Game-theoretic learning algorithm for a spatial coverage problem

Ketan Savla and Emilio Frazzoli
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
ksavla@mit.edu, frazzoli@mit.edu

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 cast the problem as a spatial game in which each agent's objective is to maximize the expected value of the "time spent alone" at the next target location and show that the Nash equilibria of the game correspond to the desired agent configurations. We propose learning-based control strategies that, while making minimal or no assumptions on communications between agents as well as the underlying distribution, provide the same level of steady-state performance achieved by the best known decentralized strategies.**

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.

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 [1], [2], [3], [4], [5]. 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 [6], [7]. 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 [8], 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 [9], [10], and decentralized strategies were presented in [11]. This problem has connections to the Persistent Area Denial (PAD) and area coverage problems discussed, e.g., in [3], [12], [13], [14].

In this paper we cast a dynamic vehicle routing problem as a spatial game in which each agent's objective is to maximize the expected value of the "time spent alone" at the next target location and show that the Nash equilibria of the game correspond to the desired agent configurations. We propose learning-based control strategies that, while making minimal or no assumptions on communications between agents as well as the underlying distribution, provide the same level of steady-state performance achieved by the best known decentralized strategies. In this sense, the problem can 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.

II. Problem Formulation

Let $\mathcal{Q} \subset \mathbb{R}^2$ be a convex and compact domain on the plane; we will refer to $\mathcal{Q}$ as the *workspace*. A stochastic process generates *service requests* over time, which are associated to points in $\mathcal{Q}$; 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 $\varphi : \mathcal{Q} \to \mathbb{R}_+$, with support $\mathcal{Q}$ (i.e., $\varphi(q) > 0 \Leftrightarrow q \in \mathcal{Q}$). The spatial density function φ is normalized in such a way that $\int_{\mathcal{Q}} \varphi(q)\, dq = 1$. Both λ and φ are not necessarily known.

A spatio-temporal Poisson point process is a collection of functions $\{\mathcal{P} : \overline{\mathbb{R}}_+ \to 2^{\mathcal{Q}}\}$ such that, for any $t > 0$, $\mathcal{P}(t)$ is a random collection of points in $\mathcal{Q}$, 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 $\mathcal{S} \subseteq \mathcal{Q}$ that satisfies
$$\Pr\left[\mathrm{card}\left((\mathcal{P}(s+t) - \mathcal{P}(s)) \cap \mathcal{S}\right) = k\right] = \\ = \frac{\exp(-\lambda t \cdot \varphi(\mathcal{S}))(\lambda t \cdot \varphi(\mathcal{S}))^k}{k!}, \quad \forall k \in \mathbb{N},$$
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 $S \subseteq \mathcal{Q}$ during a time interval of length Δt is given by:

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

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. Let $p(t) = (p_1(t), p_2(t), \ldots, p_m(t)) \in \mathcal{Q}^m$ be a vector describing the positions of the agents at time t. The agents are free to move, with bounded speed, within the workspace $\mathcal{Q}$; 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

$$\dot{p}_i(t) = u_i(t), \quad \text{with } \|u_i(t)\| \leq 1, \quad \forall t \geq 0, \qquad (1)$$

for each $i \in \{1, \ldots, m\}$. The agents are identical, and have unlimited range and target-servicing capability.

Let $\mathcal{B}_i(t) \subset \mathcal{Q}$ 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).

In this paper we concentrate our investigation on the light load case, in which the target generation rate is very small, i.e., when $\lambda \to 0^+$. In that case, the agents will spend most of the time staying idle, say at locations $\pi = \{\pi_1, \ldots, \pi_m\}$ and upon the arrival of a new target the closest one goes toward the target location.

In the limit as $\lambda \to 0^+$, at most one service request is outstanding at any given time with probability one. 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

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

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

Let T_j be the time elapsed between the issuance of the j-th service request, and the time it is fulfilled, and let $\overline{T}_\pi := \lim_{j \to \infty} \mathrm{E}[T_j]$ be 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 loitering policy that yields a quality of service (i.e., system time) achieving, or approximating, the theoretical optimal performance given by

$$\overline{T}_{\mathrm{opt}} = \min_{\pi} \overline{T}_\pi. \qquad (2)$$

III. On the properties of the cost function

If all service requests are generated with the agents at their reference positions, the average service time (for small λ) can be evaluated as

$$\begin{aligned}\overline{T}_{\mathrm{opt}} &= \int_{\mathcal{Q}} \min_{i=1,\dots,m} \|\hat{p}_i^* - q\| \; \varphi(q)dq \\ &= \sum_{i=1}^{m} \int_{\mathcal{V}_i(p^*)} \|\hat{p}_i^* - q\| \; \varphi(q)dq.\end{aligned}$$

