2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Privacy-Preserving Map-Free Exploration for Confirming the Absence of a Radioactive Source

Eric Lepowsky[1,2,†] David Snyder[1,3,†] Alexander Glaser[1,2] Anirudha Majumdar[1,3]

***Abstract*— Performing an inspection task while maintaining the privacy of the inspected site is a challenging balancing act. In this work, we are motivated by the future of nuclear arms control verification, which requires both a high level of privacy and guaranteed correctness. For scenarios with limitations on sensors and stored information due to the potentially secret nature of observable features, we propose a robotic verification procedure that provides map-free exploration to perform a source verification task without requiring, nor revealing, any task-irrelevant, site-specific information. We provide theoretical guarantees on the privacy and correctness of our approach, validated by extensive simulated and hardware experiments.**

I. INTRODUCTION

Autonomous robots observe and process information from their operating environments, giving rise to privacy concerns. Privacy is commonly discussed in the context of protecting personal information – relating to a person's identity, behaviors, or health – for instance, in assistive, social, and home robotics. Similar concerns also may arise when considering remote monitoring of mutually distrustful parties, where sensitive or compromising information must be protected [1]. In this work, we consider the future of nuclear arms control, which is predicated on the collection of information to verify compliance with agreed-upon limits while contending with the secrecy of the nuclear enterprise, as one such sensitive and high-stakes setting which demands privacy [2, 3].

One method for achieving privacy is "forgetting," either by deleting previously acquired information or by compression and abstraction of the information [4, 5]. For settings where the information at stake requires even more protection, such as in arms control, and inspired by the concept of "forgetting," we invoke a more extreme alternative for achieving privacy: never learning or remembering in the first place. In this regard, techniques from reduced- and minimal-information decision making can help to achieve a higher level of privacy by not requiring, nor revealing, any sensitive information or any information which is not strictly task-relevant.

While it is difficult to anticipate the objectives of future international agreements, they are likely to require new verification approaches which preserve aspects of onsite inspections – traditionally essential in monitoring and verification – while resolving concerns about intrusiveness [6]. Onsite inspections commonly include radiation detection as a verification tool. In scenarios where no sources (e.g., nuclear warheads or fissile material) are declared, the inspector must confirm the declared absence of sources or identify an anomalous source, if present [7, 8]. Completing this "absence confirmation" task with high confidence remains challenging, particularly within this high-privacy paradigm.

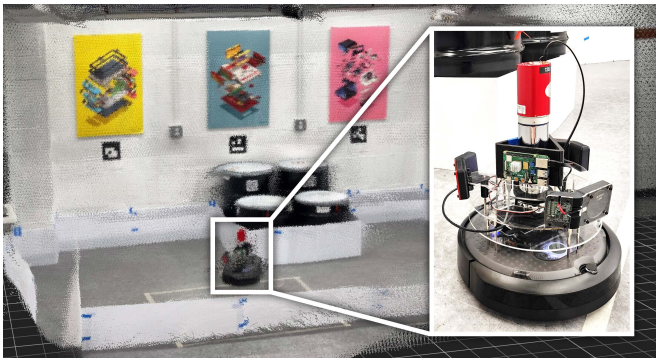

Fig. 1: Robotic inspector in a representative laboratory search environment. The environment is approximately 15 m^2 with steel drum obstacles and dividers to reconfigure the space. The robotic inspector, a Create 3 platform fitted with gamma-ray detectors, explores the unknown environment and confirms, with high probability, the absence or presence of a radioactive source using only non-sensitive information.

The introduction of robotics to nuclear verification has the potential to fundamentally transform relevant inspection approaches [9–12], including by affording a higher level of privacy for the inspected party. Consider what a human inspector may observe during an onsite inspection: as they survey a site, they can mentally catalogue its contents, acquire measurements, or more, inevitably seeing and learning things that are sensitive yet irrelevant to the inspection task at hand. Limiting access to these observations is impossible with human-based inspections, but it can be accomplished with a carefully designed robotic inspector (see Figure 1).

Accordingly, we address the absence confirmation task while considering any observable, site-specific, task-irrelevant information from the search environment to be sensitive (and hence limited), including imagery, dimensions, and even the site layout since mapping would reveal design characteristics of the site's contents. Many robot-compatible source detection methods, such as [13–16], are in contention with such a minimal-information constraint by either using *a priori* knowledge of the environment or, by consequence of the algorithmic design, revealing the site's configuration or radiation field. By considering this stringent information constraint, and demonstrating the plausibility of the inspection task in a provably-private and correct manner, we contribute to a dialogue on what may be possible in the future.

Corresponding Author: dasnyder@princeton.edu. [1]Mechanical and Aerospace Engineering, Princeton University, NJ, USA. [2]Program on Science and Global Security (SGS), Princeton, NJ, USA. [3]Intelligent Robot Motion (IRoM) Lab, Princeton, NJ, USA. [†]E. L. and D. S. contributed equally; other authors listed alphabetically. Project GitHub: github.com/elepowsky/verification. Mathematical proofs for all theoretical claims are presented in the extended version: arXiv:2402.17130.

Statement of contributions. The primary contribution of this work is the development of a verification algorithm, which guarantees exploration while encoding only non-sensitive information, to confirm the absence of sources. We provide theoretical guarantees on privacy preservation and bounds on the false positive rate, and characterize the false negative rate in terms of parameters fundamental to the verification task. The proposed algorithm is validated in simulation and through extensive hardware experiments.

