Persistent Surveillance of Events with Unknown, Time-varying Statistics

Cenk Baykal, Guy Rosman, Sebastian Claici, and Daniela Rus

***Abstract*— We consider the problem of monitoring stochastic, time-varying events occurring at discrete locations. Our problem formulation extends prior work in persistent surveillance by considering the objective of maximizing event detections in unknown, dynamic environments where the rates of events are time-inhomogeneous and may be subject to abrupt changes. We propose a novel monitoring algorithm that effectively strikes a balance between exploration and exploitation as well as a balance between remembering and discarding information to handle temporal variations in unknown environments. We present an analysis proving the long-run average optimality of the policies generated by our algorithm under the assumption that the total temporal variations are sub-linear. We present simulation results demonstrating the effectiveness of our algorithm in several monitoring scenarios inspired by real-world applications, and its robustness to both continuous-random and abrupt changes in the statistics of the observed processes.**

I. Introduction

Persistent surveillance tasks often require the agent to monitor stochastic, spatially-distributed events of interest in unknown, dynamic environments over long periods of time. Uncertainty over time-varying event statistics necessitates the robot to travel from one landmark to another, identify the regions of importance, and adapt to the temporal variations in the environment over time. The overarching objective is to maximize the number of events observed in order to enable efficient data collection, which may be imperative for a successful surveillance mission. Applications include monitoring of wildlife, natural phenomena (e.g., floods, volcanic eruptions), and friendly and unfriendly activities.

We consider a novel persistent surveillance problem in which a mobile robot is tasked with monitoring transient events that occur in discrete, spatially-distributed landmarks according to station-specific Poisson processes with unknown, time-varying statistics. We assume that the monitoring task is conducted by a single robot equipped with a limited-range sensor that can only record measurements when the robot is stationary at a location, i.e., it cannot make measurements while traveling. Hence, the robot must travel to each location and wait for transient events to occur for an appropriately generated amount of time before traveling to another location. The persistent surveillance problem is to generate an optimal sequence of location-time pairs that maximizes the total sightings of events.

This paper contributes the following:

1) A persistent surveillance problem formulation that bridges the monitoring objective of maximizing event observations with the objective of minimizing regret by introducing a new definition of weak regret for persistent surveillance.
2) A novel monitoring algorithm for generating appropriate policies to monitor transient events in unknown, dynamic environments where the total variation over time is bounded by a variation budget V_T that is known a priori.
3) An analysis proving that under the assumption that the total variation of the event rates is bounded by a variation budget $V_T = o(T)$, our algorithm generates long-run average optimal policies.
4) Simulation results that characterize our algorithm's effectiveness in minimizing regret (i.e., maximizing the number of event observations) in several dynamic and random environments and comparing its performance to adaptive monitoring algorithms.

Cenk Baykal, Guy Rosman, Sebastian Claici, and Daniela Rus are with the Computer Science and Artificial Intelligence Lab, Cambridge, MA 02139, USA {baykal,rosman,sclaici,rus}@csail.mit.edu

II. Related Work

Our work leverages and builds upon prior work in persistent surveillance, mobile sensor scheduling, and stochastic optimization. The problem of persistent surveillance has been studied in a variety of real-world inspired monitoring applications such as underwater marine monitoring and detection of natural phenomena [1]–[11]. These approaches generally assume that the robot can obtain measurements while moving and generate paths that optimize an application-specific monitoring objective, such as mutual information.

The problem of persistent surveillance can be formulated as a mobile sensor scheduling problem and has been studied extensively in this context [12]–[15]. Mobile sensor scheduling in environments with discrete landmarks are of particular relevance to our work. For instance, [3] considers monitoring discrete locations such as buildings, windows, and doors using a team of autonomous micro-aerial vehicles (MAVs). In [16], the authors present an approach to the min-max latency walk problem and [17] extends this work to the multi-objective mobile sensor scheduling problem for surveillance of transient, stochastic events via cyclic policies. Recently, [18] extended the work of [17] by introducing a method with provable guarantees that relaxed the assumption of known event statistics.

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 [19]. Yu et al. present an extension of OP to monitor spatially-correlated locations within a predetermined time [20]. In [21] and [22] the authors consider the OP problem in which the reward accumulated is characterized by a known function

of the time spent at each point of interest. In contrast to our work, approaches in OP predominantly consider known, static environments and budget-constrained policies that visit each location at most once.

As exemplified by the aforementioned works, a variety of monitoring algorithms have been presented and shown to perform well empirically. However, literature on methods with theoretical performance guarantees in unknown, time-varying environments has been sparse and limited. The added complexity stems from the inherent exploration and exploitation trade-off, which has been rigorously addressed and analyzed in canonical Multi-armed Bandit (MAB) literature [23]. Our work differs from the traditional MAB formulation in that we consider optimization in the face of travel costs, non-stationary processes, distributions with infinite support, 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 a monitoring algorithm with provable regret guarantees with respect to the number of events observed.

MAB formulations that relax the assumptions of the traditional MAB problem are of particular pertinence to our work. Besbes et al. present a non-stationary MAB formulation where the variation of the rewards are bounded by a variation budget V_T and present policies that achieve a regret of order $(KV_T)^{1/3}T^{2/3}$, which is long-run average optimal if the variation V_T is sub-linear with respect to the time horizon T [24]. The authors mathematically show the difficulty of this problem by proving the lower bound of $\Omega((KV_T)^{1/3}T^{2/3})$, which implies that long-run average optimality is not achievable whenever V_T is linear in T.

Garivier et al. consider a non-stationary MAB setting where the distributions of the rewards change abruptly at unknown time instants, but the number of changes up to time T, Υ_T, is bounded and known in advance [25]. The authors present discounted and sliding window variants of the Upper Confidence Bound (UCB) algorithm [26], [27] that achieve a regret of $O(\sqrt{T\Upsilon_T \log T})$ and also prove the lower-bound of $\Omega(\sqrt{T\Upsilon_T})$ on the achievable regret in this setting, which is linear if the number of abrupt changes grows linearly with time. Prior work on the MAB formulation with switching costs tells a similar story regarding the difficulty of the aforementioned MAB extensions: Dekel et al. prove the lower-bound of $\tilde{\Omega}(T^{2/3})$ on the achievable regret in the presence of switching costs [28].

