substituting this in (10) we obtain

$$W_s = s\lambda(W_d + W_s) + \frac{\lambda s^2}{2}.$$

This implies

$$W_s = \frac{\rho}{1-\rho}W_d + \frac{\lambda s^2}{2(1-\rho)}, \tag{11}$$

Substituting (9) and (11) in (8) yields (7). $\hspace{1cm} \square$

In addition to the average waiting time of a classical M/G/1 queue given in (1), the queueing delay also increases due to the collector's travel.

2.4 Collector Policies

We derive upper bounds on delay by analyzing policies for the collector. The TSPN policy analyzed in the previous section is stable for all loads $\rho < 1$ and has $O(\frac{1}{1-\rho})$ delay scaling. Since the lower bound in Section 2.3 also scales with the load as $\frac{1}{1-\rho}$, *the TSPN policy has optimal delay scaling*. In the following we consider the First Come First Serve (FCFS) and the Partitioning policies that can have better delay properties than the TSPN policy. In particular, the FCFS policy is delay-optimal at light loads and the Partitioning policy has delay performance that is very close to the lower bound when the travel and reception times are comparable.

2.4.1 First Come First Serve (FCFS) Policy

A straightforward policy is the FCFS policy where the messages are served in the order of their arrival times. A version of the FCFS policy, call FCFS', where the receiver has to return to the center of the network region (the median of the region for general network regions) after each message reception was shown to be optimal at light loads for the DTRP problem [8]. This is because the center of the network region is the location that minimizes the expected distance to a uniformly distributed arrival. Since in our system we can do at least as good as the DTRP by setting $r^* = 0$, FCFS' is optimal also for our system at light loads. Furthermore, the FCFS policy is not stable for all loads $\rho < 1$, namely, there exists a value $\hat{\rho}$ such that the system is unstable under FCFS policy for all $\rho > \hat{\rho}$. This is because under the FCFS policy the average travel component of the service time is fixed, which makes the average arrival rate greater than the average service rate as $\rho \to 1$. Therefore, it is better for a policy to serve more messages in the same "neighborhood" in order to reduce the amount of time spent on mobility.

2.4.2 Partitioning Policy

Next we propose a policy based on partitioning the network region into subregions and the collector performing a cyclic service of the subregions. This policy is an adaptation of the Partitioning policy of [8] to the case of a system with wireless transmission. We explicitly derive the delay expression for this policy and show that it scales with the load as $O(\frac{1}{1-\rho})$ as in the TSPN policy.

We divide the network region into ($\sqrt{2}r^*$ x $\sqrt{2}r^*$) squares as shown in Fig. 2. This choice ensures us that every location in the square is within the communication distance r^* of the center of the square. The number of subregions in such a partitioning is given by[3] $n_s = A/(2(r^*)^2)$. The partitioning in Fig. 2 represents the case of $n_s = 16$ subregions. The collector services the subregions in a cyclic order as displayed in Fig. 2 by receiving the messages in each subregion from its center using an FCFS order. The messages within each subregion are served exhaustively, i.e., all the messages in a subregion are received before moving to the next subregion. The collector then receives the messages in the next subregion exhaustively using FCFS order and repeats this process. The distance traveled

---

[3] Note that such a partitioning requires $\sqrt{n_s} = \sqrt{A/(2(r^*)^2)}$ to be an integer. This may not hold for a given area A and a particular choice of r^*. In that case one can partition the region using the largest reception distance $\underline{r}^* < r^*$ such that this integer condition is satisfied.
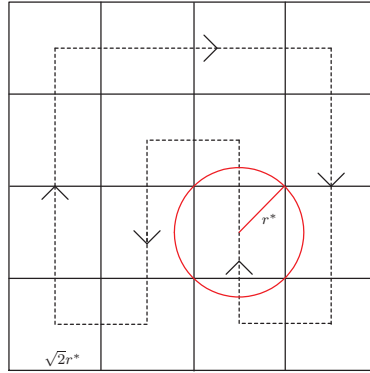
Fig. 2: The partitioning of the network region into square subregions of side $\sqrt{2}r^*$. The circle with radius r^* represents the communication range and the dashed lines represent the collector's path.

by the collector between each subregion is a constant equal to $\sqrt{2}r^*$. It is easy to verify that the Partitioning policy behaves as a multiuser M/G/1 system with reservations (see [7, p. 198]) where the n_s subregions correspond to users and the travel time between the subregions corresponds to the reservation interval. Using the delay expression for multiuser M/G/1 queue with reservations in [7, p. 200] we obtain,

$$T_{part} = \frac{\lambda s^2}{2(1-\rho)} + \frac{n_s - \rho}{2v(1-\rho)}\sqrt{2}r^* + s, \tag{12}$$

where $\rho = \lambda s$ is the system load. Combining this result with (7) and noting that the above expression is finite for all loads $\rho < 1$, we have established the following observation.

Observation 1 *The time average delay in the system scales as $\Theta(\frac{1}{1-\rho})$ with the load ρ and the Partitioning policy is stable for all $\rho < 1$.*

Despite the travel component of the service time, we can achieve $\Theta(\frac{1}{1-\rho})$ delay as in classical queuing systems (e.q., G/G/1 queue). This is the fundamental difference between this system and the corresponding system where wireless transmission is not used, as in the latter system the delay scaling with load is $\Theta(\frac{1}{(1-\rho)^2})$ [8]. This difference can be explained intuitively as follows. Denote by N the average number of departures in a waiting time. It is easy to see from the P-K formula that in a classical M/G/1 queue, N scales with the load as $\Theta(\frac{1}{1-\rho})$. We argue that this scaling for N is preserved in our system but not in [8]. The W_s expression as a function of W_d in (11) implies that for any given policy with its corresponding W_d, N can be lower bounded by $\frac{\lambda W_d}{1-\rho}$. For the system in [8], the minimum per-message distance the collector moves in the high load regime scales as $\Omega(\frac{\sqrt{A}}{\sqrt{N}})$ [8]. Intuitively, this is due to the observation that the nearest neighbor distance among N uniformly distributed points on a square region of area A scales as $\frac{\sqrt{A}}{\sqrt{N}}$. Therefore, for this system we have $W_d \approx N\Omega(\frac{\sqrt{A}}{\sqrt{N}}) \approx \Omega(\sqrt{NA})$ which gives $N \approx \Omega(\frac{\lambda^2 A}{(1-\rho)^2})$. Namely, W_d increases with the load and this results in an extra $1/(1-\rho)$ scaling in delay in addition to the $1/(1-\rho)$ factor of classical G/G/1 queues. However, with the wireless reception capability, the collector does not need to move for messages that are inside a disk of radius r^* around it. Since a finite (constant) number of such disks cover the network region, W_d can be upper bounded by a constant independent of the system load, for example, for the Partitioning policy an easy upper bound on W_d is the length of one cyclic tour around the network. Therefore, in our system N scales as $1/(1-\rho)$ as in classical queueing systems.

It is interesting to note that [15] considered the case where messages were transmitted to the collector according to a random access scheme, i.e., transmissions occur with probability p in each time slot. There the delay scaling of $\Omega(\frac{1}{(1-\rho)^2})$ with load ρ was obtained as in the system without
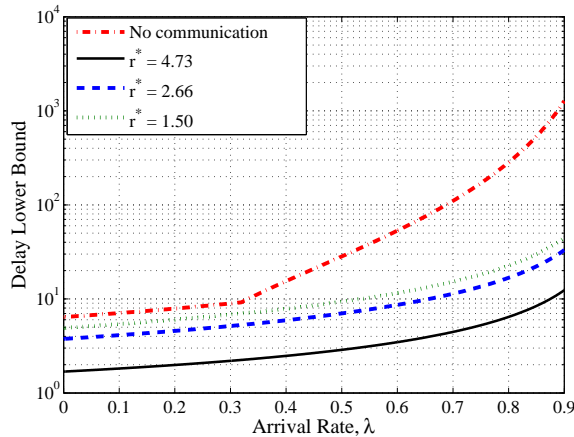
Fig. 3: Delay lower bound vs. network load using different communication ranges for $A = 200$, $\beta = 2$, $\alpha = 4$, $v = 1$ and $s = 1$.

wireless transmission. The reason for this is that in order to have successful transmissions under the random access interference of neighboring nodes, the reception distance should be of the same order as the nearest neighbor distances [15], [23].

2.4.3 Numerical Results-Single Collector

Here we present numerical results corresponding to the analysis in the previous sections. We lower bound the delay expression in (7) using $\mathbb{E}[(\|U\| - r^*)^+] \geq \mathbb{E}[\|U\|] - r^*$, where $[\|U\|] = 0.383\sqrt{A}$ is the expected distance of a uniform arrival to the center of square region of area A [8]. Fig. 3 shows the delay lower bound as a function of the network load for increased values of the communication range r^* [4]. As the communication range increases, the message delay decreases as expected. For heavy loads, the delay in the system is significantly less than the delay in the corresponding system without wireless transmission in [8], demonstrating the difference in the delay scaling between the two systems. For light loads and small communication ranges, the delay performance of the wireless network tends to the delay performance of [8].

