Efficient routing algorithms for multiple vehicles

with no explicit communications

Alessandro Arsie   Ketan Savla   Emilio Frazzoli

Abstract

In this paper we consider a class of dynamic vehicle routing problems, in which a number of mobile agents in the plane must visit target points generated over time by a stochastic process. It is desired to design motion coordination strategies in order to minimize the expected time between the appearance of a target point and the time it is visited by one of the agents. We propose control strategies that, while making minimal or no assumptions on communications between agents, provide the same level of steady-state performance achieved by the best known decentralized strategies. In other words, we demonstrate that inter-agent communication does not improve the efficiency of such systems, but merely affects the rate of convergence to the steady state. Furthermore, the proposed strategies do not rely on the knowledge of the details of the underlying stochastic process. Finally, we show that our proposed strategies yield an efficient, pure Nash equilibrium in a game theoretic formulation of the problem, in which each agent's objective is to maximize the expected value of the "time spent alone" at the next target location. Simulation results are presented and discussed.

I. INTRODUCTION

A very active research area today addresses coordination of several mobile agents: groups of autonomous robots and large-scale mobile networks are being considered for a broad class of applications, ranging from environmental monitoring, to search and rescue operations, and national security.

An area of particular interest is concerned with the generation of efficient cooperative strategies for several mobile agents to move through a certain number of given target points, possibly avoiding obstacles or threats [2]–[6]. Trajectory efficiency in these cases is understood in terms of cost for the agents: in other words, efficient trajectories minimize the total path length, the time needed to complete the task, or the fuel/energy expenditure. A related problem has been investigated as the Weapon-Target

Assignment (WTA) problem, in which mobile agents are allowed to team up in order to enhance the probability of a favorable outcome in a target engagement [7], [8]. In this setup, targets locations are known and an assignment strategy is sought that maximizes the global success rate. In a biological setting, the closest parallel to many of these problems is the development of foraging strategies, and of territorial vs. gregarious behaviors [9], in which individuals choose to identify and possibly defend a hunting ground.

In this paper we consider a class of cooperative motion coordination problems, to which we can refer as *dynamic vehicle routing*, in which service requests are not known a priori, but are dynamically generated over time by a stochastic process in a geographic region of interest. Each service request is associated to a target point in the plane, and is fulfilled when one of a team of mobile agents visits that point. For example, service requests can be thought of as threats to be investigated in a surveillance application, events to be measured in an environmental monitoring scenario, and as information packets to be picked up and delivered to a user in a wireless sensor network. It is desired to design a control strategy for the mobile agents that provably minimizes the expected waiting time between the issuance of a service request and its fulfillment. In other words, our focus is on the quality of service as perceived by the "end user," rather than, for example, fuel economies achieved by the mobile agents. Similar problems were also considered in [10], [11], and decentralized strategies were presented in [12]. This problem has connections to the Persistent Area Denial (PAD) and area coverage problems discussed, e.g., in [4], [13]–[15].

A common theme in cooperative control is the investigation of the effects of different communication and information sharing protocols on the system performance. Clearly, the ability to access more information at each single agent can not decrease the performance level; hence, it is commonly believed that by providing better communication among agents will improve the system's performance. In this paper, we prove that there are certain dynamic vehicle routing problems which can, in fact, be solved (almost) optimally without any explicit communication between agents; in other words, the no-communication constraint in such cases is not binding, and does not limit the steady-state performance. The main contribution of this paper is the introduction of a motion coordination strategy that does not require any explicit communication between agents, while achieving provably optimal performance in certain conditions.

The paper is structured as follows: in Section II we set up and formulate the problem we investigate in the paper. In Section III we introduce the proposed solution algorithms, and discuss their characteristics. Section IV is the technical core of the paper, in which we prove the convergence of the performance provided by the proposed algorithms to a critical point (either a local minimum or a saddle point) of the global performance function. Moreover, we show that any optimal configuration corresponds to a class of tessellations of the plane that we call Median Voronoi Tessellations. Section V is devoted to a game-theoretic interpretation of our result in which the agents are modeled as rational autonomous decision makers trying to maximize their

own utility function. We prove that, our algorithms yield an efficient, pure Nash equilibrium in a game theoretic formulation of the problem. In Section VI we present some numerical results, while Section VII is dedicated to final remarks and further extensions of this line of research.

II. PROBLEM FORMULATION

Let $\Omega \subset \mathbb{R}^2$ be a convex domain on the plane, with non-empty interior; we will refer to Ω as the *workspace*. A stochastic process generates *service requests* over time, which are associated to points in Ω; these points are also called *targets*. The process generating service requests is modeled as a spatio-temporal Poisson point process, with temporal intensity $\lambda > 0$, and an absolutely continuous spatial distribution described by the density function $\varphi : \Omega \to \mathbb{R}_+$, with bounded and convex support within Ω (i.e., $\varphi(q) > 0 \Leftrightarrow q \in \mathcal{Q} \subseteq \Omega$, with $\mathcal{Q}$ bounded and convex). The spatial density function φ is normalized in such a way that $\int_\Omega \varphi(q)\, dq = 1$. Both λ and φ are not necessarily known.

A spatio-temporal Poisson point process is a collection of functions $\{\mathcal{P} : \overline{\mathbb{R}}_+ \to 2^\Omega\}$ such that, for any $t > 0$, $\mathcal{P}(t)$ is a random collection of points in Ω, representing the service requests generated in the time interval $[0, t)$, and such that

- The total numbers of events generated in two disjoint time-space regions are *independent* random variables;
- The total number of events occurring in an interval $[s, s + t)$ in a measurable set $\mathcal{S} \subseteq \Omega$ that satisfies

$$\Pr\left[\operatorname{card}\left((\mathcal{P}(s + t) - \mathcal{P}(s)) \cap \mathcal{S}\right) = k\right] = \frac{\exp(-\lambda t \cdot \varphi(\mathcal{S}))(\lambda t \cdot \varphi(\mathcal{S}))^k}{k!}, \quad \forall k \in \mathbb{N},$$

where $\varphi(\mathcal{S})$ is a shorthand for $\int_\mathcal{S} \varphi(q)\, dq$.

Each particular function $\mathcal{P}$ is a realization, or trajectory, of the Poisson point process. A consequence of the properties defining Poisson processes is that the *expected* number of targets generated in a measurable region $S \subseteq \Omega$ during a time interval of length Δt is given by:

$$\mathrm{E}[\operatorname{card}\left((\mathcal{P}(t + \Delta t) - \mathcal{P}(t)) \cap \mathcal{S}\right)] = \lambda \Delta t \cdot \varphi(\mathcal{S}).$$

Without loss of generality, we will identify service requests with targets points, and label them in order of generation; in other words, given two targets $e_i, e_j \in \mathcal{P}(t)$, with $i < j$, the service request associated with these target have been issued at times $t_i \leq t_j \leq t$ (since events are almost never generated concurrently, the inequalities are in fact strict almost surely).

A service request is fulfilled when one of m mobile agents, modeled as point masses, moves to the target point associated with it; m is a possibly large, but finite number. Let $p(t) = (p_1(t), p_2(t), \ldots, p_m(t)) \in \Omega^m$ be a vector describing the positions of the agents at time t. (We will tacitly use a similar notation throughout the paper). The agents are free to move, with bounded speed, within the workspace Ω; without loss of generality, we will assume that the maximum speed is unitary. In other words,

the dynamics of the agents are described by differential equations of the form

$$\frac{d\,p_i(t)}{dt} = u_i(t), \quad \text{with } \|u_i(t)\| \leq 1, \quad \forall t \geq 0, i \in \{1, \ldots, m\}. \tag{1}$$

The agents are identical, and have unlimited range and target-servicing capability.

Let $\mathcal{B}_i(t) \subset \Omega$ indicate the set of targets serviced by the i-th agent up to time t. (By convention, $\mathcal{B}_i(0) = \emptyset$, $i = 1, \ldots, m$). We will assume that $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ if $i \neq j$, i.e., that service requests are fulfilled by at most one agent. (In the unlikely event that two or more agents visit a target at the same time, the target is arbitrarily assigned to one of them).

Let $\mathcal{D} : t \to 2^\Omega$ indicate (a realization of) the stochastic process obtained by combining the service request generation process $\mathcal{P}$ and the removal process caused by the agents servicing outstanding requests; in other words,

$$\mathcal{P}(t) = \mathcal{D}(t) \cup \mathcal{B}_1(t) \cup \ldots \cup \mathcal{B}_m(t), \qquad \mathcal{D}(t) \cap \mathcal{B}_i(t) = \emptyset, \ \forall i \in \{1, \ldots, m\}.$$

The random set $\mathcal{D}(t) \subset \Omega$ represents the *demand*, i.e., the service requests outstanding at time t; let $n(t) := \operatorname{card}(\mathcal{D}(t))$.

Our objective in this paper will be the design of motion coordination strategies that allow the mobile agents to fulfill service requests efficiently (we will make this more precise in the following). In particular, in this paper we will concentrate on motion coordination strategies of the following two forms:

$$\pi_i : (p_i, \mathcal{B}_i, \mathcal{D}) \mapsto u_i, \qquad i \in \{1, \ldots, m\}, \tag{2}$$

and

$$\pi_i : (p_1, \ldots, p_m, \mathcal{B}_i, \mathcal{D}) \mapsto u_i, \qquad i \in \{1, \ldots, m\}. \tag{3}$$

An agent executing a control policy of the form (2) relies on the knowledge of its own current position, on a record of targets it has previously visited, and on the current demand. In other words, such control policies do not need any explicit information exchange between agents; as such, we will refer to them as *no communication* (nc) policies. Such policies are trivially decentralized.

Following the first control policy, it is certainly true that agent i must know the set $\mathcal{D}(t)$ of outstanding targets, but this does not mean that this agent must know *any* of the sets $\mathcal{B}_j(t)$, for $j \neq i$. It is true that agents implicitly share/update the set of tasks that must be completed, but this does not mean that there is any communication among the agents. This is how the agents can have complete knowledge of $\mathcal{D}(t)$ without any communication among themselves (notice that it is possible to find out other ways to achieve a complete knowledge of $\mathcal{D}(t)$ without communication, what follows is simply one of the possible ways). Indeed, suppose that the targets broadcast their service requests and their positions to the whole environment and they immediately stop broadcasting as soon as they are visited by one of the agents. Suppose also that each agent is able to receive

the signal broad-casted by each target and is able to detect the corresponding target position. In this set-up each agent knows precisely what are the outstanding targets, without exchanging directly any information with other agents. On the other hand, each agent seeing that a target has stopped broadcasting its service request can infer that that particular target has been visited. This is however not a communication among the agents, but simply an inference (and moreover the identity of the agent which has serviced that particular target is unknown to all other agents). Therefore the agents are not connected among them, they are instead connected with the targets, in the sense that they are able to receive and locate the signals broad-casted by all targets. The no-communication hypothesis is referred to the lack of direct communication among the agents, not to the inferences that each single agent can perform observing how the service requests of the outstanding targets are satisfied.

Another possibility to get a complete knowledge of $\mathcal{D}(t)$ without explicit communication among the agents is as follows. Suppose there exists a satellite broadcasting the positions of all outstanding targets to the agents (the satellite simply broadcasts a list of available targets with their positions). Each time a target is serviced the satellite recognizes that that specific target is no more available (for instance because it has been destroyed) and drops it from the list of available targets. Each agent receives the same list from the satellite and also in this case, if a target disappear from the list, an agent which has not serviced that target can infer that it has been serviced by some other agent.

On the other hand, an agent executing a control policy of the form (3) can sense the current position of other agents, but still has information only on the targets itself visited in the past. We call these *sensor-based* (sb) policies, to signify the fact that only factual information is exchanged between agents—as opposed to information related to intent and past history. Note that both families of coordination policies rely, in principle, on the knowledge of the locations of all outstanding targets. However only local target sensing will be necessary in practice. This last claim is difficult to justify from a theoretical point of view, and it is better understood in terms of simulations. For a more complete treatment of this issue, in particular the effect of limiting sensing radius for target sensing and its effect on the system performance, see the review [16].

A policy $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$ is said to be *stabilizing* if, under its effect, the expected number of outstanding targets does not diverge over time, i.e., if

$$\overline{n}_\pi = \lim_{t \to +\infty} \mathrm{E}[n(t) \| \dot{p}_i(t) = \pi_i(p(t), \mathcal{B}_i(t), \mathcal{D}(t)), i \in \{1, \ldots, m\}] < \infty. \tag{4}$$

Intuitively, a policy is stabilizing if the mobile agents are able to visit targets at a rate that is—on average—at least as fast as the rate at which new service requests are generated.

Let T_j be the time elapsed between the issuance of the j-th service request, and the time it is fulfilled. If the system is

stable, then the following balance equation (also known as Little's formula [17]) holds:

$$\overline{n}_\pi = \lambda \overline{T}_\pi, \tag{5}$$

where $\overline{T}_\pi := \lim_{j \to \infty} \mathrm{E}[T_j]$ is the system time under policy π, i.e., the expected time a service request must wait before being fulfilled, given that the mobile agents follow the strategy defined by π. Note that the system time $\overline{T}_\pi$ can be thought of as a measure of the quality of service, as perceived by the "user" issuing the service requests.