Recently, Srivastava et al. presented an approach with a provable upper bound on the number of visits to sub-optimal regions that bridges surveillance and MAB for monitoring phenomena in an unknown, abruptly changing environment [29]. However, their approach considers a discrete state and parameter space (i.e., the generation of observation times is not considered), assumes Gaussian distributed random variables –which may be less suitable for monitoring instantaneous events (such as arrivals), assumes that the number of abrupt changes are bounded and known in advance, and does not explicitly take travel cost into consideration.

We build upon prior work and consider an unknown, dynamic environment where the robot is tasked with visiting each location more than once, observing stochastic, instantaneous events for an appropriately generated time, and adapting to the temporal variations in the environment over an unbounded amount of time. Unlike prior work in persistent surveillance which has focused on environments with a bounded number of abrupt changes, our problem formulation extends to continuously-varying as well as to abruptly-changing environments, as long as the total variation is bounded [24]. We introduce a novel monitoring algorithm with provable guarantees with respect to the number of event sightings and present simulation results of real-world inspired monitoring scenarios that support our theoretical claims.

III. Problem Definition

Let there be $n \in \mathbb{N}_+$ discrete stations in the environment where transient events of interest occur according to inhomogeneous Poisson processes. The temporal variations at each station $i \in [n]$ are governed by an integrable rate function $\lambda_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ that is station-specific and independent of those of other stations. We assume that the rate functions can exhibit an unbounded number of abrupt changes, however, we require that the total variation of each function λ_i within the time horizon $T \in \mathbb{R}_+$ be bounded by a variation budget $V_T \in \mathbb{R}_+$ [24], i.e.,

$$\sup_{P \in \mathcal{P}} \sum_{j=1}^{n_p - 1} \max_{i \in [n]} \left| \lambda_i(p_{j+1}) - \lambda_i(p_j) \right| \leq V_T, \tag{1}$$

where $\mathcal{P} = \left\{ P = \{p_1, \ldots, p_{n_P}\} \mid P \text{ is a partition of } [0, T] \right\}$. We note that since our problem is intimately linked with the MAB problem, we address the case of a known surveillance duration T as is common in MAB literature.

We assume that there exists a travel cost, $c : [n] \times [n] \to \mathbb{R}_{\geq 0}$, associated with going from one station to another. Due to sensor constraints that mandate the robot to be stationary to make accurate measurements, the robot cannot make observations while traveling. Our overarching monitoring objective is to generate an optimal sequence of station-time pairs that dictates the appropriate station visit order and respective observation windows in order to maximize the number of sighted events.

More formally, a policy $\pi = \big((s_1, t_1), \ldots, (s_m, t_m)\big)$ is a sequence of $m \in \mathbb{N}_+$ ordered pairs where each ordered pair, (s, t), denotes an observation window of $t \in \mathbb{R}_{\geq 0}$ time at station $s \in [n]$. For any non-negative reals a, b such that $a \leq b$, let $N_i(a, b]$ denote the random number of events that occur in the time interval $(a, b]$ at station $i \in [n]$. It follows then that $\mathbb{E}\big[N_i(a, b]\big] = \int_a^b \lambda_i(\tau)\, d\tau$ by definition of an inhomogeneous Poisson process at each station i. The expected number of events obtained for any policy $\pi = \big((s_1, t_1), \ldots, (s_m, t_m)\big)$ constrained by the total surveillance time T can then be computed as follows:

$$\mathbb{E}[N(\pi, T)] := \sum_{j=1}^{m} \int_{o_j(\pi)}^{o_j(\pi)+t_j} \lambda_{s_j}(\tau)\, d\tau, \tag{2}$$

where $o_j(\pi)$ denotes the start of the j^{th} observation window, i.e., $o_1(\pi) = 0$ and for any integral value $j > 1$,

$$o_j(\pi) := \sum_{k=1}^{j-1} t_k + c(s_k, s_{k+1}).$$

Our notion of weak regret is defined relative to the maximum number of *expected* events, N_T^* at a single best station after an allotted monitoring time of $T \in \mathbb{R}_+$:

$$N_T^* = \max_{i\in[n]} \mathbb{E}\big[N_i(0,T)\big] = \max_{i\in[n]} \int_0^{\mathrm{T}} \lambda_i(\tau)\ d\tau.$$

We seek to generate policies that minimize the *expected regret* with respect to the quantity N_T^*. We let $R(\pi, T)$ denote the regret accrued by policy $\pi = \big((s_1,t_1),\ldots,(s_m,t_m)\big)$ after time T

$$R(\pi, T) := N_T^* - N(\pi, T) \tag{3}$$

and define our optimization problem with respect to the expectation of $R(\pi, T)$.

Problem 1 (Persistent Surveillance Problem)**.** *Compute the optimal monitoring policy, $\pi^* = \big((s_1^*,t_1^*),\ldots,(s_m^*,t_m^*)\big)$, that minimizes the expected regret with respect to the allotted monitoring time T*

$$\pi^* = \operatorname*{argmin}_{\pi}\ \mathbb{E}[R(\pi, T)]. \tag{4}$$

In what follows, we seek to minimize a long-term variant of Eqn. 4. We define the long-run average optimal policy as

Definition 1 (Long-run Average Optimal Policy)**.** A policy $\pi = \big((s_1,t_1),\ldots,(s_m,t_m)\big)$ is called a long-run average optimal policy if and only if

$$\limsup_{T\to\infty} \frac{\mathbb{E}[R(\pi,T)]}{T} = \frac{\mathbb{E}[N_T^*] - \mathbb{E}[N(\pi,T)]}{T} \le 0. \tag{5}$$

IV. Methods

In this section, we describe the intuition behind our approach, and present a monitoring algorithm (Alg. 1). Our approach trades off exploration and exploitation by leveraging information gained within bounded time steps. Specifically, we partition our alloted time into equal length intervals called epochs. Within each epoch we reason about the currently known best station and attempt to cleverly remove stations that are suboptimal with high probability. By removing suboptimal stations, future passes through the list of remaining stations are expected to yield better long-term rewards and require less time to be spent on traveling relative to observing stations.

