Multiple-Agent Probabilistic Pursuit-Evasion Games[1]

João P. Hespanha
hespanha@usc.edu
Electrical Eng.–Systems
Univ. of Southern California

Hyoun Jin Kim
jin@eecs.berkeley.edu
Electrical Eng. & Comp. Science
Univ. of California at Berkeley

Shankar Sastry
sastry@eecs.berkeley.edu
Electrical Eng. & Comp. Science
Univ. of California at Berkeley

Abstract

In this paper we develop a probabilistic framework for pursuit-evasion games. We propose a "greedy" policy to control a swarm of autonomous agents in the pursuit of one or several evaders. At each instant of time this policy directs the pursuers to the locations that maximize the probability of finding an evader at that particular time instant. It is shown that, under mild assumptions, this policy guarantees that an evader is found in finite time and that the expected time needed to find the evader is also finite. Simulations are included to illustrate the results.

1 Introduction

This paper addresses the problem of controlling a swarm of autonomous agents in the pursuit of one or several evaders. To this effect we develop a probabilistic framework for pursuit-evasion games involving multiple agents. The problem is nondeterministic because the motions of the pursuers/evaders and the devices they use to sense their surroundings require probabilistic models. It is also assumed that when the pursuit starts only an *a priori* probabilistic map of the region is known. A probabilistic framework for pursuit-evasion games avoids the conservativeness of deterministic worst-case approaches.

Pursuit-evasion games arise in numerous situations. Typical examples are search and rescue operations, localization of (possibly moving) parts in a warehouse, search and capture missions, etc. In some cases the evaders are actively avoiding detection (e.g., search and capture missions) whereas in other cases their motion is approximately random (e.g., search and rescue operation). The latter problems are often called *games against nature*.

Deterministic pursuit-evasion games on finite graphs have been well studied [1, 2]. In these games, the region in which the pursuit takes place is abstracted to be a finite collection of nodes and the allowed motions for the pursuers and evaders are represented by edges connecting the nodes. An evader is "captured" if he and one of the pursuers occupy the same node. A question often studied within the context of pursuit-evasion games on graphs is the computation of the search number $s(G)$ of a given graph G. By the "search number" it is meant the smallest number of pursuers needed to capture a single evader in finite time, regardless of how the evader decides to move. It turns out that determining if $s(G)$ is smaller than a given constant is NP-hard [2, 3]. Pursuit-evasion games on graphs have been limited to worst-case motions of the evaders.

When a region in which the pursuit takes place is abstracted to a finite graph, the sensing capabilities of each pursuer becomes restricted to a single node: the node occupied by the pursuer. The question then arises of how to decompose a given continuous space F into a finite number of regions, each to be mapped to a node in a graph G, so that the game on the resulting finite graph G is equivalent to the original game played on F [4, 5]. LaValle *et al.* [5] propose a method for this decomposition based on the principle that an evader is captured if it is in the line-of-sight of one of the pursuers. They present algorithms that build finite graphs that abstract pursuit-evasion games for known polygonal environments [5] and simply-connected, smooth-curved, two-dimensional environment [6].

So far the literature on pursuit-evasion games always assumed the region on which the pursuit takes place (be it a finite graph or a continuous terrain) is known. When the region is unknown *a priori* a "map-learning" phase is often proposed to precede the pursuit. However, systematic map learning is time consuming and computationally hard, even for simple two-dimensional rectilinear environments with each side of the obstacles parallel to one of the coordinate axis [7]. In practice, map learning is further complicated by the fact that the sensors used to acquire the data upon which the map is built are not accurate. In [8] an algorithm is proposed for maximum likelihood estimation of the map of a region from noisy observations obtained by a mobile robot.

Our approach differs from others in the literature in that we combine exploration (or map-learning) and pursuit in a single problem. Moreover, this is done in a probabilistic framework to avoid the conservative-

[1]This research was supported by the Office of Naval Research (grant N00014-97-1-0946).

ness inherent to worst-case assumptions on the motion of the evader. A probabilistic framework is also natural to take into account the fact that sensor information is not precise and that only an inaccurate *a priori* map of the terrain may be known [8].

The remaining of this paper is organized as follows: A probabilistic pursuit-evasion game is formalized in Section 2, and performance measures for pursuit policies are proposed. In Section 3 it is shown that pursuit policies with a certain "persistency" property are guaranteed to find an evader in finite time with probability one. Moreover, the expected time needed to do this is also finite. In Section 4 specific persistent policies are proposed for simple multi-pursuers/single-evader games with inaccurate observations and obstacles. Simulation results are shown in Section 5 for a two-dimensional pursuit game and Section 6 contains some concluding remarks and directions for future research. The reader is referred to [9] for the proofs of some of the results presented here.

