2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 27 - October 1, 2021. Prague, Czech Republic

2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-1714-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IROS51168.2021.9636029

Optimization-Based Robot Team Exploration Considering Attrition and Communication Constraints*

Matthew A. Schack† John G. Rogers‡ Qi Han† Neil T. Dantam†

***Abstract*— Exploring robots may fail due to environmental hazards. Thus, robots need to account for the possibility of failure to plan the best exploration paths. Optimizing expected utility enables robots to find plans that balance achievable reward with the inherent risks of exploration. Moreover, when robots rendezvous and communicate to exchange observations, they increase the probability that at least one robot is able to return with the map. Optimal exploration is NP-hard, so we apply a constraint-based approach to enable highly-engineered solution techniques. We model exploration under the possibility of robot failure and communication constraints as an integer, linear program and a generalization of the Vehicle Routing Problem. Empirically, we show that for several scenarios, this formulation produces paths within 50% of a theoretical optimum and achieves twice as much reward as a baseline greedy approach.**

I. Introduction

Using multiple robots to explore an unknown area has the potential to construct maps more efficiently by exploring multiple regions simultaneously. Yet robots face hazards in many scenarios [1]; conditions in the environment may cause robots to get stuck, lost, or otherwise fail. Robots that fail before communicating new observations will not contribute to the team's map, so robots may need to form subteams that explore together to ensure that at least one robot in each subteam transmits the map updates. Moreover, wireless communication between robots itself presents challenges due to communication range limits or obstacles, so the team may not know that a robot outside communication range has failed. To best explore, a robot team must balance efficiency and speed of concurrent exploration with robustness of forming subteams with multiple robots. Furthermore, exploration will typically reveal new areas to explore; thus, exactly computing optimal explorations paths from limited initial knowledge is not generally possible. Instead, each robot subteam must be able to quickly update plans with new information.

To address these aforementioned issues, we adopt the game-theoretic notion of *expected utility* [2], providing a metric for reward (i.e., new information) over a path balanced with the likelihood of achieving the reward (i.e., the probability that the robot will not fail). While the expected utility of a particular path can be directly and efficiently computed using a probabilistic graphical model, planning the utility-optimal path presents a computational challenge. We prove that finding the utility-optimal path is NP-hard through a reduction from the NP-hard Traveling Salesman Problem [3].

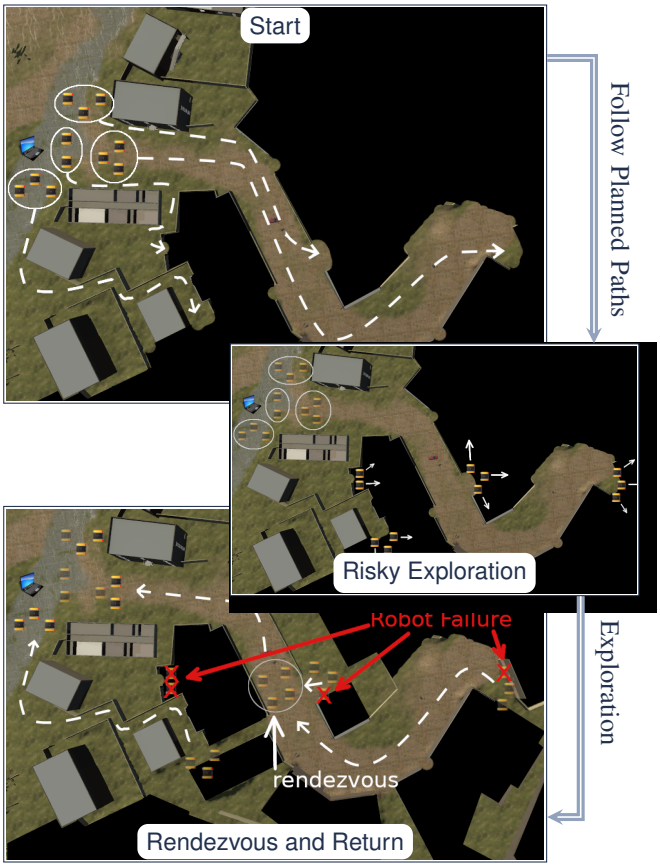


Fig. 1: Overview of exploration with communication and attrition. All robots start within communication range and plan assignments of unknown space to explore as well as rendezvous points and times to communicate observations. Robots may fail, so they work in subteams to improve reliability. At the scheduled time, robots rendezvous to share info with other subteams and the base station.

We present an optimization-based approach for multi-robot exploration that maximizes expected utility while accounting for the possibility of robot attrition and communication constraints (Fig. 1). The exact optimization problem is inherently nonlinear, so it is generally not possible to efficiently find the true optimum (see Sec. III). Instead, we design an integer, linear program to efficiently approximate the optimal solution (see Sec. IV), and we develop a recursive subteaming approach to explore newly revealed areas (see

* This work was supported in part by the ARL DCIST CRA [W911NF-17-2-0181] and the NSF [CNS-1823245]. This work relates to Department of Navy award N00014-211-2418 issued by the Office of Naval Research.

† Department of Computer Science, Colorado School of Mines, USA. {mschack,qhan,ndantam}@mines.edu.

‡ United States Army Research Laboratory, Adelphi, MD, USA. john.g.rogers59.civ@mail.mil

Sec. IV-G). We empirically evaluate the optimality of our approximation and its computational performance in Sec. V.

II. Related Work

This paper addresses robust exploration using multiple robots and under the possibility of robot failure. We next review related work in robot exploration, coverage, and navigation under uncertainty.

Leading approaches in exploration focus on efficient management of a group of robots and scalability to many robots. Decentralized approaches [4], [5], [6] have shown the ability to scale to many robots, but they may get stuck in local optima. Conversely, algorithms that incorporate information gain [7], [8], [9], approaches for active information acquisition [10], [11], or goal assignment [12], [13] produce results closer to a globally optimal path but at a cost to scalability. Additionally, exploration algorithms that consider communication requirements [14], [15] provide insight applicable to cluttered or otherwise constrained environments. Our work differs from existing works by explicitly addressing the possibility of robots failing during exploration.

