Reinforcement Learning for Joint Optimization of Sensing, Communications, and Control

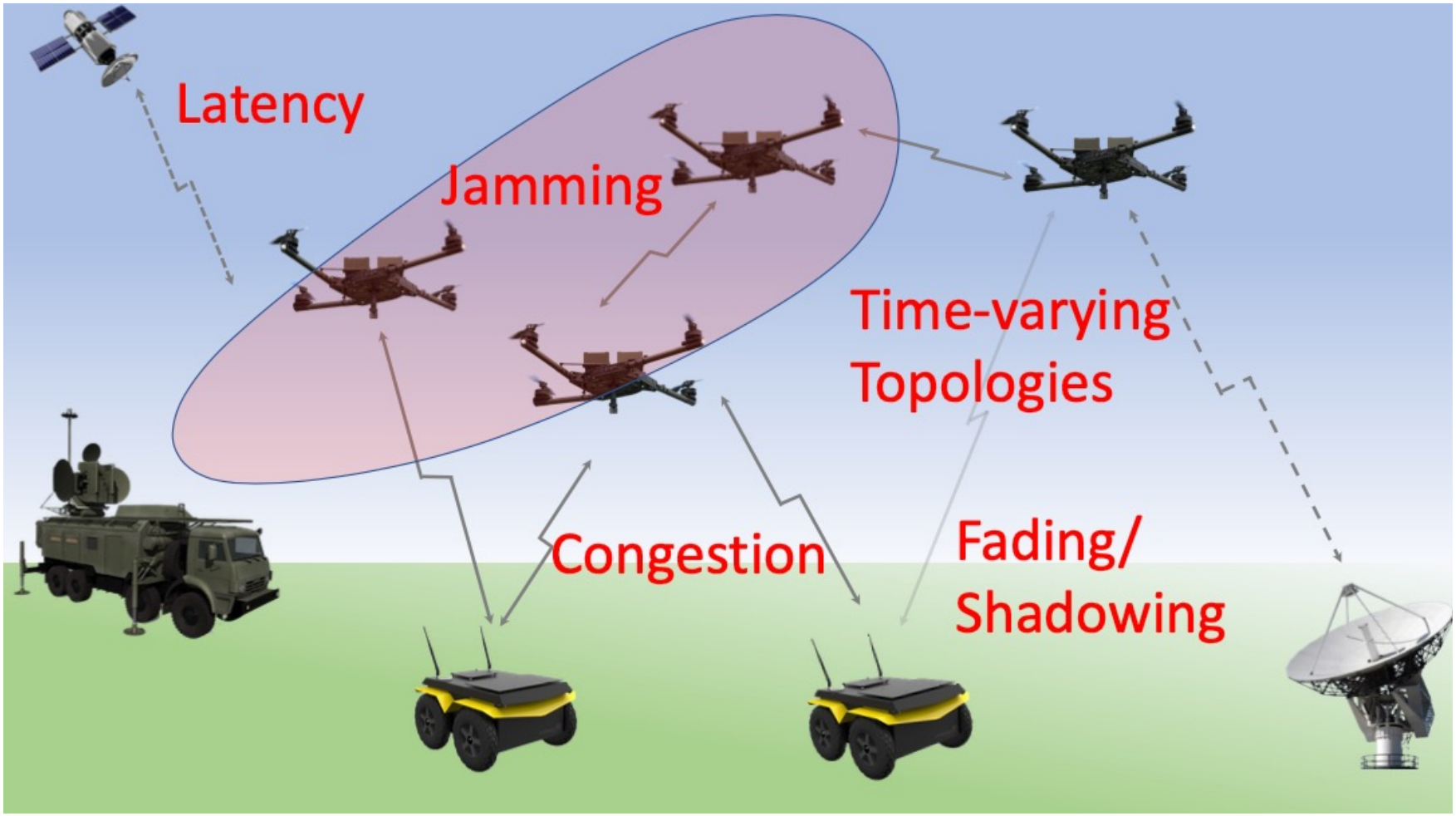
John M. Shea

UF UNIVERSITY of FLORIDA
Duke UNIVERSITY
TEXAS The University of Texas at Austin
UC SANTA CRUZ

Overview

- Focusing on problems that required joint optimization across:
  - Sensing
  - Communications and Networks
  - Control and Navigation

Big Picture


Latency
Jamming
Time-varying
Topologies
Congestion
Fading/
Shadowing

Reinforcement Learning (RL)

- Problems are inherently:
  - Stochastic
    - Noise, interference, jamming
    - Mobility
  - Partially observable
    - State is only known through noisy and/or delayed measurements
    - GPS may not be available for localization
  - Multi-objective
    - Network used by diverse control, sensing, and communication systems
  - Distributed
    - No centralized base stations for communications
    - Sensors and control agents may have to operate with limited coordination

Example: Optimal Jamming

- Early work: jamming to disrupt static networks

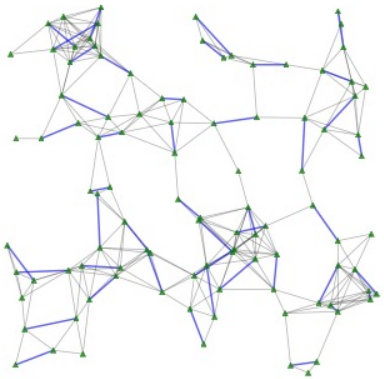
- **Objective:** Place mobile jammers to partition a wireless network "as much as possible" so its nodes can no longer cooperate
  - → Number of clusters should be bounded below
  - Want to avoid leaving large connected subgraph in which most nodes can still cooperate
  - → Cluster size should be bounded above

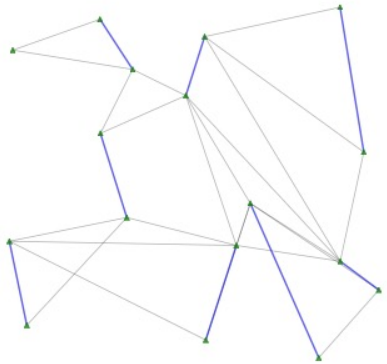
- Formulate as optimization problem on graph:

