49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

Persistent Patrol with Limited-range On-Board Sensors

Vu Anh Huynh, John J. Enright, and Emilio Frazzoli

Abstract— **We propose and analyze the Persistent Patrol Problem (PPP). An unmanned aerial vehicle (UAV) moving with constant speed and unbounded acceleration patrols a bounded region of the plane where localized incidents occur according to a renewal process with known time intensity and spatial distribution. The UAV can detect incidents using on-board sensors with a limited visibility radius. We want to minimize the expected waiting time between the occurrence of an incident, and the time that it is detected. First, we provide a lower bound on the achievable expected detection time of any patrol policy in the limit as the visibility radius goes to zero. Second, we present the Biased Tile Sweep policy whose upper bound shows i) the lower bound's tightness, ii) the policy's asymptotic optimality, and iii) that the desired spatial distribution of the searching vehicle's position is proportional to the square root of the underlying spatial distribution of incidents it must find. Third, we present two online policies: i) a policy whose performance is provably within a constant factor of the optimal called TSP Sampling, ii) and the TSP Sampling with Receding Horizon heuristically yielding better performance than the former in practice. Fourth, we present a decision-theoretic approach to the PPP that attempts to solve for optimal policies offline. In addition, we use numerical experiments to compare performance of the four approaches and suggest suitable operational scenarios for each one.**

I. INTRODUCTION

Persistent patrol missions arise in many contexts such as crime prevention, search and rescue, post-conflict stability operations, and peace keeping. In these situations, military or police units are not only effective deterrents to would-be adversaries but also a speedy task force to intercept any trespassers or provide swift security and assistance. With recent advances in technology, unmanned aerial vehicles (UAVs) are well-suited for these tasks because they possess a large bird's-eye view and are unhindered by ground obstacles. The path planning algorithms used in such missions play a critical role in minimizing the required resources, and maximizing the quality of service provided.

In this work, we propose and analyze the persistent patrol problem (PPP), a generic mathematical model for UAVs with limited sensors to perform such a mission in stochastic environments. Incidents occur dynamically and stochastically according to a general renewal process with known time intensity and spatial distribution in the environment $\mathcal{A} \subset \mathbb{R}^2$. The UAV is modeled as a point mass traveling at a constant speed with unbounded acceleration. The UAV detects incidents within the footprint of its on-board sensors, i.e., within its visibility range σ. We want to minimize the expected waiting time between the occurrence of an incident, and its detection epoch.

Related research focuses on search and rescue missions in which the number of searched objects is known at the beginning of the missions [1]–[3]. In other words, these works present problems in which the set of objects to be found is static. In the PPP, by contrast, incidents of interest arrive continuously with unknown arrival times, and therefore the search effort must be persistent and preventive over an infinite-time horizon.

A closely related problem is the Dynamic Traveling Repairman Problem (DTRP), see, e.g., [4]–[7]. Both the PPP and the DTRP have queuing phenomena in which demands or incidents arrive in the system and wait for service. However, unlike the PPP, the DTRP assumes that all demands are known upon their arrival epochs. Thus, few works have been devoted to analyzing policies for the PPP other than our own [8]–[10]. Other studies such as [11] investigate the use of approximate dynamic programming to construct policies for a team of vehicles to perform persistent patrol or monitoring. However, the work does not provide a connection with the queuing nature of the system. In our work, we highlight this connection by using Little's theorem to convert the minimization of the expected waiting time into a dynamic programming formulation. Song, Kim and Yi considered the problem of searching for a static object emitting intermittent stochastic signals under a limited sensing range, and analyze the performance of standard algorithms such as systematic sweep and random walks [12]. Due to the intermittent signals from the object, searching robots need to perform a persistent search, thus making the work relevant to our problem. However, the authors assumed no prior information about the location of the target object is available; hence, their setting is equivalent to the assumption of a uniform spatial distribution. In our work, we explicitly consider non-uniform spatial distributions, which lead to different kinds of optimal policies. Mathew and Mezic presented an algorithm named Spectral Multiscale Coverage (SMC) to devise trajectories such that the spatial distribution of the patrol vehicle's position asymptotically matches a given function [13]. We show that when attempting to minimize discovery time, the desired spatial distribution of the patrol vehicle's position is dependent on, but not equivalent to the underlying spatial distribution of incidents it must find.

In this paper, we introduce the mathematical model of the problem, prove bounds on achievable performance of any algorithm for the PPP, and propose a variety of policies. The approaches discussed include periodic path coverage sweeps of subregions of the environment, sampling the known spatial distribution of the incident generation process and performing Traveling Salesman Problem (TSP) tours over them, and an application of approximate dynamic programming (ADP) [14]. The contribution of this paper is fourfold. First, we provide a lower bound on the achievable expected detection time of any patrol policy in the limit as the visibility radius goes to zero. Second, we present the Biased Tile Sweep policy whose upper bound shows i) the lower bound's tightness, ii) the policy's asymptotic optimality, and iii) that the desired spatial distribution of the

V. A. Huynh and E. Frazzoli are members of the Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. vuhuynh@mit.edu, frazzoli@mit.edu
J. J. Enright is with Kiva Systems, 225 Wildwood Ave, Woburn, MA 01801. jenright@kivasystems.com

searching vehicle's position is porportional to the square root of the underlying spatial distribution of incidents it must find. Third, we present two online and easy-to-implement policies: i) a policy whose performance is provably within a constant factor of the optimal called TSP Sampling (TSP-S), ii) and the TSP Sampling with Receding Horizon (TSP-SRH(η)) heuristically yielding better performance than the former in practice. All three are adaptive to time intensity of the incident process. While the TS policy is deterministic, TSP-S and TSP-SRH(η) policies are stochastically driven. Fourth, we present a decision-theoretic model using a Markov Decision Process (MDP) that attempts to solve for optimal policies offline using ADP. In addition, we use numerical experiments to compare performance of the four approaches and suggest suitable conditions for each one.

The organization of this paper is as follows. In Section II, we formally state the problem. We prove a lower bound for the class of policies of interest in Section III and present our policies in Section IV. In Section V, we present an alternative MDP model of the problem. We discuss results in Section VI and conclude with final remarks in Section VII.

II. PROBLEM STATEMENT

Consider a planar region $\mathcal{A} \subset \mathbb{R}^2$ of unit area. Incidents arrive according to a renewal process with time intensity λ, and upon arrival are independently and identically assigned a location, according to an absolutely continuous distribution supported on $\mathcal{A}$, with spatial density φ. For simplicity, we model φ as a piecewise constant function, over K subregions (such a model can approximate any spatial density of interest arbitrarily well):

$$\varphi(x) = \varphi_k, \quad \text{for } x \in \mathcal{A}_k, \quad k = 1, 2, \ldots, K, \tag{1}$$

with $\sum_{k=1}^{K} \varphi_k A_k = 1$, where $A_k > 0$ is the area of the subregion $\mathcal{A}_k$. Let $\overline{\varphi} = \max_{x \in \mathcal{A}} \varphi(x)$ be the maximum value attained by φ over the entire region $\mathcal{A}$.

