# The Efficacy of Dense Local Incentives vs. Sparse Global Penalties in Multi-Agent Wildfire Suppression

Abdul Rahman Hussain Siddique (ah244)
Anirudh Ramesh (aramesh8)
Yashwanth Pulimi (ypulimi)

November 22, 2025

**Abstract**

In Multi-Agent Reinforcement Learning (MARL) for disaster response, it is often assumed that penalizing agents for the global state of the disaster (e.g., total fire size) is necessary to induce cooperative behavior. We investigated this by training a squad of 3 helicopters in the FirecastRL environment using two distinct reward structures: *Cooperative* (Global Penalty) and *Greedy* (Local Reward only). Contrary to our initial hypothesis, the Greedy agents significantly outperformed the Cooperative agents, reducing the average burnt area by $\approx 39\%$ and extinguishing the fire $\approx 40\%$ faster. This result suggests that in high-dimensional spatial environments, dense local feedback is a more effective learning signal than sparse global objectives.

## 1 Introduction

Wildfire suppression is a race against time. While single-agent RL has shown promise, it struggles to cover large geographical areas. This project explores scaling to multi-agent systems. The core challenge in MARL is reward shaping: how do we encourage agents to work together? We tested the hypothesis that agents sharing a global penalty for fire spread would learn superior containment strategies compared to agents acting purely on self-interest (maximizing individual drops).

## 2 Methodology

### 2.1 Environment and Agent Setup

We modified the `FirecastRL` environment to support $N = 3$ agents controlled by a centralized PPO policy. The state space includes a $160 \times 240$ grid of fire ignition times and the coordinates of all agents. The action space is a `MultiDiscrete` vector allowing simultaneous control of all helicopters.

### 2.2 Reward Functions

We trained two distinct models for 100,000 timesteps each:

- **Cooperative Agent:** Received +10 for extinguishing a cell, -2 for wasting water, and a continuous penalty of $-0.1 \times$ (Total Burning Cells). This penalty was intended to create urgency regarding the fire's spread.

- **Greedy Agent:** Received +10 for extinguishing and -2 for wasting water. Crucially, this agent received **zero penalty** for the fire spreading. It only cared about its own successful hits.

# 3 Results

We evaluated both models over 10 deterministic episodes. The results refuted our initial hypothesis.

| **Strategy** | **Avg Cells Burnt** (↓) | **Avg Steps to Finish** (↓) | **Reliability (Std Dev)** |
|---|---|---|---|
| Cooperative | 1600.90 | 728.1 | 61.04 |
| **Greedy** | **973.60** | **432.1** | 510.92 |

Table 1: Comparative performance statistics. The Greedy agent saved significantly more forest.

## 3.1 Analysis of Damage (Final Impact)

As shown in Figure 1, the Cooperative agent consistently converged to a sub-optimal solution (high damage), whereas the Greedy agent achieved a much lower median damage.

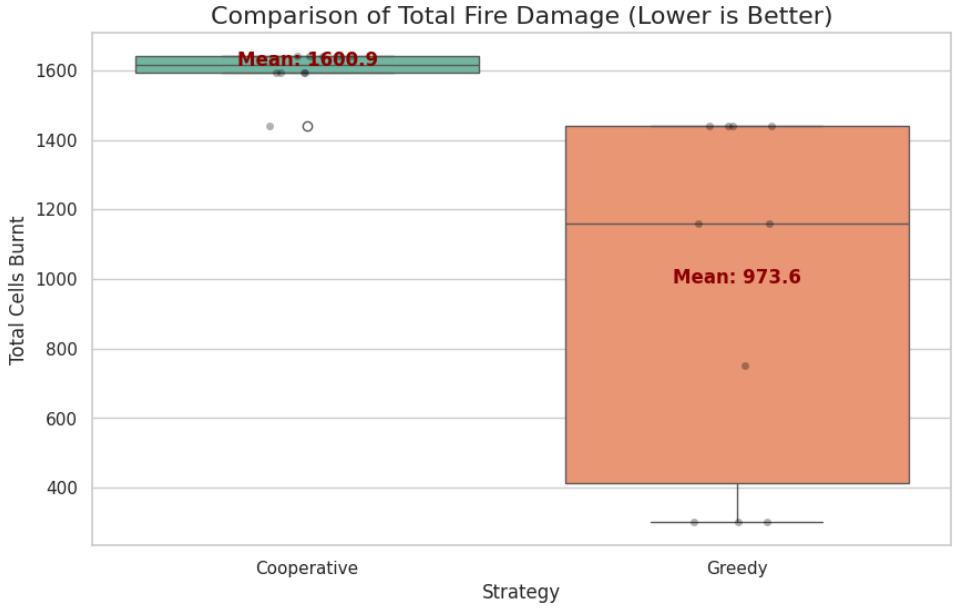


Figure 1: **Total Fire Damage.** The Box Plot illustrates that the Greedy strategy (Orange) consistently resulted in fewer burnt cells compared to the Cooperative strategy (Green/Blue).