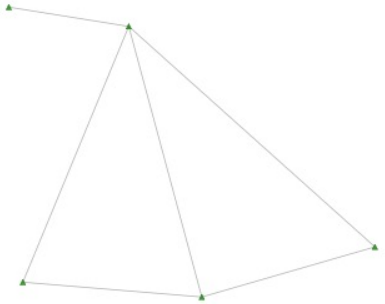
- Problem formulation — specify:
  - Minimum number of clusters
  - Maximum cluster size

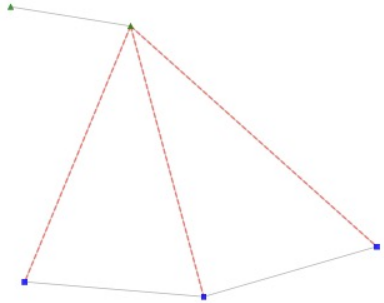
- Solutions:
  - Linear program (NP-Hard) if jammers limited to locations of network devices
  - Optimal search across real plane by developing set of “candidate jammer locations” – similar complexity
  - Fast solutions: spectral clustering, multilevel graph-cuts

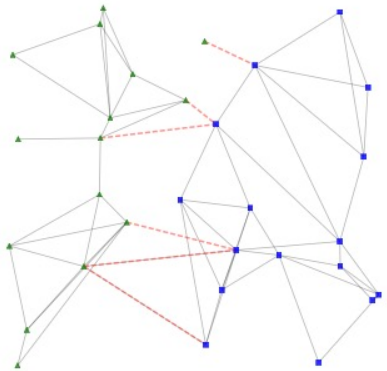
Multiresolution Graph Cuts

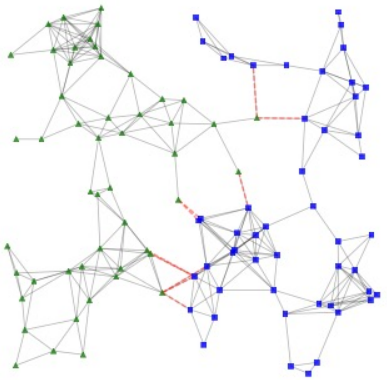










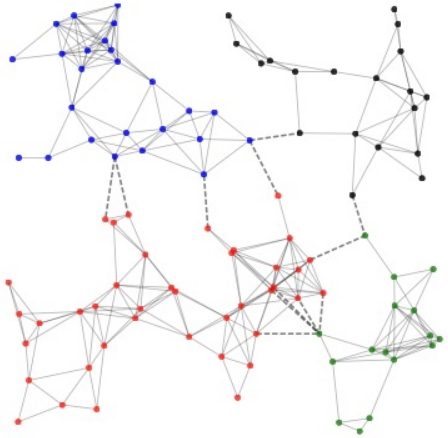


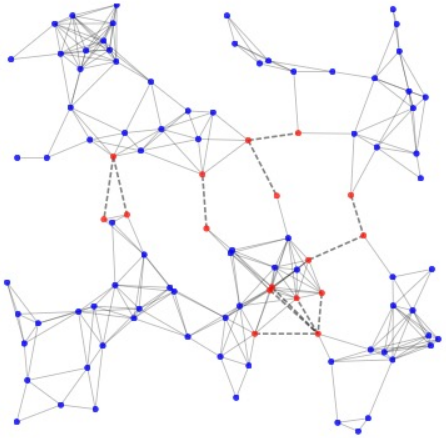
UF UNIVERSITY of FLORIDA

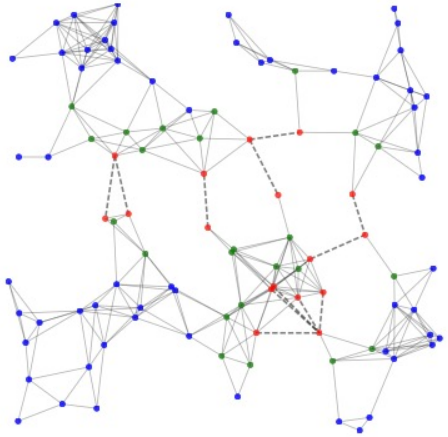
Duke UNIVERSITY

