Persistent Surveillance of Events with Unknown Rate Statistics

Cenk Baykal[1], Guy Rosman[1], Kyle Kotowick[1], Mark Donahue[2], and Daniela Rus[1]

[1] Massachusetts Institute of Technology (MIT), Cambridge MA 02139, USA,
{baykal, rosman, kotowick, rus}@csail.mit.edu,
[2] MIT Lincoln Laboratory, Lexington MA 02421, USA,
mark.donahue@ll.mit.edu

Abstract. We present a novel algorithm for persistent monitoring of stochastic events that occur at discrete locations in the environment with unknown event rates. Prior research on persistent monitoring assumes knowledge of event rates, which is often not the case in robotics applications. We consider the multi-objective optimization of maximizing the total number of events observed in a balanced manner subject to real-world autonomous system constraints. We formulate an algorithm that quantifies and leverages uncertainty over events statistics to greedily generate adaptive policies that simultaneously consider learning and monitoring objectives. We analyze the favorable properties of our algorithm as a function of monitoring cycles and provide simulation results demonstrating our methods effectiveness in real-world inspired monitoring applications.

Keywords: persistent monitoring, optimization and optimal control, probabilistic reasoning, mobile robots, sensor planning, machine learning

1 Introduction

Robotic surveillance missions often require a mobile robot to navigate an unknown environment and monitor stochastic events of interest over a long period of time. Equipped with limited a priori knowledge, the agent is tasked with exploring the environment by traveling from one landmark to the other and identifying regions of importance in an efficient way. The overarching monitoring objective is to observe as many events as possible in a uniform, balanced manner so that sufficient heterogeneous information can be collected across different parts of the environment. Applications include people and vehicle surveillance of friendly and unfriendly activity and environmental monitoring of natural phenomena.

In our formulation, we consider monitoring stochastic, instantaneous events of interest that occur at discrete stations, i.e., locations, over an infinite time horizon. Our robot is equipped with a limited-range sensor that can only record accurate measurements when the robot is stationary, e.g., a microphone on a

UAV. Hence, the robot must travel to each location and listen for events for a predetermined amount of time before traveling to another location.

We are given a fixed, cyclic path for the robot to traverse, but do not know the dwell time at each station. The persistent monitoring problem is to compute the optimal *observation time* for each station with respect to problem-specific optimality criteria. In particular, we consider maximizing the number of events observed in a maximally balanced way to be the overarching monitoring objective. Our multi-objective problem formulation extends previous work by relaxing the assumption that the rates of events are known, which introduces the notorious exploration and exploitation trade-off.

This paper contributes the following:

1. A novel per-cycle monitoring problem that hinges on an *uncertainty constraint*, a hard constraint that enables computation of policies conducive to balancing the exploration and exploitation trade-off.
2. A persistent monitoring algorithm with provable per-cycle guarantees that quantifies and employs the uncertainty over events' statistics to generate per-cycle optimal policies.
3. An analysis proving probabilistic bounds on the accuracy of rate approximations and the per-cycle optimality of the generated policies as a function of monitoring cycles.
4. Simulation results that characterize our algorithm's effectiveness in robotic surveillance scenarios and compare its performance to state-of-the-art monitoring algorithms.

1.1 Related Work

We build on important prior work in persistent surveillance, sensor scheduling, and machine learning. Robotic surveillance missions have been considered for a variety of applications and objectives such as ecological monitoring, underwater marine surveillance, and detection of natural phenomena [1–10]. Examples of monitoring objectives include facilitating high-value data collection for autonomous underwater vehicles [2], keeping a growing spatio-temporal field bounded using speed controllers [5], and generating the shortest watchman routes along which every point in a given space is visible [11].

Surveillance of discrete landmarks is of particular relevance to our work. Monitoring discrete locations such as buildings, windows, doors using a team of autonomous micro-aerial vehicles (MAVs) is considered in [3]. [12] presents different approaches to the min-max latency walk problem in a discrete setting. [13] extends this work to include multiple objectives, i.e. [13] considers the objective of minimizing the maximum latency and maximizing balance of events across stations using a single mobile robot. The authors show a reduction of the optimization problem to a quasi-convex problem and prove that a globally optimal solution can be computed in $O(\text{poly}(n))$ time where n is the number of discrete landmarks. Persistent surveillance in a discrete setting can be extended to the case of reasoning over different trajectories as shown in [5, 9, 14]. However, most prior work assumes that the rates of events are known prior to the surveillance mission, which is very often not the case in real world robotics applications. In this paper, we relax the assumption of known rates and present an algorithm with provable guarantees to generate policies conducive to learning event rates and optimizing the monitoring objectives.

[10] considers controlling multiple agents to minimize an uncertainty metric in the context of a 1D spatial domain. Decentralized approaches to controlling a network of robots for purposes of sensory coverage are investigated in [9], where a control law to drive a network of mobile robots to an optimal sensing configuration is presented. Persistent monitoring of dynamic environments has studied in [4, 5, 7]. For instance, [7] considers optimal sensing in a time-changing Gaussian Random Field and proposes a new randomized path planning algorithm to find the optimal infinite horizon trajectory. [15] presents a surveillance method based on Partially Observable Markov Decision Processes (POMDPs), however, POMDP-based approaches are often computationally intractable, especially when the action set includes continuous parameters, as in our case.

Persistent surveillance is closely related to sensor scheduling [16], sensor positioning [17], and coverage [18]. Previous approaches have considered persistent monitoring in the context of a mobile sensor [19]. Other related work includes variants and applications of the Orienteering Problem (OP) to generate informative paths that are constrained to a fixed length or time budget [20]. Yu et al. present an extension of OP to monitor spatially-correlated locations within a predetermined time [21]. In [22] and [23] the authors consider the OP problem in which the reward is a known function of the time spent at each point of interest. In contrast to our work, approaches in OP predominantly consider known environments and budget-constrained policies that visit each location at most once and optimize only a single objective.

The main challenge for the problem we address in this paper stems from the inherent exploration and exploitation trade-off, which has been rigorously analyzed in the form of regret bounds in Reinforcement Learning [24–26] and more relevantly, in Multi-armed Bandit (MAB) literature [27–29]. However, the traditional MAB problem considers minimizing regret with respect to the accumulated reward by appropriately pulling one of the $K \in \mathbb{N}_+$ levers at each discrete time step to obtain a stochastic reward that is generally assumed to bounded or subgaussian.