Fig. 4 compares the delay in the Partitioning Policy to the delay lower bound for two different cases. When the travel time dominates the reception time, the delay in the Partitioning policy is about 10.6 times the delay lower bound. For a more balanced case, i.e., when the reception time is comparable to the travel time, the delay ratio drops to 2.4.

3 Multiple collectors - Interference-Free Networks

In this section we extend our analysis to a wireless network with multiple identical collectors that do not interfere with each other. An arriving message is transmitted when one of the m collectors comes within the reception distance of the message location and grants access for the message's transmission. Therefore, at a given time there can be at most m transmissions in the network. We consider policies that partition the network region into m subregions. Each collector is assigned to one of the subregions and is allowed to operate only in its own subregion. We call this class of policies the *network partitioning policies*. In such a case, there is no interference from nodes within the subregion

---

[4] For the delay plot of the no-communication system, the point that is not smooth arises since the plot is the maximum of two delay lower bounds proposed in [8].
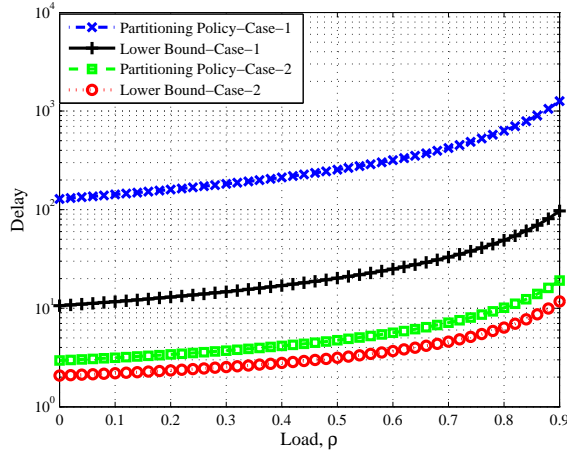
Fig. 4: Delay in the Partitioning policy vs the delay lower bound for $r^* = 2.2$, $\beta = 2$ and $\alpha = 4$. Case-1: Dominant travel time ($A = 800$, $v = 1$, $s = 2$). Case-2: Comparable travel and reception times ($A = 60$, $v = 10$, $s = 2$).

where the transmission is taking place. The only source of interference can be due to transmissions in other subregions.

The interference-free assumption holds for example if the signaling schemes used in different subregions are orthogonal to each other. This can be achieved for example by having different frequency bands for transmissions in different subregions. Furthermore, for networks deployed on a large area, even if orthogonal signaling is not utilized, the interference between subregions can be negligible due to signal attenuation with distance.

Note that the interference-free assumption is consistent for the purposes of a lower bound on delay since the interference from neighboring nodes can only increase the message reception time. Similar to the previous section, we assume a *communication range* r^* for each collector and each reception takes time s.

3.1 Stability

Here we show that $\rho = \lambda s/m < 1$ is a necessary and sufficient condition for stability of the system.

3.1.1 Necessary Condition for Stability

A necessary condition for stability of the multi-collector system is given by $\rho = \lambda s/m < 1$. We prove this by showing that the system stochastically dominates the corresponding system with zero travel times (i.e., an M/D/m queue, a queue with Poisson arrivals, constant service time and m servers) similar to Section 2.2.

Theorem 4 *A necessary condition for the stability of any policy is* $\rho = \lambda s/m < 1$. *Furthermore, the optimal steady state time average delay* T_m^* *is lower bounded by*

$$T_m^* \geq \frac{\lambda s^2}{2m^2(1-\rho)} - \frac{m-1}{m}\frac{s^2}{2s} + s, \tag{13}$$

where $\rho = \lambda s/m$ *is the system load.*

The proof is similar to the proof of Theorem 1 and is given in Appendix-B. It makes use of the fact that the steady state time average delay in the system is at least as big as the delay in the equivalent system in which travel times are considered to be zero (i.e., $v = \infty$).

3.1.2 Sufficient Condition for Stability

Next we establish that $\rho < 1$ is also sufficient for stability. This can be seen by dividing the network region into m identical square subregions, and performing a single-collector TSPN policy in each sub-region[5]. Since the arrival process is Poisson, each subregion receives an independent Poisson arrival process of intensity λ/m. Furthermore, each collector performs a TSPN policy independently of the other collectors. Therefore, using the stability result of the single-collector TSPN policy, the systems in each subregion are stable if $\rho < 1$. We state this fact in the following theorem:

Theorem 5 *The system is stable under the multi-collector TSPN policy for all loads $\rho = \lambda s/m < 1$.*

Note that a similar delay analysis to the single-collector TSPN case shows that the multi-collector TSPN policy has $O(\frac{1}{1-\rho})$ delay scaling with the load ρ.

3.2 Delay Lower Bound

The delay lower bound in (13) neglects the travel component of the delay. Therefore, we provide another lower bound for the optimal delay T_m^* and take their simple average. The following theorem states the second lower bound on delay. It is based on the convexity argument that when the travel component of the waiting time is lower bounded by a constant, the equal area partitioning of the network region minimizes the resulting delay expression over all area partitionings.

Theorem 6 *For the class of network partitioning policies, the optimal steady state time average delay T_m^* is lower bounded by*

$$T_m^* \geq \frac{\max\left(0, \frac{2}{3}\sqrt{\frac{A}{m\pi}} - r^*\right)}{v(1-\rho)} + s, \tag{14}$$

where $\rho = \lambda s/m$ is the system load.

Proof Here we use an approach similar to the proof of Theorem 3. We divide the average delay T into three components:

$$T = W_d + W_s + s. \tag{15}$$

The lemma below provides a bound for W_d, the average message waiting time due to the collectors's travel, using a result from [25] for the m-median problem.

Lemma 1

$$W_d \geq \frac{\max\left(0, \frac{2}{3}\sqrt{\frac{A}{m\pi}} - r^*\right)}{v}. \tag{16}$$

Proof Let Ω be *any* set of points in $\Re$ with $|\Omega| = m$. Let U be a uniformly distributed location in $\Re$ independent of Ω and define $Z^* \triangleq \min_{\nu \in \Omega} \| U - \nu \|$. Let the random variable Y be the distance from the center of a disk of area A/m to a uniformly distributed point within the disk. Then it is shown in [25] that

$$\mathbb{E}[f(Z^*)] \geq \mathbb{E}[f(Y)] \tag{17}$$

---

[5] For simplicity we assume that it is possible to divide the region into m identical square subregions.

for any nondecreasing function $f(.)$. Using this result we obtain $\mathbb{E}[\max(0, Z^* - r^*)] \geq \mathbb{E}[\max(0, Y - r^*)]$. Note that W_d can be lower bounded by the expected distance of a uniform arrival to the closest collector at the time of arrival less r^*. Because the travel distance is nonnegative, we have

$$W_d \geq \mathbb{E}[\max(0, Y - r^*)]/v \geq \max(0, \mathbb{E}[Y] - r^*)/v,$$

where the second bound is due to Jensen's inequality. Substituting $\mathbb{E}[Y] = \frac{2}{3}\sqrt{\frac{A}{m\pi}}$ into the above expression completes the proof. $\square$

Intuitively the best a priori placement of m points in $\Re$ in order to minimize the distance of a uniformly distributed point in the region to the closest of these points is to cover the region with m disjoint disks of area A/m and place the points at the centers of the disks. Such a partitioning of the region is not possible, however, using this idea we can lower bound the expected distance as in (16).