II. RELATED WORK

In developing a framework for accomplishing privacy-preserving absence confirmation, we take inspiration from work on reduced- or minimal-information decision making. Techniques like dimensionality reduction [17–20] and control-theoretic methods [21, 22] promote robustness via lossy, compressed representations of the sensory feedback and regularization of information usage. Analytical frameworks capturing the notion of "available information" quantify and formalize a relation between available and utilized information and performance [23, 24]. While these methods are generally concerned with improving robustness and preventing overfitting, our task here is to set information usage to be absolutely minimal, then characterize the feasible system performance. From this vantage point, the literature on differential privacy is closely related in providing sufficient notions of information security [5, 25, 26]; however, this does not provide guarantees on minimal algorithm performance.

To provide map-free exploration, the algorithm we propose leverages random walk processes. Of particular interest for guaranteed exploration is the cover time of random walks, or essentially how long it takes for a random walker to visit all regions of a given domain. Worst-case results for the expected cover time on undirected graphs have been given in [27, 28]. General properties for finite graphs of various structures have been shown in [29–32], with classical Perron-Frobenius theory summarized in [33, 34]. There has also been interest in developing approximations for expected cover times, as in [35–37]. For sampling from sparse fields, Lévy flights, which exhibit heavy-tailed distributions over step size, are one means of efficient, multi-scale, bio-inspired exploration [38–40]. While this particular type of random walk could expedite exploration, high-probability coverage properties would require additional analytical scaffolding, complicating the use of such multi-scale strategies.

III. PROBLEM FORMULATION

The problem is twofold, requiring definitions for both the search environment and the source verification task. Although the goal is to confirm the absence of sources, this framework requires that a source must be detected, if present.

Specifically, assume that a site is declared to contain no sources. The inspection task is to verify the declaration by confirming the absence of sources (or their presence in a non-compliance situation) by traversing the free space and measuring the observed scalar field. Complicating this is our information constraint: the robotic inspector's retained information must be kept to a minimum. Ideal verification methods must provide both *calibrated correctness* (the ability to choose the probability of returning the correct inspection result) and *provable privacy* (minimizing the robot's capacity to "leak" information). Notably, these goals are generally in opposition; correctness would benefit from more information to better characterize the space, while privacy would require less information be available to the robot.

A. Defining the Environment

For the search environment $E(\mathcal{I}, s, M)$, the robotic inspector $\mathcal{I}$ is tasked with determining the absence or presence of a source of emission strength $s \geq 0$ in the map M, where $s = 0$ corresponds to "no source". The robot $\mathcal{I}$ has a fundamental length $r_I > 0$ (e.g., its diameter). The map $M(l_x, l_y, B)$ is physically bounded by positive length constants l_x, l_y with an unknown occupancy function defining the free space. l_x, l_y are assumed to be non-sensitive if they may be determined without access to the search environment, for instance through open-source information or outside observation. For radiation detection, the environment also has a Poisson-distributed background of mean $B \geq 0$.

B. Map Compression

For this work, we must restrict ourselves to the class of maps with a single, traversable (i.e., contiguous) region of free space. We regularize the set of valid maps by discretizing the problem into a directed graph representation, where each node is a region of space. Specifically, the inequalities of Eq. 1 must hold for the discretization length ϵ_M of map M, where r_D is the detector range, or the distance from which a source is readily detectable above background:

$$r_I \leq \epsilon_M \leq \frac{r_D}{\sqrt{2}}. \tag{1}$$

The left-hand inequality ensures traversability after discretization, while the right-hand inequality ensures that if the robot enters a particular bin, it can detect a source from anywhere in that bin. This latter statement implicitly assumes that sources are detectable, which is dependent on various factors, including any intervening material between the source and the detector; this problem has been explored in the context of radiation detection for treaty verification, for example in [7], and will be momentarily neglected, such that excessively shielded sources are considered to be incompatible with the present formulation. We conservatively take r_D to be the distance at which the signal-to-background ratio reaches unity. Therefore, each time the discretized space is covered, there must be at least one potential anomalous measurement if a source is present.

Henceforth, we will refer to a *compressed map* as $M(l_x, l_y, B, \epsilon_M)$ and define a class of compressed maps as $\mathbb{M}(l_x, l_y, B, \epsilon_{\underline{M}}) = \{M(l_x, l_y, B, \epsilon_M) : \epsilon_M \geq \epsilon_{\underline{M}}\}$. The property $\epsilon' \geq \epsilon \implies \mathbb{M}(l_x, l_y, B, \epsilon') \subseteq \mathbb{M}(l_x, l_y, B, \epsilon)$ follows directly. Incorporating Eq. 1 with this definition, the valid set of maps for a given inspector is the set difference given by $\mathbb{M}_{\mathcal{I}} = \{\mathbb{M}(l_x, l_y, B, r_I) \setminus \mathbb{M}(l_x, l_y, B, \frac{r_D}{\sqrt{2}})\}$.

Critically, although the inspector acts in $M \in \mathbb{M}_{\mathcal{I}}$, it does not see nor construct a representation of the underlying map. Furthermore, to maintain the privacy of the site, the inspector does not collect or store information (e.g., a state history) which would be sufficient to deduce the map, nor does it store information (e.g., scalar measurements) which would be sufficient to characterize the radiation field. We emphasize that all map-dependent results (coverage times and fractions) are from an omniscient view unavailable to the inspector.