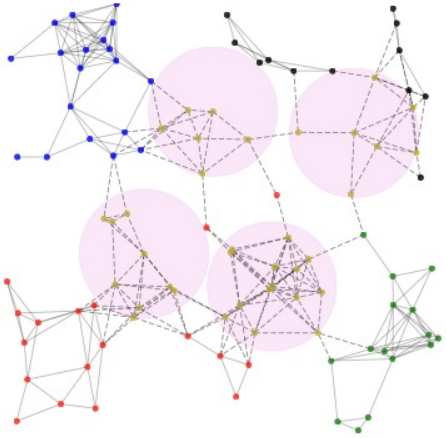
TEXAS The University of Texas at Austin

UC SANTA CRUZ









UF UNIVERSITY of FLORIDA

Duke UNIVERSITY

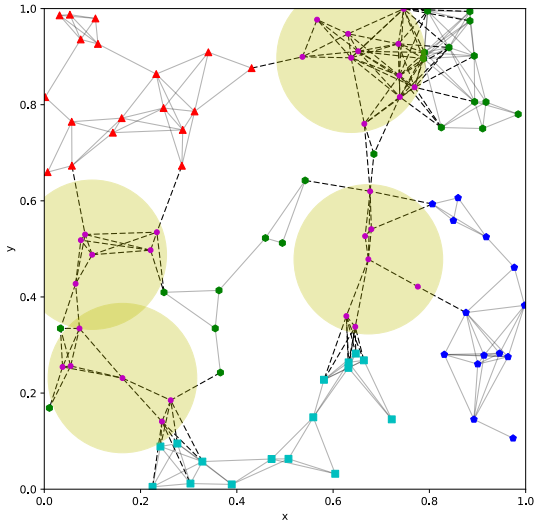
TEXAS The University of Texas at Austin

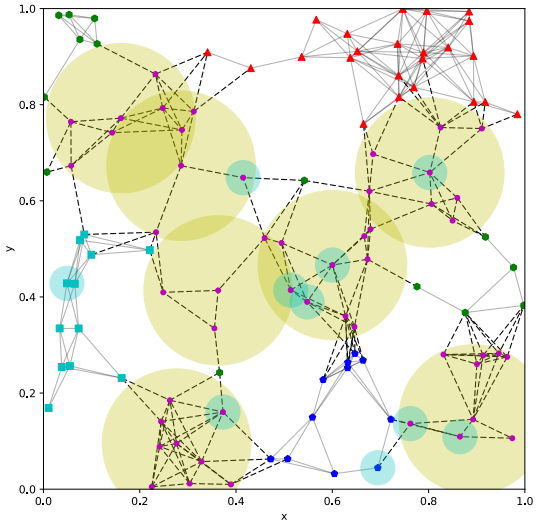
MUNITIONS DIRECTORATE

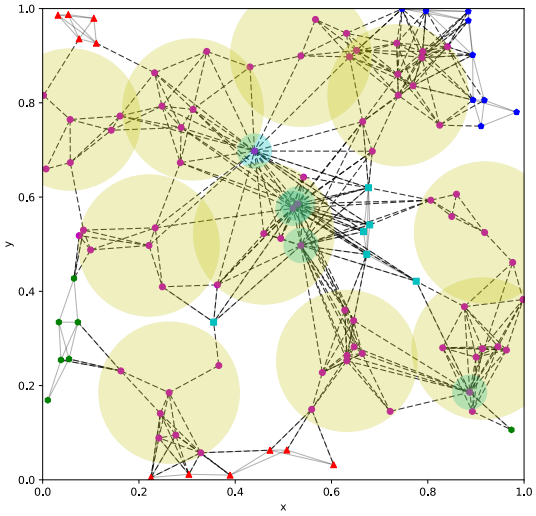
UC SANTA CRUZ

Reinforcing Networks Against Jammer Attacks

- Use mobile routers to reduce vulnerability of network to partitioning
  - Place routers to maximize number of jammers needed to partition network
- Find router placement via heuristic search







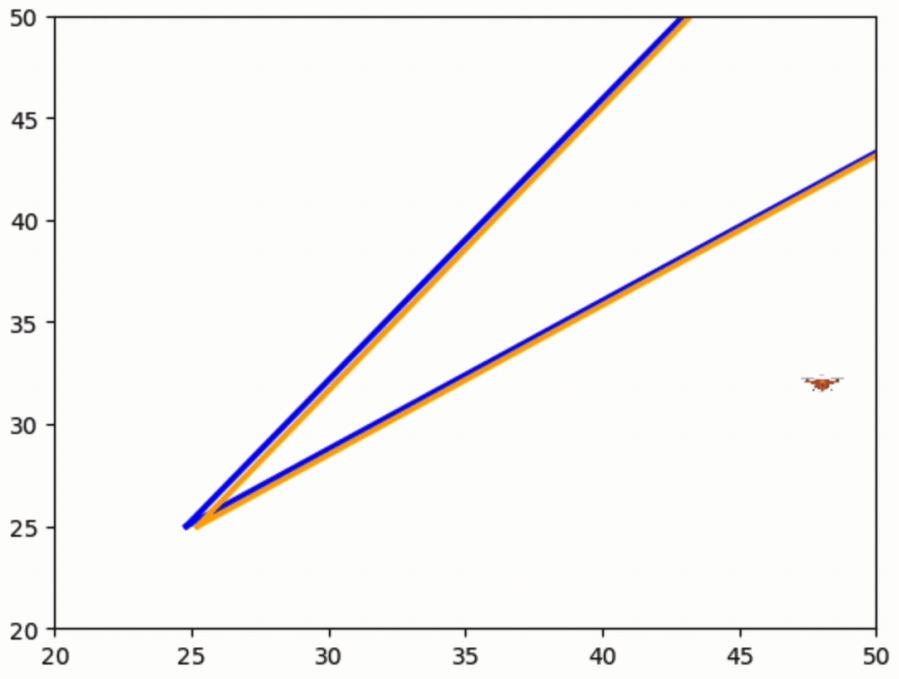
UF UNIVERSITY of FLORIDA · Duke UNIVERSITY · TEXAS The University of Texas at Austin · UC SANTA CRUZ