Our work differs from the canonical MAB formulation in that we consider a multi-objective optimization problem, i.e. we consider both the number and balance of event observations, in the face of travel costs, distributions with infinite support, cyclic policy structure, and continuous state and parameter space. To the best of our knowledge, this paper presents the first treatment of a MAB variant exhibiting all of the aforementioned complexities and an adaptive algorithm with provable guarantees as a function of monitoring cycles.

2 Problem Definition

Let there be $n \in \mathbb{N}_+$ spatially-distributed stations in the environment whose locations are known. At each station $i \in [n]$, stochastic events of interest occur according to a Poisson process with an unknown, station-specific rate parameter λ_i that is independent of other stations' rates. We assume that the robot executes a given cyclic path, taking $d_{i,j} > 0$ time to travel from station i to station j and let $D := \sum_{i=1}^{n-1} d_{i,i+1} + d_{n,1}$ denote the total travel time per cycle. The robot can only observe events at one station at any given time and cannot make observations while traveling.

We denote each complete traversal of the cyclic path as a *monitoring cycle*, indexed by $k \in \mathbb{N}_+$. We denote the observations times for all stations $\pi_k := (t_{1,k}, \ldots, t_{n,k})$ as the *monitoring policy* at cycle k. Our monitoring objective is to generate policies that maximize the number of events observed in a balanced manner across all stations within the allotted monitoring time $T_{\max}$ that is assumed to be unknown and unbounded. We introduce the function $f_{\text{obs}}(\Pi)$ that computes the total number of expected observations for a sequence of policies $\Pi := (\pi_k)_{k\in\mathbb{N}_+}$: $f_{\text{obs}}(\Pi) := \sum_{\pi_k} \sum_{i\in[n]} \mathbb{E}[N_i(\pi_k)]$, where $N_i(\pi_k)$ is the Poisson random variable, with realization $n_{i,k}$, denoting the number of events observed at station i under policy π_k and $\mathbb{E}[N_i(\pi_k)] := \lambda_i t_{i,k}$ by definition.

To reason about balanced attention, we let $f_{\text{bal}}(\Pi)$ denote as in [13] the expected observations ratio taken over the sequence of policies Π:

$$f_{\text{bal}}(\Pi) := \min_{i\in[n]} \frac{\sum_{\pi_k} \mathbb{E}[N_i(\pi_k)]}{\sum_{\pi_k} \sum_{j=1}^{n} \mathbb{E}[N_j(\pi_k)]}. \tag{1}$$

The *idealized* persistent surveillance problem is then:

Problem 1 (Idealized Persistent Surveillance Problem). Generate the optimal sequence of policies $\Pi^* = \operatorname{argmax}_{\Pi\in S} f_{\text{obs}}(\Pi)$ where S is the set of all possible policies that can be executed within the allotted monitoring time $T_{\max}$.

Generating the optimal solution Π^* at the beginning of the monitoring process is challenging due to the lack of knowledge regarding both the upper bound $T_{\max}$ and the station-specific rates. Hence, instead of optimizing the entire sequence of policies at once, we take a greedy approach and opt to subdivide the problem into multiple, *per-cycle* optimization problems. For each cycle $k \in \mathbb{N}_+$, our goal is to adaptively generate the policy π_k^* that optimizes the monitoring objectives with respect to the most up-to-date knowledge of event statistics. We let $\hat{f}_{\text{bal}}$ represent the per-cycle counterpart of f_{bal}

$$\hat{f}_{\text{bal}}(\pi_k) := \min_{i\in[n]} \frac{\mathbb{E}[N_i(\pi_k)]}{\sum_{j=1}^{n} \mathbb{E}[N_j(\pi_k)]}.$$

We note that the set of policies that optimize $\hat{f}_{\text{bal}}$ is uncountably infinite and policies of all possible lengths belong to this set [13]. To generate observation times that are conducive to exploration, we impose the hard constraint $t_{i,k} \geq t_{i,k}^{\text{low}}$

on each observation time, where $t_{i,k}^{\text{low}}$ is a lower bound that is a function of our uncertainty of the rate parameter λ_i (see Sec. 3). The optimization problem that we address in this paper is then of the following form:

Problem 2 (Per-cycle Monitoring Optimization Problem). At each cycle $k \in \mathbb{N}_+$, generate a per-cycle optimal policy π_k^* satisfying

$$\pi_k^* \in \underset{\pi_k}{\operatorname{argmax}}\, \hat{f}_{\text{bal}}(\pi_k) \quad \text{s.t.} \quad \forall i \in [n] \quad t_{i,k} \geq t_{i,k}^{\text{low}}. \tag{2}$$

3 Methods

In this section, we present our monitoring algorithm and detail the main subroutines employed by our method to generate dynamic, adaptive policies and interleave learning and approximating of event statistics with policy execution.

3.1 Algorithm for Monitoring Under Unknown Event Rates

The entirety of our persistent surveillance method appears as Alg. 1 and employs Alg. 2 as a subprocedure to generate adaptive, uncertainty-reducing policies for each monitoring cycle.

Algorithm 1: Core monitoring algorithm

$\alpha_i \leftarrow \alpha_{i,0}$; $\beta_i \leftarrow \beta_{i,0}$; $\hat{\lambda}_i \leftarrow \alpha_i/\beta_i$;
Loop
  $\pi^* \leftarrow \text{Algorithm2}(\alpha_i, \beta_i)$;
  for $i \in [n]$ **do**
    Observe for t_i^* time to obtain n_i observations;
    $\alpha_i \leftarrow \alpha_i + n_i$,; $\beta_i \leftarrow \beta_i + t_i^*$;
    $\hat{\lambda}_i \leftarrow \alpha_i/\beta_i$;

Algorithm 2: Generates a per-cycle optimal policy π^*

for $i \in [n]$ **do**
  Compute t_i^{low} using (8);
$\pi_{\text{low}} \leftarrow (t_1^{\text{low}}, \ldots, t_n^{\text{low}})$;
$N_{\max} \leftarrow \max_{i \in [n]} t_i^{\text{low}} \alpha_i/\beta_i$;
for $i \in [n]$ **do**
  Compute t_i^* using $N_{\max}$ according to (10);
return $\pi^* = (t_1^*, \ldots, t_n^*)$;

3.2 Learning and Approximating Event Statistics

We use the Gamma distribution as the conjugate prior for each rate parameter because it provides a closed-form expression for updating the posterior distribution after observing events. We let $\text{Gamma}(\alpha_i, \beta_i)$ denote the Gamma distribution with hyper-parameters $\alpha_i, \beta_i \in \mathbb{R}_+$ that are initialized to user-specified values $\alpha_{i,0}, \beta_{i,0}$ for all stations i and are updated as new events are observed.