A single UAV, moving at a constant speed v, has on-board sensors that detect incidents within a distance σ of the vehicle. The motion of the UAV is determined by a search policy, i.e., an algorithm that determines the flight direction based on the vehicle's position and on the available information about the locations of the incidents (e.g., an estimate of the spatial distribution of the currently outstanding incidents or prior statistics on the incident generation process). The detection time T_i of the i^{th} incident is defined as the elapsed time from its arrival to the moment the incident is detected by the UAV. Given a search policy π, the system detection time is thus defined as

$$\overline{T}_\pi = \lim_{i \to \infty} \mathbb{E}[T_i : \text{ the UAV executes policy } \pi], \tag{2}$$

where we assume the limits exits.

A policy is called stable if the expected number of undetected incidents is uniformly bounded at all times. Let $\mathcal{P}$ be a set of all causal, stable, and stationary policies. The optimal detection time is denoted as $\overline{T}^* = \inf_{\pi \in \mathcal{P}} \overline{T}_\pi$. The objective of this work is to find a policy that provably achieves either the optimal detection time, or an approximation thereof. In particular, we say that a search policy π achieves a constant factor approximation κ if $\overline{T}_\pi \leq \kappa \overline{T}^*$.

In addition to the problem statement, we require the following classic result throughout our own proofs. For a set of n points independently sampled from an absolutely continuous distribution with spatial density φ with compact support $\mathcal{A} \subset \mathbb{R}^2$, there exists a constant $\beta \in \mathbb{R}$ such that the length of the Euclidean Traveling Salesman Problem (ETSP) tour through all n points satisfies the following limit, almost surely [15]:

$$\lim_{n \to +\infty} \frac{\text{ETSP}(n)}{\sqrt{n}} = \beta \int_{\mathcal{A}} \sqrt{\varphi(x)}\, dx, \qquad \text{a.s.} \tag{3}$$

The current best estimate of the constant is $\beta = 0.7120 \pm 0.0002$ [16], [17].

III. LOWER BOUND

We now investigate the performance limits of any stabilizing policy for the PPDP. The following result can be directly extended to the multi-vehicle case. However, for the scope of this paper, we reduce it to the single-vehicle case.

Theorem 1: The optimal detection time for the PPDP satisfies

$$\lim_{\sigma \to 0^+} \overline{T}^* \sigma \geq \frac{1}{4v} \left(\int_{\mathcal{A}} \sqrt{\varphi(x)}\, dx \right)^2. \tag{4}$$

We will refer to the RHS of Eq. 4 divided by σ as T_{LB}.

Proof: In the following, we denote the sensor footprint of a vehicle at position p as $\mathcal{S}_\sigma(p)$. The probability that an incident's location is within the sensor footprint at the time of arrival is bounded by $\Pr[x \in \mathcal{S}_\sigma(p)] \leq \overline{\varphi}\pi\sigma^2$. In this case, the detection time for the incident is zero. However, for any given distribution φ,

$$\lim_{\sigma \to 0^+} \Pr[x \in \mathcal{S}_\sigma(p)] \leq \lim_{\sigma \to 0^+} \overline{\varphi}\pi\sigma^2 = 0, \quad \forall x \in \mathcal{A}$$

and therefore, $\lim_{\sigma \to 0^+} \Pr[x \notin \mathcal{S}_\sigma(p)] = 1$. We note that in this limit, from the perspective of a point $x \in \mathcal{A}$, the actions of any stabilizing policy π can be described by the following (possibly nondeterministic) sequence of variables: the lengths of the time intervals during which the point is not contained in the sensor footprint, $Y_j(x)$. To denote the mean of the sequence under the actions of a given policy, we use

$$\mathbb{E}\left[Y_\pi(x)\right] = \lim_{j \to \infty} \mathbb{E}\left[Y_j(x) : \text{ the UAV executes policy } \pi\right],$$

and similarly to denote the variance and second moment of such a sequence, we use $\operatorname{var}\left[Y_\pi(x)\right]$ and $\mathbb{E}\left[Y_\pi^2(x)\right]$, respectively. Due to random incidences [18], [19], an incident's detection time, conditioned upon its location, is written as

$$\lim_{\sigma \to 0^+} \mathbb{E}\left[T_i | x_i = x\right] = \lim_{\sigma \to 0^+} \Pr[x \notin \mathcal{S}_\sigma(p)] \cdot \frac{\mathbb{E}\left[Y_\pi^2(x)\right]}{2\,\mathbb{E}\left[Y_\pi(x)\right]}$$
$$= \frac{\mathbb{E}\left[Y_\pi(x)\right]^2 + \operatorname{var}\left[Y_\pi(x)\right]}{2\,\mathbb{E}\left[Y_\pi(x)\right]} \geq \frac{1}{2}\,\mathbb{E}\left[Y_\pi(x)\right].$$

In other words, for fixed $\mathbb{E}\left[Y_\pi(x)\right]$, the detection time is minimized if $\operatorname{var}\left[Y_\pi(x)\right]$ is zero.

The remainder of the proof assumes the limit as $\sigma \to 0^+$. Let us define $f(x)$ as the frequency at which point x is searched, i.e., $f(x) = 1/\,\mathbb{E}\left[Y_\pi(x)\right]$, and so

$$\overline{T} = \int_{\mathcal{A}} \varphi(x)\,\mathbb{E}\left[T_i | x_i = x\right]\, dx \geq \frac{1}{2} \int_{\mathcal{A}} \frac{\varphi(x)}{f(x)}\, dx.$$

The vehicle is capable of searching at a maximum rate of $2v\sigma$ (area per unit time), and so the average searching frequency is bounded by $\int_{\mathcal{A}} f(x)/A \ dx \leq 2v\sigma/\mathrm{A}$.

Thus, we have: $\overline{T}^* \geq \frac{1}{2}\overline{T}_f$, where

$$\overline{T}_f = \min_f \int_{\mathcal{A}} \frac{\varphi(x)}{f(x)} \ dx$$

$$\text{subject to} \qquad \int_{\mathcal{A}} f(x) \ dx \leq 2v\sigma \quad \text{ and } \quad f(x) > 0.$$

Since the objective function is convex in $f(x)$ and the constraints are linear, the above is an infinite-dimensional convex program. Relaxing the constraint with a multiplier, we arrive at the Lagrange dual:

$$\begin{aligned}&\min_{f(x)>0}\left[\int_{\mathcal{A}} \frac{\varphi(x)}{f(x)} \ dx + \Gamma\left(\int_{\mathcal{A}} f(x) \ dx - 2v\sigma\right)\right] \\ &= \int_{\mathcal{A}} \min_{f(x)>0}\left[\frac{\varphi(x)}{f(x)} + \Gamma f(x)\right] \ dx - 2v\sigma\Gamma.\end{aligned}$$

Differentiating the integrand with respect to $f(x)$ and setting it equal to zero, we find the pair

$$f^*(x) = \sqrt{\frac{\varphi(x)}{\Gamma^*}}, \quad \Gamma^* = \left(\frac{1}{2v\sigma}\int_{\mathcal{A}} \sqrt{\varphi(x)} \ dx\right)^2 \qquad (5)$$

satisfy the Kuhn-Tucker necessary conditions for optimality [20], [21]. Since it is a convex program, these conditions are sufficient to insure global optimality. Upon substitution, (4) is proved. ∎

Oftentimes, a tight lower bound offers insight into the optimal solution of a problem. Assuming that this lower bound is tight, Eq. (5) suggests that the optimal policy searches the neighborhood of a point x at identical intervals at a relative frequency proportional to $\sqrt{\varphi(x)}$. In fact, this lower bound is shown to be asymptotically tight through a constructive proof using one of the policies presented in this paper.