The algorithm begins by computing the length of each epoch, τ, as a function of the total time T, variation budget V_T, and maximum travel time between each station $T_{\text{travel}} = \max_{i,j\in[n]:i\neq j} c(i,j)$. The variables N_i and T_i, denoting the total number of observations and the total time spent at station i respectively, are reset at the beginning of each epoch. Discarding out-dated information in this way enables us to balance remembering and forgetting by computing the average rate, $\hat{\lambda}_i = N_i/T_i$, for each station using only information obtained within that epoch. For each epoch, our method employs an algorithm based on the Improved UCB Algorithm [27] and seeks to balance the inherent exploration/exploitation trade-off.

Algorithm 1: Dynamic Upper Confidence Bound Monitoring Algorithm

Input: Time horizon T, variation budget V_T, number of stations n, and travel costs $c : [n] \times [n] \to \mathbb{R}_{\ge 0}$.
Effect: Monitors locations of interest for T time.

$T_{\text{travel}} \leftarrow \max_{i,j\in[n]:i\neq j} c(i,j)$;
// Compute the length of each epoch
$\tau \leftarrow (n\lambda_{\max} T/V_T)^{\frac{2}{3}}$;
while $t_{current} \le T$ **do**
  // Initialize parameters, discarding all previously obtained information from previous epochs
  $T_i \leftarrow 0 \ \ \forall i \in [n]; \quad N_i \leftarrow 0 \ \ \forall i \in [n]$;
  // Initialize the set of all station indices
  $S \leftarrow \{1,\ldots,n\}$;
  $\tilde{\Delta} \leftarrow \lambda_{\max}$;
  // Determine the end point of the current epoch
  $T_{\text{end}} \leftarrow t_{\text{current}} + \tau$;
  while $t_{current} \le T_{end}$ **do**
    // Compute the goal observation time
    $T_{\text{obs}} \leftarrow \frac{8\lambda_{\max}\log\left(\tau\tilde{\Delta}^2\right)}{3\tilde{\Delta}^2}$;
    if $|S| > 1$ **then**
      for $i^* \in S$ *such that* $T_{i^*} < T_{obs}$ **do**
        $t_{i^*} \leftarrow \min\left\{T_{\text{end}} - t_{\text{current}}, T_{\text{obs}} - T_{i^*}\right\}$;
        Observe at station i^* for t_{i^*} time;
        $T_{i^*} \leftarrow T_{i^*} + t_{i^*}$;
    else
      // Only one station remains in S
      $t_{i^*} \leftarrow T_{\text{end}} - t_{\text{current}}$;
      Observe at the sole station $i^* \in S$ until T_{end};
      $T_{i^*} \leftarrow T_{i^*} + t_{i^*}$;
    // Identify and remove suboptimal stations
    $\xi \leftarrow \sqrt{\frac{8\lambda_{\max}\log\left(\tau\tilde{\Delta}^2\right)}{3T_{\text{obs}}}}$;
    $\hat{\lambda}^* \leftarrow \max_{i\in S} \hat{\lambda}_i - \xi$;
    $B \leftarrow \{i \in S \mid \hat{\lambda}_i + \xi < \hat{\lambda}^*\}$;
    $S \leftarrow S \setminus B$;
    $\tilde{\Delta} \leftarrow \frac{\tilde{\Delta}}{2}$;

V. Analysis

In this section, we present a regret bound analysis proving that the policy π generated by Alg. 1 is long-run average optimal with respect to our definition of weak regret. To establish our result, we proceed by bounding the total regret in each epoch of length τ, and then sum the regret over

all $\lceil \frac{T}{\tau} \rceil$ epochs to obtain an upper bound on the entire monitoring horizon of length T.

A. Preliminaries

Assumption 1 (Bounded Rates)**.** *For a given time horizon $T \in \mathbb{R}_+$, the rate parameters $\forall i \in [n] \;\; \lambda_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ are bounded above by a known constant λ_{max}.*

Define a *stage* indexed by $m \in \mathbb{N}$ as the completion of the inner while loop of Alg. 1 (i.e., execution of lines 13-30) and denote the partition of the time horizon T into $k = \lceil \frac{T}{\tau} \rceil$ epochs as $\tau_1, \ldots, \tau_k$ of length τ each (with the possible exception of τ_k). For a given stage m, we define $\tilde{\Delta}(m)$, $T_{\text{obs}}(m)$, $S(m)$, and $\xi(m)$ as the values of each variable at stage m (see Alg. 1). Let $w_{i,0}, \ldots, w_{i,m}$ be the $m+1 \in \mathbb{N}$ observation windows at station $i \in [n]$, where each observation window $w_{i,j}$ is defined by the time interval $(a_{i,j}, b_{i,j})$. Note that $\sum_{j=0}^{m}(b_{i,j} - a_{i,j}) = T_{\text{obs}}(m)$.

For an arbitrary epoch τ_j, we let $\hat{\lambda}_i(m)$ denote the *sample mean* of the rate parameter and let $\bar{\lambda}_i(m)$ denote the ground-truth mean rate after observing at station i for m stages. We define $\bar{\lambda}_i$ to denote the average rate of a station over the specific epoch τ_j (that is clear from the context) and let $\bar{\lambda}^* = \max_{i \in [n]} \bar{\lambda}_i$ denote the epoch-specific optimal rate of the best station $*$. Finally, we let $\Delta_i = \bar{\lambda}^* - \bar{\lambda}_i$ denote the difference in the rates of a station i in comparison to that of the optimal station over epoch τ_j.