## 3.2 Suppression Dynamics (Speed)

To understand *how* the Greedy agents achieved lower damage, we analyzed the fire size over time. Figure 2 shows that Greedy agents aggressively reduce the active fire count early in the episode, whereas Cooperative agents allow the fire to linger, leading to runaway spread.

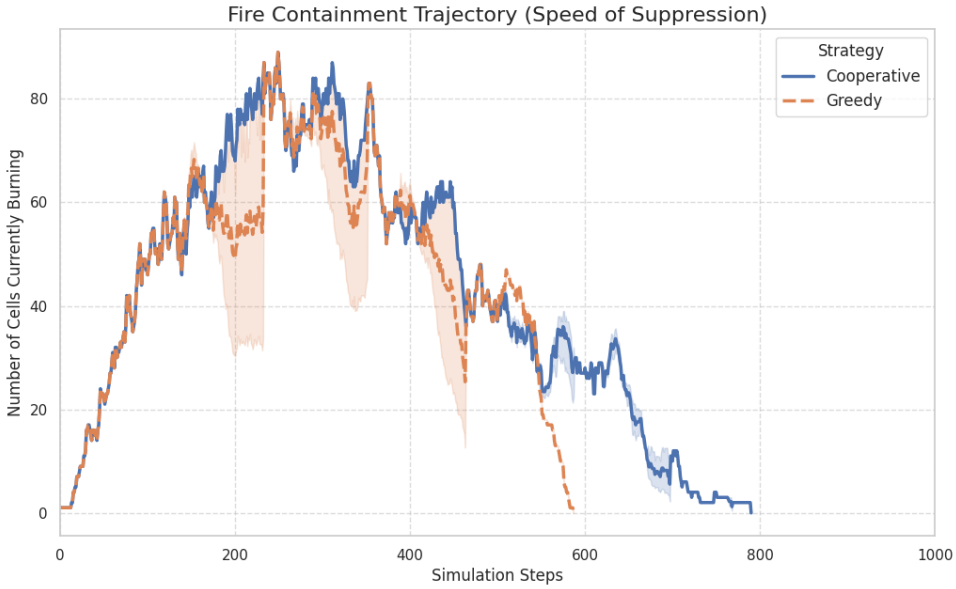


Figure 2: **Containment Trajectory.** The Greedy agent (Orange Dashed) creates a steeper drop in active fire cells, extinguishing the fire significantly faster than the Cooperative agent.

### 3.3 Efficiency Analysis

We examined the trade-off between episode duration and total damage. Figure 3 reveals a clear cluster of Greedy episodes in the bottom-left corner (Short Duration, Low Damage), indicating that "rushing" to score points actually resulted in better conservation of the environment.