C. Source Detection with Limited Information

The verification task has two distinct failure modes: a false negative occurs if the robotic inspector incorrectly returns "absence confirmed," and a false positive occurs if the robotic inspector incorrectly returns "anomaly detected." To minimize the false negative rate (FNR), the robot needs to guarantee exploration of the space, such that a source would be detected, if present. Additionally, each indicator of source-presence must individually have a guaranteed false positive rate (FPR). This is similar to a standard suite of problems in robotics, including out-of-distribution detection, anomaly detection, and failure prediction [41–45]. What distinguishes our setting is the fundamental constraint that the stored information $\mathcal{G}_t$ be exclusively non-sensitive, that it not allow for reconstruction of the underlying map, and that this property holds uniformly across all $t \in \mathbb{N}$.

Given r_I and r_D, it is assumed that the map M is drawn from the class of valid maps $\mathbb{M}_{\mathcal{I}}$. This assumption is not too onerous given that human inspectors also have non-zero extent, and therefore would struggle in an overly obstacle-dense map. Granted, we cannot rival humans in being able to flag certain "adversarial" maps—though they would use (sensitive) sensory information to do so. Second, the measurement model $h(x_t, y_t; E)$ must be predictable, such that anomalous detections may be differentiated from the background. Critically, the actual measurements h_t are sensitive and cannot be stored directly. The robot's position also cannot be known during operation, as this would reveal information about the map. The verification algorithm $\mathcal{A}$ must take physical actions and make decisions d_t ("absence confirmed," "anomaly detected," or "continue") that rely only on non-sensitive accumulated information $\mathcal{G}_t$.

For radiation detection, $h \sim \mathcal{P}(B + g(s, x, y))$ is Poisson-distributed. The non-negative function g is 0 if $s = 0$ or the robot position (x, y) does not have line-of-sight to the source; this assumes that obstacles are completely attenuating, which is a simplification of the absorption and scattering which occurs in real-life. Otherwise, $g \propto \frac{s}{r^2}$, for r as the Euclidean distance from the inspector to the source; the inverse-square law is merely an approximation, which we replace with an experiment-based model in our presented demonstration.

IV. METHODOLOGY

The algorithm we propose takes inspiration from randomized, sampling-based motion planners [46, 47] and out-of-distribution detection [48]. Our random walk policy encodes the scalar measurements as physical actions, guaranteeing exploration of the search environment while simultaneously accumulating $\mathcal{G}_t$, which is a non-sensitive proxy of the measurement history; the history itself is *never* stored.

Consider a robot that can translate forward and rotate in place, detect imminent collisions, and accurately acquire scalar (radiation) measurements; such a system can run Alg. 1. When a measurement is consistent with source-absence, the robot moves according to a "reference" random walk (maximum step size c_U) that explores the space; otherwise, if consistent with source-presence, it moves according to an "out-of-distribution" random walk (maximum step size c_L). Since the actions depend only on the measured scalar, the resulting distribution over actions (step sizes) for any source-free map is theoretically identical; we refer to this source-free action distribution as the reference, V_r. Detection of a shift in the realized action distribution, denoted V_e, is accomplished by Kolmogorov-Smirnov (KS) testing [49]. We set the confidence parameter p^* based on the KS test P-value to determine if the distributions are more likely distinct.

Algorithm 1 Random walk absence confirmation.

Input: Estimated background B, Outer dimensions l_x, l_y, Confidence parameter p^*, Run time T, Test count n, Threshold level z, Step size constants $0 \leq c_L < c_U$, Reference distribution V_r
Output: Inspection result
Initialize P-value $\underline{p} = 1.0$, Time step $t = 1$, Realized action distribution $V_e = \{\emptyset\}$, Starting pose x_0, y_0, θ_0
while $t \leq T$ **do**
  $N_t \sim h(x_t, y_t; E)$ {Field measurement}
  $c \leftarrow c_L + (c_U - c_L)\mathbb{1}[N_t \leq B + z\sqrt{B}]$ {Set max step}
  $ds,\ d\theta \sim \mathcal{U}[0, c],\ \mathcal{U}[0, 2\pi]$ {Step length, rotation}
  Rotate by $d\theta$ rad. and move forward ds distance
  Append ds to memory V_e
  if $t \equiv 0 \pmod{T/n}$ **then**
    $\underline{p} = \min\{\underline{p}, \mathbf{KS}(V_e, V_r)\}$ {Perform KS test}
  end if
  if $\underline{p} \leq p^*/n$ **then**
    return 1 {Result: Anomaly detected}
  end if
end while
return 0 {Result: Absence confirmed}

The robot takes smaller steps when anomalously high counts are detected (if the observed counts exceed z standard deviations above the expected background, as is common practice in radiation detection). We set the reduced step size to $c_L = \frac{c_U}{10}$; for $c_L \lesssim \epsilon_{\underline{M}}$, the inspector is likely to stay in the vicinity of an anomalous source so that evidence of source-presence is self-reinforcing. The step size (ds) recorded is the randomly-selected distance between measurement points; when the robot encounters an obstacle, it randomly redirects then travels the remaining distance. Alg. 1 terminates once the maximum number of steps is reached ($d_t = 0$, "Absence confirmed") or if the KS test P-value reaches the threshold of p^*/n ($d_t = 1$, "Anomaly detected").

Note that Alg. 1 requires an estimate of the background. For this proof-of-concept, we assume that the background has been previously characterized (for example, when the environment was first initialized). A more robust approach could consider pseudo-online learning of the background, which we defer to future work. A possible complication would be a non-uniform background field, for which more thorough characterization would violate the information constraint. However, as a workaround, the estimated background, B, can be re-interpreted as B_{max} which captures any fluctuations, such that "absence confirmed" means that nothing over the maximally acceptable background is present.

V. PRIVACY AND CORRECTNESS

The proposed methodology is now characterized theoretically to confirm that it satisfies the two required properties: provable privacy and calibrated correctness.