Lemma 1 (Concentration Inequalities)**.** For any station $i \in [n]$ and arbitrary sequence of observation windows $w_{i,0}, \ldots, w_{i,m}$ such that $T_{\text{obs}}(m) = \sum_{j=0}^{m}(b_{i,j} - a_{i,j})$ and $\xi \in \mathbb{R}_{\geq 0}$:

$$\mathbb{P}\left(\hat{\lambda}_i(m) > \bar{\lambda}_i(m) + \xi\right) \leq \exp\left(-\frac{3T_{\text{obs}}(m)\xi^2}{8\lambda_{\text{max}}}\right)$$

and for $\xi \in [0, \lambda_{\text{max}}]$

$$\mathbb{P}\left(\hat{\lambda}_i(m) < \bar{\lambda}_i(m) - \xi\right) \leq \exp\left(-\frac{3T_{\text{obs}}(m)\xi^2}{8\lambda_{\text{max}}}\right).$$

B. Regret over an epoch

Our proof employs results established in, and follows a similar structure as the proof given by [27] and [24]. Consider an arbitrary epoch τ_j and let V_j denote the total variation in the rates during this epoch

$$V_j = \sup_{P \in \mathcal{P}_j} \sum_{k=1}^{n_p - 1} \max_{i \in [n]} \left|\lambda_i(p_{k+1}) - \lambda_i(p_k)\right|, \tag{6}$$

where $\mathcal{P}_j$ is a partition of epoch τ_j. Summing over all epochs $j = 1, \ldots, \lceil \frac{T}{\tau} \rceil$, note that $\sum_{j=1}^{\lceil T/\tau \rceil} V_j \leq V_T$.

Let $m_i = \min\{m : \tilde{\Delta}(m) < \Delta_i/8\}$ denote the first stage index in which our guess $\tilde{\Delta}(m)$ is close to the actual difference in the rates for stations $i \in S$. The following inequalities follow by definition (Alg. 1)

$$\tilde{\Delta}(m_i) = \frac{\lambda_{\text{max}}}{2^{m_i}} < \frac{\Delta_i}{8} \leq 2\tilde{\Delta}(m_i) = \frac{2\lambda_{\text{max}}}{2^{m_i}} \tag{7}$$

and

$$\xi(m_i) = \sqrt{\frac{8\lambda_{\text{max}} \log\left(\tau\tilde{\Delta}^2(m_i)\right)}{3T_{\text{obs}}(m_i)}} = \tilde{\Delta}(m_i) < \Delta_i/8.$$

We will consider bounding the regret incurred by monitoring *clearly suboptimal* locations, i.e., $\mathcal{B} = \{i \in S \mid \Delta_i > \Delta\}$, instead of monitoring the optimal station $*$, where $\Delta = \max\{4V_j, \sqrt{\frac{8\exp(1-3/(5\lambda_{\text{max}}))}{\tau}}\}$. Note that suboptimal in this context refers to stations in $\mathcal{B}$, i.e., those stations with significantly high optimality gaps, in contrast to any station i with $\Delta_i > 0$. Let $\mathcal{B}(m) = \mathcal{B} \cap S(m)$ denote the set of suboptimal stations still in the set $S(m)$ after m stages. Let $R^m(i,j)$ denote the event that a station $i \in S(m)$ is removed at the end of stage m (or before) by station j (Alg. 1, Lines 27-29).

We decompose the total expected regret over an arbitrary epoch τ_j of length τ, $\mathbb{E}[R(\pi, \tau)]$, and consider the regret incurred by observing and traveling separately, i.e., $\mathbb{E}[R(\pi,\tau)] = \mathbb{E}[R_{\text{obs}}(\pi,\tau)] + \mathbb{E}[R_{\text{travel}}(\pi,\tau)]$. The following lemmas establish bounds on the probability that an optimal station removes a suboptimal station and vice-versa.

Lemma 2 (Conditional Probability of $R^{m_i}(i,*)$)**.** Given that the optimal station $*$ is in the set $S(m_i)$, the probability of removing a suboptimal station $i \in \mathcal{B}(m_i)$ from $S(m_i)$ at stage m_i (or before) by the optimal station $*$ is given by

$$\mathbb{P}\left(R^{m_i}(i,*) \mid i^* \in S(m_i)\right) \geq 1 - \frac{2}{\tau\tilde{\Delta}^2(m_i)}$$

where $m_i = \min\{m : \tilde{\Delta}(m) < \Delta_i/8\}$.

Proof. We will proceed by finding an upper bound for the probability that an arbitrary suboptimal station is *not* removed during the duration of the epoch τ_j. Consider the following inequalities for some stage m

$$\hat{\lambda}_i(m) \leq \bar{\lambda}_i(m) + \xi(m) \tag{8}$$
$$\hat{\lambda}^*(m) \geq \bar{\lambda}^*(m) - \xi(m). \tag{9}$$

If conditions (8) and (9) hold at stage $m = m_i$ under the assumption that $* \in S(m_i)$, then it follows that i will be removed from $S(m_i)$ at stage m_i

$$\begin{aligned}
\hat{\lambda}_i(m_i) + \xi(m_i) &\leq \bar{\lambda}_i(m_i) + 2\xi(m_i) && \text{by (8)} \\
&\leq \bar{\lambda}_i + V_j + 2\xi(m_i) && \text{by (6)} \\
&< \bar{\lambda}_i + \Delta_i - V_j - 2\xi(m_i) \\
&\leq \bar{\lambda}^* - V_j - 2\xi(m_i) \\
&\leq \bar{\lambda}^*(m_i) - 2\xi(m_i) && \text{by (6)} \\
&\leq \hat{\lambda}^*(m_i) - \xi(m_i) && \text{by (9)}
\end{aligned}$$

where we used the fact that $\Delta_i > 2V_j + 4\xi(m_i)$.