IV. ONLINE POLICIES

A. Biased Tile Sweep (BTS)

We begin with a description of a subroutine used by the BTS policy.

SWEEP-SERVICE

Given a subregion, $\mathcal{S}$, partition it into strips of width 2σ and execute a path running along the longitudinal bisector of each strip, visiting all strips from top-to-bottom, connecting adjacent strip bisectors by their endpoints. We now state a bound on the length $L_{\text{ss}}(\mathcal{S})$ of the path planned by the algorithm.

Proposition 2: The length $L_{\text{ss}}(\mathcal{S})$ of the path planned by SWEEP-SERVICE for region $\mathcal{S}$ satisfies

$$\lim_{\sigma\to 0^+} L_{\text{ss}}(\mathcal{S})\sigma \leq \frac{\mathrm{A}_{\mathcal{S}}}{2}.$$

Although the full proof is omitted for brevity, it is a simple extension of the following fact: given a grid of squares whose side-lengths are 2σ, the number of squares with nonzero intersection with $\mathcal{S}$ satisfies [22]: $\lim_{\sigma\to 0^+} N_{\text{sq}}(\mathcal{S})\sigma^2 = \mathrm{A}_{\mathcal{S}}/4$.

This policy requires a tiling of the environment $\mathcal{A}$ with the following properties. For some chosen positive integer $N_c \in \mathbb{N}$, partition each subset $\mathcal{A}_k$ into $N_k = N_c/\sqrt{\varphi_k}$ tiles, each of

Algorithm 1 Biased Tile Sweep Policy

procedure BTS
  Initialize $\ell_k \leftarrow 1$ for $k = 1, 2, \ldots, K$
  for $i \leftarrow 1, 2, ..., \infty$ **do**
    for $k \leftarrow 1, 2, ..., K$ **do**
      Execute SWEEP-SERVICE on tile $\mathcal{S}_{k,\ell_k}$
      if $\ell_k < N_k$ **then**
        $\ell_k \leftarrow (\ell_k + 1)$
      else
        $\ell_k \leftarrow 1$
      end if
    end for
  end for
end procedure

area $\mathrm{A}_k/N_k = \mathrm{A}_k\sqrt{\varphi_k}/N_c$. We assume N_c is chosen large enough that an integer N_k can be found such that N_c/N_k is sufficiently close to $\sqrt{\varphi_k}$. For example, one way of achieving such a tiling is to partition $\mathcal{A}_k$ into strip-like tiles of equal measure with $N_k - 1$ parallel lines. Let us give the tiles of $\mathcal{A}_k$ an ordered labeling $\mathcal{S}_{k,1}, \mathcal{S}_{k,2}, \ldots, \mathcal{S}_{k,N_k}$.

The BTS policy is defined in Algorithm 1, where the index i is a label for the current phase of the policy. The aim of the algorithm is to perform a periodic sweep of every point in the domain with the following two properties. The intervals between sweeps are identical in length, i.e., they have zero variance. The frequency with which a point is searched is proportional to the square root of its density. During each phase, the UAV sweeps one tile from each constant-density subregion. Since all the tiles of a given subregion are equal in area, each phase is equal in duration (asymptotically as $\sigma \to 0^+$). The number of phases a point must wait between sweeps is equal to the number of tiles in its subregion (N_k), which is chosen to be inversely proportional to the square root of that regions density. This causes the frequency with which a point is searched to be proportional to the square root of its density. Since the phase durations are identical, the intervals between sweeping any given tile are identical. To ensure that each point is searched at identical intervals, SWEEP-SERVICE should always execute the same path on a given tile.

Theorem 3: The system detection time of an agent operating on $\mathcal{A}$ under the BTS policy satisfies $\lim_{\sigma\to 0^+} \frac{\overline{T}_{\text{BTS}}}{\overline{T}^*} = 1$.

Proof: The total distance traveled between tiles during a phase is no more than $K \operatorname{diam}(\mathcal{A})$. The duration of a single phase T^{phase} satisfies $T^{\text{phase}} \leq \frac{\sum_{k=1}^K L_{\text{ss}}(\mathcal{S}_{k,\ell_k}) + K \operatorname{diam}(\mathcal{A})}{v}$. Applying Proposition 2, $\lim_{\sigma\to 0^+} T^{\text{phase}}\sigma$ is smaller than:

$$\begin{aligned}&\lim_{\sigma\to 0^+} \frac{\sum_{k=1}^K L_{\text{ss}}(\mathcal{S}_{k,\ell_k})\sigma}{v} + \lim_{\sigma\to 0^+} \frac{K \operatorname{diam}(\mathcal{A})\sigma}{v} \\ &= \frac{\sum_{k=1}^K \mathrm{A}_k/N_k}{2v} = \frac{\sum_{k=1}^K \mathrm{A}_k\sqrt{\varphi_k}}{2N_c v}.\end{aligned}$$

Conditioned upon its location $x \in \mathcal{A}_k$, a target waits one half of N_k phases to be serviced,

$$\mathbb{E}\left[\overline{T}_{\text{BTS}} | x \in \mathcal{A}_k\right] = \frac{1}{2} N_k T^{\text{phase}} = \frac{1}{2}\frac{N_c}{\sqrt{\varphi_k}} T^{\text{phase}}.$$

Algorithm 2 Distribution Sampling TSP Policy

procedure TSP-S($\lambda_1^{\rm s}, \lambda_2^{\rm s}, ..., \lambda_K^{\rm s}, v, \sigma$)
  Initialize TSP constant $\beta := 0.712$.
  while true **do**
    for $k \leftarrow 1, 2, ..., K$ **do**
      $\overline{n}_k := \lambda_k^{\rm s}\beta^2(\sum_{j=1}^{K}\sqrt{\lambda_j^{\rm s}A_j})^2$.
      Sample $\overline{n}_k$ virtual targets uniformly in $\mathcal{A}_k$.
    end for
    Compute $\mathcal{N} := \overline{n}_1 + ... + \overline{n}_K$.
    Compute the TSP tour through $\mathcal{N}$ virtual targets.
    Traverse the TSP tour with random direction with speed v.
    Detect real incidents within sensor radius σ along the tour.
  end while
end procedure