The robot coverage problem tasks a group of robots to revisit an already-explored area with the possibility of robot failures. Leading approaches account for failures by decomposing the space into regions and adding a robot to existing regions if a robot fails [16], taking over a failed robot's work once a different robot completes its assigned task [17], or immediately reassigning robots when a robot fails [18]. However, in the coverage problem there is no unknown area. Also, communication is typically assumed to be perfect, which implies that the robot group is immediately aware of a robot failure. Our work focuses on exploring new areas under communication constraints, i.e., a robot failure may not be immediately detected by other robots.

Works that consider uncertain exploration can also be applied to mitigate risk during exploration. Leading approaches for uncertain exploration construct a Partially Observable Markov Decision Process (POMDP) [19], [20], which computes a policy for what the robots should do in all possible scenarios. While such approaches generate optimal solutions, as the size of the map and the number of robots grow, the computational time to make the policy increases—*the curse of dimensionality*. In contrast, our approach computes a single path, which while not guaranteed to be optimal, can be calculated more quickly.

III. Problem Definition and Probabilistic Model

We define *the attrition-aware multi-robot exploration problem* (*AAMREP*) where we seek to maximize the expected utility at the base station—i.e., maximize the information gained at the base station while accounting for the possible robot failures during exploration. Robots communicate observations to teammates and the base station; however, a robot that fails during exploration loses all the untransmitted information—i.e., we adopt a conservative approach and assume all failure is catastrophic. Thus, optimal paths need to balance the risk of failure during exploration with the estimated information gain reward. We next define the variables and functions for cost and information gain that will be used in our problem definition.

Definition 1. $\Sigma = \left(\mathcal{R}, \mathcal{X}, x^{[0]}, b, \mathcal{W}_c, \mathcal{W}_g, \mathcal{W}_k, a, \mathbb{M}_{known}\right)$ *where,*

- $\mathcal{R}$ *is a finite set of robots,*
- $\mathcal{X} = \mathcal{M}^{|\mathcal{R}|}$ *is the multi-robot configuration space, where each* $\mathcal{M}$ *is the space of a single robot—e.g.* $\mathcal{SE}(2)$*,* $\mathcal{SE}(3)$*, or* $\mathbb{R}^n$*,*
- $x^{[0]} \in \mathcal{X}$ *is the initial multi-robot configuration.*
- $b \in \mathcal{M}$ *is the position of the base station,*
- $\mathcal{W}_c : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}$ *is a signed communication function, where positive values indicate that communication is possible between the two positions,*
- $\mathcal{W}_g : \mathcal{M} \mapsto \mathbb{R}^+$ *is a function that maps from an observed point to information gained,*
- $\mathcal{W}_k : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}^+$ *is a cost function to move between two points—e.g., distance, time, or energy.*
- $a \in [0, 1]$ *is an attrition probability per unit cost,*
- $\mathbb{M}_{known} \subseteq \mathcal{M}$ *is the known map.*

Single-robot space $\mathcal{M}$ consists of disjoint free space $\mathcal{M}_{\texttt{free}}$ and obstacle region $\mathcal{M}_{\texttt{obs}}$. A multi-robot configuration is valid if the position of each individual robot is free: $\mathcal{X}_{\text{valid}} = \left\{ (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{|\mathcal{R}|}) \in \mathcal{X} \;\middle|\; \text{each } \boldsymbol{m}_i \in \mathcal{M}_{\texttt{free}} \right\}$.

Initially, the robots have knowledge only of $\mathbb{M}_{\texttt{known}}$, and the rest of the map is unknown: $\mathbb{M}_{\texttt{unknown}} = \mathcal{M} \setminus \mathbb{M}_{\texttt{known}}$. We assume a discrete, finite representation of the map, e.g., an occupancy grid, roadmap, or octree. Note that positions in $\mathbb{M}_{\texttt{unknown}}$ can still be in free space $\mathcal{M}_{\texttt{free}}$ or the obstacle region $\mathcal{M}_{\texttt{obs}}$, thus attempting to plan through $\mathbb{M}_{\texttt{unknown}}$ could result in infeasible paths.

The base station gains information by communicating with robots, and robots gain information either by direct observation (i.e., traveling to an unknown area and sensing the area) or by obtaining information via communicating with other robots. Information gained at the base station is the sum of $\mathcal{W}_g$ over observations received. However, we must account for the possibility that robots fail before communicating observations. Thus, we find optimal paths based not on reward—i.e., information gain, but on game-theoretic *expected utility*—i.e., reward scaled by likelihood of achieving that reward.

A. Observation Likelihoods

The likelihood of a robot or the base station receiving an observation depends on (1) a robot reaching a position to make the observation and (2) two robots or the base station communicating the observation. We model likelihoods using random variables representing robot i arriving at a position, x_i; observing an unknown point, o_i; or communicating with robot or base station j, $c_{i,j}$. Since we assume a discrete map of $\mathcal{M}$, there are finite number of positions and observations to consider. For any specific path $\boldsymbol{\sigma} : [0, 1] \mapsto \mathcal{X}$, there are finite number of events, which occur when robots reach a position to observe or communicate, so we need to only consider random variables over a finite number of timesteps. We model the

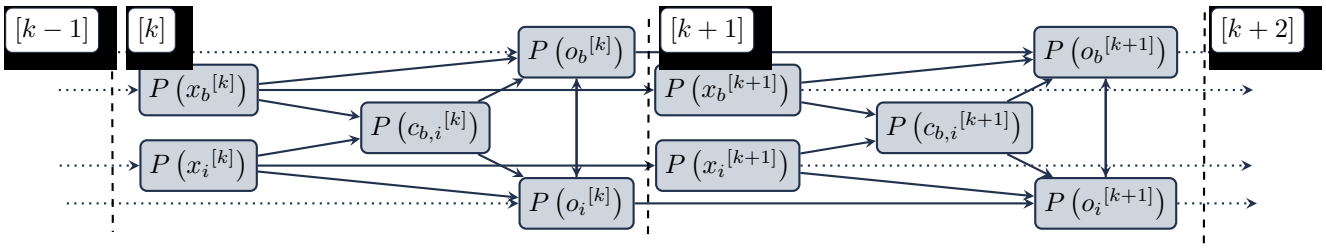


Fig. 2: Dynamic Bayesian Network fragment for a robot i and base station b over two timesteps, k, $k+1$. The variables represent the likelihood of a robot or the base station arriving at a position (x), communicating with one another (c), and observing a single point (o). Subscripts denote the robot the random variable refers to and arrows on the DBN correspond to dependencies between the random variables. Horizontal dotted lines show dependencies from preceding to future timesteps, and the dashed lines separate the timesteps.

random variables and dependencies as a *Dynamic Bayesian Network* (DBN) in Fig. 2.