Using Lemma 1, the probability that either (8) or (9) does not hold is as follows

$$\mathbb{P}\left(\hat{\lambda}_i(m_i) > \bar{\lambda}_i(m_i) + \xi(m_i)\right) \leq \frac{1}{\tau\tilde{\Delta}^2(m_i)},$$

and similarly for condition (9)

$$\mathbb{P}\left(\hat{\lambda}^*(m_i) < \bar{\lambda}^*(m_i) - \xi(m_i)\right) \leq \frac{1}{\tau\tilde{\Delta}^2(m_i)}.$$

By the union bound, the probability that the suboptimal station is not eliminated in stage m_i (or before) is bounded above by $\frac{2}{\tau\tilde{\Delta}^2(m_i)} \leq \frac{512}{\tau\Delta_i^2}$. □

Lemma 3 (Probability Bound on $R^{m_i}(*,i)$)**.** The probability that the optimal station $*$ is removed by a suboptimal station $i \in \mathcal{B}$ at stage m_i (or before) is bounded above by

$$\mathbb{P}\left(R^{m_i}(*,i)\right) \leq \frac{1}{\tau\tilde{\Delta}^2(m_i+1)}.$$

Proof. Consider the event that the suboptimal station $i \in \mathcal{B}$ removes the optimal station $*$ at stage m_*. This implies that the following removal condition holds at m_*

$$\hat{\lambda}_i(m_*) - \xi(m_*) > \hat{\lambda}^*(m_*) + \xi(m_*). \tag{10}$$

If we assume that the inequalities (8) and (9) hold at stage m_*, then (10) leads to the contradiction $\bar{\lambda}_i + 2V_j > \bar{\lambda}^*$ since $\Delta_i > 4V_j$:

$$\begin{aligned}
\bar{\lambda}_i + V_j &\geq \bar{\lambda}_i(m_*) \\
&\geq \hat{\lambda}_i(m_*) - \xi(m_*) && \text{by (8)} \\
&> \hat{\lambda}^*(m_*) + \xi(m_*) && \text{by (10)} \\
&\geq \bar{\lambda}^*(m_*) && \text{by (9)} \\
&\geq \bar{\lambda}^* - V_j.
\end{aligned}$$

Thus, it follows that if (8) and (9) hold at stage m_*, the optimal station $*$ will not be removed at this stage. Thus, using previously established results, we have that the probability that an arbitrary suboptimal station $i \in \mathcal{B}$ removes $*$ at stage m_* is at most $\frac{2}{\tau\tilde{\Delta}^2(m_*)}$.

Summing over all stages preceding m_i, we have:

$$\begin{aligned}
\mathbb{P}\left(R^{m_i}(*,i)\right) &\leq \sum_{m_*=0}^{m_i} \frac{2}{\tau\tilde{\Delta}^2(m_*)} = 2\sum_{m_*=0}^{m_i} \frac{2^{2m_*}}{\tau\lambda_{\max}^2} \\
&\leq \frac{2^{2(m_i+1)}}{\tau\lambda_{\max}^2} = \frac{1}{\tau}\left(\frac{2^{m_i+1}}{\lambda_{\max}}\right)^2 \\
&\leq \frac{1}{\tau\tilde{\Delta}^2(m_i+1)} \leq \frac{1024}{\tau\Delta_i}.
\end{aligned}$$

□

Using Lemmas 2 and 3 in conjunction with standard regret bound techniques (such as the one outlined in [27]) yields the following per-epoch regret.

Lemma 4 (Distribution-independent Epoch Regret)**.** The per-epoch expected regret of our algorithm, $\mathbb{E}[R_{\text{obs}}(\pi,\tau)]$, with respect to an arbitrary epoch j of length τ and with total variation budget V_j is at most

$$\mathbb{E}[R_{\text{obs}}(\pi,\tau)] = \mathcal{O}\left(V_j\tau + n\sqrt{\tau}\lambda_{\max}\right).$$

Lemma 5 (Bound on Travel Time Per Epoch)**.** In an epoch length of duration τ, the total regret incurred by traveling from one station to the other is bounded above by

$$\mathbb{E}[R_{\text{travel}}(\pi,\tau)] = \mathcal{O}\left(\log(\tau)n\lambda_{\max}T_{\text{travel}}\right).$$

Proof. By definition of our algorithm, no more than $\mathcal{O}(\log(\tau))$ stages can be executed within an epoch of length τ. Moreover, since the cardinality of S is at most n at each stage, the regret incurred per stage is bounded above by $n\lambda_{\max}T_{\text{travel}}$, which yields the result. □

C. Total Regret

Theorem 1 (Long-run Average Optimality)**.** The total expected regret of Alg. 1, $\mathbb{E}[R(\pi,T)]$, over the entire monitoring duration T and total variation budget V_T is bounded by

$$\mathbb{E}[R(\pi,T)] = \mathcal{O}\left(V_T^{1/3}(Tn\lambda_{\max})^{2/3}\right),$$

for a choice of epoch length $\tau = \mathcal{O}(n\lambda_{\max}T/V_T)^{2/3}$ as long as $n\lambda_{\max}T_{\text{travel}}$ is negligible relative to T, i.e., $n\lambda_{\max}T_{\text{travel}} = \mathcal{O}(1)$.

Proof. Invoking Lemmas 4 and 5 and summing over $\lceil\frac{T}{\tau}\rceil$ epochs yields

$$\begin{aligned}
&\mathbb{E}[R(\pi,T)] \\
&\quad= \sum_{j=1}^{\lceil T/\tau\rceil} \mathcal{O}\left(V_j\tau + n\lambda_{\max}(\sqrt{\tau} + \log(\tau)T_{\text{travel}})\right) \\
&\quad= \mathcal{O}(V_T\tau) + \mathcal{O}\left((\frac{T}{\tau}+1)(n\lambda_{\max}\sqrt{\tau})\right) \\
&\quad= \mathcal{O}\left(V_T\tau + \frac{Tn\lambda_{\max}}{\sqrt{\tau}}\right).
\end{aligned}$$