Noting that $\Pr[x \in \mathcal{A}_k] = \varphi_k \, \mathrm{A}_k$ and unconditioning on $x \in \mathcal{A}_k$ to find the system time,

$$\overline{T}_{\rm BTS} = \sum_{k=1}^{K} \Pr[x \in \mathcal{A}_k] \cdot \mathbb{E}\left[\overline{T}_{\rm BTS} | x \in \mathcal{A}_k\right]$$
$$= \sum_{k=1}^{K} (\mathrm{A}_k \, \varphi_k) \cdot \left(\frac{1}{2}\frac{N_c}{\sqrt{\varphi_k}} T^{\rm phase}\right) = \frac{N_c T^{\rm phase}}{2} \sum_{k=1}^{K} \mathrm{A}_k \sqrt{\varphi_k}.$$

Thus,

$$\lim_{\sigma \to 0^+} \overline{T}_{\rm BTS}\sigma = \left(\frac{N_c}{2}\sum_{k=1}^{K}\mathrm{A}_k\sqrt{\varphi_k}\right)\lim_{\sigma\to 0^+} T^{\rm phase}\sigma$$
$$= \frac{1}{4v}\left(\sum_{k=1}^{K}\mathrm{A}_k\sqrt{\varphi_k}\right)^2.$$

Combining with the lower bound on the optimal system time in Theorem 1, the claim is proved. ■

Theorem 3 shows that the optimal policy for small sensor-range searches a point x in the environment at identical intervals at a relative frequency proportional to $\sqrt{\varphi(x)}$, as suggested by Eq. (5) in the proof of the corresponding lower bound in Theorem 1.

B. TSP-Sampling (TSP-S)

The key idea of the TSP-S policy, shown in Algorithm 2, is to simulate an arrival process for incidents, by sampling a distribution in the environment to create virtual targets. We sample virtual targets in $\mathcal{A}$ with a rate $\lambda^{\rm s} = \sum_{k=1}^{K}\lambda_k^{\rm s}$ and a spatial distribution

$$\phi(x) = \frac{\lambda_k^{\rm s}}{\lambda^{\rm s}A_k}, \; \textit{for } x \in \mathcal{A}_k. \tag{6}$$

In each sampling iteration, let us sample $\overline{n}_k = \lambda_k^{\rm s}\beta^2(\sum_{j=1}^{K}\sqrt{\lambda_j^{\rm s}A_j})^2$ virtual targets in the subregion $\mathcal{A}_k$ parameterized by the sampling rate λ_k^s, and up on their arrivals, virtual targets are distributed independently uniformly in the subregion $\mathcal{A}_k$. The path is constructed by computing a series of TSP solutions through sampled virtual targets. An incident is detected when the UAV visits a virtual target whose distance to the incident is less than the radius of sensor range σ. Although solving an exact TSP tour through n points is an NP-hard problem, efficient solvers such as Linkern [23], [24] can obtain approximate TSP tours of length within a constant factor of optimal tours for n in the order of 10,000 in real time. We note that in Subsections IV-B and IV-C, positions of virtual targets are additional information to estimate the spatial distribution of the currently outstanding incidents.

Optimal sampling rates and upper bound

In the TSP-S policy, we can compute detection time as a function of the sampling rates $\lambda_1^{\rm s}$,...,$\lambda_K^{\rm s}$, which are design parameters. Thus, we can optimize the sampling rate parameters to minimize the system detection time. We represent the sampling rates as a function of the visibility radius:

$$\lambda_k^{\rm s} = \frac{l_k}{\sqrt{\pi}\beta\sigma}, \text{ where } l_k \in \mathbb{R}^+ \quad k = 1, 2, ..., K. \tag{7}$$

In each iteration, the number of sampled virtual targets is $\mathcal{N} = \sum_{k=1}^{K}\overline{n}_k = \lambda^{\rm s}\beta^2(\sum_{j=1}^{K}\sqrt{\lambda_j^{\rm s}A_j})^2$. We note that

$$\left(\int_{\mathcal{A}}\phi(x)^{1/2}dx\right)^2 = \frac{1}{\lambda^{\rm s}}\left(\sum_{j=1}^{K}\sqrt{\lambda_j^{\rm s}A_j}\right)^2 = \frac{1}{(\lambda^{\rm s})^2\beta^2}\mathcal{N}.$$

When $\mathcal{N}$ is large, the length of a TSP tour can be approximated by

$$\mathrm{ETSP}(\mathcal{N}) = \sqrt{\mathcal{N}}\beta\int_{\mathcal{A}}\phi(x)^{1/2}dx = \beta^2\left(\sum_{j=1}^{K}\sqrt{\lambda_j^{\rm s}A_j}\right)^2. \tag{8}$$

Now, let us consider the $i^{\rm th}$ incident. If the UAV does not discover the incident in the current TSP tour, the vehicle needs to complete the current tour, sample new virtual targets, and traverse a new TSP tour. This process is repeated until there is a nearby sampled virtual target whose distance to the i^{th} incident is no more than σ. Thus, the i^{th} detection time can be bounded by the sum of three components: (i) waiting time d_i between arrival of the actual target and the first virtual target sampling iteration to occur after its arrival epoch, (ii) waiting time f_i from the first sampling iteration after its arrival epoch until a virtual target appears within the circular neighborhood of radius σ, and (iii) waiting time w_i for the vehicle to visit the virtual target. We have:

$$T_i \leq d_i + f_i + w_i. \tag{9}$$

Since virtual targets are identically and independently sampled according to the spatial distribution $\phi(x)$ in each sampling iteration, we have

$$\mathbb{E}[d_i] \leq \frac{\mathrm{ETSP}(\mathcal{N})}{2v}, \quad \mathbb{E}[w_i] \leq \frac{\mathrm{ETSP}(\mathcal{N})}{2v}. \tag{10}$$

In addition, let $V_{i,k}$ be the event that the i^{th} incident appears in the subregion $\mathcal{A}_k$. Conditioned on $V_{i,k}$, the incident needs to wait for the next sampling iteration if no virtual target appears in its neighborhood in the current sampling iteration, which occurs with probability

$$p(A_k, \overline{n}_k) = \left(1 - \frac{\pi\sigma^2}{A_k}\right)^{\overline{n}_k}. \tag{11}$$

We have:

$$\mathbb{E}[f_i|V_{i,k}] = p(A_k, \overline{n}_k)\left(\frac{\mathrm{ETSP}(\mathcal{N})}{v} + \mathbb{E}[f_i|V_{i,k}]\right), \tag{12}$$

$$\mathbb{E}[f_i|V_{i,k}] = \frac{p(A_k, \overline{n}_k)}{1 - p(A_k, \overline{n}_k)}\frac{\mathrm{ETSP}(\mathcal{N})}{v}. \tag{13}$$