The likelihood that robot i reaches position x at timestep k and for path $\boldsymbol{\sigma}_i$ depends on accumulated cost ($\mathcal{W}_k$) and the attrition rate (a):

$$P\left(x_i{}^{[k]} \middle| x_i{}^{[k-1]}\right) = (1-a)^{\mathcal{W}_k(\boldsymbol{\sigma}_i(k),\boldsymbol{\sigma}_i(k-1))} \ . \quad (1)$$

The likelihood that two robots can communicate is the probability that a network path exists between them. A network path exists when $\mathcal{W}_c \geq 0$ for all robots in the path. $\mathcal{W}_c$ is deterministic, so the uncertainty arises from the probability of a robot reaching a position according to attrition rate (1)—i.e., the probability that the robot has not failed.

For a specific path $\boldsymbol{\sigma}$, we construct and evaluate a DBN to determine the probabilities that the base station has each observation o at final timestep h',

$$L(o, \boldsymbol{\sigma}) = P\left(o_b{}^{[h']}\right) \ . \quad (2)$$

B. Expected Utility

The expected utility $\mathcal{W}_e$ for path $\boldsymbol{\sigma}$ depends on likelihoods of observations L and corresponding information gain $\mathcal{W}_g$,

$$\mathcal{W}_e\left(\boldsymbol{\sigma}\right) = \sum_{o \in \mathbb{M}_{\texttt{unknown}}} L\left(o, \boldsymbol{\sigma}\right) \mathcal{W}_g\left(o\right) \ . \quad (3)$$

We assume that the gain from each observed point, o, is independent of other points, consistent with other information theoretic works [21] and typical occupancy grid assumptions.

The solution to the problem defined by Def. 1 is a path $\boldsymbol{\sigma}$ that is feasible and maximizes expected utility.

Definition 2. *An information-optimal path σ solves,*

$$\max_{\boldsymbol{\sigma}} \quad \mathcal{W}_e\left(\boldsymbol{\sigma}\right)$$
$$\text{s.t.} \quad \boldsymbol{\sigma}[0,1] \in \mathcal{X}_{\text{valid}} \wedge \boldsymbol{\sigma}(0) = x^{[0]} \ .$$

Def. 2 specifies optimal paths over the entire space $\mathcal{M}$. However, paths that move into unknown space could be infeasible ($\mathbb{M}_{\texttt{unknown}} \cap \mathcal{M}_{\texttt{obs}}$), and paths that stay in known map $\mathbb{M}_{\texttt{known}}$ may be suboptimal, as the optimal path could move through the unknown but still valid space ($\mathbb{M}_{\texttt{unknown}} \cap \mathcal{M}_{\texttt{free}}$). Consequently, we must incorporate new observations during exploration. Precomputing a policy to respond to all combinations of free and obstacle portions of $\mathbb{M}_{\texttt{unknown}}$ would be computationally challenging, if not intractable. Instead, we develop a recursive approach to explore and plan.

IV. Approach

We develop an approach to find paths that optimize expected utility (i.e., expected information gain) according to Def. 2. When the entire robot team is in communication, the team divides into subteams, and they plan paths to the boundaries of the known space—i.e., frontiers [22]—along with rendezvous locations to communicate their observations after exploration. Then, when subteams reach frontiers, they recursively plan based on the newly explored space. Finally, subteams rejoin at rendezvous points and ultimately relay observations to the base station.

Because most of the information comes from unknown points, we only consider reward from unknown space, rather than over the entire path, to improve scalability. To estimate the maximum reward from exploring unknown space, we partition the unknown space into frontier regions. There are many partition approaches, such as using k-means clustering [23] or Voronoi diagrams [24]. Empirically, we found a Voronoi-based partitioning worked well for our experiments. Based on the partitioning, we estimate the maximum reward for a single frontier by summing information gain $\mathcal{W}_g$ over all unknown points—i.e., we assume unknown space is free.

Our approach plans paths for robots by choosing which robots should explore which frontiers in which order before rendezvous with other robots and finally returning to the base station. The Vehicle Routing Problem (VRP) [25] addresses robots that visit a sequence of points and return to a starting location—without communicating. Thus, we formulate our approach as an extension of the VRP to include the ability to communicate.

A. Background on the Vehicle Routing Problem

We briefly summarize the typical formulation for the VRP; please see [25] for a more thorough discussion. The VRP is defined as $(\mathcal{R}, \mathcal{Q})$, where $\mathcal{R}$ is a set of vehicles (robots) and $\mathcal{Q}$ is an undirected, weighted graph of points to visit (goals) as well as start s and end e locations. s and e represent the same physical location but are represented separately in $\mathcal{Q}$ to track when robots arrive at the end. Every graph edge has two weights, a path cost $c_{qq'}$—given by $\mathcal{W}_k$—and a time, $t_{qq'}$, needed to travel between any two points. The vehicles must visit every goal in $\mathcal{Q}$ and return to e before the final time h while minimizing the total travel cost.

The VRP is typically modeled as an integer, linear program (ILP) with decision variables specifying paths for each robot. The typical model for the VRP contains continuous variables x_{rq} representing when robot r arrives at point q and binary decision variables encoding paths, $\alpha_{rqq'}$, defined to be one if and only if robot r travels between from q to q'. Robots can remain at a point, so we add another continuous variable y_{rq} representing the time a robot leaves a point.

Solutions to the VRP minimize the path cost, while ensuring that some robot visits every goal before the final time, h. The constraints ensure that a path starts at s, visits some sequence of points, and then returns to e. The solution is a continuous time path given by the arrival and leaving times (x_{rq} and y_{rq}) for every point.