Jamming for Mobile Networks

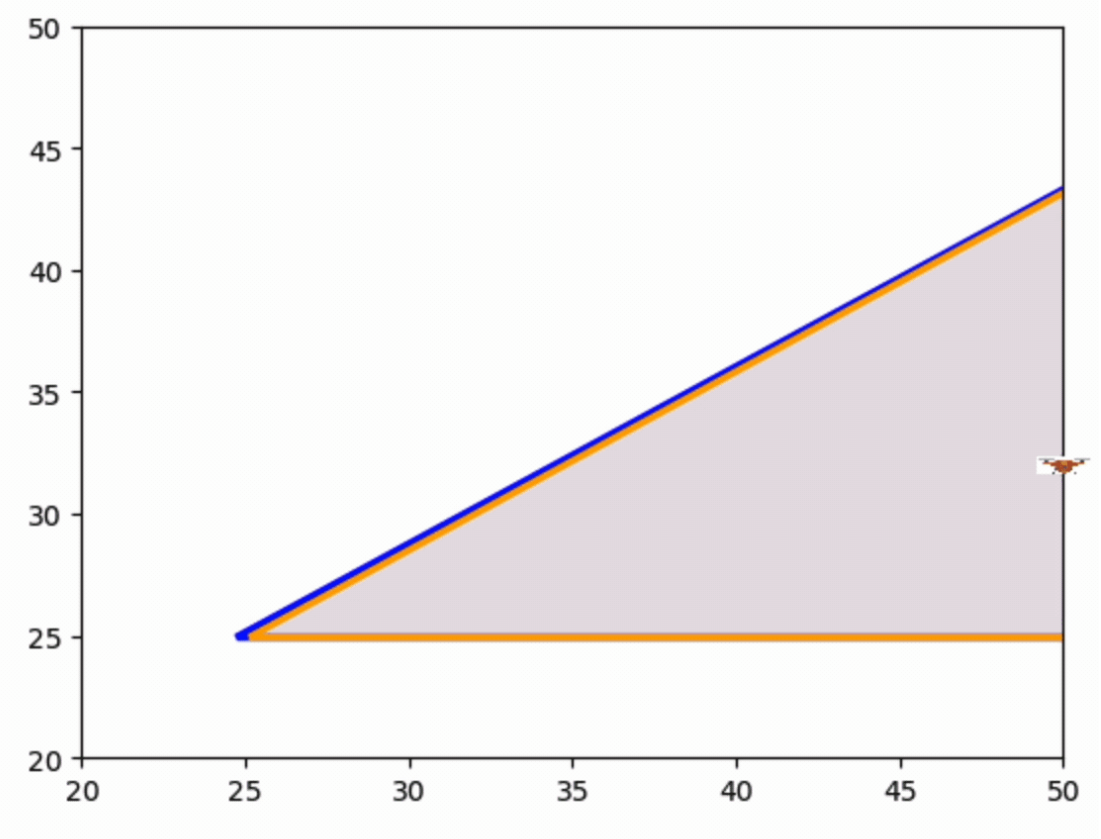
- Adapting jamming to disrupt mobile networks is challenging
  - Small changes to positions of mobile nodes can have large impact on network connectivity

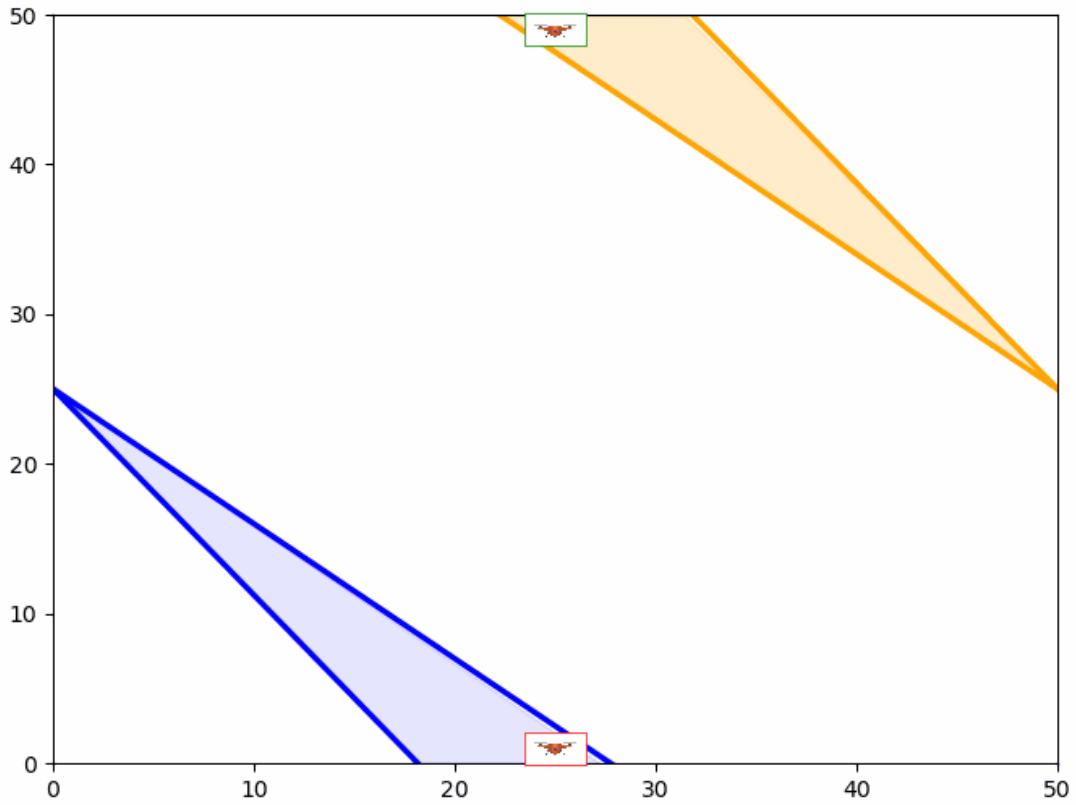
- Focus on simpler scenario: Protect a region from surveillance by drones
- Jammers use steerable antennas to disrupt drone communication
- Jammers may have limited ability to coordinate – consider distributed decision making

- With distributed decision making, need way to allow jammers to make separate decision
- **Example:** 2 jammers protecting against 1 drone with fixed beam directions
  - If both jammers point at same location, drone may move out of predicted beam and be able to establish communications