We now derive a lower bound on W_s. Let $R^1, R^2, ..., R^m$ be the network partitioning with areas $A^1, A^2, ..., A^m$ respectively ($\sum_{j=1}^{m} A^j = A$). Consider the message receptions in steady state that are received by collector j eventually. Let λ^j be the fraction of the arrival rate served by collector j. Due to the uniform distribution of the message locations we have

$$\frac{\lambda^j}{\lambda} = \frac{A^j}{A}.$$

Let N^j be the average number of message receptions for which the messages that are served by collector j waits in steady state. Similarly let W_s^j and W_d^j be the average waiting times for messages served by collector j due to the time spent on message receptions and collector j's travel respectively. Using (10) and lower bounding the residual time by zero we have

$$W_s^j \geq sN^j.$$

Using Little's law ($N^j = \lambda^j(W_s^j + W_d^j)$) in a similar way as for (11) we have

$$W_s^j \geq \frac{\lambda^j s}{1 - \lambda^j s} W_d^j. \tag{18}$$

The fraction of messages served by collector j is A^j/A. Therefore, we can write W_s as

$$W_s = \sum_{j=1}^{m} \frac{A^j}{A} W_s^j \geq \sum_{j=1}^{m} \frac{A^j}{A} \frac{\lambda^j s}{1 - \lambda^j s} W_d^j. \tag{19}$$

For a given region R^j with area A^j, W_d^j is lower bounded by (similar to the derivation of (9)) the distance of a uniform arrival to the median of the region less r^*.

$$W_d^j \geq \frac{\mathbb{E}[\max(0, ||U - \nu|| - r^*)]}{v} \geq \frac{\max(0, \mathbb{E}[||U - \nu||] - r^*)}{v}, \tag{20}$$

where ν is the median of R^j and $||U - \nu||$ is the distance of U, a uniformly distributed location inside R^j, to ν. The inequality in (20) is due to Jensen's inequality for convex functions. A disk shaped region yields the minimum expected distance of a uniform arrival to the median of the region [25]. Using this we further lower bound W_d by noting that for a disk shaped region of area A_j, $\mathbb{E}[||U - \nu||]$ is just the expected distance of a uniform arrival to the center of the disk given by $\frac{2}{3}\sqrt{\frac{A^j}{\pi}}$. Hence

$$W_d^j \geq \frac{\max(0, \frac{2}{3}\sqrt{\frac{A^j}{\pi}} - r^*)}{v} = \frac{\max(0, c_1\sqrt{A^j} - r^*)}{v}, \tag{21}$$

where $c_1 = \frac{2}{3\sqrt{\pi}} = 0.376$. Letting $f(A^j) = \frac{\lambda \frac{A^j}{A} s}{1 - \lambda \frac{A^j}{A} s}$, which is a convex and increasing function of A^j, we rewrite (19) as

$$W_s \geq \sum_{j=1}^{m} \frac{f(A^j)}{vA} A^j \max(0, c_1 \sqrt{A^j} - r^*).$$ (22)

Next we show that the function $f(A^j) A^j \max(0, c_1 \sqrt{A^j} - r^*)$ is a convex function of A^j via the two lemmas below.

Lemma 2 *Let $f(.)$ and $g(.)$ be two convex and increasing functions. The function $h(.) = f(.)g(.)$ is also convex and increasing.*

Proof See Appendix C.   □

Lemma 3 $h(x) = x \max(0, c_1 \sqrt{x} - c_2)$ *with domain $[0, \infty)$ is a convex and increasing function of x.*

Proof See Appendix D.   □

Letting $g(A^j) \doteq f(A^j) A^j \max(0, c_1 \sqrt{A^j} - r^*)$, we have from the lemmas 2 and 3 that the function $g(A^j)$ is convex. Now rewriting (22) we have

$$W_s \geq \left(\frac{m}{vA}\right) \frac{1}{m} \sum_{j=1}^{m} g(A^j).$$

Using the convexity of the function $g(A^j)$ we have

$$\begin{aligned} W_s &\geq \left(\frac{m}{vA}\right) g\left(\frac{\sum_{j=1}^{m} A^j}{m}\right) = \frac{m}{vA} g\left(\frac{A}{m}\right) \\ &= \frac{\frac{\lambda s}{m}}{1 - \frac{\lambda s}{m}} \frac{\max\left(0, c_1 \sqrt{\frac{A}{m}} - r^*\right)}{v} \\ &= \frac{\rho}{1 - \rho} \frac{\max\left(0, c_1 \sqrt{\frac{A}{m}} - r^*\right)}{v}. \end{aligned}$$ (23)

The above analysis essentially implies that the W_s expression in (22) is minimized by the *equitable partitioning* of the network region. Finally combining (15), (16) and (23) we obtain (14).   □

Finally, taking the simple average of (13) and (14) we arrive at the following theorem.

Theorem 7 *For the class of network partitioning policies, the optimal steady state time average delay T_m^* is lower bounded by*

$$T_m^* \geq \frac{\lambda s^2}{4m^2(1-\rho)} + \frac{max\left(0, \frac{2}{3}\sqrt{\frac{A}{m\pi}} - r^*\right)}{2v(1-\rho)} - \frac{m-1}{m} \frac{s^2}{4s} + s,$$ (24)

where $\rho = \lambda s/m$ is the system load.

This expression incorporates both the travel time and the message reception time components of the expected delay in the system. Theorem 7 is valid for the class of network partitioning policies. For the system without wireless transmission, it has been shown that partitioning the region into m equal size disjoint subregions (one for each collector) preserves optimality in the high load limit [9], [51]. We conjecture that this optimality is also preserved in our system.
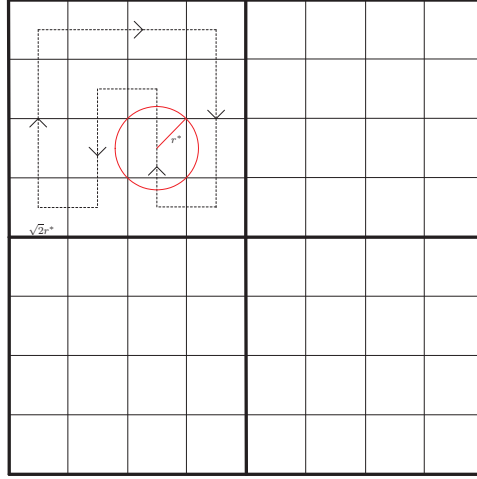
Fig. 5: The partitioning of the network region into square subregions of side $\sqrt{2}r^*$ for the case of multiple collectors in the network. The circle with radius r^* represents the communication range and the dashed lines represent one of the collector's path.

3.3 Multiple Collector Policies

It is easy to show that a generalization of the FCFS policy in which we partition the network region into m subregions and assign each collector to perform the single-collector FCFS Policy in its own subregion has optimal delay at light loads. The reason for this results is similar to the light load optimality of the FCFS Policy for the single collector case. The area partitioning has to be done as follows: We create m Voronoi regions with centers of the regions given by the m-median locations[6] of the network region. This policy is not stable as $\rho \to 1$ due to the same reason as in Section 2.4.

3.3.1 Generalized Partitioning Policy

Next we propose a policy based on dividing the network region into m equal size subregions. For simplicity we assume that it is possible to divide the region into m identical square subregions. Each collector is assigned to one of the subregions and is responsible for receiving messages that arrive into its own subregion using the single collector partitioning policy analyzed in Section 2.4.2. Namely, first the network region is divided into subregions of area A/m and then each subregion is divided into $\sqrt{2}r^*$ x $\sqrt{2}r^*$ squares. The number of $\sqrt{2}r^*$ x $\sqrt{2}r^*$ squares in each subregion is given by[7] $n_s = \frac{A/m}{2(r^*)^2}$. Fig. 5 represents such a partitioning for the case of four collectors in the network with $n_s = 16$ squares in each subregion. Since each subregion behaves identically, the average delay of this policy is the average delay of the single collector Partitioning policy applied to a subregion with arrival rate λ/m, area A/m, and $n_s = \frac{A/m}{2(r^*)^2}$:

$$T_{part} = \frac{\lambda s^2}{2m(1-\rho)} + \frac{\frac{A}{2m(r^*)^2} - \rho}{2v(1-\rho)}\sqrt{2}r^* + s, \qquad (25)$$

where $\rho = \lambda s/m$ is the system load. This result, when combined with (13), establishes that for the case of multiple collectors in the system the delay scaling with the load is $\Theta(\frac{1}{1-\rho})$. This is again a

---

[6] The set of m-median locations for a region is the set of the best m a priori locations in the region that minimizes the expected distance to a uniform arrival.

[7] As for the single collector Partitioning policy, we note that if $\sqrt{n_s} = \sqrt{(A/m)/(2(r^*)^2)}$ is not an integer, one can partition the region using the largest reception distance $\underline{r}^* < r^*$ such that this integer condition is satisfied.
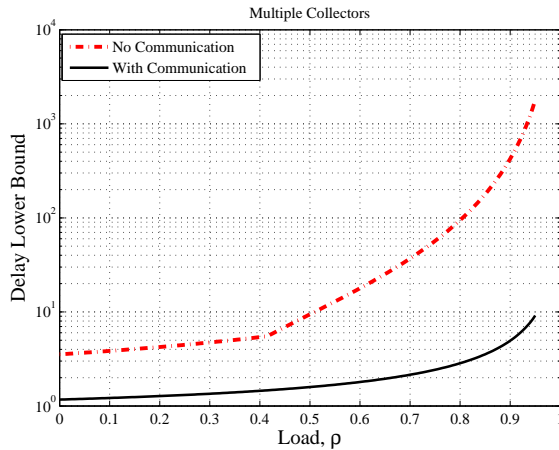
Fig. 6: Delay lower bound vs. network load for m=2 collectors, $r^* = 4.7$, $A = 400$, $\beta = 2$, $\alpha = 4$, $v = 1$ and $s = 1$.
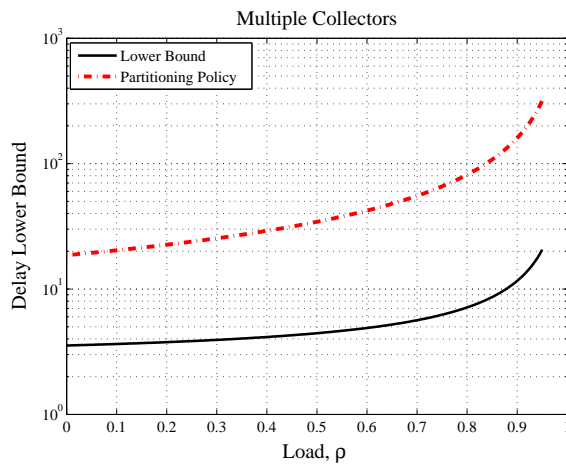

Fig. 7: Delay of the Partitioning policy vs the delay lower bound for $m = 4$ collectors, $r^* = 2.6$, $A = 500$, $\beta = 2$, $\alpha = 4$, $v = 1$ and $s = 2$.

fundamental improvement compared to the $\Theta(\frac{1}{(1-\rho)^2})$ delay scaling in the system without wireless transmission and with multiple collectors in [9].