Some requirements of the VRP couple continuous and binary decision variables, which could, result in a nonlinear constraint if binary variables are multiplied by continuous ones. To formulate such requirements as linear constraints—e.g., (10), [25] formulates all constraints comparing binary and continuous decision variables as inequality constraints, and multiplies the binary variable by a constant large enough that the constraint is always satisfied for one value of $\alpha_{rqq'}$. We use this technique—i.e., big M [26], in our formulation.

Fig. 3 summarizes the objectives and constraints of the VRP. Constraints (5) and (6) ensure that every robot begins at starting location s and ends at final location e. Constraint (7) ensures each robot enters and leaves a point through exactly one path, and constraint (8) ensures that each point is visited exactly once. Constraints (9) and (10) ensure that the time when a robot leaves a point, y_{rq}, is after it arrives, x_{rq}, and before it arrives at the next. Constraints (11) and (12) ensure that no points are visited after the final time, h.

B. Extensions to the VRP and NP-Hardness

We extend the VRP formulation in Sec. IV-A to address the exploration problem of Def. 2. The globally optimal path may move through unknown space, so it is not possible to exactly compute such a path. Instead, we finds paths through known space to reach frontiers while deciding the amount of time for exploration of frontiers, a decision not addressed by the VRP. Additionally, the VRP explicitly constrains each robot to return to the ending location; though returning to an end location is not an explicit requirement in our problem, we prove in Sec. IV-C there is an optimal path where every robot returns to the base station. By planning for time to explore frontiers and eventually returning to the base station, robot subteams can recursively plan when they arrive at the frontier, finding a sequence of near-optimal paths.

Next, we prove that *AAMREP* is NP-Hard through a reduction from the NP-hard Traveling Salesman Problem [3]. We reduce the TSP to *AAMREP* by constructing an *AAMREP* with one robot and every TSP goal point as a frontier with information gain large enough that the optimal solution is to visit every frontier. Since there are no other robots to communicate with, the solution to Def. 2 would be to visit every frontier at the lowest cost. Thus an answer to *AAMREP* would be an answer to TSP, therefore *AAMREP* is NP-hard.

$$\min \quad \sum_{r\in\mathcal{R}}\sum_{q\in\mathcal{Q}}\sum_{q'\in\mathcal{Q}} c_{qq'}\alpha_{rqq'} \tag{4}$$

$$\text{s.t.} \quad \sum_{q\in\mathcal{Q}} \alpha_{rsq} = 1 \qquad \forall r \in \mathcal{R} \tag{5}$$

$$\sum_{q\in\mathcal{Q}} \alpha_{rqe} = 1 \qquad \forall r \in \mathcal{R} \tag{6}$$

$$\sum_{q'\in\mathcal{Q}} \alpha_{rq'q} = \sum_{q'\in\mathcal{Q}} \alpha_{rqq'} \qquad \forall r \in \mathcal{R}, q \in \mathcal{Q}\setminus\{s,e\} \tag{7}$$

$$\sum_{r\in\mathcal{R}}\sum_{q'\in\mathcal{Q}} \alpha_{rq'q} = 1 \qquad \forall q \in \mathcal{Q}\setminus\{s,e\} \tag{8}$$

$$x_{rq} \le y_{rq} \qquad \forall r \in \mathcal{R}, q \in \mathcal{Q} \tag{9}$$

$$y_{rq} + t_{qq'} - x_{rq'} \le (h + t_{qq'})(1 - \alpha_{rqq'}) \qquad \forall r \in \mathcal{R}, q, q' \in \mathcal{Q} \tag{10}$$

$$y_{rq} \le h \sum_{q'\in\mathcal{Q}} \alpha_{rqq'} \qquad \forall r \in \mathcal{R}, q \in \mathcal{Q} \tag{11}$$

$$y_{re} = h \qquad \forall r \in \mathcal{R} \tag{12}$$

Fig. 3: Objective and constraints for the VRP [25]. $\mathcal{R}$ is the set of robots and $\mathcal{Q}$ is the set of points. $\alpha_{rqq'}$ represents the choice for robot r to take the path from q to q', where $c_{qq'}$ and $t_{qq'}$ are the cost and travel time of that path respectively. We track the arrival time, x_{rq}, and departure time, y_{rq}, for each robot r at each point q. h represents the final possible time, and points s and e represent the start and end locations.

The NP-hardness of *AAMREP* means that optimal solutions will, in general, be computationally intractable, even when planning through $\mathbb{M}_{\texttt{known}}$. Thus, the rest of our approach focuses on tractable approximations of an optimal solution.

C. Ending paths at the base station

Though *AAMREP* does not explicitly require robots to return to the base station, we prove that an optimal path must exist in which every robot returns to the base station. Since there is such an optimal path ending at the base station, we are able to retain constraint (6) for robots to reach the base station. We prove robots may optimally complete paths at the base station by showing, for all cases where a robot ends its path elsewhere, there is a path where the robot ends at the base station with the same or greater expected utility.

Proposition 1. *A robot, whose (1) path does not end at the base station and (2) observations are all received by the base station, has expected utility equal to a path where the robot ends at the base station.*

Proof. Expected utility increases either by (1) increasing likelihood a particular observation is received by the base station; or (2) new observations being received by the base station. The base station has all the robot's observations, so the robot cannot (1) increase (or decrease) likelihood of the base station having an observation or (2) provide new observations to the base station. □

Proposition 2. *If a robot's path ends with observations not received by the base station, then there exists a path with greater or equal expected utility ending at the base station.*

Proof. Likelihood is strictly non-negative. Regardless of path cost, returning to the base station will never decrease likelihood of the base station receiving the untransmitted observation. Thus, it is always optimal to return to the base station with untransmitted observations. □

Proposition 3. *An optimal path must exist where every robot ends its path at the base station.*

Proof. A robot that does not end at the base station either has or has not transmitted all of its observations to the base station. Prop. 1 and Prop. 2 prove for both cases that returning to the base station provides equal or greater expected utility. □

D. Robot Teaming and Exploration

Unlike the VRP, *AAMREP* permits multiple robots to visit frontiers for exploration or a rendezvous points for communication. We model multiple robots in a subteam as the set of robots visiting the same frontier. The behavior of a subteam exploring a frontier is different than their behavior at a rendezvous point—i.e., they recursively plan to explore unknown space at a frontier, and they wait and communicate at a rendezvous points. To distinguish frontiers and regular space, we create a new set of points, $f \in \mathcal{F} \subseteq \mathcal{Q}$, that contain just the frontiers—i.e., points where robots can explore. We keep non-frontier points for possible communication locations—e.g., the base station. We capture the decision for a robot to explore a frontier with β_{rf} which is one if and only if robot r explores frontier f. We define a subteam as the set of robots r exploring the same frontier f, $\{r \in \mathcal{R} \mid \beta_{rf} = 1\}$.