Setting $\tau = (n\lambda_{\max}T/V_T)^{2/3}$ we have

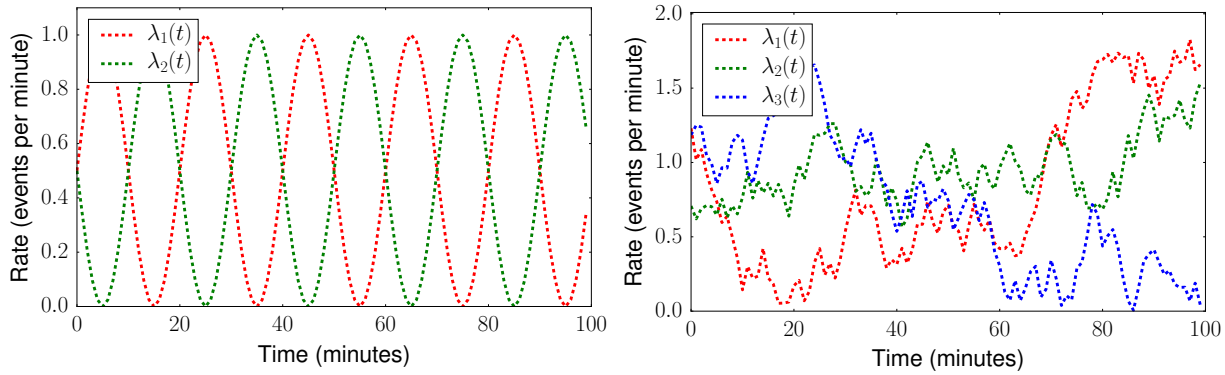
$$\begin{aligned}
\mathbb{E}[R(\pi,T)] &= \mathcal{O}\left(V_T^{1/3}(Tn\lambda_{\max})^{2/3}\right) \\
&= o(T),
\end{aligned}$$

which establishes the long-run average optimality of our algorithm. □

VI. Results

We evaluate the performance of our algorithm in simulated environments subject to temporal variations and compare its effectiveness in maximizing the number of observations within the allotted monitoring time. In particular, we compare our algorithm (Alg. 1) to the following baseline and adaptive procedures that are inspired by state-of-the-art methods:

1) Random Choice & Time: picks a station i^* uniformly at random and observes for a random time $t_i^* \sim \text{Exp}(\lambda_{\text{exp}}(t_{\text{current}}, i^*))$.
2) ϵ-greedy: explores with probability $\epsilon(t_{\text{current}})$ (see Random Choice & Time) and exploits otherwise, i.e. $i^* = \text{argmax}_{i\in[n]}\,\bar{\lambda}_i$ and $t_i^* \sim \text{Exp}(\lambda_{\text{exp}}(t_{\text{current}}, i^*))$.
3) Discounted ϵ-greedy: same procedure as ϵ-greedy except discounted sample means, $\tilde{\lambda}_i$, are used instead.



a) b)

Fig. 1: The two scenarios explored in our experiments. a) The sinusoidal rates of each Poisson process as a function of time with $V_T = \sqrt{T}$ and $T = 100$ minutes. b) The rates of each Poisson process as a function of time generated by a discrete random walk as described in Sec. VI-B. The figure depicts the rates of three stations over a time horizon $T = 100$ minutes and variation budget $V_T = T^{2/3}$.

4) Discounted Cyclic Policy: generates cyclic policies using an extended version of the algorithm introduced by [17] where discounted sample means, $\tilde{\lambda}_i$, are used to ensure adaptiveness.

where $\lambda_{\text{exp}}(t,i) = \bar{\lambda}_i/(n\epsilon(t))$ and $\epsilon : \mathbb{R}_{\geq 0} \to [0,1]$ denotes the exploration function defined as $\epsilon(t) = 1/\log t$. Methods 3 and 4 employ discounted sample means which is computed as follows at time t_{current}

$$\forall i \in [n] \quad \tilde{\lambda}_i = \frac{\sum_{t=0}^{\lceil t_{\text{current}} \rceil} \gamma^{(\lceil t_{\text{current}} \rceil - t)} N_i(t)}{\sum_{t=0}^{\lceil t_{\text{current}} \rceil} \gamma^{(\lceil t_{\text{current}} \rceil - t)} T_i(t)}, \tag{11}$$

where $\gamma = 0.99$ and $N_i(t)$ and $T_i(t)$ denote the sum of events observed and the total observation time spent at station i up to time t respectively.

A. Sinusoidal Variations

We consider the simulated scenario involving the surveillance of two spatially-distributed stations where events occur according to unknown event statistics that are subject sinusoidal temporal variations. The rate functions of the two stations are given as a function of the variation budget V_T where V_T depends sub-linearly on the allotted surveillance time, $V_T = \sqrt{T}$ (see Fig. 1)(a):

$$\lambda_1(t) = \frac{1}{2} + \frac{1}{2}\sin\left(\frac{\pi V_T t}{T}\right)$$
$$\lambda_2(t) = \frac{1}{2} + \frac{1}{2}\sin\left(\frac{\pi V_T t}{T} + \pi\right).$$

The cost of travel from one station to the other is assumed to be 3 minutes of travel during which the robot is unable to record any observations.

Figures 2(a)-(b) show the average performance of each monitoring algorithm given a time horizon $T = 20,000$ minutes, over 100 trials. Our algorithm (shown in cyan) achieves sub-linear regret over time, reaffirming the theoretical property of our algorithm established in Sec. V. In comparison to the other adaptive monitoring algorithms, our algorithm is the only procedure that achieves $\mathbb{E}[R(\pi,T)]/T \approx 0$ (see Def. 1). Furthermore, Fig. 2(c) shows that our algorithm observes the highest percentage of event sighting with respect to the cumulative number of all events that occurred across all stations.

B. Discrete Random Walk

In the previous subsection, we considered environments with temporal variations with a relatively small variation budget $V_T = \sqrt{T}$ where the changes in the environment were continuous and sinusoidal. In this subsection, we consider a significantly more erratic and challenging scenario in which we increase the budget to $V_T = T^{2/3}$ and allow discontinuous, abrupt temporal variations in the event rates. In particular, we consider monitoring 3 stations where the rate functions follow a bounded discrete random walk that is generated as a function of the variation budget $V_T = T^{2/3}$.