At this point we can finally state our problem: we wish to devise a policy that is (i) stabilizing, and (ii) yields a quality of service (i.e., system time) achieving, or approximating, the theoretical optimal performance given by

$$\overline{T}_{\mathrm{opt}} = \inf_{\pi \text{ stabilizing}} \overline{T}_\pi \tag{6}$$

Centralized and decentralized strategies are known that optimize or approximate (6) in a variety of cases of interest [11], [12], [18], [19]. However, all such strategies rely either on a central authority with the ability to communicate to all agents, or on the exchange of certain information about each agent's strategy with other neighboring agents. In addition, these policies require the knowledge of the spatial distribution φ; decentralized versions of these implement versions of Lloyd's algorithm for vector quantization [20].

In the remainder of this paper, we will investigate how the additional constraints posed on the exchange of information between agents by the models (2) and (3) impact the achievable performance and quality of service. Remarkably, the policies we will present do not rely on the knowledge of the spatial distribution φ, and are a generalized version of MacQueen's clustering algorithm [21].

III. CONTROL POLICY DESCRIPTION

In this section, we introduce two control policies of the forms, respectively, (2) and (3). An illustration of the two policies is given in Figure 1.

A. A control policy requiring no explicit communication

Let us begin with an informal description of a policy π_{nc} requiring no explicit information exchange between agents. At any given time t, each agent computes its own control input according to the following rule:

1) If $\mathcal{D}(t)$ is not empty, move towards the nearest outstanding target.

2) If $\mathcal{D}(t)$ is empty, move towards the point minimizing the average distance to targets *serviced in the past* by each agent. If there is no unique minimizer, then move to the nearest one.
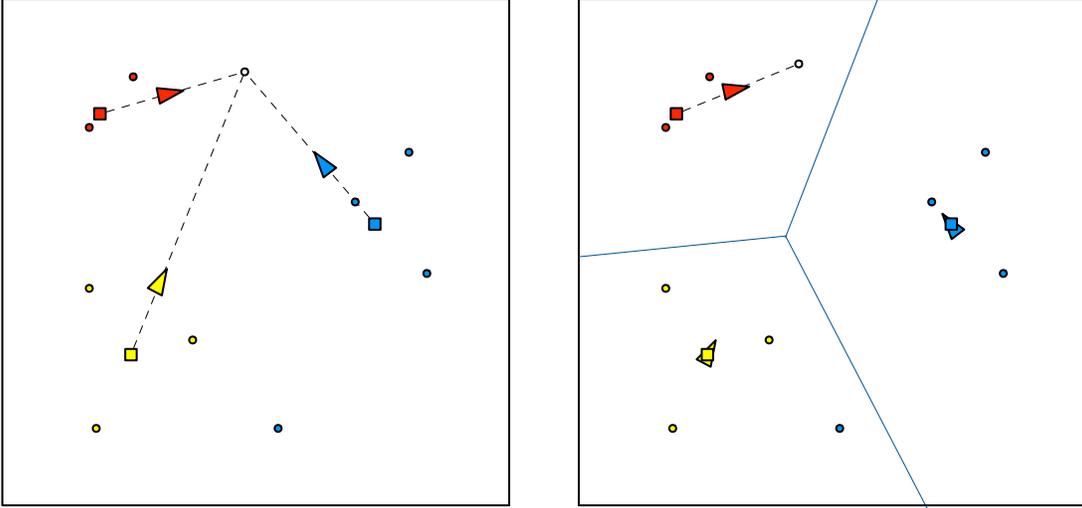
Fig. 1. An illustration of the two control policies proposed in Section III. While no targets are outstanding, vehicles wait at the point that minimizes the average distance to targets they have visited in the past; such points are depicted as squares, while targets are circles and vehicles triangles. In the no-communication policy, at the appearance of a new target, all vehicles pursue it (left). In the sensor-based policy, only the vehicle that is closest to the target will pursue it (right).

In other words, we set

$$\pi_{\mathrm{nc}}(p_i(t), \mathcal{B}_i(t), \mathcal{D}(t)) = \mathrm{vers}(F_{\mathrm{nc}}(p_i(t), \mathcal{B}_i(t), \mathcal{D}(t)) - p_i(t)), \tag{7}$$

where

$$F_{\mathrm{nc}}(p_i, \mathcal{B}_i, \mathcal{D}) = \begin{cases} \arg\min_{q \in \mathcal{D}} \|p_i - q\|, & \text{if } \mathcal{D} \neq \emptyset, \\ \arg\min_{q \in \Omega} \sum_{e \in \mathcal{B}_i} \|e - q\|, & \text{otherwise}, \end{cases} \tag{8}$$

$\|\cdot\|$ is the Euclidean norm, and

$$\mathrm{vers}(v) = \begin{cases} v/\|v\|, & \text{if } v \neq 0, \\ 0 & \text{otherwise}. \end{cases}$$

The convex function $W : q \mapsto \sum_{e \in \mathcal{B}} \|q - e\|$, often called the (discrete) Weber function in the facility location literature [22], [23] (modulo normalization by $\mathrm{card}(\mathcal{B})$), is not strictly convex only when the point set $\mathcal{B}$ is empty—in which case we set $W(\cdot) = 0$ by convention— or contains an even number of collinear points. In such cases, the minimizer nearest to p_i in (8) is chosen. We will call the point $p_i^*(t) = F_{\mathrm{nc}}(\cdot, \mathcal{B}_i(t), \emptyset)$ the *reference point* for the i-th agent at time t.

In the π_{nc} policy, whenever one or more service requests are outstanding, all agents will be pursuing a target; in particular, when only one service request is outstanding, all agents will move towards it. When the demand queue is empty, agents will either (i) stop at the current location, if they have visited no targets yet, or (ii) move to their reference point, as determined by the set of targets previously visited.

B. A sensor-based control policy

The control strategy in the previous section can be modified to include information on the current position of other agents, if available (e.g., through on-board sensors). In order to present the new policy, indicate with $\mathcal{V}(p) = \{\mathcal{V}_1(p), \mathcal{V}_2(p), \ldots, \mathcal{V}_m(p)\}$ the Voronoi partition of the workspace Ω, defined as:

$$\mathcal{V}_i(p) = \{q \in \Omega : \|q - p_i\| \leq \|q - p_j\|, \forall j = 1 \ldots m\}. \tag{9}$$

As long as an agent has never visited any target, i.e., as long as $\mathcal{B}_i(t) = \emptyset$, it executes the π_{nc} policy. Once an agent has visited at least one target, it computes its own control input according to the following rule:

1) If $\mathcal{D}(t) \cap \mathcal{V}_i(t)$ is not empty, move towards the nearest outstanding target in the agent's own Voronoi region.

2) If $\mathcal{D}(t) \cap \mathcal{V}_i(t)$ is empty, move towards the point minimizing the average distance to targets in $\mathcal{B}_i(t)$. If there is no unique minimizer, then move to the nearest one.

In other words, we set

$$\pi_{\mathrm{sb}}(p(t), \mathcal{B}_i(t), \mathcal{D}(t)) = \mathrm{vers}(F_{\mathrm{sb}}(p(t), \mathcal{B}_i(t), \mathcal{D}(t)) - p_i(t)), \tag{10}$$

where

$$F_{\mathrm{sb}}(p, \mathcal{B}_i, \mathcal{D}) = \begin{cases} \arg\min_{q \in \mathcal{D}} \|p_i - q\|, & \text{if } \mathcal{D} \neq \emptyset, \text{ and } \mathcal{B}_i = \emptyset \\[2mm] \arg\min_{q \in \mathcal{D} \cap \mathcal{V}_i(p)} \|p_i - q\|, & \text{if } \mathcal{D} \cap \mathcal{V}_i \neq \emptyset, \text{ and } \mathcal{B}_i \neq \emptyset \\[2mm] \arg\min_{q \in \Omega} \sum_{e \in \mathcal{B}_i} \|e - q\|, & \text{otherwise.} \end{cases} \tag{11}$$

In the π_{sb} policy, at most one agent will be pursuing a given target, at any time after an initial transient that terminates when all agents have visited at least one target each. The agents' behavior when no outstanding targets are available in their Voronoi region is similar to that determined by the π_{nc} policy previously discussed, i.e., they move to their reference point, determined by previously visited targets.

Remark 1: While we introduced Voronoi partitions in the definition of the control policy, the explicit computation of each agent's Voronoi region is not necessary. In fact, each agent only needs to check whether it is the closest agent to a given target or not. In order to check whether a target point q is in the Voronoi region of the i-th agent, it is necessary to know the current position only of agents within a circle or radius $\|p_i - q\|$ centered at q (see Figure 2). For example, if such circle is empty, then q is certainly in $\mathcal{V}_i$; if the circle is not empty, distances of the agents within it to the target must be compared. This provides a degree of spatial decentralization—with respect to other agents—that is even stronger than that provided by restricting communications to agents sharing a boundary in a Voronoi partition (i.e., neighboring agents in the Delaunay graph, dual to the partition (9)).
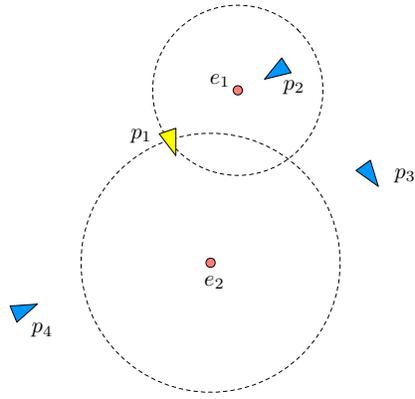
Fig. 2. Implicit computation of Voronoi regions: Even though target e_1 is the nearest target to p_1, it is not in the Voronoi region of the 1st agent. In fact, the circle of radius $\|e_1 - p_1\|$ centered at e_1 contains p_2, and the 2nd agent is closer to e_1. However, the circle of radius $\|e_2 - p_1\|$ centered at e_2 does not contain any other agent, ensuring that e_2 is in the Voronoi region generated by p_1.

Remark 2: The sensor-based policy is more efficient than the no-communication policy in terms of the length of the path traveled by each agent, since there is no duplication of effort as several agents pursue the same target (unless a target is on the boundary of two or more Voronoi regions). However, in terms of "quality of service," we will show that there is no difference between the two policies, for low target generation rates. Numerical results show that the sensor-based policy is more efficient in a broader range of target generation rates, and in fact provides almost optimal performance both in light and heavy load conditions.

IV. PERFORMANCE ANALYSIS IN LIGHT LOAD

In this section we analyze the performance of the control policies proposed in the previous section. In particular, we concentrate our investigation on the light load case, in which the target generation rate is very small, i.e., as $\lambda \to 0^+$. This will allow us to prove analytically certain interesting and perhaps surprising characteristics of the proposed policies. The performance analysis in the general case is more difficult; we will discuss the results of numerical investigation in Section VI, but no analytical results are available at this time.

A. Overview of the system behavior in the light load regime

Before starting a formal analysis, let us summarize the key characteristics of the agents' behavior in light load, i.e., for small values of λ.

1) At the initial time the m agents are assumed to be deployed at general positions in Ω, and the demand queue is empty, $\mathcal{D}(0) = \emptyset$.

2) The agents do not move until the first service request appears. At that time, if the policy π_{nc} is used, all agents will start moving towards the first target. If the sensor-based policy π_{sb} is used, only the closest agent will move towards the target.

3) As soon as one agent reaches the target, all agents start moving towards their current reference point, and the process continues.

For small λ, with high probability (i.e., with probability approaching 1 as $\lambda \to 0^+$) at most one service request is outstanding at any given time. In other words, new service requests are generated so rarely that most of the time agents will be able to reach a target and return to their reference point before a new service request is issued. This is just the content of the following.

Proposition 3: Each agent will be able to return to its reference point before the generation of a new service request infinitely often with high probability as $\lambda \to 0^+$.

Proof: Let t_1 be such that $\mathcal{B}_i(t_1) \neq \emptyset$, for all $i \in \{1, \ldots, m\}$. Such a time exists, almost surely, basically because of the fairness of the proposed policies. At time t_1 all agents will be within $\mathcal{Q}$. Let $n_1 = \mathrm{card}(\mathcal{D}(t_1))$ be the total number of outstanding targets at time t_1. An upper bound on the time needed to visit all targets in $\mathcal{D}(t_1)$ is $n_1(\mathrm{diam}(\mathcal{Q}))$. When there are no outstanding targets, agents move to their reference points, reaching them in at most $\mathrm{diam}(\mathcal{Q})$ units of time.