For any arbitrary number of events $n_{i,k} \in \mathbb{N}$ observed in $t_{i,k}$ time, the posterior distribution is given by $\text{Gamma}(\alpha_i + n_{i,k}, \beta_i + t_{i,k})$ for any arbitrary station $i \in [n]$ and cycle $k \in \mathbb{N}_+$. For notational convenience, we let $X_i^k := (n_{i,k}, t_{i,k})$

represent the summary of observations for cycle $k \in \mathbb{N}_+$ and define the aggregated set of observations up to any arbitrary cycle as $X_i^{1:k} := \{X_i^1, X_i^2, \ldots, X_i^k\}$ for all stations $i \in [n]$. After updating the posterior distribution using the hyperparameters, i.e. $\alpha_i \leftarrow \alpha_i + n_{i,k}$, $\beta_i \leftarrow \beta_i + t_{i,k}$, we use the maximum probabiliy estimate of the rate parameter λ_i, denoted by $\hat{\lambda}_{i,k}$ for any arbitrary station i:

$$\hat{\lambda}_{i,k} := E[\lambda_i | X_i^{1:k}] = \frac{\alpha_{i,0} + \sum_{k=1}^{n} n_{i,k}}{\beta_{i,0} + \sum_{k=1}^{n} t_{i,k}} = \frac{\alpha_i}{\beta_i}. \tag{3}$$

3.3 Per-cycle Optimization and the Uncertainty Constraint

Inspired by confidence-based MAB approaches [28–30], our algorithm adaptively computes policies by reasoning about the uncertainty of our rate approximations. We introduce the *uncertainty-constraint*, an optimization constraint that enables the generating a station-specific observation time based on uncertainty of each station's parameter. The constraint helps bound the policy lengths adaptively over the course of the monitoring process so that approximation uncertainty decreases uniformly across all stations. We use the posterior variance of the rate parameter λ_i, $\mathrm{Var}(\lambda_i | X_i^{1:k})$, as our uncertainty measure of each station i after executing k cycles. We note that in our Gamma-Poisson model, $\mathrm{Var}(\lambda_i | X_i^{1:k}) := \frac{\alpha_i}{\beta_i^2}$ by definition of the Gamma distribution.

Uncertainty constraint For a given $\delta \in (0,1), \epsilon \in \left(0, 2(1 + 2e^{1/\pi})^{-1}\right)$ and arbitrary cycle $k \in \mathbb{N}_+$, π_k must satisfy the following

$$\forall i \in [n] \quad \mathbb{P}\left(\mathrm{Var}(\lambda_i | X_i^{1:k}, \pi_k) \leq \delta \mathrm{Var}(\lambda_i | X_i^{1:k-1}) \middle| X_i^{1:k-1}\right) > 1 - \epsilon. \tag{4}$$

We incorporate the uncertainty constraint as a hard constraint and recast the per-cycle optimization problem from Sec. 2 in terms of the optimization constraint.

Problem 3 (Recast Per-cycle Monitoring Optimization Problem). For each monitoring cycle $k \in \mathbb{N}_+$ generate a per-cycle optimal policy π_k^* that simultaneously satisfies the uncertainty constraint (4) and maximizes the balance of observations, i.e.,

$$\begin{aligned} &\pi_k^* \in \operatorname*{argmax}_{\pi_k} \hat{f}_{\mathrm{bal}}(\pi_k) \qquad (5) \\ &\text{s.t.} \;\; \forall i \in [n] \quad \mathbb{P}\left(\mathrm{Var}(\lambda_i | X_i^{1:k}, \pi_k) \leq \delta \mathrm{Var}(\lambda_i | X_i^{1:k-1}) \middle| X_i^{1:k-1}\right) > 1 - \epsilon. \end{aligned}$$

3.4 Controlling Approximation Uncertainty

We outline an efficient method for generating observation times that satisfy the uncertainty constraint and induce uncertainty reduction at each monitoring cycle. We begin by simplifying (4) to obtain

$$\mathbb{P}\left(N_i(t_{i,k}) \leq \delta k(t_{i,k}) | X_i^{1:k-1}\right) > 1 - \epsilon \tag{6}$$

where $N_i(t_{i,k}) \sim \text{Pois}(\lambda_i t_{i,k})$ by definition of Poisson process and $k(t_{i,k}) := \delta\alpha_i(\beta_i+t_{i,k})^2/\beta_i^2-\alpha_i$. Given that the distribution of the random variable $N_i(t_{i,k})$ is a function of the unknown parameter λ_i, we use interval estimation to reason about the cumulative probability distribution of $N_i(t_{i,k})$.

For each monitoring cycle $k \in \mathbb{N}_+$ we utilize previously obtained observations $X_i^{1:k-1}$ to construct the equal-tail credible interval for each parameter λ_i, $i \in [n]$ defined by the open set $(\lambda_i^{\text{l}}, \lambda_i^{\text{u}})$ such that

$$\forall \lambda_i \in \mathbb{R}_+ \quad \mathbb{P}\left(\lambda_i \in (\lambda_i^{\text{l}}, \lambda_i^{\text{u}}) |\, X_i^{1:k-1}\right) = 1 - \epsilon$$

where $\epsilon \in (0, 2(1+2e^{1/\pi})^{-1})$. By leveraging the relation between the Poisson and Gamma distributions, we compute the end-points of the equal-tailed credible interval:

$$\lambda_i^{\text{l}} := \frac{Q^{-1}(\alpha_i, \frac{\epsilon}{2})}{\beta_i} \quad \lambda_i^{\text{u}} := \frac{Q^{-1}(\beta_i, 1-\frac{\epsilon}{2})}{\beta_i}$$

where $Q^{-1}(a, s)$ is the Gamma quantile function and α_i and β_i are the posterior hyper-parameters after observations $X_i^{1:k-1}$. Given that we desire our algorithm to be *cycle-adaptive* (Sect. 2), we seek to generate the minimum feasible observation time satisfying the uncertainty constraint for each station $i \in [n]$, i.e.,

$$t_{i,k}^{\text{low}} = \inf_{t_{i,k} \in \mathbb{R}_+} t_{i,k} \quad \text{s.t.} \quad \mathbb{P}\left(N_{i,k}(t_{i,k}) \leq \delta k(t_{i,k}) | X_i^{1:k-1}\right) > 1 - \epsilon. \tag{7}$$