3.3.2 Numerical Results

We compare the delay lower bound in (24) to the delay lower bound in the corresponding system without wireless transmission in [9] for the case of two collectors in Fig. 6. The delay in the two-collector system is significantly below the delay in the system without wireless transmission and this difference is more pronounced for high loads. Fig. 7 displays the delay lower bound in (24) and the delay of the Partitioning policy in (25) as functions of the network load ρ. The delay of the Partitioning policy is about 7 times the delay lower bound.

4 Multiple Collectors - Systems with Interference Constraints

In this section we consider systems in which simultaneous transmissions to different collectors interfere with each other. Interference between simultaneous transmissions occurs if the collectors do not use orthogonal signaling. At each point in time, the problem is to dynamically determine *message pick up locations* for the collectors and also to efficiently route the collectors to these pick up locations based on the current collector configuration and the number of messages present in different parts of the network region. The objective is to minimize the expected message waiting time in the system. This is a joint scheduling and euclidian vehicle routing problem which has not been considered previously.

Here we obtain preliminary results for this problem by emphasizing the scheduling aspects of the problem through fairly general interference constraints and simplifying the mobility aspect of the problem by discretizing the collectors' motion. We characterize the stability region of the system in terms of interference constraints. We show that a frame-based version of the Max-Weight scheduling policy [11], [46], can stabilize the system whenever it is stabilizable and we derive an upper bound on the average delay under this policy.

First we explain the model in more details. Similar to the previous section, consider m collectors in a square region $\mathcal{R}$ of area A receiving data messages that arrive in time according to a Poisson process of intensity λ. Upon arrival the messages are distributed independently and uniformly in $\mathcal{R}$ and an arriving message is transmitted when a collector comes within the *reception distance*, r^*, of the message location and grants access for the message's transmission. We assume that time is slotted, $t = 0, 1, 2, ...$, where the slot length is equal to one message transmission time s. We consider a partitioning of the network region into a grid $\mathcal{G}$ of $\left(\frac{r^*}{\sqrt{2}} \times \frac{r^*}{\sqrt{2}}\right)$ squares, i.e., $K \doteq \frac{A}{(r^*)^2/2}$ square cells of diameter r^* as shown in Fig. 8 [8]. The collectors are confined to move on the grid $\mathcal{G}$ and simultaneous transmissions to different collectors are subject to interference constraints defined below.

Definition 3 (Cell Interference Model) *Given a collector that is at the intersection of (at most) 4 adjacent cells, a transmission to the collector from one of the cells is successfully received if there is no other transmission within the other cells adjacent to the collector.*

The Cell Interference Model essentially creates an *exclusion region* of 4 cells around a collector receiving a message. Similar interference models have been considered in literature. For example, the Protocol Model considered in [24], [43] assumes successful transmission if a disk region around the receiver has no other transmission. Similarly, the Vulnerability Circle Model considered in [16] or the Disk Reception Model considered in [15] require an exclusion region around receivers for successful reception.

The Cell Interference Model essentially creates K cells which can be treated as "users" in a multi-user queue with m servers. Assume a fixed ordering of these subregions. Due to the splitting property of the Poisson process, each cell $i = 1, 2, ..., K$, receives Poisson arrivals with rate $\lambda_i \doteq \lambda/K$. Let $A_i(t)$ denote the number of messages that arrive into cell i in time slot t. The expected *load* entering cell i per time slot is given by $\rho_i = \lambda_i s$. Furthermore, let $N_i(t)$ be the number of messages present in cell i at the beginning of time slot t. We have $\mathbf{N}(t) \doteq [N_1(t), ..., N_K(t)]$ and $N(t) = N_1(t) + ... + N_K(t)$, where $N(t)$ is the total number of messages in the system at the beginning of time slot t. We assume that the system is initially empty, therefore $N(0) = 0$.

Next we characterize the interference constraints of the system in terms of *activation vectors*. We call a cell *active* if at least 1 message in the cell is scheduled for transmission, and we assume that each cell $k = 1, 2, ..., K$ is associated with exactly one pick up location on the grid $\mathcal{G}$. For instance, the pick up location for each cell could be the upper left corner of the cell as shown in Fig. 8. Therefore, specifying the set of cells to activate also specifies the locations of the collectors. A feasible activation

---

[8] Similar to the previous sections, if $K = \frac{A}{(r^*)^2/2}$ is not an integer, one can partition the region using the largest reception distance $\underline{r}^* < r^*$ such that this integer condition is satisfied.
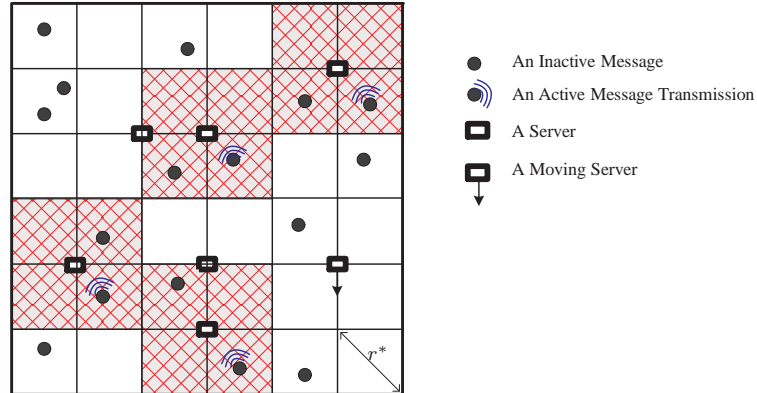
Fig. 8: The network model. Red regions are the exclusion zones for the servers currently in service. 2 Servers are forced to be inactive since the queue in their vicinity are in the exclusion zone of another server.

vector $\mathbf{I} \in \mathcal{I}$ is the one under which the transmissions from the set of active cells do not interfere with each other, where $\mathcal{I}$ is the set of all feasible activation vectors. If cell k, $k = 1, ..., K$, is active under activation vector $\mathbf{I}$, then we have $\mathbf{I}(k) = 1$, otherwise $\mathbf{I}(k) = 0$. The set $\mathcal{I}$ consists of K-dimensional vectors of at most m nonzero entries. Furthermore, because transmissions require an exclusion zone of up to 4 adjacent cells, the number of nonzero entries of vectors in $\mathcal{I}$ is typically less than m. Note that we include the zero vector $\mathbf{I} = \mathbf{0}$ in $\mathcal{I}$ for convenience.

An alternative way of modeling the interference constraints is to have a 4 by K matrix for each activation $\mathbf{I}$. In this representation every column of the matrix $\mathbf{I}$ corresponds to a single cell and the different rows in a column vector specifies the position of the collector responsible for serving the cell. Therefore, for a given matrix $\mathbf{I}$, at most one element of a column vector can be nonzero. The matrix activation model is more complex, however, it does not require the extra assumption that each cell is associated with a single pick up location. Here we present our results in terms of the simpler vector activation model for ease of exposition.

Let T_r denote the number of time slots required for a collector to move from the lower right corner of the grid $\mathcal{G}$ to the upper left corner of $\mathcal{G}$. We call T_r the *reconfiguration time* of the network. Consider the corresponding system with infinite speed, i.e., $T_r = 0$. This system is essentially a parallel queuing system with with multiple servers and interference constraints, which is a special case of the system considered in [38] or [46]. When $T_r = 0$, the stability region of this system, $\mathbf{\Lambda}^0$, consists of the closure of all load vectors $\boldsymbol{\rho} = [\rho_1, \rho_2, ..., \rho_K]' = s[\lambda_1, \lambda_2, ..., \lambda_K]'$ in the convex hull of the vectors in $\mathcal{I}$ [12], [37], i.e.,

$$\mathbf{\Lambda}^0 = \{\boldsymbol{\rho} | \boldsymbol{\rho} \in Conv\{\mathcal{I}\}\}. \tag{26}$$

The celebrated Max-Weight scheduling algorithm was introduced in [46] and was shown to be throughput optimal for the system with zero reconfiguration time. Specifically, the Max-Weight policy activates the set of users in $\mathbf{I}^*(t)$ where

$$\mathbf{I}^*(t) = \arg \max_{\mathbf{I} \in \mathcal{I}} \mathbf{N}(t).\mathbf{I}. \tag{27}$$

When $T_r > 0$, we lose service opportunities during the reconfiguration times. Therefore, the stability region of our system can be no larger than $\mathbf{\Lambda}^0$:

Lemma 4 *An outer bound on the stability region is given by*

$$\mathbf{\Lambda} \subseteq \mathbf{\Lambda}^0.$$

The analysis in the next section shows that we have $\mathbf{\Lambda} = \mathbf{\Lambda}^0$. The intuition behind this is that, under throughput-optimal policies, as the load approaches the boundary of the stability region, the fraction of time spent to reconfiguration tends to zero.