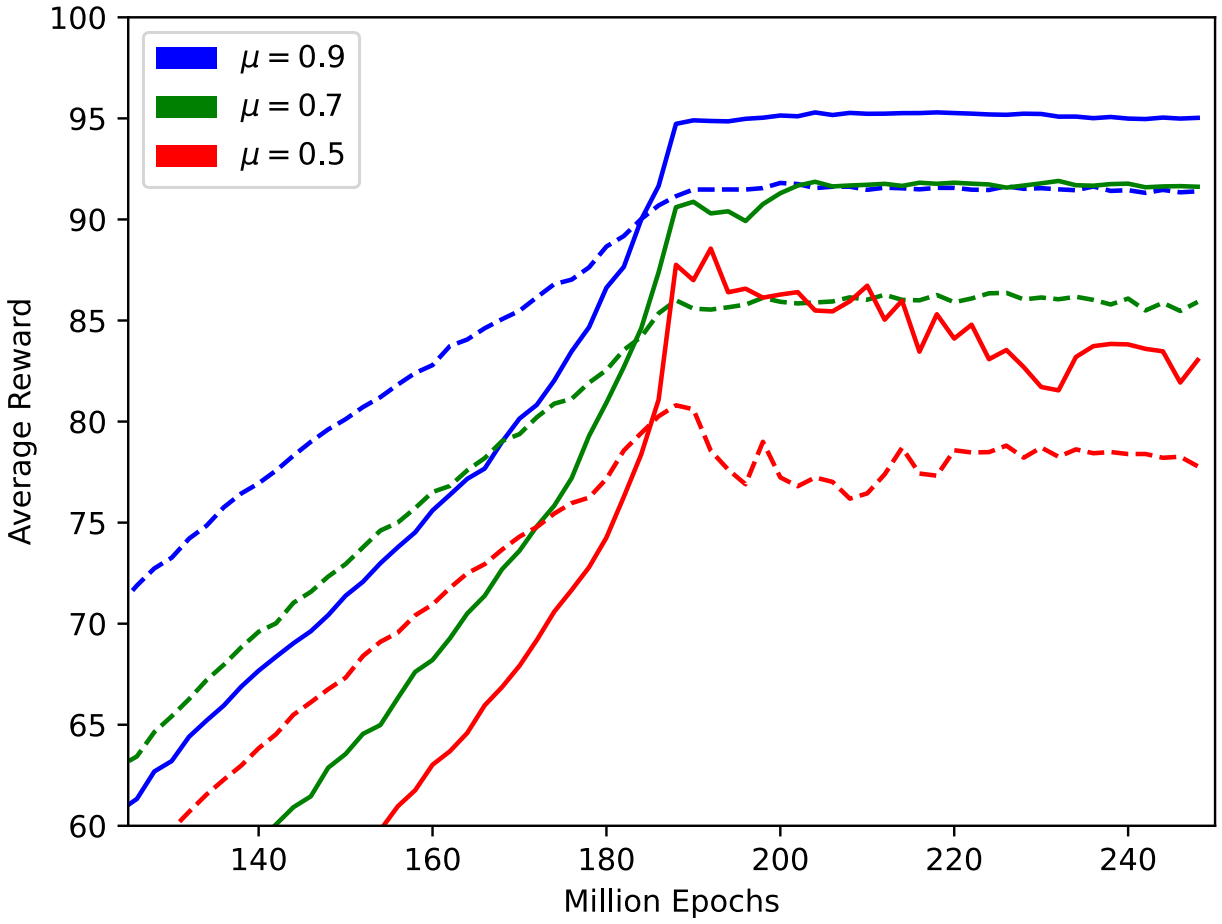
- Our approach: use RL to learn stochastic policies
  - In each state, learn PMF over actions
  - Action taken is randomly chosen according to PMF





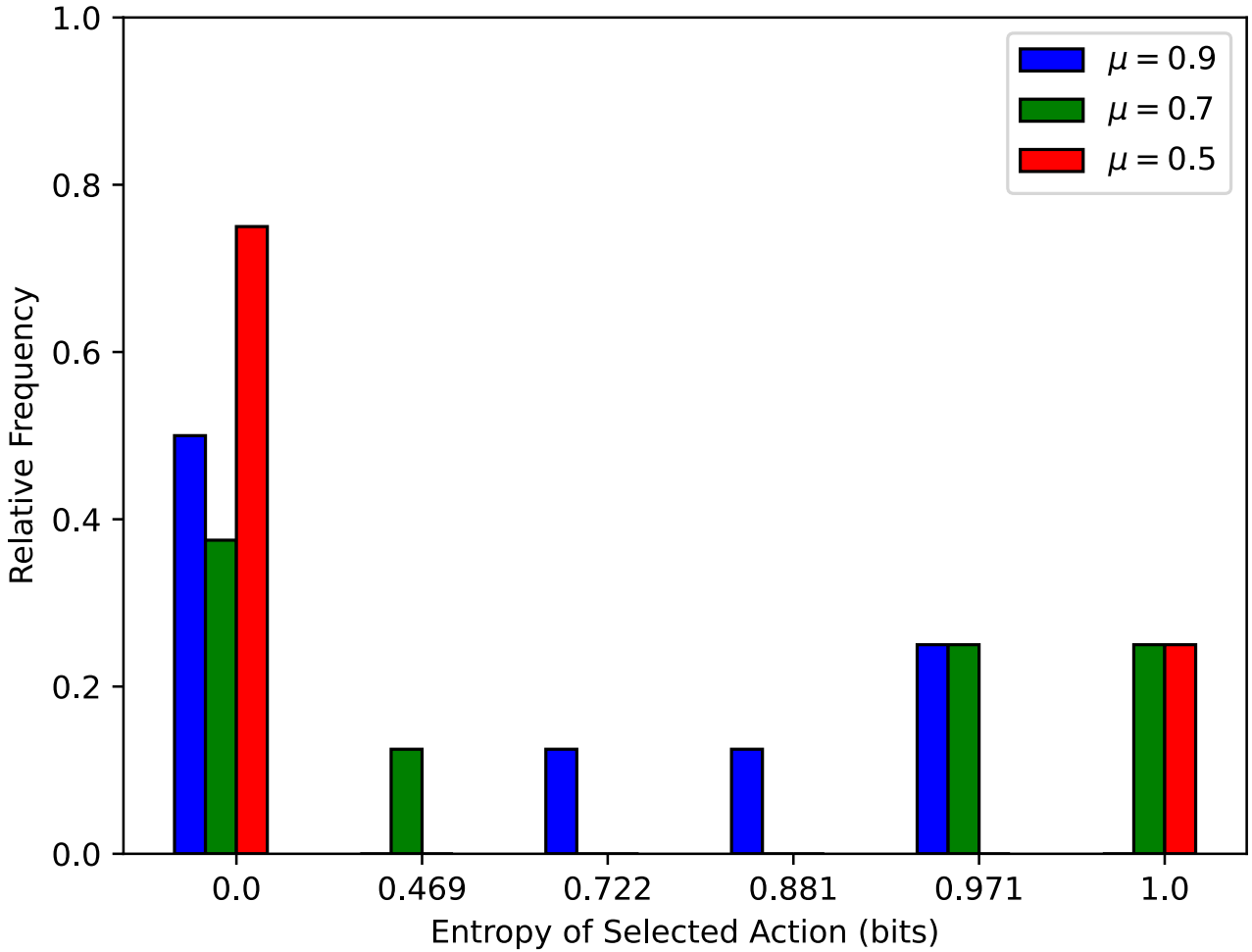
- Stochastic policies provide for weak form of cooperation, needed with distributed decision making and partial observability

