Multiple-Agent Probabilistic Pursuit-Evasion Games*

João P. Hespanha Hyoun Jin Kim Shankar Sastry

Dept. Electrical Eng. & Comp. Science
University of California at Berkeley
275M Cory Hall #1770, Berkeley, CA 94720-1770
Phone: (510) 643-5190, Fax: (510) 642-1341
{hespanha,jin,sastry}@eecs.berkeley.edu

March 6, 1999

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 a *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.

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

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 conservativeness 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.

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 $\mathcal{F}$. We assume that the σ-algebra $\mathcal{F}$ is rich enough so that all the probabilities considered below are well defined. 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*, i.e., $\mathrm{P}(A|B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$. Bold face symbols are used to denote random variables. Following the usual abuse of notation, given a random variable $\boldsymbol{\xi} : \Omega \to E$ and some $A \subset E$ we write $\mathrm{P}(\boldsymbol{\xi} \in A)$ for $\mathrm{P}(\{\omega \in \Omega : \boldsymbol{\xi}(\omega) \in A\})$. Similar notation is used for conditional probabilities.

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 $\mathcal{X} \triangleq \{1, 2, \ldots, n_c\}$ 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, $\mathcal{T} \triangleq \{1, 2, \dots\}$.

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 $\tau \leq t$. 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 $\bar{\mathcal{T}} \triangleq \mathcal{T} \cup \{+\infty\}$. We denote by $F_{\bar{g}} : \bar{\mathcal{T}} \to [0, 1]$ its distribution function, i.e., $F_{\bar{g}}(t) \triangleq \mathrm{P}_{\bar{g}}(\mathbf{T}^* \leq t)$. Given any $t > 1$,

$$F_{\bar{g}}(t) = \mathrm{P}_{\bar{g}}(\mathbf{T}^* \leq t) = \mathrm{P}_{\bar{g}}(\mathbf{T}^* < t) + \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t). \tag{2}$$

Now, $\mathrm{P}_{\bar{g}}(\mathbf{T}^* = t)$ can be regarded as the probability of finding an evader at time t and not having found any up to that time, therefore

$$\mathrm{P}_{\bar{g}}(\mathbf{T}^* = t) = f_{\bar{g}}(t)\, \mathrm{P}_{\bar{g}}(\mathbf{T}^* \geq t) = f_{\bar{g}}(t)\big(1 - \mathrm{P}_{\bar{g}}(\mathbf{T}^* < t)\big), \tag{3}$$

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., $f_{\bar{g}}(t) \triangleq \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t \mid \mathbf{T}^* \geq t)$. From (2) and (3) we then conclude that

$$F_{\bar{g}}(t) = F_{\bar{g}}(t-1) + f_{\bar{g}}(t)\big(1 - F_{\bar{g}}(t-1)\big), \qquad t > 1.$$

The previous expression can also be written as

$$1 - F_{\bar{g}}(\tau) = \big(1 - f_{\bar{g}}(\tau)\big)\big(1 - F_{\bar{g}}(\tau - 1)\big), \qquad \tau > 1,$$

[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.

which can be iterated from $\tau = t_0 + 1 > 1$ to $\tau = t \geq t_0$ to conclude that[2]

$$1 - F_{\bar{g}}(t) = \big(1 - F_{\bar{g}}(t_0)\big) \prod_{\tau=t_0+1}^{t} \big(1 - f_{\bar{g}}(\tau)\big), \qquad t \geq t_0 \geq 1.$$

Since $F_{\bar{g}}(1) = f_{\bar{g}}(1)$, from the previous expression we also conclude that

$$1 - F_{\bar{g}}(t) = \prod_{\tau=1}^{t} \big(1 - f_{\bar{g}}(\tau)\big), \qquad t \geq 1, \tag{4}$$

and therefore

$$F_{\bar{g}}(t) = 1 - \prod_{\tau=1}^{t} \big(1 - f_{\bar{g}}(\tau)\big), \qquad t \in \mathcal{T}. \tag{5}$$

Suppose now that the probability of $\mathbf{T}^*$ being finite is equal to one and therefore that $\mathrm{P}_{\bar{g}}(\mathbf{T}^* = +\infty) = 0$. The expected value of $\mathbf{T}^*$ is then equal to

$$E_{\bar{g}}[\mathbf{T}^*] \triangleq \sum_{t=1}^{\infty} t\, \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t) = F_{\bar{g}}(1) + \sum_{t=2}^{\infty} t\big(F_{\bar{g}}(t) - F_{\bar{g}}(t-1)\big).$$

From this and (4) one obtains

$$\begin{aligned} E_{\bar{g}}[\mathbf{T}^*] &= f_{\bar{g}}(1) - \sum_{t=2}^{\infty} t \left(\prod_{\tau=1}^{t} \big(1 - f_{\bar{g}}(\tau)\big) - \prod_{\tau=1}^{t-1} \big(1 - f_{\bar{g}}(\tau)\big) \right) \\ &= f_{\bar{g}}(1) - \sum_{t=2}^{\infty} t \Big(\big(1 - f_{\bar{g}}(t)\big) - 1 \Big) \left(\prod_{\tau=1}^{t-1} \big(1 - f_{\bar{g}}(\tau)\big) \right), \end{aligned}$$

which can be simply written as

$$E_{\bar{g}}[\mathbf{T}^*] = \sum_{t=1}^{\infty} t f_{\bar{g}}(t) \left(\prod_{\tau=1}^{t-1} \big(1 - f_{\bar{g}}(\tau)\big) \right). \tag{6}$$

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 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 $\mathcal{Y}_\tau^{\neg\mathrm{fnd}} \subset \mathcal{Y}^*$, $\tau \in \mathcal{T}$, the set of all sequences of measurements of length τ, associated with an evader not being found up to that time, i.e., $\mathcal{Y}_\tau^{\neg\mathrm{fnd}} = \mathbf{Y}_\tau(\{\omega \in \Omega : \mathbf{T}^*(\omega) > \tau\})$. Since the decision to whether or not an evader was found up to time τ is purely

[2]In this paper we use the notation that, for every integer m and every sequence $\{a_k\}$, $\prod_{k=m}^{m-1} a_k \triangleq 1$.

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

$$f_{\bar{g}}(t) = \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t \mid \mathbf{T}^* \geq t) = \sum_{Y_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}} h_{\bar{g}}(Y_{t-1})\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t-1} = Y_{t-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}), \qquad t \in \mathcal{T}, \tag{7}$$

where $h_{\bar{g}} : \mathcal{Y}^* \to [0,1]$ is a function that maps each sequence $Y \in \mathcal{Y}^*$ of $\tau \triangleq |Y|$ 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., $h_{\bar{g}}(Y) \triangleq \mathrm{P}_{\bar{g}}(\mathbf{T}^* = |Y| + 1 \mid \mathbf{Y}_{|Y|} = Y)$. Equation (7) shows that $f_{\bar{g}}(t)$ is equal to the expected value of $h_{\bar{g}}(\mathbf{Y}_{t-1})$ when the probability measure is conditioned by $\mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}$. This can be written compactly as

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

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. In Sections 4 and 5 we show how to efficiently compute $h_{\bar{g}}(Y)$ from these models.

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

$$f_{\bar{g}}(t) \geq \epsilon, \qquad \forall t \in \mathcal{T}. \tag{8}$$

From (5) 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 (8). 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 (6) and Lemma 4 in the appendix). 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 was proved:

Lemma 1. *For a persistent pursuit policy* $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$, $\mathrm{P}_{\bar{g}}(\mathbf{T}^* < \infty) = 1$, $F_{\bar{g}}(t) \geq 1 - (1-\epsilon)^t$, $t \in \mathcal{T}$, *and* $E_{\bar{g}}[\mathbf{T}^*] \leq \epsilon^{-1}$, *with* ϵ *as in* (8).