For computational efficiency in the optimization above, we opt to use a tight and efficiently-computable lower bound for approximating the Poisson cumulative distribution function that improves upon the Chernoff-Hoeffding inequalities by a factor of at least two [31]. As demonstrated rigorously in Lemma 1, the expression for an approximately-minimal observation time satisfying constraint (4) is given by

$$t_{i,k}^{\text{low}} := t \in \mathbb{R}_+ \;|\; D_{\text{KL}}\big(\text{Pois}(\lambda_i^{\text{u}} t) \,||\, \text{Pois}(k(t))\big) - W_\epsilon = 0 \tag{8}$$

where $D_{\text{KL}}\big(\text{Pois}(\lambda_1) \,||\, \text{Pois}(\lambda_2)\big)$ is the Kullback-Leibler (KL) divergence between two Poisson distributions with mean λ_1 and λ_2 respectively and W_ϵ is defined using the Lambert W function [32]: $W_\epsilon = \frac{1}{2} W\big(\frac{(\epsilon-2)^2}{2\epsilon^2\pi}\big)$. An appropriate value for $t_{i,k}^{\text{low}}$ can be obtained by invoking a root-finding algorithm such as Brent's method on the equation above [33].

The constant factor $\delta \in (0, 1)$ is the exploration parameter that influences the rate of uncertainty decay. Low values of δ lead to lengthy, and hence less cycle-adaptive policies, whereas high values lead to shorter, but also less efficient policies due to incurred travel time. We found that values generated by a logistic function with respect to problem-specific parameters as input worked well in practice for up to 50 stations: $\delta(n) := (1 + \exp(-n/D))^{-1}$ where D is the total travel time per cycle.

3.5 Generating Balanced Policies that Consider Approximation Uncertainty

We build upon the method introduced in the previous section to generate a policy π_k^* that simultaneously satisfies the uncertainty constraint and balances attention given to all stations in approximately the minimum time possible. The key insight is that the value of $t_{i,k}^{\mathrm{low}}$ given by (8) acts as a lower bound on the observation time for each station $i \in [n]$ for satisfying the uncertainty constraint (see Lemma 2). We also leverage the following fact from [13] regarding the optimality of the balance objective for a policy π_k:

$$\mathbb{E}[N_1(\pi_k)] = \cdots = \mathbb{E}[N_n(\pi_k)] \Leftrightarrow \pi_k \in \operatorname*{argmax}_{\pi} \hat{f}_{\mathrm{bal}}(\pi). \tag{9}$$

We use a combination of this result and the fact that any observation time satisfying $t_{i,k} \geq t_{i,k}^{\mathrm{low}}$ also satisfies the uncertainty constraint to arrive at an expression for the optimal observation time for each station. In constructing the optimal policy $\pi_k^* = (t_{1,k}^*, \ldots, t_{n,k}^*)$, we first identify the "bottleneck" value, $N_{\max}$, which is computed using the lower bounds for each $t_{i,k}$, i.e., $N_{\max} := \max_{i \in [n]} \hat{\lambda}_{i,k} t_{i,k}^{\mathrm{low}}$. Given (9), we use the bottleneck value $N_{\max}$ to set the value of each observation time $t_{i,k}^*$ appropriately so that each $t_{i,k}^* \geq t_{i,k}^{\mathrm{low}}$ and the policy defined by $\pi_k^* := (t_{1,k}^*, \ldots, t_{n,k}^*)$ maximizes the balance objective function. Namely, the optimal observation times for all stations which constitute the per-cycle optimal policy $\pi_k^* = (t_{1,k}^*, \ldots, t_{n,k}^*)$ are computed individually:

$$\forall k \in \mathbb{N}_+ \quad \forall i \in [n] \quad t_{i,k}^* := \frac{N_{\max}}{\hat{\lambda}_{i,k}} = N_{\max} \frac{\beta_i}{\alpha_i}. \tag{10}$$

4 Analysis

The outline of results in this section is as follows: we begin by proving the uncertainty-reducing property and per-cycle optimality of policies generated by Alg. 2 with respect to the rate approximations. We present a probabilistic bound on posterior variance and error of our rate approximations with respect to the ground-truth rates by leveraging the properties of each policy. We use the previous results to establish a probabilistic bound on the per-cycle optimality of any arbitrary policy generated by Alg. 2 with respect to the ground-truth optimal solution of Problem 3.

We impose the following assumption on user-specified input.

Assumption 1. *The parameters* ϵ *and* δ *are confined to the intervals* $(0, 2(1 + 2e^{1/\pi})^{-1})$ *and* $(0,1)$ *respectively, i.e.,* $\epsilon \in (0, 2(1 + 2e^{1/\pi})^{-1}),\ \delta \in (0,1)$.

A policy π_k is said to be *approximately-optimal* at cycle $k \in \mathbb{N}_+$ if π_k is an optimal solution to Problem 3 with respect to the rate approximations $\hat{\lambda}_{1,k}, \ldots, \hat{\lambda}_{n,k}$, i.e., if it is optimal under the approximation of expectation: $\mathbb{E}[N_i(\pi_k)] \approx \hat{\lambda}_{i,k} t_{i,k}\ \forall i \in [n]$. In contrast, a policy π_k is *ground-truth optimal* if it is an optimal solution to Problem 3 with respect to the ground-truth rates $\lambda_1, \ldots, \lambda_n$. For sake of notational brevity, we introduce the function $g : \mathbb{R} \to \mathbb{R}$ denoting

$$g(x) := 1 - \frac{e^{-x}}{\max\left\{2, 2\sqrt{\pi x}\right\}},$$

and note the bound established by [31] for a Poisson random variable Y with mean m and $k \in \mathbb{R}_+$ such that $k \geq m$

$$\mathbb{P}\left(Y \leq k\right) > g\big(D_{\mathrm{KL}}(\mathrm{Pois}(m) \,||\, \mathrm{Pois}(k))\big). \tag{11}$$

We begin by proving that each policy generated by Alg. 2 is optimal with respect to the per-cycle optimization problem (Problem 3).