We remove constraint (8) so multiple robots may visit the same point, and add (13) so any robot visiting a frontier, given by α_{rqf}, is in the subteam exploring it.

$$\sum_{q' \in \mathcal{Q}} \alpha_{rq'f} = \beta_{rf} \qquad \forall r \in \mathcal{R} f \in \mathcal{F} \quad (13)$$

We ensure that robot subteams remain together by restricting teams to arrive and leave at the same time.

E. Exploration Constraints and Objective

When a subteam explores a frontier, it gains information dependent upon its observations. We model the available information and how much information a subteam obtains from exploring a frontier. However, we do not know the composition of unknown space (by definition), so we estimate the maximum information available at a frontier.

We estimate the maximum information gained from a frontier with the constant, d_f, defined as the sum of $\mathcal{W}_g$ over an entire frontier, and add the decision variable z_{rf} as the amount of information robot r gains from frontier f. Tracking which robot has information about which frontier is not necessary for computing the reward at the base station (as all robots will return to the base station); however, we must know which robots have observations about which frontiers for correct communication in Sec. IV-F.

We construct a linear approximation of (3) for an objective function in the ILP, approximating the optimal solution to Def. 2. We consider paths through only known space, though the true optimal solution may move through unknown space, which is not possible to find. However, our experiments show that the linear approximation was within around 50% of the theoretical upper bound of optimal paths through known space. Expected utility (3) can increase three ways: new information arriving at the base station, increased likelihood from more robots transmitting observations, and increased likelihood of a robot arriving at a point (given by (1)) by choosing a lower cost path. Thus our objective function seeks to increase the reward at the base station (z_{bf}) and obtained by every robot (z_{rf}) while minimizing the cost accrued ($c_{qq'}$). We scale a robot's reward and the path cost by the attrition rate, a, as the change in likelihood from more transmissions or lower cost is directly related to the attrition rate.

$$\max \sum_{f \in \mathcal{F}} \left(z_{bf} + a \sum_{r \in \mathcal{R} \setminus \{b\}} z_{rf} \right) - a \sum_{r \in \mathcal{R}} \sum_{q \in \mathcal{Q}} \sum_{q' \in \mathcal{Q}} c_{qq'} \alpha_{rqq'} \quad (14)$$

We consider the base station as a stationary robot and add it to $\mathcal{R}$. We further add constraints dictating that the base station immediately travels and stays at the ending point.

We plan for additional time for the robots to explore a frontier. When more robots are sent to a frontier, or more time is spent exploring, the reward increases up to the maximum estimated amount, d_f. The benefit from adding more robots does not necessarily scale linearly with the number of robots—e.g., exploring a narrow hallway is not faster with more robots. Thus, we add a hyperparameter representing a set of exploration rates, $j \in \mathcal{J}$, describing information gain per unit time. If at least n_j robots explore, they achieve rate m_j. We decide rates using binary variable, γ_{fj} which is one if and only if exploration rate j is used to explore frontier f. Ideal values of m_j depend on the environment. We would expect the exploration rate to increase linearly for open environments due to simultaneous exploration, but have diminishing returns in cluttered environments.

We define constraints for the reward from frontier exploration. Constraint (15) limits the reward for a robot from frontier f, described by β_{rf}, or unconstrained otherwise. Constraints (16)-(18) ensure we choose at most one exploration rate with the proper number of robots if exploring.

$$z_{rf} \leq m_j (y_{rf} - x_{rf}) + d_f (1 - \gamma_{fj}) + d_f (1 - \beta_{rf}) \qquad \forall r \in \mathcal{R}, f \in \mathcal{F}, j \in \mathcal{J} \quad (15)$$

$$n_j \gamma_{fj} \leq \sum_{r \in \mathcal{R}} \beta_{rf} \qquad \forall f \in \mathcal{F}, j \in \mathcal{J} \quad (16)$$

$$||\mathcal{R}|| \sum_{j \in \mathcal{J}} \gamma_{fj} \geq \sum_{r \in \mathcal{R}} \beta_{rf} \qquad \forall f \in \mathcal{F} \quad (17)$$

$$\sum_{j \in \mathcal{J}} \gamma_{fj} \leq 1 \qquad \forall f \in \mathcal{F} \quad (18)$$

F. Communication Constraints

The likelihood of the base station receiving an observation increases when multiple robots attempt to send the same observation. Each robot has some probability of failure, so communication from multiple robots increases the likelihood that any one robot arrives back at the base station with the information. We model information exchange between robots and communication with the base station.

We model rendezvous points to communicate by expanding $\mathcal{Q}$ to include more points within $\mathcal{M}_{\texttt{known}}$. There is an optimality-scalability trade-off in adding rendezvous points; more rendezvous points will better approximate the optimal solution at the cost of additional decision variables. In our experiments we expanded $\mathcal{Q}$ to include the midpoint on the path between any two frontiers and the midpoint between the base station and any frontier. Alternatively, rendezvous points could be chosen by sampling the space or choosing specific features, such as intersections, on the map.

We model communication with a binary variable, $\eta_{rr'fq}$, that is one if and only if robot r communicates information about frontier f to r' at point q.