Notation: We denote by $(\Omega, \mathcal{F}, \mathrm{P})$ the relevant *probability space* with Ω the set of all possible events related to the pursuit-evasion game, $\mathcal{F}$ a family of subsets of Ω forming a σ-algebra, and $\mathrm{P} : \mathcal{F} \to [0,1]$ a probability measure on $\mathcal{F}$. Given two sets of events $A, B \in \mathcal{F}$ with $\mathrm{P}(B) \neq 0$, we write $\mathrm{P}(A|B)$ for the *conditional probability of A given B*. Bold face symbols are used to denote random variables.

2 Pursuit policies

For simplicity we assume that both space and time are quantized. The region in which the pursuit takes place is then regarded as a finite collection of cells $\mathcal{X} := \{1, 2, \dots, n_c\}$ and all events take place on a set of discrete times $\mathcal{T}$. Here, the events include the motion and collection of sensor data by the pursers/evaders. For simplicity of notation we take equally spaced event times. In particular, $\mathcal{T} := \{1, 2, \dots\}$.

For each time $t \in \mathcal{T}$, we denote by $\mathbf{y}(t)$ the set of all measurements taken by the pursuers at time t. Every $\mathbf{y}(t)$ is assumed a random variable with values in a measurement space $\mathcal{Y}$. At each time $t \in \mathcal{T}$ it is possible to execute a control action $\mathbf{u}(t)$ that, in general, will affect the pursuers sensing capabilities at times $\tau \geq t$. Each control action $\mathbf{u}(t)$ is a function of the measurements before time t and should therefore be regarded as a random variable taking values in a control action space $\mathcal{U}$.

For each time $t \in \mathcal{T}$ we denote by $\mathbf{Y}_t \in \mathcal{Y}^*$ the sequence[1] of measurements $\{\mathbf{y}(1), \dots, \mathbf{y}(t)\}$ taken up to time t. By the *pursuit policy* we mean the function $\mathbf{g} : \mathcal{Y}^* \to \mathcal{U}$ that maps the measurements taken up to some time to the control action executed at the next time instant, i.e.,

$$\mathbf{u}(t+1) = \mathbf{g}(\mathbf{Y}_t), \qquad t \in \mathcal{T}. \tag{1}$$

Formally, we regard the pursuit policy $\mathbf{g}$ as a random variable and, when we want to study performance of a specific function $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$ as a pursuit policy, we condition the probability measure to the event $\mathbf{g} = \bar{g}$. To shorten the notation, for each $A \in \mathcal{F}$, we abbreviate $\mathrm{P}(A \mid \mathbf{g} = \bar{g})$ by $\mathrm{P}_{\bar{g}}(A)$. The goal of this paper is to develop pursuit policies that guarantee some degree of success for the pursuers. We defer a more detailed description of the nature of the control actions and the sensing devices to later.

Take now a specific pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$. Because the sensors used by the pursuers are probabilistic, in general it may not be possible to guarantee with probability one that an evader was found. In practice, we say that an evader was *found at time $t \in \mathcal{T}$* when one of the pursuers is located at a cell for which the (conditional) posterior probability of the evader being there, given the measurements $\mathbf{Y}_t$ taken by the pursuers up to t, exceeds a certain threshold $p_{\text{found}} \in (0,1]$. At each time instant $t \in \mathcal{T}$ there is then a certain probability of one of the evaders being found. We denote by $\mathbf{T}^*$ the first time instant in $\mathcal{T}$ at which one of the evaders is found, if none is found in finite time we set $\mathbf{T}^* = +\infty$. $\mathbf{T}^*$ can be regarded as a random variable with values in $\bar{\mathcal{T}} := \mathcal{T} \cup \{+\infty\}$. We denote by $F_{\bar{g}} : \bar{\mathcal{T}} \to [0,1]$ its distribution function, i.e., $F_{\bar{g}}(t) := \mathrm{P}_{\bar{g}}(\mathbf{T}^* \leq t)$. One can show [9] that

$$F_{\bar{g}}(t) = 1 - \prod_{\tau=1}^{t} \big(1 - f_{\bar{g}}(\tau)\big), \qquad t \in \mathcal{T}, \tag{2}$$

where, for each $t \in \mathcal{T}$, $f_{\bar{g}}(t)$ denotes the conditional probability of finding an evader at time t, given that none was found up to that time, i.e., $f_{\bar{g}}(t) := \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t \mid \mathbf{T}^* \geq t)$. Moreover, when the probability of $\mathbf{T}^*$ being finite is equal to one,

$$E_{\bar{g}}[\mathbf{T}^*] = \sum_{t=1}^{\infty} t f_{\bar{g}}(t) \left(\prod_{\tau=1}^{t-1} \big(1 - f_{\bar{g}}(\tau)\big) \right). \tag{3}$$