Performance with 2 Jammers, 2 Drones



- Solid = stochastic
- Dashed = deterministic
- μ = momentum
- Rewards:
  - 100 if both drones are jammed
  - 25 if one drone jammed
  - -100 if neither drone jammed

Effects of Momentum on Policies



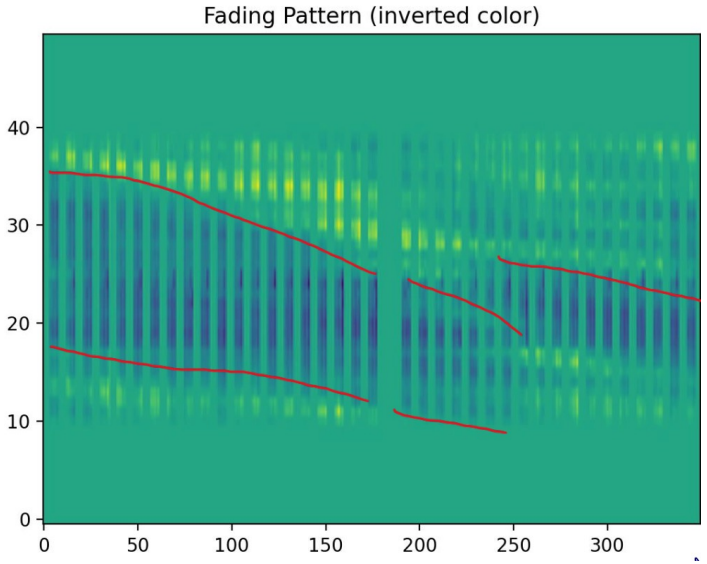
Implementations at UF Autonomy Park

- Undergrad researchers working on:
- Mobile jamming of ground networks using UAVs
- Jammer payloads consist of
  - Raspberry Pi
  - USRP B210 SDR
  - LP directional antenna

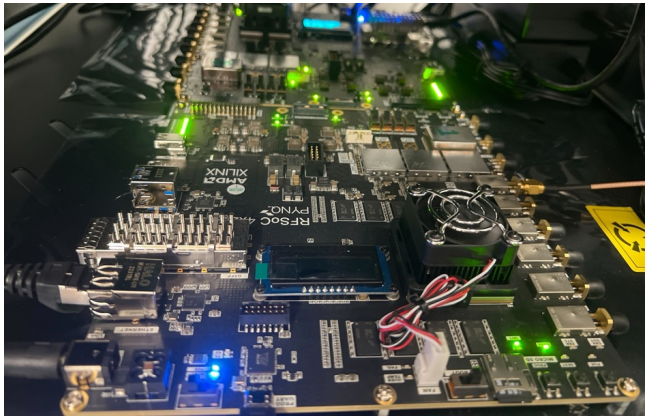
UF UNIVERSITY of FLORIDA · Duke UNIVERSITY · TEXAS The University of Texas at Austin · UC SANTA CRUZ

Implementations at UF Autonomy Park

- Building algorithms for drone detection based on propagation characteristics of drone channel
- Observed time-varying deep nulls in frequency domain



- Working on new real-time detection algorithms using RFSoC boards



Distributed Sensing and Delivery

- Data from sensors often must be delivered to sensor fusion centers (SFCs) for processing
- We consider two main scenarios
  - For fixed sensor networks, a field of sensors must deliver data to SFC(s) over a wireless channel
  - For mobile sensing platforms, the mobile agents may have to collect data and then travel within the comms range of one or more access points (APs)
- In the fixed sensor network, sensors must decide how to use wireless channel
  - Ex: whether or when to transmit to AP

Distributed Sensing and Delivery

- Mobile sensing platforms must decide
  - When to collect data vs deliver data
  - Where to collect data and where to deliver data
  - How to use wireless channel (such as which AP to associate with)

Fixed Sensing Network

Distributed Sensing and Coordination: Who senses and transmits? $\Rightarrow$ Partially observable, multi-agent MDP

UF UNIVERSITY of FLORIDA · Duke UNIVERSITY · TEXAS The University of Texas at Austin · UC SANTA CRUZ

Sensing Model

- Example: use distributed sensors to localize a moving vehicle
- Sensors listen to channel for signal energy (e.g., audio or RF)
- Received signal energy decreases with distance
- Signal energy measurements corrupted by noise and/or interference

- Sensors relay data to a single SFC over a **shared** wireless channel
  - collisions occur if two sensors transmit simultaneously

Optimization Formulation

- Assume slotted time
- Sensors decide individually whether to transmit in each slot
- **Cooperative goal:** minimize the mean-squared error of the location estimate at the SFC

- **Optimal rule is likely stochastic**:
  - For example, always transmitting from sensor with highest SNR prevents ability to triangulate vehicle

- Problem formulated as a **decentralized, partially-observable Markov decision process (DEC-POMDP)**

Fusion Model

- Sensor network updates beliefs in each interval (using data or model-based)