To ensure robots have an observation before they communicate it, we add variable w_{rf} for the time that robot r obtains observations from f, and add constraint (19) to limit when a robot obtains new information when exploring a frontier.

$$y_{rf} \leq w_{rf} + h\,(1 - \beta_{rf}) \qquad \forall r \in \mathcal{R}, f \in \mathcal{F} \quad (19)$$

The reward a robot gains from communication, $z_{r'f}$, is limited to what information was transmitted. We add constraint (20) to limit reward from communication based upon what the communicating robot knows or unconstrained (i.e., d_f) if not obtaining information by communication.

$$z_{r'f} \leq z_{rf} + d_f \left(1 - \sum_{q \in \mathcal{Q}} \eta_{rr'fq}\right) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F} \quad (20)$$

To bound the maximum reward a robot can receive, we define (21) to be the maximum amount available at a frontier if it receives any information and zero otherwise. This constraint combined with the other two that limit reward ((15) and (20)) ensures that the reward a robot receives reflects the actions it performs—e.g., exploration, communication, or neither. Constraint (22) limits the ways a robot can receive information to either communication or exploration, as a robot that has explored a frontier already knows the information.

$$z_{rf} \leq r_f \left(\sum_{r' \in \mathcal{R}} \sum_{q \in \mathcal{Q}} \eta_{r'rfq} + \beta_{rf}\right) \qquad \forall r \in \mathcal{R}, f \in \mathcal{F} \quad (21)$$

$$\sum_{r' \in \mathcal{R}} \sum_{q \in \mathcal{Q}} \eta_{r'rfq} + \beta_{rf} \leq 1 \qquad \forall r \in \mathcal{R}, f \in \mathcal{F} \quad (22)$$

When two robots rendezvous to communicate, we must ensure that they arrive at the same point, given by $\eta_{rr'fq}$, at the same time (given by the arrival, x_{rq} and departure, y_{rq}, time), after one robot has learned the information, w_{rf}. We model these three conditions with the constraints (23)-(27).

(23) limits the communication to at most one rendezvous point. (24) and (25) constrain both robots to be at rendezvous point at the same time. (26) ensures the transmitting robot has the information before arriving, and (27) enforces the receiving robot has the information after leaving.

$$\sum_{q} \eta_{rr'fq} \leq 1 \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F} \quad (23)$$

$$x_{r'q} \leq y_{rq} + h\,(1 - \eta_{rr'fq}) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F}, q \in \mathcal{Q} \quad (24)$$

$$x_{rq} \leq y_{r'q} + h\,(1 - \eta_{rr'fq}) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F}, q \in \mathcal{Q} \quad (25)$$

$$w_{r'f} \geq w_{rf} - h\left(1 - \sum_{q} \eta_{rr'fq}\right) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F} \quad (26)$$

$$w_{r'f} \geq y_{r'q} - h\,(1 - \eta_{rr'fq}) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F}, q \in \mathcal{Q} \quad (27)$$

Communication can occur from any robot that knows information to one that does not; it does not need to be from the robot that directly explored the frontier. To ensure that the origin of any information is from exploration, we model flow of information through the variable $v_{rr'f}$ representing network flow from r to r' about f.

Constraint (28) enforces that information about a frontier is only generated by exploring a frontier, and constraint (29) ensures information only flows between robots that have communicated. These constraints ensure that all communicated information starts with an exploring robot.

$$(||\mathcal{R}|| - 1)\beta_{rf} + \sum_{r'} v_{r'rf} \geq \sum_{r'} v_{rr'f} + \sum_{r'} \sum_{q} \eta_{r'rfq} \qquad \forall r \in \mathcal{R}, f \in \mathcal{F} \quad (28)$$

$$v_{rr'f} \leq ||\mathcal{R}|| \left(\sum_{q} \eta_{r'rfq}\right) \qquad \forall r, r' \in \mathcal{R}, f \in \mathcal{F} \quad (29)$$

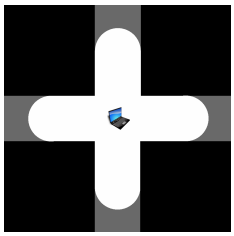
We limit communication at the base station to ensure any rendezvous point increases the likelihood of the base station receiving information, rather than just increase (14).

$$\eta_{rr'fe} = 0 \qquad \forall r \in \mathcal{R}, r' \in \mathcal{R} \setminus \{b\}, f \in \mathcal{F} \quad (30)$$

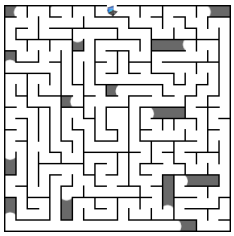
Lastly, we add typical non-negativity and binary constraints.

G. Recursive Exploration

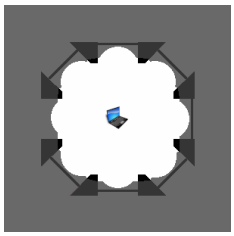
We develop a recursive exploration strategy (Alg. 1) using the model in Sec. IV. Each subteam plans division into new subteams, frontier exploration, movement, and rendezvous (line 2). Subteams execute their plans in parallel (line 10). The robots record new observations during movement (line 17), and recursively plan (line 15) with new observations at frontiers, before returning to any previously planned rendezvous.



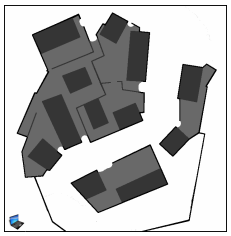
(a) A Four Way Intersection



(b) A Maze Environment



(c) Caged Environment



(d) Camp Lejeune

Fig. 4: Maps of the simulated environments used in our experiments. Dark and light gray points represent occluded and free unknown space respectively. The laptop represents the base station and every robot's starting location.

Algorithm 1: Recursive subteam exploration.