A. Information Privacy

A satisfactory verification approach must be private, meaning that it does not "leak" any sensitive information, as formalized by Theorem V-A.

Theorem V-A *(Information Privacy of Compliant Hosts):* Consider the class of compliant (source-free) maps, denoted by $\mathbb{M}^-(l_x, l_y, B, \epsilon_{\underline{M}})$. Alg. 1 is considered to be private, for all time, t, with respect to any map, $M \in \mathbb{M}^-$, in that the mutual information ($\mathcal{MI}$) between any stored data point (namely, the step size between measurements) and the particular compliant (source-free) map is zero [50]. This mutual information is mathematically expressed by Eq. 2, for each data point $\{\mathcal{G}_t \setminus \mathcal{G}_{t-1}\}$ (equivalent to ds_t) and class of maps $\mathbb{M}^-$. In other words, at no time in its operation can Alg. 1 distinguish between any pair of compliant maps:

$$\mathcal{MI}(\{\mathcal{G}_t \setminus \mathcal{G}_{t-1}\}, \mathbb{M}^-) = \mathcal{MI}(ds_t, \mathbb{M}^-) = 0 \ \forall t \geq 1. \quad (2)$$

To further demonstrate this result, Figure 2 shows (in simulation) that Alg. 1, which stores only the step size between measurements, yields a result that depends *only* on the absence or presence of a source. The equivalency between the realized action distribution for empty and obstacle-filled environments concretely demonstrates that there is zero mutual information, per Theorem V-A. Conversely, as a counter-example, a seemingly similar information storage scheme which stores the step size between turns "leaks" enough information to differentiate between environments.

B. Inspection Correctness

Correctness requires guaranteeing a low false negative rate (FNR) without compromising the false positive rate (FPR), and vice versa. We begin with an immediate characterization of the false-positive calibration by the setting of p^* and n.

Remark V-B *(Calibrated False Positive Rate):* The FPR of Alg. 1, equivalently the probability of incorrectly detecting an out-of-distribution anomaly, is less than or equal to p^*. This follows from a union bound applied to the outcomes of n pre-specified Kolmogorov-Smirnov (KS) tests, each individually performed at a significance of p^*/n.

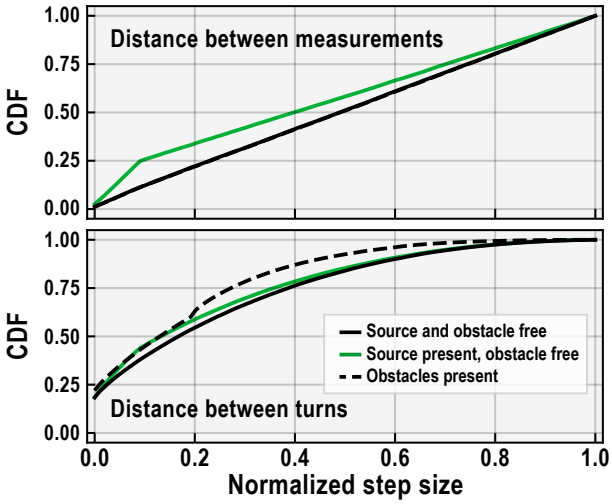

Fig. 2: Step size distributions for information storage scheme privacy and counterexample. Cumulative density functions over step size for our algorithm (distance between measurements) and a "leaky" alternative (distance between turns). Our algorithm (above) is only dependent on the presence/absence of a source, whereas the seemingly similar information storage scheme (below) leaks information which can differentiate between environments of differing occupancy. Note that the solid and dashed black lines in the upper plot are overlapped; this particular curve is equivalent to the reference distribution, V_r, which is independent of the environment.

To address the FNR, full coverage of the environment with sufficiently high resolution is necessary to eliminate the possibility of source-presence. Unfortunately, standard coverage algorithms typically rely on detailed knowledge of the environment [51–53]. Even planners which don't require a map *a priori* typically maintain a state history, forming a representation of the space that is incompatible with the minimal-information constraint. Although less time-efficient, we use random walk processes to provide the necessary exploration without requiring any environmental information.

In essence, our absence confirmation algorithm is a random walk in continuous space, reduced to a discrete graph for tractability. To calibrate the FNR, we require a bound on coverage time (T) that guarantees full coverage of the environment (i.e., every discretized bin has been visited). Formally, we desire a function $\mathcal{T}(N) = \max_M T_M \ \forall M \in \mathbb{M}(l_x, l_y, B, \epsilon_{\underline{M}})$ which provides an upper bound on coverage time for a given class of maps. The following lemma characterizes the tail behavior of the coverage distribution.

Lemma V-B *(Passage Times in Exponential Family [30]):* Consider any compressed map with N nodes and graph diameter D_G, radius r_G. The distribution of first passage times to node i from any other node $j \neq i$ is a member of the exponential family; the first passage time from node $j^* \neq i$ to node i is distributed geometrically in non-dimensionalized time $\tau = \frac{t}{r_G(i)}$, where $r_G(i) \leq D_G \leq N$. This ensures that relatively tight high-probability bounds on coverage can be obtained; it also reflects how the particular structure overcomes several worst-case coverage time results.

Achieving a calibrated FNR also requires quantifying the conditional probability of source detection *given* full graph coverage. We assume that potential sources are sufficiently strong (or the detector is sufficiently sensitive within r_D)

such that the conditional probability is essentially equal to one, which is consistent with the simulations and hardware experiments. Although simplified here, the notion of detectability is a multi-faceted yet analytical problem [54]; generally, notwithstanding excessive shielding, higher efficiency detectors and longer sampling periods improve detectability.

