

# Adaptive and Distributed Algorithms for Vehicle Routing in a Stochastic and Dynamic Environment

Marco Pavone   Emilio Frazzoli   Francesco Bullo

**Abstract**—In this paper we present adaptive and distributed algorithms for motion coordination of a group of  $m$  vehicles. The vehicles must service demands whose time of arrival, spatial location and service requirement are stochastic; the objective is to minimize the average time demands spend in the system. The general problem is known as the  $m$ -vehicle Dynamic Traveling Repairman Problem ( $m$ -DTRP). The best previously known control algorithms rely on centralized task assignment and are not robust against changes in the environment. In this paper, we first devise new control policies for the 1-DTRP that: (i) are provably optimal both in light-load conditions (i.e., when the arrival rate for the demands is small) and in heavy-load conditions (i.e., when the arrival rate for the demands is large), and (ii) are adaptive, in particular, they are robust against changes in load conditions. Then, we show that specific partitioning policies, whereby the environment is partitioned among the vehicles and each vehicle follows a certain set of rules within its own region, are optimal in heavy-load conditions. Building upon the previous results, we finally design control policies for the  $m$ -DTRP that (i) are adaptive and distributed, and (ii) have strong performance guarantees in heavy-load conditions and stabilize the system in any load condition.

## I. INTRODUCTION

Advances in computer technology and wireless communications have opened up tremendous possibilities for deploying large networks of autonomous mobile vehicles capable of interacting directly with the environment. Potential applications are numerous and include surveillance and exploration missions, where vehicles are required to visit points of interest, and transportation and distribution tasks, where vehicles are required to provide service or goods to people. In this context, one of the major scientific challenges is the design of possibly distributed multi-vehicle policies to assign and schedule demands for service (or targets) that are defined over an extended geographical area.

In the recent past, considerable efforts have been devoted to the problem of how to cooperatively assign and schedule demands in multi-vehicle systems [1]–[6]. In these papers, the underlying mathematical model is *static* and fits within the framework of the Vehicle Routing Problem (see [7] for a thorough introduction to the VRP), whereby: (i) a team of  $m$  vehicles is required to service a set of  $n$  demands in

a 2-dimensional space; (ii) each demand requires a certain amount of on-site service; (iii) the goal is to compute a set of routes that optimizes the cost of servicing (according to some quality of service metric) the demands. The VRP is *static* in the sense that vehicle routes are computed assuming that no new demands arrive.

Despite its richness, inherent elegance and frequent occurrence in practical problems, the VRP is often a *static approximation* to problems that are in reality *dynamic*, and therefore it fails to be an appropriate model in many applications of interest [9]. In fact, demands often arrive *sequentially* in time, and planning algorithms should actually provide *policies* (in contrast to pre-planned routes) that prescribe how the routes should evolve as a function of those inputs that evolve in real-time. From a methodological point of view, dynamic demands (i.e., demands that vary over time) add *queueing phenomena* to the combinatorial nature of the VRP. As a motivating example, consider a surveillance mission where a team of Unmanned Aerial Vehicles (UAVs) must ensure continued coverage of a certain area; as events occur, i.e., as new targets are detected by on-board sensors or other assets (e.g., intelligence, high-altitude or orbiting platforms, etc.), UAVs must proceed to the location of the new event and provide close-range information about the target. Each request for close-range observation might also require an on-site service time (e.g., a time interval to identify the target), which might be known a priori *only* through prior statistics. The UAV mission control is then a *continuous* process of detecting targets, planning routes and sending UAVs. In such a dynamic and stochastic setting, the routing process might rely on a priori statistics about the future stream of targets, and stability of the overall system is an additional issue to be considered.

Accordingly, in the present work we wish to address the assignment and scheduling problem in the more general framework where new demands for service are generated *over time* by a stochastic process. In [10], the authors present a setting where demands arrive sequentially over time and routes are generated in real time, however the optimization problem is over a finite horizon and the analysis only addresses a number of vehicles  $m \leq 2$ . Arguably, the most general model for vehicle routing problems that have both a *dynamic* and a *stochastic* component is the  $m$ -vehicle Dynamic Traveling Repairman Problem ( $m$ -DTRP) [8], [9], [11], [12]. In the  $m$ -DTRP,  $m$  vehicles operating in a bounded environment and traveling with bounded velocity must service demands whose time of arrival, location and on-site service are stochastic (in the UAV example, the inter-arrival times might be exponential random variables, the demands' locations might be uniformly distributed within the workspace, and the time intervals required to identify the targets might be exponential random

The research leading to this work was partially supported by the National Science Foundation through grants #0705451 and #0705453, and by the Air Force Office of Scientific Research through grant AFOSR MURI #FA9550-07-1-0528.

**Marco Pavone** is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 (Marco.Pavone@jpl.nasa.gov).

**Emilio Frazzoli** is with the Laboratory for Information and Decision Systems, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 (frazzoli@mit.edu).

**Francesco Bullo** is with the Center for Control Engineering and Computation, University of California at Santa Barbara, Santa Barbara, CA 93106 (bullo@engineering.ucsb.edu).

## Report Documentation Page

*Form Approved  
OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>18 NOV 2010</b>	2. REPORT TYPE	3. DATES COVERED <b>00-00-2010 to 00-00-2010</b>			
4. TITLE AND SUBTITLE <b>Adaptive And Distributed Algorithms For Vehicle Routing In A Stochastic And Dynamic Environment</b>		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, 91109</b>		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>IEEE Transactions on Automatic Control, Volume: PP Issue:99, On page(s): 1 - 1 ISSN: 0018-9286, Government or Federal Purpose Rights License.</b>					
14. ABSTRACT <b>In this paper we present adaptive and distributed algorithms for motion coordination of a group of m vehicles. The vehicles must service demands whose time of arrival, spatial location and service requirement are stochastic; the objective is to minimize the average time demands spend in the system. The general problem is known as the m-vehicle Dynamic Traveling Repairman Problem (m-DTRP). The best previously known control algorithms rely on centralized task assignment and are not robust against changes in the environment. In this paper, we first devise new control policies for the 1-DTRP that: (i) are provably optimal both in light-load conditions (i.e., when the arrival rate for the demands is small) and in heavy-load conditions (i.e., when the arrival rate for the demands is large), and (ii) are adaptive, in particular, they are robust against changes in load conditions. Then, we show that specific partitioning policies whereby the environment is partitioned among the vehicles and each vehicle follows a certain set of rules within its own region, are optimal in heavy-load conditions. Building upon the previous results, we finally design control policies for the m-DTRP that (i) are adaptive and distributed, and (ii) have strong performance guarantees in heavy-load conditions and stabilize the system in any load condition.</b>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>16</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

TABLE I  
ADAPTIVE POLICIES FOR THE 1-DTRP

Properties	DC Policy, $r \rightarrow \infty$	DC Policy, $r = 1$	RH Policy	UTSP Policy [8], $r \rightarrow \infty$
Light-load performance	optimal	optimal	optimal	not optimal
Heavy-load performance	optimal	within 100% of the optimum	within 100% of the optimum*	optimal
Adaptive	yes, except for $f$	yes	yes	no

TABLE II  
DISTRIBUTED AND ADAPTIVE POLICIES FOR THE  $m$ -DTRP

Properties	$m$ -DC Policy, $r \rightarrow \infty$	$m$ -RH Policy	UTSP Policy [8], $r \rightarrow \infty$
Light-load performance	“locally” optimal	“locally” optimal	not optimal
Heavy-load performance:			
a) uniform $f$	optimal	within 100% of optimal unbiased*, <sup>†</sup>	optimal
b) general $f$ and $\bar{s} = 0$	optimal	within 100% of optimal unbiased*, <sup>†</sup>	optimal
c) general $f$ and $\bar{s} > 0$	within $m$ of optimal unbiased <sup>†</sup>	within $2m$ of optimal unbiased*, <sup>†</sup>	optimal
Adaptive	yes, except for $f$	yes, except for $f$	no
Distributed	yes	yes	no

variables as well). The objective is to find a policy to service demands over an infinite horizon that minimizes the expected system time (wait plus service) of the demands. The best previously known control policies for the  $m$ -DTRP rely on centralized task assignment and are not robust against changes in the environment, in particular changes in load conditions; therefore, they are of limited applicability in scenarios involving ad-hoc networks of autonomous vehicles operating in a time-varying environment.

In this paper we devise adaptive and distributed policies for the  $m$ -DTRP. We mainly focus on policies that in heavy load are constrained to provide *spatially-unbiased* service, i.e., the same quality of service to all outstanding demands independently of their locations (we refer to such policies as *unbiased* policies). We argue in Section III how such constraint is indeed natural for applications. The contribution is threefold: First, we present a new class of policies for the 1-DTRP that are adaptive (e.g., to the load conditions) and satisfy the aforementioned constraint. In particular, we propose the Divide & Conquer (DC) policy, whose performance depends on a design parameter  $r \in \mathbb{N}_{>0}$ . If  $r \rightarrow +\infty$ , the policy is (i) provably optimal both in light- and in heavy-load conditions, and (ii) adaptive with respect to changes in the load conditions and in the statistics of the on-site service requirement; if, instead,  $r = 1$ , the policy is (i) provably optimal in light-load conditions and within a factor 2 of the optimal performance in heavy-load conditions, and (ii) adaptive with respect to all problem data, in particular, and perhaps surprisingly, it does not require *any* knowledge about the demand generation process. Moreover, by applying ideas of receding-horizon control to dynamic vehicle routing problems, we introduce the Receding Horizon (RH) policy, that also does not require any knowledge about the demand generation process; we show that the RH policy is optimal in light load and stable in any load condition, and we heuristically argue that its performance is close to optimality in heavy-load conditions (in particular, we heuristically argue that the RH policy is the best available unbiased and adaptive policy for the 1-DTRP). Second, we show that specific partitioning policies, whereby

the environment is partitioned among the vehicles and each vehicle follows a certain set of rules within its own region, are optimal in heavy-load conditions. Finally, by combining the DC policy with suitable partitioning policies, we design a routing policy for the  $m$ -DTRP (called  $m$ -DC policy) that (i) is spatially distributed, scalable to large networks, and adaptive to network changes, (ii) is within a constant factor of the optimal unbiased performance in heavy-load conditions (in particular, it is an optimal unbiased policy when demands are uniformly dispersed over the environment or when the average on-site service time  $\bar{s}$  is zero) and stabilizes the system in *any* load condition. Here, by network changes we mean changes in the number of vehicles, in the arrival rate of demands, and in the characterization of the on-site service requirement. We mention that the problems addressed in this paper, and the proposed algorithms, can not be cast within the framework of “traditional” queueing systems, and thus their analysis requires the use of non-standard techniques. Tables I and II provide a synoptic view about the main properties of the routing policies that we propose and analyze in this paper; our policies are compared with the best spatially-unbiased policy available in the literature, i.e., the UTSP policy with  $r \rightarrow \infty$  [8]. (In Tables I and II an asterisk \* signals that the result is heuristic, a dagger † means that the spatially-unbiased constraint might be violated in heavy load,  $f$  denotes the spatial probability density for the demands’ locations,  $m$  is the number of vehicles, and “optimal unbiased” stands for optimal performance for policies providing spatially-unbiased service.)

The paper is structured as follows. In Section II we provide some background on the Euclidean Traveling Salesman Problem, and on the Continuous Multi-Median Problem. In Section III we formulate the problem we wish to address, and we review existing results. In Sections IV and V we present and analyze, respectively, the Divide & Conquer policy and the Receding Horizon policy. In Section VI we discuss partitioning policies and we design distributed algorithms for the  $m$ -DTRP. In Section VII we discuss the particular case when the on-site service time is zero, and in Section VIII we present results from numerical experiments. Finally, in Section IX we draw

some conclusions and discuss some directions for future work.

## II. PRELIMINARIES

In this section, we briefly describe some known concepts from combinatorics and locational optimization, on which we will rely extensively later in the paper.

### A. Equitable Partitions

Let  $\mathcal{Q} \subset \mathbb{R}^d$  be a compact, convex set. An  $r$ -partition of  $\mathcal{Q}$  (where  $r \in \mathbb{N}$ ,  $r \geq 1$ ) is a collection of  $r$  closed subsets  $\{\mathcal{Q}_k\}_{k=1}^r$  with disjoint interiors whose union is  $\mathcal{Q}$ . Given a measurable function  $f : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$ , an  $r$ -partition  $\{\mathcal{Q}_k\}_{k=1}^r$  is *equitable* with respect to  $f$  if  $\int_{\mathcal{Q}_k} f(x) dx = \int_{\mathcal{Q}} f(x) dx / r$  for all  $k \in \{1, \dots, r\}$ . Similarly, given two measurable functions  $f_j : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$ ,  $j \in \{1, 2\}$ , an  $r$ -partition  $\{\mathcal{Q}_k\}_{k=1}^r$  is *simultaneously equitable* with respect to  $f_1$  and  $f_2$  if  $\int_{\mathcal{Q}_k} f_j(x) dx = \int_{\mathcal{Q}} f_j(x) dx / r$  for all  $k \in \{1, \dots, r\}$  and  $j \in \{1, 2\}$ . Theorem 12 in [13] shows that, given two measurable functions  $f_j : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$ ,  $j \in \{1, 2\}$ , there *always* exists an  $r$ -partition of  $\mathcal{Q}$  that is simultaneously equitable with respect to  $f_1$  and  $f_2$  and where the subsets  $\mathcal{Q}_k$  are convex.

### B. Asymptotic and Worst-Case Properties of the Euclidean Traveling Salesman Problem

The Euclidean Traveling Salesman Problem (for short, TSP) is formulated as follows: given a set  $D$  of  $n$  points in  $\mathbb{R}^d$ , find a minimum-length tour (i.e., a cycle that visits all nodes exactly once) of  $D$ ; the length of a tour is the sum of all Euclidean distances on the tour. The decision version of the TSP belongs to the class of NP-complete problems, which suggests that there is no general algorithm capable of finding the optimal tour in an amount of time polynomial in the size of the input. Let  $\text{TSP}(D)$  denote the minimum length of a tour through all the points in  $D$ ; by convention,  $\text{TSP}(\emptyset) = 0$ . Assume that the locations of the  $n$  points are random variables independently and identically distributed in a compact set  $\mathcal{Q} \subset \mathbb{R}^d$ ; in [14] it is shown that there exists a constant  $\beta_{\text{TSP},d}$  such that

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{E}[\text{TSP}(D)]}{n^{1-1/d}} = \beta_{\text{TSP},d} \int_{\mathcal{Q}} \bar{f}^{1-1/d}(x) dx, \quad (1)$$

where  $\bar{f}$  is the density of the absolutely continuous part of the point distribution. For the case  $d = 2$ , the constant  $\beta_{\text{TSP},2}$  has been estimated numerically as  $\beta_{\text{TSP},2} \simeq 0.7120 \pm 0.0002$  [15].