The expected value of $\mathbf{T}^*$ provides a good measure of the performance of a pursuit policy. However, since the dependence of the $f_{\bar{g}}$ on the specific pursuit policy $\bar{g}$ is, in general, complex, it may be difficult to minimize $E_{\bar{g}}[\mathbf{T}^*]$ by choosing an appropriate pursuit policy. In the next section we concentrate on pursuit policies that, although not minimizing $E_{\bar{g}}[\mathbf{T}^*]$, guarantee upper bounds for this expected value.

Before proceeding we discuss—for the time being at an abstract level—how to compute $f_{\bar{g}}$ from known models

[1] Given a set $\mathcal{A}$ we denote by $\mathcal{A}^*$ the set of all finite sequences of elements of $\mathcal{A}$ and, given some $a \in \mathcal{A}^*$, we denote by $|a|$ the length of the sequence a.

for the sensors and the motion of the evader. A more detailed discussion for a specific game is deferred to Sections 4 and 5. Since the decision to whether or not an evader was found at some time t is completely determined by the measurements taken up to that time, it is possible to compute the conditional probability $f_{\bar{g}}(t)$ of finding an evader at time t, given that none was found up to $t-1$, as a function of the conditional probability of finding an evader for the first time at t, given the measurements taken up to $t-1$. Suppose we denote by $\mathcal{Y}_\tau^{\neg\text{fnd}} \subset \mathcal{Y}^*$, $\tau \in \mathcal{T}$, the set of all sequences of measurements of length τ, associated with an evader not being found up to that time, i.e., $\mathcal{Y}_\tau^{\neg\text{fnd}} = \mathbf{Y}_\tau(\{\omega \in \Omega : \mathbf{T}^*(\omega) > \tau\})$. Since the decision to whether or not an evader was found up to time τ is purely a function of the measurements $\mathbf{Y}_\tau$ taken up to τ, we have that $\{\omega \in \Omega : \mathbf{T}^*(\omega) \geq \tau\} = \{\omega \in \Omega : \mathbf{Y}_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}}\}$, $\tau \in \mathcal{T}$. We can then expand $f_{\bar{g}}(t)$ as

$$f_{\bar{g}}(t) = E_{\bar{g}}[h_{\bar{g}}(\mathbf{Y}_{t-1}) \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}], \tag{4}$$

where $h_{\bar{g}} : \mathcal{Y}^* \to [0,1]$ is a function that maps each sequence $Y \in \mathcal{Y}^*$ of $\tau := |Y|$ measurements to the conditional probability of finding an evader for the first time at $\tau+1$, given the measurements $\mathbf{Y}_\tau = Y$ taken up to τ, i.e., $h_{\bar{g}}(Y) := \mathrm{P}_{\bar{g}}(\mathbf{T}^* = |Y|+1 \mid \mathbf{Y}_{|Y|} = Y)$ [9]. This equation allows one to compute the probabilities $f_{\bar{g}}(t)$ using the function $h_{\bar{g}}$. The latter effectively encodes the information relevant for the pursuit-evasion game that is contained in the models for the sensors and for the motion of the evader.

3 Persistent pursuit policies

A specific pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$ is said to be *persistent* if there is some $\epsilon > 0$ such that

$$f_{\bar{g}}(t) \geq \epsilon, \qquad \forall t \in \mathcal{T}. \tag{5}$$

From (2) it is clear that, for each $t \in \mathcal{T}$, $F_{\bar{g}}(t)$ is monotone nondecreasing with respect to any of the $f_{\bar{g}}(\tau)$, $\tau \in \mathcal{T}$. Therefore, for a persistent pursuit policy $\bar{g}$, $F_{\bar{g}}(t) \geq 1-(1-\epsilon)^t$, $t \in \mathcal{T}$, with ϵ as in (5). From this we conclude that $\sup_{t<\infty} F_{\bar{g}}(t) = 1$ and therefore the probability of $\mathbf{T}^*$ being finite must be equal to one. The expected value $E_{\bar{g}}[\mathbf{T}^*]$, on the other hand, is monotone nonincreasing with respect to any of the $f_{\bar{g}}(t)$, $t \in \mathcal{T}$ (cf. equation (3) and Lemma 4 in [9]). Therefore, for the same pursuit policy we also have that $E_{\bar{g}}[\mathbf{T}^*] \leq \epsilon \sum_{t=1}^{\infty} t(1-\epsilon)^{t-1} = \epsilon^{-1}$. The following can then be stated:

Lemma 1 *For a persistent pursuit policy $\bar{g}$, $\mathrm{P}_{\bar{g}}(\mathbf{T}^* < \infty) = 1$, $F_{\bar{g}}(t) \geq 1-(1-\epsilon)^t$, $t \in \mathcal{T}$, and $E_{\bar{g}}[\mathbf{T}^*] \leq \epsilon^{-1}$, with ϵ as in* (5).

Often pursuit policies are not persistent in the way defined above but they are *persistent on the average*. By this we mean that there is an integer T and some $\epsilon > 0$ such that, for each $t \in \mathcal{T}$, the conditional probability of finding an evader on the set of T consecutive time instants starting at t, given that none was found up to that time, is greater or equal to ϵ. In particular,

$$\bar{f}_{\bar{g}}(t) := \mathrm{P}_{\bar{g}}(\mathbf{T}^* \in \{t, t+1, \dots, t+T-1\} \mid \mathbf{T}^* \geq t) \geq \epsilon, \qquad \forall t \in \mathcal{T}. \tag{6}$$

We call T the *period of persistence*. The following extension of Lemma 1 is proved in [9]:

Lemma 2 *For a persistent on the average pursuit policy $\bar{g}$, with period T, $\mathrm{P}_{\bar{g}}(\mathbf{T}^* < \infty) = 1$, $F_{\bar{g}}(t) \geq 1-(1-\epsilon)^{\lfloor \frac{t}{T} \rfloor}$, $t \in \mathcal{T}$, and $E_{\bar{g}}[\mathbf{T}^*] \leq T\epsilon^{-1}$, with ϵ as in* (6).

Lemmas 1 and 2 show that, with persistent policies, the probability of finding the evader in finite time is equal to one and the expected time needed to find it is always finite. Moreover, these lemmas give simple bounds for the expected value of the time at which the evader is found. This makes persistent policies very attractive. It turns out that often it is not hard to design policies that are persistent. The next section describes a pursuit-evasion game for which this is the case. Before proceeding note that a sufficient condition for (5) to hold—and therefore for $\bar{g}$ to be persistent—is that

$$h_{\bar{g}}(Y) \geq \epsilon, \qquad \forall t \in \mathcal{T}, Y \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}. \tag{7}$$

This is a direct consequence of (4). A somewhat more complex condition, also involving $h_{\bar{g}}$, can be found for persistency on the average [9].

4 Pursuit-evasion games with partial observations and obstacles

In the game considered in this section, n_p pursuers try to find a single evader. We denote by $\mathbf{x}_e$ the position of the evader and by $\mathbf{x} := \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_p}\}$ the positions of the pursuers. Both the evader and the pursuers can move and therefore $\mathbf{x}_e$ and $\mathbf{x}$ are time-dependent quantities. At each time $t \in \mathcal{T}$, $\mathbf{x}_e(t)$ and each $\mathbf{x}_i(t)$ are random variables taking values on $\mathcal{X}$.

Some cells contain fixed obstacles and neither the pursuers nor evader can move to these cells. The positions of the obstacles are represented by a function $\mathbf{m} : \mathcal{X} \to \{0,1\}$ that takes the value 1 precisely at those cells that contain an obstacle. The function $\mathbf{m}$ is called the *obstacle map* and for each $x \in \mathcal{X}$, $\mathbf{m}(x)$ is a random variable. All the $\mathbf{m}(x)$ are assumed independent. When the game starts only an "*a priori* obstacle map" is known. By an *a priori obstacle map* we mean a function $p_m : \mathcal{X} \to [0,1]$ that maps each $x \in \mathcal{X}$ to the probability of cell x containing an obstacle, i.e.,

$p_m(x) = \mathrm{P}\big(\mathbf{m}(x) = 1\big)$, $x \in \mathcal{X}$. This probability is assumed independent of the pursuit policy.

