Exploration of Multi-Agent Reinforcement Learning for Flight Path Planning

Student: Lynphone Mark Xie

Student Email: Xie.34@wright.edu

Faculty: Faathi Amsaad

Faculty Email: Faculty email

AFRL Sponsor: Huaining Cheng

AFRL Directorate: AFRL

PA #: RH-23-9

Exploration of Multi-Agent Reinforcement Learning for Flight Path Planning

Mark Xie[2], Emily Conway[1], Huaining Cheng[1] , Fathi Amsaad[2]

[1]Air Force Research Laboratory, [2]Wright State University

Problem Statement

The current airborne ISR collection planning is optimized through an assembly line-like manual process, which can be cumbersome and time-consuming, hence lacking planning efficiency and operation optimality and agility.

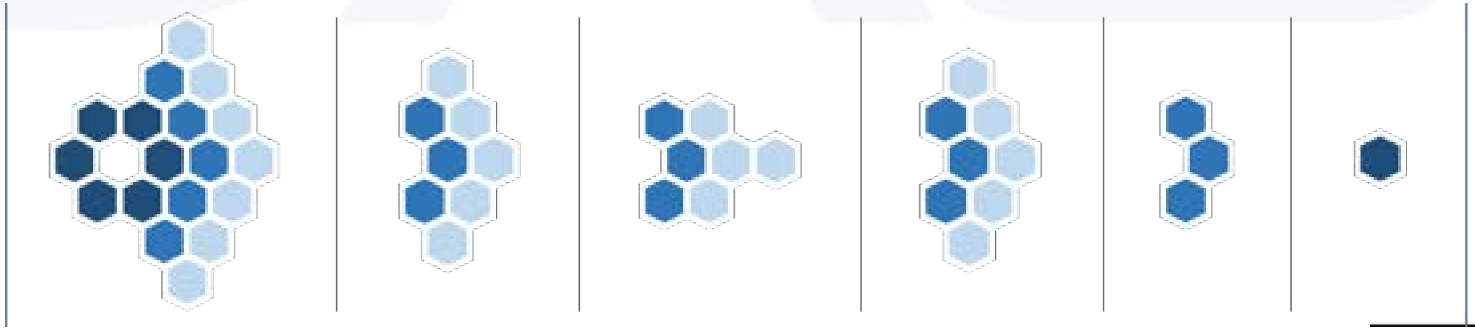
These problems could be amplified significantly in potential future conflicts with a near-peer adversary because the complexity, scale, and intensity of an airborne ISR operation could be several orders higher.

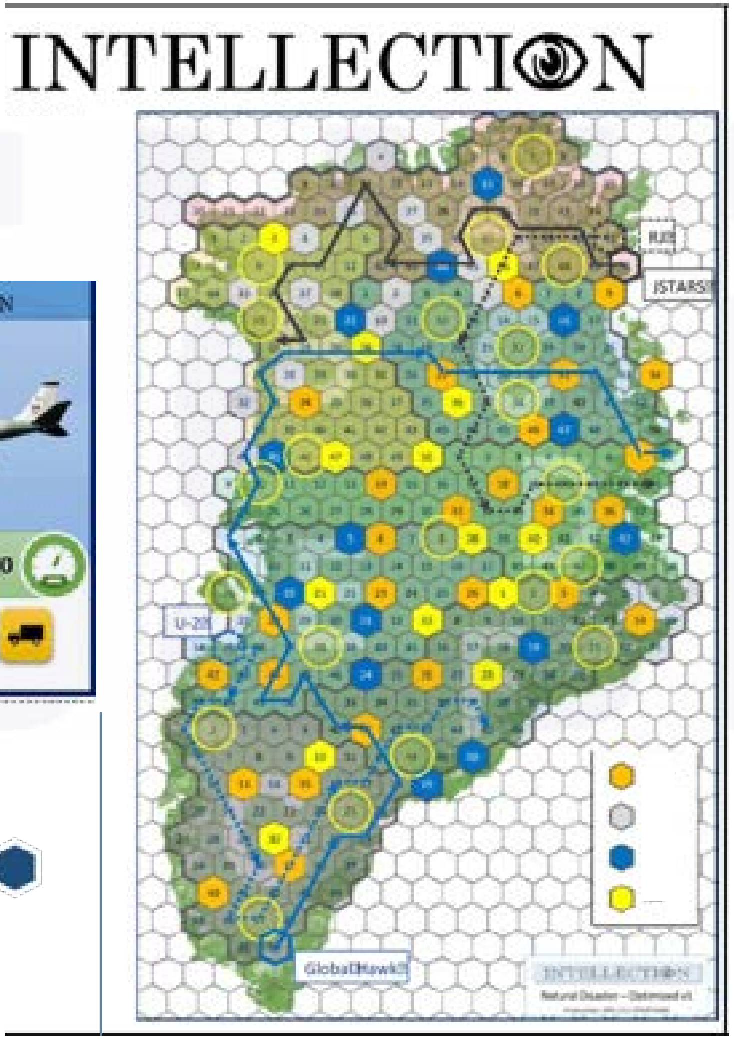
We propose using Reinforcement Learning to assist in automating and expediting ISR collection planning.