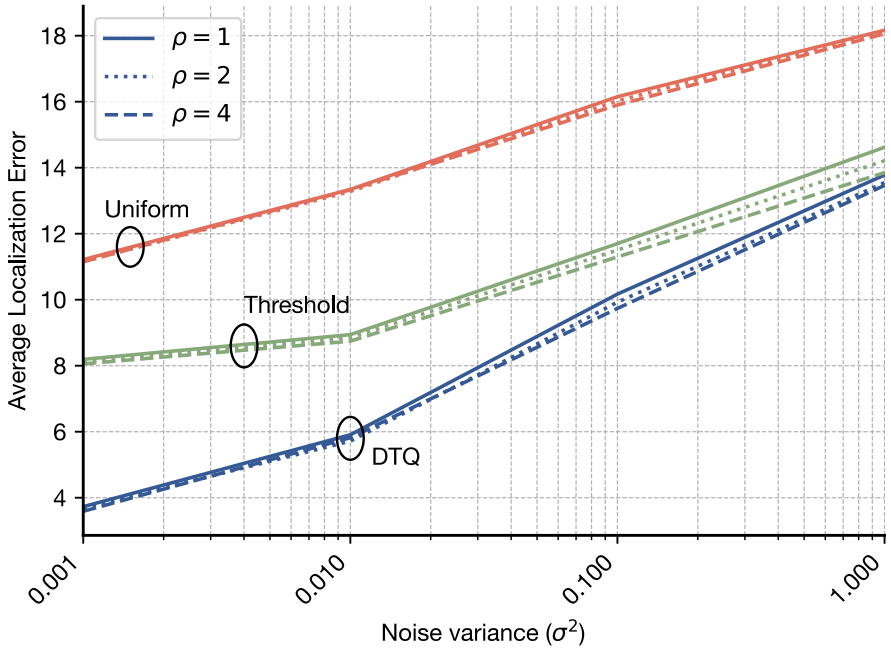
MAC Transmission Control

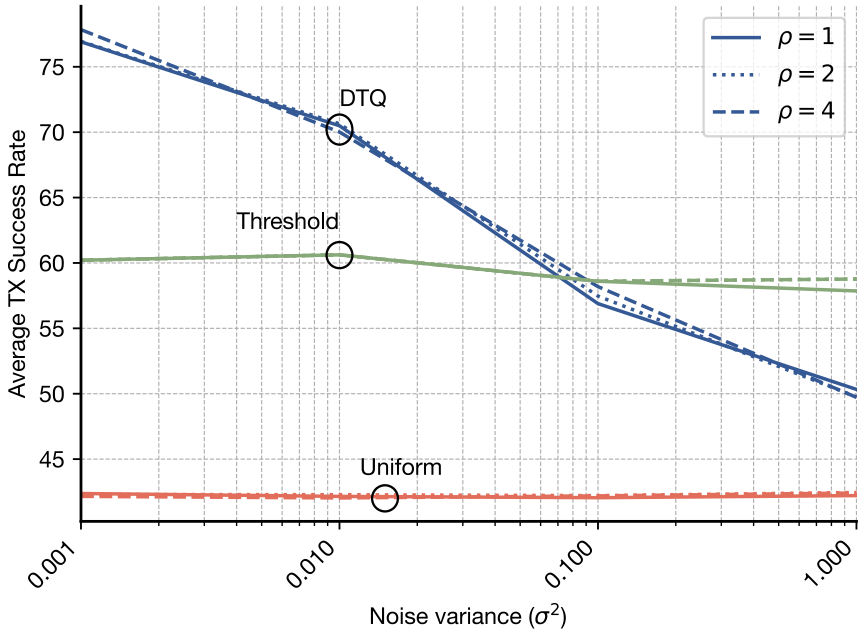
- Information agents may use in making this decision:
  - Received signal strength
  - Result of last channel access (success/failure)
  - Current beliefs (need to be broadcast by fusion agent)
  - # slots since last (successful) transmission by this agent

- Use RL to learn *probability* node should transmit given current state and received energy
- For current results, use tabular Q-learning where beliefs are compressed into MAP state estimate and quantized entropy

Baseline Algorithms

- Compare performance to non-ML approaches:
  - Optimal slotted-ALOHA: equal probability of transmission of $1/m$ for all m sensors
  - Threshold-based policy: transmit if received power greater than threshold; optimal threshold is found via search


$\rho = 1$
$\rho = 2$
$\rho = 4$
Uniform
Threshold
DTQ
Average Localization Error
Noise variance (σ^2)
18
16
14
12
10
8
6
4
0.001
0.010
0.100
1.000

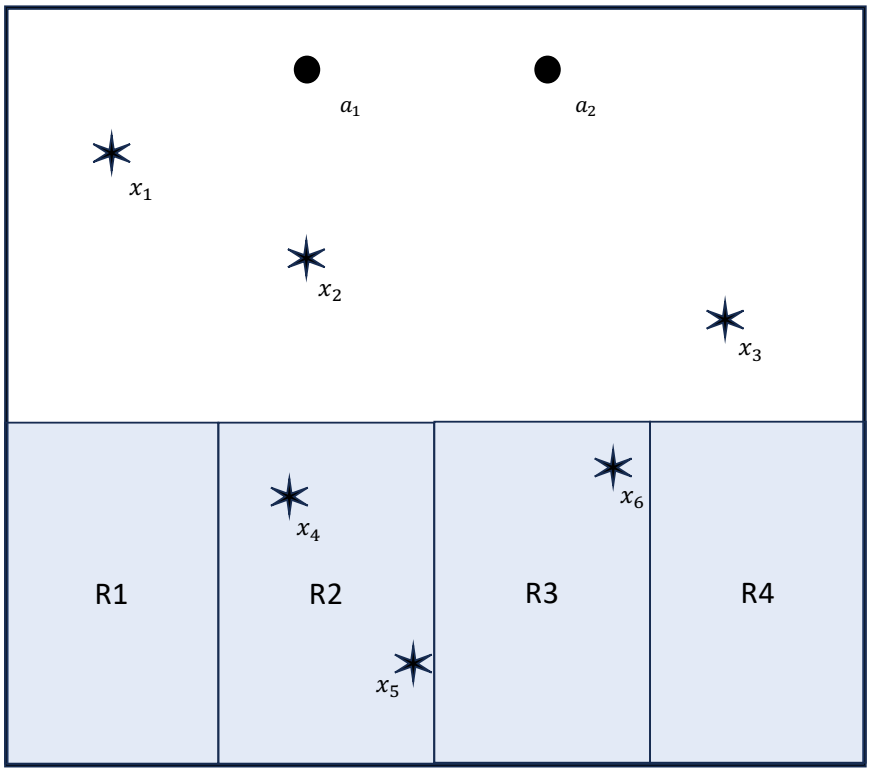

$\rho = 1$
$\rho = 2$
$\rho = 4$
DTQ
Threshold
Uniform
Average TX Success Rate
Noise variance (σ^2)
75
70
65
60
55
50
45
0.001
0.010
0.100
1.000