The time needed to service the initial targets and go to the reference configuration is hence bounded by $t_{\mathrm{ini}} \leq t_1 + (n_1 + 1)\mathrm{diam}(\mathcal{Q})$. The probability that at the end of this initial phase the number of targets is reduced to zero is

$$P\left[n(t_{\mathrm{ini}}) = 0\right] = \exp(-\lambda(t_{\mathrm{ini}} - t_1)) \geq \exp(-\lambda(n_1 + 1)\mathrm{diam}(\mathcal{Q})),$$

that is, $P\left[n(t_{\mathrm{ini}}) = 0\right] \to 1^-$ as $\lambda \to 0^+$. As a consequence, after an initial transient, all targets will be generated with all agents waiting at their reference points, and an empty demand queue, with high probability as $\lambda \to 0^+$. ∎

Consider the j-th service request, generated at time t_j. Assuming that, at t_j, all agents are at their reference positions, the expected system time $\mathrm{E}[T_j]$ can be computed as

$$\mathrm{E}[T_j] = \int_\Omega \min_{i=1,\ldots,m} \|p_i^*(t_j) - q\| \, \varphi(q)dq.$$

Assume for now that the sequences $\{p_i^*(t_j) : j \in \mathbb{N}\}$ converge, and let

$$\lim_{j \to +\infty} p_i^*(t_j) = \hat{p}_i^*.$$

Note that $\hat{p}_i^*$ is a random variable, the value of which depends in general on the particular realization of the target generation process. If all service requests are generated with the agents at their reference positions, the average service time (for small

λ) can be evaluated as

$$\overline{T}_{\mathrm{nc}} = \overline{T}_{\mathrm{sb}} = \int_{\Omega} \min_{i=1,\ldots,m} \|\hat{p}_i^* - q\| \; \varphi(q)dq = \sum_{i=1}^{m} \int_{\mathcal{V}_i(p^*)} \|\hat{p}_i^* - q\| \; \varphi(q)dq. \tag{12}$$

Since the system time depends on the random variable $\hat{p}^* = (\hat{p}_1^*, \ldots, \hat{p}_m^*)$, it is itself a random variable. The function appearing

on the right hand side of the above equation, relating the system time to the asymptotic location of reference points, is called

the continuous multi-median function [22]. This function admits a global minimum (in general not unique) for all non-singular

density functions φ, and in fact it is known [11] that the optimal performance in terms of system time is given by

$$\overline{T}_{\mathrm{opt}} = \min_{p \in \Omega^m} \sum_{i=1}^{m} \int_{\mathcal{V}_i(p)} \|p_i - q\| \; \varphi(q)dq. \tag{13}$$

In the following, we will investigate the convergence of the reference points as new targets are generated, in order to draw

conclusions about the average system time $\overline{T}$ in light load. In particular, we will prove not only that the reference points

converge with high probability (as $\lambda \to 0^+$) to a local critical point (more precisely, either local minima or saddle points)

for the average system time, but also that the limiting reference points $\hat{p}^*$ are *generalized medians* of their respective Voronoi

regions, where

Definition 4 (Generalized median): The *generalized median* of a set $\mathcal{S} \subset \mathbb{R}^n$ with respect to a density function $\varphi : \mathcal{S} \to \overline{\mathbb{R}}_+$

is defined as

$$\overline{p} := \arg \min_{p \in \mathbb{R}^n} \int_{\mathcal{S}} \|p - q\|\varphi(q) \; dq.$$

We call the resulting Voronoi tessellation *Median Voronoi Tessellation* (MVT for short), in analogy with what is done with

Centroidal Voronoi Tessellations. A formal definition is as follows:

Definition 5 (Median Voronoi Tessellation): A Voronoi tessellation $\mathcal{V}(p) = \{\mathcal{V}_1(p), \ldots, \mathcal{V}_m(p)\}$ of a set $\mathcal{S} \subset \mathbb{R}^n$ is called

a *Median Voronoi Tessellation* of $\mathcal{S}$ with respect to the density function φ if the ordered set of generators p is equal to the

ordered set of generalized medians of the sets in $\mathcal{V}(p)$ with respect to φ, i.e., if

$$p_i = \arg \min_{s \in \mathbb{R}^n} \int_{\mathcal{V}_i(p)} \|s - q\|\varphi(q) \; dq, \qquad \forall i \in \{1, \ldots, m\}.$$

Since the proof builds on a number of intermediate results, we provide an outline of the argument as a convenience to the

reader.

1) First, we prove that the reference point of any agent that visits an unbounded number of targets over time converges

   almost surely.

2) Second, we prove that, if $m \geq 1$ agents visit an unbounded number of targets over time, their reference points will

   converge to the generators of a MVT almost surely, as long as agents are able to return to their reference point infinitely

   often.

3) Third, we prove that all agents will visit an unbounded number of targets (this corresponds to a property of distributed algorithms that is often called *fairness* in computer science).

4) Finally, we prove that agents are able to return to their reference points infinitely often with high probability as $\lambda \to 0^+$.

Combining these steps, together with (6), will allow us to state that the reference points converge to a local critical point of the system time, with high probability as $\lambda \to 0^+$.

B. Convergence of reference points

Let us consider an agent i, such that

$$\lim_{t \to +\infty} \text{card}(\mathcal{B}_i(t)) = +\infty,$$

i.e., an agent that services an unbounded number of requests over time. Since the number of agents m is finite, and the expected number of targets generated over a time interval $[0, t)$ is proportional to t, at least one such agent will always exist. In the remainder of this section, we will drop the subscript i, since we will consider only this agent, effectively ignoring all others for the time being.

For any finite t, the set $\mathcal{B}(t)$ will contain a finite number of points. Assuming that $\mathcal{B}(t)$ contains at least three non-collinear points, the discrete Weber function $p \mapsto \sum_{q \in \mathcal{B}(t)} \|p - q\|$ is strictly convex, and has a unique optimizer $p^*(t) = \arg\min_{p \in \Omega} \sum_{q \in \mathcal{B}(t)} \|p - q\|$. The optimal point $p^*(t)$ is called the Fermat-Torricelli (FT) point—or the Weber point in the location optimization literature—associated with the set $\mathcal{B}(t)$; see [23]–[25] for a historical review of the problem and for solution algorithms.

It is known that the FT point is unique and algebraic for any set of non-collinear points. While there are no general analytic solutions for the location of the FT point associated to more than 4 points, numerical solutions can be easily constructed relying on the convexity of the Weber function, and on the fact that it is differentiable for all points not in $\mathcal{B}$. Polynomial-time approximation algorithms are also available (see, e.g., [26], [27]). Remarkably, a simple mechanical device can be constructed to solve the problem, based on the so-called Varignon frame, as follows. Holes are drilled on a horizontal board, at locations corresponding to the points in $\mathcal{B}$. A string attached to a unit mass is passed through each of these holes, and all strings are tied together at one end. The point reached by the knot at equilibrium is a FT point for $\mathcal{B}$.

Some useful properties of FT points are summarized below. If there is a $q_0 \in \mathcal{B}$ such that

$$\left\| \sum_{q \in \mathcal{B} \setminus q_0} \text{vers}(q_0 - q) \right\| \le 1 \tag{14}$$

then $p^* = q_0$ is a FT point for $\mathcal{B}$. If no point in $\mathcal{B}$ satisfies such condition, then the FT point p^* can be found as a solution of

the following equation:

$$\sum_{q \in \mathcal{B}} \text{vers}(p^* - q) = 0. \tag{15}$$

In other words, p^* is simply the point in the plane at which the sum of unit vectors starting from it and directed to each of the points in $\mathcal{B}$ is equal to zero; this point is unique if $\mathcal{B}$ contains non-collinear points. Clearly, the FT point is in the convex hull of $\mathcal{B}$.

Note that the points in $\mathcal{B}(t)$ are randomly sampled from an unknown absolutely continuous distribution, described by a spatial density function $\tilde{\varphi}$—which is not necessarily the same as φ, and in general is time-varying, depending on the past actions of all agents in the system. Even though $\tilde{\varphi}$ is not known, it can be expressed as

$$\tilde{\varphi}(q, t) = \begin{cases} \dfrac{\varphi(q)}{\int_{\mathcal{I}(t)} \varphi(q)\, dq} & \text{if } q \in \mathcal{I}(t) \\[2mm] 0 & \text{otherwise,} \end{cases}$$

for some convex set $\mathcal{I}(t)$ containing $p(t)$. (In practical terms, such set will be the Voronoi region generated by $p(t)$, in any case this remark is given as an aside and it is not important in the analysis of the proposed policies).

The function $t \mapsto p^*(t)$ is piecewise constant, i.e., it changes value at the discrete time instants $\{t_j : j \in \mathbb{N}\}$ at which the agent visits new targets. As a consequence, we can concentrate on the sequence $\{p^*(t_j) : j \in \mathbb{N}\}$, and study its convergence.

Definition 6: For any $t > 0$, let the *solution set* $C(t)$ be defined as

$$C(t) := \left\{ p \in \Omega : \left\| \sum_{q \in \mathcal{B}(t)} \text{vers}(p - q) \right\| \le 1 \right\}.$$

An example of such set is shown in Figure 3. Before exploring the properties of these sets, let us define e_j the target point corresponding to the j-th service request. The reason for introducing such solution sets is that they have quite remarkable properties as shown by the following result.

Proposition 7: For any $j \in \mathbb{N}$, $p^*(t_{j+1}) \in C(t_j)$. More specifically, if $e_{j+1} \notin C(t_j)$ (i.e., the target point associated to the j-th service request is outside the solution set) then the FT point $p^*(t_{j+1})$ is on the boundary of $C(t_j)$. If $e_{j+1} \in C(t_j)$, then $p^*(t_{j+1}) = e_{j+1}$.

Proof: If e_{j+1} lies outside $C(t_j)$, we search for $p^*(t_{j+1})$ as the solution of the equation

$$\sum_{q \in \mathcal{B}(t_j)} \text{vers}(p - q) + \text{vers}(p - e_{j+1}) = 0,$$

from which it turns out immediately

$$\left\| \sum_{q \in \mathcal{B}(t_j)} \text{vers}(p - q) \right\| = \| -\text{vers}(p - e_{j+1}) \| = 1,$$
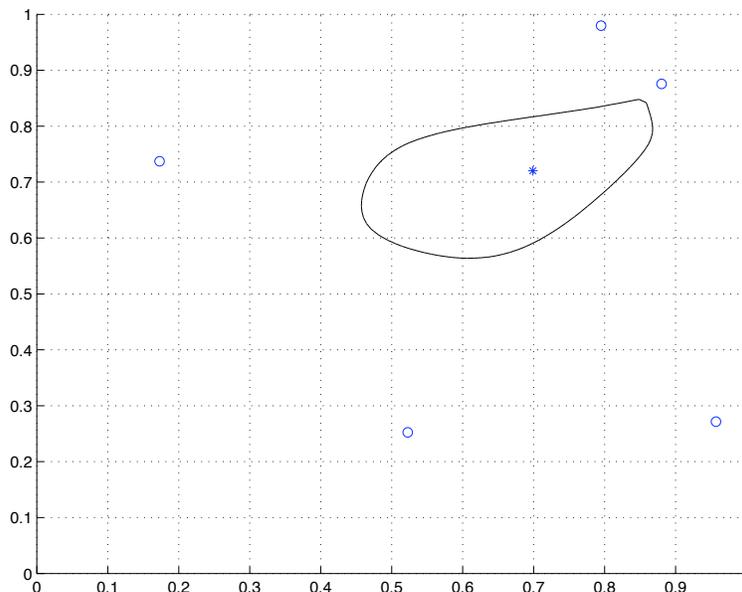
Fig. 3. Example of a Fermat-Torricelli point (star) and solution set corresponding to five target points (circles). Upon the addition of an arbitrarily chosen sixth target point, the Fermat-Torricelli is guaranteed to remain within the region bounded by the curve.

thus $p^*(t_{j+1}) \in \partial C(t_j)$. Notice that is is not true in general that the solution $p^*(t_{j+1})$ will lie on the line connecting $p^*(t_j)$ with the new target e_{j+1}. In the other case, if e_{j+1} lies in $C(t_j)$, then it satisfies condition (14), and is the new FT point. ∎

Now, in order to prove that the $\{p^*(t_j)\}_{j \in \mathbb{N}}$ converges to a point $\hat{p}^*$, we will prove that the diameter of the solution set $C(t_j)$ vanishes almost surely as j tends to infinity. First we need the following result.

Proposition 8: If $\mathcal{Q} = \mathrm{Supp}(\varphi)$ is convex with non-empty interior, then $p^*(t) \in \mathrm{int}(\mathcal{Q})$ almost surely, for all t such that $\mathcal{B}(t) \neq \emptyset$.

Proof: For any non-empty target set $\mathcal{B}(t)$, the FT point lies within the convex hull of $\mathcal{B}(t)$. All points in the set $\mathcal{B}(t)$ are contained within the interior of $\mathcal{Q}$ with probability one, since the boundary of $\mathcal{Q}$ is a set of measure zero. Since $\mathrm{int}(\mathcal{Q})$ is convex, and $\mathcal{B}(t) \subset \mathrm{int}(\mathcal{Q})$ almost surely, $p^*(t) \in \mathrm{co}(\mathcal{B}(t)) \subset \mathrm{int}(\mathcal{Q})$, almost surely. ∎

Proposition 9: If the support of φ is convex and bounded,

$$\lim_{j \to +\infty} \mathrm{diam}(C(t_j)) = 0, \qquad \text{a.s.}$$

Proof: Consider a generic point $p \in C(t_j)$, and let $\delta = p - p^*(t_j)$, $\alpha_q = \arccos\left[\mathrm{vers}(p - p^*(t_j)) \cdot \mathrm{vers}(q - p^*(t_j))\right]$, and $\alpha'_q = \arccos\left[\mathrm{vers}(p - p^*(t_j)) \cdot \mathrm{vers}(q - p)\right]$, see Figure 4.