Intellection

The board game designed to train new personnel.

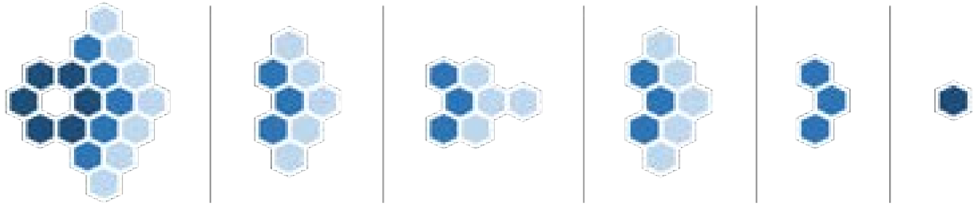
Personal Objective: Collect the most targets.
Team Objective: Collect 80% of all targets

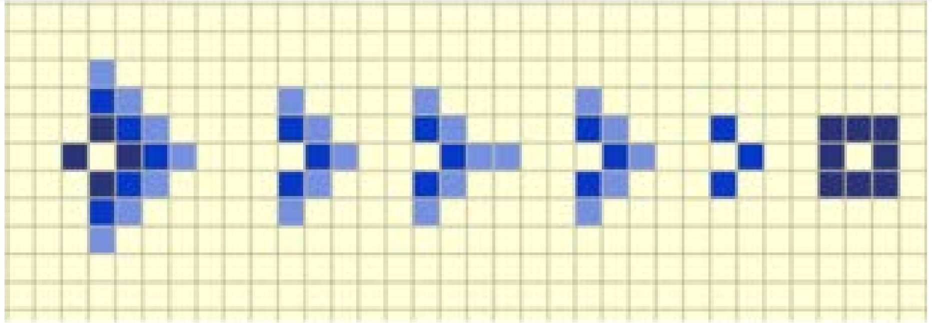




Environment

- Changed to Grid system
- Randomly allocated target location
- Random entry points
- Adjustments for fuel capacity and sensor footprints