4.1 Framed-Max-Weight Policy

For systems with nonzero reconfiguration times, the Max-Weight policy is not throughput-optimal [11], [12]. The intuitive reason behind this is that the Max-Weight policy gives decisions to change the schedule too frequently, resulting in throughput-loss during the reconfiguration intervals. A frame based version of the Max-Weight policy where the same schedule is used throughout the frame incurs less throughput loss during the reconfiguration intervals. Indeed, single hop optical networks with interference constraints and nonzero reconfiguration times were considered in [11], where a frame-based version of the Max-Weight policy was shown to be throughput-optimal asymptotically in the frame length. Fluid limit of the system was considered in [11] and throughput-optimality was established under rate stability[9]. Furthermore, an NxN switch with matching constraints and reconfiguration delays was considered in [12], where a policy based on servicing batches of arrivals over frames according to the Max-Weight activation vector was shown to be throughput optimal asymptotically in the frame length. In fact, we show that the frame-based Max-Weight (FMW) policy proposed in [11] is throughput-optimal asymptotically in the frame length also for the system considered in this section. Different from the analysis in [11] and [12], we prove this result using classical quadratic Lyapunov drift techniques. The reason that the FMW policy stabilizes the system is that as the system load approaches the boundary of the stability region, the policy employs good schedules, i.e., maximum-weight schedules, over longer frame lengths, effectively decreasing the fraction of throughput lost to reconfiguration to zero.

Under the FMW policy the time is divided into intervals of length T slots. The FMW policy picks the activation vector corresponding to the Max-Weight configuration at the beginning of each frame. Then it idles the system for T_r slots to ensure that all the servers travel to their assigned locations. Then the policy applies this activation vector for $T - T_r$ slots. The choice of the frame length T depends on the load $\boldsymbol{\rho}$. Specifically, the policy requires $T > T_r/\epsilon$ where ϵ is determined by solving the Linear Program below [11].

$$\epsilon(\boldsymbol{\rho}) \triangleq \max \left(1 - \sum_{\mathbf{I} \in \mathcal{I}} \alpha_{\mathbf{I}} \right)$$

$$\text{subject to} \quad \rho_i = \lambda_i s \leq \sum_{\mathbf{I} \in \mathcal{I}} \alpha_{\mathbf{I}} \mathbf{I}(i), \ i \in \{1, ..., K\}$$

$$\sum_{\mathbf{I} \in \mathcal{I}} \alpha_{\mathbf{I}} \leq 1$$

$$\alpha_{\mathbf{I}} \geq 0, \ \forall \mathbf{I} \in \mathcal{I}. \tag{28}$$

Note that ϵ is a measure of distance of the load vector to the boundary of the stability region [11]. The FMW policy is described in Algorithm 2 in detail.

Theorem 8 *For any $\boldsymbol{\rho} = [\rho_1, ..., \rho_K]'$ strictly inside $\mathbf{\Lambda_0}$, the FMW policy stabilizes the system as long as $T > \frac{T_r}{\epsilon}$.*

The proof is in Appendix E. A similar result is proved in [12] for the fluid limit of this system establishing the rate stability. The proof in Appendix E is based on a quadratic Lyapunov drift argument over

---

[9] A queue of length $N_i(t)$ at time t is rate stable if $\lim_{t \to \infty} N_i(t)/t = 0$. This is a weaker notion of stability as compared to the stability definition in 1 which implies bounded first moments of a stationary measure.

Algorithm 2 *Framed-Max-Weight Policy*

1: Find the Max-Weight configuration for the servers based on the queue lengths at the beginning of the frame: Assuming the system is at the kth frame, find

$$\mathbf{I}^*(t) = \arg\max_{\mathbf{I}\in\mathcal{I}} \mathbf{N}(kT).\mathbf{I}. \tag{29}$$

2: Let the users be idle for T_r slots and reconfigure the collectors to their new locations.
3: Apply the activation vector $\mathbf{I}^*$ for $T - T_r$ slots where $T > \frac{T_r}{\epsilon}$.

---

a sufficiently long frame of duration T. Let t_k be the first slot of the kth frame. The proof establishes that the T-step expected drift of the queue lengths satisfies

$$\mathbb{E}\left[L(\mathbf{N}(t_k + T)) - L(\mathbf{N}(t_k))\big|\mathbf{N}(t_k)\right] \leq KBT^2 - \frac{2T}{K}(\epsilon - \frac{T_r}{T})\sum_i N_i(t_k), \tag{30}$$

where $B = 1 + \frac{\lambda s}{K} + \frac{\lambda^2 s^2}{K^2}$ is a constant. This means that the queue sizes tend to decrease if they are outside a bounded set for a sufficiently large frame length which implies bounded expected queue sizes [37].

The following corollary follows from Theorem 8 and Lemma 4.

Corollary 1 *We have*

$$\mathbf{\Lambda} = \mathbf{\Lambda_0}.$$

Note that the overall arrival rate λ is given by $\lambda = \lambda_1 + ... + \lambda_K$ where $\lambda_k = \lambda/K$ for $k = 1, ..., K$. Therefore, the necessary and sufficient stability condition on $\rho = \lambda s$ is $\rho < \hat{\rho}$ where $\hat{\rho}$ is the intersection of the K-dimensional region $\mathbf{\Lambda}$ with the line $\rho_1 = ... = \rho_K = \rho/K$.

4.2 Delay Analysis

In this section we derive an upper bound on the expected number of messages in the system. Note that through Little's law, the expected delay in the system is proportional to the expected number of messages in the system. Taking the expectation of (30) over the distribution of $\mathbf{N}(kT)$ we have

$$\mathbb{E}\left[L(\mathbf{N}(t_k + T))\right] - \mathbb{E}\left[L(\mathbf{N}(t_k))\right] \leq KBT^2 - \frac{2T}{K}(\epsilon - \frac{T_r}{T})\sum_i \mathbb{E}\left[N_i(t_k)\right].$$

We write a similar drift expression for $k = 0, 1, ..., M - 1$, and sum over k to obtain a telescoping series that gives

$$\mathbb{E}\left[L(\mathbf{N}(t_M))\right] - \mathbb{E}\left[L(\mathbf{N}(0))\right] \leq MKBT^2 - \frac{2T}{K}(\epsilon - \frac{T_r}{T})\sum_{k=0}^{M-1}\sum_i \mathbb{E}\left[N_i(t_k)\right],$$

where $t_{k+1} = t_k + T$, $k = 0, 1, ...$ Dividing by M and using $L(\mathbf{N}(t_M)) \geq 0$ and $L(\mathbf{N}(0)) = 0$ we have

$$\frac{1}{M}\left(\frac{2T}{K}(\epsilon - \frac{T_r}{T})\sum_{k=0}^{M-1}\sum_i \mathbb{E}\left[N_i(t_k)\right]\right) \leq KBT^2.$$

Taking the $\limsup$ as $M \to \infty$ we have

$$\limsup_{M\to\infty} \frac{1}{M}\sum_{k=0}^{M-1}\sum_i \mathbb{E}\left[N_i(t_k)\right] \leq \frac{K^2 TB}{2(\epsilon - T_r/T)}.$$

Noting that $\sum_i N_i(t) = N(t)$ that is the total number of messages in the system at time slot t we have

$$\limsup_{M \to \infty} \frac{1}{M} \sum_{k=0}^{M-1} \mathbb{E}\left[N(t_k)\right] \leq \frac{K^2 T B}{2(\epsilon - T_r/T)}. \tag{31}$$

Now for any given $t \in \{t_k, t_{k+1} - 1\}$ we have

$$N(t) \leq N(t_k) + \sum_{\tau=0}^{T-1} \sum_i A_i(t_k + \tau).$$

Summing over the time slots within the kth frame and taking conditional expectation we have for any $t \in \{t_k, t_{k+1} - 1\}$

$$\sum_{\tau=0}^{T-1} \mathbb{E}\left[N(t_k + \tau) | \mathbf{N}(t_k)\right] \leq T N(t_k) + T^2 \sum_i \lambda_i s,$$

where we used the fact that arrival processes are i.i.d. and independent of the queue lengths. Using $\lambda = \sum_i \lambda_i$, taking expectation with respect to $\mathbf{N}(t_k)$, writing similar expressions for M frames and summing them we obtain

$$\frac{1}{MT} \sum_{t=0}^{MT-1} \mathbb{E}\left[N(t)\right] \leq \frac{1}{M} \sum_{k=0}^{M-1} \mathbb{E}\left[N(t_k)\right] + T\rho.$$

Finally, taking the $\limsup$ as $M \to \infty$ and using (31) we have

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left[N(t)\right] \leq \frac{K^2 T B}{2(\epsilon - T_r/T)} + T\rho,$$

Using this expression it can be shown that [37]

$$\limsup_{t \to \infty} \mathbb{E}\left[N(t)\right] \leq \frac{K^2 T B}{2(\epsilon - T_r/T)} + T\rho.$$

For large frame lengths such that $T >> T_r$, the delay scaling is approximately proportional to $1/\epsilon$, where ϵ is a measure of the distance of arrival rate to the boundary of the stability region. Note that when the reconfiguration interval is large, large frame lengths $T >> T_r$ may be required for achieving throughput-optimality, which, on the other hand, increases the expected delay as the delay under the FMW policy is proportional to the frame length.