Lemma 1 (Satisfaction of the uncertainty constraint). *The observation time $t_{i,k}^{low}$ given by* (8) *satisfies the uncertainty constraint* (4) *for any arbitrary station $i \in [n]$ and iteration $k \in \mathbb{N}_+$.*

Proof. We consider the left-hand side of (6) from Sect. 3 and marginalize over the unknown parameter $\lambda_i \in \mathbb{R}_+$:

$$\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}\right) = \int_0^\infty \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda\right) \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda$$

where the probability is with respect to the random variable $N_i(t_{i,k}) \sim \mathrm{Pois}(\lambda t_{i,k})$ $\forall \lambda \in \mathbb{R}_+$ by definition of a Poisson process with parameter λ. Using the equal-tails credible interval constructed in Alg. 2, i.e. the interval $(\lambda_i^{\mathrm{l}}, \lambda_i^u)$ satisfying

$$\forall i \in [n] \quad \forall \lambda_i \in \mathbb{R}_+ \; \mathbb{P}\left(\lambda_i^{\mathrm{l}} > \lambda_i | X_i^{1:k-1}\right) = \mathbb{P}\left(\lambda_i^{\mathrm{u}} < \lambda_i | X_i^{1:k-1}\right) = \frac{\epsilon}{2},$$

we establish the inequalities:

$$\begin{aligned}
\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}\right) &> \int_0^{\lambda_i^u} \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda\right) \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda \\
&\geq \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^u\right) \int_0^{\lambda_i^u} \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda \\
&= (1 - \frac{\epsilon}{2})\, \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^u\right). \qquad (12)
\end{aligned}$$

where we utilized the fact that $\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^{\mathrm{u}}\right)$ is monotonically decreasing with respect to λ. By construction, $t_{i,k}^{\mathrm{low}}$ satisfies $D_{\mathrm{KL}}(\mathrm{Pois}(\lambda_i^{\mathrm{u}} t_{i,k}^{\mathrm{low}}) \,||\, \mathrm{Pois}(k(t_{i,k}^{\mathrm{low}}))) = W_\epsilon$ which yields $1 - g(W_\epsilon) = 1 - \frac{\epsilon}{2-\epsilon}$ by definition and thus by (11) we have:

$$\mathbb{P}\left(N_i(t_{i,k}^{\mathrm{low}}) \leq k(t_{i,k}^{\mathrm{low}}) | X_i^{1:k-1}, \lambda_i^{\mathrm{u}}\right) > 1 - g(W_\epsilon) = 1 - \frac{\epsilon}{2 - \epsilon}.$$

Combining this inequality with the expression of (12) establishes the result. □

Lemma 2 (Monotonicity of solutions satisfying (4)**).** *For any arbitrary station $i \in [n]$ and monitoring cycle $k \in \mathbb{N}_+$, the observation time $t_{i,k}$ satisfying $t_{i,k} \geq t_{i,k}^{low}$, where $t_{i,k}^{low}$ is given by* (8)*, satisfies the uncertainty constraint.*

Theorem 1 (Per-cycle approximate-optimality of solutions). *For any arbitrary cycle $k \in \mathbb{N}_+$, the policy $\pi_k^* := (t_{1,k}^*, \ldots, t_{n,k}^*)$ generated by Alg. 2 is an approximately-optimal solution with respect to Problem 3.*

Proof. By definition of (10), we have for any arbitrary cycle $k \in \mathbb{N}_+$ and station $i \in [n]$, $t^*_{i,k} = N_{\max}/\hat{\lambda}_{i,k} \geq t^{\text{low}}_{i,k}$ by definition of $N_{\max} := \max_{i\in[n]} \hat{\lambda}_{i,k} t^{\text{low}}_{i,k}$. Applying Lemma 2 and observing that

$$\hat{\lambda}_{1,k} t^*_{1,k} = N_{\max},\ \hat{\lambda}_{2,k} t^*_{2,k} = N_{\max},\ \ldots,\ \hat{\lambda}_{n,k} t^*_{n,k} = N_{\max}$$

implies that the uncertainty constraint is satisfied for all stations $i \in [n]$ and that $\pi^*_k \in \operatorname{argmax}_{\pi_k} \hat{f}_{\text{bal}}(\pi_k)$, which establishes the optimality of π_k with respect to Problem 3. □

Using the fact that each policy satisfies the uncertainty constraint, we establish probabilistic bounds on uncertainty, i.e. posterior variance, and rate approximations.

Lemma 3 (Bound on posterior variance). *After executing an arbitrary number of cycles $k \in \mathbb{N}_+$, the posterior variance $Var(\lambda_i|X_i^{1:k})$ is bounded above by $\delta^k\, Var(\lambda_i)$ with probability at least $(1-\epsilon)^k$, i.e.,*

$$\forall i \in [n] \quad \forall k \in \mathbb{N}_+ \quad \mathbb{P}\left(Var(\lambda_i|X_i^{1:k}) \leq \delta^k\, Var(\lambda_i)|X_i^{1:k}\right) > (1-\epsilon)^k$$

for all stations $i \in [n]$ where $Var(\lambda_i) := \alpha_{i,0}/\beta^2_{i,0}$ is the prior variance.

Proof. Iterative application of the inequality $\operatorname{Var}(\lambda_i|X_i^{1:k}) \leq \delta \operatorname{Var}(\lambda_i|X_i^{1:k-1})$ each with probability $1-\epsilon$ by the uncertainty constraint (4) yields the result. □

Corollary 1 (Bound on variance of the posterior mean). *After executing an arbitrary number of cycles $k \in \mathbb{N}_+$, the variance of our approximation $Var\big(\hat{\lambda}_{i,k}|X_i^{1:k-1}\big)$ is bounded above by $\delta^{k-1}\, Var(\lambda_i)$ with probability greater than $(1-\epsilon)^{k-1}$, i.e.,*

$$\forall i \in [n] \quad \mathbb{P}\left(Var(\hat{\lambda}_{i,k}|X_i^{1:k-1}) \leq \delta^{k-1}\, Var(\lambda_i)|X_i^{1:k-1}\right) > (1-\epsilon)^{k-1}.$$

Proof. Application of the law of total conditional variance and invoking Lemma 3 yields the result. □

Theorem 2 (ξ-bound on approximation error). *For all $\xi \in \mathbb{R}_+$ and cycles $k \in \mathbb{N}_+$, the inequality $|\hat{\lambda}_{i,k} - \lambda_i| < \xi$ holds with probability at least $(1-\epsilon)^{k-1}(1 - \frac{\delta^{k-1}\, Var(\lambda_i)}{\xi^2})$, i.e.,*

$$\forall i \in [n] \quad \mathbb{P}\left(|\hat{\lambda}_{i,k} - \lambda_i| < \xi | X_i^{1:k-1}\right) > (1-\epsilon)^{k-1}\big(1 - \frac{\delta^{k-1}\, Var(\lambda_i)}{\xi^2}\big).$$