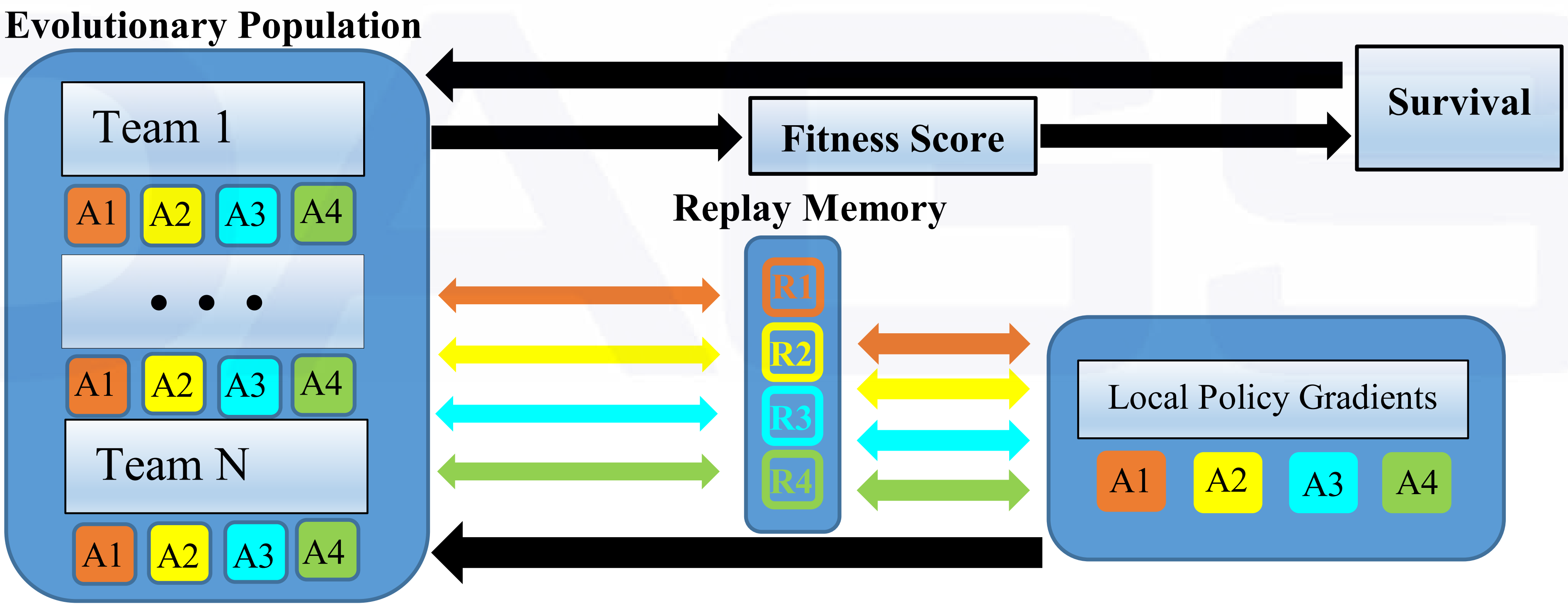


Reinforcement Learning Methods and Evolutionary Algorithm

- DDPG
  - Deep Deterministic Policy Gradient
  - Bellman Equation: $V(s)=\max_a(R(s,a)+\gamma V(s'))$
- TD3
  - Twin Delayed Deep Deterministic Policy Gradient is a version of DDPG that was created to address estimation errors. It uses two Q-functions and selects smaller Q-value as the target, reducing overestimation bias.
- Evolutionary Algorithm
  - The evolutionary algorithm is inspired by evolutionary theory and is used to solve NP-hard problems.

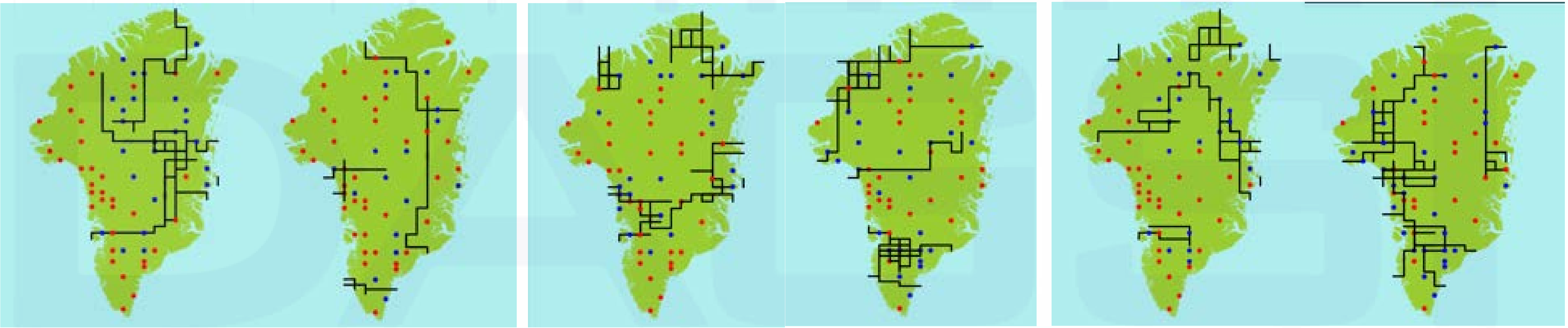
Multiagent Evolutionary Reinforcement Learning (MERL)