For any compact environment  $\mathcal{Q} \subset \mathbb{R}^d$ , the following (deterministic) bound holds on the length of the optimal TSP tour, uniformly on  $n$  [14]:

$$\text{TSP}(D) \leq \beta_{\mathcal{Q},d} |\mathcal{Q}|^{1/d} n^{1-1/d}, \quad (2)$$

where  $|\mathcal{Q}|$  is the hypervolume of  $\mathcal{Q}$  and  $\beta_{\mathcal{Q},d}$  is a constant generally larger than  $\beta_{\text{TSP},d}$ .

In the following, we will present algorithms that require on-line solutions of possibly large TSPs. Practical implementations of the algorithms will rely on heuristics or on polynomial-time approximation schemes, such as Lin-Kernighan's [16] or Christofides' [17]. A modified version of the Lin-Kernighan heuristic [16] is implemented in

linkern<sup>1</sup>; this powerful solver yields approximations in the order of 5% of the optimal tour cost very quickly for many instances. If a constant-factor approximation algorithm is used, the effect on the asymptotic performance guarantees of our algorithms can be simply modeled as a scaling of the constant  $\beta_{\text{TSP},d}$ .

### C. The Continuous Multi-Median Problem

Given a compact set  $\mathcal{Q} \subset \mathbb{R}^d$  and a vector  $P = (p_1, \dots, p_m)$  of  $m$  distinct points in  $\mathcal{Q}$ , the expected distance between a random point  $x$ , generated according to a probability density function  $f$ , and the closest point in  $P$  is given by

$$\begin{aligned} H_m(P, \mathcal{Q}) &\doteq \mathbb{E} \left[ \min_{k \in \{1, \dots, m\}} \|p_k - x\| \right] \\ &= \sum_{k=1}^m \int_{\mathcal{V}_k(P, \mathcal{Q})} \|p_k - x\| f(x) dx, \end{aligned}$$

where  $\mathcal{V}(P, \mathcal{Q}) = (\mathcal{V}_1(P, \mathcal{Q}), \dots, \mathcal{V}_m(P, \mathcal{Q}))$  is the Voronoi partition of the set  $\mathcal{Q}$  generated by the points  $P$ . In other words,  $x \in \mathcal{V}_k(P, \mathcal{Q})$  if and only if  $\|x - p_k\| \leq \|x - p_j\|$ , for all  $j \in \{1, \dots, m\}$ . The set  $\mathcal{V}_k$  is referred to as the Voronoi cell of the generator  $p_k$ . The function  $H_m$  is known in the locational optimization literature as the continuous Weber function or the continuous multi-median function; see [18] and references therein.

The  $m$ -median of the set  $\mathcal{Q}$ , with respect to the measure induced by  $f$ , is the global minimizer  $P_m^*(\mathcal{Q}) = \arg \min_{P \in \mathcal{Q}^m} H_m(P, \mathcal{Q})$ . We let  $H_m^*(\mathcal{Q}) = H_m(P_m^*(\mathcal{Q}), \mathcal{Q})$  be the global minimum of  $H_m$ . It is straightforward to show that the map  $P \mapsto H_1(P, \mathcal{Q})$  is differentiable and strictly convex on  $\mathcal{Q}$ . Therefore, it is a simple computational task to compute  $P_1^*(\mathcal{Q})$ . We refer to  $P_1^*(\mathcal{Q})$  as the median of  $\mathcal{Q}$ . On the other hand, when  $m > 1$ , the map  $P \mapsto H_m(P, \mathcal{Q})$  is differentiable (whenever  $(p_1, \dots, p_m)$  are distinct) but not convex, thus making the solution of the continuous  $m$ -median problem hard in the general case. The set of critical points of  $H_m$  contains all configurations  $(p_1, \dots, p_m)$  with the property that each point  $p_k$  is the generator of the Voronoi cell  $\mathcal{V}_k(P, \mathcal{Q})$  as well as the median of  $\mathcal{V}_k(P, \mathcal{Q})$  (we refer to such Voronoi diagrams as *median Voronoi diagrams*); in particular, the global minimum of  $H_m$  must be one of these configurations.

## III. PROBLEM FORMULATION AND EXISTING RESULTS

In this section we formulate the problem we wish to address and we review existing results.

### A. Problem Formulation

The problem we consider in this paper is known as the  $m$ -vehicle Dynamic Traveling Repairman Problem ( $m$ -DTRP), and was introduced by Psaraftis in [11] and later studied by Bertsimas and van Ryzin in [8], [9], [12].

Let the workspace  $\mathcal{Q} \subset \mathbb{R}^d$  be a compact, convex set, and let  $\|\cdot\|$  denote the Euclidean norm in  $\mathbb{R}^d$ . We define the

<sup>1</sup>linkern is freely available for academic research use at [www.tsp.gatech.edu/concorde/index.html](http://www.tsp.gatech.edu/concorde/index.html).

diameter of  $\mathcal{Q}$  as:  $\text{diam}(\mathcal{Q}) \doteq \sup\{\|p - q\| \mid p, q \in \mathcal{Q}\}$ . For simplicity, we will only consider the planar case, i.e.,  $d = 2$ , with the understanding that extensions to higher dimensions are possible. Consider  $m$  holonomic vehicles, modeled as point masses, and let  $P(t) = (p_1(t), \dots, p_m(t)) \in \mathcal{Q}^m$  describe the locations of the vehicles at time  $t$ . The vehicles are free to move, traveling at a constant velocity  $v$ , within the environment  $\mathcal{Q}$ . The vehicles are identical, and have unlimited range and demand servicing capacity.

Demands are generated according to a homogeneous (i.e., time-invariant) spatio-temporal Poisson process<sup>2</sup>, with time intensity  $\lambda \in \mathbb{R}_{>0}$ , and spatial density  $f : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$ . In other words, demands arrive to  $\mathcal{Q}$  according to a Poisson process with intensity  $\lambda$  and their locations  $\{X_j; j \geq 1\}$  are i.i.d. (i.e., independent and identically distributed) and distributed according to a density  $f$  whose support is  $\mathcal{Q}$ . A demand's location becomes known (is realized) at its arrival epoch; thus, at time  $t$  vehicles know with *certainty* the locations of demands that arrived prior to time  $t$ , but future demand locations form an i.i.d. sequence. The density  $f$  satisfies:

$$\mathbb{P}[X_j \in \mathcal{S}] = \int_{\mathcal{S}} f(x) dx \quad \forall \mathcal{S} \subseteq \mathcal{Q}, \quad \text{and} \quad \int_{\mathcal{Q}} f(x) dx = 1.$$

We assume that the number of initial outstanding demands is a random variable with finite first and second moments.

At each demand location  $X_j$  vehicles spend some time  $s_j$  in on-site service that is i.i.d. and generally distributed with finite first and second moments denoted by  $\bar{s}$  and  $\bar{s}^2$  (we also assume  $\bar{s} > 0$ ; we will discuss in Section VII the case  $\bar{s} = 0$ ). A realized demand is removed from the system after one of the vehicles has completed its on-site service. We define the *load factor*  $\rho \doteq \lambda \bar{s} / m$ .

The system time of demand  $j$  (demands are labeled in an increasing order with respect to time of arrival), denoted by  $T_j$ , is defined as the elapsed time between the arrival of demand  $j$  and the time one of the vehicles completes its service. The waiting time of demand  $j$ ,  $W_j$ , is defined by  $W_j = T_j - s_j$ . The steady-state system time is defined by  $\bar{T} \doteq \limsup_{j \rightarrow \infty} \mathbb{E}[T_j]$ . A policy for routing the vehicles is said to be *stable* if the expected number of demands in the system is uniformly bounded at all times. A necessary condition for the existence of a stable policy is that  $\rho < 1$ ; we shall assume  $\rho < 1$  throughout the paper. When we refer to *light-load* conditions, we consider the case  $\rho \rightarrow 0^+$ , in the sense that  $\lambda \rightarrow 0^+$ ; when we refer to *heavy-load* conditions, we consider the case  $\rho \rightarrow 1^-$ , in the sense that  $\lambda \rightarrow [m/\bar{s}]^-$ .

Let  $\mathcal{P}$  be the set of all causal, stable and stationary routing policies. Letting  $\bar{T}_\pi$  denote the system time of a particular policy  $\pi \in \mathcal{P}$ , the  $m$ -DTRP is then defined as the problem of finding a policy  $\pi^*$  (if one exists) such that

$$\bar{T}_{\pi^*} = \inf_{\pi \in \mathcal{P}} \bar{T}_\pi.$$

<sup>2</sup>There are three main reasons behind the choice of modeling the demand arrival process with a Poisson process: First, the system we deal with is a spatial queueing system, and the arrival process to a queueing system is often well approximated by a Poisson process (see, e.g., [19]). Second, the Poisson process leads to a model that is analytically tractable. Third, the Poisson process, being "the most random" way to distribute events in time, leads to useful "worst-case" scenario evaluations (worst-case with respect to the possible types of arrival processes).

We let  $\bar{T}^*$  denote the infimum in the right hand side above.

We finally need two definitions; in these definitions,  $X$  is the location of a randomly chosen demand and  $W$  is its waiting time.

*Definition 3.1:* A policy  $\pi$  is called *spatially unbiased* if for every pair of sets  $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{Q}$  it holds  $\mathbb{E}[W \mid X \in \mathcal{S}_1] = \mathbb{E}[W \mid X \in \mathcal{S}_2]$ ; a policy  $\pi$  is instead called *spatially biased* if there exist sets  $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{Q}$  such that  $\mathbb{E}[W \mid X \in \mathcal{S}_1] > \mathbb{E}[W \mid X \in \mathcal{S}_2]$ .

## B. Existing Results for the $m$ -DTRP

As in many queueing problems, the analysis of the  $m$ -DTRP for all values of  $\rho \in (0, 1)$  is difficult. In [8], [9], [12], [20], lower bounds for the optimal steady-state system time are derived for the light-load case (i.e.,  $\rho \rightarrow 0^+$ ), and for the heavy-load case (i.e.,  $\rho \rightarrow 1^-$ ). Subsequently, policies are designed for these two limiting regimes, and their performance is compared to the lower bounds.

1) *Lower Bounds:* For the light-load case, a tight lower bound on the system time is derived in [12]. In the light-load case, the lower bound on the system time is strongly related to the solution of the multi-median problem:

$$\bar{T}^* \geq \frac{1}{v} H_m^*(\mathcal{Q}) + \bar{s}, \quad \text{as } \rho \rightarrow 0^+. \quad (3)$$

This bound is tight: there exist policies whose system times, in the limit  $\rho \rightarrow 0^+$ , attain this bound; we present such asymptotically optimal policies in Section III-B2.

The following two lower bounds hold in the heavy-load case [8], [20]. Within the class in  $\mathcal{P}$  of *spatially-unbiased* policies the optimal system time  $\bar{T}_U^*$  is lower bounded by

$$\bar{T}_U^* \geq \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m^2 v^2 (1 - \rho)^2}, \quad \text{as } \rho \rightarrow 1^-. \quad (4)$$

Within the class in  $\mathcal{P}$  of *spatially-biased* policies the optimal system time  $\bar{T}_B^*$  is lower bounded by

$$\bar{T}_B^* \geq \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{2/3}(x) dx \right]^3}{m^2 v^2 (1 - \rho)^2}, \quad \text{as } \rho \rightarrow 1^-. \quad (5)$$

Both bounds (4) and (5) are tight: there exist policies whose system times attain these bounds, in the limit  $\rho \rightarrow 1^-$ . Therefore the inequalities in (4) and (5) should indeed be replaced with equalities. We present asymptotically optimal policies for the heavy-load case in Section III-B3. It is shown in [8] that the lower bound in equation (5) is always lower than or equal to the lower bound in equation (4) for all densities  $f$ .

*Remark 3.2:* In equations (4) and (5), the right-hand side approaches  $+\infty$  as  $\rho \rightarrow 1^-$ . Thus, one should more formally write the inequalities with  $\bar{T}^*(1 - \rho)^2$  on the left-hand side, so that the right-hand side is finite. However, this makes the presentation less readable, and thus, in the remainder of the paper, we adhere to the less formal but more transparent style of equations (4) and (5). Furthermore, this is the style generally used in the literature on the  $m$ -DTRP (see, e.g., [8] Section 4 and [21] Proposition 2).

2) *Optimal Policies for the Light-Load Case:* For the light-load case, to the best of our knowledge, there is only one policy that is known to be optimal; this policy is as follows.

**The  $m$  Stochastic Queue Median (SQM) Policy** [12] — Locate one vehicle at each of the  $m$ -median locations for the environment  $\mathcal{Q}$ . When demands arrive, assign them to the nearest median location and its corresponding vehicle. Have each vehicle service its respective demands in First-Come, First-Served (FCFS) order returning to its median after each service is completed.

This policy, although optimal in light load, has two characteristics that limit its application to large-scale vehicle networks: First, it quickly becomes unstable as the load increases, i.e., there exists  $\varrho_c < 1$  such that for all  $\varrho > \varrho_c$  the system time  $\bar{T}_{\text{SQM}}$  is infinite (hence, this policy is not adaptive with respect to load conditions). Second, a *central entity* needs to compute the  $m$ -median locations and assign them to the vehicles (hence, it is centralized).

3) *Optimal Policies for the Heavy-Load Case:* For the heavy-load case, to the best of our knowledge, there is only one spatially-unbiased policy that has been proven to achieve the lower bound in equation (4); this policy is as follows.

**The Unbiased TSP (UTSP) Policy** [8] — Let  $r$  be a fixed positive, large integer. From a central point in the interior of  $\mathcal{Q}$ , subdivide the service region into  $r$  wedges  $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_r$  such that  $\int_{\mathcal{Q}_k} f(x) dx = 1/r$ ,  $k \in \{1, \dots, r\}$ . Within each subregion, form sets of size  $n/r$  ( $n$  is a design parameter). As sets are formed, deposit them in a queue and service them FCFS with the first available vehicle by forming a TSP tour on the set and following it in an arbitrary direction. Optimize over  $n$ .

It is possible to show [8] that  $\lim_{r \rightarrow \infty} \lim_{\varrho \rightarrow 1^-} \bar{T}_{\text{UTSP}}(r) / \bar{T}_{\text{U}}^* = 1$ , hence the UTSP policy with  $r \rightarrow +\infty$  is an optimal spatially-unbiased policy in heavy load.

The same paper [8] presents an optimal spatially-biased policy. This policy, called Biased TSP (BTSP) Policy, relies on an even finer partition of the environment, requires  $f$  to be piecewise constant, and also depends on a parameter  $n$ .

Although both the UTSP and the BTSP policies are optimal within their respective classes, they have two major drawbacks: First, in the UTSP policy, to ensure stability,  $n$  should be chosen so that (see [8], page 961)  $n > \lambda^2 \beta_{\text{TSP},2}^2 \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2 / (m^2 v^2 (1 - \varrho)^2)$ ; therefore, to ensure stability over a wide range of values of  $\varrho$ , the system designer is forced to select a *large* value for  $n$ . However, if during the execution of the policy the load factor turns out to be only moderate, demands have to wait for an excessively large set to be formed, and the overall system performance deteriorates significantly. Similar considerations hold for the BTSP policy. Hence, these two policies are not adaptive with respect to load conditions. Second, both policies require a centralized data structure (the demands' queue is shared by the vehicles); hence, both policies are centralized.