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,

$$\bar{f}_{\bar{g}}(t) \triangleq \mathrm{P}_{\bar{g}}(\mathbf{T}^* \in \{t, t+1, \ldots, t+T-1\} \mid \mathbf{T}^* \geq t) \geq \epsilon, \qquad \forall t \in \mathcal{T}. \tag{9}$$

We call T the *period of persistence.* To analyze policies that are persistent on the average with period T we define

$$\bar{F}(k) \triangleq F_{\bar{g}}(kT), \qquad \bar{f}(k) \triangleq \bar{f}_{\bar{g}}\big((k-1)T + 1\big), \qquad k \in \{1, 2, \ldots\}.$$

This function could be interpreted as the distribution function of $\mathbf{T}^*$ for an auxiliary game that runs T times faster. Reasoning as in Section 2 we can then conclude that

$$\bar{F}(k) = 1 - \prod_{i=1}^{k} (1 - \bar{f}(i)), \qquad k \in \{1, 2, \dots\}.$$

Because the pursuit policy is persistent on the average, each $\bar{f}(i) \geq \epsilon$, which means that the auxiliary game running T times faster is persistent. Reasoning as in the proof of Lemma 1 with $\bar{F}$ and $\bar{f}$ playing the roles of $F_{\bar{g}}$ and $f_{\bar{g}}$, respectively, we then conclude that for every $k \in \{1, 2, \dots\}$,

$$F_{\bar{g}}(kT) \geq 1 - (1-\epsilon)^k, \qquad \sum_{k=1}^{\infty} k \, \mathrm{P}_{\bar{g}} \left(\mathbf{T}^* \in \{(k-1)T + 1, (k-1)T + 2, \dots, kT\} \right) \leq \epsilon^{-1}.$$

But then[3] $F_{\bar{g}}(t) \geq 1 - (1-\epsilon)^{\lfloor \frac{t}{T} \rfloor}$, $t \in \mathcal{T}$, and

$$\begin{aligned} E_{\bar{g}}[\mathbf{T}^*] &= \sum_{k=1}^{\infty} \sum_{i=1}^{T} \left((k-1)T + i \right) \mathrm{P}_{\bar{g}}(\mathbf{T}^* = (k-1)T + i) \\ &\leq T \sum_{k=1}^{\infty} k \sum_{i=1}^{T} \mathrm{P}_{\bar{g}}(\mathbf{T}^* = (k-1)T + i) \leq T\epsilon^{-1}. \end{aligned}$$

We proved the following:

Lemma 2. *For a persistent on the average pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$, 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 (9).*

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 we develop basic tools that can be used to show that a specific pursuit policy $\bar{g}$ is persistent, eventually only on the average. In particular, we determine (sufficient) conditions for persistency of $\bar{g}$ in terms of the conditional probabilities $h_{\bar{g}}$ of finding the evader, given the measurements. As we will see in Section 4, conditions in terms of $h_{\bar{g}}$ are often more convenient to verify than conditions in terms of $f_{\bar{g}}$. A sufficient condition for (8) to hold—and therefore for $\bar{g}$ to be persistent—is that

$$h_{\bar{g}}(Y) \geq \epsilon, \qquad \forall t \in \mathcal{T}, Y \in \mathcal{Y}_{t-1}^{\neg \mathrm{fnd}}. \tag{10}$$

This is a direct consequence of (7) and the fact that

$$\sum_{Y \in \mathcal{Y}_{t-1}^{\neg \mathrm{fnd}}} \mathrm{P}(\mathbf{Y}_{t-1} = Y \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg \mathrm{fnd}}) = 1, \qquad \forall t \in \mathcal{T}.$$

An analogous condition (proved in the Appendix) can be found for persistency on the average:

[3] Given a scalar a, we denote by $\lfloor a \rfloor$ the largest integer smaller or equal to a.

Lemma 3. *A sufficient condition for $\bar{g}$ to be persistent on the average with period T is the existence of some $\delta > 0$ such that, for each $t \in \mathcal{T}$ and each $Y \in \mathcal{Y}^{\neg\text{fnd}}_{t+T-2}$, there is some $\tau \in \{t-1, t, \dots, t+T-2\}$ for which $h_{\bar{g}}(Y_\tau) \geq \delta$, where Y_τ denotes the sequence consisting of the first τ measurements in Y. In this case* (9) *holds with*

$$\epsilon \triangleq \begin{cases} \frac{1}{T}\left(1-\frac{1}{T}\right)^{T-1} & \delta \geq \frac{1}{T} \\ \delta(1-\delta)^{T-1} & \delta < \frac{1}{T} \end{cases} \tag{11}$$

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 $\mathbf{x} \triangleq \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_p}\}$ 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 therefore 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., $p_m(x) = \mathrm{P}\left(\mathbf{m}(x) = 1\right)$, $x \in \mathcal{X}$. This probability is assumed independent of the pursuit policy.

At each time $t \in \mathcal{T}$, the control action $\mathbf{u}(t) \triangleq \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_p}\}$ 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

$$\mathcal{U} \triangleq \left\{\{v_1, v_2, \dots, v_{n_p}\} : v_i \in \mathcal{X},\ v_i \neq v_j \text{ for } i \neq j\right\}.$$

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 $t \triangleq |Y|$. We also assume that this probability is independent of the pursuit policy being used, i.e., for every specific pursuit policy $\bar{g} : \mathcal{Y}^* \to \mathcal{U}$,

$$p_e(x, Y) = \mathrm{P}_{\bar{g}}\left(\mathbf{x}_e(|Y|+1) = x \mid \mathbf{Y}_{|Y|} = Y\right), \qquad \forall x \in \mathcal{X},\ Y \in \mathcal{Y}^*. \tag{12}$$

In practice, this amounts to saying that the motion of the evader is independent of the motions of the pursuers. This would happen, for example, in games against nature. In Section 5 we show how the function p_e can be efficiently computed for which the motion of the evader follows a Markov model. In fact, we shall see that for every sequence $Y_t \in \mathcal{Y}^*$ of $t \triangleq |Y| \in \mathcal{T}$ 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 Markov game [9].

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.,

$$g_u(Y) = \arg \max_{\{v_1, v_2, \ldots, v_{n_p}\} \in \mathcal{U}} \sum_{k=1}^{n_p} p_e(v_k, Y), \qquad Y \in \mathcal{Y}^*. \tag{13}$$

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 $t \triangleq |Y|$ 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

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

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

$$h_{g_u}(Y) = \sum_{k=1}^{n_p} \mathrm{P}_{g_u}\big(\mathbf{x}_e(t+1) = \mathbf{u}_k(t+1) \mid \mathbf{Y}_t = Y\big). \tag{14}$$

Let now $\{v_1, v_2, \ldots, v_{n_p}\} \triangleq g_u(Y)$. Because of (1) we have $\mathbf{u}_k(t+1) = v_k$, $k \in \{1, 2, \ldots, n_p\}$, given that $\mathbf{Y}_t = Y$ and $g = g_u$. From this, (12), and (14) we conclude that

$$h_{g_u}(Y) = \sum_{k=1}^{n_p} p_e(v_k, Y).$$

Because of (13) and the fact that $\sum_{x=1}^{n_c} p_e(x, Y) = 1$, we then obtain

$$h_{g_u}(Y) = \sum_{k=1}^{n_p} p_e(v_k, Y) \geq \frac{n_p}{n_c} \sum_{x=1}^{n_c} p_e(x, Y) = \epsilon \triangleq \frac{n_p}{n_c}. \tag{15}$$

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 (15) one concludes that g_u is persistent (cf. (10)) 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. One could, for example, 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 $v \triangleq \{v_1, v_2, \ldots, v_{n_p}\} \in \mathcal{U}$, 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:

Assumption 1. For any $v_{\text{init}}, v_{\text{final}} \in \mathcal{U}$, there exists a finite sequence

$$\{v(0), v(1), \ldots, v(t) : v(0) = v_{\text{init}},\ v(t) = v_{\text{final}},\ t \in \mathcal{T}\} \in \mathcal{U}^*$$

such that each $v(\tau)$, $\tau \in \mathcal{T}$, is reachable from $v(\tau-1)$ in a single time step, i.e., $v(\tau) \in \mathcal{U}(v(\tau-1))$.