- Multi-Agent Evolutionary Reinforcement Learning (MERL)
  - A hybrid algorithm from Intel Labs using gradient-based optimizers to maximize individual agent rewards and a gradient-free optimizer to maximize the team reward through neuro evolution



Training

- In the case of training 5 Agents
- Randomized entry points and exit points



Starting Iteration

Final Iteration

Starting Iteration

Final Iteration

Starting Iteration

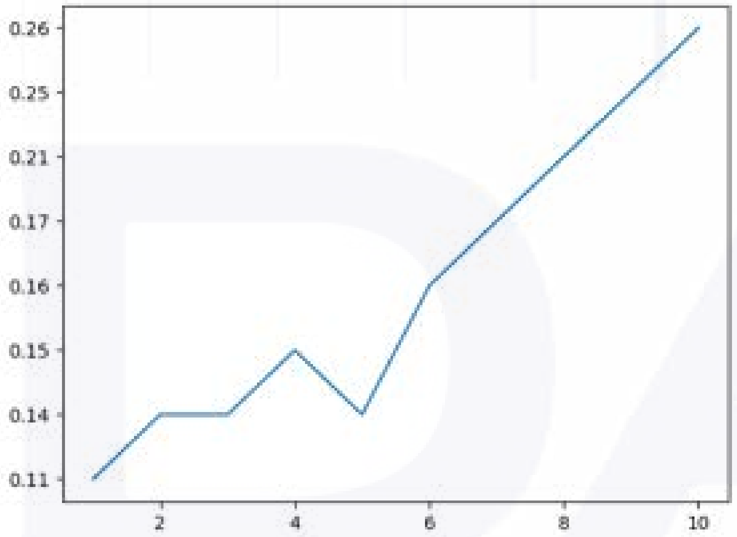
Final Iteration

Generation 0 Team Test

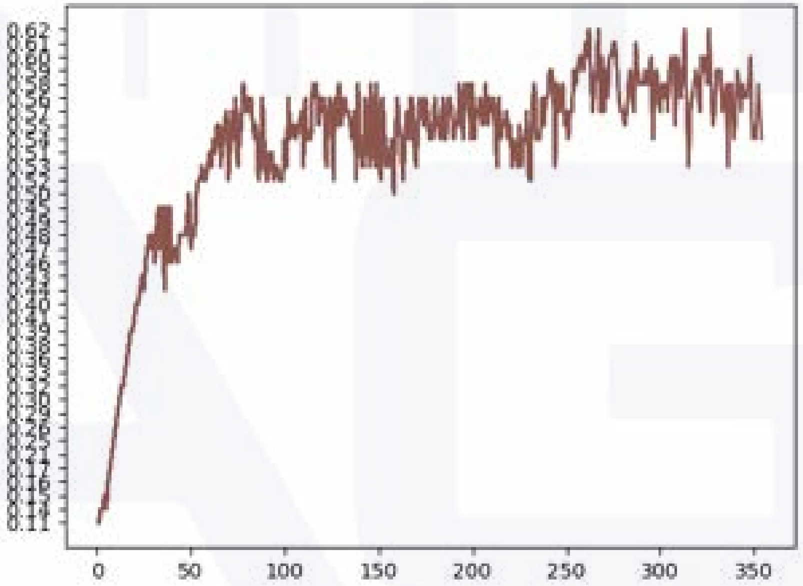
Generation 180 Team Test

Generation 354 Team Test

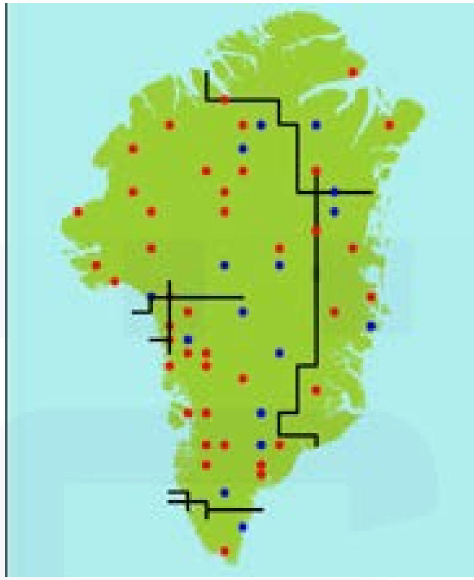
Results



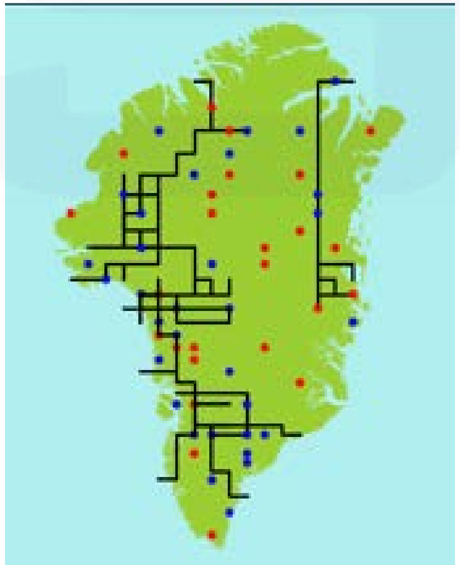
Generation 0-10 Learning Rate