At each time $t \in \mathcal{T}$, the control action $\mathbf{u}(t) := \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_p}\}$ consists of a list of desired positions for the pursuers at time t. Two pursuers should not occupy the same cell, therefore each $\mathbf{u}(t)$ must be an element of the control action space $\mathcal{U} := \big\{\{v_1, v_2, \dots, v_{n_p}\} : v_i \in \mathcal{X},\ v_i \neq v_j \text{ for } i \neq j\big\}$.

Each pursuer is capable of determining its current position and sensing a region around it for obstacles or the evader but the sensor readings may be inaccurate. In particular, there is a nonzero probability that a pursuer reports the existence of an evader/obstacle in a nearby cell when there is none, or vice-versa. However, we assume that the information the pursuers report regarding the existence of evaders in the cell that they are occupying is accurate. In this game we then say that the evader was *found* at some time $t \in \mathcal{T}$, only when a pursuer is located at a cell for which the conditional probability of the evader being there, given the measurements $\mathbf{Y}_t$ taken up to t, is equal to one.

For the results in this section we do not need to specify precise probabilistic models for the pursuers sensors nor for the motion of the evader. However, we will assume that, for each $x \in \mathcal{X}$, $Y \in \mathcal{Y}^*$, it is possible to compute the conditional probability $p_e(x, Y)$ of the evader being in cell x at time $t+1$, given the measurements $\mathbf{Y}_t = Y$ taken up to $t := |Y|$. We also assume that this probability is independent of the pursuit policy being used, i.e., for every specific pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$, and every $x \in \mathcal{X}$, $Y \in \mathcal{Y}^*$,

$$p_e(x, Y) = \mathrm{P}_{\bar{g}}\big(\mathbf{x}_e(|Y| + 1) = x \mid \mathbf{Y}_{|Y|} = Y\big). \tag{8}$$

In practice, this amounts to saying that a model for the motion of the evader is known, or can be estimated. In [9] it is shown how the function p_e can be efficiently computed when the motion of the evader follows a Markov model. In fact, we shall see that for every sequence $Y_t \in \mathcal{Y}^*$ of $t := |Y| \in \mathcal{T}$ measurements, the $p_e(x, Y_t)$, $x \in \mathcal{X}$, can be computed as a deterministic function of the the last measurement $y(t)$ in Y_t and the $p_e(x, Y_{t-1})$, $x \in \mathcal{X}$, with $Y_{t-1} \in \mathcal{Y}^*$ denoting the first $t-1$ measurements in Y_t. The function p_e can therefore be interpreted as an "information state" for the game [10].

4.1 Greedy policies with unconstrained motion

We start by assuming that the pursuers are fast enough to move from any cell to any other cell in a single time step. When this happens we say that the motion of the pursuers is unconstrained. By the *greedy pursuit policy with unconstrained motion* we mean the policy $g_u : \mathcal{Y}^* \to \mathcal{U}$ that, at each instant of time t, moves the pursuers to the positions that maximize the (conditional) posterior probability of finding the evader at time $t+1$, given the measurements $\mathbf{Y}_t = Y$ taken by the pursuers up to t, i.e., for each $Y \in \mathcal{Y}^*$,

$$g_u(Y) = \arg \max_{\{v_1, v_2, \dots, v_{n_p}\} \in \mathcal{U}} \sum_{k=1}^{n_p} p_e(v_k, Y). \tag{9}$$

We show next that this pursuit policy is persistent. To this effect consider an arbitrary sequence $Y \in \mathcal{Y}_t^{\neg\text{fnd}}$ of $t := |Y|$ measurements for which the evader was not found. Since the data regarding the existence of an evader in the same cell as one of the pursuers is assumed accurate, finding an evader at $t+1$ for the first time is precisely equivalent to having the evader in one of the cells occupied by a pursuer at $t+1$. The conditional probability $h_{g_u}(Y)$ of finding the evader for the first time at $t+1$, given the measurements $\mathbf{Y}_t = Y \in \mathcal{Y}_t^{\neg\text{fnd}}$ taken up to t, is then given by

$$h_{g_u}(Y) = \mathrm{P}_{g_u}\big(\exists k \in \{1, 2, \dots, n_p\} : \\ \mathbf{x}_e(t+1) = \mathbf{u}_k(t+1) \mid \mathbf{Y}_t = Y\big).$$

Moreover, since there is only one evader and all the $\mathbf{u}_k$ are distinct, we further conclude that

$$h_{g_u}(Y) = \sum_{k=1}^{n_p} \mathrm{P}_{g_u}\big(\mathbf{x}_e(t+1) = \mathbf{u}_k(t+1) \mid \mathbf{Y}_t = Y\big). \tag{10}$$