5 CONCLUSION

In this paper we considered the use of dynamic vehicle routing in order to improve the throughput and delay performance of wireless networks where messages arriving randomly in time and space are gathered by mobile collectors via wireless communications. For the case of a single collector, we characterized the stability region of this system to be all system loads $\rho < 1$. We developed fundamental lower bounds on time average expected delay and derived upper bounds on delay by analyzing TSPN and Partitioning policies. For the case of multiple collectors whose communications do not interfere with each other, we extended the stability and delay scaling results of the single collector case. Our results show that combining controlled mobility and wireless transmission results in $\Theta(\frac{1}{1-\rho})$ delay scaling with load ρ. This is the fundamental difference between our system and the system without wireless transmission (DTRP) analyzed in [8] and [9] where the delay scaling with the load is $\Theta(\frac{1}{(1-\rho)^2})$. Finally, for the the case where simultaneous transmissions to different collectors

interfere with each other we formulate a scheduling problem and characterize the stability region of the system in terms of interference constraints. We show that a frame-based version of the Max-Weight policy is throughput-optimal asymptotically in the frame length and derive an upper bound on average delay under this policy.

This work is a first attempt towards utilizing a combination of controlled mobility and wireless transmission for data collection in stochastic and dynamic wireless networks. Therefore, there are many related open problems. In this paper we have utilized a simple wireless communication model based on a communication range. In the future we intend to study more advanced wireless communication models such as modeling the transmission rate as a function of the transmission distance. For the case of multiple-collectors whose transmissions are subject to interference constraints, we intend to study interference models that do not restrict the collectors' motion to a grid. Note that such a joint server routing and scheduling problem is significantly more involved. For instance, the stability region of such a system depends on the interference constraints, and it is unknown since there are uncountably many possible activation vectors.

Appendix A - Proof of Theorem 1

We first show that the unfinished work and the delay experienced by a message in the system stochastically dominates that in the equivalent system with zero travel times for the collector.

Lemma 5 *The steady state time average delay in the system is at least as big as the delay in the equivalent system in which travel times are considered to be zero (i.e., $v = \infty$).*

Proof Consider the summation of per-message reception and travel times, s and d_i, as the total service requirement of a message in each system. Since d_i is zero for all i in the infinite speed system and since the reception times are constant equal to s for both systems, the total service requirement of each message in our system is deterministically greater than that of the same message in the infinite speed system. Let D_i and D_i', $i = 1, 2...$, be the departure instant of the ith message in the original and the infinite speed system respectively. Similarly let A_i, $i = 1, 2...$ be the arrival time of the ith message in both systems. We will use induction to prove that $D_i \geq D_i'$ for all i. We trivially have $D_1 \geq D_1'$. Furthermore,

$$A_{n+1} \leq D_{n+1} - s, \tag{32}$$

hence the $n + 1^{th}$ message is available before the time $D_{n+1} - s$. Using the induction hypothesis, $D_n \geq D_n'$, we have

$$D_n' \leq D_n \leq D_{n+1} - s,$$

where the second inequality is because we need at least s amount of time between the nth and $n + 1^{th}$ transmissions. Hence the collector is available in the infinite speed system before the time $D_{n+1} - s$. Combining this with (32) proves the induction.

Now let $D(t)$ and $D'(t)$ be the total number of departures by time t in our system and the infinite speed system respectively. Similarly let $N(t)$ and $N'(t)$ be the total number of messages in the two systems at time t. Finally let $A(t)$ be the total number of arrivals by time t in both systems. We have $A(t) = N(t) + D(t) = N'(t) + D'(t)$. Using the result of the above induction we have $D(t) \leq D'(t)$ and therefore

$$N(t) \geq N'(t).$$

Since this is true at all times, we have that the time average number of customers in the system is greater than that in the infinite speed system. Finally using Little's law proves the lemma. □

Since the infinite speed system is an M/G/1 queue (an M/G/1 queue is a queue with Poisson arrivals, general i.i.d. service times and 1 server and an M/D/1 queue has constant service times), the average waiting time in this system is given by the Pollaczek-Khinchin (P-K) formula for M/G/1 queues [7, p.

189], i.e., $\lambda s^2/(2(1-\lambda s))$. Therefore we have (1). Furthermore, a direct consequence of this lemma is that a necessary condition for stability in the infinite speed system is also necessary for our system. It is well-known that the necessary (and sufficient) condition for stability in the M/G/1 queue is given by $\rho < 1$ (see e.g., [7] or [22]). □

Appendix B - Proof of Theorem 5

The proof is similar to the proof of Theorem 1. First consider the following lemma.

Lemma 6 *The steady state time average delay in the system is at least as big as the delay in the equivalent system in which travel times are considered to be zero (i.e., $v = \infty$).*

Proof Similar to the proof of Lemma 5, the total service requirement of a given message i, $s + d_i$, is deterministically greater than that of the same message in the infinite speed system. Similarly, let D_i and $D_i^{'}$, and A_i, $i = 1, 2, ...$, be the departure and arrival instants of the ith message in the original and the infinite speed system. We use *complete induction* to prove that $D_i \geq D_i^{'}$ for all i. We trivially have $D_1 \geq D_1^{'}$. Assume we have $D_i \geq D_i^{'}$ for all $i \leq n$. We need to show that $D_{n+1} \geq D_{n+1}^{'}$ in order to complete the complete induction. We have

$$A_{n+1} \leq D_{n+1} - s, \tag{33}$$

hence the $n + 1^{th}$ message is available at time $D_{n+1} - s$. We also have

$$D_{n+1-m}^{'} \leq D_{n+1-m} \leq D_{n+1} - s.$$

The first inequality is due to the complete induction hypothesis and the second inequality is due the fact that the mth last departure before the $n + 1^{th}$ departure has to occur before the time $D_{n+1} - s$. Hence there is at least one collector available in the infinite speed system before the time $D_{n+1} - s$. Combining this with (33) proves the complete induction.

Similar to the proof of Lemma 5, let $D(t)$, $D^{'}(t)$ and $N(t)$, $N^{'}(t)$ be the total number of departures from the two systems by time t and the total number of messages waiting for service in the two systems at time t respectively. Also let $A(t)$ be the total number of arrivals by time t in both systems. We have $A(t) = N(t) + D(t) = N^{'}(t) + D^{'}(t)$. From the above induction we have $D(t) \leq D^{'}(t)$ and therefore $N(t) \geq N^{'}(t)$. Since this is true at all times, we have that the time average number of customers in the system is greater than that in the infinite speed system. Finally using Little's law proves the lemma. □

When the travel time is considered to be zero, the system becomes an M/D/m queue (a queue with Poisson arrivals, constant service time and m servers). Therefore we can bound T_m^* using bounds for general G/G/m systems. In particular, the waiting time $W_{G/G/m}$ in a G/G/m queue with service time s is bounded from below by [28, p. 48]

$$W_{G/G/m} \geq \hat{W} - \frac{m-1}{m}\frac{s^2}{2s}, \tag{34}$$

where $\hat{W}$ is the waiting time in a single server system with the same arrivals as in the G/G/m queue and service time s/m. Since in our case the infinite speed system behaves as an M/D/m system, $\hat{W}$ has an exact expression given by the P-K formula: $\hat{W} = \lambda s^2/(2m^2(1-\rho))$ where $\rho = \lambda s/m$. Substituting this in (34) and using Lemma 6 we have (13).