Since $p \in C(t_j)$, the magnitude of the sum of unit vectors $\sum_{q \in \mathcal{B}(t_j)} \mathrm{vers}(p - q)$ is no more than one, and the following
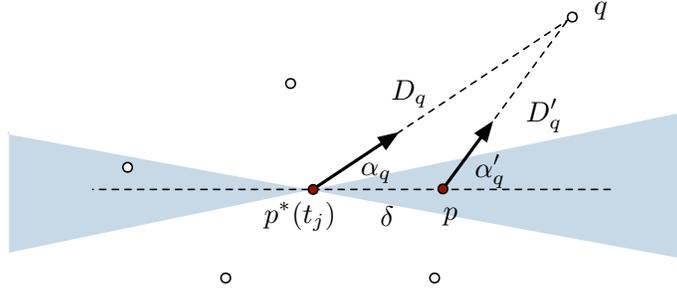
Fig. 4. Geometric constructions in the proof of Proposition 9.

inequality is true:

$$\left|\left(\sum_{q\in\mathcal{B}(t_j)}\mathrm{vers}(p-q)\right)\cdot\mathrm{vers}(p-p^*(t_j))\right|=\left|\sum_{q\in\mathcal{B}(t_j)}(\mathrm{vers}(p-q)\cdot\mathrm{vers}(p-p^*(t_j)))\right|=\left|\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q')\right|\leq 1. \qquad (16)$$

Using elementary planar geometry, we obtain that

$$\alpha_q'-\alpha_q\geq\sin(\alpha_q'-\alpha_q)=\frac{\delta\sin(\alpha_q)}{\|q-p\|}.$$

Pick a small angle $0<\alpha_{\min}<\pi/2$, and let

$$\mathcal{B}_{\alpha_{\min}}(t_j)=\{q\in\mathcal{B}(t_j):\sin(\alpha_q)\geq\sin(\alpha_{\min})\}.$$

(In other words, $\mathcal{B}_{\alpha_{\min}}(t)$ contains all points in $\mathcal{B}$ that are not in a conical region of half-width $\alpha_{\min}$, as shown in Figure 4).

For all $q\in\mathcal{B}(t_j)$, $\cos(\alpha_q')\leq\cos(\alpha_q)$; moreover, for all $q\in\mathcal{B}_{\alpha_{\min}}$,

$$\cos(\alpha_q')\leq\cos(\alpha_q)-\sin(\alpha_{\min})(\alpha_q'-\alpha_q)\leq\cos(\alpha_q)-\frac{\delta\sin(\alpha_{\min})^2}{\|q-p\|}\leq\cos(\alpha_q)-\frac{\delta\sin(\alpha_{\min})^2}{\mathrm{diam}(\mathcal{Q})+\delta}.$$

Hence, summing over all $q\in\mathcal{B}(t_j)$, we get:

$$\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q')\leq\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q)-\sum_{q\in\mathcal{B}_{\alpha_{\min}}(t_j)}\frac{\delta\sin(\alpha_{\min})^2}{\mathrm{diam}(\mathcal{Q})+\delta} \qquad (17)$$

Observe now that in any case

$$\left|\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q)\right|\leq 1,$$

(it is zero in case $p^*(t_j)\notin\mathcal{B}(t_j)$, and bounded in absolute value by one if $p^*(t_j)\in\mathcal{B}(t_j)$). Therefore, rearranging equation (17):

$$\mathrm{card}(\mathcal{B}_{\alpha_{\min}}(t_j))\frac{\delta\sin(\alpha_{\min})^2}{\mathrm{diam}(\mathcal{Q})+\delta}\leq\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q)-\sum_{q\in\mathcal{B}(t_j)}\cos(\alpha_q')\leq 2$$

Solving this inequality with respect to δ we get:

$$\delta\leq\frac{2\,\mathrm{diam}(\mathcal{Q})}{\mathrm{card}(\mathcal{B}_{\alpha_{\min}}(t_j))\sin(\alpha_{\min})^2-2}. \qquad (18)$$

Since (i) $\alpha_{\min}$ is a positive constant, (ii) $p^*(t_j)$ is in the interior of $\mathcal{Q}$, and (iii) $\lim_{j\to\infty} \text{card}(\mathcal{B}(t_j)) = +\infty$, the right hand side of (18) converges to zero with probability one. Since the bound holds for all points $p \in C(t_j)$, for all $j \in \mathbb{N}$, the claim follows. ∎

In the previous proposition, we have proven that $\|p^*(t_{j+1}) - p^*(t_j)\|$ tends to zero a.s. as $j \to \infty$, under some natural assumptions on the distribution φ and its support. Unfortunately this is not sufficient to prove that the sequence $\{p^*(t_j)\}_{t_j \in \mathbb{N}}$ is Cauchy; convergence of the sequence is however ensured by the following

Proposition 10: The sequence $\{p^*(t_j)\}_{j \in \mathbb{N}}$ converges almost surely.

Proof: Since the sequence $\{p^*(t_j)\}_{j \in \mathbb{N}}$ takes value in a compact set (the closure of $\text{Supp}(\varphi)$), by the Bolzano-Weirstrass theorem there exists a subsequence converging to a limit point $\hat{p}^*$ in the compact set. Construct from $\{p^*(t_j)\}_{j \in \mathbb{N}}$ a maximal subsequence converging to $\hat{p}^*$, and call J the set of indices of this maximal subsequence. If the original sequence $\{p^*(t_j)\}_{j \in \mathbb{N}}$ is not converging to $\hat{p}^*$, then there exists an $L > 0$ such that $\|p^*(t_j) - \hat{p}^*\| \geq L$, for any $j \in \mathbb{N} \setminus J$, and this set of indices is unbounded. Take $\epsilon := L/3 > 0$. We have that $\|p^*(t_{k-1}) - \hat{p}^*\| \leq \epsilon$ for any sufficiently large $(k-1) \in J$; moreover, $\|p^*(t_{k-1}) - p^*(t_k)\| \leq \epsilon$, a.s. by Proposition (9). Choose a sufficiently large $(k-1) \in J$, such that $k \in \mathbb{N} - J$ (this is always possible since the complementary set of J is unbounded by the assumption of non-convergence). But now, we have

$$L \leq \|p^*(t_k) - \hat{p}^*\| \leq \|p^*(t_k) - p^*(t_{k-1})\| + \|p^*(t_{k-1}) - \hat{p}^*\| \leq \frac{2}{3}L,$$

which is a contradiction. ∎

C. Convergence to the generalized median

By the discussion in the previous section, we know that the reference points of all agents that visit an unbounded number of targets converge to a well-defined limit. So do, trivially, the reference points of all agents that visit a bounded set of targets. Hence, we know that the sequence of reference points $p_i^*(t_j)$ converges to a limit $\hat{p}_i^*$, almost surely for all $i \in \{1, \ldots, m\}$. Let us denote by $\mathcal{V}_i(p^*(t_j))$ the Voronoi region associated to the generator $p_i^*(t_j)$, and by $\mathcal{V}_i(\hat{p}^*)$ the Voronoi region corresponding to the limit point $\hat{p}_i^*$.

Proposition 11: If the limit reference points $\hat{p}^* = (\hat{p}_1^*, \ldots, \hat{p}_2^*)$ are distinct, then the sequence of Voronoi partitions $\{\mathcal{V}(p^*(t_j))\}_{j \in \mathbb{N}}$ converges to the Voronoi partition generated by the limit of reference points, i.e.,

$$\lim_{j \to \infty} \mathcal{V}_i(p^*(t_j)) = \mathcal{V}_i(\hat{p}^*), \qquad \text{a.s.}$$

Proof: The boundaries of regions in a Voronoi partition are algebraic curves that depend continuously on the generators, as long as these are distinct. Hence, under this assumption, almost sure convergence of the generators implies the almost sure convergence of the Voronoi regions. ∎

As a next step, we wish to understand what is the relation between the asymptotic reference positions and their associated Voronoi regions. More precisely, let $\mathcal{A} \subset \{1, \ldots, m\}$ be the subset of indices of agents that visit an unbounded number of targets; we want to prove that $\hat{p}_i^*$ is indeed the generalized median $\overline{p}_i$ associated to agent i, with respect to the limiting set $\mathcal{V}_j(\hat{p}^*)$ and distribution $\varphi(x)$, $\forall i \in \mathcal{A}$. First we need the following technical result.

Lemma 12: Let $\{f_i\}_{i \in \mathbb{N}} : \mathcal{Q} \to \mathbb{R}$ be a sequence of strictly convex continuous functions, defined on a common compact subset $\mathcal{Q} \subset \mathbb{R}^n$. Assume that each f_i has a unique $x_i := \arg\min_x f_i$ belonging to the interior of Ω for any i and that this sequence of function converges uniformly to a continuous strictly convex function f admitting a unique minimum point $\overline{x}$ belonging also to the interior of Ω. Then $\lim_{i \to \infty} x_i = \overline{x}$.

Proof: Since $\{f_i\}_{i \in \mathbb{N}}$ converges uniformly to f, then for any $\epsilon > 0$, there exists an $I(\epsilon)$ such that for any $i \geq I(\epsilon)$, $\|f_i - f\| \leq \epsilon$ uniformly in $x \in \Omega$. Let $m := f(\overline{x})$, the minimum value achieved by f. Consider the set

$$U_\epsilon := \{x \in \Omega \text{ such that } f(x) \leq m + 2\epsilon\}.$$

Since f is strictly convex, for ϵ sufficiently small, U_ϵ will be a compact subset contained in Ω. Moreover, since f is strictly convex, we have that U_ϵ is strictly included in $U_{\epsilon'}$, whenever $\epsilon' > \epsilon$ and both are sufficiently small. It is also clear that $\lim_{\epsilon \to \infty} U_\epsilon = \overline{x}$ (nested strictly decreasing sequence of compact subsets all containing the point $\overline{x}$). If $\|f_i - f\| \leq \epsilon$, we claim that $x_i \in U_\epsilon$; we prove this by contradiction. Since $f_i \geq f - \epsilon$, if x_i does not belong to U_ϵ, then $\min(f_i) = f_i(x_i) > m + \epsilon$ (this is just because U_ϵ is simply the set where the function $(f - \epsilon) \leq m + \epsilon$). But since $f_i \leq f + \epsilon$ it turns out that $f_i(\overline{x}) \leq m + \epsilon < \min(f_i)$, which is a contradiction. ∎

We conclude this section with the following:

Proposition 13: Assume that all agents in $\mathcal{A}$ are able to return infinitely often to their reference point between visiting two targets. Then, the limit reference points of such agents coincide, almost surely, with the generalized medians of their limit Voronoi regions, i.e.,

$$\hat{p}_i^* = \arg\min_{p \in \Omega} \int_{\mathcal{V}_i(\hat{p}^*)} \|p - q\| \varphi(q) dq, \qquad \text{a.s.}, \qquad \forall i \in \mathcal{A}.$$

Proof: For any $i \in \mathcal{A}$, define the functions $f_{i,t_j}(p) := \frac{1}{j} \sum_{q \in \mathcal{B}_i(t_j)} \|p - q\|$ and $f_i(p) := \int_{\mathcal{V}_i(\hat{p}^*)} \|p - q\| \varphi(q) \, dq$. These functions are continuous and well defined over Ω. We restrict their domains of definition to the compact set $\mathcal{Q} = \text{Supp}(\varphi)$. These functions are also strictly convex and have unique minima in the interior of $\mathcal{Q}$. Let us notice that, with our previous notations, we have that $p_i^*(t_j) = \arg\min f_{i,t_j}(p)$ and $\overline{p}_i = \arg\min f_i(p)$. Observe that the functions $f_{i,t_j}(p)$ and $f_i(p)$ can be considered random variables with respect to a probability space whose space of events coincide with all possible realizations of target sequences. Consider a restriction of these random variables to a new probability space whose space of events coincide

with all possible realizations of target sequences, for which the corresponding FT points converge to a limiting point. On this new probability space the random variable $f_i(p)$ becomes a deterministic function which is the expected value of the random variables $f_{i,t_j}(p)$. Since $\mathcal{Q}$ is compact, it is immediate to see that $f_{i,t_j}(p)$ has finite expectation and variance over this reduced probability space, and by the Strong Law of Large Numbers we can conclude that almost surely (over this reduced probability space) $f_{i,t_j}(p)$ converge pointwise to $f_i(p)$. To show that $f_{i,t_j}(p)$ converges pointwise to $f_i(p)$ over the original probability space, it is sufficient to observe the following. The original probability space is the probability space whose space of events coincide with all possible realizations of target sequences. We already know that almost surely the FT points associated to any possible realization of target sequences will converge. So we can fiber the space of events of the first probability space into spaces of events of reduced probability spaces, except for a set of measure zero. This is sufficient to prove that $f_{i,t_j}(p)$ converge pointwise to $f_i(p)$ almost surely with respect to all possible realizations of target sequences.

Now that we have proved that almost surely the sequence $\{f_{i,t_j}(p)\}_{t_j \in \mathbb{R}}$ converges pointwise to $f_i(p)$, we prove that it does converge uniformly. To do this, we use a theorem from [28], which is usually devoted to Dini-Arzelà. The theorem states the following: an equicontinuous sequence of functions converges uniformly to a continuous function on a compact set $\mathcal{Q}$ if and only if it converges pointwise to a continuous function on the same compact set. Our sequence $f_{i,t_j}(p)$ is equicontinuous if $\forall \epsilon > 0$ and $\forall p \in \mathcal{Q}$ there exists a $\delta > 0$ such that for all $j \in \{1, \ldots, n\}$ and for all $p' \in \mathcal{Q}$ with $\|p' - p\| \leq \delta$, we have $\|f_{i,t_j}(p) - f_{i,t_j}(p')\| \leq \epsilon$; observe that δ is independent on j, while in general it will depend on ϵ and on p. Now we have

$$\|f_{i,t_j}(p) - f_{i,t_j}(p')\| \leq \frac{1}{j} \sum_{q \in \mathcal{B}_i(t_j)} \left| \|p - q\| - \|p' - q\| \right|.$$

Using

$$\|q - p\| = \|q - p' + p' - p\| \leq \|q - p'\| + \|p - p'\|,$$

it is immediate to see that

$$\|f_{i,t_j}(p) - f_{i,t_j}(p')\| \leq \frac{1}{j} \sum_{q \in \mathcal{B}_i(t_j)} \|p - p'\| \leq \|p - p'\|.$$

So it is sufficient to take $\delta = \epsilon$ in the previous definition and δ does not depend on j. So the sequence is equicontinuous and the pointwise convergence is upgraded to uniform convergence.

