# Final Project Proposal: Validating the Model of a Fire-Detecting Poorly-Communicating Pair of Drones

**Daniel T. Simpson** datsimp@stanford.edu

*AA228V/CS238V, Stanford University*

## 1. Goal

Researchers have been developing fire-detecting drone swarms to address recent increases in extreme wildfire frequency [1; 2]. In these safety-critical scenarios, drones must quickly and reliably detect the location of a fire within a forest area. One group of students attempted to model this problem as a Final Project for Stanford University's course AA228: Decision Making Under Uncertainty [3]. The model involves two drones searching for a fixed fire cell within a square gridworld. The drones can observe only within a 3x3 observation window, move to adjacent cells, and have communication actions, which penalize the global reward function but allow each drone to gain the other's history of states. Thus, the goal of this paper is to validate the safety of this two-drone, costly-communications model.

If the model was safe for fire-detection, it must satisfy specifications including a near-zero failure rate, a low time to find the fire, and a low trajectory cost. If these specifications are met, then there may be an optimal starting position for the agents to minimize the probability of failure, minimize the time to finding the fire, and minimize the trajectory cost. This would reveal failure distributions over the configuration space, and some reachability analysis could be completed. This would reveal the safe set of initial configurations and identify which starting states inevitably lead to failure. Thus, the safest positions and trajectories for the two-drone costly-comms system would be found.

## 2. System

The model is a multi-agent network where agents collaboratively search for an objective under unreliable communications. The aim of this project was to aid in wildfire surveillance, based on Mahdi Al-Husseini's work [4]. The multi-agent model mimics a fire-search scenario with two drones in a simple environment (e.g. an open field). Each agent has six actions to choose from including: stay, move one space left, move one space right, move one space up, move one space down, or communicate. Each agent also can partially observe the world within a 3x3 observation window, centered at the agent's position. In the gridworld, there is a fire at a fixed location and the agents are not allowed to pass beyond the world borders. The results of the project showed that in a scenario where two surveying agents have costly communication between each other, they still successfully reach the objective, within a similar time as a scenario with free communications (i.e. no costly communications). Costly communication is the communication-failure mode between agents, wherein using a communication action adds a penalty to the reward. During a communication action, both drones engage in a "sync" method of communication, which allows each drone to know the full state of the other [5]. Thus, each communication action gives each agent maximum possible information gain from the other agent, while creating a penalty on the global reward.

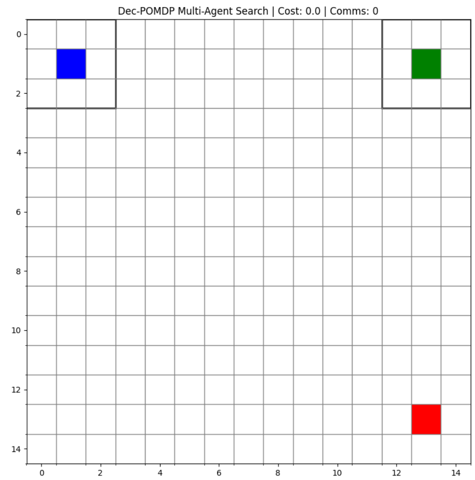


Figure 1: Possible setup of the multi-agent grid-world problem where blue and green squares are agents with an observation window bounded by a black square, and red squares represent the location of a fire.

The current model has no disturbances, except for an exploration bonus that stochastically allows the drone to explore more terrain rather than commit to a greedy best Q-value action. To perform safety validation, I will implement disturbances such as stochastic communication noise (distorting the belief map transfer) and 'wind' (positional offsets).

The system already has built-in simulation methods that render the positions of the drones relative to the fire and their observation windows. Each simulation runs $N_{trials}$ and reports the total time it takes for the drone(s) to reach the objective, the total number of communication actions taken by the drones, the total cost of the trajectory of the trial, and the failure rate, defined as the rate at which the drone(s) total run time reached a threshold time, $t_f$. The simulation has set configurations as shown in the table below:

## 3. Specification

To validate the safety and effectiveness of the multi-agent system, Signal Temporal Logic (STL) and probabilistic risk thresholds could be used to define model specifications.

The primary specification is Time-Bounded Reachability. The system is considered "successful" if and only if at least one agent's observation window overlaps with the fire location within the maximum simulation time $t_f$. If $s_t$ represents the state of the system at time $t$, and $\mathcal{S}$ represents the set of states where the fire is detected, the specification $\phi_{find}$ is:

$$\phi_{find} = \Diamond_{[0,t_f]}(s_t \in \mathcal{S}) \tag{1}$$

where $\Diamond$ denotes the temporal "eventually" operator. Given the stochastic nature of the initialization and exploration bonuses, we cannot guarantee this property holds for every

Table 1: Simulation Parameters

| **Parameter** | **Value** |
|---|---|
| Grid Size | 15 |
| Observation Window Size | 3 |
| Initial Time ($t_i$) | 0.0 s |
| Time Step ($\Delta t$) | 0.05 s |
| Max Simulation Time ($t_f$) | 20.0 s |
| Gamma $\gamma$ | 0.95 |
| Communication Threshold | 0.2 |
| Exploration Bonus | 2.0 |
| Movement Cost | 2.0 |
| Time Cost ($\kappa$) | 1.0 |
| Communication Noise | 0.05 |
| Uncertainty Growth Rate | 0.1 |
| Minimum Communication Interval | 5 |

trajectory. Therefore, we define a probabilistic specification:

$$P(s_{0:t_f} \models \phi_{find}) \geq 1 - \epsilon \tag{2}$$

where $\epsilon$ is the maximum allowable probability of failure (e.g., $\epsilon = 0.05$).

A secondary specification is a Resource Safety Constraint to validate the cost of the "bad comms" model. The system fails if the accumulated cost $J(t)$ (a sum of movement, time, and communication penalties) exceeds a critical budget $J_{max}$ (e.g. a battery or limited bandwidth). This is defined as the global property $\phi_{cost}$:

$$\phi_{cost} = \Box_{[0,t_f]}(J(t) < J_{max}) \tag{3}$$

where $\Box$ denotes the "always" operator.

The goal of this project is to estimate the failure probabilities for $\phi_{find}$ and $\phi_{cost}$ using falsification techniques to identify initial configurations or disturbance sequences that induce violations.

## References

[1] C. X. Cunningham, G. J. Williamson, and D. M. J. S. Bowman, "Increasing frequency and intensity of the most extreme wildfires on earth," *Nature Ecology & Evolution*, vol. 8, no. 8, pp. 1420–1425, aug 2024. [Online]. Available: https://doi.org/10.1038/s41559-024-02452-2

[2] Mongabay. (2025, jul) Drone swarms and AI take aim at stopping wildfires in 10 minutes. Mongabay. Accessed: 2026-01-20. [Online]. Available: https://news.mongabay.com/short-article/2025/07/drone-swarms-and-ai-take-aim-at-stopping-wildfires-in-10-minutes/

[3] M. J. Kochenderfer, T. A. Wheeler, and K. H. Wray, *Algorithms for Decision Making.* MIT Press, 2022.

[4] M. Al-Husseini, K. H. Wray, and M. J. Kochenderfer, "Hierarchical framework for optimizing wildfire surveillance and suppression using human-autonomous teaming," *Journal of Aerospace Information Systems*, vol. 21, no. 10, pp. 790–811, 2024. [Online]. Available: https://doi.org/10.2514/1.I011368

[5] P. Xuan, V. Lesser, and S. Zilberstein, "Communication in multi-agent markov decision processes," *UMass Computer Science Technical Report 2000-01*, 2000.