Let now $\{v_1, v_2, \dots, v_{n_p}\} := g_u(Y)$. Because of (1) we have $\mathbf{u}_k(t+1) = v_k$, $k \in \{1, 2, \dots, n_p\}$, given that $\mathbf{Y}_t = Y$ and $g = g_u$. From this, (8), and (10) we conclude that $h_{g_u}(Y) = \sum_{k=1}^{n_p} p_e(v_k, Y)$. Because of (9) and the fact that $\sum_{x=1}^{n_c} p_e(x, Y) = 1$, we then obtain

$$h_{g_u}(Y) = \sum_{k=1}^{n_p} p_e(v_k, Y) \geq \frac{n_p}{n_c} \sum_{x=1}^{n_c} p_e(x, Y) = \epsilon := \frac{n_p}{n_c}. \tag{11}$$

Here we used the fact that, given any set of n_c numbers, the sum of the largest $n_p \leq n_c$ of them, is larger or equal to n_p/n_c times the sum of all of them. From (11) one concludes that g_u is persistent (cf. (7)) and, because of Lemma 1, we can state the following:

Theorem 1 *The greedy pursuit policy with unconstrained motion g_u is persistent. Moreover,* $\mathrm{P}_{g_u}(\mathbf{T}^* < \infty) = 1$, $F_{g_u}(t) \geq 1 - \left(1 - \frac{n_p}{n_c}\right)^t$, $t \in \mathcal{T}$, *and* $E_{g_u}[\mathbf{T}^*] \leq \frac{n_c}{n_p}$.

The upper bound on $E_{g_u}[\mathbf{T}^*]$ provided by Theorem 1 is independent of the specific model used for the motion of the evader. In particular, if the evader moves according to a Markov model (cf. Section 5), the bound for $E_{g_u}[\mathbf{T}^*]$ is independent of the probability ρ of the evader moving from his present cell to a distinct one (ρ can be viewed as measure of the "speed" of the evader). This constitutes an advantage of g_u over simpler policies. For example, with a Markov evader, one could be tempted to use a "stay-in-place" policy defined by $g_{x^*}(Y) = x^*$, $Y \in \mathcal{Y}^*$, for some fixed $x^* \in \mathcal{X}$. However, for such a policy $E_{g_{x^*}}[\mathbf{T}^*]$ would increase as the "speed" of the evader ρ decreases. In fact, in the extreme case of $\rho = 0$ (evader not moving), the probability of finding the evader in finite time would actually be smaller than one.

4.2 Greedy policies with constrained motion

Suppose now that the motion of each pursuer is constrained by that, in a single time step, it can only move to cells close to its present position. Formally, if at a time $t \in \mathcal{T}$ the pursuers are positioned in the cells $v := \{v_1, v_2, \dots, v_{n_p}\} \in \mathcal{U}$, we denote by $\mathcal{U}(v)$ the subset of $\mathcal{U}$ consisting of those lists of cells to which the pursuers could move at time $t+1$, were these cells empty. We say that the lists of cells in $\mathcal{U}(v)$ are *reachable from v in a single time step.* A pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$ is called *admissible* if, for every sequence of measurements $Y \in \mathcal{Y}^*$, $\bar{g}(Y)$ is reachable in a single time step from the positions v of the pursuers specified in the last measurement in Y, i.e., $\bar{g}(Y) \in \mathcal{U}(v)$. Although the motion of the pursuers in a single time step is constrained, we shall assume that their motions are not constrained over a sufficiently large time interval, i.e., that the cells without obstacles form a connected region, with connectivity defined in terms of the allowed motions for the pursuers.

When the motion of the pursuers is constrained, greedy policies similar to the one defined in Section 4.1 may not yield a persistent pursuit policy. For example, it could happen that the probability of existing an evader in any of the cells to which the pursuers can move is exactly zero. With constrained motion, the best one can hope for is to design a pursuit policy that is persistent on the average. To do this we need the following assumption:

Assumption 1 *There is a positive constant $\gamma \le 1$ such that for any sequence $Y_t \in \mathcal{Y}_t^{\neg\text{fnd}}$ of $t \in \mathcal{T}$ measurements for which the evader was not found,*

$$p_e(x, Y_t) \ge \gamma p_e(x, Y_{t-1}), \tag{12}$$

for any $x \in \mathcal{X}$ for which (i) x is not in the list of pursuers positions specified in the last measurement in Y_t and (ii) $\mathrm{P}_{\bar{g}}\big(\mathbf{m}(x) = 1 \mid \mathbf{Y}_t = Y_t\big) < 1$, for any pursuit policy $\bar{g}$. In (12), Y_{t-1} denotes the sequence consisting of the first $t-1$ elements in Y_t.