We already know that almost surely the points $p_i^*(t_j)$ do converge to points $\hat{p}_i^*$ (Proposition (10)); therefore we can claim that $\hat{p}_i^* = \overline{p}_i$, simply applying Lemma (12), which requires the uniform convergence. Thus we can claim that the reference position of each agent which services infinitely many targets following our algorithm converges to the generalized median of its Voronoi region, almost surely. ∎

D. Fairness

In this section, we prove that, as long as φ is strictly positive over a convex set, both policies introduced in Section III are fair.

Proposition 14 (Fairness): If $\mathcal{Q} = \mathrm{Supp}(\varphi)$ is convex, all agents eventually visit an unbounded number of targets, almost surely, i.e.,

$$\lim_{t \to +\infty} \mathrm{card}(\mathcal{B}_i(t)) = +\infty, \qquad \text{a.s.}, \qquad \forall i \in \{1, \ldots, m\}.$$

Proof: Under either policy, each agent will pursue and is guaranteed to service the nearest target in its own Voronoi region. Hence, in order to show that an agent services an unbounded number of targets, it is sufficient to show that the probability that the next target be generated within its Voronoi region remains strictly positive, i.e., that

$$\int_{\mathcal{V}_i(p(t))} \varphi(q)dq > 0.$$

Since $\mathcal{Q}$ is convex, and all agents move towards the nearest target, at least initially, all agents will eventually enter $\mathcal{Q}$, and remain within it. Let us denote with P_{ij} the probability that agent i visits the j-th target. For any i this probability is always strictly positive. Indeed, even if it happens that some agents are servicing simultaneously the same target (simply because the service request appears at the boundary of two different Voronoi regions), this does not mean that their reference points have to coincide. For the reference points of two agents eventually to coincide, it must happen that they are servicing infinitely many often and simultaneously the same target. Since the boundaries of Voronoi regions have measure zero and φ is a continuous distribution without singular components, we can claim that $\lim_{j \to +\infty} P_{ij} > 0$ almost surely. Now call $P = \min_{i=1,\ldots,m} \lim_{j \to +\infty} P_{ij}$. Then P is strictly positive almost surely. Therefore, the probability that the i-th agent *does not* visit an unbounded number of targets is bounded from above by $\lim_{j \to +\infty} \prod_{k=1}^{j}(1 - P)^k = 0$, almost surely. ∎

E. Efficiency

In this section, we will prove that the system time provided by both the algorithms in Section III converges to a critical point (either a saddle point or a local minimum) with high probability as $\lambda \to 0^+$.

In the preceding sections, we have proved that—as long as each agent is able to return to its reference point between servicing two targets, infinitely often—the reference points $p_i^*(t_j)$ converge to points $\hat{p}_i^*$, which generate a MVT. In such case, we know that the average time of service will converge to

$$\overline{T}_\pi = \int_\Omega \min_{i=1,\ldots,m} \|\hat{p}_i^* - q\| \; \varphi(q)dq = \sum_{i=1}^{m} \int_{\mathcal{V}_i(\hat{p}^*)} \|\hat{p}_i^* - q\| \; \varphi(q)dq. \tag{19}$$

Consider now functions $\mathcal{H}_m$ of the form:

$$\mathcal{H}_m(p_1, \ldots, p_m) = \int_\Omega \min_{i=1,\ldots,m} \|p_i - q\| \ \varphi(q)dq = \sum_{i=1}^m \int_{\mathcal{V}_i(p)} \|p_i - q\| \ \varphi(q)dq. \qquad (20)$$

Observe that $\overline{T}_\pi$ belongs to the class of functions of the form $\mathcal{H}_m$ where each point p_i is constrained to be the generalized median of the corresponding Voronoi region (i.e., $\mathcal{V}(p)$ is a MVT).

We want to prove that $\overline{T}_\pi$ is a critical point of $\mathcal{H}_m$. To do so, we consider an extension of $\mathcal{H}_m$, i.e. a functional $\mathcal{K}_m$ defined as follows:

$$\mathcal{K}_m(p_1, \ldots, p_m, \mathcal{V}_1, \ldots, \mathcal{V}_m) := \sum_{i=1}^m \int_{y \in \mathcal{V}_i} \|y - p_i\| \ \varphi(y)dy.$$

Observe that in this case the regions $\{\mathcal{V}_i\}_{i=1,\ldots,m}$ are not restricted to form a MVT with respect to the generators $\{p_i\}_{i=1,\ldots,m}$. Thus we can view the functional $\mathcal{H}_m$ we are interested in as a constrained form of the unconstrained functional $\mathcal{K}_m$. It turns out therefore that critical points of $\mathcal{K}_m$ are also critical points of $\mathcal{H}_m$. With respect to critical points of $\mathcal{K}_m$ we have the following result:

Proposition 15: Let $\{p_i\}_{i=1,\ldots,m}$ denote *any set* of m points belonging to $\mathrm{Supp}(\varphi)$ and let $\{\mathcal{V}_i\}_{i=1,\ldots,m}$ denote *any tessellation* of $\mathrm{Supp}(\varphi)$ into m regions. Moreover, let us define $\mathcal{K}_m$ as above. Then a sufficient condition for $\{p_1, \ldots, p_m, \mathcal{V}_1, \ldots, \mathcal{V}_m\}$ to be a critical point (either a saddle point or a local minimum), is that the $\mathcal{V}_i$'s are the Voronoi regions corresponding to the p_i's, and, *simultaneously*, the p_i's are the generalized median of the corresponding $\mathcal{V}_i$'s.

Proof: Consider first the variation of $\mathcal{K}_m$ with respect to a single point, say p_i. Now let v be a vector in $\mathbb{R}^2$, such that $p_i + \epsilon v \in \Omega$. Then we have

$$\mathcal{K}_m(p_i + \epsilon v) - \mathcal{K}_m(p_i) = \int_{y \in \mathcal{V}_i} \{\|y - p_i - \epsilon v\| - \|y - p_i\|\} \ \varphi(y)dy,$$

where we have not listed the other variables on which $\mathcal{K}_m$ depends since they remain constant in this variation. By the very form of this variation, it is clear that if the point p_i is the generalized median for the *fixed* region $\mathcal{V}_i$, we will have that $\mathcal{K}_m(p_i + \epsilon v) - \mathcal{K}_m(p_i) > 0$, for any v. Now consider the points $\{p_i\}_{i=1,\ldots,m}$ fixed and consider a tessellation $\{\mathcal{U}_i\}_{i=1,\ldots,m}$ different from the Voronoi regions $\{\mathcal{V}_i\}_{i=1,\ldots,m}$ generated by the points p_i's. We compare the value of $\mathcal{K}_m(p_1, \ldots, p_m, \mathcal{V}_1, \ldots, \mathcal{V}_m)$, with the value of $\mathcal{K}_m(p_1, \ldots, p_m, \mathcal{U}_1, \ldots, \mathcal{U}_m)$. Consider those y which belong to the Voronoi region $\mathcal{V}_j$ generated by p_j, and possibly not to the Voronoi region of another p_i. Anyway, since $\mathcal{U}_i$ is not a Voronoi tessellation, it can happen that in any case these y belong to $\mathcal{U}_i$. Thus for these particular y's we have $\varphi(y)\|y - p_j\| \leq \varphi(y)\|y - p_i\|$. Moreover, since $\{\mathcal{U}_i\}_{i=1,\ldots,m}$ are not the Voronoi tessellation associated to the p_i's, the last inequality must be strict over some set of positive measure. Thus we have that $\mathcal{K}_m(p_1, \ldots, p_m, \mathcal{V}_1, \ldots, \mathcal{V}_m) < \mathcal{K}_m(p_1, \ldots, p_m, \mathcal{U}_1, \ldots, \mathcal{U}_m)$, and therefore $\mathcal{K}_m$ is minimized, keeping fixed the p_i's exactly when the subset $\mathcal{V}_i$'s are chosen to be the Voronoi regions associated with the point p_i's. ∎

By the previous proposition and by the fact that critical points of the unconstrained functional $\mathcal{K}_m$ are also critical points of the constrained functional $\mathcal{H}_m$, we have that the MVT are always critical points for the functional $\mathcal{H}_m$, and in particular $\overline{T}_\pi$ is either a saddle point or a local minimum for the functional $\mathcal{H}_m$.

We can now conclude with following:

Theorem 16 (Efficiency): The system time provided by the no-communication policy π_{nc} and by the sensor-based policy π_{sb} converges to a critical point (either a saddle point or a local minimum) with high probability as $\lambda \to 0^+$.

Proof: Combining results in Propositions 10 and 13 we conclude that the reference points of all agents that visit an unbounded number of targets converge to a MVT, almost surely—provided agents can return to the reference point between visiting targets. Moreover, the fairness result in Proposition 14 shows that in fact all agents do visit an unbounded number of targets almost surely; as a consequence of Proposition 15 the limit configuration is indeed a critical point for the system time. Since agents return infinitely often to their reference positions with high probability as $\lambda \to 0$, the claim is proven. ∎

Thus we have proved that the suggested algorithm enables the agents to realize a coordinated task, such that "minimizing" the cost function without explicit communication, or with mutual position knowledge only. Let us underline that, in general, the achieved critical point strictly depends on the initial positions of the agents inside the environment Ω. It is known that the function $\mathcal{H}_m$ admits (not unique, in general) *global minima*, but the problem to find them is *NP-hard*.

Remark 17: We can not exclude that the algorithm so designed will converge indeed to a saddle point instead of a local minimum. This is due to the fact that the algorithm provides a sort of implementation of the *steepest descent* method, where, unfortunately we are not following the steepest direction of the gradient of the function $\mathcal{H}_m$, but just the gradient with respect to one of the variables. For a broader point of view of steepest descent in this framework see for instance [29].

On the other hand, since the algorithm is based on a sequence of targets and at each phase we are trying to minimize a different cost function, it can be proved that the critical points reached by this algorithm are *no worse* than the critical points reached knowing a priori the distribution φ. This is a remarkable result proved in a different context in [30], where an example is presented in which the use of a sample sequence provides a better result (with probability one) than the a priori knowledge of φ. In that specific example the algorithm with the sample sequence does converge to a global minimum, while the algorithm based on the a priori knowledge of the distribution φ gets stuck in a saddle point.

F. A comparison with algorithms for vector quantization and centroidal Voronoi tessellations

The use of Voronoi tessellations is ubiquitous in many fields of science, ranging from operations research, animal ethology (territorial behaviour of animals), computer science (design of algorithms), to numerical analysis (construction of adaptive grids for PDEs and general quadrature rules), and algebraic geometry (moduli spaces of abelian varieties). For a detailed account of

possible applications see for instance the book [29]. In the available literature, most of the analysis is devoted to applications of centroidal Voronoi tessellations, i.e., Voronoi tessellations such that

$$p_i = \arg\min_{s \in \mathbb{R}^n} \int_{\mathcal{V}_i(p_1, p_2, \ldots, p_m)} \|s - q\|^2 \varphi(q) \, dq, \qquad \forall i \in \{1, \ldots, m\}.$$

A popular algorithm due to Lloyd [20] is based on the iterative computation of centroidal Voronoi tessellations. The algorithm can be summarized as follows. Pick m generator points, and consider a large number n of samples from a certain distribution. At each step of the algorithm generators are moved towards the centroid of the samples inside their respective Voronoi region. The algorithm terminates when each generator is within a given tolerance from the centroid of samples in its region, thus obtaining a centroidal Voronoi tessellation weighted by the sample distribution. There is also a continuous version of the algorithm, which requires the a priori knowledge of a spatial density function, and computation of the gradient of the polar moments of the Voronoi regions with respect to the positions of the generators. An application to coverage problems in robotics and sensor networks of Lloyd's algorithm is available in [14].

The algorithms we introduced in this paper are more closely related to an algorithm due to MacQueen [21], originally designed as a simple online adaptive algorithm to solve clustering problems, and later used as the method of choice in several vector quantization problems where little information about the underlying geometric structure is available. MacQueen's algorithm can be summarized as follows. Pick m generator points. Then iteratively sample points according to the probability density function φ. Assign the sampled point to the nearest generator, and update the latter by moving it in the direction of the sample. In other words, let q_j be te j-th sample, let $i^*(j)$ be the index of the nearest generator, and let $c = (c_1, \ldots, c_m)$ be a counter vector, initialized to a vector of ones. The update rule takes the form

$$p_{i^*(j)} \leftarrow \frac{c_{i^*(j)} q_j + p_{i^*(j)}}{c_{i^*(j)} + 1},$$

$$c_{i^*(j)} \leftarrow c_{i^*(j)} + 1.$$

The process is iterated until some termination criterion is reached. Compared to Lloyd's algorithm, MacQueen's algorithm has the advantage to be a learning adaptive algorithm, not requiring the a priori knowledge of the distribution of the objects, but rather allowing the online generation of samples. It can be recognized that the update rule in MacQueen's algorithm corresponds to moving the generator points to the centroids of the samples assigned to them.