Proof. Applying Corollary 1 and using Chebyshev's inequality gives the result. □

Theorem 3 (Δ-bound on optimality with respect to Problem 3). *For any $\xi_i \in \mathbb{R}_+$, $i \in [n], k \in \mathbb{N}_+$, given that $|\hat{\lambda}_{i,k} - \lambda_i| \in (0, \xi_i)$ with probability as given in Theorem 2, let $\sigma_{min} := \sum_{i=1}^n (\lambda_i - \xi_i)^{-1}$ and $\sigma_{max} := \sum_{i=1}^n (\lambda_i + \xi_i)^{-1}$. Then, the objective value of the policy π^*_k at iteration k is within a factor of Δ of the ground-truth optimal solution, where $\Delta := \frac{\sigma_{min}}{\sigma_{max}}$ with probability greater than $(1-\epsilon)^{n(k-1)}\big(1 - \frac{\delta^{k-1}\, Var(\lambda_i)}{\xi^2}\big)^n$.*

Proof. Note that for any arbitrary total observation time $T \in \mathbb{R}_+$, a policy $\pi_k = (t^*_{1,k}, \ldots, t^*_{n,k})$ satisfying

$$\forall i \in [n] \quad t^*_{i,k} := \frac{T}{\lambda_i \sum_{l=1}^{n} \frac{1}{\lambda_l}}. \tag{13}$$

optimizes the balance objective function $\hat{f}_{\text{bal}}$ [13]. Using the fact that $|\hat{\lambda}_{i,k} - \lambda_i| < \xi_i$ with probability given by Theorem 2, we arrive at the following inequality for $\hat{f}_{\text{bal}}(\pi^*_k)$

$$\hat{f}_{\text{bal}}(\pi^*_k) > \frac{\frac{T}{\sum_{l=1}^{n}(\lambda_i+\xi_l)^{-1}}}{\frac{nT}{\sum_{l=1}^{n}(\lambda_l-\xi_l)^{-1}}} = \frac{\sum_{l=1}^{n}(\lambda_l - \xi_l)^{-1}}{n\sum_{l=1}^{n}(\lambda_l + \xi_l)^{-1}}$$

with probability at least $(1-\epsilon)^{n(k-1)}\left(1 - \frac{\delta^{k-1}\mathrm{Var}(\lambda_i)}{\xi^2}\right)^n$. □

5 Results

We evaluate the performance of Alg. 1 in two simulated scenarios modeled after real-world inspired monitoring tasks: (i) a synthetic simulation in which events at each station precisely follow a station-specific Poisson process and (ii) a scenario simulated in Armed Assault (ARMA) [34], a military simulation game, involving detections of suspicious agents. We note the statistics do not match our assumed Poisson model, and yet our algorithm performs well compared to other approaches. We compare Alg. 1 to the following monitoring algorithms:

1. Equal Time, Min. Delay (ETMD): computes the total cycle time to minimize latency T_{obs} [13] and partitions T_{obs} evenly across all stations.
2. Bal. Events, Min. Delay (BEMD): the algorithm introduced by [13] which generates policies that minimize latency and maximize observation balance.
3. Incremental Search, Bal. Events (ISBE): generates policies to maximize balance that increase in length by a fixed amount $\Delta_{\text{obs}} \in \mathbb{R}_+$ after each cycle.
4. Oracle Algorithm (Oracle Alg.): an omniscient algorithm assuming perfect knowledge of ground-truth rates and monitoring time $T_{\max}$ where each observation time is generated according to (13).

5.1 Synthetic Scenario

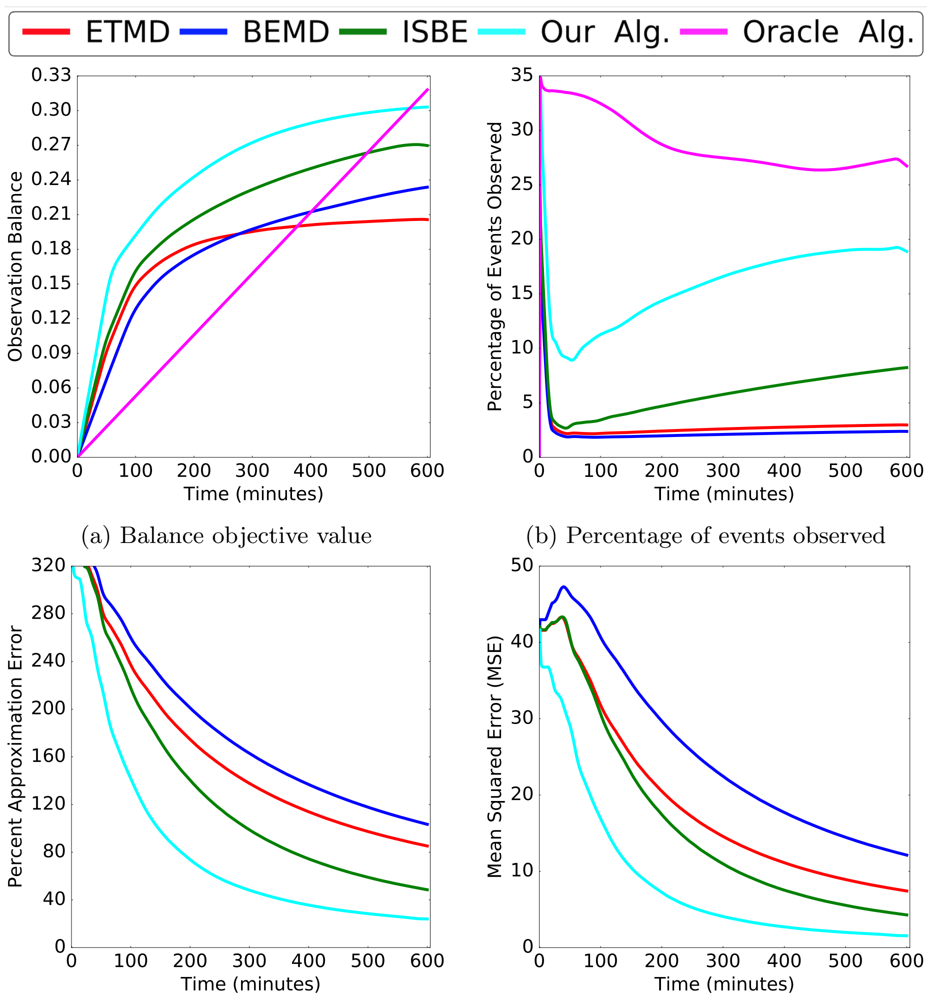
We consider the monitoring scenario involving the surveillance of events in three discrete stations over a monitoring period of 10 hours. We characterize the average performance of each monitoring algorithm with respect to 10,000 randomly generated problem instances with the following statistics:

1. Prior hyper-parameters: $\alpha_{i,0} \sim \text{Uniform}(1, 20)$ and $\beta_{i,0} \sim \text{Uniform}(0.75, 1.50)$.
2. Rate parameter of each station: $\mu_{\lambda_i} = 2.23$ and $\sigma_{\lambda_i} = 1.02$ events per minute.
3. Initial percentage error of the rate estimate $\lambda_{i,0}$, denoted by ρ_i: $\mu_{\rho_i} = 358.29\%$ and $\sigma_{\rho_i} = 221.32\%$.