Input: $\mathcal{R}, \mathbb{M}, \mathcal{F}, h$ // team, map, frontiers, final time
Output: $\mathcal{R}, \mathbb{M}.\mathcal{F}$ // new team, map, and frontiers
function $\texttt{Plan}(\mathcal{R}, \mathbb{M}, \mathcal{F}, h)$ **is**
  // $\mathcal{L}$ is a start time ordered list of actions for $\mathcal{R}$
  $\mathcal{L} \leftarrow \texttt{optimize}(\mathcal{R}, \mathbb{M}, \mathcal{F}, h)$ // Sec. IV
  return $\texttt{Execute}(\mathcal{R}, \mathbb{M}, \mathcal{F}, \mathcal{L})$
function $\texttt{Execute}(\mathcal{R}, \mathbb{M}, \mathcal{F}, \mathcal{L})$ **is**
  if $\mathcal{L} = ()$ **then return** $\mathbb{M}$
  $(l,\ \mathcal{R}',\ h',\ q) \leftarrow \texttt{first}(\mathcal{L})$
  if $\mathcal{R} \cap \mathcal{R}' = \emptyset$ **then** // Skip others' actions
    return $\texttt{Execute}(\mathcal{R}, \mathbb{M}, \mathcal{F}, \texttt{rest}(\mathcal{L}))$
  else if $\mathcal{R} \not\subseteq \mathcal{R}'$ **then** // Fork subteam
    fork $\texttt{Execute}(\mathcal{R}', \mathbb{M}, \mathcal{F}, \mathcal{L})$
    return $\texttt{Execute}(\mathcal{R} \setminus \mathcal{R}', \mathbb{M}, \mathcal{F}, \texttt{rest}(\mathcal{L}))$
  else // $\mathcal{R}' \supseteq \mathcal{R}$
    switch l **do** // Perform action
      case `PLAN` **do**
        // Recursively explore the frontier
        $\mathcal{R}, \mathbb{M}, \mathcal{F} \leftarrow \texttt{Plan}(\mathcal{R}, \mathbb{M}, \mathcal{F}, h')$
      case `MOVE` **do** // Robots can fail here
        $\mathcal{R}, \mathbb{M}, \mathcal{F} \leftarrow \texttt{Move}(\mathcal{R}, \mathbb{M}, q)$
      case `COMMUNICATE` **do** // Merge subteams
        $\mathcal{R}, \mathbb{M}, \mathcal{F} \leftarrow \texttt{Join}(\mathcal{R}', \mathbb{M})$
    return $\texttt{Execute}(\mathcal{R}, \mathbb{M}, \mathcal{F}, \texttt{rest}(\mathcal{L}))$

V. Experiment

We evaluate the optimality of a single iteration of our algorithm and simulate exploration in the scenarios in Fig. 4. Scenarios Fig. 4a, Fig. 4b, and Fig. 4c test basic execution, scalability to many frontiers, and performance with an inaccurate estimate of information gain respectively. Fig. 4d is a map of Camp Lejeune, a military training center in North Carolina, testing performance in a real world scenario. As a baseline, we compare against a theoretical upper bound on optimality (Sec. V-A) and a greedy search.

A. Upper Bound

We compare our results against an upper bound on the expected utility from a single iteration. We construct the upper bound by assuming the likelihood a robot can arrive and return from a frontier is independent of the likelihood it can arrive and return from any other frontier. With this assumption, the optimal solution is for every robot to form one subteam and go from the base station to each frontier and return with the maximum reward, d'_f:

$$\sum_{f \in \mathcal{F}} \left(1 - \prod_{r \in \mathcal{R}} \left(1 - (1-a)^{2c_{fe}}\right)\right) d'_f \,. \tag{31}$$

The likelihood the robots obtain the reward from a frontier is still bounded by the cost to go to the frontier, c_{fe}, and the attrition rate, a, so we expect the upper bound to be close to the true optimal answer. For Fig. 4a, our approach was within 0.0% of the upper bound, meaning the upper bound was the optimum for this problem.

B. Greedy Search Baseline

We compare against a greedy baseline that assigns robots to the frontier that maximizes the increase in expected utility and returns after exploring for the full amount of time.

Past work on robot exploration has not explicitly considered the possibility of failure [12], [13]. Instead, they assume robots are independent, whereas attrition couples decisions since we must consider the probability of any one robot communicating an observation. Thus, using expected utility as the objective for such algorithms is not possible.

C. Experiments and Results

We solved the ILP in Gurobi [27] on an Intel Xeon CPU at 3.40Ghz, and computed costs to go between points using a Probabilistic Road Map [28], [29]. We found frontiers using [22] and expanded them by constructing Voronoi cells around the frontiers. We find the optimization starting point by iteratively assigning each robot to the highest reward frontier with the least amount of robots assigned to it, and greedily assigning communication points between pairs of frontiers with robots. We post-process to explore for the full available time until the next scheduled rendezvous. For each recursive call, planning had a timeout of 10% of the exploration time.

We specify possible communication points as the midpoint on the path between any two frontiers or from a frontier to the base station. We used a cost and time to go as distance, attrition rate of 0.005—i.e., a 0.5% chance of failure per meter—and maximum reward as the number of unknown 0.1x0.1m cells in the frontier. We assume subteams have exploration rates of $\{4, 5.5, 6.5, 7\}\frac{m^2}{s}$ for $\{1,2,3,4\}$ robots.

	Upper Bound	Greedy Search	Our Work	Greedy Search	Our Work
10 robots	Initial Expected Utility			Simulated Information Gain $\pm$ Std. Dev.	
Four Way	980	**979 (99.9%)**	**979 (99.9%)**	$\mathbf{980 \pm 0}$	$\mathbf{980 \pm 0}$
Maze	4785	2165 (45.3%)	**2393 (50.0%)**	1940 ± 530	$\mathbf{2005 \pm 300}$
Cage	59968	56009 (93.4%)	**57192 (95.3%)**	$\mathbf{8144 \pm 7362}$	7326 ± 3005
Camp Lejeune	185109	72003 (38.9%)	**108008 (58.3%)**	10459 ± 9819	$\mathbf{21513 \pm 16392}$
20 robots	Initial Expected Utility			Simulated Information Gain $\pm$ Std. Dev.	
Four Way	980	**980 (100.0%)**	**980 (100.0%)**	$\mathbf{980 \pm 0}$	$\mathbf{980 \pm 0}$
Maze	4879	2914 (59.7%)	**3635 (74.5%)**	2887 ± 481	$\mathbf{3173 \pm 617}$
Cage	59969	59746 (99.6%)	**59959 (100.0%)**	18351 ± 8108	$\mathbf{19218 \pm 8795}$
Camp Lejeune	191705	115289 (60.1%)	**175470 (91.5%)**	35456 ± 13036	$\mathbf{45162 \pm 16011}$

TABLE I: Initial Expected Utility and Information Gain during Simulation. We compared our approach to a greedy search and list the % difference ($\frac{\texttt{ours}}{\texttt{bound}}$) from the theoretical upper bound, and standard deviation of information gain.