### C. Objective of the Paper

In many applications it is desirable to provide spatially-unbiased service, i.e., the same quality of service to all of the outstanding demands independently of their locations; e.g., in the UAV example discussed in the introduction, outstanding “threats” at the periphery of  $\mathcal{Q}$  should receive the same quality of service as the outstanding “threats” in the middle of  $\mathcal{Q}$ . However, in light load, when most of the time the vehicles are idling, it is more reasonable to consider spatially-biased policies that aim at positioning the vehicles in the best a priori locations, i.e., at the locations that minimize the expected distance to the next demand.

Accordingly, in this paper we focus on policies for the  $m$ -DTRP that are constrained to provide spatially-unbiased service *in heavy load*, but can provide spatially-biased service in light load. With a slight abuse, we refer to policies satisfying this constraint as *unbiased policies*. The objective is to devise *adaptive, distributed, scalable* unbiased policies with *provable performance guarantees*, that rely only on local exchange of information between neighboring vehicles, and do not require explicit knowledge of the global structure of the network.

The key idea that we will pursue in this paper is that of *partitioning policies*. Given a policy  $\pi$  for the 1-DTRP and  $m$  vehicles, a  $\pi$ -partitioning policy is a multi-vehicle policy such that 1) the environment  $\mathcal{Q}$  is partitioned according to some partitioning scheme into  $m$  openly disjoint subregions  $\mathcal{Q}_k$ ,  $k \in \{1, \dots, m\}$ , whose union is  $\mathcal{Q}$ , 2) one vehicle is assigned to each subregion (thus, there is a one-to-one correspondence between vehicles and subregions), and 3) each vehicle executes the single-vehicle policy  $\pi$  to service demands that fall within its own subregion. The SQM policy, which is optimal *in light load*, is indeed a partitioning policy whereby  $\mathcal{Q}$  is partitioned according to a median Voronoi diagram that “minimizes”  $H_m$ , and each vehicle executes inside its own Voronoi cell the single-vehicle policy “service FCFS and return to the assigned median after each service completion”. Moreover, in Section VI, we will prove that *in heavy load*, given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, there exists an unbiased  $\pi^*$ -partitioning policy that is optimal.

The above discussion leads to the following strategy: First, for the 1-DTRP, we design unbiased and adaptive control policies with provable performance guarantees. Then, by using recently developed *distributed* algorithms for environment partitioning, we extend these single-vehicle policies to distributed multi-vehicle policies.

## IV. THE SINGLE-VEHICLE DIVIDE & CONQUER POLICY

We describe the Divide & Conquer (DC) policy in Algorithm 1, where  $D$  is the set of outstanding demands waiting for service and  $p$  is the current position of the vehicle. An  $r$ -partition that is simultaneously equitable with respect to  $f$  and  $f^{1/2}$  is guaranteed to exist (see discussion on equitable partitions in Section II). The number of subregions  $r$  is a design parameter whose choice will be discussed after the analysis of the policy. One should note that  $\hat{P}_1^*$  (which is defined in Algorithm 1) is indeed the empirical median. Next, we characterize the DC policy. One can easily show that the DC policy is *unbiased*.

---

**Algorithm 1: Divide & Conquer (DC) Policy**


---

**Assumes:** An  $r$ -partition  $\{\mathcal{Q}_k\}_{k=1}^r$  of  $\mathcal{Q}$  that is simultaneously equitable with respect to  $f$  and  $f^{1/2}$ .

```

1 if  $D = \emptyset$  then
2   Let  $\tilde{P}_1^*$  be the point minimizing the sum of distances to
   demands serviced in the past (the initial condition for  $\tilde{P}_1^*$  is
   a random point in  $\mathcal{Q}$ ); if  $p \neq \tilde{P}_1^*$ , move to  $\tilde{P}_1^*$ , otherwise
   stop.
3 else
4    $k \leftarrow k_0$  ( $k_0$  is the subregion with the largest number of
   outstanding demands).
5   repeat
6     Move to the median of subregion  $\mathcal{Q}_k$ .
7     Compute a TSP tour through all demands in subregion
      $\mathcal{Q}_k$ . Service all demands by following the TSP tour,
     starting at a random demand on the TSP tour.
8      $k \leftarrow k + 1$  modulo  $r$ .
9   until  $D = \emptyset$ 
10 Repeat.

```

---

### A. Analysis of the DC Policy in Light Load

We first study the light-load case (i.e.,  $\varrho \rightarrow 0^+$ ).

*Theorem 4.1 (Performance of DC policy in light load):*

As  $\varrho \rightarrow 0^+$ , for every  $r$ , the DC policy is asymptotically optimal, that is,  $\bar{T}_{\text{DC}}(r) \rightarrow \bar{T}^*$ , as  $\varrho \rightarrow 0^+$ .

*Proof:* Using arguments similar to those in [21], consider generic initial conditions in  $\mathcal{Q}$  for the vehicle's position and for the positions of the outstanding demands (denote the initial number of demands with  $n_0$ ). Under the DC policy, an upper bound on the time  $C_0$  needed to service all of the initial demands and move to the empirical median  $\tilde{P}_1^*$  is:  $C_0 \leq (n_0 + 1) \text{diam}(\mathcal{Q})/v + \sum_{j=1}^{n_0} s_j$ . Thus, we have  $\mathbb{E}[C_0] \leq \mathbb{E}[n_0](\text{diam}(\mathcal{Q})/v + \bar{s}) + \text{diam}(\mathcal{Q})/v$ . Given the time interval  $C_0 = t$ ,  $t \in \mathbb{R}_{\geq 0}$ , the probability that no new demand arrives during such time interval is  $\mathbb{P}(N(C_0) = 0 | C_0 = t) = \exp(-\lambda t)$ , where  $N(t)$  is the counting variable associated to a Poisson process with rate  $\lambda$ . Hence, by applying the law of total probability and Jensen's inequality for convex functions one obtains:  $\mathbb{P}[N(C_0) = 0] = \int_0^\infty \mathbb{P}(N(C_0) = 0 | C_0 = t) f_{C_0}(t) dt = \int_0^\infty \exp(-\lambda t) f_{C_0}(t) dt \geq \exp\left(-\lambda \int_0^\infty t f_{C_0}(t) dt\right) = \exp\left(-\lambda \mathbb{E}[C_0]\right) \geq \exp\left(-\lambda \mathbb{E}[n_0] \left(\frac{\text{diam}(\mathcal{Q})}{v} + \bar{s}\right) - \lambda \frac{\text{diam}(\mathcal{Q})}{v}\right)$ , where  $f_{C_0}$  is the density function of random variable  $C_0$ . Since, by assumption,  $\mathbb{E}[n_0] < \infty$  and  $\varrho \rightarrow 0^+$  (i.e.,  $\lambda \rightarrow 0^+$ ), we obtain  $\mathbb{P}[N(C_0) = 0] \rightarrow 1^-$  as  $\varrho \rightarrow 0^+$ . As a consequence, after an initial transient, all demands will be generated with the vehicle at the empirical median  $\tilde{P}_1^*$  and with an empty demand queue, with probability 1. Then, in light load, the DC policy becomes identical to the light-load routing policy presented in [22]; since such routing policy is optimal, in light load, for the 1-DTRP [22], we obtain the desired result (the proof in [22] basically shows that  $\tilde{P}_1^* \rightarrow P_1^*(\mathcal{Q})$ ). ■

### B. Analysis of the DC Policy in Heavy Load

Next, we consider the DC policy in heavy load, i.e., when  $\varrho \rightarrow 1^-$ . The performance of the DC policy in heavy load is characterized by the following theorem.

*Theorem 4.2 (Performance of DC policy in heavy load):*

As  $\varrho \rightarrow 1^-$ , the system time for the DC policy satisfies

$$\bar{T}_{\text{DC}}(r) \leq \left(1 + \frac{1}{r}\right) \frac{\beta_{\text{TSP},2}^2 \lambda \left[\int_{\mathcal{Q}} f^{1/2}(x) dx\right]^2}{2 v^2 (1 - \varrho)^2}, \quad (6)$$

where  $r$  is the number of subregions.

The proof of Theorem 4.2 relies on techniques similar to those we developed in [23], [24], and builds on a number of intermediate results. We start with the following lemma, similar to Lemma 1 in [21], characterizing the number of outstanding demands in heavy load.

*Lemma 4.3 (Number of demands in heavy load):* In heavy load (i.e.,  $\varrho \rightarrow 1^-$ ), after a transient, the number of demands serviced in a single tour of the vehicle in the DC policy is very large with high probability (i.e., the number of demands tends to  $+\infty$  with probability that tends to 1, as  $\varrho$  approaches  $1^-$ ).

*Proof:* Consider the case where the vehicle moves with infinite velocity (i.e.,  $v \rightarrow +\infty$ ); then the system is reduced to an M/G/1 queue (i.e., a queue with exponentially distributed inter-arrival times, generally distributed service times, and a single server). The infinite-velocity system has fewer demands waiting in queue. A known result on M/G/1 queues [25] states that, after transients, the total number of demands, as  $\varrho \rightarrow 1^-$ , is very large with high probability. Due to the way  $\mathcal{Q}$  is partitioned, the number of outstanding demands in each subregion is also very large with high probability. In particular, after transients, the number of demands is very large with high probability at the time instants when the vehicle starts a new TSP tour. ■

Lemma 4.3 and its proof imply that in heavy load  $D \neq \emptyset$  with high probability. Accordingly, in the following heavy-load analysis we consider the condition  $D = \emptyset$  always false, and the DC policy then reduces to lines 5-9 of Algorithm 1.

We refer to the time instant  $t_i$ ,  $i \geq 0$ , in which the vehicle starts a new TSP tour in subregion 1 as the epoch  $i$  of the policy; we refer to the time interval between epoch  $i$  and epoch  $i + 1$  as the  $i$ th iteration. Let  $n_i^k$ ,  $k \in \{1, \dots, r\}$ , be the number of outstanding demands serviced in subregion  $k$  during iteration  $i$ . Finally, let  $C_i^k$  be the time interval between the time instant the vehicle starts to service demands in subregion  $k$  during iteration  $i$  and the time instant the vehicle starts to service demands in the same subregion  $k$  during next iteration  $i + 1$ . Demands arrive in subregion  $\mathcal{Q}_k$  according to a Poisson process with rate  $\hat{\lambda} \doteq \lambda/r$ , where we use the fact that the partition  $\{\mathcal{Q}_k\}_{k=1}^r$  is equitable with respect to  $f$ ; then, we have  $\mathbb{E}[n_{i+1}^k] = \hat{\lambda} \mathbb{E}[C_i^k]$ . The time interval  $C_i^k$  is the sum of three components: 1) the time for the  $r$  travels from the end of one TSP tour to the start of next one; 2) the time to perform  $r$  TSP tours; and 3) the time to provide on-site service while performing the  $r$  TSP tours.

The first component is trivially upper bounded by  $2 \text{diam}(\mathcal{Q})/v \cdot r$ . Given that a demand falls in subregion  $\mathcal{Q}_k$ ,

the conditional density for its location (whose support is  $\mathcal{Q}_k$ ) is  $f(x)/\left(\int_{\mathcal{Q}_k} f(x) dx\right)$ . By equation (1), we have that the expected length of a TSP tour through  $n$  demands independently distributed according to the density  $f(x)/\left(\int_{\mathcal{Q}_k} f(x) dx\right)$  satisfies (with a slight abuse of notation, we call such length  $\text{TSP}(n)$ )

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{\mathbb{E}[\text{TSP}(n)]}{\sqrt{n}} &= \beta_{\text{TSP},2} \left( \int_{\mathcal{Q}_k} \sqrt{\frac{f(x)}{\int_{\mathcal{Q}_k} f(x) dx}} dx \right) \\ &= \beta_{\text{TSP},2} \sqrt{r} \frac{\int_{\mathcal{Q}} f^{1/2}(x) dx}{r} \doteq \beta, \end{aligned} \quad (7)$$

where we have exploited the fact that, by definition of the DC policy, the partition  $\{\mathcal{Q}_k\}_{k=1}^r$  is simultaneously equitable with respect to  $f$  and  $f^{1/2}$ . Pick now an arbitrarily small  $\zeta > 0$ . By equation (7), there exists an  $\tilde{n} \in \mathbb{N}_{>0}$  such that for all  $n > \tilde{n}$  it holds  $\mathbb{E}[\text{TSP}(n)] \leq (\beta + \zeta) \sqrt{n}$ . Then, the expected length of a TSP tour through  $n_i^k$  demands (with a slight abuse of notation, we call such length  $\text{TSP}(n_i^k)$ ) can be upper bounded as:

$$\begin{aligned} \mathbb{E}[\text{TSP}(n_i^k)] &= \sum_{n=0}^{+\infty} \mathbb{E}[\text{TSP}(n_i^k) | n_i^k = n] \mathbb{P}[n_i^k = n] \\ &\leq \sum_{n=\tilde{n}+1}^{+\infty} (\beta + \zeta) \sqrt{n} \mathbb{P}[n_i^k = n] + \sum_{n=0}^{\tilde{n}} \tilde{n} \text{diam}(\mathcal{Q}) \mathbb{P}[n_i^k = n] \\ &\leq (\beta + \zeta) \mathbb{E}[\sqrt{n_i^k}] + \tilde{n} \text{diam}(\mathcal{Q}) \mathbb{P}[n_i^k \leq \tilde{n}] \\ &\leq (\beta + \zeta) \sqrt{\mathbb{E}[n_i^k]} + \tilde{n} \text{diam}(\mathcal{Q}) \mathbb{P}[n_i^k \leq \tilde{n}], \end{aligned}$$

where in the first equality we have used the law of total expectation, and in the last inequality we have applied Jensen's inequality for concave functions in the form  $\mathbb{E}[\sqrt{X}] \leq \sqrt{\mathbb{E}[X]}$ . By Lemma 4.3, there exists  $\bar{\varrho} \in (0, 1)$  such that for all  $\varrho \in [\bar{\varrho}, 1)$  there exists an integer  $\bar{i}(\varrho)$  (possibly dependent on  $\varrho$ ) such that  $\mathbb{P}[n_i^k \leq \tilde{n}] < 1/\tilde{n}$ . Consider any  $\varrho \in [\bar{\varrho}, 1)$  and assume that  $i \geq \bar{i}(\varrho)$ , then  $\mathbb{E}[\text{TSP}(n_i^k)] \leq (\beta + \zeta) \sqrt{\mathbb{E}[n_i^k]} + \text{diam}(\mathcal{Q})$ , for  $i \geq \bar{i}(\varrho)$  and  $\varrho \in [\bar{\varrho}, 1)$ .