Under this assumptions it is always possible to construct a "navigation policy" that takes a desired list of final positions for the pursuers $v_{\text{final}} \in \mathcal{U}$, together with the measurements $\mathbf{Y}_t$ taken up to some time $t \in \mathcal{T}$, and either produces a position reachable in a single time step that is "one step closer" to v_{final} or concludes with probability one that the final position is not reachable in a single time step from anywhere in $\mathcal{U}$. The latter can only happen when there are obstacles precisely on the cells specified in v_{final}. Here we require such a policy to be prioritized in that only a particular position in $v_{\text{final}} \in \mathcal{U}$, for example the kth one, needs to be reached in finite time (provided there is no obstacle there). Such a navigation policy implicitly defines a "distance function" that measures how many time steps are needed for the kth pursuer to either reach its final position in v_{final} or conclude that this position is unreachable. Formally, a *navigation policy* is the a function $\text{nav} : \mathcal{U} \times \{1, 2, \ldots, n_p\} \times \mathcal{Y}^* \to \mathcal{U}$ for which there is a bounded *distance function* $\text{dist} : \mathcal{U} \times \{1, 2, \ldots, n_p\} \times \mathcal{Y}^* \to \mathbb{R}$ with the properties that, for each $\{v_{\text{final}}, k, Y_t\}$ in the domain of nav, the following is true:

1. $\text{nav}(v_{\text{final}}, k, Y_t) \in \mathcal{U}(v(t))$ whenever $\mathrm{P}_{\bar{g}}\left(\mathbf{m}(x_k) = 1 \mid \mathbf{Y}_t = Y_t\right) < 1$, for any pursuit policy $\bar{g}$;

2. $\text{dist}(v_{\text{final}}, k, Y) = 0$ whenever $x_k = v_k(t)$;

3. $\text{dist}(v_{\text{final}}, k, Y) \leq \text{dist}(v_{\text{final}}, k, Y_{t-1}) - 1$ whenever $v(t) = \text{nav}(v_{\text{final}}, k, Y_{t-1})$.

In the above, $v(t)$ denotes the positions of the pursuers specified in the last element of Y_t, Y_{t-1} the sequence consisting of the first $t-1$ elements in Y, and $v_k(t)$ and x_k the kth elements of $v(t)$ and v_{final}, respectively.

Property 1 guarantees that the navigation policy always tries to move the pursuers to positions reachable in a single time step (unless this is impossible for the kth pursuers); Property 2 states that the distance to the cell currently occupied by the kth pursuer must be zero; and Property 3 requires that each move produced by the navigation policy will get the pursuers one step closer to the goal, which is to make the kth pursuer reach its final position. For simple arrangements of the obstacles, navigation policies can be constructed along the lines of the "bug" algorithms in [10]. Since the focus of this paper is not on the navigation problem we take as given a navigation policy nav together with its distance function dist.

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 it to design a pursuit policy that is persistent on the average. To this effect we need the following assumption:

Assumption 2. There is a positive constant $\gamma \leq 1$ 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,

$$p_e(x, Y_t) \geq \gamma p_e(x, Y_{t-1}), \tag{16}$$

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) $\mathrm{P}_{\bar{g}}\left(\mathbf{m}(x) = 1 \mid \mathbf{Y}_t = Y_t\right) < 1$, for any pursuit policy $\bar{g}$. In (16), Y_{t-1} denotes the sequence consisting of the first $t-1$ elements in Y_t.

Assumption 2 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.

Theorem 2. *Suppose Assumptions 1 and 2 hold. There exists an admissible pursuit policy* $g_c : \mathcal{Y}^* \to \mathcal{U}$ *that is persistent on the average with period* $T \triangleq d + n_o(d-1)$*, where* n_o *denotes the number of obstacles and* d *a (non strict) upper bound for the distance function* dist*. Moreover,* $\mathrm{P}_{g_c}(\mathbf{T}^* < \infty) = 1$, $F_{g_c}(t) \geq 1 - (1-\epsilon)^{\lfloor \frac{t}{T} \rfloor}$, $t \in \mathcal{T}$, *and* $E_{g_c}[\mathbf{T}^*] \leq T\epsilon^{-1}$*, with* ϵ *given by* (11) *and* $\delta \triangleq \frac{\gamma^{d-1}}{n_c}$.

Before proving Theorem 2 it should be emphasized that the upper bounds given here for F_{g_c} and $E_{g_c}[\mathbf{T}^*]$ may be very conservative if the parameters γ and d on which they are based are also very conservative. This happens often because γ and d correspond to worst-case bounds.

Proof of Theorem 2. We prove this lemma by constructing an admissible pursuit policy $g_c : \mathcal{Y}^* \to \mathcal{U}$ that is persistent on the average with period $T \triangleq d + n_o(d-1)$. To this effect take an arbitrary sequence of measurements

$Y \in \mathcal{Y}^*$ and let $\mathcal{R}(Y,m) \subset \mathcal{X}$ denote the set of cells that can be reached in m steps by one of the pursuer, when they start at the positions specified in the last measurement in Y, i.e.,

$$\mathcal{R}(Y,m) \triangleq \Big\{x \in \mathcal{X} : \exists v \in \mathcal{U} \text{ such that } x \text{ is an element of } v \text{ and } \min_k \operatorname{dist}(v,k,Y) \le m\Big\}.$$

Since dist is bounded by d, $\mathcal{R}(Y,d) = \mathcal{X}$. Let now $m(Y)$ be the smallest positive integer for which

$$\exists x^* \in \mathcal{R}\big(Y, m(Y)\big) \quad : \quad p_e(x^*, Y) \ge \frac{\gamma^{d-m(Y)}}{n_c}. \tag{17}$$

The integer $m(Y)$ will be, at most, as large as d. This is because (17) will always hold when $m(Y) = d$, since at least one cell $x \in \mathcal{X} = \mathcal{R}\big(Y,d\big)$ must have $p(x,Y) \ge 1/n_c$. $m(Y)$ corresponds to the smallest number of steps needed to reach a cell x^* for which $p_e(x^*,Y) \ge \frac{\gamma^{d-m(Y)}}{n_c}$. Let us now define $x_{\text{final}}(Y)$ to be the cell with highest probability of having an evader that can be reached in $m(y)$ steps, i.e., $x_{\text{final}}(Y) \triangleq \arg\max_{x \in \mathcal{R}\left(Y, m(Y)\right)} p_e(x,Y)$. Because of the way $m(Y)$ was defined

$$p_e(x_{\text{final}}(Y), Y) \ge \frac{\gamma^{d-m(Y)}}{n_c}. \tag{18}$$

To specify g_c we need to consider the list of cells $v_{\text{final}}(Y) \in \mathcal{U}$ that are reachable in $m(Y)$ steps and contain $x_{\text{final}}(Y)$, for which the probability of finding an evader by the group of pursuers is maximal. Formally,

$$v_{\text{final}}(Y) = \{v_1, v_2, \ldots, v_{n_p}\} \triangleq \arg \max_{\{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_{n_p}\} \in \mathcal{N}} \sum_{k=1}^{n_p} p_e(\bar{v}_k, Y). \tag{19}$$

where $\mathcal{N} \triangleq \big\{v \in \mathcal{U} : x_{\text{final}}(Y) \text{ is an element of } v \text{ and } \min_k \operatorname{dist}(v,k,Y) \le m(Y)\big\}$. We shall then define $g_c(Y)$ as follows

$$g_c(Y) \triangleq \operatorname{nav}\big(v_{\text{final}}(Y), k_{\text{final}}(Y), Y\big), \tag{20}$$

where $k_{\text{final}}(Y) \in \{1, 2, \ldots, n_p\}$ is the integer for which $x_{\text{final}}(Y) = v_{k_{\text{final}}(Y)}$. Because of (18) there must be a non zero probability of existing an evader at position $x_{\text{final}}(Y)$. Therefore, the probability of existing an obstacle at $x_{\text{final}}(Y)$ must be smaller than one. From this and Property 1 of the navigation policy, we conclude that the pursuit policy g_c must be admissible.