Algorithm 3 TSP Sampling with Receding Horizon

```
1: procedure TSP-SRH(η)(λ^s_1, λ^s_2, ..., λ^s_K, η, v, σ)
2:     Initialize TSP constant β := 0.712.
3:     for k ← 1, 2, ..., K do
4:         n̄_k := (λ^s_k β^2 / 2v^2) (Σ_{j=1}^K √(λ^s_j A_j))^2.
5:         Initialize n_k virtual targets uniformly in A_k.
6:     end for
7:     Compute N := n̄_1 + ... + n̄_K.
8:     while true do
9:         Compute the TSP tour through N virtual targets.
10:        Traverse the η portion of TSP tour with speed v.
11:        Choose the direction that visits more virtual targets.
12:        Count the number of cleared virtual targets N_clear.
13:        Detect real incidents with sensor of radius σ along the tour.
14:        for k ← 1, 2, ..., K do
15:            ΔN_k := λ^s_k η ETSP(N).
16:            Sample additional ΔN_k virtual targets in subregion A_k.
17:        end for
18:        Compute N := N − N_clear + ΔN_1 + ... + ΔN_K.
19:    end while
20: end procedure
```

Thus, using the iterative expectation rule and upon substituting Eqs. 10, 13 into Eq. 9, we have

$$\mathbb{E}[T_i] \le \left(\sum_{k=1}^{K} \frac{\varphi_k A_k}{1-p(A_k,\overline{n}_k)}\right)\frac{\mathrm{ETSP}(\mathcal{N})}{v}. \qquad (14)$$

Using Eq. 7, we can express $\overline{n}_k$ and $TSP(\mathcal{N})$ in terms of l_k:

$$\overline{n}_k = \frac{l_k}{\pi\sigma^2}S(l), \quad \mathrm{ETSP}(\mathcal{N}) = \frac{\beta}{\sqrt{\pi}\sigma}S(l), \qquad (15)$$

where $S(l) = \left(\sum_{j=1}^{K}\sqrt{l_j A_j}\right)^2, \quad l = (l_1, ..., l_K).$

We also note that $\lim_{\sigma\to 0^+} p(A_k, \overline{n}_k) = e^{-\frac{l_k}{A_k}S(l)}$, and from Eq. 14, we obtain an upper bound for detection time of TSP-S policy in terms of $l_1, ..., l_K$:

$$\lim_{\sigma\to 0^+} \overline{T}_S\sigma \le \left(\sum_{k=1}^{K}\frac{\varphi_k A_k}{1-e^{-\frac{l_k}{A_k}S(l)}}\right)\frac{\beta S(l)}{\sqrt{\pi}v}. \qquad (16)$$

Considering the RHS of Eq. 16, which is named the TSP-S upper bound, as a function of $l_1, ..., l_K$, we recognize that though this function is not convex, it has a local minimum that is also the global minimum. Therefore, we can use well-known numerical methods such as Newton's method to minimize the RHS of Eq. 16 and solve for the optimal parameters l_k^*, and hence the optimal sampling rates $(\lambda_k^{\mathrm{s}})^*$. For a given incident spatial distribution $\varphi(x)$, the optimal sampling rates $(\lambda_k^{\mathrm{s}})^*$ yield a constant bound factor of the optimal performance. We will evaluate these constant bound factors experimentally in Section VI.

C. TSP Sampling with Receding Horizon (TSP-SRH(η))

The TSP-SRH(η) policy, shown in Algorithm 3, modifies the TSP-S policy by reducing the waiting time to the next sampling iteration. We have observed that in the TSP-S policy, the vehicle needs to wait for a duration of d_i until the next sampling iteration, thus we can reduce this time if the UAV follows only a fraction η of a TSP tour from the current position before computing the next TSP tour. In particular, we choose the fraction of the TSP tour which contains as many virtual targets as possible. In other words, we generate the sampling process as a Poisson process and then apply the receding horizon policy to them, much like the standard DTRP. In [7], the authors heuristically argue that the receding horizon policy is optimal for the DTRP when the parameter η is small. The experimental value of η is within $[0.1, 0.2]$. We conjecture that the receding horizon policy is optimal for the DTRP, and at steady state, the variance of outstanding virtual targets is small. Thus, the optimal waiting time of any virtual target for suitably small parameter η is conjectured to be:

$$\lim_{i\to\infty}\mathbb{E}[w_i] = \frac{\beta^2\lambda^{\mathrm{s}}}{2v^2}\left(\int_{\mathcal{A}}\phi^{1/2}(x)dx\right)^2 = \frac{\beta^2}{2v^2}\left(\sum_{k=1}^{K}\sqrt{\lambda_k^{\mathrm{s}}A_k}\right)^2.$$

Using Little's theorem, we can obtain the steady expected number of outstanding virtual targets

$$\overline{n} = \lambda^{\mathrm{s}}\lim_{i\to\infty}\mathbb{E}[w_i] = \frac{\lambda^{\mathrm{s}}\beta^2}{2v^2}\left(\sum_{k=1}^{K}\sqrt{\lambda_k^{\mathrm{s}}A_k}\right)^2, \qquad (17)$$

$$\overline{n}_k = \frac{\lambda_k^{\mathrm{s}}}{\lambda^{\mathrm{s}}}\overline{n} = \frac{\lambda_k^{\mathrm{s}}\beta^2}{2v^2}\left(\sum_{k=1}^{K}\sqrt{\lambda_k^{\mathrm{s}}A_k}\right)^2. \qquad (18)$$

Therefore, in Algorithm 3, we initialize the system at steady state at Line 5 and sample additional virtual targets as expected values of Poisson random variables at Line 16.

Similar to the TSP-S policy, we consider sampling rates as a function of the visibility radius: $\lambda_k^{\mathrm{s}} = l_k\sqrt{\frac{2}{\pi}\frac{v}{\beta\sigma}}$. When parameter η is small, we can approximately bound $\mathbb{E}[d_i + f_i|V_{i,k}]$ from above by $\frac{A_k}{\lambda_k^{\mathrm{s}}\pi\sigma^2}$, which is the expected value of the first arrival time of the sub-Poisson process in the subregion $\mathcal{A}_k$. We can carry out a similar procedure as in the TSP-S policy to obtain an upper bound for TSP-SRH(η):

$$\lim_{\sigma\to 0^+}\overline{T}_{SRH}\sigma \le \left[\sum_{k=1}^{K}\left(\frac{\varphi_k A_k^2}{l_k}e^{-\frac{l_k}{A_k}S(l)}\right) + S(l)\right]\frac{\beta}{v\sqrt{2\pi}}. \quad (19)$$

We refer to the RHS of Eq. 19 as the TSP-SRH(η) upper bound. Now, we can find optimal sampling rates $(\lambda_k^{\mathrm{s}})^*$ to minimize the above upper bound by numerical methods, which in turn yields a constant bound factor compared to the optimal performance given an incident spatial distribution. In Section VI, we will carry out extensive simulation to support the above conjecture. Furthermore, similar to the BTS policy, both TSP-S and TSP-SRH(η) with the optimal sampling rates $(\lambda_k^{\mathrm{s}})^*$ also patrol regions with higher spatial density more often.