Assumption 1 basically demands that, in a single time step, the conditional probability of the evader being at a cell $x \in \mathcal{X}$, given the measurements taken up to that time, does not decay by more than a certain amount. That is, unless one pursuer reaches x—in which case the probability of the evader being at x may decay to zero if the evader is not there—or if it is possible to conclude from the measured data that an obstacle is at x with probability one. Such an assumption holds for most sensor models. The following can then be stated:

Theorem 2 *There exists an admissible pursuit policy that is persistent on the average with period $T := d + n_o(d-1)$, where n_o denotes the number of obstacles and d the maximum number of steps needed to travel from one cell to any other.*

The proof of Theorem 2 can be found in [9]. A specific admissible pursuit policy is also given in this reference.

5 Example

In this section we describe a specific pursuit-evasion game with partial observations and obstacles to which the greedy pursuit policies developed in Section 4 can be applied. In this game the pursuit takes place in a rectangular two-dimensional grid with n_c square cells numbered from 1 to n_c. We say that two distinct cells $x_1, x_2 \in \mathcal{X} := \{1, 2, \cdots, n_c\}$ are *adjacent* if they share one side or one corner. In the sequel we denote by $\mathcal{A}(x) \subset \mathcal{X}$ the set of cells adjacent to some cell $x \in \mathcal{X}$. Each $\mathcal{A}(x)$ will have, at most, 8 elements. The motion of the pursuers is constrained in that each pursuer can only remain in the same cell or move to a cell adjacent to its present position. This means that if at a time $t \in \mathcal{T}$ the pursuers are positioned in the cells $v := \{v_1, v_2, \dots, v_{n_p}\} \in \mathcal{U}$, then the subset of $\mathcal{U}$ consisting of those lists of cells to which the pursuers could move at time $t+1$, were these cells empty, is given by

$$\mathcal{U}(v) := \Big\{\{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_{n_p}\} \in \mathcal{U} : \bar{v}_i \in \{v_i\} \cup \mathcal{A}(v_i)\Big\}.$$

We assume a Markov model for the motion of the evader. The model is completely determined by a scalar parameter $\rho \in [0, 1/8]$ that represents the probability of the evader moving from its present position to an adjacent cell with no obstacles.

Each pursuer is capable of determining its current position and sensing the cells adjacent to the one it occupies for obstacles/evader. Each measurement $\mathbf{y}(t)$, $t \in \mathcal{T}$, therefore consists of a triple $\{\mathbf{v}(t), \mathbf{e}(t), \mathbf{o}(t)\}$ where $\mathbf{v}(t) \in \mathcal{U}$ denotes the measured positions of the pursuers, $\mathbf{e}(t) \subset \mathcal{X}$ a set of cells where an evader was detected, and $\mathbf{o}(t) \subset \mathcal{X}$ a set of cells where obstacles were detected. For this game we then have $\mathcal{Y} := \mathcal{U} \times 2^{\mathcal{X}} \times 2^{\mathcal{X}}$, where $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$, i.e., the set of all subsets of $\mathcal{X}$. For simplicity, we shall assume that $\mathbf{v}(t)$ reflect accurate measurements and therefore $\mathbf{v}(t) = \mathbf{x}(t)$, $t \in \mathcal{T}$. We also assume that the detection of the evader is perfect for the cells in which the pursuers are located, but not for adjacent ones. The sensor model for evader detection is a function of two parameters The probability $p \in [0, 1]$ of a pursuer detecting an evader in a cell adjacent to its current position, given that none was there, and the probability

$q \in [0, 1]$ of not detecting an evader, given that it was there. We call p the *probability of false positives* and q the *probability of false negatives*. These probabilities being nonzero reflect the fact that the sensors are not perfect. For simplicity we shall assume that the sensors used for obstacle detection is perfect in that $\mathbf{o}(t)$ contains precisely those cells adjacent to the pursuers that contain an obstacle. The reader is referred to [9] for a detailed description of how to compute the (conditional) posterior probability $p_e(x, Y_t)$ of the evader being in cell x at time $t+1$, given the measurements $\mathbf{Y}_t = Y_t$ taken up to time $t := |Y_t|$.