Finally, since on-site service times are independent of the number of outstanding demands, the expected on-site service requirement for  $n_i^k$  demands is  $\bar{s} \mathbb{E}[n_i^k]$ .

Then, we obtain the following recurrence relation (where we define  $\bar{n}_i^k \doteq \mathbb{E}[n_i^k]$ ):

$$\begin{aligned} \bar{n}_{i+1}^k &= \hat{\lambda} \mathbb{E}[C_i^k] \leq \hat{\lambda} \left( \text{diam}(\mathcal{Q}) \frac{3r}{v} + \sum_{j=k}^r \left( \bar{s} \bar{n}_i^j + \frac{\beta + \zeta}{v} \sqrt{\bar{n}_i^j} \right) \right. \\ &\quad \left. + \sum_{j=1}^{k-1} \left( \bar{s} \bar{n}_{i+1}^j + \frac{\beta + \zeta}{v} \sqrt{\bar{n}_{i+1}^j} \right) \right), \end{aligned} \quad (8)$$

where  $k \in \{1, \dots, r\}$ ,  $i \geq \bar{i}(\varrho)$ , and  $\varrho \in [\bar{\varrho}, 1)$ .

The  $r$  inequalities above describe a system of recurrence relations that allows us to find an upper bound on  $\limsup_{i \rightarrow +\infty} \bar{n}_i^k$ . The following lemma bounds the values to which the limits  $\limsup_{i \rightarrow +\infty} \bar{n}_i^k$  converge.

**Lemma 4.4 (Bound on steady-state number of demands):**

In heavy load, for every set of initial conditions

$\{\bar{n}_1^k\}_{k \in \{1, \dots, r\}}$ , the trajectories  $i \mapsto \bar{n}_i^k$ ,  $k \in \{1, \dots, r\}$ , satisfy

$$\bar{n}_i^k \doteq \limsup_{i \rightarrow +\infty} \bar{n}_i^k \leq \beta_{\text{TSP},2}^2 \frac{\lambda^2 \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{r v^2 (1 - \varrho)^2}.$$

*Proof:* Define the constants:  $\tilde{\beta} \doteq \beta + \zeta$  and  $a \doteq \text{diam}(\mathcal{Q}) 3r/v$ . Henceforth we assume that  $\varrho \in [\bar{\varrho}, 1)$  and that  $i \geq \bar{i}(\varrho)$ , and we simply denote  $\bar{i}(\varrho)$  as  $\bar{i}$ . Next we define two auxiliary systems, System-Y and System-Z, whose trajectories will be used to bound the trajectories  $i \mapsto \bar{n}_i^k$ . We define System-Y (with state  $y(i) \in \mathbb{R}^r$ ) as

$$\begin{aligned} y_k(i+1) &= \hat{\lambda} a + \hat{\lambda} \left( \sum_{j=k}^r \left( \bar{s} y_j(i) + \frac{\tilde{\beta}}{v} \sqrt{y_j(i)} \right) \right. \\ &\quad \left. + \sum_{j=1}^{k-1} \left( \bar{s} y_j(i+1) + \frac{\tilde{\beta}}{v} \sqrt{y_j(i+1)} \right) \right), \text{ for } i \geq \bar{i}, \end{aligned} \quad (9)$$

with  $y_k(\bar{i}) = \bar{n}_{\bar{i}}^k$ , and where  $k \in \{1, \dots, r\}$ . System-Y is obtained by replacing the inequality in equation (8) with an equality. Pick, now, any  $\varepsilon > 0$ . From Young's inequality  $\sqrt{w} \leq 1/(4\varepsilon) + \varepsilon w$ , for all  $w \in \mathbb{R}_{\geq 0}$ . By applying Young's inequality in equation (9) we obtain

$$\begin{aligned} y_k(i+1) &\leq \hat{\lambda} \left( \bar{s} + \varepsilon \frac{\tilde{\beta}}{v} \right) \left( \sum_{j=k}^r y_j(i) + \sum_{j=1}^{k-1} y_j(i+1) \right) \\ &\quad + \hat{\lambda} \left( \frac{r \tilde{\beta}}{4v\varepsilon} + a \right), \quad k \in \{1, \dots, r\}. \end{aligned} \quad (10)$$

Next, define System-Z (with state  $z(i) \in \mathbb{R}^r$ ) as

$$\begin{aligned} z_k(i+1) &= \hat{\lambda} \left( \bar{s} + \varepsilon \frac{\tilde{\beta}}{v} \right) \left( \sum_{j=k}^r z_j(i) + \sum_{j=1}^{k-1} z_j(i+1) \right) \\ &\quad + \hat{\lambda} \left( \frac{r \tilde{\beta}}{4v\varepsilon} + a \right), \quad \text{for } i \geq \bar{i}, \end{aligned} \quad (11)$$

with  $z_k(\bar{i}) = \bar{n}_{\bar{i}}^k$ , and where  $k \in \{1, \dots, r\}$ . The proof now proceeds as follows. First, we show that the condition  $\bar{n}_{\bar{i}}^k = y_k(\bar{i}) = z_k(\bar{i})$  for all  $k$  implies that

$$\bar{n}_i^k \leq y_k(i) \leq z_k(i), \quad \text{for all } k \text{ and for all } i \geq \bar{i}. \quad (12)$$

Second, we show that the trajectories of System-Z are bounded; this fact, together with equation (12), implies that also trajectories of variables  $\bar{n}_i^k$  and  $y_k(i)$  are bounded. Third, and last, we will compute  $\limsup_{i \rightarrow +\infty} y_k(i)$ ; this quantity, together with equation (12), will yield the desired result.

We prove the first fact by induction on  $i$ , i.e., we prove that if  $\bar{n}_i^k \leq y_k(i) \leq z_k(i)$  for all  $k \in \{1, \dots, r\}$ , then  $\bar{n}_{i+1}^k \leq y_k(i+1) \leq z_k(i+1)$  for all  $k \in \{1, \dots, r\}$ , where  $i \geq \bar{i}$ . Indeed, it is immediate to show that, under the assumptions, it holds:  $\bar{n}_{i+1}^1 \leq y_1(i+1) \leq z_1(i+1)$ ; moreover, it is easy to show that, under the assumptions, and if  $\bar{n}_{i+1}^j \leq y_j(i+1) \leq z_j(i+1)$  for all  $j \leq k < r$ , then  $\bar{n}_{i+1}^{k+1} \leq y_{k+1}(i+1) \leq z_{k+1}(i+1)$ . The inductive step then follows immediately.



We now turn our attention to the second issue, namely boundedness of trajectories for System-Z (described in equation (11)). Define  $\delta \doteq \hat{\lambda}(\bar{s} + \varepsilon \tilde{\beta}/v)$ . System-Z is a discrete-time system and can be rewritten (by adding and subtracting  $z_{k-1}(i)$  in the second term between parentheses in equation (11), when  $k > 1$ ) as

$$z_k(i+1) = \begin{cases} \delta \sum_{j=1}^r z_j(i) + \hat{\lambda} \left( \frac{r \tilde{\beta}}{4v\varepsilon} + a \right), & \text{for } k=1, \\ (1+\delta)z_{k-1}(i+1) - \delta z_{k-1}(i), & \text{for } k \in \{2, \dots, r\}. \end{cases}$$

By using the above equation and by simple inductive arguments, it is possible to show that System-Z can be written in canonical form as

$$z_k(i+1) = (1+\delta)^{k-1} \left( \delta \sum_{j=1}^r z_j(i) + \hat{\lambda} \left( \frac{r \tilde{\beta}}{4v\varepsilon} + a \right) \right) - \sum_{j=1}^{k-1} \delta (1+\delta)^{k-j-1} z_j(i), \quad k \in \{1, \dots, r\}.$$

Hence, System-Z can be written in compact form as  $z(i+1) = (A + B(\varepsilon))z(i) + c(\varepsilon)$ , where  $A$  is a matrix in  $\mathbb{R}^{r \times r}$  and is independent of  $\varepsilon$ ,  $B(\varepsilon) \in \mathbb{R}^{r \times r}$  and each element of  $B(\varepsilon)$ , say  $b(\varepsilon)_{kj}$ , has the property  $\lim_{\varepsilon \rightarrow 0^+} b(\varepsilon)_{kj} = 0$  for all  $k, j \in \{1, \dots, r\}$ , and  $c(\varepsilon) \in \mathbb{R}^{r \times 1}$  (notice that  $c(\varepsilon)$  is well defined for every  $\varepsilon > 0$ ). It is easy to see that the entries of matrix  $A$  are

$$a_{kj} = \begin{cases} \hat{\varrho} \left[ (1+\hat{\varrho})^{k-1} - (1+\hat{\varrho})^{k-j-1} \right], & \text{for } 1 \leq j \leq k-1, \\ \hat{\varrho} (1+\hat{\varrho})^{k-1}, & \text{otherwise,} \end{cases}$$

where  $\hat{\varrho} \doteq \hat{\lambda} \bar{s}$ . Notice that, since  $\hat{\varrho} > 0$ , we have  $|a_{kj}| = a_{kj}$  for all  $k$  and  $j$ . Then, for each  $k \in \{1, \dots, r\}$ , we have

$$\begin{aligned} \sum_{j=1}^r |a_{kj}| &= \sum_{j=1}^r a_{kj} = r \hat{\varrho} (1+\hat{\varrho})^{k-1} - \hat{\varrho} \sum_{j=1}^{k-1} (1+\hat{\varrho})^{k-j-1} \\ &= (1+\hat{\varrho})^{k-1} \underbrace{(r \hat{\varrho} - 1)}_{< 0} + 1 < 1, \end{aligned}$$

since  $\hat{\varrho} = \lambda \bar{s}/r$  and  $\varrho = \lambda \bar{s} < 1$  by assumption. Therefore, the  $L^\infty$ -induced norm of matrix  $A$  satisfies:  $\|A\|_\infty = \max_k \sum_{j=1}^r |a_{kj}| < 1$ . Since any induced norm  $\|\cdot\|$  satisfies the inequality  $\rho(A) \leq \|A\|$ , where  $\rho(A)$  is the spectral radius of  $A$ , we conclude that  $A \in \mathbb{R}^{r \times r}$  has eigenvalues strictly inside the unit disk (i.e.,  $A$  is a stable matrix). Since the eigenvalues of a matrix depend continuously on the matrix entries, there exists a sufficiently small  $\varepsilon > 0$  such that the matrix  $A + B(\varepsilon)$  has eigenvalues strictly inside the unit disk. Accordingly, having selected a sufficiently small  $\varepsilon$ , the solution  $i \mapsto z(i) \in \mathbb{R}_{\geq 0}^r$  of System-Z converges exponentially fast to the unique equilibrium point  $z^*(\varepsilon) = (I_r - A - B(\varepsilon))^{-1} c(\varepsilon)$ , where  $I_r$  is the identity matrix of size  $r$ . Combining equation (12) with the previous statement, we see that the solutions  $i \mapsto \bar{n}(i)$  (where  $\bar{n}(i) \doteq (\bar{n}_1^i, \dots, \bar{n}_r^i)$ ) and  $i \mapsto y(i)$  are bounded. Thus

$$\limsup_{i \rightarrow +\infty} \bar{n}(i) \leq \limsup_{i \rightarrow +\infty} y(i) < +\infty. \quad (13)$$

Finally, we turn our attention to the third issue, namely the computation of the limit  $y \doteq \limsup_{i \rightarrow +\infty} y(i)$ . Taking the lim sup of the left- and right-hand sides of equation (9), and noting that, for  $j \in \{1, \dots, r\}$ ,  $\limsup_{i \rightarrow +\infty} \sqrt{y_j(i)} = \sqrt{\limsup_{i \rightarrow +\infty} y_j(i)}$ , since  $\sqrt{\cdot}$  is continuous and strictly monotone increasing on  $\mathbb{R}_{> 0}$ , we obtain that

$$y_k = \hat{\lambda} a + \hat{\lambda} \sum_{j=1}^r \left( \bar{s} y_j + \frac{\tilde{\beta}}{v} \sqrt{y_j} \right); \quad (14)$$

therefore we have  $y_k = y_j$  for all  $k, j \in \{1, \dots, r\}$ . Substituting in equation (14) and solving for  $y_k$  we obtain  $y_k = \frac{1}{4} \left( \frac{\lambda \tilde{\beta}}{v(1-\varrho)} + \sqrt{\frac{\lambda^2 \tilde{\beta}^2}{v^2(1-\varrho)^2} + \frac{4\hat{\lambda}a}{1-\varrho}} \right)^2$ . Recalling that this analysis holds for every  $\varrho \in [\bar{\varrho}, 1)$ , by taking  $\varrho$  arbitrarily close to 1 we obtain

$$y_k = \frac{\lambda^2 \tilde{\beta}^2}{v^2(1-\varrho)^2}, \quad \text{as } \varrho \rightarrow 1^-. \quad (15)$$

(Formally, we should have written  $\lim_{\varrho \rightarrow 1^-} y_k(1-\varrho)^2 = \lambda^2 \tilde{\beta}^2/v^2$ , however, as discussed in Remark 3.2, we prefer to adhere to the less formal but more transparent style of equation (15)). Noting that from equation (13)  $\limsup_{i \rightarrow +\infty} \bar{n}_i^k \leq y_k$ , and recalling that  $\tilde{\beta} = \frac{\beta_{\text{TSP},2} \int_{\mathcal{Q}} f^{1/2}(x) dx}{\sqrt{r}} + \zeta$ , where  $\zeta$  is an arbitrarily small constant, we obtain the desired result. ■

We are now in a position to prove Theorem 4.2

*Proof of Theorem 4.2:* Define  $\bar{C}^k \doteq \limsup_{i \rightarrow \infty} \mathbb{E}[C_i^k]$ ; then we have, by using the upper bound on  $\mathbb{E}[C_i^k]$  in equation (8) and neglecting constant terms,

$$\begin{aligned} \bar{C}^k &\doteq \limsup_{i \rightarrow \infty} \mathbb{E}[C_i^k] \leq \frac{\tilde{\beta}}{v} \sum_{j=1}^r \sqrt{\bar{n}^j} + \bar{s} \sum_{j=1}^r \bar{n}^j \\ &\leq \frac{r \lambda \tilde{\beta} \beta}{v^2(1-\varrho)} + \frac{r \varrho \lambda \beta^2}{v^2(1-\varrho)^2}. \end{aligned}$$

Hence, in the limit  $\varrho \rightarrow 1^-$ , we have

$$\bar{C}^k \leq \frac{r \lambda \beta^2}{v^2(1-\varrho)^2} = \frac{\lambda \beta_{\text{TSP},2}^2 \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{v^2(1-\varrho)^2}.$$

Consider a random demand that arrives in subregion  $k$ . Its expected steady-state system time,  $\bar{T}^k$ , will be upper bounded, as  $\varrho \rightarrow 1^-$ , by  $\bar{T}^k \leq \frac{1}{2} \bar{C}^k + \frac{1}{2} \bar{s} \bar{n}^k \leq \frac{1}{2} \bar{C}^k + \frac{1}{2} \frac{\lambda \beta^2}{v^2(1-\varrho)^2}$ , where we used the fact that, as  $\varrho \rightarrow 1^-$ , the travel time along a TSP tour is negligible compared to the on-site service time requirement. Unconditioning, we obtain the claim. ■

### C. Discussion

The DC policy is optimal in light load (Theorem 4.1); moreover, if we let  $r \rightarrow \infty$ , the DC policy achieves the heavy-load lower bound (4) for unbiased policies (Theorem 4.2). Therefore the DC policy is *both* optimal in light load *and* arbitrarily close to optimality in heavy load, and stabilizes the system in *every* load condition (since it stabilizes the system in the limit  $\varrho \rightarrow 1^-$ ). Notice that with  $r = 10$  the DC policy is already guaranteed to be within 10% of the optimal (for unbiased policies) performance in heavy load.