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 [15]. This function admits a global minimum (in general not unique) for all non-singular density functions φ, and in fact it is known [10] that the optimal performance in terms of system time is given by

$$\overline{T}_{\mathrm{opt}} = \min_{p\in\mathcal{Q}^m} \sum_{i=1}^{m} \int_{\mathcal{V}_i(p)} \|p_i - q\| \; \varphi(q)dq. \qquad (3)$$

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 3.1 (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 3.2 (Median Voronoi Tessellation): A Voronoi tessellation $\mathcal{V}(p) = \{\mathcal{V}_1(p), \dots, \mathcal{V}_m(p)\}$ 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

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

IV. Spatial Game Formulation in Strategic Form

In this section we cast the coverage problem as a spatial game. In particular, we frame our presentation along the works [7], [16], 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.

A. Game formulation in the strategic form

We first formulate the game in the strategic form [17] as follows. Consider a scenario with the same workspace $\mathcal{Q}$, 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 [18], 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 $i \in \{1, \dots, m\}$. 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., $\pi_{-i} := (\pi_1, \dots, \pi_{i-1}, \pi_{i+1}, \dots, \pi_m)$. 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

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \mathrm{E}_q[r_i(q, \pi)]. \qquad (4)$$

Equation (4) 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

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \mathrm{E}_q[\max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\}] =$$
$$= \int_{\mathcal{Q}} \max\{0, \min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|\} \varphi(q) dq. \qquad (5)$$

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 (5), we derive the following expression for the utility function.

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \int_{\mathcal{V}_i(\pi)} (\min_{j \neq i} \|\pi_j - q\| - \|\pi_i - q\|) \varphi(q) dq. \qquad (6)$$

B. Properties of the Game

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 [19] is as follows.

Definition 4.1: 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 $\psi : \times_{i \in \{1, \dots, m\}} \Pi_i \to \mathbb{R}$,

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

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

Proposition 4.2: $\mathcal{G}$ is a potential game.

Proof: The expression for the utility function of agent i, as given by Equation (6), can be rewritten as:

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

$$= \int_Q \min_{j\neq i} \|\pi_j - q\|\varphi(q)dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\|\varphi(q)dq. \qquad (7)$$

In the integrand of the first term in Equation (7), $\min_{j\neq i} \|\pi_j - q\|$ is the distance from point q to the closest among all the agents, except the i^{th} 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 (7) can then be written as

$$\mathcal{U}_i(\pi_i, \pi_{-i}) = \sum_{\substack{j=1 \\ j\neq i}}^{m} \int_{\mathcal{V}_j(\pi_{-i})} \|\pi_j - q\|\varphi(q)dq - \sum_{j=1}^{m} \int_{\mathcal{V}_j(\pi)} \|\pi_j - q\|\varphi(q)dq. \qquad (8)$$

Note that the first term on the right hand side of Equation (8) is independent of π_i. With this observation, consider the following potential function:

$$\psi(\pi) = -\sum_{i=1}^{m} \int_{\mathcal{V}_i(\pi)} \|\pi_i - q\|\varphi(q)dq. \qquad (9)$$

The proposition then follows by combining Equations (8) and (9) 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 4.3: 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 $\psi : \times_{i\in\{1,\dots,m\}}\Pi_i \to \mathbb{R}$,

$$\mathcal{U}_i(\pi_i^{'}, \pi_{-i}) - \mathcal{U}_i(\pi_i^{"}, \pi_{-i}) > 0$$

if and only if

$$\psi(\pi_i^{'}, \pi_{-i}) - \psi(\pi_i^{"}, \pi_{-i}) > 0,$$

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

Remark 4.4: Note that every potential game is also an ordinal potential game.