Generation 0-354 Learning Rate



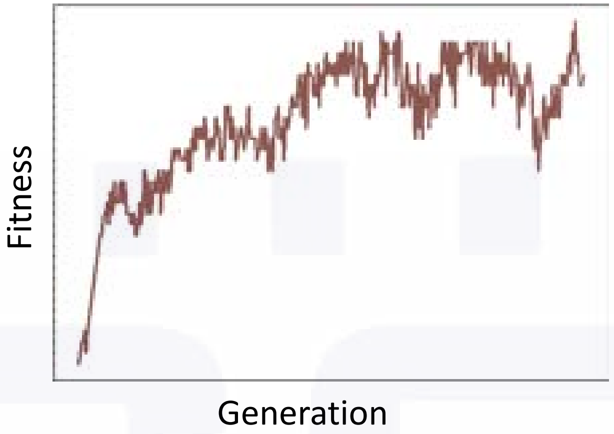
Testing Starting Generation

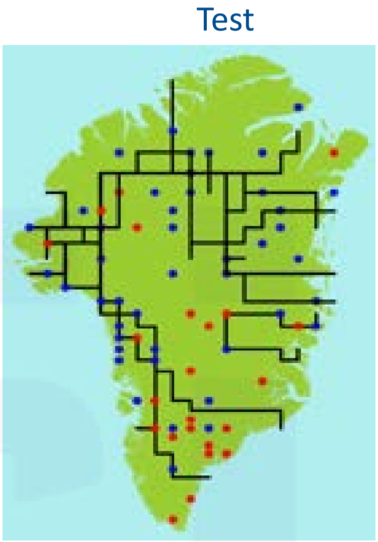


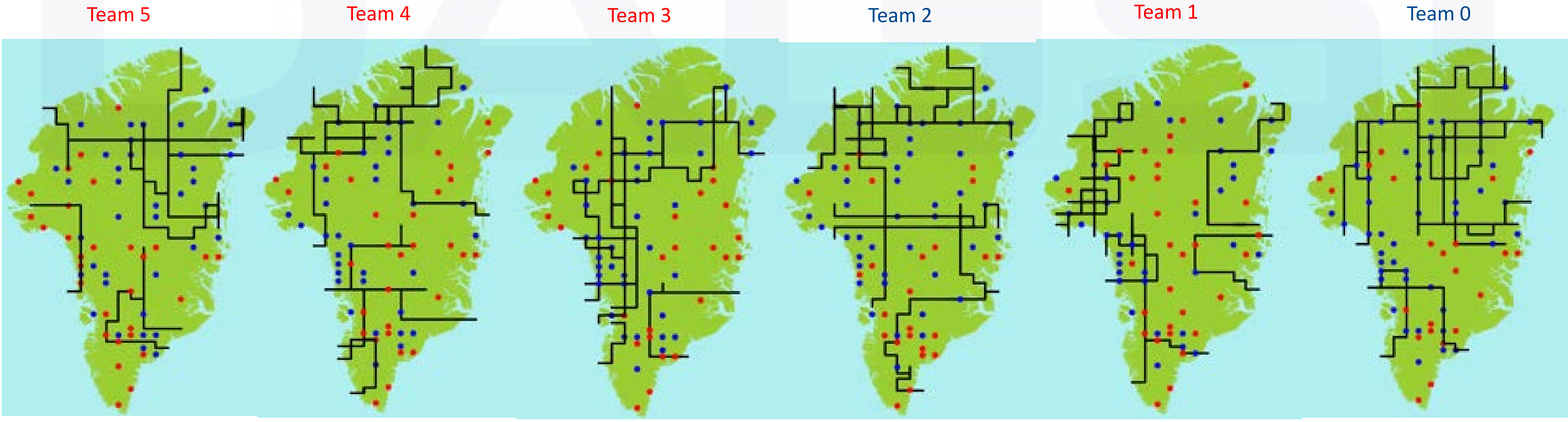
Testing Final Generation

Training Additional Agents

This is seven agent. The training progresses rapidly initially, then slowly increased as the margin between the team fitness scores grew closer and closer together.

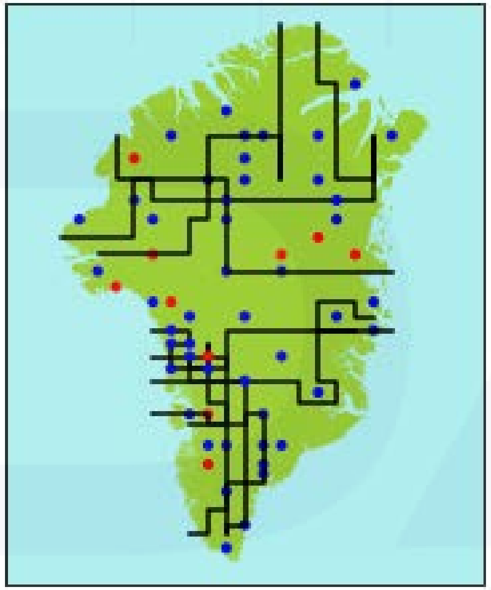






Results

Greenland

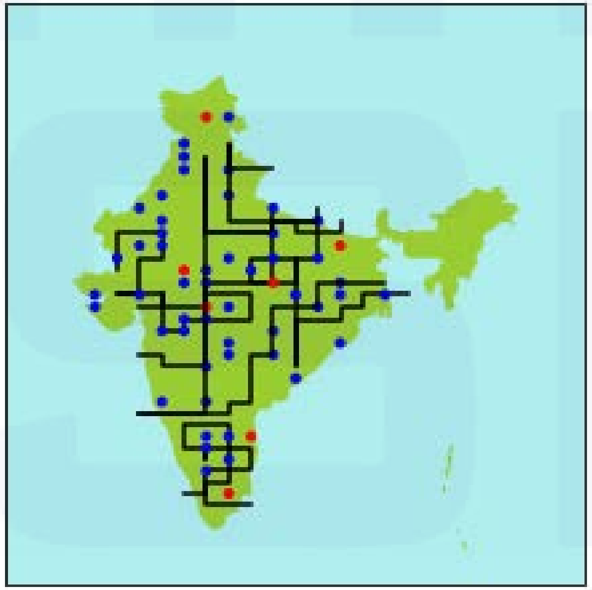


82% of all targets collected

Given roughly the same grid space, the model exceeds the collection threshold of 80% of all targets.

It successfully earned to explore the entirety of the map and learned to explore sections several times with different aircrafts

India



87% of all targets collected

Note: Unlike humans playing the intellection board game, the model frequently took advantage of moving backwards.

Summary

- Shows potential usage for ISR usage
- We found that there is a need for a benchmark to fully assess this method
- We found an improvement will be to the turn based grid space to continuous space

Future Work

- Human Machine Teaming
  - Human examples in Training
  - Human + Machine Team training
  - Human Feedback
- More Complex Parameters
  - Emerging Targets
  - Unexpected Aircraft Malfunction
  - Timed Fly/No-Fly Zones
  - Mandatory entry and/or exit locations
- Continuous Space
  - Realistic flight path simulation
  - Time based calculation