The implementation of the DC policy does *not* require the knowledge of  $\lambda$ ,  $\bar{s}$ , and  $v$ ; however, if one chooses  $r > 1$ , the computation of a simultaneously equitable  $r$ -partition of  $\mathcal{Q}$  requires the knowledge of  $f$ . Hence, if  $r \rightarrow +\infty$ , the policy is (i) provably optimal in both light and heavy load, and (ii) adaptive with respect to arrival rate  $\lambda$ , expected on-site service  $\bar{s}$ , and vehicle's velocity  $v$ . If, instead,  $r = 1$ , the policy is (i) provably optimal in light load and within a factor 2 of the optimal performance in heavy load, and (ii) adaptive with respect to arrival rate  $\lambda$ , expected on-site service  $\bar{s}$ , vehicle's velocity  $v$ , and spatial density  $f$ ; in other words, when  $r = 1$ , the DC policy adapts to *all* problem data. (These results should be compared with the discussion in Section III about the non-adaptive nature of the UTSP policy, which, for example, might become unstable if  $\lambda$  increases.)

An approximate algorithm to compute, with *arbitrarily small* error, an  $r$ -partition that is simultaneously equitable with respect to two measures can be found in [13] (specifically, see discussion on page 621); in the particular case when the density  $f$  is uniform, the problem of finding an  $r$ -partition that is simultaneously equitable with respect to  $f$  and  $f^{1/2}$  becomes trivial. Given a simultaneously equitable  $r$ -partition of  $\mathcal{Q}$ , the subregions can be indexed using the following procedure: (i) a TSP tour through the medians of the subregions is computed, (ii) subregions are indexed according to the order induced by this TSP tour. In practice, the DC policy would be implemented by allowing the vehicle to skip subregions with no demands. Note that if we assume that  $f$  is known (as it must be the case for the implementation of the DC policy with  $r > 1$ ), then instead of using the empirical median  $\tilde{P}_1^*$  one should directly use the exact median  $P_1^*$  (which is the solution of a strictly convex problem).

The DC policy requires on-line solutions of possibly large TSPs; therefore, practical implementations of this policy should rely on heuristics or on polynomial-time approximation schemes for the solution of TSPs, such as Lin-Kernighan's or Christofides' (see Section II-B). We will further discuss computation times in Section VIII.

Finally, the DC policy with  $r = 1$  is similar to the generation policy presented in [21]; in particular, the generation policy has the same performance guarantees of the DC policy with  $r = 1$ . However, the generation policy is analyzed only for the case of *uniform* spatial density  $f$ , and its implementation *does* require the knowledge of  $f$  (while, as discussed before, the DC policy with  $r = 1$  adapts to all problem data, including  $f$ ). The part of Algorithm 1 in lines 5-9 is similar to the PART-TSP policy presented in [20]; however, the PART-TSP policy is only *informally* analyzed by assuming steady-state conditions, in particular no proof of stability and convergence is provided.

In the next section we present and analyze another single-vehicle policy, the Receding Horizon (RH) policy, that applies ideas of receding-horizon control to dynamic vehicle routing problems.

## V. THE SINGLE-VEHICLE RECEDING HORIZON POLICY

We describe the Receding Horizon (RH) policy in Algorithm 2, where  $D$  is the set of outstanding demands waiting for

service and  $p$  is the current position of the vehicle; moreover, given a tour  $T$  of  $D$ , a fragment of  $T$  is a connected subgraph of  $T$ . If  $D \neq \emptyset$ , the vehicle selects a random  $\eta$ -fragment

---

### Algorithm 2: Receding Horizon (RH) Policy

---

**input:** Scalar  $\eta \in (0, 1)$

- 1 **if**  $D = \emptyset$  **then**
- 2     Let  $\tilde{P}_1^*$  be the point minimizing the sum of distances to demands serviced in the past (the initial condition for  $\tilde{P}_1^*$  is a random point in  $\mathcal{Q}$ ); if  $p \neq \tilde{P}_1^*$ , move to  $\tilde{P}_1^*$ , otherwise stop.
- 3 **else**
- 4     **repeat**
- 5         Compute the TSP tour through all demands in  $D$ .
- 6         Uniformly randomly choose, independently from the past, a fragment of length  $\eta$  TSP( $D$ ) of such TSP tour.
- 7         Service the selected fragment starting from the endpoint closest to the current position.
- 8     **until**  $D = \emptyset$
- 9 Repeat.

---

of the TSP tour through the demands in  $D$  (i.e., a fragment of length  $\eta$  TSP( $D$ ) of such tour); note that the horizon is not fixed, but it is adjusted according to the cost of servicing the outstanding demands. In general, the performance of the system will depend on the choice of the parameter  $\eta$ , which will be discussed after the analysis of the policy. One can easily show that the RH policy is *unbiased*.

#### A. Stability and Performance of the RH Policy

The RH policy is attractive since it can be implemented without any knowledge of the underlying demand generation process (in particular, without any knowledge of  $f$ ). In this section, we study the stability of the RH policy and we discuss its performance.

1) *Stability of the RH Policy:* In general, the stability properties of receding horizon controllers deteriorate as the horizon becomes shorter; remarkably, the RH policy is stable in every load for every  $0 < \eta < 1$ .

*Theorem 5.1 (Stability of RH policy):* As  $\varrho \rightarrow 1^-$ , the system time for the RH policy satisfies

$$\bar{T}_{\text{RH}}(\eta) \leq \frac{\lambda \beta_{\mathcal{Q},2}^2 |\mathcal{Q}|}{v^2 (1 - \varrho)^2}, \quad \text{for all } \eta \in (0, 1),$$

where  $|\mathcal{Q}|$  is the area of  $\mathcal{Q}$  and  $\beta_{\mathcal{Q},2}$  is the constant appearing in equation (2).

*Proof:* By following the same arguments as in Lemma 4.3, one can show that, in heavy load,  $D \neq \emptyset$  with high probability. Accordingly, in the following heavy-load analysis we consider the condition  $D = \emptyset$  *always* false, and the RH policy then *reduces* to lines 4-8 of Algorithm 2.

We refer to the time instant  $t_i$ ,  $i \geq 0$ , in which the vehicle computes a new TSP tour as the epoch  $i$  of the policy; we refer to the time interval between epoch  $i$  and epoch  $i + 1$  as the  $i$ th iteration and we refer to its length as the cost  $C_i$ . Let

$n_i$  be the number of outstanding demands at epoch  $i$ ; with a slight abuse of notation, we denote by  $\text{TSP}(n_i)$  the length of the TSP tour through the outstanding demands at epoch  $i$ . Since an  $\eta$ -fragment is randomly chosen, the expected number of demands left unserved during iteration  $i$  is  $(1 - \eta) \mathbb{E}[n_i]$ ; then,

$$\mathbb{E}[n_{i+1}] = \underbrace{\lambda \mathbb{E}[C_i]}_{\text{newly arrived demands}} + \underbrace{(1 - \eta) \mathbb{E}[n_i]}_{\text{demands left unserved during iteration } i}.$$

The expected number of demands that receive service during iteration  $i$  is  $\eta \mathbb{E}[n_i]$ . Moreover, the time to reach the starting endpoint of the selected fragment is bounded by  $\text{diam}(\mathcal{Q})/v$ . Then, the expected time length of iteration  $i$  can be upper bounded as

$$\begin{aligned} \mathbb{E}[C_i] &\leq \text{diam}(\mathcal{Q})/v + \frac{\eta}{v} \mathbb{E}[\text{TSP}(n_i)] + \bar{s} \eta \mathbb{E}[n_i] \\ &\leq \text{diam}(\mathcal{Q})/v + \frac{\eta}{v} \beta_{\mathcal{Q},2} \sqrt{|\mathcal{Q}|} \sqrt{\mathbb{E}[n_i]} + \bar{s} \eta \mathbb{E}[n_i], \end{aligned} \quad (16)$$

where we have applied the deterministic bound in equation (2) and Jensen's inequality for concave functions. Therefore, we obtain

$$\begin{aligned} \mathbb{E}[n_{i+1}] &\leq \lambda \text{diam}(\mathcal{Q})/v + \lambda \frac{\eta}{v} \beta_{\mathcal{Q},2} \sqrt{|\mathcal{Q}|} \sqrt{\mathbb{E}[n_i]} \\ &\quad + \varrho \eta \mathbb{E}[n_i] + (1 - \eta) \mathbb{E}[n_i]. \end{aligned}$$

By using the same techniques as in the proof of Lemma 4.4 (in particular, the eigenvalue of the "bounding" discrete-time linear system is  $\mu = 1 - \eta(1 - \varrho)$ , whose magnitude is strictly less than one for all  $\eta \in (0, 1)$ ), it is possible to show that  $\limsup_{i \rightarrow \infty} \mathbb{E}[n_i] \doteq \bar{n}$  can be upper bounded as

$$\bar{n} \leq \frac{1}{4} \left( \frac{\lambda \beta_{\mathcal{Q},2} \sqrt{|\mathcal{Q}|}}{v(1 - \varrho)} + \sqrt{\frac{\lambda^2 \beta_{\mathcal{Q},2}^2 |\mathcal{Q}|}{v^2 (1 - \varrho)^2} + \frac{4 \lambda \text{diam}(\mathcal{Q})}{v \eta (1 - \varrho)}} \right)^2,$$

for all  $\eta \in (0, 1)$ . Define  $\bar{C} \doteq \limsup_{i \rightarrow +\infty} \mathbb{E}[C_i]$ . Consider a random demand in steady state; the expected number of iterations before such demand is scheduled for service is given by  $\sum_{p=0}^{+\infty} p \eta (1 - \eta)^p$ . Hence, the expected steady-state system time of such demand,  $\bar{T}$ , will be upper bounded for all  $\eta \in (0, 1)$  and as  $\varrho \rightarrow 1^-$  according to:

$$\bar{T} \leq \frac{1}{2} \bar{C} + \frac{1}{2} \bar{s} \eta \bar{n} + \bar{C} \sum_{p=0}^{+\infty} p \eta (1 - \eta)^p \leq \frac{\lambda \beta_{\mathcal{Q},2}^2 |\mathcal{Q}|}{v^2 (1 - \varrho)^2},$$

where we used the fact that, as  $\varrho \rightarrow 1^-$ , the travel time along an  $\eta$ -fragment is negligible compared to the on-site service time requirement. This concludes the proof. ■

Since, by Theorem 5.1, the RH policy stabilizes the system in the limit  $\varrho \rightarrow 1^-$ , it also stabilizes the system for every load factor  $\varrho$ .

2) *Performance of the RH Policy:* In light load, the RH policy becomes identical to the DC policy, and therefore it is *optimal* (see Theorem 4.1). For the case  $\varrho \rightarrow 1^-$ , because of the dependencies that arise among demands' locations, we were unable to obtain rigorous upper bounds on the system time that are well matched by numerical experiments. However, we now present a heuristic analysis of the RH policy

that provides interesting insights on its behavior and suggests bounds on its heavy-load system time that are well matched by simulation results.

The service of an  $\eta$ -fragment introduces *dependencies* among the locations of the demands that are left unserved; therefore, even though the number of demands, as  $\varrho \rightarrow 1^-$ , becomes large with probability 1, the result in equation (1) (that requires a set of *independently* distributed demands) formally does *not* hold. However, due to the randomized selection of the  $\eta$ -fragments, it is quite natural to conjecture that equation (1) still provides a good estimate on the average lengths of the TSP tours that are computed. The rationale behind this conjecture is the following: When  $\eta$  is close to 1, most of the outstanding demands are serviced during one iteration, and the next TSP tour is mostly through newly arrived demands, that are independent by the assumptions on the demand generation process; when, instead,  $\eta$  is close to zero, the randomized selection of *short* fragments only introduce "negligible" correlations.

The above observations motivate us to use  $\beta_{\text{TSP},2} \int_{\mathcal{Q}} f^{1/2}(x) dx$  (as it would be dictated by equation (1)) in equation (16), instead of  $\beta_{\mathcal{Q},2} \sqrt{|\mathcal{Q}|}$  (as it is dictated by equation (2)). Then, by repeating the same steps as in the proof of Theorem 4.2, Lemma 4.4, and Theorem 5.1, we would obtain the following result for all  $\eta \in (0, 1)$

$$\bar{T}_{\text{RH}}(\eta) \leq \beta_{\text{TSP},2}^2 \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{v^2 (1 - \varrho)^2}, \quad \text{as } \varrho \rightarrow 1^-. \quad (17)$$

The upper bound in equation (17) is generally tighter than the upper bound in Theorem 5.1 (notice that by Jensen's inequality for concave functions  $\int_{\mathcal{Q}} f^{1/2}(x) dx \leq \sqrt{|\mathcal{Q}|} \left[ \int_{\mathcal{Q}} f(x) dx \right]^{1/2} = \sqrt{|\mathcal{Q}|}$ ), but unlike the upper bound in Theorem 5.1 it has not been established formally. Simulation results (see Section VIII) indeed confirm the upper bound in equation (17).

Finally, note that when  $\eta$  is close to zero, the RH policy is conceptually similar to the DC policy with  $r \rightarrow \infty$ , so we also speculate that in heavy load the performance of the RH policy improves as the horizon becomes shorter. Simulation results (see Section VIII) indeed confirm this behavior.

## B. Discussion

The RH policy is optimal in light load and stabilizes the system in *every* load condition (these two statements have been established rigorously). The implementation of the RH policy does *not* require the knowledge of  $\lambda, \bar{s}, v$  and  $f$ . Therefore, the RH policy *adapts* to arrival rate  $\lambda$ , expected on-site service  $\bar{s}$ , vehicle's velocity  $v$ , and spatial density  $f$ ; in other words, the RH policy adapts to *all* problem data.