Definition 4.5: The set of agents utilities $\{\mathcal{U}_i\}_{i=1,\dots,m}$ is *aligned* with the global utility $\mathcal{U}_g$ if and only if the game with utility functions $\{\mathcal{U}_i\}_{i=1,\dots,m}$ 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

$$\mathcal{U}_g(\pi) = -\sum_{i=1}^{m} \int_{\mathcal{V}_i(\pi)} \|\pi_i - q\|\varphi(q)dq. \qquad (10)$$

Proposition 4.6: The utility functions of the agents in $\mathcal{G}$ are aligned with its global utility function.

Proof: Comparing Equation (10) with the expression of the potential function in Equation (9), we observe that $\mathcal{U}_g(\pi) = \psi(\pi)$. Proposition 4.2 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 4.5. ■

We now show that our utility function belongs to a class of utility functions called Wonderful Life Utility Functions [20].

Definition 4.7: 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 4.8: The Wonderful Life utility function measures the marginal contribution of an agent towards the global utility.

Proposition 4.9: The local utility function, as defined in Equation (8) is a Wonderful Life utility function with respect to the global utility function defined in Equation (10).

Proof: We refer to Equation (8) in the proof of Proposition 4.2, 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 (10), 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 $\hat{p}^* = \{\hat{p}_1^*, \dots, \hat{p}_m^*\}$ such that, for all $i \in \{1,\dots,m\}$, $\hat{p}_i^*$ 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 [7]:

Definition 4.10: A strategy $\tilde{\pi}$ is called a *pure Nash equilibrium* if, for all $i = \{1,\dots,m\}$,

$$\mathcal{U}_i(\tilde{\pi}_i, \tilde{\pi}_{-i}) = \max_{\pi_i\in\Pi_i} \mathcal{U}_i(\pi_i, \tilde{\pi}_{-i}). \qquad (11)$$

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

Proposition 4.11: The $\hat{p}^*$ strategy is an efficient pure Nash equilibrium for the game $\mathcal{G}$.

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

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

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

Remark 4.12: Note that we do *not* claim in Proposition 4.11 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.

V. Learning Algorithm

In this section, we propose a learning based control algorithm for the spatial game $\mathcal{G}$.

A. The complete information case

A similar algorithm can be given for the problem of achieving Median Voronoi Tessellations. Indeed, assume that the agents update generators' locations (in this case, the position of the generators should coincide with the position of the agents) and weights according to the following law.

$$\dot{\pi}_i = -\int_{\mathcal{V}_i(\pi)} \frac{\pi_i - q}{\|\pi_i - q\|} \varphi(q) dq, \quad \pi_i(0) \in \mathcal{Q} \qquad (12)$$

The success of the previous strategy depends on the fidelity of communication channels between agents. In fact, a common theme in cooperative control is the investigation of the effects of different communication and information sharing protocols on the system performance.

B. The limited information case

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:

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

(ii) 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,

where $\mathcal{D}(t)$ is the set of locations of the outstanding tasks. In other words, we set

$$\begin{aligned}&\pi_{\mathrm{nc}}(p_i(t), \mathcal{B}_i(t), \mathcal{D}(t)) \\ &\qquad = \mathrm{vers}(F_{\mathrm{nc}}(p_i(t), \mathcal{B}_i(t), \mathcal{D}(t)) - p_i(t)), \quad (13)\end{aligned}$$

where

$$F_{\mathrm{nc}}(p_i, \mathcal{B}_i, \mathcal{D}) = \begin{cases} \arg\min_{q \in \mathcal{D}} \|p_i - q\|, & \text{if } \mathcal{D} \neq \emptyset, \\ \arg\min_{q \in \Omega} \sum_{e \in \mathcal{B}_i} \|e - q\|, & \text{otherwise,} \end{cases} \qquad (14)$$

$\|\cdot\|$ 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 [15], [21] (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 (14) 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.

Theorem 5.1: The system time provided by the learning algorithm converges to a critical point (either a saddle point or a local minimum) with high probability as $\lambda \to 0^+$.

The proof of Theorem 5.1 is reported in [22]. We provide a brief outline of the proof here.

(i) First, we prove that the reference point of any agent that visits an unbounded number of targets over time converges almost surely.

(ii) 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.

(iii) 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).

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

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

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

[16] 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.

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

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

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

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

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

[22] A. Arsie, K. Savla, and E. Frazzoli, "Efficient routing algorithms for multiple vehicles with no explicit communications," *IEEE Trans. on Automatic Control*, 2009. In press.