V. MDP MODEL

In this section, we attempt to solve the PPP optimally by solving an MDP model offline. To enable a tractable model, we further assume that the incident process is Poisson. This assumption is reasonable for the purpose of comparing computational cost with the above policies. We partition the region $\mathcal{A}$ into $\mathcal{C}$ square cells of side $\sqrt{2}\sigma$ such that each cell can be covered within the sensor visibility range so that when the UAV enters a cell, it can detect all outstanding incidents in that cell.

Let T_i be the detection time of the i^{th} incident and let $\mathcal{N}(t)$ be number of outstanding incidents at time instant t. From Little's Theorem, we have the following relation when considering the class of stable policies:

$$\lim_{t\to\infty}\mathbb{E}[\mathcal{N}(t)] = \lambda\lim_{i\to\infty}\mathbb{E}[T_i], \qquad (20)$$

assuming that the two limits exist. Thus, minimizing system detection time, $\lim_{i\to\infty} E[T_i]$, is equivalent to minimizing $\lim_{t\to\infty} E[\mathcal{N}(t)]$, the steady state number of outstanding incidents. If we further discretize the time axis with discrete time index k, we can arrive at:

$$\inf_{\pi\in\mathcal{P}} \lim_{i\to\infty} \mathbb{E}[T_i] = \frac{1}{\lambda} \inf_{\pi\in\mathcal{P}} \lim_{N\to\infty} \mathbb{E}\left[\sum_{k=0}^{N} n_k\right], \tag{21}$$

where n_k is the number of outstanding incidents added (possibly negative if more are serviced than generated) during the k-th interval.

In this section, we define λ_j as the arrival rate of the sub-Poisson process characterizing real incidents in the j-th cell:

$$\lambda_j = \lambda \int_{\mathcal{C}_j} \varphi(x)dx. \tag{22}$$

The components of the MDP are defined as below:

- State: $x_k = (p, r)_k$, where p is the current position index of the vehicle with value from 1 to $\mathcal{C}$, and $r \in \mathbb{R}^{\mathcal{C}}$ is a vector contained the elapsed times since the previous visit to each cell. Because the number of new incidents in the j-th cell generated since the previous visit is Poisson($\lambda_j r_j$), r_j is a sufficient statistic for the j-th cell.
- Control: u_k is one of the neighboring cells (at most four in our case) that share edges with the current cell.
- State transition:

$$x_{k+1} = f_k(x_k, u_k) = (p, r)_{k+1}, \tag{23}$$

$$p_{k+1} = u_k,\ \Delta T = \frac{\sqrt{2}\sigma}{v}, \tag{24}$$

$$r_{i,k+1} = r_{i,k} + \Delta T \text{ if } i \neq p_{k+1},\ r_{p_{k+1},k+1} = 0. \tag{25}$$

- Stage cost/reward:

$$h_k(x_k, u_k) = \mathbb{E}[\frac{n_k}{\lambda}] = \Delta T - g_k(x_k, u_k), \tag{26}$$

$$g_k(x_k, u_k) = \frac{\lambda_{u_k}}{\lambda}(r_{u_k,k} + \Delta T), \tag{27}$$

  where $h_k(x_k, u_k)$, $g_k(x_k, u_k)$ are the normalized (divided by λ) expected number of additional outstanding incidents, and the normalized expected number of detected incidents. While $h_k(x_k, u_k)$ is considered as stage cost, $g_k(x_k, u_k)$ is considered as stage reward.
- Objective function:
  We introduce a discount factor $\alpha \in (0, 1)$ to approximate the original problem as an infinite discounted sum. When $\alpha \to 1$, we have the original problem, and when $\alpha \to 0$, the model prefers large stage reward or small stage cost:

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \alpha^k \frac{n_k}{\lambda}\middle| x_0 = x\right] = \frac{\Delta T}{1-\alpha} - J_\pi(x), \tag{28}$$

$$J_\pi(x) = \mathbb{E}\left[\sum_{k=0}^{\infty} \alpha^k g_k(x_k, u_k)\middle| x_0 = x\right]. \tag{29}$$

  Thus, we have converted the original problem of minimizing the system detection time to maximizing an infinite sum of discounted number of detected incidents.

  The objective is to find an optimal stable stationary policy π^* that maximize $J_\pi(x)$: $\pi^* = \text{argmax}_{\pi\in\mathcal{P}} J_\pi(x)$.

The advantage of this approach is the MDP model can be extended for more general objective functions that considers trade-off among factors. However, the model has an infinite space dimension, and thus solving it exactly is impossible. To solve this MDP, we use policy iteration with linear approximate cost structure. From [11], under a stationary policy π, at state $x = (p, r)$, the cost $J_\pi(x)$ is approximated by $\tilde{J}_\pi(x)$:

$$\tilde{J}_\pi(x = (p, r)) = \gamma_p + \xi_p' r, \tag{30}$$

where $\gamma_p \in \mathbb{R}, \xi_p \in \mathbb{R}^{\mathcal{C}}$. We start the policy iteration process with a greedy algorithm that visits the neighbor cell with the largest expected number of incidents. The value of discount factor α is set to be in the range of $[0.4\ 0.6]$ experimentally to obtain solutions that converge. Because the objective function is an approximation of detection time, we need to simulate a policy to compute actual detection time.

VI. RESULTS

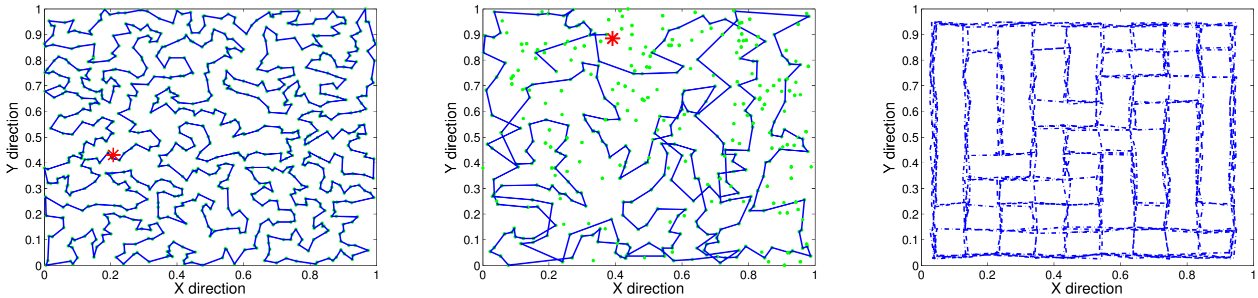
In the following experiments, we assume that the UAV travels at a unit speed, $v = 1$ (unit length per second). In Fig. 1, we show examples of a path generated by different policies. Incidents arrives with temporal intensity $\lambda = 1$ and a uniform spatial distribution. We assume that the vehicle has visibility radius $\sigma = 0.05$. As we can see in Fig. 1.a and Fig. 1.b, both TSP-S and TSP-SRH($\eta = 0.2$) are stochastic by following TSP tours induced by sampled virtual targets. While the TSP-S policy follows a complete tour before a new tour, the TSP-SRH($\eta = 0.2$) policy travels 20 percent of a tour before sampling new targets and considering outstanding targets in the next TSP computation. In Fig. 1.c, the MDP policy with $\alpha = 0.6$ and 10 policy iterations generates a deterministic path and constrains the vehicle to travel in a grid pattern. When the vehicle enters a cell, it goes to the center of the cell.