In Section IV we saw that also the DC policy with  $r = 1$  adapts to all problem data. Simulation results (see Section VIII) show that in heavy load the RH policy with a short horizon  $\eta$  (say,  $\eta \approx 0.2$ ) performs better than the DC policy with  $r = 1$  (the light-load behavior is clearly the same). Therefore, to date, the best available policy for the 1-DTRP

that is unbiased and adapts to all problem data seems to be the RH policy with  $\eta \approx 0.2$ .

It is natural to wonder how the RH policy performs if the  $\eta$ -fragment is not selected randomly, but it is instead selected according to the rule: Find the fragment of length  $\eta$  TSP( $D$ ) that maximizes the number of reached demands. In other words, the vehicle now looks for a maximum-reward  $\eta$ -fragment. Clearly, Theorem 5.1 still applies. Simulation results (see [26]) show that this spatially-biased variant of the RH policy appears to perform better in heavy load than an optimal unbiased policy but not as well as an optimal biased policy.

As for the DC policy, the RH policy requires on-line solutions of possibly large TSPs; therefore, practical implementations of this policy should rely on heuristics or on polynomial-time approximation schemes for the solution of TSPs, such as Lin-Kernighan's or Christofides'. We will further discuss computation times in Section VIII.

## VI. ADAPTIVE AND DISTRIBUTED POLICIES FOR THE $m$ -DTRP

In this section we extend the previous single-vehicle policies to the multi-vehicle case. First, we prove that in heavy load, given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, there exists an unbiased  $\pi^*$ -partitioning policy that is optimal. This result and the optimality of the SQM policy (which is a partitioning policy that is optimal in light load) lead us to propose, for the multi-vehicle case, *partitioning policies* that use the DC policy or the RH policy as single-vehicle policies. Finally, we discuss how a partition can be computed in a distributed fashion, so that the proposed policies are amenable to distributed implementation.

### A. Optimality of Partitioning Policies in Heavy Load

The following theorem shows the optimality of a specific type of partitioning policy.

*Theorem 6.1 (Optimal partitioning in heavy load):* As  $\varrho \rightarrow 1^-$ , given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, a  $\pi^*$ -partitioning policy using an  $m$ -partition which is simultaneously equitable with respect to  $f$  and  $f^{1/2}$  is an optimal unbiased policy for the  $m$ -DTRP.

*Proof:* Let  $\pi^*$  be an optimal unbiased policy (e.g., the DC policy with  $r \rightarrow \infty$ ). Construct an  $m$ -partition  $\{\mathcal{Q}_k\}_{k=1}^m$  of  $\mathcal{Q}$  that is simultaneously equitable with respect to  $f$  and  $f^{1/2}$  (such partition is guaranteed to exist as discussed in Section II); assign one vehicle to each subregion. Each vehicle executes the single-vehicle unbiased policy  $\pi^*$  to service demands that fall within its own subregion. The probability that a demand falls in subregion  $\mathcal{Q}_k$  is equal to  $\int_{\mathcal{Q}_k} f(x) dx = 1/m$ . Notice that the arrival rate  $\lambda_k$  in subregion  $\mathcal{Q}_k$  is reduced by a factor  $\int_{\mathcal{Q}_k} f(x) dx$ , i.e.,  $\lambda_k = \lambda \int_{\mathcal{Q}_k} f(x) dx = \lambda/m$ ; therefore, the load factor in subregion  $\mathcal{Q}_k$  (where only one vehicle provides service) is  $\varrho_k = \lambda_k \bar{s} = \lambda \bar{s}/m = \varrho < 1$ , in particular the necessary condition for stability is satisfied in every subregion. Finally, given that a demand falls in subregion  $\mathcal{Q}_k$ , the conditional density for its location (whose support is  $\mathcal{Q}_k$ ) is  $f(x)/\left(\int_{\mathcal{Q}_k} f(x) dx\right) = m f(x)$ . Then, by the law of

total expectation we have

$$\begin{aligned} \bar{T} &= \sum_{k=1}^m \left( \int_{\mathcal{Q}_k} f(x) dx \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda_k \left[ \int_{\mathcal{Q}_k} \sqrt{\frac{f(x)}{\int_{\mathcal{Q}_k} f(x) dx}} dx \right]^2}{v^2 (1 - \varrho_k)^2} \right) \\ &= \sum_{k=1}^m \left( \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}_k} f^{1/2}(x) dx \right]^2}{v^2 (1 - \varrho)^2} \right) \\ &= \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m^2 v^2 (1 - \varrho)^2}, \quad \text{as } \varrho \rightarrow 1^-. \end{aligned} \tag{18}$$

Thus, the bound in equation (4) is achieved. We finally show that such multi-vehicle policy is indeed unbiased. In fact, assume steady-state conditions, and let  $X$  be the location of a randomly chosen demand,  $T$  be its system time and  $\mathcal{S}$  be an arbitrary subset of  $\mathcal{Q}$ ; then, by the law of total expectation we have:  $\mathbb{E}[T|X \in \mathcal{S}] = \mathbb{E}\left[\mathbb{E}[T|X \in \mathcal{S} \cap \mathcal{Q}_k] | X \in \mathcal{S}\right]$ . Since  $\pi^*$  is unbiased, it must hold  $\mathbb{E}[T|X \in \mathcal{S} \cap \mathcal{Q}_k] = \mathbb{E}[T|X \in \mathcal{Q}_k]$ ; moreover, from the previous analysis, we have  $\mathbb{E}[T|X \in \mathcal{Q}_k] = \bar{T}_{\pi^*}$ , which is a constant independent of  $\mathcal{Q}_k$ . Hence, we finally obtain  $\mathbb{E}[T|X \in \mathcal{S}] = \bar{T}_{\pi^*}$ . The claim then follows from the fact that  $\mathcal{S}$  was chosen arbitrarily. ■

*Remark 6.2:* Theorem 6.1 proves Conjecture 1 in [12] (made only for uniform spatial densities).

*Remark 6.3:* It is immediate to generalize Theorem 6.1 as follows. Let  $\pi$  be a single-vehicle unbiased policy such that in heavy load  $\bar{T}_{\pi}/\bar{T}_{\pi^*} \leq \gamma$ . As  $\varrho \rightarrow 1^-$ , given such policy  $\pi$  and  $m$  vehicles, a  $\pi$ -partitioning policy using an  $m$ -partition which is simultaneously equitable with respect to  $f$  and  $f^{1/2}$  is always within a factor  $\gamma$  of the optimal unbiased performance.

The proof of Theorem 6.1 relies on  $m$ -partitions that are simultaneously equitable with respect to  $f$  and  $f^{1/2}$ . We next show in Theorem 6.4 that a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f^{1/2}$  may be unstable; moreover, we show in Theorem 6.5 that, in general, a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f$  does not achieve the optimal unbiased performance, however it is always within a factor  $m$  of it.

*Theorem 6.4 (Unstable partitioning policies):* Given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f^{1/2}$  may be unstable for values of  $\varrho$  strictly less than one.

*Proof:* The arrival rate in subregion  $\mathcal{Q}_k$ ,  $k \in \{1, \dots, m\}$ , is  $\lambda_k = \lambda \int_{\mathcal{Q}_k} f(x) dx$ , and therefore the load factor in subregion  $\mathcal{Q}_k$  (where only one vehicle provides service) is  $\varrho_k = \lambda \int_{\mathcal{Q}_k} f(x) dx \bar{s}$ . In general, an  $m$ -partition that is equitable with respect to  $f^{1/2}$  is *not* equitable with respect to  $f$ . If this is the case, there exists  $\bar{k} \in \{1, \dots, m\}$  and  $\varepsilon > 0$  such that  $\int_{\mathcal{Q}_{\bar{k}}} f(x) dx = 1/m + \varepsilon$ ; then,  $\varrho_{\bar{k}} = \varrho + \varepsilon \lambda \bar{s} = \varrho(1 + \varepsilon m)$ . Hence, for every value of  $\varrho$  such that  $1/(1 + \varepsilon m) \leq \varrho < 1$ , such multi-vehicle policy is unstable (since in subregion  $\mathcal{Q}_{\bar{k}}$  one has  $\varrho_{\bar{k}} \geq 1$ ), even though the load factor  $\varrho$  is strictly less than one. ■

*Theorem 6.5 (Suboptimal partitioning in heavy load):* As  $\varrho \rightarrow 1^-$ , given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f$  does not achieve, in general, the optimal unbiased performance, however it is always within a factor  $m$  of it.

*Proof:* Let  $\pi^*$  be an optimal unbiased policy. The environment  $\mathcal{Q}$  is partitioned into  $m$  subregions  $\mathcal{Q}_k$ ,  $k \in \{1, \dots, m\}$ , such that  $\int_{\mathcal{Q}_k} f(x) dx = 1/m$  for all  $k$ , and one vehicle is assigned to each subregion. Each vehicle executes the single-vehicle policy  $\pi^*$  to service demands that fall within its own subregion. By arguments similar to those in the proof of Theorem 6.1, one can show that

$$\bar{T} = \sum_{k=1}^m \left( \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}_k} f^{1/2}(x) dx \right]^2}{v^2 (1-\varrho)^2} \right), \text{ as } \varrho \rightarrow 1^-. \quad (19)$$

Define the vector

$$z = \left( \int_{\mathcal{Q}_1} f^{1/2}(x) dx, \dots, \int_{\mathcal{Q}_m} f^{1/2}(x) dx \right) \in \mathbb{R}^m.$$

We will denote by  $\|z\|_2$  and  $\|z\|_1$  the 2-norm and the 1-norm, respectively, of vector  $z$ . Notice that  $\|z\|_1 = \int_{\mathcal{Q}} f^{1/2}(x) dx$ . Since  $\|z\|_1 \leq \sqrt{m} \|z\|_2$  for any vector  $z \in \mathbb{R}^m$ , equation (19) becomes

$$\begin{aligned} \bar{T} &= \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \|z\|_2^2}{v^2 (1-\varrho)^2} \geq \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \|z\|_1^2}{m v^2 (1-\varrho)^2} \\ &= \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m^2 v^2 (1-\varrho)^2} = \bar{T}_U^*, \text{ as } \varrho \rightarrow 1^-, \end{aligned}$$

where  $\bar{T}_U^*$  is the optimal unbiased performance. In general, an  $m$ -partition that is equitable with respect to  $f$  is *not* equitable with respect to  $f^{1/2}$ ; therefore, in general, the inequality  $\|z\|_1 \leq \sqrt{m} \|z\|_2$  is strict and  $\bar{T}$  is larger than the optimal unbiased performance.

On the other hand,  $\|z\|_2 \leq \|z\|_1$  for any vector  $z \in \mathbb{R}^m$ , therefore by using the same reasoning as in equation (19) we obtain

$$\begin{aligned} \bar{T} &= \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \|z\|_2^2}{v^2 (1-\varrho)^2} \leq \frac{1}{m} \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \|z\|_1^2}{v^2 (1-\varrho)^2} \\ &= \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m v^2 (1-\varrho)^2} = m \bar{T}_U^*, \text{ as } \varrho \rightarrow 1^-, \end{aligned}$$

where, again,  $\bar{T}_U^*$  is the optimal unbiased performance. Note that in this case the resulting multi-vehicle policy might be spatially-biased. ■

*Remark 6.6:* It is immediate to generalize Theorem 6.5 as follows. Let  $\pi$  be a single-vehicle unbiased policy such that in heavy load  $\bar{T}_\pi / \bar{T}_U^* \leq \gamma$ . As  $\varrho \rightarrow 1^-$ , given such policy  $\pi$  and  $m$  vehicles, a  $\pi$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f$  is always within a factor  $\gamma m$  of the optimal unbiased performance.

## B. Distributed Policies for the $m$ -DTRP and Discussion

The optimality of the SQM policy and Theorem 6.1 suggest the following distributed multi-vehicle version of the DC policy: 1) the vehicles compute in a distributed way an  $m$ -median of  $\mathcal{Q}$  that induces a Voronoi partition that is equitable with respect to  $f$  and  $f^{1/2}$ ; 2) the vehicles assign themselves to the subregions (thus, there is a one-to-one correspondence between vehicles and subregions); and 3) each vehicle executes the single-vehicle DC policy to service demands that fall within its own subregion, by using the median of the subregion instead of  $\bar{P}_1^*$ . (In an identical way, one can obtain a multi-vehicle version of the RH policy.)

If there exists, given  $\mathcal{Q}$  and  $f$ , an  $m$ -median of  $\mathcal{Q}$  that induces a Voronoi partition that is equitable with respect to  $f$  and  $f^{1/2}$ , then the above policy is optimal both in light load and arbitrarily close to optimality in heavy load, and stabilizes the system in every load condition. There are two main issues with the above policy, namely (i) existence of an  $m$ -median of  $\mathcal{Q}$  that induces a Voronoi partition that is equitable with respect to  $f$  and  $f^{1/2}$ , and (ii) how to compute it in a distributed way. In [27], we showed that for some choices of  $\mathcal{Q}$  and  $f$  a median Voronoi diagram that is equitable with respect to  $f$  and  $f^{1/2}$  *fails* to exist. On the other hand, in [28], we presented a synchronous, distributed partitioning algorithm that, for any possible choice of  $\mathcal{Q}$  and  $f$ , provides a partition of  $\mathcal{Q}$  that is equitable with respect to  $f$  and represents a “good” approximation of a median Voronoi diagram (see [28] for details on the metrics that we use to judge “closeness” to median Voronoi diagrams). If an  $m$ -median of  $\mathcal{Q}$  that induces a Voronoi partition that is equitable with respect to  $f$  exists, the algorithm will locally converge to it. Accordingly, we define the multi-vehicle Divide & Conquer policy as follows:

---

### Algorithm 3: Multi-Vehicle DC ( $m$ -DC) Policy

---

- 1 The vehicles run the distributed partitioning algorithm defined as a set of differential equations in [28], by using as input measure  $f$ .
  - 2 Simultaneously, each vehicle executes the single-vehicle DC policy inside its own subregion (the assignment of vehicles to subregions is a by-product of the algorithm in [28]).
- 

According to Remark 6.6 the  $m$ -DC policy is within a factor  $(1+1/r)m$  of the optimal unbiased performance in heavy load (since the algorithm in [28] *always* provides a partition of  $\mathcal{Q}$  that is equitable with respect to  $f$ ), and stabilizes the system in every load condition. In general, the  $m$ -DC policy is only *suboptimal* in light load; note, however, that the computation of the global minimum of the Weber function  $H_m$  (which is non-convex for  $m > 1$ ) is difficult for  $m > 1$  (it is NP-hard for the discrete version of the problem, see [18]); therefore, for  $m > 1$ , suboptimality has also to be expected from *any* practical implementation of the SQM policy. If an  $m$ -median of  $\mathcal{Q}$  that induces a Voronoi partition that is equitable with respect to  $f$  exists, the partitioning algorithm used by the  $m$ -DC will locally converge to it, thus we say that the  $m$ -DC policy is “locally” optimal in light load.