Appendix C - Proof of Lemma 2

Clearly $h(.) = f(.).g(.)$ is increasing. Let x and y be two points in the domain of h and let $\alpha \in (0, 1)$ be a real number.

$$
\begin{aligned}
h(\alpha x + (1 - \alpha)y) &= f(\alpha x + (1 - \alpha)y)g(\alpha x + (1 - \alpha)y) \\
&\leq (\alpha f(x) + (1 - \alpha)f(y))(\alpha g(x) + (1 - \alpha)g(y)) \\
&= \alpha^2 f(x)g(x) + (1 - \alpha)^2 f(y)g(y) + \alpha(1-\alpha)f(x)g(y)+\alpha(1-\alpha)f(y)g(x),
\end{aligned}
$$

where the inequality is due to the convexity of f and g. We add and subtract $\alpha f(x)g(x)$ and after some algebra obtain

$$
\begin{aligned}
h(\alpha x + (1 - \alpha)y) &\leq \alpha h(x) + (1 - \alpha)h(y) + \alpha(1 - \alpha)(f(x) - f(y))(g(y) - g(x)) \\
&\leq \alpha h(x) + (1 - \alpha)h(y),
\end{aligned}
$$

where the last inequality is due to the fact that f and g are increasing functions.

Appendix D - Proof of Lemma 3

It is clear that $h(x)$ is an increasing function of x. Let $x, y \geq 0$ be two points in the domain of h and let $\alpha \in (0, 1)$ be a real number.

$$
\begin{aligned}
h(\alpha x + (1 - \alpha)y) = &= (\alpha x + (1 - \alpha)y)\max(0, c_1\sqrt{\alpha x + (1 - \alpha)y} - c_2) \\
&= \max(0, c_1(\alpha x + (1 - \alpha)y)^{\frac{3}{2}} - c_2(\alpha x + (1 - \alpha)y)) \\
&\leq \max\big(0, c_1(\alpha x^{\frac{3}{2}} + (1 - \alpha)y^{\frac{3}{2}}) - c_2(\alpha x + (1 - \alpha)y)\big) \\
&= \max\big(0, \alpha x(c_1\sqrt{x} - c_2) + (1 - \alpha)y(c_1\sqrt{y} - c_2)\big) \\
&\leq \max\big(0, \alpha x(c_1\sqrt{x}-c_2)\big)+\max\big(0, (1-\alpha)y(c_1\sqrt{y}-c_2)\big) \\
&= \alpha h(x) + (1 - \alpha)h(y),
\end{aligned}
$$

where the first inequality is due to the convexity of the function $x^{\frac{3}{2}}$. This shows that $h(x)$ is a convex and increasing function.

Appendix E - Proof of Theorem 8

We prove Theorem 8 for a broader class of arrival processes. We assume that each cell i has an arrival process $A_i(t)$ that is i.i.d. over time and that satisfies $\mathbb{E}[A_i(t)^2] \leq A_{\max}^2$ independent of the number of messages in the system, which is satisfied if the overall arrival process into the system is Poisson. Note that we have $\mathbb{E}[A_i(t)] = \lambda_i s$ independent of the number of messages in the system. Let t_k, $k = 0, 1, ...$, be the first time slot of the kth frame. Let $D_i(t)$, $t \in \{t_k + T_r, t_{k+1} - 1\}$, be 1 if cell i is scheduled to be active during the kth frame and zero otherwise. Note that $D_i(t)$ is the *service opportunity* given to cell i at time slot t and not the actual departure process. Let $N_i(t)$ be the number of messages in cell i at the beginning of the time slot t. Recall that for simplicity we assume that the departures occur before the arrivals which takes place at the end of time slots. We have the following queue evolution relation.

$$
N_i(t + 1) = \max\{N_i(t) - D_i(t), 0\} + A_i(t).
$$

Similarly, the following T-step queue evolution expression holds:

$$
N_i(t_k + T) \leq \max\left\{N_i(t_k) - \sum_{\tau=0}^{T-1} D_i(t_k + \tau), 0\right\} + \sum_{\tau=0}^{T-1} A_i(t_k + \tau).
$$

The inequality is due to the fact that cell i might become empty and that some arrivals depart during the frame. Squaring both sides we have,

$$(N_i(t_k+T))^2 - (N_i(t_k))^2 \leq \Big(\sum_{\tau=0}^{T-1} D_i(t_k+\tau)\Big)^2 + \Big(\sum_{\tau=0}^{T-1} A_i(t_k+\tau)\Big)^2 - 2N_i(t_k)\Big(\sum_{\tau=0}^{T-1} D_i(t_k+\tau) - \sum_{\tau=0}^{T-1} A_i(t_k+\tau)\Big).$$
(35)

Define the quadratic Lyapunov function

$$L(\mathbf{N}(t_k)) = \sum_{i=1}^{K} N_i^2(t_k),$$

and the T-step conditional Lyapunov drift

$$\Delta_T(t_k) \triangleq \mathbb{E}\left\{L(\mathbf{N}(t_k+T)) - L(\mathbf{N}(t_k))\big|\mathbf{N}(t_k)\right\}.$$

Summing (35) over the queues, taking conditional expectation, using $D_i(t) \leq 1$ for all time slots t, $\mathbb{E}\{A_i(t)^2\} \leq A_{\max}^2$ and $\mathbb{E}\{A_i(t_1)A_i(t_2)\} \leq \sqrt{\mathbb{E}\{A_i(\tau_1)\}^2 \mathbb{E}\{A_i(\tau_2)\}^2} \leq A_{\max}^2$ for all τ_1 and τ_2 we have

$$\Delta_T(t_k) \leq KBT^2 + 2\mathbb{E}\left\{\sum_i N_i(t_k) \sum_{\tau=0}^{T-1} [A_i(t_k+\tau) - D_i(t_k+\tau)] \big|\mathbf{N}(t_k)\right\}$$

$$= KBT^2 + 2T \sum_i N_i(t_k)\lambda_i s - 2\sum_i N_i(t_k)\mathbb{E}\left\{\sum_{\tau=0}^{T-1} D_i(t_k+\tau)\big|\mathbf{N}(t_k)\right\}$$

where $B = 1 + A_{\max}^2$ is a constant. Note that $D_i(t+\tau) = 0, \forall i \in \{1, ..., K\}$ for $\tau \in \{0, 1, ..., T_r-1\}$ since the system is idle for the first T_r slots of the frame under the FMW policy. Therefore,

$$\Delta_T(t_k) \leq NBT^2 + 2T \sum_i N_i(t_k)\lambda_i s - 2\sum_i \sum_{\tau=T_r}^{T-1} N_i(t_k)\mathbb{E}\left\{D_i(t_k+\tau)\big|\mathbf{N}(t_k)\right\}$$

Now using the fact that for any load vector $\boldsymbol{\rho} = \boldsymbol{\lambda}s$ that is strictly inside $\boldsymbol{\Lambda}^0$, there exist real numbers $\alpha_1, ..., \alpha_{|\mathcal{I}|}$ such that $\alpha_j > 0, \forall j \in 1, ..., |\mathcal{I}|, \sum_{j=1}^{|\mathcal{I}|} \alpha_j = 1 - \epsilon$ for some $\epsilon > 0$ and

$$\boldsymbol{\rho} = \sum_{j=1}^{|\mathcal{I}|} \alpha_j \mathbf{I}^j$$

where $\mathbf{I}^j$ is a K dimensional vector in $\mathcal{I}$. Over the time interval $[t+T_r, t+T-1]$, the FMW policy applies the activation vector that has the property

$$\mathbf{I}^*(t_k) = \arg\max_{\mathbf{I}\in\mathcal{I}} \mathbf{N}(t_k).\mathbf{I}.$$
(36)

Therefore $\sum_i N_i(t_k)D_i(t_k+\tau) = \mathbf{N}(t_k).\mathbf{I}^*(t_k)$. Hence we have

$$\Delta_T(t_k) \leq KBT^2 + 2T\mathbf{N}(t_k).\left(\sum_{j=1}^{|\mathcal{I}|} \alpha_j \mathbf{I}^j\right) - 2T(1 - \frac{T_r}{T})\mathbf{N}(t_k).\mathbf{I}^*(t_k)$$

$$= KBT^2 - 2T\sum_{j=1}^{|\mathcal{I}|} \alpha_j(\mathbf{N}(t_k).\mathbf{I}^*(t_k) - \mathbf{N}(t_k).\mathbf{I}^j) - 2T(1 - \sum_{j=1}^{|\mathcal{I}|} \alpha_j)\mathbf{N}(t_k).\mathbf{I}^*(t_k) + 2T_r\mathbf{N}(t_k).\mathbf{I}^*(t_k)$$

$$\leq KBT^2 - 2T\epsilon\mathbf{N}(t_k).\mathbf{I}^*(t_k) + 2T_r\mathbf{N}(t_k).\mathbf{I}^*(t_k)$$

$$= KBT^2 - 2T(\epsilon - \frac{T_r}{T})\mathbf{N}(t_k).\mathbf{I}^*(t_k)$$
(37)

Note that we have $\mathbf{N}(t_k).\mathbf{I}^*(t_k) \geq \frac{1}{K}\sum_i N_i(t_k)$ since the maximum weight schedule has more weight than the average. Therefore, for $T > \frac{T_r}{\epsilon}$ we have

$$\Delta_T(t_k) \leq KBT^2 - 2T(\epsilon - \frac{T_r}{T})\frac{1}{K}\sum_i N_i(t_k). \tag{38}$$

Therefore, the T-step conditional Lyapunov drift is negative if $T > \frac{T_r}{\epsilon}$ and if the queue sizes are outside a bounded set. The stability result now follow from [38] or Section 4.2.