Namely, for each station $i \in [n]$, we construct a station-specific random sequence defined by $X_0 \sim \text{Uniform}(0,1)$ and X_t for $t \in \mathbb{N}_+$, $t \leq T$:

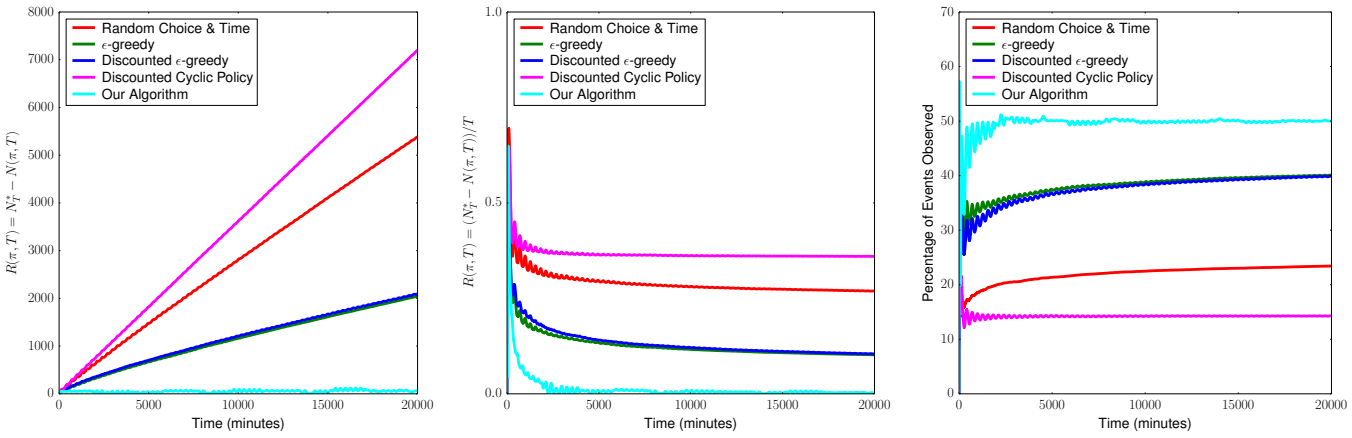
$$X_t = \begin{cases} X_{t-1} + U_t & \text{if } X_{t-1} + U_t > 0 \\ X_{t-1} + |U_t| & \text{otherwise} \end{cases} \tag{12}$$

where $U_t \sim \text{Uniform}(-V_T/T, V_T/T)$. Then, the rate function associated with station $i \in [n]$ is defined as

$$\lambda_i(t) = X_{\lceil t \rceil}. \tag{13}$$

Figure 1(b) depicts an example generated by this construction performed with a curtailed time horizon of $T = 100$ minutes. The travel time between one station i to the other j, $i \neq j$, is uniformly drawn, i.e., $c(i,j) \sim \text{Uniform}(1,5)$ minutes.

Figures 3(a)-(c) show the performance of each monitoring algorithm for a time horizon $T = 20,000$ minutes, averaged over 100 trials. The figures tell the same story as did those from the previous subsection: our algorithm (cyan) is the only method to achieve sub-linear regret over time (Figs. 3)(a)-(b), which reaffirms the long-run average opti-



a) b) c)

Fig. 2: a) Plot of total regret $R(\pi, T) = N_T^* - N(\pi, T)$ over time. The figure depicts sub-linear growth of regret over time for our algorithm (cyan), as expected from our theoretical results (Sec. V). b) Growth of total regret over time expressed as the quotient $R(\pi, T)/T$. Our algorithm achieves sub-linear regret over time and that $R(\pi, T)/T \to 0$. c) Percentage of events observed with respect to the sum of events that occurred across all stations in the environment subject to sinusoidal variation over time. Our algorithm approximately attains optimal number of expected events in this setting consisting of 2 stations.

mality of the policies generated by our algorithm, depicted in Fig. 3(b). In addition to minimizing the regret metric formalized in Sec. III, the policies generated by our method achieve the highest percentage of observed events taken with respect to all of the transpired events at 3 stations, as shown in Fig. 3(c).

VII. Conclusion

In this paper, we presented a novel algorithm for monitoring transient events in unknown, dynamic environments over a long period of time. The algorithm proposed in this paper builds upon and extends the state-of-the-art in persistent surveillance by introducing a method of constructing policies that are provably long-run average optimal even in scenarios subject to discontinuous, abrupt temporal variations. Our method hinges on novel connections between persistent surveillance and the Multi-armed Bandit (MAB) problem variants, which may be of independent theoretical interest.

Our favorable simulation results in both continuously and abruptly changing environments reaffirm our theoretical results and show the potential applications of our algorithm to a wide range of monitoring applications. We envision that our algorithm may be employed to facilitate persistent surveillance missions, such as detection and tracking efforts at a large scale.

ACKNOWLEDGMENT

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. We are also grateful for the support provided by ONR grant N00014-12-1-1000.

References

[1] J. Binney, A. Krause, and G. S. Sukhatme, "Informative path planning for an autonomous underwater vehicle," in *Robotics and automation (icra), 2010 IEEE international conference on*. IEEE, 2010, pp. 4791–4796.

[2] R. N. Smith, M. Schwager, S. L. Smith, B. H. Jones, D. Rus, and G. S. Sukhatme, "Persistent ocean monitoring with underwater gliders: Adapting sampling resolution," *Journal of Field Robotics*, vol. 28, no. 5, pp. 714–741, 2011.

[3] N. Michael, E. Stump, and K. Mohta, "Persistent surveillance with a team of mavs," in *IROS*, 2011.

[4] S. L. Smith, M. Schwager, and D. Rus, "Persistent robotic tasks: Monitoring and sweeping in changing environments," *Robotics, IEEE Transactions on*, vol. 28, no. 2, pp. 410–426, 2012.

[5] D. E. Soltero, M. Schwager, and D. Rus, "Generating informative paths for persistent sensing in unknown environments," in *IROS*, 2012, pp. 2172–2179.

[6] G. A. Hollinger, S. Choudhary, P. Qarabaqi, C. Murphy, U. Mitra, G. S. Sukhatme, M. Stojanovic, H. Singh, and F. Hover, "Underwater data collection using robotic sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 899–911, 2012.

[7] X. Lan and M. Schwager, "Planning periodic persistent monitoring trajectories for sensing robots in gaussian random fields," in *ICRA*. IEEE, 2013, pp. 2415–2420.

[8] G. A. Hollinger and G. S. Sukhatme, "Sampling-based motion planning for robotic information gathering." in *Robotics: Science and Systems*. Citeseer, 2013, pp. 72–983.