Note that, when the density is uniform, a partition that is equitable with respect to  $f$  is also equitable with respect to  $f^{1/2}$ ; therefore, when the density is uniform the  $m$ -DC policy (with  $r \rightarrow +\infty$ ) is an optimal unbiased policy in heavy load (see Theorem 6.1). The  $m$ -DC policy *adapts* to arrival rate  $\lambda$ , expected on-site service  $\bar{s}$ , and vehicle's velocity  $v$ ; however, it requires the knowledge of  $f$ . If this is not available, the vehicles might use an empirical distribution function for the demand locations (which might be computed either through a centralized algorithm or through a distributed algorithm). Since the demand locations are i.i.d., the empirical distribution will converge to the true distribution, and therefore after a transient (whose length depends on the arrival rate  $\lambda$ ) the vehicles will provide a system time that agrees with the bounds derived under the assumption that  $f$  is known by the vehicles. The key feature of the partitioning algorithm in [28] is that it is *distributed*; therefore, the  $m$ -DC policy is indeed a distributed control policy. The multi-vehicle version of the RH policy (which we call  $m$ -RH policy) can be defined in an identical way. Its properties are summarized in Table II in Section I.

## VII. ON THE CASE WITH ZERO ON-SITE SERVICE TIME

In the formulation of the  $m$ -DTRP we assumed that  $\bar{s} > 0$  (see Section III). However, it is of interest to also study the case  $\bar{s} = 0$ , since in some applications the on-site service time might actually be zero (or negligible). We show that, perhaps surprisingly, when  $\bar{s} = 0$  a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f^{1/2}$  is an optimal unbiased policy in heavy load. Hence, when  $\bar{s} = 0$ , optimality does *not* require simultaneous equitability, a considerable simplification.

Specifically, when  $\bar{s} = 0$ , the light-load regime is defined as  $\lambda \rightarrow 0^+$ , while the heavy-load regime is defined as  $\lambda \rightarrow +\infty$ . Then, it is possible to verify that all the performance bounds in Sections III, IV, and V hold by simply substituting  $\varrho = 0$ , and by formally changing the limits  $\varrho \rightarrow 0^+$  and  $\varrho \rightarrow 1^-$  with  $\lambda \rightarrow 0^+$  and  $\lambda \rightarrow +\infty$ . For example, equation (4) reads  $\bar{T}_U \geq \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m^2 v^2}$ , as  $\lambda \rightarrow +\infty$ . The major difference is about which type of partition one should use in heavy load (i.e., when  $\lambda$  is large). We have the following result.

*Theorem 7.1 (Optimal partitioning when  $\bar{s} = 0$ ):* Assume  $\bar{s} = 0$ . As  $\lambda \rightarrow +\infty$ , given a single-vehicle, optimal unbiased policy  $\pi^*$  and  $m$  vehicles, a  $\pi^*$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f^{1/2}$  is an optimal unbiased policy for the  $m$ -DTRP.

*Proof:* We only sketch the proof, since it is very similar to the proof of Theorem 6.1. Let  $\pi^*$  be an optimal unbiased policy; by considering an  $m$ -partition  $\{\mathcal{Q}_k\}_{k=1}^m$  of  $\mathcal{Q}$  that is equitable with respect to  $f^{1/2}$ , and by using arguments similar to the ones in the proof of Theorem 6.1, one can show that, as  $\lambda \rightarrow +\infty$ ,

$$\begin{aligned} \bar{T} &= \sum_{k=1}^m \left( \int_{\mathcal{Q}_k} f(x) dx \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \frac{\int_{\mathcal{Q}} f^{1/2}(x) dx}{m} \right]^2}{v^2} \right) \\ &= \frac{\beta_{\text{TSP},2}^2}{2} \frac{\lambda \left[ \int_{\mathcal{Q}} f^{1/2}(x) dx \right]^2}{m^2 v^2}, \end{aligned} \quad (20)$$

where in the last equality we have used the fact that  $\sum_{k=1}^m \int_{\mathcal{Q}_k} f(x) dx = 1$ . Thus, the bound in equation (4) is achieved. The proof that this multi-vehicle policy is unbiased requires arguments identical to those presented in the proof of Theorem 6.1, and therefore it is omitted.

*Remark 7.2:* It is immediate to generalize Theorem 7.1 as follows. Let  $\pi$  be a single-vehicle unbiased policy such that in heavy load  $\bar{T}_\pi / \bar{T}_U^* \leq \gamma$ . Assume  $\bar{s} = 0$ . As  $\lambda \rightarrow +\infty$ , given such policy  $\pi$  and  $m$  vehicles, a  $\pi$ -partitioning policy using an  $m$ -partition which is equitable with respect to  $f^{1/2}$  is always within a factor  $\gamma$  of the optimal unbiased performance. ■

Theorem 7.1 is very important from a practical viewpoint, since it means that when  $\bar{s} = 0$  optimality does not require simultaneous equitability. In light of Theorem 7.1, when  $\bar{s} = 0$  the  $m$ -DC policy should be defined in the same way as Algorithm 3, with the exception that the partitioning algorithm in [28] should partition  $\mathcal{Q}$  according to  $f^{1/2}$  (instead of  $f$ ). Since the partitioning algorithm in [28] is then guaranteed to converge to a partition that is equitable with respect to  $f^{1/2}$ , we conclude that when  $\bar{s} = 0$  the distributed  $m$ -DC policy (with the modification that equitability should be with respect to  $f^{1/2}$ , and with  $r \rightarrow +\infty$ ) is an optimal unbiased policy in heavy load for *any* density  $f$ .

## VIII. SIMULATION EXPERIMENTS

In this section we discuss, through the use of simulations, the heavy-load behavior of both the DC and the RH policy, and we comment on their relative performance. All simulations are performed on a machine with a 2.4GHz Intel Core Duo processor and 4GB of RAM. The code is written in Matlab<sup>®</sup>7.4 with external calls to the program `linkern`, for which we set a running time bound of one second. In all simulations we consider a circular environment  $\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1/\pi\}$  (hence, the area of  $\mathcal{Q}$  is 1), a vehicle's velocity  $v = 1$ , and an on-site service time uniformly distributed in the interval  $[0, 1]$  (thus  $\bar{s} = 0.5$ ). The spatial density  $f : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$  used in the simulation experiments is the piecewise constant function  $f(x) = \frac{1-\delta}{\varepsilon} \chi_{\mathcal{Q}_1}(x) + \frac{\delta}{1-\varepsilon} \chi_{\mathcal{Q}_2}(x)$ , where  $\chi_{\mathcal{S}}(x)$  is the indicator function for the set  $\mathcal{S} \subseteq \mathcal{Q}$  (having the value 1 when  $x \in \mathcal{S}$  and 0 otherwise),  $\mathcal{Q}_1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq \varepsilon/\pi\}$ ,  $\mathcal{Q}_2 = \mathcal{Q} \setminus \mathcal{Q}_1$ , and  $\varepsilon, \delta \in (0, 1)$ . If  $1 - \delta = \varepsilon$ , then the density  $f$  is uniform; if, instead,  $1 - \delta > \varepsilon$ , then the density  $f$  has a peak in subregion  $\mathcal{Q}_1$ . In all simulation experiments we set  $\varepsilon = 0.1$ .

### A. Heavy-Load Performance of the DC Policy

We first consider a uniform spatial density, i.e.,  $\delta = 1 - \varepsilon = 0.9$ . We consider 5 values of the load factor, namely  $\varrho = 0.9, 0.93, 0.95, 0.97, 0.99$ . Circles in the top figure of Figure 1 represent the ratios between the experimental system time  $\bar{T}_{\text{DC}}(1)$  (i.e.,  $r = 1$  in the DC policy) and  $\bar{T}_U^*$  (whose expression is given in equation (4)), for the different values of  $\varrho$ . Squares in the top figure of Figure 1 represent the ratios between the experimental system time  $\bar{T}_{\text{DC}}(16)$  (i.e.,  $r = 16$  in the DC policy) and  $\bar{T}_U^*$ , for the different values of  $\varrho$ . It can be observed that for  $r = 1$  the ratios  $\bar{T}_{\text{DC}}(1)/\bar{T}_U^*$  decrease as

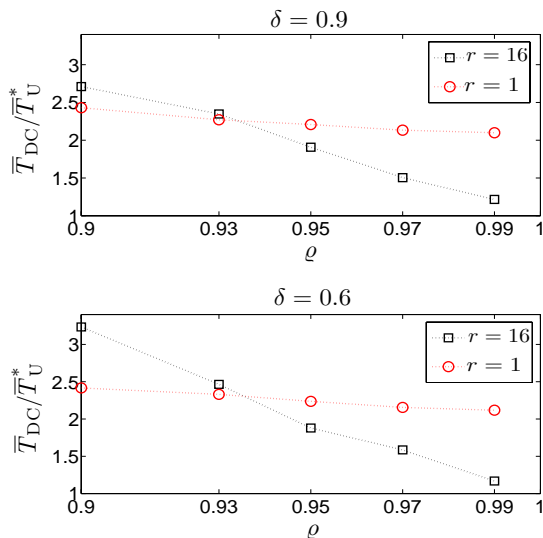


Fig. 1. Top Figure: Ratio between experimental system time under the DC policy and  $\bar{T}_U^*$  (whose expression is given in equation (4)) in the case of uniform density (i.e.,  $\delta = 0.9$ ). Bottom Figure: Ratio between experimental system time under the DC policy and  $\bar{T}_U^*$  in the case of non-uniform density ( $\delta = 0.6$ ). Circles correspond to the DC policy with  $r = 1$ , while squares correspond to the DC policy with  $r = 16$ .

TABLE III  
COMPUTATION TIMES FOR DC POLICY

	$\varrho=0.9$	$\varrho=0.93$	$\varrho=0.95$	$\varrho=0.97$	$\varrho=0.99$
$r = 1$	$\simeq 0.3s$	$\simeq 0.8s$	$\simeq 1.1s$	$\simeq 1.7s$	$\simeq 5s$
$r = 16$	$\simeq 0.03s$	$\simeq 0.1s$	$\simeq 0.15s$	$\simeq 0.3s$	$\simeq 1.3s$

$\varrho$  approaches one and tend to 2, in accordance with the upper bound in Theorem 4.2 (in fact,  $1 + 1/r = 2$  in this case). Similarly, for  $r = 16$  the ratios  $\bar{T}_{DC}(16)/\bar{T}_U^*$  decrease as  $\varrho$  approaches one and tend to  $(1 + 1/16) \approx 1.06$ , in accordance again with the upper bound in Theorem 4.2. These results confirm the theoretical analysis in Section IV and suggest that the upper bound (6) is indeed tight for the various values of  $r$ . Note that for all values of  $\varrho$  the experimental system time is *larger* than the upper bound; this is due to one or a combination of the following reasons: First, the upper bound formally holds only in the limit  $\varrho \rightarrow 1^-$ . Second, we are using an approximate solution for the optimal TSP.

In Table III we show for both  $r = 1$  and  $r = 16$  the average time required for the computation of a TSP tour. Specifically, the computation time includes (i) the time to read the points from an input file, (ii) the TSP computation time, and (iii) the time to save the tour on a file. Recall that we implemented the DC policy by using `linkern` as a solver to generate approximations to the optimal TSP tour, and we set a running time bound of one second for this solver. One can observe that the computation times are rather small even for values of  $\varrho$  very close to one; moreover, since the experimental results are very close to the theoretical bounds, we argue that `linkern` (with a running time bound of one second) computes TSP tours very close to the optimal ones. Hence, we conclude that the DC policy can be effectively implemented in real-world applications by using `linkern` as a subroutine.

The same set of simulations is performed for a non-uniform

spatial density; in particular we consider  $\delta = 0.6$ , i.e., a peak in the small subregion  $\mathcal{Q}_1$ . Note that in this case a simultaneously equitable partition can be trivially obtained by using radial cuts. Results are shown in the bottom figure of Figure 1. Circles represent the ratios between the experimental system time and  $\bar{T}_U^*$  for  $r = 1$ , while squares represent the ratios between the experimental system time and  $\bar{T}_U^*$  for  $r = 16$ . As before, the results are in accordance with the upper bound in Theorem 4.2. The computation times are very similar to those shown in Table III and therefore they are omitted.

Finally, from Figure 1 it can be observed that as  $\varrho \rightarrow 1^-$  the DC policy with  $r = 16$  performs better than the DC policy with  $r = 1$ , precisely by a factor 2 in accordance with the upper bound in Theorem 4.2. On the other hand, for moderate values of  $\varrho$ , say  $\varrho \approx 0.9$ , the DC policy with  $r = 1$  appears to perform slightly better than the DC policy with  $r = 16$  (recall that the performance bounds hold in the limit  $\varrho \rightarrow 1^-$ ).

### B. Heavy-Load Performance of the RH Policy

We first consider a uniform spatial density. We consider 3 values of the load factor, namely  $\varrho = 0.95, 0.97, 0.99$ , and 7 values of  $\eta$ , namely  $\eta = 0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 1$ . Formally, the RH policy is not defined for  $\eta = 1$ , however, by setting  $\eta = 1$  in the definition of the RH policy, we simply obtain the DC policy with  $r = 1$ . The top figure of Figure 2 shows the ratios between the experimental system time  $\bar{T}_{RH}(\eta)$  and  $\bar{T}_U^*$  (whose expression is given in equation (4)) for each pair  $(\varrho, \eta)$ . The same set of simulations is performed for a non-uniform spatial density; in particular we consider  $\delta = 0.6$ . Results are shown in the bottom figure of Figure 2. The results confirm that the system time  $\bar{T}_{RH}$  follows the  $1/(1 - \varrho)^2$  growth predicted by Theorem 5.1. In Section V we argued that in heavy load  $T_{RH}(\eta) \leq 2\bar{T}_U^*$  (see equation (17)) for all  $\eta \in (0, 1)$ ; from Figure 2, it can be observed that as  $\varrho$  approaches one this asymptotic bound seems to be confirmed, both for a uniform  $f$  and for a non-uniform  $f$ . We also argued in Section V that the performance of the RH policy should improve as the horizon becomes shorter: this is indeed the case, in particular simulation results show that optimal performance is achieved for  $\eta \approx 0.2$ . Finally, the computation times for the RH policy with any  $\eta$  are very similar to those of the DC policy with  $r = 1$  and therefore they are omitted.

The RH policy should be compared with the version of the DC policy that is also adaptive with respect to  $f$ , i.e., with the DC policy with  $r = 1$ . Indeed, as observed before, the DC policy with  $r = 1$  is the same as the RH policy when we set  $\eta = 1$ . Therefore, we can compare the two policies by looking at Figure 2. One should note that the RH policy with  $\eta \approx 0.2$  performs consistently better than the DC policy with  $r = 1$  (which, again, is the same as the RH policy with  $\eta = 1$ ), in particular it decreases the system time by  $\approx 20\%$ .