A. Optimal sampling rates in TSP-SRH(η)

In the first experiment, we supported our conjecture on the theoretical upper bound of the TSP-SRH(η) policy in Section IV-C by comparing the TSP-SRH(η) upper bound (RHS of Eq. 19) and extensively simulated detection time. Fig. 2 illustrates the performance of the sampling TSP-like policy in a two-region piecewise uniform environment where 60 percent of incidents arrive in the 20-percent subregion on the left hand side of the unit square. The visibility radius σ is set to 0.05. The TSP-SRH(η) upper bound from the conjecture is a function of the sampling rates λ_1^{s} and λ_2^{s} for the two subregions. The optimal sampling rates $(\lambda_1^{\text{s}})^*$ and $(\lambda_2^{\text{s}})^*$ can be computed accordingly. We observed that the optimal experimental sampling rates $(\lambda_1^{\text{s}})^*$ and $(\lambda_2^{\text{s}})^*$ are close to the conjectured optimal values.

B. Performance of policies as the radius σ approaches zero

To evaluate the performance of the policies in the regime of small visibility range, we compared BTS, TSP-S, TSP-SRH(η) and MDP for various values of σ from 0.1 to 1/1280 as shown in Fig. 3. The spatial distribution is uniform. As we can see, the MDP policy can obtain detection time that is almost identical to the lower bound for $\sigma \geq 0.0125$. For smaller values of σ, the gap between the MDP policy and



(a) TSP-S policy. (b) TSP-SRH(η) policy. (c) MDP policy.

Fig. 1. Examples of policy behavior in a unit square with arrival time intensity $\lambda = 1$ and a uniform spatial distribution. In (a) and (b), dots are virtual targets, and the star is a tagged incident. The incident will be detected when the UAV visits a virtual target whose distance to the incident is less than the radius of sensor visibility $\sigma = 0.05$. In the TSP-S policy, the vehicle travels a complete a TSP tour before beginning a new iteration. In the TSP-SRH(η) policy, the vehicle only travels a fraction η of a TSP tour before computing a new tour. In (c), an example of a path generated by the MDP policy, $\alpha = 0.6$ and 10 policy iterations. When the vehicle enters a cell, it goes to the center of the cell. Lines are drawn with some random offsets around cell centers to depict multiple paths. As we can see, because MDP policy is deterministic, the path possesses a regular pattern, as compared to the randomness of paths produced by TSP-S and TSP-SRH(η).

TABLE I

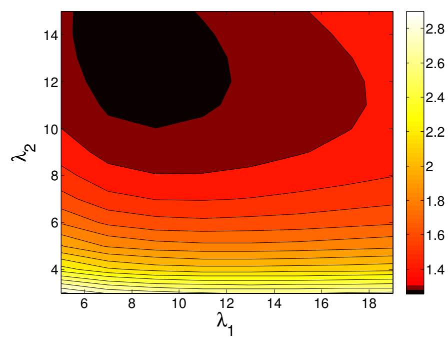
PROPERTIES OF POLICIES

Properties	BTS	TSP-S	TSP-SRH(η)	MDP
Adaptive to λ	Yes	Yes	Yes	Yes
Quality of policy as $\sigma \to 0$	Optimal	Within constant bound factor	Within constant bound factor	Optimal
Online	Yes	Yes	Yes	No
Stochastic	No	Yes	Yes	No

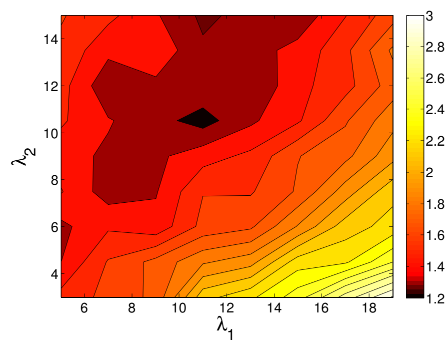
the lower bound is larger. This is due to the approximation of the cost function in the MDP model via discount factor α. In terms of computational requirement, the MDP policy requires offline computation via extensive sampling to ensure approximation quality. Thus, for small values of σ, the MDP policy provides poor performance with early termination. In contrast, the BTS policy almost achieves the lower bound for very small values of σ while keeping the online computation cost small. This observed performance agrees with our analysis. Moreover, we observe that the TSP-SRH(η) policy consistently outperforms the TSP-S policy. In practice, the TSP-SRH(η) policy has good performance in situations where we prefer stochastic trajectories.

C. Performance of policies in non-uniform distributions

In this experiment, we tested the performance of the policies in an environment made up of a unit square with the following spatial distribution: the leftmost 10% of the square has density value $1 + 10\epsilon$, and the other 90% has density value $1 - 10\epsilon/9$, where $\epsilon \in [0, 0.89]$. This environment varies from uniform to nonuniform as ϵ increases from zero. The sensing radius is σ= 0.00625. In Fig. 4, we plot the lower bound, upper bounds and experimental detection times for different values of ϵ. As we can see, the simulated system detection time curves are bounded between the theoretical lower bound and the upper bound curves. The result indicates that for a given spatial distribution, optimal sampling rates ensure that the TSP-S and TSP-SRH(η) policies are within a constant bound factor compared to the optimal performance. Moreover, the performance of BTS is very close to the lower bound and is better than TSP-S and TSP-SRH(η) as we expected. Table I summarizes the above discussion.



(a) Conjectured constant bound factor.



(b) Simulated ratio of system detection time to the lower bound.

Fig. 2. Performance of the TSP-SRH(η) policy in a two-region piecewise uniform environment (60% density in 20% of the area on the left hand side). In (a), the conjectured constant bound factor is plotted with respect to sampling rates λ_1^s and λ_2^s ($\sigma = 0.05$). The optimal sampling rates $(\lambda_1^s)^*$ and $(\lambda_2^s)^*$ can be computed accordingly. In (b), the ratio of detection time to the lower bound (for this distribution) is plotted under the same settings as in (a). The optimal experimental sampling rates $(\lambda_1^s)^*$ and $(\lambda_2^s)^*$ are close to the optimal conjectured values.

VII. CONCLUSIONS

In this paper, we have introduced, analyzed and compared four approaches to the PPP. We consider a UAV with limited sensing capability on a search mission to detect incidents that arrive continuously. We have proved a lower bound for the class of stable policies in the limit as the sensor range shrinks infinitesimally small. We presented the BTS policy and showed an upper bound on its performance that is tight with our proven lower bound. This tightness has shed light on the following fact: in a persistent search scenario, the

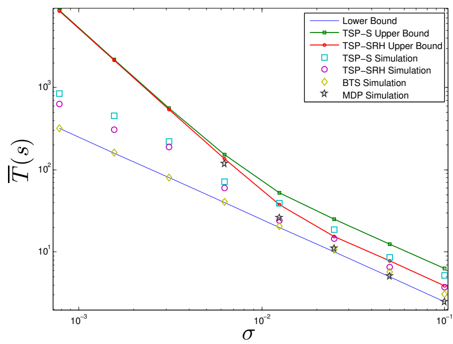


Fig. 3. Detection time of the policies compared to the lower bound as radius σ approaches zero on a log-log plot. The spatial distribution is uniform. MDP has good performance for large values of σ while BTS provides excellent performance for very small values of σ. The TSP-SRH(η) policy consistently outperforms the TSP-S policy.