The algorithm we propose is very similar in spirit to MacQueen's algorithm, however, there is a remarkable difference. MacQueen's algorithm deals with centroidal Voronoi tessellations, thus with the computation of k-means. Our algorithm instead is based on MVT, and on the computation of k-medians. In general, very little is known for Voronoi diagrams generated using simply the Euclidean distance instead of the square of the Euclidean distance. For instance, the medians of a sequence of

points can exhibit a quite peculiar behavior if compared to the one of the means. Consider the following example. Given a sequence of points $\{q_i\}_{i\in\mathbb{N}}$ in a compact set $K \subset \mathbb{R}^2$, we can construct the induced sequence of means:

$$m_N := \frac{1}{N}\sum_{i=1}^{N}q_j$$

and analogously the induce sequence of FT points we considered in the previous sections. Call FT_N the FT point corresponding to the first N points of the sequence $\{q_i\}_{i\in\mathbb{N}}$. We want to point out that induced sequence $\{m_j\}$ and $\{FT_j\}$ have a very different behaviour. Indeed, the induced sequence of means will always converge as long as the points q_js belong to a compact set. To see this, just observe that if $\mathrm{diam}(K) \leq L$, then $\|m_j\| \leq L$. Moreover, it is immediate to see that $\|m_{N+1}-m_N\| \leq \frac{2L}{N}$. Then one can conclude using the same argument of Proposition 10. On the other hand, one can construct a special sequence of points q_js in a compact set K for which the induced sequence of FT points does not converge. This is essentially due to the fact that while the contribution of each single point q_j in moving the position of the mean decreases as j increases, this could not happen in the case of the median. To give a simple example, start with the following configuration of points in $\mathbb{R}^2$: $q_1 = (1,0), q_2 = (-1,0), q_3 = (0,1)$ and $q_4 = (0,-1)$. Then the sequence of points q_js continue in the following way: $q_k = (0,1)$ if $k > 4$, and k odd, $q_k = (0,-1)$ if $k > 4$ and k is even. Using the characterization of FT points, it is immediate to see that $FT_k = (0,0)$ for $k > 4$ and k even, while $FT_k = (0,\tan(\pi/6))$ for $k > 4$ and k odd, so the induced sequence can not converge. This phenomenon can not happen to the sequence of means, which is instead always convergent. Therefore, it should be clear that the use of MVT instead of centroidal Voronoi tessellations makes much more difficult to deal with the technical aspects of the algorithm such as its convergence.

V. A GAME-THEORETIC POINT OF VIEW

In previous sections, we proposed two policies, π_{sb} and π_{nc}, which yield the $\hat{p}^*$ reference points under light load. A key component of the two policies was the update rule for the reference points. In this section we provide a game-theoretic interpretation of the $\hat{p}^*$ reference points. In particular, we frame our presentation along the works [8], [31], in which game-theoretic point of view has been introduced in the study of cooperative control and strategic coordination of decentralized networks of multi-agents systems. We prove that the $\hat{p}^*$ reference points, under light load conditions, are an efficient pure Nash equilibria in a multi-player game setting where each agent is interested in maximizing its own utility. It turns out that, by just trying to maximize their own utility function, the agents will indeed maximize a *common global utility function*. The update rule for the reference point proposed as part of the two policies in the earlier sections can then be regarded as a mechanism for the vehicles to *learn* the $\hat{p}^*$ strategy.

We first formulate the game in the strategic form [32] as follows. Consider a scenario with the same workspace Ω, density

function φ with support $\mathcal{Q}$ and the same stochastic process for generating service requests as described in the previous sections. We replace the terms *service requests* and *targets* with *resources* to fit better in the context of this section. The players of the game are the m vehicles. The vehicles are assumed to be *rational autonomous decision makers* trying to maximize their own utility function. The resources offer rewards in a continuous fashion and the vehicles can collect these rewards by traveling to the resource locations. Every resource offers reward at a rate, which depends on the number of vehicles present at its location: the reward rate is unity when there is one vehicle and it is zero when there are more than one vehicles. Moreover, the life of the resource ends as soon as more than one vehicles are present at its location. The sensing and communication capabilities of the vehicles are as follows: (i) the location of a resource is broadcast to all the vehicles as soon as it is generated; similarly, every vehicle is notified as soon as a resource ceases to exist, and (ii) there is no explicit communication between the vehicles, i.e., the vehicles do not have knowledge about each other's position and no vehicle knows the identities of other vehicles visiting any resource location.

This setup can be understood to be an extreme form of congestion game [33], where the resource cannot be shared between agents and that the resource is cut off at the first attempt to share it. The total reward for vehicle i from a particular resource is the time difference between its arrival and the arrival of the next vehicle, if i is the first vehicle to reach the location of the resource, and zero otherwise. (Note that, since a vehicle cannot communicate with any other vehicle, it cannot determine if it will be the first one to reach the resource location when the location is broadcast to it).

We focus our analysis on the light load case here, i.e., when $\lambda \to 0^+$. Moreover, we assume that all the vehicles are present in $\mathcal{Q}$ at time $t = 0$ and that there are no resources at time $t = 0$. Hence, there will be utmost one active resource at any time almost surely. Note that these assumptions on the initial conditions are without any loss of generality in light of the discussion in the earlier sections. Since our focus is on the light load scenario, we let the strategy space of agent i to be $\mathcal{Q}$ for all $i \in \{1, \ldots, m\}$. Specifically, the strategy of an agent i denoted as π_i is identified with a *reference point* in $\mathcal{Q}$ which the agent approaches in the absence of outstanding service requests. On the generation of a resource, the agents move directly towards its location and return back to the reference location once the resource ceases to be active. We will use $\pi_{-i} \in \mathcal{Q}^{m-1}$ to denote the strategy specification of all the agents, except agent i, i.e., $\pi_{-i} := (\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_m)$. Hence, we may write strategy vector π as (π_i, π_{-i}). Let $\mathcal{U}_i : \mathcal{Q}^m \to \mathbb{R}$ be the utility function of vehicle i. For a given strategy vector π, let $r_i(q, \pi)$ be the reward collected by agent i for resource generated at location $q \in \mathcal{Q}$. In light of the discussion above, the utility function of an agent i is defined as

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \mathrm{E}_q[r_i(q, \pi)]. \tag{21}$$

Equation (21) implies that the goal of every vehicle is to maximize the expected value of the reward from the next resource.

With this, we complete the formal description of the game at hand. For brevity in notation, we use $\mathcal{G}$ to denote this game. We now derive a working expression for the utility function.

As mentioned before, the reward for agent i is the time till the arrival of the second agent at point q if agent i is the first to reach q and zero otherwise. Since the vehicles move at unit speed, the reward for an agent i can be written as $r_i(q, \pi) = \max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\}$. The utility function for agent i can then be written as

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \mathrm{E}_q[\max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\}] = \int_{\mathcal{Q}} \max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\}\varphi(q)dq. \tag{22}$$

However, we know that

$$\max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\} = \begin{cases} \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|, & \text{if } q \in \mathcal{V}_i(\pi) \\ 0, & \text{otherwise .} \end{cases}$$

Substituting this into Equation (22), we derive the following expression for the utility function.

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \int_{\mathcal{V}_i(\pi)} (\min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|)\varphi(q)dq. \tag{23}$$

We now prove that $\mathcal{G}$ belongs to a class of multi-player games called potential games. In a potential game, the difference in the value of the utility function for any agent for two different strategies, when the strategies of the other agents are kept fixed, is equal to the difference in the values of a potential function that depends only on the strategy vector and not on the label of any agent. The formal definition [34] is as follows.

Definition 18: A finite m-player game with strategy spaces $\{\Pi_i\}_{i=1}^m$ and utility functions $\{\mathcal{U}_i\}_{i=1}^m$ is a *potential game* if, for some potential function $\psi : \times_{i \in \{1, \ldots, m\}} \Pi_i \to \mathbb{R}$,

$$\mathcal{U}_i(\pi_i', \pi_{-i}) - \mathcal{U}_i(\pi_i'', \pi_{-i}) = \psi(\pi_i', \pi_{-i}) - \psi(\pi_i'', \pi_{-i}),$$

for every player $i \in \{1, \ldots, m\}$, for every $\pi_{-i} \in \times_{j \neq i} \Pi$ and for every $\pi_i', \pi_i'' \in \Pi_i$.

Proposition 19: $\mathcal{G}$ is a potential game.

Proof: The expression for the utility function of agent i, as given by Equation (23), can be rewritten as:

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \int_{\mathcal{V}_i(\pi)} \min_{j\neq i} \|\pi_j - q\| \varphi(q) dq - \int_{\mathcal{V}_i(\pi)} \|\pi_i - q\| \varphi(q) dq$$

$$+ \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq - \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq$$

$$= \int_{\mathcal{V}_i(\pi)} \min_{j\neq i} \|\pi_j - q\| \varphi(q) dq + \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq$$

$$= \int_{\mathcal{V}_i(\pi)} \min_{j\neq i} \|\pi_j - q\| \varphi(q) dq + \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi)} \min_{k\neq i} \|\pi_k - q\| \varphi(q) dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq$$

$$= \int_{\mathcal{Q}} \min_{j\neq i} \|\pi_j - q\| \varphi(q) dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq. \tag{24}$$

In the integrand of the first term in Equation (24), $\min_{j\neq i} \|\pi_j - q\|$ is the distance from point q to the closest among all the agents, except the ith agent. We then consider the Voronoi partition with generators $\pi_{-i} = \pi \setminus \pi_i$, and let $\mathcal{V}_j(\pi_{-i})$ be the corresponding Voronoi cell belonging to agent j. Equation (24) can then be written as

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi_{-i})} \|\pi_j - q\| \varphi(q) dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\| \varphi(q) dq. \tag{25}$$

Note that the first term on the right hand side of Equation (25) is independent of π_i. With this observation, consider the following potential function:

$$\psi(\pi) = -\sum_{i=1}^{m} \int_{\mathcal{V}_i(\pi)} \|\pi_i - q\| \varphi(q) dq. \tag{26}$$

The proposition then follows by combining Equations (25) and (26) with the definition of a potential game.

∎

It turns out that an agent's motive to maximize its own utility function is *aligned* with the global objective of minimizing the average system time. To formally establish this fact, we start with a couple of definitions. First, we extend the concept of a potential game.

Definition 20: A finite m-player game with strategy spaces $\{\Pi_i\}_{i=1}^{m}$ and utility functions $\{\mathcal{U}_i\}_{i=1}^{m}$ is an *ordinal potential game* if, for some potential function $\psi : \times_{i\in\{1,\ldots,m\}} \Pi_i \to \mathbb{R}$,

$$\mathcal{U}_i(\pi_i', \pi_{-i}) - \mathcal{U}_i(\pi_i'', \pi_{-i}) > 0 \text{ if and only if } \psi(\pi_i', \pi_{-i}) - \psi(\pi_i'', \pi_{-i}) > 0,$$

for every player $i \in \{1, \ldots, m\}$, for every $\pi_{-i} \in \times_{j\neq i} \Pi$ and for every $\pi_i', \pi_i'' \in \Pi_i$.

Remark 21: Note that every potential game is also an ordinal potential game.

Definition 22: The set of agents utilities $\{\mathcal{U}_i\}_{i=1,\ldots,m}$ is *aligned* with the global utility $\mathcal{U}_g$ if and only if the game with utility functions $\{\mathcal{U}_i\}_{i=1,\ldots,m}$ is an ordinal potential game with $\mathcal{U}_g$ as a potential function.

For the game $\mathcal{G}$, define the global utility function to be the negative of the average system time under policy π, i.e., $\mathcal{U}_g(\pi) \equiv -\overline{T}_\pi$, which can be rewritten as

$$\mathcal{U}_g(\pi) = -\sum_{i=1}^m \int_{\mathcal{V}_i(\pi)} \|\pi_i - q\|\varphi(q)dq. \tag{27}$$

Proposition 23: The utility functions of the agents in $\mathcal{G}$ are aligned with its global utility function.

Proof: Comparing Equation (27) with the expression of the potential function in Equation (26), we observe that $\mathcal{U}_g(\pi) = \psi(\pi)$. Proposition 19 implies that $\mathcal{G}$ is a potential game, and hence an ordinal potential game, with ψ as the potential function. The proposition then follows from Definition 22. ∎

We now show that our utility function belongs to a class of utility functions called Wonderful Life Utility Functions [35].

Definition 24: Given a global utility function $\mathcal{U}_g(\pi_i, \pi_{-i})$, the Wonderful Life (local) utility function is given by

$$\mathcal{U}_i(\pi_i.\pi_{-i}) = \mathcal{U}_g(\pi_i, \pi_{-i}) - \mathcal{U}_g(\pi_{-i}),$$

where $\mathcal{U}_g(\pi_{-i})$ is the global utility in absence of agent i.

Remark 25: The Wonderful Life utility function measures the marginal contribution of an agent towards the global utility.

Proposition 26: The local utility function, as defined in Equation (25) is a Wonderful Life utility function with respect to the global utility function defined in Equation (27).