The above game is of the type described in Section 4 with constrained motion for the pursuers. It therefore admits the pursuit policy g_c described in [9] and whose existence is guaranteed by Theorem 2. Figure 1 shows a simulation of this pursuit-evasion game with $n_c := 400$ cells, $n_p := 3$ pursuers, $\rho = 5\%$, $p = q = 1\%$, and $p_m(x) = 10/400$, $x \in \mathcal{X}$. In Figure 1, the pursuers are represented by light stars, the evader is represented by a light circle, and the obstacles detected by the pursuers are represented by dark asterisks. The background color of each cell $x \in \mathcal{X}$ encodes $p_e(x|\mathbf{Y}_t)$, with a light color for low probability and a dark color for high probability. In some images (e.g., for $t = 7$) one can see very high values for $p_e(x|\mathbf{Y}_t)$ near one of the pursuers, even though the evader is far away. This is due to false positives given by the sensors.

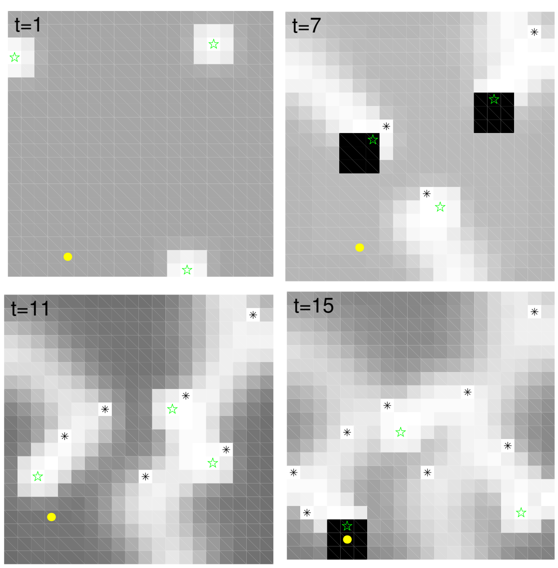

Figure 1: Pursuit with the constrained greedy policy.

6 Conclusion

In this paper we propose a probabilistic framework for pursuit-evasion games that avoids the conservativeness inherent to deterministic worst-case assumptions on the motion of the evader. A probabilistic framework is also natural to take into account the fact that sensor information is not precise and that only an inaccurate *a priori* map of the terrain may be known. We showed that greedy policies can be used to control a swarm of autonomous agents in the pursuit of one or several evaders. These policies guarantee that an evader is found in finite time and that the expected time needed to find the evader is also finite. Our current research involves the design of pursuit policies that are optimal in the sense that they minimize the expected time needed to find the evader or that they maximize the probability of finding the evader in a given finite time interval. We are also applying the results presented here to games in which the evader is actively avoiding detection.

References

[1] T. D. Parsons, "Pursuit-evasion in a graph," in *Theory and Application of Graphs* (Y. Alani and D. R. Lick, eds.), pp. 426–441, Berlin: Springer-Verlag, 1976.

[2] N. Megiddo, S. L. Hakimi, M. R. Garey, D. S. Johnson, and C. H. Papadimitriou, "The complexity of searching a graph," *Journal of the ACM*, vol. 35, pp. 18–44, Jan. 1988.

[3] A. S. Lapaugh, "Recontamination does not help to search a graph," *Journal of the ACM*, vol. 41, pp. 224–245, Apr. 1993.

[4] I. Suzuki and M. Yamashita, "Searching for a mobile intruder in a polygonal region," *SIAM J. Comput.*, vol. 21, pp. 863–888, Oct. 1992.

[5] S. M. Lavalle, D. Lin, L. J. Guibas, J.-C. Latombe, and R. Motwani, "Finding an unpredictable target in a workspace with obstacles," in *Proc. of IEEE Int. Conf. Robot. & Autom.*, IEEE, 1997.

[6] S. M. Lavalle and J. Hinrichsen, "Visibility-based pursuit-evasion: The case of curved environments." Submitted to the IEEE Int. Conf. Robot. & Autom., 1999.

[7] X. Deng, T. Kameda, and C. Papadimitriou, "How to learn an unknown environment I: The rectilinear case," *Journal of the ACM*, vol. 45, pp. 215–245, Mar. 1998.

[8] S. Thrun, W. Burgard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Machine Learning* and *Autonomous Robots* (joint issue), vol. 31, no. 5, pp. 1–25, 1998.

[9] J. P. Hespanha, H. J. Kim, and S. Sastry, "Multiple-agent probabilistic pursuit-evasion games," tech. rep., Dept. Electrical Eng. & Comp. Science, University of California at Berkeley, Mar. 1999.

[10] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.