It remains to prove that g_c is persistent on the average with period $T \triangleq d + n_o(d-1)$. This is done by showing that the sufficient condition in Lemma 3 holds. In particular, we will show that given any sequence $Y \in \mathcal{Y}^{\neg\text{fnd}}_{t+T-2}$ of $t+T-2$, $t \in \mathcal{T}$, measurements for which the evader was not found, and compatible with the pursuit policy g_c, there is some $\tau \in \{t-1, t, \ldots, t+T-2\}$ for which $h_{g_c}(Y_\tau) \ge \delta \triangleq \frac{\gamma^{d-1}}{n_c}$. Here, $Y_\tau \in \mathcal{Y}^{\neg\text{fnd}}_\tau$ denotes the sequence consisting of the first τ elements of Y. To this effect pick some $s \in \{t-1, t, \ldots, t+T-3\}$ for which $m(Y_s) > 1$ and

$$\mathrm{P}_{g_c}(\mathbf{m}(x_{\text{final}}(Y_s)) = 1 \mid \mathbf{Y}_{s+1} = Y_{s+1}) < 1. \tag{21}$$

Since $m(Y_s) > 1$,

$$p_e(x, Y_s) < \frac{\gamma^{d-1}}{n_c}, \qquad \forall x \in \mathcal{R}\big(Y_s, 1\big). \tag{22}$$

Because of (18) and the fact that $\gamma \leq 1$, we also have that

$$p_e(x_{\text{final}}(Y_s), Y_s) \geq \frac{\gamma^{d-m(Y_s)}}{n_c} \geq \frac{\gamma^{d-1}}{n_c}.$$

From this and (22) we conclude that $x_{\text{final}}(Y_s) \notin \mathcal{R}(Y_s, 1)$. This means that no pursuer can be at cell $x_{\text{final}}(Y_s)$ at time $s+1$. Because of this and (21), by Assumption 2, we must have

$$p_e(x_{\text{final}}(Y_s), Y_{s+1}) \geq \gamma p_e(x_{\text{final}}(Y_s), Y_s) \geq \frac{\gamma^{d-(m(Y_s)-1)}}{n_c}. \tag{23}$$

Here, we also used (18). Since the pursuit policy g_c is being used to move the pursuers, at time $s+1$ the cell $x_{\text{final}}(Y_s)$ is within reach of one of the pursuers in $m(Y_s)-1$ steps, therefore $x_{\text{final}}(Y_s) \in \mathcal{R}(Y_{s+1}, m(Y_s)-1)$. This, together with (17) and (23) imply that $m(Y_{s+1}) \leq m(Y_s) - 1$.

We have just shown that for any $s \in \{t-1, t, \ldots, t+T-3\}$ for which $m(Y_s) > 1$ and (21) holds, $m(Y_{s+1}) \leq m(Y_s) - 1$. Now, $\mathrm{P}_{g_c}(\mathbf{m}(x_{\text{final}}(Y_s)) = 1 \mid \mathbf{Y}_s = Y_s) < 1$, because otherwise an obstacle was known to be at position $x_{\text{final}}(Y_s)$ with probability one, given the measurements $\mathbf{Y}_s = Y_s$, and therefore $p_e(x_{\text{final}}(Y_s), Y_s) = 0$, which would contradict (18). This means that (21) can only be violated at the precise times s at which a new obstacle is found with probability one. Since there are at most n_o obstacles we conclude that (21) can be violated at most n_o times in $\{t-1, t, \ldots, t+T-2\}$. We therefore conclude that m will always decrease, as long as it is larger than 1, except at a finite collection of, at most, n_o times. Since $m \leq d$, at least for one $\tau \in \{t-1, t, \ldots, t+T-2\}$, with $T \triangleq d + n_o(d-1)$, we must then have $m(Y_\tau) = 1$. Consider such a time τ and let $\{v_1, v_2, \ldots, v_{n_p}\} \triangleq g_c(Y_\tau)$. Because of (1) we have $\mathbf{u}_k(\tau+1) = v_k$, $k \in \{1, 2, \ldots, n_p\}$, given that $\mathbf{Y}_\tau = Y_\tau$ and $\mathbf{g} = g_c$. Therefore the conditional probability of finding the evader for the first time at $\tau + 1$, given the measurements $\mathbf{Y}_\tau = Y_\tau$ taken up to τ, is given by

$$h_{g_c}(Y_\tau) = \mathrm{P}_{g_c}\left(\mathbf{x}_e(\tau+1) = \mathbf{u}_k(\tau+1),\ \exists k \in \{1, 2, \ldots, n_p\} \mid \mathbf{Y}_\tau = Y_\tau\right) = \sum_{k=1}^{n_p} p_e(v_k, Y_\tau). \tag{24}$$

But since $m(Y_\tau) = 1$, $\mathrm{dist}(v_{\text{final}}(Y_\tau), k_{\text{final}}(Y_\tau), Y) \leq 1$ because of (19). This means that the k_{final}-th pursuer can reach x_{final} in a single step and therefore $v_{k_{\text{final}}} = x_{\text{final}}$. From (24) and (18) one then concludes that

$$h_{g_c}(Y_\tau) \geq p_e(x_{\text{final}}(Y_\tau), Y_\tau) \geq \delta \triangleq \frac{\gamma^{d-1}}{n_c} > 0.$$

A straightforward application of Lemmas 2 and 3 finishes the proof. ∎

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 $x_1, x_2 \in \mathcal{X} \triangleq \{1, 2, \cdots, n_c\}$ are *adjacent* if they share one side or one corner (cf. Figure 1). 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

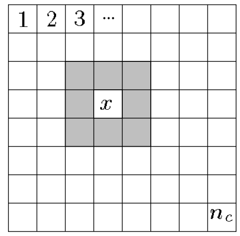

Figure 1: Pursuit region. The shaded cells are those adjacent to x.

$v \triangleq \{v_1, v_2, \ldots, v_{n_p}\} \in \mathcal{U}$, 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

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

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. This probability is independent of the specific pursuit policy being used. This means that, for each $x, \bar{x} \in \mathcal{X}$, $n \in \{0, 1, \ldots, 8\}$,

$$\mathrm{P}(\mathbf{x}_e(t+1) = x \mid \mathbf{x}_e(t) = \bar{x},\ \mathbf{m}(x) = 0,\ \mathbf{n}_0(\bar{x}) = n) = \begin{cases} \rho & x \in \mathcal{A}(\bar{x}) \text{ and } \mathbf{m}(x) = 0 \\ 1 - \big(|\mathcal{A}(\bar{x})| - n\big)\rho & x = \bar{x} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

where $|\mathcal{A}(x)| \in \{3, 5, 8\}$ denotes the number of cells adjacent to x and $\mathbf{n}_o(\bar{x}) \in \{0, 1, \ldots, 8\}$ the number of obstacles in $\mathcal{A}(\bar{x})$. This models a situation in which the moving evader is not actively avoiding detection.

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 $\mathcal{Y} \triangleq \mathcal{U} \times 2^{\mathcal{X}} \times 2^{\mathcal{X}}$, 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.

We describe next 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 $t \triangleq |Y_t|$. The derivations below are independent of the specific pursuit policy $\bar{g}$ in use and we therefore drop the subscript $\bar{g}$ in the probability measure. For computational efficiency $p_e(x, Y_t)$ is computed recursively in two steps:

1. A *measurement step* in which the probability $\mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_t = Y_t\right)$ of the evader being in cell x at time t, given the measurements $\mathbf{Y}_t = Y_t$ taken up to t, is computed based on the probability $p_e(x, Y_{t-1}) \triangleq \mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_{t-1} = Y_{t-1}\right)$ of the evader being in cell x at time t, given the measurements $\mathbf{Y}_{t-1} = Y_{t-1}$ taken up to $t-1$, and the last measurement $y(t)$ in the sequence Y_t. The sensor model is used in this step.