Proof: We refer to Equation (25) in the proof of Proposition 19, where we derived an alternate expression for the local utility function. Comparing the two terms individually with the expression for the global utility function in Equation (27), we get that

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \mathcal{U}_g(\pi_i, \pi_{-i}) - \mathcal{U}_g(\pi_{-i}).$$

The proposition then follows from the definition of the Wonderful Life utility function. ∎

The π_{nc} policy yields the *equilibrium* strategy $\hat{p}^* = \{\hat{p}_1^*, \ldots, \hat{p}_m^*\}$ such that, for all $i \in \{1, \ldots, m\}$, $\hat{p}_i^*$ is the median of the Voronoi region $\mathcal{V}_i(\hat{p}^*)$. We now state and prove results regarding the efficiency and equilibrium status of the $\hat{p}^*$ strategy in the game theoretic setting of this section. We first state the following definition adapted from [8]:

Definition 27: A strategy $\tilde{\pi}$ is called a *pure Nash equilibrium* if, for all $i = \{1, \ldots, m\}$,

$$\mathcal{U}_i(\tilde{\pi}_i, \tilde{\pi}_{-i}) = \max_{\pi_i \in \Pi_i} \mathcal{U}_i(\pi_i, \tilde{\pi}_{-i}). \tag{28}$$

Moreover, a strategy π is called *efficient* if there is no other strategy that yields higher utilities to all the agents.

Proposition 28: The $\hat{p}^*$ strategy is an efficient pure Nash equilibrium for the game $\mathcal{G}$.

Proof: For any $\pi_{-i} = \{p_{-i}, t_{-i}\}$,

$$\hat{p}_i^* = \operatorname{argmin}_{p_i \in \mathcal{Q}} \int_{\mathcal{V}_i(p)} \|p_i - q\| \varphi(q) dq.$$

Combining this with the definition of an efficient Nash equilibrium, one arrives at the proposition. ■

Remark 29: Note that we do *not* claim in Proposition 28 that the π^* strategy provides the unique efficient pure Nash equilibrium for the game $\mathcal{G}$. For instance to find other efficient pure Nash equilibria it is sufficient to modify the algorithm during the initial phases, when the FT points are not uniquely determined. These modifications produce different policy vectors which are anyway all efficient pure Nash equilibria, as it is immediate to see.

In general, it is true that alignment does not prevent pure Nash equilibria from being suboptimal from the point of view of the global utility. Moreover, even *efficient* pure Nash equilibria (i.e. pure Nash equilibria which yield the highest utility to all agents) can be suboptimal from the perspective of the global utility function. Such a phenomenon is indeed what happens in our construction.

In the earlier sections, we proved that the update rule for the reference point which was part of both the policies converges to $\hat{p}^*$ almost surely under light load conditions. That update rule can then be thought of a *learning* algorithm for the agents to arrive at the efficient Nash equilibrium $\hat{p}^*$, even without explicit knowledge of the history of policies played by the other agents at every iteration.

VI. NUMERICAL RESULTS

In this section, we present simulation results showing the performance of the proposed policies for various scenarios.

A. *Uniform distribution, light load*

In the numerical experiments, we first consider $m = 9$, choose $\mathcal{Q}$ as a unit square, and set $\varphi = 1$ (i.e., we consider a spatially uniform target-generation process). This choice allows us to determine easily the optimal placement of reference points, at the centers of a tessellation of $\mathcal{Q}$ into nine equal squares, and compute analytically the optimal system time. In fact, it is known that the expected distance of a point q randomly sampled from a uniform distribution within a square of side L from the center of the square c is

$$\mathrm{E}[\|q - c\|] = \frac{\sqrt{2} + \log(1 + \sqrt{2})}{6} L \approx 0.3826L.$$

The results for a small value of λ, i.e., $\lambda = 0.5$, are presented in Figure 5. The average service time converges to a value that is very close to the theoretical limit computed above, taking $L = 1/\sqrt{m} = 1/3$. In both cases, the reference points converge—albeit very slowly—to the generators of a MVT, while the average system time quickly approaches the optimal value.
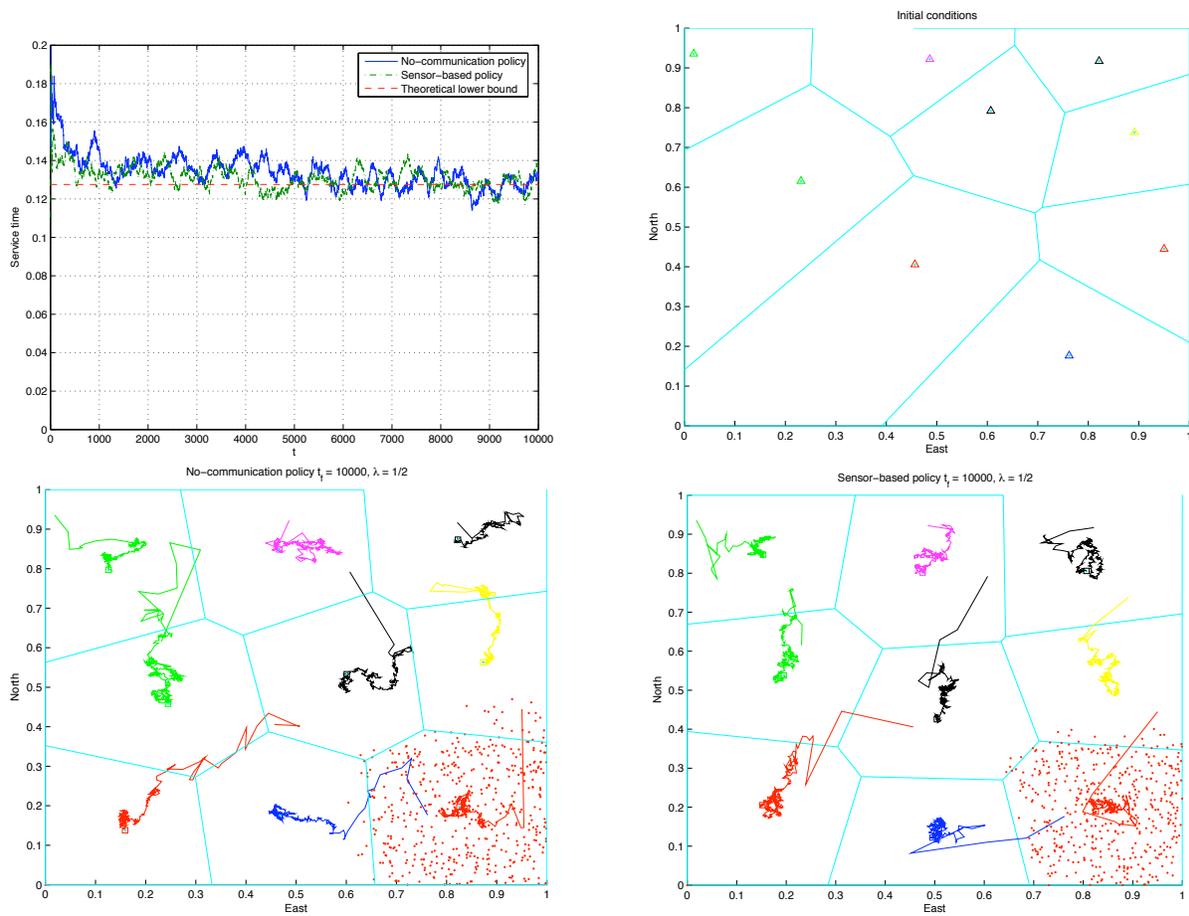
Fig. 5. Numerical simulation in the light-load case, for a uniform spatial distribution. Top left: the actual service times as a function of time, for the two policies, compared with the optimal system time. Top right: the initial configuration of the nine agents. Bottom left and right: paths followed by the reference points up to $t = 10^4$ (corresponding to approximately 5,000 targets), using the two policies. The locations of all targets visited by one of the agents are also shown.

B. Non-uniform distribution, light load

We also present in Figure 6 results of similar numerical experiments with a non-uniform distribution, namely an isotropic normal distribution centered at $(0.25, 0.25)$, with standard deviation equal to 0.25.

C. Uniform distribution, dependency on the target generation rate

An interesting set of numerical experiments evaluates the performance of the proposed policies over a large range of values of the target generation rate λ. In Section IV, we proved the convergence of the system's behavior to an efficient steady state, with high probability as $\lambda \to 0^+$, as confirmed by the simulations discussed above. For large values of λ however, the assumption that vehicles are able to return to their reference point breaks down, and the convergence result is no longer valid. In figure 7 we report results from numerical experiments on scenarios involving $m = 3$ agents, and values of λ ranging from $1/2$ to 32. In the figure, we also report the known (asymptotic) lower bounds on the system time (with 3 agents), as derived
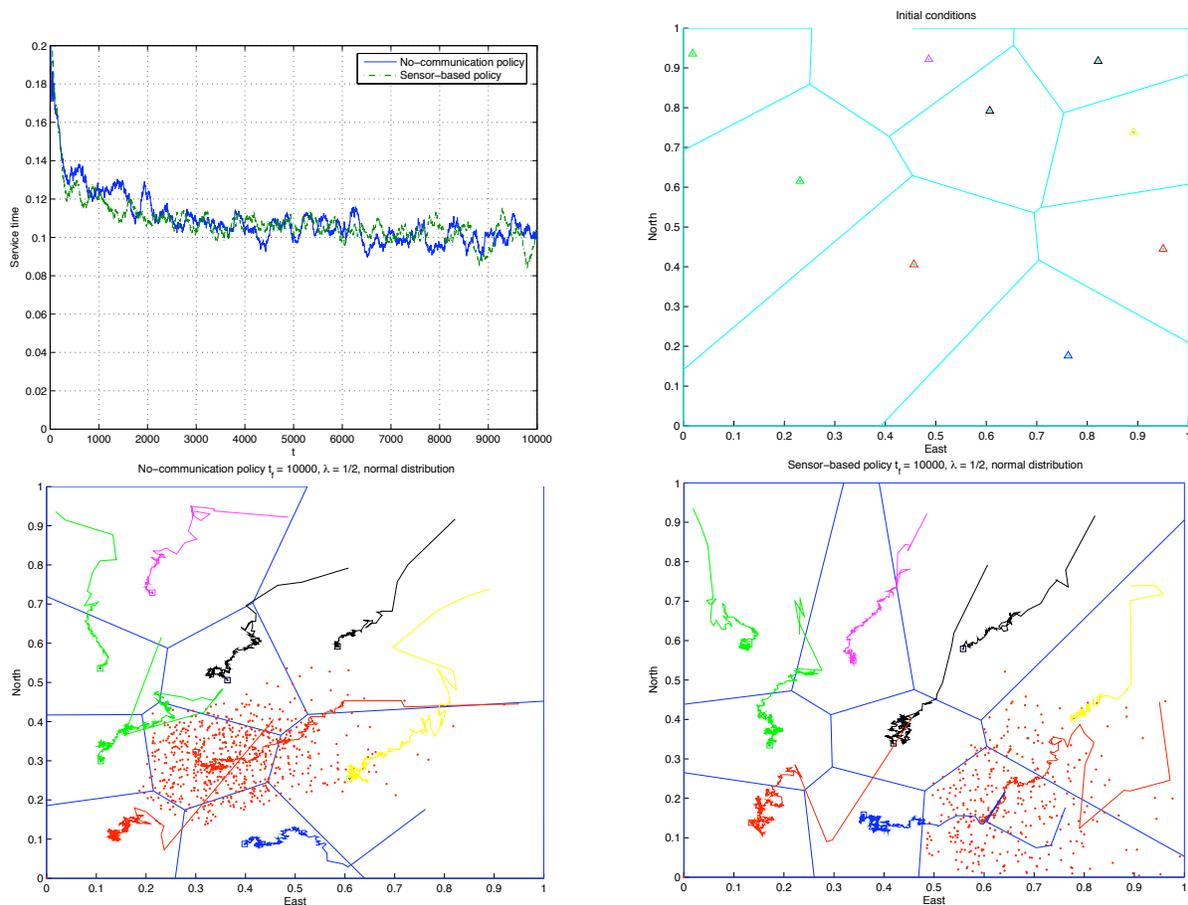
Fig. 6. Numerical simulation in the light-load case, for a normal spatial distribution. Top left: the actual service times as a function of time, for the two policies. Top right: the initial configuration of the nine agents. Bottom left and right: paths followed by the reference points up to $t = 10^4$ (corresponding to approximately 5,000 targets), using the two policies. The locations of all targets visited by one of the agents are also shown.

in [11], and the system time obtained with the proposed policies in a single-agent scenario.

The performance of both proposed policies is close to optimal for small λ, as expected. The sensor-based policy behaves well over a large range of target generation rates; in fact, the numerical results suggest that the policy provides a system time that is a constant-factor approximation of the optimum, by a factor of approximately 1.6.