4. Travel cost from station i to another j: $\mu_{d_{i,j}} = 9.97$ and $\sigma_{d_{i,j}} = 2.90$ minutes.

where μ and σ refer to standard deviation and variance of each parameter respectively and the transient events at each station $i \in [n]$ are simulated precisely according to $\text{Pois}(\lambda_i)$.

The performance of each algorithm with respect to the the monitoring objectives defined in Sect. 2 is shown in Figs. 1a and 1b respectively. The figures show that our algorithm is able to generate efficient policies that enable the robot



(a) Balance objective value

(b) Percentage of events observed

(c) Percentage error

(d) Mean squared error (MSE)

Fig. 1: Results of the synthetic simulation averaged over 10,000 trials that characterize and compare our algorithm to the four monitoring algorithms in randomized environments containing three discrete stations.

to observe significantly more events that achieve a higher balance in comparison to those computed by other algorithms (with exception of Oracle Alg.) at all times of the monitoring process. Figs. 1c and 1d depict the efficiency of each monitoring algorithm in rapidly learning the events' statistics and generating accurate approximations. The error plots show that our algorithm achieves lower measures of error at any given time in comparison to those of other algorithms and supports our method's practical efficiency in generating exploratory policies conducive to rapidly obtaining accurate approximations of event statistics. Figs. 1a-1d show our algorithm's dexterity in balancing the inherent trade-off between exploration vs. exploitation.

Fig. 2: Viewpoints from two stations in the ARMA simulation of the yellow backpack scenario. Agents wearing yellow backpacks whose detections are of interest appear in both figures.

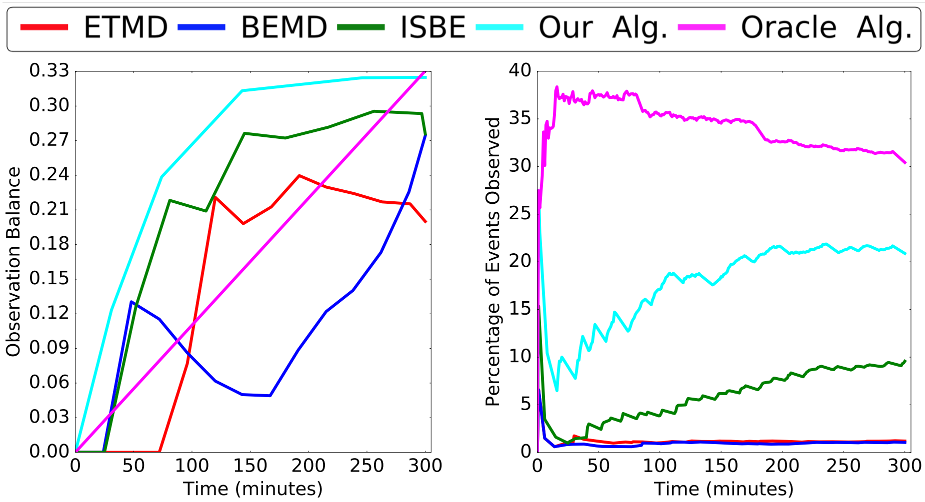
5.2 Yellow Backpack Scenario

In this subsection, we consider the evaluation of our monitoring algorithm in a real-world inspired scenario, labeled the *yellow backpack scenario*, that entails monitoring of suspicious events that do not adhere to the assumed Poisson model (Sect. 2). Using the military strategy game ARMA, we simulate human agents that wander around randomly in a simulated town. A subset of the agents wear yellow backpacks (see Fig. 2). Under this setting, our objective is to optimally monitor the yellow backpack-wearing agents using three predesignated viewpoints, i.e. stations. We considered a monitoring duration of 5 hours under the following simulation configuration:

1. Environment dimensions: 250 meters x 250 meters ($62,500$ meters^2).
2. Number of agents with a yellow backpack: 10 out of 140 ($\approx 7.1\%$ of agents).
3. Travel cost (minutes): $d_{1,2} = 3$, $d_{2,3} = 2$, $d_{3,1} = 12$.

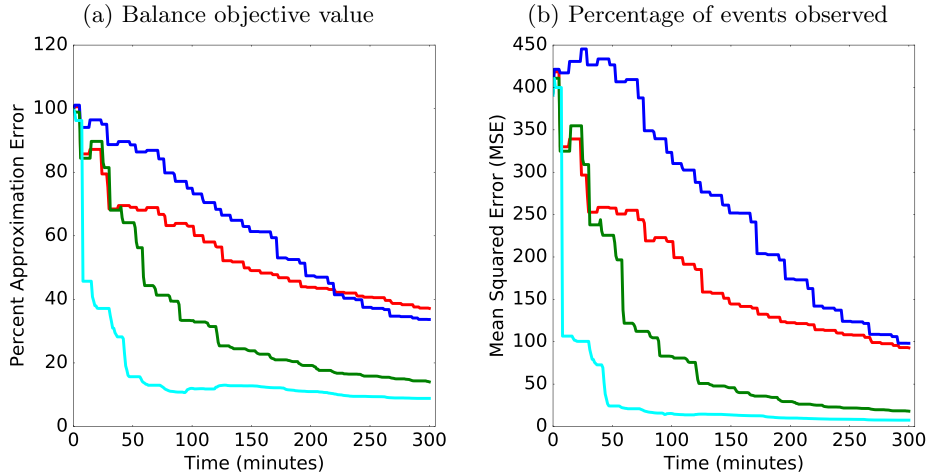
We used the Faster Region-based Convolutional Neural Network (Faster R-CNN, [35]) for recognizing yellow backpack-wearing agents in real-time at a frequency of 1 Hertz. We ran the simulation for a sufficiently long time in order to obtain estimates for the respective ground-truth rates of 23.3, 20.3, and

18.5 yellow backpack recognitions per minute, which were used to generate Figs. 3c and 3d. The results of the yellow backpack scenario, shown in Figs. 3a-3d, tell the same story as did the results of the synthetic simulation. We note that at all instances of the monitoring process, our approach that leverages uncertainty estimates outperforms others in generating balanced policies conducive to efficiently observing more events and obtaining accurate rate approximations.