2. A *prediction step* in which the probability $p_e(x, Y_t) \triangleq \mathrm{P}\left(\mathbf{x}_e(t+1) = x \mid \mathbf{Y}_t = Y_t\right)$ of the evader being in cell x at time $t+1$, given the measurements $\mathbf{Y}_t = Y_t$ taken up to time t, is then computed from $\mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_t = Y_t\right)$. The evader's motion model is used in this step.

This recursion is initialized with some *a priori* probability $p_e(x, \emptyset)$, $x \in \mathcal{X}$, for the position of the evader. Here, $\emptyset \in \mathcal{Y}^*$ denotes the empty sequence. To simplify the computations we shall assume that the obstacle density is sufficiently low so that the position of the evader at any given time is approximately independent of the positions of the obstacles.

Measurement step. For the measurement step we use Bayes' rule to write

$$\mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_t = Y_t\right) = \alpha_1 p_e(x, Y_{t-1})\, \mathrm{P}\left(\mathbf{y}(t) = y(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{Y}_{t-1} = Y_{t-1}\right), \tag{26}$$

where $\alpha_1 \triangleq 1/\,\mathrm{P}\left(\mathbf{y}(t) = y(t) \mid \mathbf{Y}_{t-1} = Y_{t-1}\right)$ is a positive normalizing constant independent of x. The last term in (26) must be computed using the sensor model. Partitioning $y(t) \triangleq \{v(t), e(t), o(t)\} \in \mathcal{Y}$, this term can be expanded as

$$\begin{aligned}
&\mathrm{P}\left(\mathbf{v}(t) = v(t),\ \mathbf{e}(t) = e(t),\ \mathbf{o}(t) = o(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{Y}_{t-1} = Y_{t-1}\right) \\
&\qquad = \mathrm{P}\left(\mathbf{e}(t) = e(t),\ \mathbf{o}(t) = o(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right)
\end{aligned}$$

Here we used the fact that $\mathrm{P}\left(\mathbf{v}(t) = v(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{Y}_{t-1} = Y_{t-1}\right) = 1$. This is because, since $\mathbf{v}(t)$ is equal to the position of the pursuers at time t, it is completely determined by Y_{t-1} and the pursuit policy. Therefore, if Y_t is compatible with a given pursuit policy $\bar{g}$, $\mathbf{v}(t) = v(t)$ must have probability one when conditioned to $\mathbf{Y}_{t-1} = Y_{t-1}$ and $\mathbf{g} = \bar{g}$. We therefore have

$$\begin{aligned}
\mathrm{P}&\left(\mathbf{y}(t) = y(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{Y}_{t-1} = Y_{t-1}\right) \\
&= \mathrm{P}\left(\mathbf{e}(t) = e(t),\ \mathbf{o}(t) = o(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right) \\
&= \mathrm{P}\left(\mathbf{e}(t) = e(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right) \\
&\qquad \mathrm{P}\left(\mathbf{o}(t) = o(t) \mid \mathbf{e}(t) = e(t),\ \mathbf{x}_e(t) = x,\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right).
\end{aligned} \tag{27}$$

Due to the low obstacle density assumption, $\mathbf{o}(t)$ is approximately independent of the position of the evader $\mathbf{x}_e(t)$, except for the constraint that no obstacle can be detected in the cell $\mathbf{x}_e(t)$ where the evader lies. In this case, from (26) and (27) we conclude that

$$\mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_t = Y_t\right) \approx \begin{cases} 0 & x \in o(t) \\ \alpha p_e(x, Y_{t-1})\, \mathrm{P}\left(\mathbf{e}(t) = e(t) \mid \mathbf{x}_e(t) = x,\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right) & x \notin o(t) \end{cases}$$

where $\alpha \triangleq \alpha_1\, \mathrm{P}\left(\mathbf{o}(t) = o(t) \mid \mathbf{e}(t) = e(t),\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right)$ is another normalizing constant independent of x. Now, according to our sensor model, the probability of obtaining a measurement $\mathbf{e}(t) = e(t)$ of where the evader was detected is purely a function of the real position of the evader $\mathbf{x}_e(t)$ and the positions of the pursuers $\mathbf{v}(t)$. Therefore[4]

$$\begin{aligned} &\mathrm{P}\left(\mathbf{e}(t)=e(t) \mid \mathbf{x}_e(t)=x,\ \mathbf{v}(t)=v(t),\ \mathbf{Y}_{t-1}=Y_{t-1}\right) \\ &\qquad = \begin{cases} 0 & x \in v(t) \setminus e(t) \text{ or } \exists \bar{x} \neq x : \bar{x} \in v(t) \cap e(t) \\ p^{k_1}(1-p)^{k_2} q^{k_3}(1-q)^{k_4} & \text{otherwise} \end{cases} \end{aligned} \tag{28}$$

where k_1 is the number of pursuers that reported in $e(t)$ seeing the evader at cells other than x that are adjacent to their one (false positives); k_2 is the number of pursuers that reported in $e(t)$ not seeing the evader at cells other than x that are adjacent to their one (true negatives); k_3 is the number of pursuers that reported in $e(t)$ not seeing the evader at the cell x adjacent to their one (false negatives); and k_4 is the number of pursuers that reported in $e(t)$ seeing the evader at the cell x adjacent to their one (true positives). $k_1+k_2+k_3+k_4$ must be equal to the number of cells adjacent to any of the pursuers positions in $v(t)$. The measurement step can then be written as

$$\begin{aligned} &\mathrm{P}\left(\mathbf{x}_e(t)=x \mid \mathbf{Y}_t=Y_t\right) \\ &\qquad \approx \alpha p_e(x, Y_{t-1}) \begin{cases} 0 & x \in o(t) \cup v(t) \setminus e(t) \text{ or } \exists \bar{x} \neq x : \bar{x} \in v(t) \cap e(t) \\ p^{k_1}(1-p)^{k_2} q^{k_3}(1-q)^{k_4} & \text{otherwise} \end{cases} \end{aligned} \tag{29}$$

where α is a normalizing constant, chosen so that $\sum_{x\in\mathcal{X}} \mathrm{P}\left(\mathbf{x}_e(t)=x \mid \mathbf{Y}_t=Y_t\right)=1$.

Prediction step. For the prediction step we expand

$$\begin{aligned} p_e(x, Y_t) &= \sum_{\bar{x}\in\{x\}\cup\mathcal{A}(x)} \mathrm{P}\left(\mathbf{x}_e(t+1)=x,\ \mathbf{x}_e(t)=\bar{x},\ \mathbf{m}(x)=0 \mid \mathbf{Y}_t=Y_t\right) \\ &\approx \mathrm{P}\left(\mathbf{x}_e(t+1)=x \mid \mathbf{x}_e(t)=x,\ \mathbf{Y}_t=Y_t\right) \mathrm{P}\left(\mathbf{x}_e(t)=x \mid \mathbf{Y}_t=Y_t\right) \\ &\quad + \sum_{\bar{x}\in\mathcal{A}(x)} \mathrm{P}\left(\mathbf{x}_e(t+1)=x \mid \mathbf{x}_e(t)=\bar{x},\ \mathbf{m}(x)=0,\ \mathbf{Y}_t=Y_t\right) \\ &\qquad\qquad \mathrm{P}\left(\mathbf{x}_e(t)=\bar{x} \mid \mathbf{Y}_t=Y_t\right) \mathrm{P}\left(\mathbf{m}(x)=0 \mid \mathbf{Y}_t=Y_t\right) \end{aligned}$$