### C. Performance of the Multi-Vehicle DC Policy

In this section we study, through the use of simulations, the behavior of the  $m$ -DC policy; we focus, in particular, on the heavy-load scenario (in light load, the system time is directly related to the convergence of the partitioning algorithm to



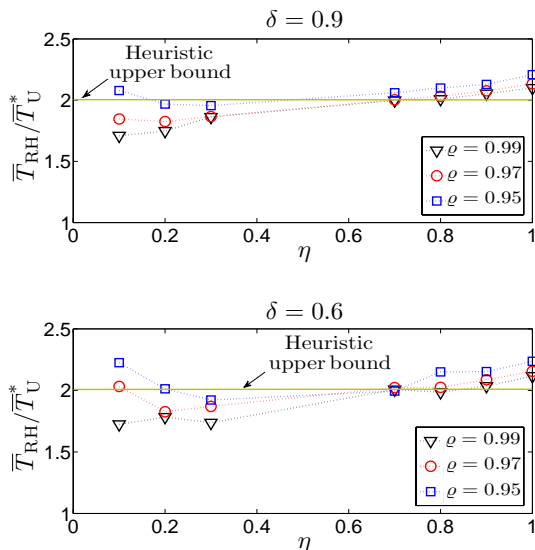


Fig. 2. Top Figure: Ratio between experimental system time under the RH policy and  $\bar{T}_U^*$  (whose expression is given in equation (4)) in the case of uniform density (i.e.,  $\delta = 0.9$ ). Bottom Figure: Ratio between experimental system time under the RH policy and  $\bar{T}_U^*$  in the case of non-uniform density ( $\delta = 0.6$ ).

a median Voronoi diagram, for which we refer to [28]). When the density  $f$  is uniform, a partition that is equitable with respect to  $f$  is also equitable with respect to  $f^{1/2}$ . By following the same arguments as in equation (18) one can show that in heavy load the steady-state system time within the entire environment follows the same behavior of the steady-state system time within each subregion. Hence, when  $f$  is uniform and the system is in heavy load the simulation results for the  $m$ -DC policy are virtually identical to the simulation results for the 1-DC policy (presented above), and therefore they are omitted.

When the density  $f$  is *not* uniform, a partition that is equitable with respect to  $f$  is not necessarily equitable with respect to  $f^{1/2}$ . Remark 6.6, however, guarantees that the  $m$ -DC policy in heavy load is within a factor  $(1 + 1/r)m$  of the optimal unbiased performance. We verify this result through simulation experiments. We consider a square environment  $\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$  (hence, the area of  $\mathcal{Q}$  is 1), vehicles' velocity  $v = 1$ , and an on-site service time uniformly distributed in the interval  $[0, 1]$  (thus  $\bar{s} = 0.5$ ). The spatial density  $f : \mathcal{Q} \rightarrow \mathbb{R}_{>0}$  used in these simulations is the piecewise constant function  $f(x) = 3\chi_{\mathcal{Q}_1}(x) + \frac{1}{3}\chi_{\mathcal{Q}_2}(x)$ , where  $\chi_{\mathcal{S}}(x)$  is the indicator function for the set  $\mathcal{S} \subseteq \mathcal{Q}$ ,  $\mathcal{Q}_1 = \{(x, y) \in \mathbb{R}^2 \mid 0.5 \leq x \leq 1, 0.5 \leq y \leq 1\}$ , and  $\mathcal{Q}_2 = \mathcal{Q} \setminus \mathcal{Q}_1$ . We consider a load factor  $\varrho = 0.9$  and 4 values for the number  $m$  of vehicles, namely  $m = 2, 4, 6, 8$ . Within each subregion, we simulate the DC policy with  $r = 1$ , which has a constant factor guarantee equal to 2 (see Theorem 4.2). The circles in Figure 3 represent the ratios between the experimental system time  $\bar{T}_{m-DC}$  and  $\bar{T}_U^*$ , for the different values of  $m$ . The squares in Figure 3 represent the upper bounds for such ratios, which, according to Remark 6.6, are equal to  $2m$ . The simulation results show that the ratios  $\bar{T}_{m-DC}/\bar{T}_U^*$  seem not to increase linearly with  $m$ , or, in other

words, that the upper bound in Remark 6.6 is conservative.

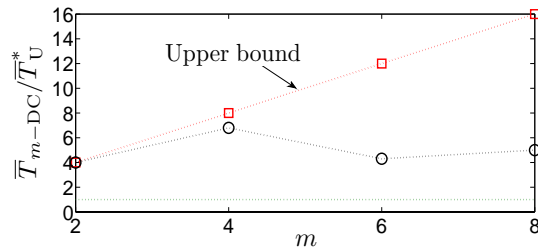


Fig. 3. The circles represent the ratios between the experimental system times under the  $m$ -DC policy (with  $r = 1$ ) and  $\bar{T}_U^*$ ; the squares represent the theoretical upper bounds on these ratios. The load factor is  $\varrho = 0.9$ .

## IX. CONCLUSION

The  $m$ -DTRP is a widely applicable model for dynamic task-assignment and scheduling problems in robotic networks. In this paper, our objective was to design unbiased policies for the  $m$ -DTRP that (i) are adaptive (in particular do *not* require the knowledge of the arrival rate  $\lambda$  and the statistics of the on-site service time), (ii) enjoy provable performance guarantees (in particular, are provably stable under any load condition), and (iii) are distributed. To this end, we introduced two new policies for the 1-DTRP, namely the DC policy and the RH policy, that were subsequently extended to the multi-vehicle case through the idea of partitioning policies. Our policies require on-line solutions of possibly large TSPs; however, we have argued in Section VIII that our policies can be effectively implemented in real-world applications by using `linkern` as a solver to generate approximations to the optimal TSP tour.

The focus of this paper was on the  $m$ -DTRP. Note, however, that in many dynamic vehicle routing problems (e.g., problems with stochastic time constraints [23] and with priorities [24]) partitioning policies provide “good” performance. Hence, following the discussion in Section VI, we argue that combining the distributed partitioning algorithms presented in [28] with appropriate partitioning and single-vehicle routing policies yields adaptive and distributed algorithms for a wide variety of dynamic vehicle routing problems. The approach presented in this paper is therefore rather general and not specifically tailored to the  $m$ -DTRP.

This paper leaves numerous important extensions open for further research. First, we plan to devise distributed algorithms that provide simultaneously equitable partitions; such algorithms would make the  $m$ -DC policy (with  $r \rightarrow +\infty$ ) optimal for any density  $f$ . Second, it is of interest to design optimal or near-optimal policies that are *biased* in heavy load and that are adaptive and distributed (recall that allowing biased service results in a strict reduction of the optimal system time for any non-uniform density  $f$ ). Third, most of the results presented in this paper hold in asymptotic regimes (either  $\varrho \rightarrow 0^+$  or  $\varrho \rightarrow 1^-$ ); addressing optimality of performance in the intermediate regime would be very important both on a theoretical and on a practical level. Finally, in our analysis, we considered omni-directional vehicles with first order dynamics; more realistic dynamical models will have to be taken into account for practical application to UAVs or other systems.



## REFERENCES

- [1] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: A game-theoretical formulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 584–596, 2007.
- [2] B. J. Moore and K. M. Passino, "Distributed task assignment for mobile agents," *IEEE Trans. on Automatic Control*, vol. 52, no. 4, pp. 749–753, 2007.
- [3] L. F. Bertuccelli, M. Alighanbari, and J. P. How, "Robust planning for coupled cooperative UAV missions," in *Proc. IEEE Conf. on Decision and Control*, vol. 3, Bahamas, Dec. 2004, pp. 2917–2922.
- [4] S. L. Smith and F. Bullo, "Monotonic target assignment for robotic networks," *IEEE Trans. on Automatic Control*, vol. 54, no. 9, pp. 2042–2057, 2009.
- [5] M. Alighanbari and J. P. How, "Cooperative task assignment of unmanned aerial vehicles in adversarial environments," in *Proc. American Control Conf.*, vol. 7, Portland, Oregon, Jun. 2005, pp. 4661–4666.
- [6] R. W. Beard, T. W. McLain, M. A. Goodrich, and E. P. Anderson, "Coordinated target assignment and intercept for unmanned air vehicles," *IEEE Trans. Robotics and Automation*, vol. 18, no. 6, pp. 911–922, 2002.
- [7] P. Toth and D. Vigo, *The vehicle routing problem*. Philadelphia, PA: SIAM Monographs on Discrete Mathematics and Applications, 2002.
- [8] D. J. Bertsimas and G. J. van Ryzin, "Stochastic and dynamic vehicle routing with general interarrival and service time distributions," *Advances in Applied Probability*, vol. 25, pp. 947–978, 1993.
- [9] —, "A stochastic and dynamic vehicle routing problem in the Euclidean plane," *Operations Research*, vol. 39, pp. 601–615, 1991.
- [10] L. Wei and C. G. Cassandras, "A cooperative receding horizon controller for multivehicle uncertain environments," *IEEE Trans. on Automatic Control*, vol. 51, no. 2, pp. 242–257, 2006.
- [11] H. N. Psaraftis, "Dynamic vehicle routing problems," in *Vehicle Routing: Methods and Studies*, B. Golden and A. Assad, Eds. Amsterdam, The Netherlands: Elsevier Science Publishers B.V., 1988, pp. 223–248.
- [12] D. J. Bertsimas and G. J. van Ryzin, "Stochastic and dynamic vehicle routing in the Euclidean plane with multiple capacitated vehicles," *Operations Research*, vol. 41, no. 1, pp. 60–76, 1993.
- [13] S. Bespamyatnikh, D. Kirkpatrick, and J. Snoeyink, "Generalizing ham sandwich cuts to equitable subdivisions," *Discrete and Computational Geometry*, vol. 24, pp. 605–622, 2000.
- [14] J. M. Steele, "Probabilistic and worst case analyses of classical problems of combinatorial optimization in Euclidean space," *Mathematics of Operations Research*, vol. 15, no. 4, p. 749, 1990.
- [15] G. Percus and O. C. Martin, "Finite size and dimensional dependence of the Euclidean traveling salesman problem," *Physical Review Letters*, vol. 76, no. 8, pp. 1188–1191, 1996.
- [16] S. Lin and B. W. Kernighan, "An effective heuristic algorithm for the traveling-salesman problem," *Operations Research*, vol. 21, pp. 498–516, 1973.
- [17] N. Christofides, "Worst-case analysis of a new heuristic for the traveling salesman problem," Graduate School of Industrial Administration, Carnegie Mellon University, Tech. Rep. 388, 1976.
- [18] P. K. Agarwal and M. Sharir, "Efficient algorithms for geometric optimization," *ACM Computing Surveys*, vol. 30, no. 4, pp. 412–458, 1998.
- [19] R. C. Larson and A. R. Odoni, *Urban Operations Research*. Prentice Hall, 1981.
- [20] H. Xu, "Optimal policies for stochastic and dynamic vehicle routing problems," PhD Thesis, Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA., 1995.
- [21] J. D. Papastavrou, "A stochastic and dynamic routing policy using branching processes with state depended immigration," *European Journal of Operational Research*, vol. 95, pp. 167–177, 1996.
- [22] A. Arsie, K. Savla, and E. Frazzoli, "Efficient routing algorithms for multiple vehicles with no explicit communications," *IEEE Trans. on Automatic Control*, vol. 54, no. 10, pp. 2302–2317, 2009.
- [23] M. Pavone, N. Bisnik, E. Frazzoli, and V. Isler, "A stochastic and dynamic vehicle routing problem with time windows and customer impatience," *ACM/Springer Journal of Mobile Networks and Applications*, vol. 14, no. 3, pp. 350–364, 2009.
- [24] S. L. Smith, M. Pavone, F. Bullo, and E. Frazzoli, "Dynamic vehicle routing with priority classes of stochastic demands," *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3224–3245, 2010.
- [25] R. Wolff, *Stochastic modeling and the theory of queues*. Prentice Hall, 1989.
- [26] M. Pavone, E. Frazzoli, and F. Bullo, "Decentralized algorithms for stochastic and dynamic vehicle routing with general demand distribution," in *Proc. IEEE Conf. on Decision and Control*, New Orleans, LA, Dec. 2007, pp. 4869–4874.
- [27] —, "Distributed policies for equitable partitioning: Theory and applications," in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 4191–4197.
- [28] M. Pavone, A. Arsie, E. Frazzoli, and F. Bullo, "Equitable partitioning policies for mobile robotic networks," *IEEE Trans. on Automatic Control*, Dec. 2008, provisionally accepted. Available at <http://arxiv.org/abs/0903.5267>.



**Marco Pavone** is a Staff Engineer/Technologist within the Advanced Robotic Controls Group at the Jet Propulsion Laboratory, California Institute of Technology. He received the Laurea degree in Systems and Control Engineering from the University of Catania, Italy, in 2004, the Diploma Engineer degree from Scuola Superiore di Catania, Italy, in 2005, and the Ph.D. degree in Aeronautics and Astronautics from the Massachusetts Institute of Technology, in 2010. His main research interests lie in the area of design and control of autonomous systems, with a focus on distributed coordination of multi-agent networks, robotic motion planning, transportation networks, and formation flying.



**Emilio Frazzoli** received the Laurea degree in aerospace engineering from the University of Rome La Sapienza, Rome, Italy, in 1994 and the Ph.D. degree in navigation and control systems from the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology (MIT), Cambridge, in 2001. From 2001 to 2004, he was an Assistant Professor of aerospace engineering at the University of Illinois at Urbana-Champaign. From 2004 to 2006, he was an Assistant Professor of mechanical and aerospace engineering at the University of California, Los Angeles. He is currently an Associate Professor of aeronautics and astronautics with the Laboratory for Information and Decision Systems, Department of Aeronautics and Astronautics, MIT. His main research interests lie in the general area of planning and control for mobile cyber-physical systems, with a particular emphasis on autonomous vehicles, mobile robotics, and transportation networks. Dr. Frazzoli received the National Science Foundation (NSF) CAREER Award in 2002.



**Francesco Bullo** received the Laurea degree "summa cum laude" in Electrical Engineering from the University of Padova, Italy, in 1994, and the Ph.D. degree in Control and Dynamical Systems from the California Institute of Technology in 1999. From 1998 to 2004, he was an Assistant Professor with the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. He is currently a Professor with the Mechanical Engineering Department and the Center for Control, Dynamical Systems and Computation at the University of California, Santa Barbara.

His main research interest is multi-agent networks with application to robotic coordination, distributed computing and power networks; he has worked on problems in vehicle routing, geometric control, and motion planning. He is the coauthor of "Geometric Control of Mechanical Systems" (Springer, 2004, 0-387-22195-6) and "Distributed Control of Robotic Networks" (Princeton, 2009, 978-0-691-14195-4). His students' papers were finalists for the Best Student Paper Award at the IEEE Conference on Decision and Control (2002, 2005, 2007), and the American Control Conference (2005, 2006, 2010). He has published more than 150 papers in international journals, books, and refereed conferences. He has served, or is serving, on the editorial boards of the "IEEE Transactions on Automatic Control", the "ESAIM: Control, Optimization, and the Calculus of Variations" and the "SIAM Journal of Control and Optimization".