Backed by these theoretical underpinnings, we now assess the coverage time empirically. A diverse set of 10 simulated environments were utilized (see Section VI-A), several of which reflect known worst-case configurations for undirected graphs, such as a barbell graph. We simulate 50 independent trials for each 10×10 m environment and each of 5 different maximum step sizes, c_U = (2, 4, 6, 8, 10 m). The coverage versus time for a range of discretization sizes (25, 100, or 400 bins of corresponding side length 2, 1, or 0.5 m) is summarized in Table I. We can use these results to approximate the upper bound on coverage time, $\mathcal{T}(N)$.

	2 m	4 m	6 m	8 m	10 m
5×5 bins	810 (3529)	305 (1861)	200 (1721)	159 (818)	145 (699)
10×10 bins	1481 (5285)	741 (2614)	611 (1932)	547 (1444)	547 (1369)
20×20 bins	3422 (12161)	2502 (6710)	2300 (6369)	2256 (6295)	2232 (5022)

TABLE I Empirical coverage time (step number) for varied maximum step size and discretization. The mean number of steps, averaged over all 10 environments and all 50 trials, is reported; the maximum over all rooms and trials is reported in parentheses to demonstrate the worst case observed.

The tabulated results are intuitive: neglecting travel time, larger maximum step sizes and larger bins (equivalently, fewer bins) result in lower coverage time. To further elucidate the results from Table I, Figure 3 visualizes the empirical coverage over time (converting from number of steps to real-world time) for the 5×5 binning; this particular size was chosen since the 25-bin compressed maps are most closely aligned with the bin number of the hardware experiments.

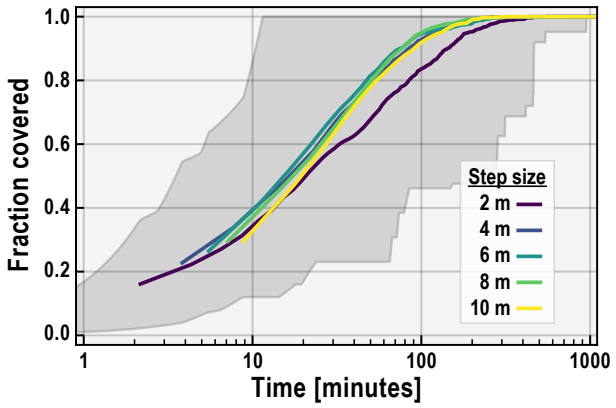

Fig. 3: Empirical coverage versus time. Evaluated for a range of maximum step sizes for the 5×5 binning. For each step size, the average over all 10 environments and all 50 trials is shown; the curves start after 10 initial time steps. The shaded region represents the full range of possible values, evaluated over all step sizes. To convert from step number to real-world time, we assumed 3-second measurements, travel speed of 10 cm/s, and neglect the time spent avoiding obstacles.

VI. EXPERIMENTS

We experimentally demonstrate the correctness of our algorithm, i.e., the ability to correctly identify the absence or presence of a source, both in diverse simulated environments and on hardware in various laboratory settings, collectively spanning a wide range of scales and configurations.

A. Simulation in PyBullet

Our simulation environment uses PyBullet [55, 56], based on the environment setup from [57]. A variety of environments were constructed (30 in total, in addition to the 10 unique environments used for Table I), each with different occupancy functions, including an assortment of maps with open space ranging from 100 m^2 (an entirely empty room) down to 20 m^2 (obstacle-dense maps). For each map, 100 independent trials were conducted: 10 with and 10 without a source present, each for 5 different maximum step sizes. For each trial, the robot and source (if present) were initialized in random positions. Ray-tracing provided a realistic, albeit simplified, measurement model, where obstacles were assumed to be fully attenuating and the spatial dependence for non-attenuated counts was experimentally-based.

The evolution of the KS test P-value for Alg. 1, averaged over all similar trials across all environments (i.e., 1,500 trials each for source absence and presence), is shown in Figure 4. We apply the KS test after every 100 measurements, using $p^* = 0.005$ and $n = 500$, assuming a conservative upper coverage time of 50,000 steps. This yields an overall confidence of 99.5%. For omniscient reference, the corresponding average coverage is included; this information is *not* acquired or inferable given the data storage of Alg. 1.

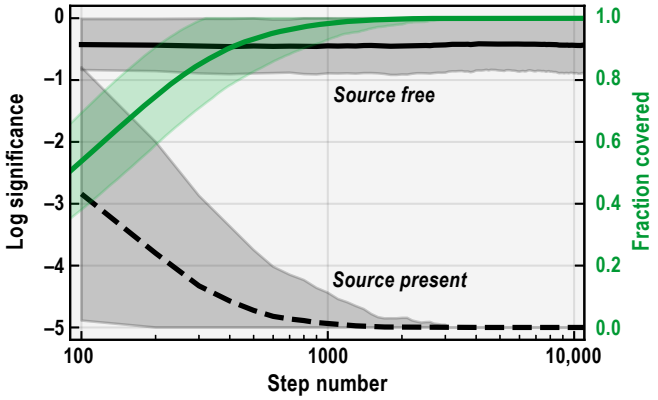

Fig. 4: Evolution of KS test significance and spatial coverage for simulated trials. The results for source-absence (solid) and source-presence (dashed) are averaged over all trials and simulated rooms. Coverage is shown for the source-absence case; coverage for the source-presence case is omitted since the algorithm terminates once a source is confirmed. The log-significance, represented in $\log_{10}$-space, is floored at the KS test trigger threshold, which was -5 for the simulated trials. For each curve, the shaded region represents one standard deviation of the full range of values.