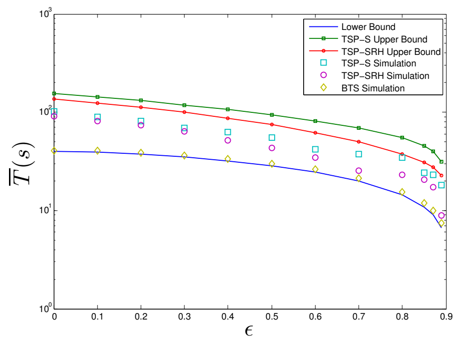


Fig. 4. Performance of the policies in comparison with the theoretical lower bound and upper bounds for varying spatial distribution on a semi-log plot. The sensing radius is $\sigma = 0.00625$. The spatial distribution varies from uniform to 99% of the incidents occuring in a subregion with 10% of the area.

desired spatial distribution of the searching vehicle's position is porportional to the square root of the underlying spatial distribution of incidents it must find.

In particular, the BTS policy scans subregions at identical, predictable intervals. While this is necessary to achieve the lower bound, it could be a weakness in a more complex problem setting in which the incidents have the ability to evade the searching vehicle. With that vision in mind, we present two stochastic policies whose executed paths less unpredictable, called TSP-S and TSP-SRH(η). Both policies work based on solving consecutive TSP tours of sampled virtual targets. We have proved an upper bound for the TS-S policy and presented guidelines for choosing optimal sampling rates to minimize the upper bound. In connection to the DTRP, we have provided a conjecture on an upper bound of the TSP-SRH(η) policy, and through simulation we have verified that the TSP-RH(η) policy has a good performance in practice.

Moreover, we have presented a general MDP formulation for the system detection time, which can be extended to general objective functions. Since the size of the state space is infinite, we use ADP methods to find near-optimal solutions. We have verified that in the limit as the visibility radius goes to zero, the performance of the TSP-SRH(η) policy and the MDP policy are comparable, but the former is more computationally efficient. This result has suggested that online sampling policies are preferred when the visibility radius is small compared to the area of supervised regions. In future work, we would like to extend and analyze sampling methods for more complex situations such as multiple vehicles with imperfect sensors, uncertain movement, and imperfect knowledge of their locations.

VIII. ACKNOWLEDGMENTS

V. A. Huynh gratefully thanks the Arthur Gelb Foundation for supporting him during this work. This research was partially supported by the Air Force Office of Scientific Research, under grants no. and FA8650-07-2-3744 and FA9550-09-1-0522.

REFERENCES

[1] D. V. Chudnovsky and G. V. Chudnovsky, *Search theory : some recent developments*. CRC Press, 1989.
[2] B. DasGupta, J. P. Hespanha, and E. Sontag, "Computational complexity of honey-pot searching with local sensory information," in *Proc. of the 2004 Amer. Contr. Conf.*, June 2004.
[3] J. N. Eagle and J. R. Yee, "An optimal branch-and-bound procedure for the constrained path, moving target search problem," *Oper. Res.*, vol. 38, no. 1, pp. 110–114, 1990.
[4] D. J. Bertsimas and G. Van Ryzin, "Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles," *Oper. Res.*, vol. 41, no. 1, pp. 60–76, 1993.
[5] D. Bertsimas and G. Van Ryzin, "Stochastic and dynamic vehicle routing with general demand and interarrival time distributions," *Advances in Applied Probability*, vol. 25, no. 4, pp. 947–978, 1993.
[6] H. Xu, "Optimal policies for stochastic and dynamic vehicle routing problems," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Aug. 1995.
[7] M. Pavone, E. Frazzoli, and F. Bullo, "Adaptive and distributed algorithms for vehicle routing in a stochastic and dynamic environment," *IEEE Trans. on Automatic Control*, 2009, provisionally accepted.
[8] J. J. Enright and E. Frazzoli, "Cooperative UAV routing with limited sensor range," in *Proceedings of the AIAA Guidance, Navigation and Control Conference*, Keystone, CO, 2006.
[9] A. Arsie, J. J. Enright, and E. Frazzoli, "On the value of information in dynamic multiple-vehicle routing problems," in *Cooperative Control of Distributed Multi-Agent Systems*, J. Shamma, Ed. John Wiley & Sons, 2007.
[10] J. J. Enright, "Efficient routing of multi-vehicle systems: limited sensing and nonholonomic motion constraints," Ph.D. dissertation, University of California at Los Angeles, LA, 2008.
[11] S. Ruan, C. Meirina, F. Yu, K. Pattipati, and R. L. Popp, "Patrolling in a stochastic environment," in *Proceedings of the International Command and Control Research Symposium*, 2005.
[12] D. Song, C. Y. Kim, and J. Yi, "Stochastic modeling of the expected time to search for an intermittent signal source under a limited sensing range," in *Proceedings of Robotics: Science and Systems*, Zaragoza, Spain, June 2010.
[13] G. Mathew and I. Mezic, "Spectral multiscale coverage: A uniform coverage algorithm for mobile sensor networks," in *Proceedings of the 48th IEEE Control and Decision Conference*, Shanghai, China, December 2009, pp. 7872–7877.
[14] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. PO Box 391, Belmont, MA 02178: Athena Scientific, 2007, vol. I–II.
[15] J. Beardwood, J. Halton, and J. Hammersley, "The shortest path through many points," *Proceedings of the Cambridge Philoshopy Society*, vol. 55, pp. 299–327, 1959.
[16] G. Percus and O. C. Martin, "Finite size and dimensional dependence of the Euclidean traveling salesman problem," *Physical Review Letters*, vol. 76, no. 8, pp. 1188–1191, 1996.
[17] D. S. Johnson, L. A. McGeoch, and E. E. Rothberg, "Asymptotic experimental analysis for the held-karp traveling salesman bound," in *Proc. 7th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1996, pp. 341–350.
[18] R. Larson and A. Odoni, *Urban Operations Research*. Englewood Cliffs, NJ: Prentice Hall, 1981.
[19] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1996.
[20] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[22] R. Kershner, "The number of circles covering a set," *American Journal of Mathematics*, vol. 61, no. 3, pp. 665–671, July 1939.
[23] S. Lin and B. W. Kernighan, "An effective heuristic algorithm for the traveling-salesman problem," *Operations Research*, vol. 21, no. 2, pp. 498–516, 1973.
[24] K. Helsgaun, "An effective implementation of the lin-kernighan traveling salesman heuristic," *European Journal of Operational Research*, vol. 126, pp. 106–130, 2000.