(a) Balance objective value

(b) Percentage of events observed



(c) Percentage error

(d) Mean squared error (MSE)

Fig. 3: The performance of each monitoring algorithm evaluated in the ARMA-simulated yellow backpack scenario.

6 Conclusion

In this paper, we presented a novel algorithm with provable guarantees for monitoring stochastic, transient events that occur at discrete stations over a long period of time. The algorithm developed in this paper advances the state of the art in persistent surveillance by removing the assumption of known event rates. Our simulation experiments show that our approach has potential applications to important real world scenarios such as detection and tracking efforts at a large scale. We conjecture that our algorithm can be extended to persistent surveillance of events in dynamic environments where event statistics are both unknown and time-varying.

Acknowledgments This material is based upon work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

References

1. Jonathan Binney, Andreas Krause, and Gaurav S Sukhatme. Informative path planning for an autonomous underwater vehicle. In *ICRA*, pages 4791–4796, 2010.
2. Ryan N Smith, Mac Schwager, Stephen L Smith, et al. Persistent ocean monitoring with underwater gliders: Adapting sampling resolution. *Journal of Field Robotics*, 28(5):714–741, 2011.
3. Nathan Michael, Ethan Stump, and Kartik Mohta. Persistent surveillance with a team of mavs. In *IROS*, 2011.
4. Stephen L Smith, Mac Schwager, and Daniela Rus. Persistent robotic tasks: Monitoring and sweeping in changing environments. *T-RO*, 28(2):410–426, 2012.
5. Daniel E. Soltero, Mac Schwager, and Daniela Rus. Generating informative paths for persistent sensing in unknown environments. In *IROS*, pages 2172–2179, 2012.
6. Geoffrey A Hollinger, Sunav Choudhary, Parastoo Qarabaqi, et al. Underwater data collection using robotic sensor networks. *JSAC*, 30(5):899–911, 2012.
7. Xiaodong Lan and Mac Schwager. Planning periodic persistent monitoring trajectories for sensing robots in gaussian random fields. In *ICRA*, pages 2415–2420. IEEE, 2013.
8. Geoffrey A Hollinger and Gaurav S Sukhatme. Sampling-based motion planning for robotic information gathering. In *RSS*, pages 72–983. Citeseer, 2013.
9. Mac Schwager, Daniela Rus, and Jean-Jacques E. Slotine. Decentralized, adaptive coverage control for networked robots. *IJRR*, 28(3):357–375, 2009.
10. Christos Cassandras, Xuchao Lin, and Xuchu Ding. An optimal control approach to the multi-agent persistent monitoring problem. *Automatic Control, IEEE Transactions on*, 58(4):947–961, 2013.
11. Wei-pang Chin and Simeon Ntafos. Optimum watchman routes. *ISO*, 28(1):39–44, 1988.
12. Soroush Alamdari, Elaheh Fata, and Stephen L Smith. Persistent monitoring in discrete environments: Minimizing the maximum weighted latency between observations. *IJRR*, 33(1):138–154, 2014.

13. J. Yu, S. Karaman, and D. Rus. Persistent monitoring of events with stochastic arrivals at multiple stations. *T-RO*, 31(3):521–535, 2015.
14. Daniel E Soltero, Mac Schwager, and Daniela Rus. Decentralized path planning for coverage tasks using gradient descent adaptive control. *IJRR*, page 0278364913497241, 2013.
15. Jesus Capitan, Luis Merino, and Anibal Ollero. Decentralized cooperation of multiple uas for multi-target surveillance under uncertainties. In *ICUAS*, pages 1196–1202, 2014.
16. Ying He and Edwin KP Chong. Sensor scheduling for target tracking in sensor networks. In *ICDC*, volume 1, pages 743–748. IEEE, 2004.
17. Alfred O Hero III, Christopher M Kreucher, and Doron Blatt. Information theoretic approaches to sensor management. In *Foundations and applications of sensor management*, pages 33–57. Springer, 2008.
18. Yoav Gabriely and Elon Rimon. Competitive on-line coverage of grid environments by a mobile robot. *Computational Geometry*, 24(3):197–224, 2003.
19. Jerome Le Ny, Munther A Dahleh, Eric Feron, et al. Continuous path planning for a data harvesting mobile server. In *ICDC*, pages 1489–1494. IEEE, 2008.
20. Aldy Gunawan, Hoong Chuin Lau, and Pieter Vansteenwegen. Orienteering problem: A survey of recent variants, solution approaches and applications. *European Journal of Operational Research*, 2016.
21. Jingjin Yu, Mac Schwager, and Daniela Rus. Correlated orienteering problem and its application to informative path planning for persistent monitoring tasks. In *IROS*, pages 342–349. IEEE, 2014.
22. Güneş Erdoğan and Gilbert Laporte. The orienteering problem with variable profits. *Networks*, 61(2):104–116, 2013.
23. Jingjin Yu, Javed Aslam, Sertac Karaman, et al. Anytime planning of optimal schedules for a mobile sensing robot. In *IROS*, pages 5279–5286. IEEE, 2015.
24. P Ortner and R Auer. Logarithmic online regret bounds for undiscounted reinforcement learning. *NIPS*, 19:49, 2007.
25. Thomas Jaksch, Ronald Ortner, and Peter Auer. Near-optimal regret bounds for reinforcement learning. *JMLR*, 11(Apr):1563–1600, 2010.
26. Kenji Kawaguchi. Bounded optimal exploration in MDP. In *AAAI*, pages 1758–1764. AAAI Press, 2016.
27. Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.
28. Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *JMLR*, 47(2-3):235–256, 2002.
29. Peter Auer and Ronald Ortner. UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica*, 61(1-2):55–65, 2010.
30. Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *JMLR*, 3(Nov):397–422, 2002.
31. Michael Short. Improved inequalities for the Poisson and binomial distribution and upper tail quantile functions. *ISRN Probability and Statistics*, 2013.
32. Robert M Corless, Gaston H Gonnet, David EG Hare, et al. On the Lambert W function. *Advances in Computational mathematics*, 5(1):329–359, 1996.
33. Richard P Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 2013.
34. Bohemia Interactive. ARMA 3. `http://arma3.com/`.
35. Shaoqing Ren, Kaiming He, Ross Girshick, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, pages 91–99, 2015.