Here, we used the fact that $\mathbf{x}_e(t)=x$ automatically implies that $\mathbf{m}(x)=0$, and also the low obstacle density assumption to conclude that $\mathbf{m}(x)$ is approximately independent of the position $\mathbf{x}_e(t+1)=\bar{x}$ of the evader, when $x \neq \bar{x}$. From (25) we conclude that

$$\mathrm{P}\left(\mathbf{x}_e(t+1)=x \mid \mathbf{x}_e(t)=\bar{x},\ \mathbf{m}(x)=0,\ \mathbf{Y}_t=Y_t\right)=\rho, \qquad \bar{x}\in\mathcal{A}(x)$$

and

$$\mathrm{P}\left(\mathbf{x}_e(t+1)=x \mid \mathbf{x}_e(t)=x,\ \mathbf{Y}_t=Y_t\right)=\sum_{n=0}^{8}\left(1-(|\mathcal{A}(x)|-n)\rho\right) \mathrm{P}\left(\mathbf{n}_o(\bar{x})=n \mid \mathbf{x}_e(t)=x,\ \mathbf{Y}_t=Y_t\right),$$

where $|\mathcal{A}(x)| \in \{3,5,8\}$ denotes the number of cells adjacent to x and $\mathbf{n}_o(\bar{x}) \in \{0,1,\ldots,8\}$ the number of obstacles in $\mathcal{A}(x)$. Because of the low obstacle density assumption, the probability of $\mathbf{n}_o(\bar{x})$ begin equal to zero is much larger than that of being larger than 0 and therefore, to simplify the computations we assume that $\mathrm{P}\left(\mathbf{n}_o(\bar{x})=0 \mid \mathbf{x}_e(t)=x,\ \mathbf{Y}_t=Y_t\right)\approx 1$. In this case,

$$\begin{aligned} p_e(x, Y_t) &\approx (1-|\mathcal{A}(x)|\rho)\, \mathrm{P}\left(\mathbf{x}_e(t)=x \mid \mathbf{Y}_t=Y_t\right) \\ &\quad + \rho\, \mathrm{P}\left(\mathbf{m}(x)=0 \mid \mathbf{Y}_t=Y_t\right) \sum_{\bar{x}\in\mathcal{A}(x)} \mathrm{P}\left(\mathbf{x}_e(t)=\bar{x} \mid \mathbf{Y}_t=Y_t\right). \end{aligned} \tag{30}$$

[4]With a slight abuse of notation, here we regard the list of pursuer positions $v(t)$ as a set.

To conclude the prediction step, it remains to show how to compute $\mathrm{P}\left(\mathbf{m}(x) = 0 \mid \mathbf{Y}_t = Y_t\right)$, $x \in \mathcal{X}$. It turns out that the computation of these conditional probabilities can be done in a recursive fashion, similar to the computation of the $p_e(x, Y_t)$. Actually, it is simpler because the obstacles do not move and therefore the prediction step is not needed. In fact, since we are assuming that the sensor used for obstacle detection is perfect, we simply have

$$\mathrm{P}\left(\mathbf{m}(x) = 0 \mid \mathbf{Y}_t = Y_t\right) = \begin{cases} 1 & x \notin o(t) \text{ and is in or adjacent to an element in } v(t) \\ 0 & x \in o(t) \text{ and is in or adjacent to an element in } v(t) \\ \mathrm{P}\left(\mathbf{m}(x) = 0 \mid \mathbf{Y}_{t-1} = Y_{t-1}\right) & \text{otherwise} \end{cases}$$

The recursion is initialized with the *a priori* obstacle map $p_m : \mathcal{X} \to [0, 1]$.

With (29) and (30) at hand it is straightforward to show that Assumption 2 holds. In fact, from (29) and (30) one concludes 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, and for any $x \in \mathcal{X}$ for which (i) x is not in the list of pursuers positions $v(t)$ specified in the last measurement in Y_t and (ii) $\mathrm{P}\left(\mathbf{m}(x) = 1 \mid \mathbf{Y}_t = Y_t\right) < 1$,

$$p_e(x, Y_t) \geq (1 - |\mathcal{A}(x)|\rho)\, \mathrm{P}\left(\mathbf{x}_e(t) = x \mid \mathbf{Y}_t = Y_t\right) \geq \alpha (1 - |\mathcal{A}(x)|\rho) p^{k_1} (1-p)^{k_2} q^{k_3} (1-q)^{k_4} p_e(x, Y_{t-1}),$$

with

$$\alpha \triangleq \frac{\mathrm{P}\left(\mathbf{o}(t) = o(t) \mid \mathbf{e}(t) = e(t),\ \mathbf{v}(t) = v(t),\ \mathbf{Y}_{t-1} = Y_{t-1}\right)}{\mathrm{P}\left(\mathbf{y}(t) = y(t) \mid \mathbf{Y}_{t-1} = Y_{t-1}\right)} = \frac{1}{\mathrm{P}\left(\mathbf{e}(t) = e(t),\ \mathbf{v}(t) = v(t) \mid \mathbf{Y}_{t-1} = Y_{t-1}\right)} \geq 1.$$

Here we used the fact that x must not be in $o(t)$, otherwise $\mathrm{P}(\mathbf{m}(x) = 1 \mid \mathbf{Y}_t = Y_t) = 1$. Since $k_1 + k_2 + k_3 + k_4$ must be equal to the number of cells adjacent to any of the pursuers, we must have $k_1 + k_2 + k_3 + k_4 \leq 8 n_p$. Therefore

$$p^{k_1} (1-p)^{k_2} q^{k_3} (1-q)^{k_4} \geq \min\{p, 1-p, q, 1-q\}^{8 n_p},$$

and we conclude that Assumption 2 holds with $\gamma \triangleq (1 - 8\rho) \min\{p, 1-p, q, 1-q\}^{8 n_p} > 0$, provided that $p, q \in (0, 1)$.

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 Section 4.2. Figure 2 shows a simulation of this pursuit-evasion game with $n_c \triangleq 400$ cells, $n_p \triangleq 3$ pursuers, $\rho = 5\%$, $p = q = 1\%$, and $p_m(x) = 10/400$, $x \in \mathcal{X}$. The navigation policy used is directly inspired in the Bug2 Algorithm in [10]. In Figure 2, 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 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.