Table I shows the initial expected utility for the theoretical upper bound (31) and the two approaches as well as the information gain at the base station averaged over five simulated trials. Our model has higher expected utility than the greedy search in more complex environments and produces results within 50% of the theoretical upper bound. Additionally, our approach has both greater achieved information gain and less deviation from simulated exploration.

The deviation in information gain from our method is significantly lower the greedy search's, due to the subteams planning rendezvous points. The likelihood that at least one robot returns to the base station with some information increases when robots communicate. Thus, the information gain is less dependent on an individual robot successfully returning to the base station, decreasing the deviation in the simulated information gain and increasing the expected utility.

Greedy search may outperform our method when our estimate of information gain is very inaccurate (Fig. 4c) and with small number of robots. However, we do outperform the greedy search in environments with sufficient accuracy in estimated information gain (Fig. 4d), implying that our approach can account for some deviation in the estimated information gain.

VI. Conclusion

We presented an optimization model and recursive approach to find robot paths that maximize expected utility under communication and attrition. Our model extends the VRP, and we solve an ILP to approximate the optimal solution. Our results show that, for tested scenarios, our approach outperforms greedy search and finds plans within 50% of a theoretical upper bound on optimality. In future work, we will evaluate this approach on physical robots and further refine choices of rendezvous locations.

References

[1] K. Nagatani *et al.*, "Emergency response to the nuclear accident at the Fukushima Daiichi nuclear power plants using mobile rescue robots," *J. of Field Robotics*, 2013.

[2] P. J. Schoemaker, "The expected utility model: Its variants, purposes, evidence and limitations," *J. of economic literature*, 1982.

[3] J. D. Little, K. G. Murty, D. W. Sweeney, and C. Karel, "An algorithm for the traveling salesman problem," *Operations research*, 1963.

[4] R. T. Vaughan, K. Stoy, G. S. Sukhatme, and M. J. Mataric, "Lost: Localization-space trails for robot teams," *T-RO*, 2002.

[5] N. R. Hoff, A. Sagoff, R. J. Wood, and R. Nagpal, "Two foraging algorithms for robot swarms using only local communication," in *Intl. Conf. on Robotics and Biomimetics*. IEEE, 2010.

[6] D. A. Shell and M. J. Mataric, "On foraging strategies for large-scale multi-robot systems," in *IROS*. IEEE, 2006, pp. 2717–2723.

[7] R. Zlot, A. Stentz, M. B. Dias, and S. Thayer, "Multi-robot exploration controlled by a market economy," in *ICRA*. IEEE, 2002.

[8] M. Corah, C. O'Meadhra, K. Goel, and N. Michael, "Communication-efficient planning and mapping for multi-robot exploration in large environments," *RA-L*, vol. 4, no. 2, 2019.

[9] R. G. Colares and L. Chaimowicz, "The next frontier: combining information gain and distance cost for decentralized multi-robot exploration," in *ACM Symposium on Applied Computing*, 2016.

[10] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Decentralized active information acquisition: Theory and application to multi-robot slam," in *ICRA*. IEEE, 2015.

[11] Y. Kantaros, B. Schlotfeldt, N. Atanasov, and G. J. Pappas, "Asymptotically optimal planning for non-myopic multi-robot information gathering." in *RSS*, 2019.

[12] J. Faigl, M. Kulich, and L. Přeučil, "Goal assignment using distance cost in multi-robot exploration," in *IROS*. IEEE, 2012.

[13] G. Best, J. Faigl, and R. Fitch, "Online planning for multi-robot active perception with self-organising maps," *Autonomous Robots*, 2018.

[14] Z. Qiao, J. Zhang, X. Qu, and J. Xiong, "Dynamic self-organizing leader-follower control in a swarm mobile robots system under limited communication," *IEEE Access*, 2020.

[15] W. Sheng, Q. Yang, S. Ci, and N. Xi, "Multi-robot area exploration with limited-range communications," in *ICRA*. IEEE, 2004.

[16] N. Hazon and G. A. Kaminka, "Redundancy, efficiency and robustness in multi-robot coverage," in *ICRA*. IEEE, 2005, pp. 735–741.

[17] N. Agmon, N. Hazon, and G. A. Kaminka, "The giving tree: constructing trees for efficient offline and online multi-robot coverage," *Annals of Mathematics and Artificial Intelligence*, 2008.

[18] J. Song and S. Gupta, "Care: Cooperative autonomy for resilience and efficiency of robot teams for complete coverage of unknown environments under robot failures," *Autonomous Robots*, 2020.

[19] H. Kurniawati, D. Hsu, and W. S. Lee, "Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces." in *RSS*. Zurich, Switzerland., 2008.

[20] N. P. Garg, D. Hsu, and W. S. Lee, "Despot-α: Online pomdp planning with large state and observation spaces," in *RSS*, 2019.

[21] B. Charrow, G. Kahn, S. Patil, S. Liu, K. Goldberg, P. Abbeel, N. Michael, and V. Kumar, "Information-theoretic planning with trajectory optimization for dense 3D mapping."

[22] N. C. Fung, C. Nieto-Granda, J. M. Gregory, and J. G. Rogers, "Autonomous exploration using an information gain metric," *US Army Research Laboratory Adelphi United States*, 2016.

[23] A. Solanas and M. A. Garcia, "Coordinated multi-robot exploration through unsupervised clustering of unknown space," in *IROS*. IEEE, 2004.

[24] L. Wu, M. Á. García García, D. Puig Valls, and A. Solé Ribalta, "Voronoi-based space partitioning for coordinated multi-robot exploration," 2007.

[25] P. Toth and D. Vigo, *The vehicle routing problem*. SIAM, 2002.

[26] J. D. Camm, A. S. Raturi, and S. Tsubakitani, "Cutting big m down to size," *Interfaces*, vol. 20, no. 5, pp. 61–66, 1990.

[27] L. Gurobi Optimization, "Gurobi optimizer reference manual," 2021. [Online]. Available: http://www.gurobi.com

[28] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *Trans. on Robotics and Automation*, 1996.

[29] I. A. Şucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *RAM*, vol. 19, no. 4, 2012.