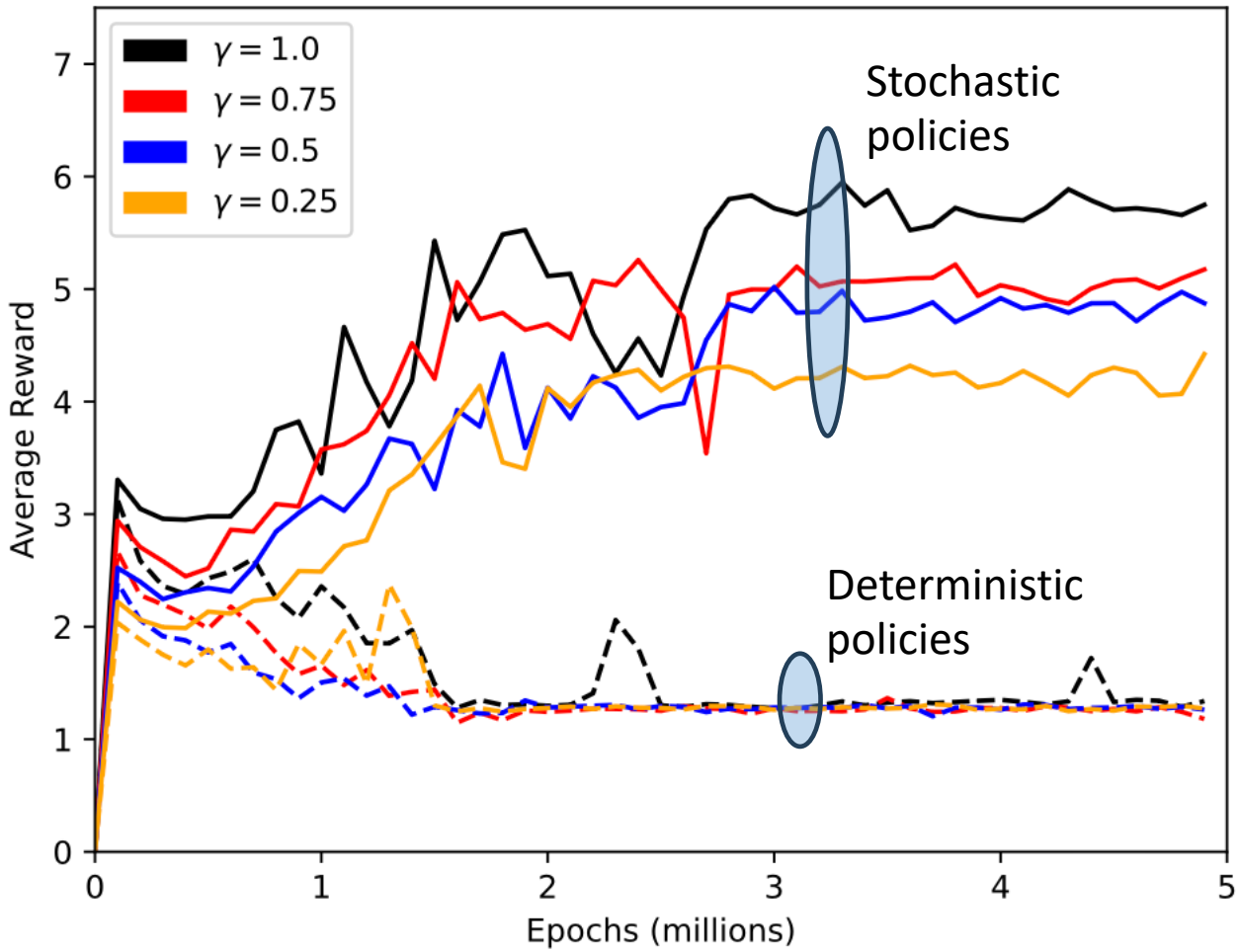
Data Collection and Delivery

- In this scenario, mobile agents monitor an area for some phenomena of interest and then deliver the data to one of multiple access points (Aps)

- Agents work independently but may observe same or related phenomena
  - Thus, agents with data to deliver may be in same geographic area
- For a single agent, choice of AP is simple because will travel to whatever AP achieves fastest data delivery (depending on distance and supported data rates)
- For multiple agents, selfish choice leads to overloading of best AP
  - Agents can use a stochastic approach to avoid all choosing the same AP

- **Example scenario:** 6 agents and 2 APs
- APs and agents have long-range, low-rate data link to share affiliation information




$\gamma = 1.0$
$\gamma = 0.75$
$\gamma = 0.5$
$\gamma = 0.25$
Stochastic policies
Deterministic policies
Average Reward
Epochs (millions)
0
1
2
3
4
5
6
7

Related Research

- Other research using RL discussed at previous Center meetings:
  - Dynamic spectrum sharing
  - Distributed timing synchronization for localization in GPS-denied areas

Conclusions

- DEC-POMDPs offer good model for many systems involving sensing, communications, and control
- However, all POMDPs are hard to solve
- Optimal policies for DEC-POMDPs are not necessarily deterministic
- Have shown in several scenarios of interest that stochastic policies can significantly outperform deterministic policies
- Because all POMDPs have continuous state spaces (beliefs), solutions via function approximation (NNs) are appropriate
  - Ongoing work on developing new approaches to learn stochastic policies using policy gradient approaches

Thank you!


UF
UNIVERSITY of
FLORIDA


Duke
UNIVERSITY


TEXAS
The University of Texas at Austin


UC SANTA CRUZ