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

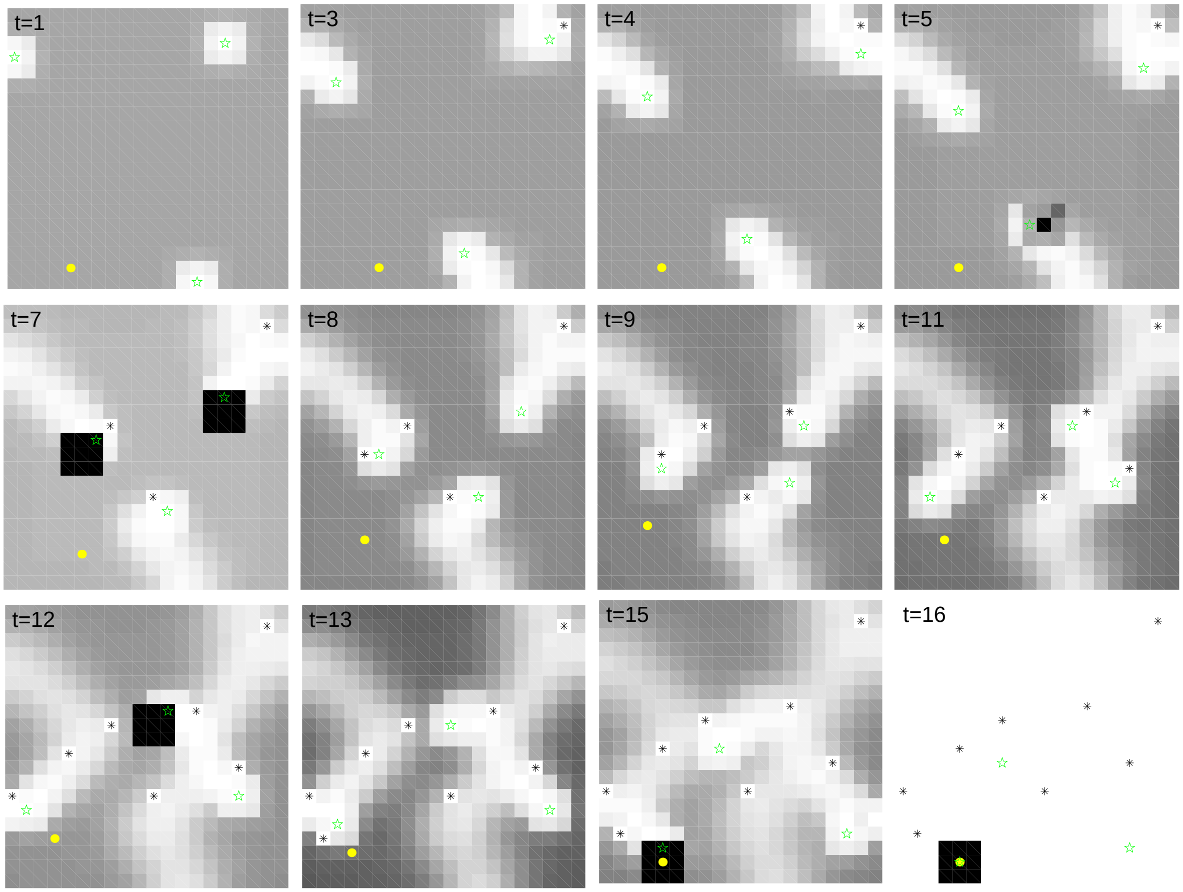

Figure 2: Pursuit using the constrained greedy pursuit policy. The pursuers are represented by light stars, the evader is represented by a light circle, and the obstacles detected by the pursuers are represented by dark asterisks. The background color of each cell $x \in \mathcal{X}$ encodes $p_e(x|\mathbf{Y}_t)$, with a light color for a low probability and a dark color for high probability.

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 extending the results presented here to games in which the evader is actively avoiding detection.

A Technical Lemmas

Lemma 4. *For a given set of discrete times* $\mathcal{T} \triangleq \{t_0, t_1 \dots\}$, *with* $t_{k+1} \geq t_k$, $k \geq 0$, *let* $f : [0,1]^* \to [0,1]$ *be defined by* $\{u_0, u_1, \dots\} \longmapsto \sum_{n=1}^{\infty} t_n u_n \left(\prod_{k=0}^{n-1} (1-u_k)\right)$. *Then, for every integer* m *for which* $\prod_{k=0, k\neq m}^{\infty}(1-u_k) = 0$,

$$\frac{\partial f}{\partial u_m} = -\sum_{n=m+1}^{\infty} (t_n - t_m) u_n \left(\prod_{k=0, k\neq m}^{n-1} (1-u_k)\right) \leq 0. \tag{31}$$

Proof of Lemma 4. Let m be an integer for which $\prod_{k=0,k\neq m}^{\infty}(1-u_k) = 0$. Taking the partial derivative of f with respect to u_m one obtains

$$\begin{aligned}\frac{\partial f}{\partial u_m} &= t_m \Big(\prod_{k=0}^{m-1} (1-u_k) \Big) - \sum_{n=m+1}^{\infty} t_n u_n \Big(\prod_{k=0,\, k\neq m}^{n-1} (1-u_k) \Big) \\ &= \Big(\prod_{k=0}^{m-1} (1-u_k) \Big) \Big(t_m - \sum_{n=m+1}^{\infty} t_n u_n \Big(\prod_{k=m+1}^{n-1} (1-u_k) \Big) \Big).\end{aligned}$$

But it is straightforward to show that $\sum_{n=m+1}^{\infty} u_n \Big(\prod_{k=m+1}^{n-1} (1-u_k) \Big) = 1 - \prod_{k=m+1}^{\infty}(1-u_k)$, therefore

$$\begin{aligned}\frac{\partial f}{\partial u_m} &= \Big(\prod_{k=0}^{m-1} (1-u_k) \Big) \Big(t_m \prod_{k=m+1}^{\infty} (1-u_k) - \sum_{n=m+1}^{\infty} (t_n - t_m) u_n \Big(\prod_{k=m+1}^{n-1} (1-u_k) \Big) \Big), \\ &= - \sum_{n=m+1}^{\infty} (t_n - t_m) u_n \Big(\prod_{k=0, k\neq m}^{n-1} (1-u_k) \Big) + t_m \prod_{k=0, k\neq m}^{\infty} (1-u_k),\end{aligned}$$

from which (31) follows. ∎

Proof of Lemma 3. For a given $t \in \mathcal{T}$, $\bar{f}_{\bar{g}}(t)$ can be written as

$$\bar{f}_{\bar{g}}(t) = \sum_{k=0}^{T-1} \mathrm{P}_{\bar{g}}(\mathbf{T}^* = t+k \mid \mathbf{T}^* \geq t) = \sum_{k=0}^{T-1} \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+k} \notin \mathcal{Y}_{t+k}^{\neg\mathrm{fnd}},\ \mathbf{Y}_{t+k-1} \in \mathcal{Y}_{t+k-1}^{\neg\mathrm{fnd}} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}}).$$

Pick now some $\bar{\delta} \in [0, \delta]$. A lower bound for $\bar{f}_{\bar{g}}(t)$ can then be constructed constraining the events on the probability measure as follows:

$$\begin{aligned}\bar{f}_{\bar{g}}(t) \geq \sum_{k=0}^{T-1} \mathrm{P}_{\bar{g}}(&\mathbf{Y}_{t+k} \notin \mathcal{Y}_{t+k}^{\neg\mathrm{fnd}},\ \mathbf{Y}_{t+k-1} \in \mathcal{Y}_{t+k-1}^{\neg\mathrm{fnd}},\ h(\mathbf{Y}_{t+k-1}) \geq \bar{\delta}, \\ & h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}}). \qquad (32)\end{aligned}$$

For each $n \in \{0, 1, \dots, T-1\}$, let us now define[5]

$$\begin{aligned}u_n \triangleq \sum_{k=0}^{T-n-2} &\mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+k-1} \in \mathcal{Y}_{t+k-1}^{\neg\mathrm{fnd}},\ h(\mathbf{Y}_{t+k-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}}) \\ &+ \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+T-n-2} \in \mathcal{Y}_{t+T-n-2}^{\neg\mathrm{fnd}},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+T-n-3 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}}).\end{aligned}$$

We show next that $\bar{f}_{\bar{g}}(t) \geq \bar{\delta} u_0$. To this effect note that for any $\tau \in \{t, t+1, \dots, t+T-1\}$,

$$\begin{aligned}&\mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau \notin \mathcal{Y}_\tau^{\neg\mathrm{fnd}},\ \mathbf{Y}_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\mathrm{fnd}},\ h(\mathbf{Y}_{\tau-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}}) \\ &= \sum_{\substack{Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\mathrm{fnd}} \\ h(Y_{\tau-1}) \geq \bar{\delta} \\ h(Y_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-2}} \Big(\sum_{\substack{y \in \mathcal{Y} \\ Y_\tau \notin \mathcal{Y}_\tau^{\neg\mathrm{fnd}}}} \mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau = Y_\tau \mid \mathbf{Y}_{\tau-1} = Y_{\tau-1}) \Big) \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} = Y_{\tau-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\mathrm{fnd}})\end{aligned}$$

where Y_τ denotes the sequence consisting of the measurements in $Y_{\tau-1}$ followed by y, and Y_s, $s < \tau - 1$, the sequence consisting of the first s measurements in Y_τ. But

$$\sum_{\substack{y \in \mathcal{Y} \\ Y_\tau \notin \mathcal{Y}_\tau^{\neg\mathrm{fnd}}}} \mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau = Y_\tau \mid \mathbf{Y}_{\tau-1} = Y_{\tau-1})$$

[5] In this paper we use the notation that, for every integer m and every sequence $\{a_k\}$, $\sum_{k=m}^{m-1} a_k \triangleq 0$.