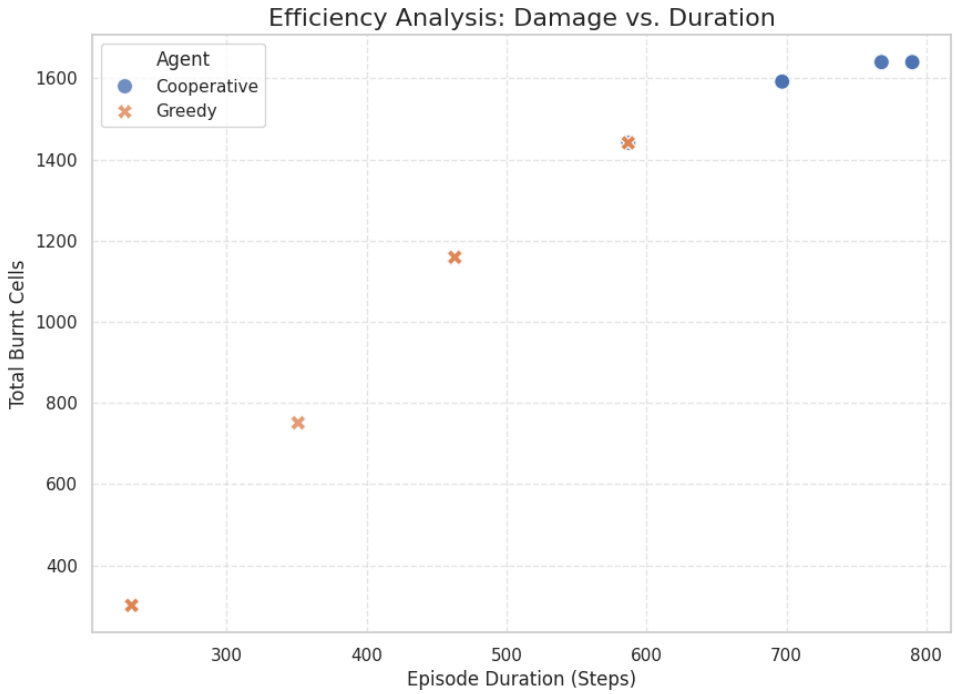


Figure 3: **Efficiency Scatter Plot.** There is a strong correlation between speed and quality. The Greedy agents do not sacrifice accuracy for speed; they achieve both.

### 3.4 Risk Profile

Finally, we analyzed the probability distribution of outcomes. Figure 4 shows that while the Cooperative agent is highly consistent (narrow peak), it is consistently *bad*. The Greedy agent has a wider variance but its probability mass is shifted significantly towards lower damage outcomes.

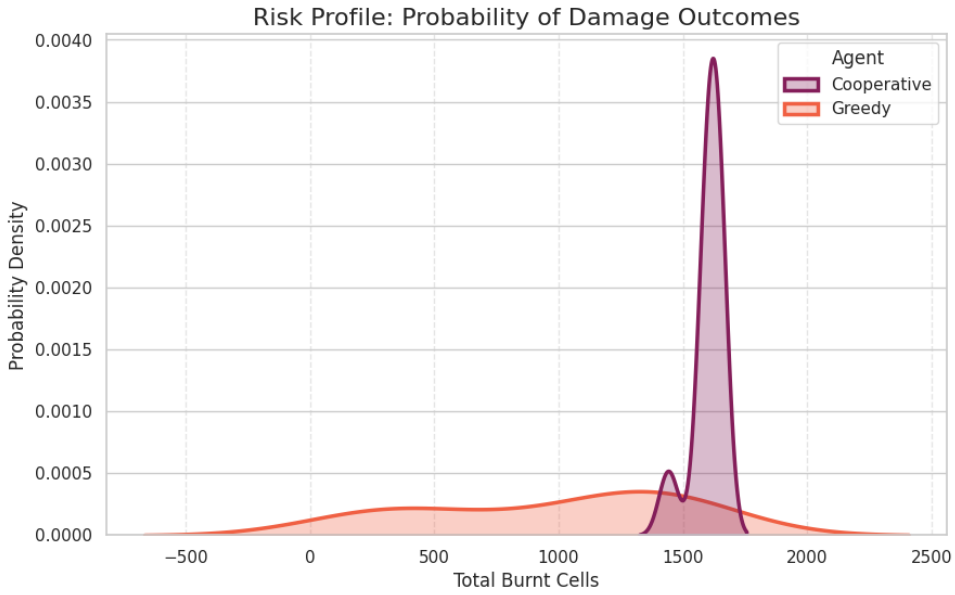


Figure 4: **Risk Density Profile.** The Greedy distribution is left-shifted, indicating a higher probability of saving more trees, despite higher variance.

## 4 Discussion: Why did Cooperation Fail?

The failure of the Cooperative reward can be attributed to the **Signal-to-Noise Ratio**.

1. **Noisy Global Signal:** The global penalty ($-0.1 \times$ Fire) fluctuates based on wind and physics, not just agent actions. This created "noise" that made it difficult for the PPO algorithm to credit specific actions.

2. **Learned Helplessness:** If the fire grew large early in the episode, the Cooperative agent accumulated massive negative rewards regardless of its actions, potentially leading to a local optimum where it ceased aggressive behavior.

3. **Direct Feedback:** The Greedy agent received a clean, dense signal: "I went to the red pixel, I dropped water, I got points." This direct cause-and-effect loop was easier to learn within the limited training window.

## 5 Conclusion

This study demonstrates that in MARL applications for wildfire suppression, **dense local incentives** are often superior to **sparse global penalties**. While intuitive to design rewards around the "ultimate goal" (saving the forest), the difficulty of credit assignment can hinder learning. Future work should explore *Curriculum Learning*, starting with greedy rewards and gradually introducing global penalties as the agents become proficient.