References

1. I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," *Ad Hoc Networks (Elsevier)*, vol. 3, no. 3, pp. 257-279, Mar. 2005.
2. E. Altman and H. Levy, "Queueing in space", *Adv. Appl. Prob.*, Vol. 26, no. 4, pp. 1095-1116, Dec. 1994.
3. E. Altman, P. Konstantopoulos, and Z. Liu, "Stability, monotonicity and invariant quantities in general polling systems," *Queuing Sys.*, vol. 11, pp. 35-57, 1992.
4. A. Al-Hanbali, A. A. Kherani, R. Groenevelt, P. Nain, and E. Altman, "Impact of mobility on the performance of relaying in ad hoc networks," In *Proc. IEEE INFOCOM'06*, Apr. 2006.
5. S. Asmussen, "Applied Probability and queues," Wiley, 1987.
6. E. M. Arkin and R. Hassin, "Approximation algorithms for the geometric covering salesman problem," *Discrete Applied Mathematics*, vol. 55, pp. 197-218, 1994.
7. D. Bertsekas and R. Gallager, "Data Networks," Prentice Hall, 1992.
8. D. J. Bertsimas and G. van Ryzin, "A stochastic and dynamic vehicle routing problem in the Euclidean plane," *Opns. Res.*, vol. 39, pp. 601-615, 1990.
9. D. J. Bertsimas and G. van Ryzin, "Stochastic and dynamic vehicle routing in the Euclidean plane with multiple capacitated vehicles," *Opns. Res.*, vol. 41, pp. 60-76, 1993.
10. D. J. Bertsimas and G. van Ryzin, "Stochastic and dynamic vehicle routing with general demand and interarrival time distributions," *Adv. App. Prob.*, vol. 20, pp. 947-978, 1993.
11. A. Brzezinski, "Scheduling algorithms for throughput maximization in data networks," Ph.D. thesis, MIT, 2007.
12. A. Brzezinski and E. Modiano, "Dynamic reconfiguration and routing algorithms for IP-over-WDM networks with stochastic traffic," *IEEE Journal of Lightwave Tech.*, vol. 23, no. 10, pp. 3188-3205, Oct. 2005
13. B. Burns, O. Brock, and B. N. Levine, "MV routing and capacity building in disruption tolerant networks," In *Proc. IEEE INFOCOM'05*, Mar. 2005.
14. G. D. Celik and E. Modiano, "Dynamic vehicle routing for data gathering in wireless networks," In *Proc. IEEE CDC'10*, Dec. 2010.
15. G. D. Celik and E. Modiano, "Random access wireless networks with controlled mobility", In *Proc. IFIP MEDHOC-NET'09*, Jun. 2009.
16. G. D. Celik, G. Zussman, W. Khan, and E. Modiano, " MAC for networks with multipacket reception capability and spatially distributed nodes", *IEEE Trans. Mobile Comp.*, vol. 9, no. 2, pp. 226–240, Feb. 2010
17. A. Dumitrescu and J. Mitchell, "Approximation algorithms for TSP with neighborhoods in the plane," *J. Algorithms*, vol. 48, pp. 135-159, 2003.
18. E. Frazzoli and F. Bullo, "Decentralized algorithms for vehicle routing in a stochastic time-varying environment," In *Proc. IEEE CDC'04*, Dec. 2004.
19. S. Foss and G. Last, "Stability of Polling systems with Exhaustive service policies and state-dependent routing", *Annals of Appl. Prob.,* vol. 6, no. 1, pp. 116-137, Feb. 1996.
20. A. E. Gammal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," In *Proc. IEEE INFOCOM'04*, Mar. 2004.
21. M. Garetto, P. Giaccone, and E. Leonardi, "On the capacity of ad hoc wireless networks under general node mobility," In *Proc. IEEE INFOCOM'07*, May 2007.
22. L. Georgiadis, M. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," Now Publishers, 2006.
23. M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks ," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 125-137, Feb. 2003.
24. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
25. M. Haimovich and T. L. Magnanti, "Extremum properties of of hexagonal partitioning and the uniform distribution in euclidian location," *SIAM J. Disc. Math.* 1, 50-64, 1988.

26. D. Jea, A. A. Somasundara, and M. B. Srivastava, "Multiple controlled mobile elements (data mules) for data collection in sensor networks," In *Proc. IEEE/ACM DCOSS'05*, Jun. 2008.
27. V. Kavitha and E. Altman, "Queueing in space: design of Message Ferry Routes in sensor networks," In *Proc. ITC'09*, Sep. 2009.
28. L. Kleinrock, "Queueing systems: Volume 2: Computer applications," John Wiley & Sons, 1976.
29. E. L. Lawler, J.Lenstra, A. Kan, D. Shmoys, " The Traveling salesman problem :a guided tour of combinatorial optimization," Wiley, 1985.
30. J. Lequay, T. Friedman, and V. Conan, "Evaluating mobility pattern space routing for DTNs," In *Proc. IEEE INFOCOM'06*, Apr. 2006.
31. H. Levy, M. Sidi, and O.L. Boxma, "Dominance relations in polling systems," *Queueing Systems*, vol. 6, pp. 155-172, Apr. 1990.
32. X. Lin, G. Sharma, R. R. Mazumdar, and N. B. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with Brownian mobility," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 2777-2784, Jun. 2006.
33. J. Luo and J. P. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," In *Proc. IEEE INFOCOM'05*, Mar. 2005.
34. J. S. B. Mitchell, "A PTAS for TSP with neighborhoods among fat regions in the plane," In *Proc. ACM-SIAM SODA'07*, Jan. 2007.
35. E. Modiano, D. Shah and G. Zussman, "Maximizing throughput in wireless networks via Gossip," In *Proc. ACM SIGMETRICS/Performance'06*, June 2006.
36. R. M. de Moraes, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "On mobility-capacity-delay trade-off in wireless ad hoc networks," In *Proc. IEEE MASCOTS'04*, Oct. 2004.
37. M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," Morgan and Claypool, 2010.
38. M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.
39. M. J. Neely and E. Modiano, " Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1917-1937, Jun. 2005.
40. M. Pavone, N. Bisnik, E. Frazzoli and V. Isler, "Decentralized vehicle routing in a stochastic and dynamic environment with customer impatience," In *Proc. RoboComm'07*, Oct. 2007.
41. J. Le Ny, M. Dahleh, E. Feron, and E. Frazzoli, "Continuous path planning for a data harvesting mobile server," In *Proc. IEEE CDC'08*, Dec. 2008.
42. R. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: Modeling a three-tier architecture for sparse sensor networks," In *Proc. IEEE SNPA'03*, May 2003.
43. V. Sharma, E. Frazzoli, and P. G. Voulgaris, "Delay in mobility-assisted constant-throughput wireless networks," In *Proc. IEEE CDC'05*, Dec. 2005.
44. Y. Shi and Y. T. Hou, "Theoretical results on base station movement problem for sensor network," In *Proc. IEEE INFOCOM'08*, Apr. 2008.
45. A. Somasundara, A. Kansal, D. Jea, D. Estrin, and M. Srivastava, "Controllably mobile infrastructure for low energy embedded networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 8, pp. 958-973, Aug. 2006.
46. L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Auto. Control*, vol. 37, no. 12, pp. 1936-1948, Dec. 1992.
47. L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Infor. Theory*, vol. 39, no. 2, pp. 466-478, Mar. 1993.
48. V. M. Vishnevskii and O. V. Semenova, "Mathematical methods to study the polling systems," *Auto. and Rem. Cont.*, vol. 67, no. 2, pp. 173-220, Feb. 2006.
49. H. Waisanen, D. Shah, M. A. Dahleh, " Minimal delay in controlled mobile relay networks," In *Proc. IEEE CDC'06*, Dec. 2006.
50. H. Waisanen, D. Shah, M. A. Dahleh, " Lower bounds for multi-stage vehicle routing," In *Proc. IEEE CDC'07*, Dec. 2007.
51. H. Xu, "Optimal policies for stochastic and dynamic vehicle routing problems," Ph.D. Thesis, MIT, 1994.
52. M. Zhao, M. Ma, and Y. Yang, "Mobile data gathering with space-division multiple access in wireless sensor networks," In *Proc. IEEE INFOCOM'08*, Apr. 2008.
53. W. Zhao, M. Ammar, and E. Zegura, "A message ferrying approach for data delivery in sparse mobile ad hoc networks," In *Proc. ACM MobiHoc'04*, May 2004.