However, as λ increases, the performance of the no-communication policy degrades significantly, almost approaching the performance of a single-vehicle system over an intermediate range of values of λ. Our intuition in this phenomenon is the following. As agents do not return to their own reference points between visiting successive targets, their efficiency decreases since they are no longer able to effectively separate regions of responsibility. In practice—unless they communicate and concentrate on their Voronoi region, as in the sensor-based policy—agents are likely to duplicate efforts as they pursue the same target, and effectively behave as a single-vehicle system. Interestingly, this efficiency loss seems to decrease for large λ, and the numerical results suggest that the no-communication policy recovers a similar performance as the sensor-based policy
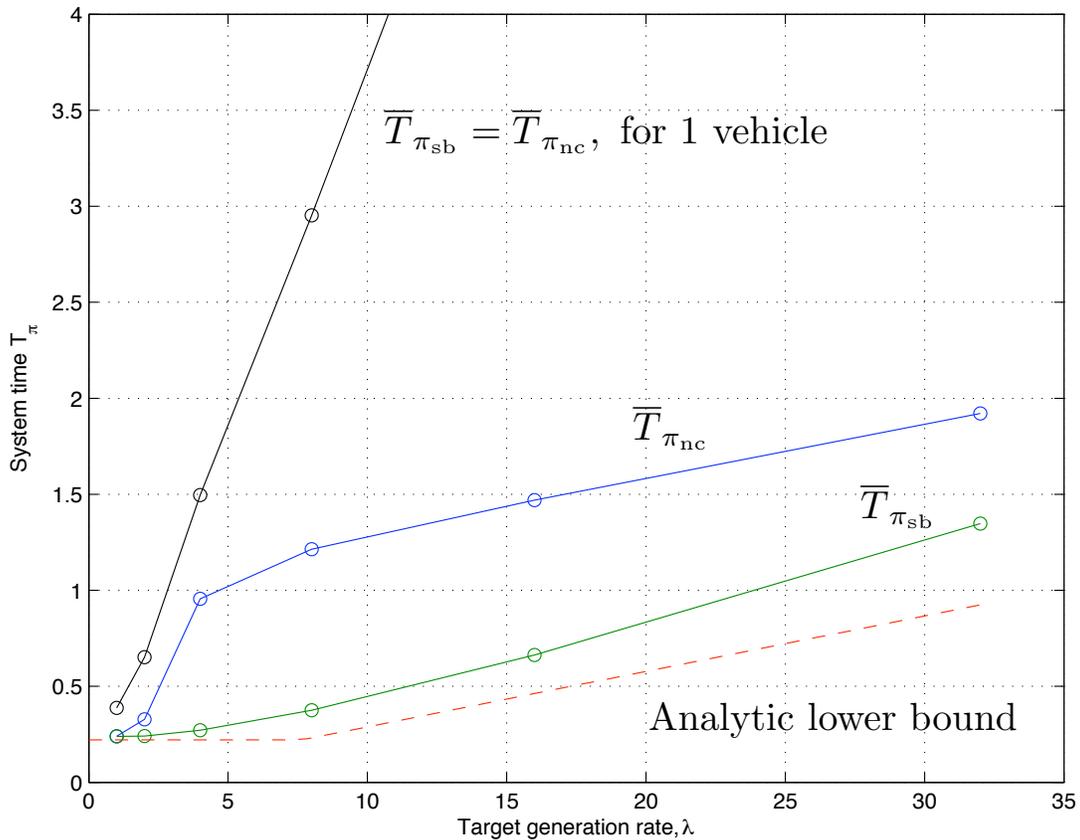
Fig. 7. System time provided by the policies proposed in this paper, as a function of the target generation rate λ. The system is composed of three vehicles, and the target points are generated uniformly in the unit square. Each data point is obtained taking the mean of the results obtained running one hundred times the system dynamics, for each specified policy and for each specified target generation rate λ. Therefore each data point is a Monte Carlo approximation of the theoretical system time corresponding to that policy and to that target generation rate.

in the heavy load limit. Unfortunately, we are not able at this time to provide a rigorous analysis of the proposed policies for general values of the target generation rate.

Finally, let us add two remarks. The first concerns what happens if the agents are slower. After a suitable rescaling, reducing the agent velocities is equivalent to increase the target generation rate λ. As it is evident from Figure 7 even when λ increases both policies are stabilizing, in the sense that the system time is finite in both cases. On the other hand, the system time is much smaller for the sensor based policy when $\lambda > 2.5$, or equivalently when the agents velocities are decreased. Therefore the sensor based policy performs better for bigger values of λ, that is to say for smaller agent velocities.

The second remark deals with the following question: what happens if $D(t)$ in not known, but only a delayed version of it is known to the agents? Assume that the agents know $D(t - t_d)$, where t_d is a positive integrable random variable, that is assume that the delay with which the information about the set of outstanding targets reaches the agents, is a positive integrable random variable t_d. Assume moreover that $E[t_d] = \Delta$. It is immediate to see that the presence of this random delay will affect

the performance of the system, namely the time of service, increasing it by an amount equal to Δ. So for instance, the plots in Figure 7 will shift upward of a quantity equal to Δ, whenever the delay is modeled in this way.

VII. Conclusions

In this paper we considered two very simple strategies for multiple vehicle routing in the presence of dynamically-generated targets, and analyzed their performance in light load conditions, i.e., when the target generation rate is very small. The strategies we addressed in this paper are based on minimal assumptions on the ability of the agents to exchange information: in one case they do not explicitly communicate at all, and in the other case, agents are only aware of other agents' current location. A possibly unexpected and striking results of our analysis is the following: the collective performance of the agents using such minimal or no-communication strategies is (locally) optimal, and is in fact as good as that achieved by the best known decentralized strategies, see for instance [12] for a comparison. The proposed no-communication algorithm does converge to a (local) optimum, even though quite slowly compared to other policies. On the other hand, the proposed strategies do not rely on the knowledge of the target generation process, and make minimal assumptions on the target spatial distribution; in fact, the convexity and boundedness assumptions on the support of the spatial distribution can be relaxed, as long as path connectivity of the support, and absolute continuity of the distribution are ensured. Also, the distribution need not be constant: Indeed, the algorithm will provide a good approximation to a local optimum for the cost function as long as the characteristic time it takes for the target generation process to vary significantly is much greater than the relaxation time of the algorithm. In summary, the proposed strategies can be seen as a learning mechanism in which the agents learn the target generation process, and the ensuing target spatial distribution, and adapt their own behavior to it.

The proposed strategies are very simple to implement, as they only require storage of the coordinates of points visited in the past and simple algebraic calculations; the "sensor-based" strategy also require a device to estimate the position of other agents. Simple implementation and the absence of active communication makes the proposed strategies attractive, for example, in embedded systems and stealthy applications. The game-theoretic interpretation of our results also provides some insight into how territorial, globally optimal behavior can arise in a population of selfish but rational individuals even without explicit mechanisms for territory marking and defense.

While we were able to prove that the proposed strategies perform efficiently for small values of the target generation rate, little is known about their performance in other regimes. In particular, we have shown numerical evidence that suggests that the first strategy we introduced, requiring no communication, performs poorly when targets are generated very frequently, whereas the performance of the sensor-based strategy is in fact comparable to that of the best known strategies for the heavy load case.

Extensions of this work will include the analysis and design of efficient strategies for general values of the target generation

rate, for different vehicle dynamics models (e.g., including differential constraints on the motion of the agents), and heterogeneous systems in which both service requests and agents can belong to several different classes with different characteristics and abilities. Initial work on reactive coverage problems for multiple agents with various kinodynamic constraints has been reported in [36], [37]. Another direction of research is to extend the game theoretic formulation for the problem in this paper to formally study various hunting and foraging strategies in animals [38].

ACKNOWLEDGMENT

REFERENCES

[1] A. Arsie and E. Frazzoli, "Efficient routing of multiple vehicles with no explicit communications," *International Journal of Robust and Nonlinear Control*, vol. 18, pp. 154–164, January 2007.

[2] R. W. Beard, T. W. McLain, M. A. Goodrich, and E. P. Anderson, "Coordinated target assignment and intercept for unmanned air vehicles," *IEEE Trans. on Robotics and Automation*, vol. 18, no. 6, pp. 911–922, 2002.

[3] A. Richards, J. Bellingham, M. Tillerson, and J. How, "Coordination and control of multiple UAVs," in *Proc. of the AIAA Conf. on Guidance, Navigation, and Control*, (Monterey, CA), 2002.

[4] C. Schumacher, P. R. Chandler, S. J. Rasmussen, and D. Walker, "Task allocation for wide area search munitions with variable path length," in *Proc. of the American Control Conference*, (Denver, CO), pp. 3472–3477, 2003.

[5] M. Earl and R. D'Andrea, "Iterative MILP methods for vehicle control problems," *IEEE Trans. on Robotics*, vol. 21, pp. 1158–1167, December 2005.

[6] W. Li and C. Cassandras, "A cooperative receding horizon controller for multivehicle uncertain environments," *IEEE Trans. on Automatic Control*, vol. 51, no. 2, pp. 242–257, 2006.

[7] R. Murphey, "Target-based weapon target assignment problems," in *Nonlinear Assignment Problems: Algorithms and Applications* (P. Pardalos and L. Pitsoulis, eds.), pp. 39–53, Kluwer Academic Publisher, 1999.

[8] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: a game theoretical formulation," *ASME Journal of Dynamic Systems, Measurement and Control*, vol. 129, no. 5, pp. 584–596, 2007.

[9] M. Tanemura and H. Hasegawa, "Geometrical models of territory I: Models for synchronous and asynchronous settlement of territories," *Journal of Theoretical Biology*, vol. 82, pp. 477–496, 1980.

[10] H. Psaraftis, "Dynamic vehicle routing problems," in *Vehicle Routing: Methods and Studies* (B. Golden and A. Assad, eds.), Studies in Management Science and Systems, Elsevier, 1988.

[11] D. J. Bertsimas and G. J. van Ryzin, "A stochastic and dynamic vehicle routing problem in the Euclidean plane," *Operations Research*, vol. 39, pp. 601–615, 1991.

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

[13] Y. Liu, J. Cruz, and A. G. Sparks, "Coordinated networked uninhabited aer vehicles for persistent area denial," in *IEEE Conf. on Decision and Control*, (Paradise Island, Bahamas), pp. 3351–3356, 2004.

[14] J. Cortés, S. Martínez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[15] B. Moore and K. Passino, "Distributed balancing of AAVs for uniform surveillance coverage," in *IEEE Conference on Decision and Control*, pp. 7060–7065, 2005.

[16] A. Arsie, J. J. Enright, and E. Frazzoli, "On the value of information in dynamic multiple-vehicle routing problems," in *Cooperative Control of Distributed Multi-Agent Systems* (J. Shamma, ed.), pp. 139–176, John Wiley & Sons, 2008. To appear.

[17] J. Little, "A proof of the queueing formula $l = \lambda w$," *Operations Research*, vol. 9, pp. 383–387, 1961.

[18] 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.

[19] 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.

[20] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[21] J. Mac Queen, "Some methods for the classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Math. Stat. and Prob.* (L. M. LeCam and J. Neyman, eds.), pp. 281–297, University of California Press, 1967.

[22] Z. Drezner, ed., *Facility Location: A Survey of Applications and Methods*. Springer Series in Operations Research, New York: Springer Verlag, 1995.

[23] P. K. Agarwal and M. Sharir, "Efficient algorithms for geometric optimization," *ACM Computing Surveys*, vol. 30, no. 4, pp. 412–458, 1998.

[24] R. Chandrasekaran and A. Tamir, "Algebraic optimization: the Fermat-Weber location problem," *Mathematical Programming*, vol. 46, pp. 219–224, 1990.

[25] G. Wesolowsky, "The Weber problem: History and perspectives," *Location Science*, vol. 1, pp. 5–23, 1993.

[26] S. P. Fekete, J. S. B. Mitchell, and K. Weinbrecht, "On the continuous Weber and k-median problems," in *Proceedings of the Sixteenth Annual Symposium on Computational Geometry (Hong Kong, 2000)*, (New York), pp. 70–79, ACM, 2000.

[27] P. Carmi, S. Har-Peled, and M. Katz, "On the fermat-weber center of a convex object," *Computational Geometry*, vol. 32, no. 3, pp. 188–195, 2005.

[28] N. L. Carothers, *Real Analysis*. Cambridge University Press, 1999.

[29] A. Okabe, B. Boots, K. Sugihara, and S. Chiu, *Spatial tessellations: Concepts and Applications of Voronoi diagrams*. Wiley Series in Probability and Statistics, Chichester, UK: John Wiley & Sons, second ed., 2000.

[30] M. Sabin and R. Gray, "Global convergence and empirical consistency of the generalized Lloyd algorithm," *IEEE Trans. on Information Theory*, vol. 32, pp. 148–155, March 1986.

[31] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Trans. on Automatic Control*, vol. 50, pp. 312–327, March 2005.

[32] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[33] R. W. Rosenthal, "A class of games possessing pure-strategy nash equilibria," *International Journal of Game Theory*, vol. 2, pp. 65–67, 1973.

[34] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[35] D. Wolpert and K. Tumer, *Collectives and the Design of Complex Systems*, ch. A Survey of Collectives, p. 142. New York: Springer Verlag, 2004.

[36] J. Enright, K. Savla, and E. Frazzoli, "Coverage control for teams of nonholonomic agents," in *IEEE Conf. on Decision and Control*, 2008. To appear.

[37] K. Savla and E. Frazzoli, "On endogenous reconfiguration for mobile robotic networks," in *Workshop on Algorithmic Foundations of Robotics (WAFR)*, (Guanajuato, Mexico), December 2008. To appear.

[38] L. A. Dugatkin and H. K. Reeve, eds., *Game Theory and Animal Behavior*. Oxford University Press, 2000.