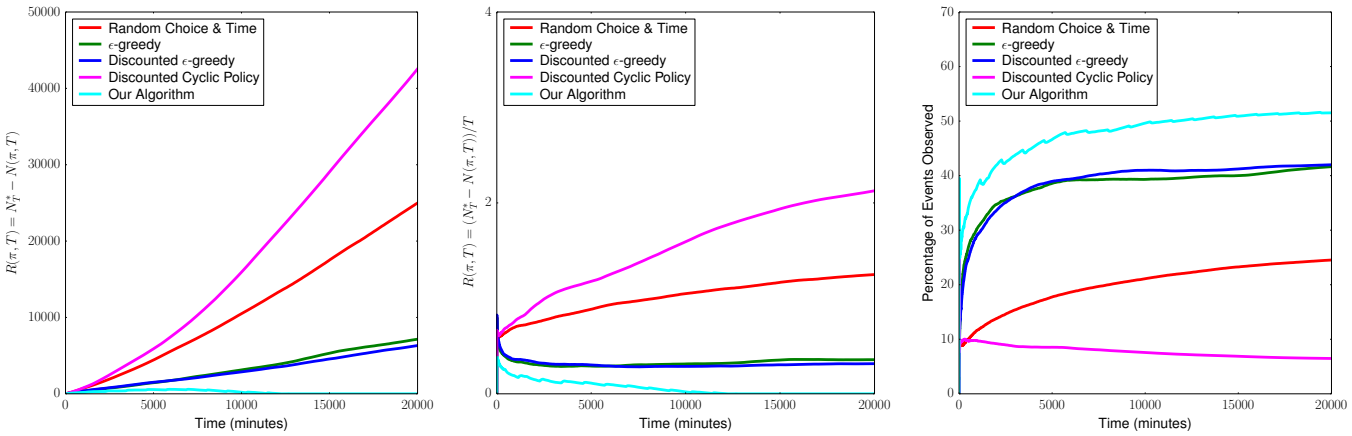
[9] M. Schwager, D. Rus, and J. E. Slotine, "Decentralized, adaptive coverage control for networked robots," *IJRR*, vol. 28, no. 3, pp. 357–375, 2009.

[10] C. Cassandras, X. Lin, and X. Ding, "An optimal control approach to the multi-agent persistent monitoring problem," *Automatic Control, IEEE Transactions on*, vol. 58, no. 4, pp. 947–961, 2013.

[11] V. Srivastava, F. Pasqualetti, and F. Bullo, "Stochastic surveillance strategies for spatial quickest detection," *The International Journal of Robotics Research*, vol. 32, no. 12, pp. 1438–1458, 2013.

[12] Y. He and E. K. Chong, "Sensor scheduling for target tracking in sensor networks," in *ICDC*, vol. 1. IEEE, 2004, pp. 743–748.

[13] A. O. Hero III, C. M. Kreucher, and D. Blatt, "Information theoretic approaches to sensor management," in *Foundations and applications of sensor management*. Springer, 2008, pp. 33–57.



(a) (b) (c)

Fig. 3: Left-to-right a) Total regret as a function of time, i.e. $R(\pi, T) = N_T^* - N(\pi, T)$, in the simulated scenario involving discontinuous, abrupt changes. Our algorithm (shown in cyan) achieves the lowest regret at all times of the allotted monitoring time $T = 20,000$ minutes. b) Growth of the total regret over time, $R(\pi, T)/T$, in an abruptly changing environment. c) Percentage of events observed with respect to all of the events that transpired in the abruptly changing environment (Sec. VI-B) at all stations during the time horizon T.

[14] Y. Gabriely and E. Rimon, "Competitive on-line coverage of grid environments by a mobile robot," *Computational Geometry*, vol. 24, no. 3, pp. 197–224, 2003.

[15] J. L. Ny, M. A. Dahleh, E. Feron, and E. Frazzoli, "Continuous path planning for a data harvesting mobile server," in *ICDC*. IEEE, 2008, pp. 1489–1494.

[16] S. Alamdari, E. Fata, and S. L. Smith, "Persistent monitoring in discrete environments: Minimizing the maximum weighted latency between observations," *IJRR*, vol. 33, no. 1, pp. 138–154, 2014.

[17] J. Yu, S. Karaman, and D. Rus, "Persistent monitoring of events with stochastic arrivals at multiple stations," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 521–535, 2015.

[18] C. Baykal, G. Rosman, K. Kotowick, M. Donahue, and D. Rus, "Persistent surveillance of events with unknown rate statistics," in *Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2016.

[19] A. Gunawan, H. C. Lau, and P. Vansteenwegen, "Orienteering problem: A survey of recent variants, solution approaches and applications," *European Journal of Operational Research*, 2016.

[20] J. Yu, M. Schwager, and D. Rus, "Correlated orienteering problem and its application to informative path planning for persistent monitoring tasks," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 342–349.

[21] G. Erdoan and G. Laporte, "The orienteering problem with variable profits," *Networks*, vol. 61, no. 2, pp. 104–116, 2013.

[22] J. Yu, J. Aslam, S. Karaman, and D. Rus, "Anytime planning of optimal schedules for a mobile sensing robot," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 5279–5286.

[23] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *arXiv preprint arXiv:1204.5721*, 2012.

[24] O. Besbes, Y. Gur, and A. Zeevi, "Non-stationary stochastic optimization," *Operations Research*, vol. 63, no. 5, pp. 1227–1244, 2015.

[25] A. Garivier and E. Moulines, "On upper-confidence bound policies for switching bandit problems," in *International Conference on Algorithmic Learning Theory*. Springer, 2011, pp. 174–188.

[26] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[27] P. Auer and R. Ortner, "Ucb revisited: Improved regret bounds for the stochastic multi-armed bandit problem," *Periodica Mathematica Hungarica*, vol. 61, no. 1-2, pp. 55–65, 2010.

[28] O. Dekel, J. Ding, T. Koren, and Y. Peres, "Bandits with switching costs: T 2/3 regret," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, 2014, pp. 459–467.

[29] V. Srivastava, P. Reverdy, and N. E. Leonard, "Surveillance in an abruptly changing world via multiarmed bandits," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 692–697.