Inspecting Figure 4, we see that full coverage of the environment was always achieved in the source-free case, as evidenced by the converged standard deviation, and there was never a statistically significant difference between the

reference and realized action distributions (no false positives). For the source-present case, we likewise see that all trials correctly converged to the KS test trigger threshold of $\log_{10}(\frac{p^*}{n}) = -5$ (no false negatives). Coverage is not shown for the source-presence case since full coverage is not guaranteed, nor is it required, when a source is present.

B. Hardware Demonstration

We built a simple gamma ray-detecting robot (shown in Figure 1) using the iRobot Create 3 platform, though we note that the proposed methodology is sensing agnostic and applicable to many possible scalar fields and sources. Collision avoidance is accomplished using the onboard infrared sensors, with outputs clipped to ensure the robot does not see beyond a small radius (≈5 cm). The radiation detection unit uses a 2-inch Mirion/Canberra NaI scintillator (Model 802) connected to an Osprey Digital MCA Tube Base. A Raspberry Pi 4 Model B reads the data over Ethernet and relays the counts via Bluetooth. For rudimentary filtering to reduce low energy noise, channels ≥ 400 of the 2048-channel spectrum are summed to yield gross counts. From experimental calibration (i.e., counts versus distance from a source), we set the detector range to be $r_D = 1$ m.

Various laboratory environments were configured. For source-present environments, a set of gamma-ray check sources were used (Cs-137, Ba-133, and Co-60, among other isotopes) totaling to around 9 μCi of activity. The hardware analog to Figure 4 is shown in Figure 5 for trials conducted in two full-scale environments (20 m^2 rectangle and 22 m^2 L-shape). For these trials, the NaI detector was utilized with 1 m^2 bins for tracking coverage. We apply the KS test after every 20 measurements, using $p^* = 0.005$ and $n = 50$, assuming an upper coverage time of 1000 steps. Accordingly, for the hardware trials, the KS test trigger threshold was $\log_{10}(\frac{p^*}{n}) = -4$. As with the simulated trials, these parameters yield an overall confidence of 99.5%.

For both environments, the KS test correctly reached the set threshold in the source-present case and full coverage was achieved without reaching the KS test threshold in the source-free case. Coverage was tracked in real-time to provide an omniscient view (unknown to the robot) of the experiment. The trials were artificially stopped once coverage was reached, but according to Alg. 1, the robot would have continued its inspection until the maximum time was reached. In addition to these large-scale environments, we conducted trials in four smaller rooms of varying complexity, summarized in Table II. For each environment, 10 trials were conducted: 5 with the gamma sources and 5 without the sources. For all trials in all environments, Alg. 1 yielded the correct result: when no source was present, the robot covered the environment without reaching the KS test threshold; when a source was present, the robot more expeditiously reached the threshold, indicative of an anomalous source.

As observed in both Figure 5 and Table II, if a source was present, Alg. 1 terminated in fewer steps than necessary to achieve full coverage of the search environment. This was an interesting empirical result which was not theoretically guaranteed, reinforcing that complete coverage is not necessary in order to identify the presence of an anomalous source; complete coverage remains a necessary condition, though, for confirming the absence of sources with high-confidence.

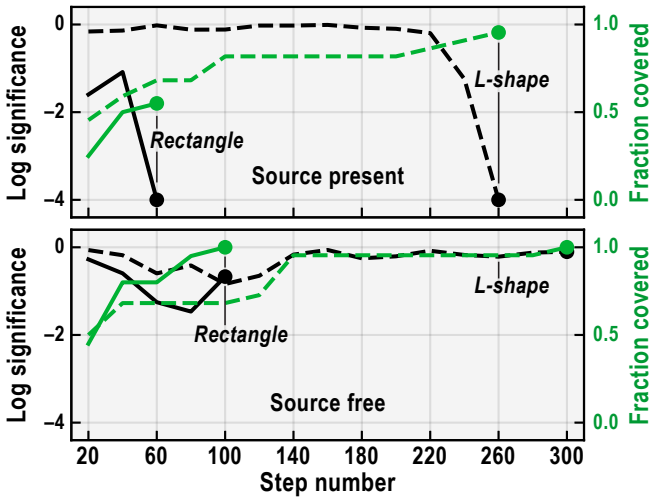

Fig. 5: Evolution of KS test significance and spatial coverage for full-scale laboratory trials. For the source-presence cases (above), the time step when the algorithm terminates is indicated (i.e., once the KS test log-significance, represented in $\log_{10}$-space, reaches -4). For the source-absence cases (below), the time step when coverage is first reached is indicated; note that the algorithm would *not* actually terminate in this case, since coverage is not known to the inspector.

	14 m^2	12 m^2	16 m^2	13 m^2
Simulated cover time	166 ± 47	125 ± 59	265 ± 131	171 ± 63
Hardware cover time	76 ± 22	82 ± 36	94 ± 22	128 ± 42
KS time for source	39 ± 15 (52 ± 16)	50 ± 17 (56 ± 20)	60 ± 37 (72 ± 35)	34 ± 6 (40 ± 0)†

TABLE II Key metrics for laboratory trials. For the four environments shown, the number of steps to reach full coverage for the source-free case and the number of steps to the first instance of $\log_{10}(P) = -4$ for the source case is reported; each table entry includes the average and standard deviation over 5 trials. For our chosen test parameters, Alg. 1 only performs a KS test after every 20 steps; the corresponding time to the algorithm terminating is in parentheses. For reference, the experimental configurations were reproduced in PyBullet; the coverage time for 10 simulated trials is reported. †For all 5 source-presence trials in the 13 m^2 environment, the KS test triggered after 40 measurements (i.e., the second KS test).