is precisely the probability $h(Y_{\tau-1})$ of finding the evader for the first time at τ, given the measurements $\mathbf{Y}_{\tau-1} = Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}}$. Therefore

$$\begin{aligned}
&\mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau \notin \mathcal{Y}_\tau^{\neg\text{fnd}},\ \mathbf{Y}_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}},\ h(\mathbf{Y}_{\tau-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad = \sum_{\substack{Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}} \\ h(Y_{\tau-1}) \geq \bar{\delta} \\ h(Y_s) < \bar{\delta},\ s=t-1,t,\dots,\tau-2}} h(Y_{\tau-1})\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} = Y_{\tau-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad \geq \bar{\delta}\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}},\ h(\mathbf{Y}_{\tau-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}).
\end{aligned}$$

Using this in (32) we conclude that

$$\begin{aligned}
\bar{f}_{\bar{g}}(t) &\geq \bar{\delta} \sum_{k=0}^{T-1} \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+k-1} \in \mathcal{Y}_{t+k-1}^{\neg\text{fnd}},\ h(\mathbf{Y}_{t+k-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&= \bar{\delta} \sum_{k=0}^{T-2} \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+k-1} \in \mathcal{Y}_{t+k-1}^{\neg\text{fnd}},\ h(\mathbf{Y}_{t+k-1}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad + \bar{\delta}\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{t+T-2} \in \mathcal{Y}_{t+T-2}^{\neg\text{fnd}},\ h(\mathbf{Y}_{t+T-2}) \geq \bar{\delta},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, t+T-3 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \qquad (33)
\end{aligned}$$

Now, due to the hypothesis of the lemma, for any realization of the random variables $\mathbf{Y}_{t-1}, \mathbf{Y}_t \dots, \mathbf{Y}_{t+T-2}$, if $h(\mathbf{Y}_s) < \bar{\delta} \leq \delta$, $s = t-1, t, \dots, t+T-3$, we must have $h(\mathbf{Y}_{t+T-2}) \geq \delta \geq \bar{\delta}$. We can therefore remove this constraint in (33) and obtain

$$\bar{f}_{\bar{g}}(t) \geq \bar{\delta} u_0. \qquad (34)$$

From the definition of the u_n it is clear that $u_{T-1} = 1$. We proceed by showing that $u_n \geq (1-\bar{\delta}) u_{n+1}$, $n \in \{0, 1, \dots, T-2\}$, from which we will conclude that $\bar{f}_{\bar{g}}(t) \geq \bar{\delta}(1-\bar{\delta})^{T-1}$. For a given $\tau \in \{t, t+1, \dots, t+T-1\}$,

$$\begin{aligned}
&\mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau \in \mathcal{Y}_\tau^{\neg\text{fnd}},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-1 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad = \sum_{\substack{Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}} \\ h(Y_s) < \bar{\delta},\ s=t-1,t,\dots,\tau-1}} \Bigg(\sum_{\substack{y \in \mathcal{Y} \\ Y_\tau \in \mathcal{Y}_\tau^{\neg\text{fnd}}}} \mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau = Y_\tau \mid \mathbf{Y}_{\tau-1} = Y_{\tau-1}) \Bigg) \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} = Y_{\tau-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}})
\end{aligned}$$

where Y_τ denotes the sequence consisting of the measurements in $Y_{\tau-1}$ followed by y, and Y_s, $s < \tau - 1$, the sequence consisting of the first s measurements in Y_τ. But

$$\sum_{\substack{y \in \mathcal{Y} \\ Y_\tau \in \mathcal{Y}_\tau^{\neg\text{fnd}}}} \mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau = Y_\tau \mid \mathbf{Y}_{\tau-1} = Y_{\tau-1})$$

is precisely the probability $1 - h(Y_{\tau-1})$ of not finding the evader for the first time at τ, given the measurements $\mathbf{Y}_{\tau-1} = Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}}$. Therefore

$$\begin{aligned}
&\mathrm{P}_{\bar{g}}(\mathbf{Y}_\tau \in \mathcal{Y}_\tau^{\neg\text{fnd}},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-1 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad = \sum_{\substack{Y_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}} \\ h(Y_s) < \bar{\delta},\ s=t-1,t,\dots,\tau-1}} \big(1 - h(Y_{\tau-1})\big)\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} = Y_{\tau-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}) \\
&\quad \geq (1-\bar{\delta})\, \mathrm{P}_{\bar{g}}(\mathbf{Y}_{\tau-1} \in \mathcal{Y}_{\tau-1}^{\neg\text{fnd}},\ h(\mathbf{Y}_s) < \bar{\delta},\ s = t-1, t, \dots, \tau-1 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}_{t-1}^{\neg\text{fnd}}).
\end{aligned}$$

Applying the above formula for $\tau = t+T-n-2$, we conclude that

$$
\begin{aligned}
u_n \geq & \sum_{k=0}^{T-n-2} \mathrm{P}_{\bar g}(\mathbf{Y}_{t+k-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t+k-1},\ h(\mathbf{Y}_{t+k-1}) \geq \bar\delta,\ h(\mathbf{Y}_s) < \bar\delta,\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1}) \\
& + (1-\bar\delta)\, \mathrm{P}_{\bar g}(\mathbf{Y}_{t+T-n-3} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t+T-n-3},\ h(\mathbf{Y}_s) < \bar\delta,\ s = t-1, t, \dots, t+T-n-3 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1}) \\
\geq & (1-\bar\delta)\Bigg(\sum_{k=0}^{T-3-n} \mathrm{P}_{\bar g}(\mathbf{Y}_{t+k-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t+k-1},\ h(\mathbf{Y}_{t+k-1}) \geq \bar\delta, \\
& \qquad h(\mathbf{Y}_s) < \bar\delta,\ s = t-1, t, \dots, t+k-2 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1}) \\
& \quad + \mathrm{P}_{\bar g}(\mathbf{Y}_{t+T-n-3} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t+T-n-3},\ h(\mathbf{Y}_{t+T-n-3}) \geq \bar\delta, \\
& \qquad h(\mathbf{Y}_s) < \bar\delta,\ s = t-1, t, \dots, t+T-n-4 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1}) \\
& \quad + \mathrm{P}_{\bar g}(\mathbf{Y}_{t+T-n-3} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t+T-n-3},\ h(\mathbf{Y}_s) < \bar\delta,\ s = t-1, t, \dots, t+T-n-3 \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1})\Bigg) \\
= & (1-\bar\delta) u_{n+1}.
\end{aligned} \tag{35}
$$

From this and (34) we then conclude that

$$
\bar f_{\bar g}(t) \geq \bar\delta(1-\bar\delta)^{T-1} u_{T-1} = \bar\delta(1-\bar\delta)^{T-1}\, \mathrm{P}_{\bar g}(\mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1} \mid \mathbf{Y}_{t-1} \in \mathcal{Y}^{\neg\mathrm{fnd}}_{t-1}) = \bar\delta(1-\bar\delta)^{T-1}.
$$

Since the above is true from any $\bar\delta \in [0, \delta]$, we also have $\bar f_{\bar g}(t) \geq \max_{\bar\delta \in [0,\delta]} \bar\delta(1-\bar\delta)^{T-1} = \epsilon$, with ϵ as in (11). ■

References

[1] T. D. Parsons. Pursuit-evasion in a graph. In Y. Alani and D. R. Lick, editors, *Theory and Application of Graphs*, pp. 426–441. 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*, 35(1):18–44, Jan. 1988.

[3] A. S. Lapaugh. Recontamination does not help to search a graph. *Journal of the ACM*, 41(2):224–245, Apr. 1993.

[4] I. Suzuki and M. Yamashita. Searching for a mobile intruder in a polygonal region. *SIAM J. Comput.*, 21(5):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*, 45(2):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), 31(5):1–25, 1998.

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

[10] V. J. Lumelsky and A. A. Stepanov. Dynamic planning for a mobile automaton with limited information on the environment. *IEEE Trans. Automat. Contr.*, 1986.