VII. DISCUSSION AND FUTURE WORK

We present an algorithm for completing a verification task without requiring, nor revealing, sensitive information from the search environment, defined to include the site layout and any observable features. We derive theoretical guarantees on privacy and correctness that are empirically validated in simulation and in physical hardware experiments.

This work also raises some interesting questions for future exploration. In particular, by more fully characterizing the fundamental trade-off between privacy and time-efficiency for the chosen task, valuable lines of inquiry emerge. Such analysis could uncover optimality conditions for varying

privacy regimes. We note that while our procedure admits privacy analysis, it does not, at present, establish optimality. It may also be possible to show that certain privacy and efficiency constraints are mutually incompatible; such "impossibility results" provide interesting insights, yet are quite rare in robotics. Exploring the broader spectrum of privacy and efficiency may inform the development of new robotic inspection approaches with varying information constraints.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2039656, and the Consortium for Monitoring, Technology, and Verification under Department of Energy National Nuclear Security Administration award number DE-NA0003920.

REFERENCES

[1] J. Tobisch et al. "Remote inspection of adversary-controlled environments". In: *Nature Communications* 14, 6566 (2023).

[2] C. Comley et al. *Confidence, Security & Verification, The Challenge of Global Nuclear Weapons Arms Control*. AWE/TR/2000/001. Atomic Weapons Establishment, 2000.

[3] Defense Threat Reduction Agency. *Radiation Detection Equipment: An Arms Control Verification Tool*. Product No. 211P. Oct. 2011.

[4] R. Aylett and P. A. Vargas. "Social Interaction: Pets, Butlers, or Companions?" In: *Living with Robots: What Every Anxious Human Needs to Know*. The MIT Press, 2023. Chap. 11.

[5] C. Dwork and A. Roth. "The Algorithmic Foundations of Differential Privacy". In: *Foundations and Trends in Theoretical Computer Science* 9.3–4 (2014), pp. 211–407.

[6] National Academies of Sciences, Engineering, and Medicine. *Nuclear Proliferation and Arms Control Monitoring, Detection, and Verification: A National Security Priority: Interim Report*. The National Academies Press, 2021.

[7] E. Lepowsky et al. "Confirming the absence of nuclear warheads via passive gamma-ray measurements". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 990, 164983 (2021).

[8] E. Lepowsky et al. "Ceci N'est Pas Une Bombe: Lessons from a Field Experiment Using Neutron and Gamma Measurements to Confirm the Absence of Nuclear Weapons". In: *Science & Global Security* (2023), pp. 1–12.

[9] F. F. Dean. *ROBIN: A Way to Collect In-Plant Safeguards Data with Minimal Inspector Access*. SAND82-1588C. Sandia National Laboratories, 1982.

[10] K. Robertson et al. "The IAEA Robotics Challenge – Demonstrating Robots for Safeguards Inspections". In: *IAEA Symposium on International Safeguards*. IAEA-CN-267. 2018.

[11] F. E. Schneider and D. Wildermuth. "Real-World Robotic Competitions for Radiological and Nuclear Inspection Tasks". In: *20th International Carpathian Control Conference (ICCC)*. 2019, pp. 1–6.

[12] B. Bird et al. "A Robot to Monitor Nuclear Facilities: Using Autonomous Radiation-Monitoring Assistance to Reduce Risk and Cost". In: *IEEE Robotics & Automation Magazine* 26.1 (2019), pp. 35–43.

[13] F. Mascarich et al. "Autonomous Distributed 3D Radiation Field Estimation for Nuclear Environment Characterization". In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 2163–2169.

[14] A. West et al. "Use of Gaussian process regression for radiation mapping of a nuclear reactor with a mobile robot". In: *Scientific Reports* 11, 13975 (1 2021).

[15] N. A. A. Rahman et al. "A coverage path planning approach for autonomous radiation mapping with a mobile robot". In: *International Journal of Advanced Robotic Systems* 19.4 (2022).

[16] K. Groves et al. "Robotic Exploration of an Unknown Nuclear Environment Using Radiation Informed Autonomous Navigation". In: *Robotics* 10.2, 78 (2021).

[17] D. P. Kingma and M. Welling. "An Introduction to Variational Autoencoders". In: *Foundations and Trends in Machine Learning* 12.4 (2019), pp. 307–392.

[18] P. Vincent et al. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion". In: *Journal of Machine Learning Research* 11 (2010), pp. 3371–3408.

[19] D. P. Kingma and M. Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: 1312.6114 [stat.ML].

[20] M. Booker and A. Majumdar. "Switching Attention in Time-Varying Environments via Bayesian Inference of Abstractions". In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2023, pp. 10174–10180.

[21] V. Pacelli and A. Majumdar. "Robust Control Under Uncertainty via Bounded Rationality and Differential Privacy". In: *International Conference on Robotics and Automation (ICRA)*. 2022, pp. 3467–3474.

[22] M. Booker and A. Majumdar. "Learning to Actively Reduce Memory Requirements for Robot Control Tasks". In: *Proceedings of Machine Learning Research*. Vol. 144. Proceedings of the 3rd Conference on Learning for Dynamics and Control. 2021.

[23] Y. Xu et al. "A Theory of Usable Information Under Computational Constraints". In: *International Conference on Learning Representations*. 2020.

[24] A. Majumdar et al. "Fundamental Limits for Sensor-Based Robot Control". In: *International Journal of Robotics Research* 42.12 (2023), pp. 1051–1069.

[25] C. Dwork et al. "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography*. Ed. by S. Halevi and T. Rabin. Vol. 3876. TCC 2006:

Lecture Notes in Computer Science. Springer, 2006, pp. 265–284.

[26] F. McSherry and K. Talwar. “Mechanism Design via Differential Privacy”. In: *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. 2007, pp. 94–103.

[27] D. Aldous. “An Introduction to Covering Problems for Random Walks on Graphs”. In: *Journal of Theoretical Probability* 2 (1989), pp. 87–99.

[28] R. Aleliunas et al. “Random walks, universal traversal sequences, and the complexity of maze problems”. In: *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*. 1979, pp. 218–223.

[29] D. Aldous and P. Diaconis. “Strong uniform times and finite random walks”. In: *Advances in Applied Mathematics* 8.1 (1987), pp. 69–97.

[30] M. Mihail. “Conductance and convergence of Markov chains-a combinatorial treatment of expanders”. In: *30th Annual Symposium on Foundations of Computer Science*. 1989, pp. 526–531.

[31] A. Sinclair and M. Jerrum. “Approximate counting, uniform generation and rapidly mixing markov chains extended abstract”. In: *Graph-Theoretic Concepts in Computer Science*. Ed. by H. Göttler and H.-J. Schneider. Vol. 314. WG 1987: Lecture Notes in Computer Science. Springer, 1988, pp. 134–148.

[32] F. Ball et al. “On the Mean and Variance of Cover Times for Random Walks on Graphs”. In: *Journal of Mathematical Analysis and Applications* 207.2 (1997), pp. 506–514.

[33] E. Seneta. “Non-negative Matrices and Markov Chains”. In: *Springer Series in Statistics*. Springer, 1981.

[34] D. Serre. “Matrices: Theory and Applications”. In: *Graduate Texts in Mathematics*. Vol. 216. Springer, 2010.

[35] J.-Q. Dong et al. “Universal cover-time distribution of heterogeneous random walks”. In: *Physical Review E* 107.2, 024128 (2023).

[36] M. Chupeau et al. “Cover times of random searches”. In: *Nature Physics* 11.10 (2015), pp. 844–847.

[37] L. Régnier et al. “Universal exploration dynamics of random walks”. In: *Nature Communications* 14.1, 618 (2023).

[38] G. M. Viswanathan et al. “Lévy flights in random searches”. In: *Physica A: Statistical Mechanics and its Applications* 282.1 (2000), pp. 1–12.

[39] A. M. Reynolds and C. J. Rhodes. “The Lévy flight paradigm: random search patterns and mechanisms”. In: *Ecology* 90.4 (2009), pp. 877–887.

[40] B. Pang et al. “Effect of random walk methods on searching efficiency in swarm robots for area exploration”. In: *Applied Intelligence* 51 (7 2021), pp. 5189–5199.

[41] A. Sharma et al. “Sketching curvature for efficient out-of-distribution detection for deep neural networks”. In: *Proceedings of Machine Learning Research*. Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence. 2021, pp. 1958–1967.

[42] R. Sinha et al. *A System-Level View on Out-of-Distribution Data in Robotics*. 2023. arXiv: 2212.14020 [cs.RO].

[43] A. Farid et al. “Failure Prediction with Statistical Guarantees for Vision-Based Robot Control”. en. In: *Robotics: Science and Systems XVIII*. June 2022. ISBN: 978-0-9923747-8-5.

[44] A. Farid et al. “Task-Relevant Failure Detection for Trajectory Predictors in Autonomous Vehicles”. In: *Proceedings of Machine Learning Research*. Vol. 205. Proceedings of The 6th Conference on Robot Learning. 2023, pp. 1959–1969.

[45] R. Luo et al. “Sample-Efficient Safety Assurances Using Conformal Prediction”. In: *Algorithmic Foundations of Robotics XV*. Ed. by S. M. LaValle et al. Vol. 25. WAFR 2022: Springer Proceedings in Advanced Robotics. Springer, 2023, pp. 149–169.

[46] S. M. LaValle and J. James J. Kuffner. “Randomized Kinodynamic Planning”. In: *The International Journal of Robotics Research* 20.5 (2001), pp. 378–400.

[47] J. James J. Kuffner and S. M. LaValle. “RRT-connect: An efficient approach to single-query path planning”. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation.* Vol. 2. 2000, pp. 995–1001.

[48] M. Basseville. “Detecting changes in signals and systems – A survey”. In: *Automatica* 24.3 (1988), pp. 309–326.

[49] A. N. Kolmogorov. “Sulla Determinazione Empirica di Una Legge di Distribuzione”. In: *Giornale dell'Istituto Italiano degli Attuari* 4 (1933), pp. 83–91.

[50] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[51] Y. Cao et al. “Optimal Coverage Path Planning Algorithm of the Tractor-formation Based on Probabilistic Roadmaps”. In: *2019 IEEE International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI)*. 2019, pp. 27–32.

[52] Z. Khanam et al. “An Offline-Online Strategy for Goal-Oriented Coverage Path Planning using A Priori Information”. In: *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*. 2021, pp. 874–881.

[53] S. A. Sadat et al. “Fractal trajectories for online non-uniform aerial coverage”. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2015, pp. 2971–2976.

[54] G. F. Knoll. *Radiation Detection and Measurement*. 4th ed. Wiley, 2010.

[55] E. Coumans and Y. Bai. *PyBullet, a Python module for physics simulation for games, robotics and machine learning*. 2021.

[56] B. Ellenberger. *PyBullet Gymperium*. 2019.

[57] Y. Kadhi et al. *Learning and generalization on a navigation task of a wheeled robot*. 2021.