Agile Waveform Design: RL to Avoid Pattern Detection

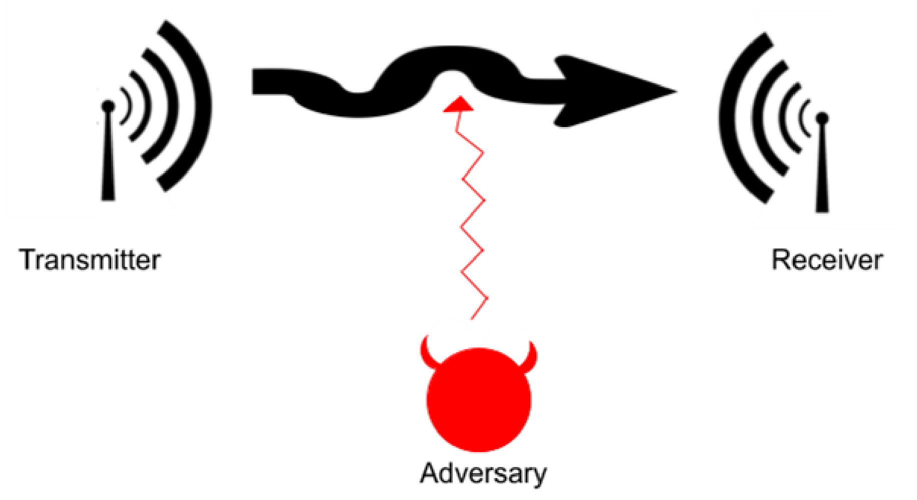
Priya R. Juarez, Jonathan Piland, Matthew Traum

Project Leads: Suya Wu, Dr. Robert Calderbank, Dr. Vahid Tarokh, Dr. Ali Pezeshki

Background

This project was inspired by the Air Force Research Laboratory and their agile waveform generator, which is used to represent various sensor data. The waveform generator communication systems respond to attacks by modifying the pattern of transmission in order to avoid jamming, but often with an associated cost. **In this project, we analyze the relative strength of various agents and test their performance in different environments.**

The Game



- Agents: **Transmitter** and **Receiver** (Team A), **Adversary** (Team B)
- M possible bandwidths to transmit over
- N transmission policies known to all players
- Goal: maximize both Team A and Team B’s points
- Rewards/costs:
  - Team A +*R1* for successful transmission
  - Team A –*R2* for switching policy
  - Team B +*R3* for successful interception

Transmitters

- **Self-Predicting:** mathematically based with an internal adversary (Gamma - Policy) to predict adversary choices
- **RL / RNN:** RNN using reinforcement learning
- **Random:** switches to a random policy with a fixed probability *p*
- **DQL:** Deep Q-Learning with linear NN

Adversaries

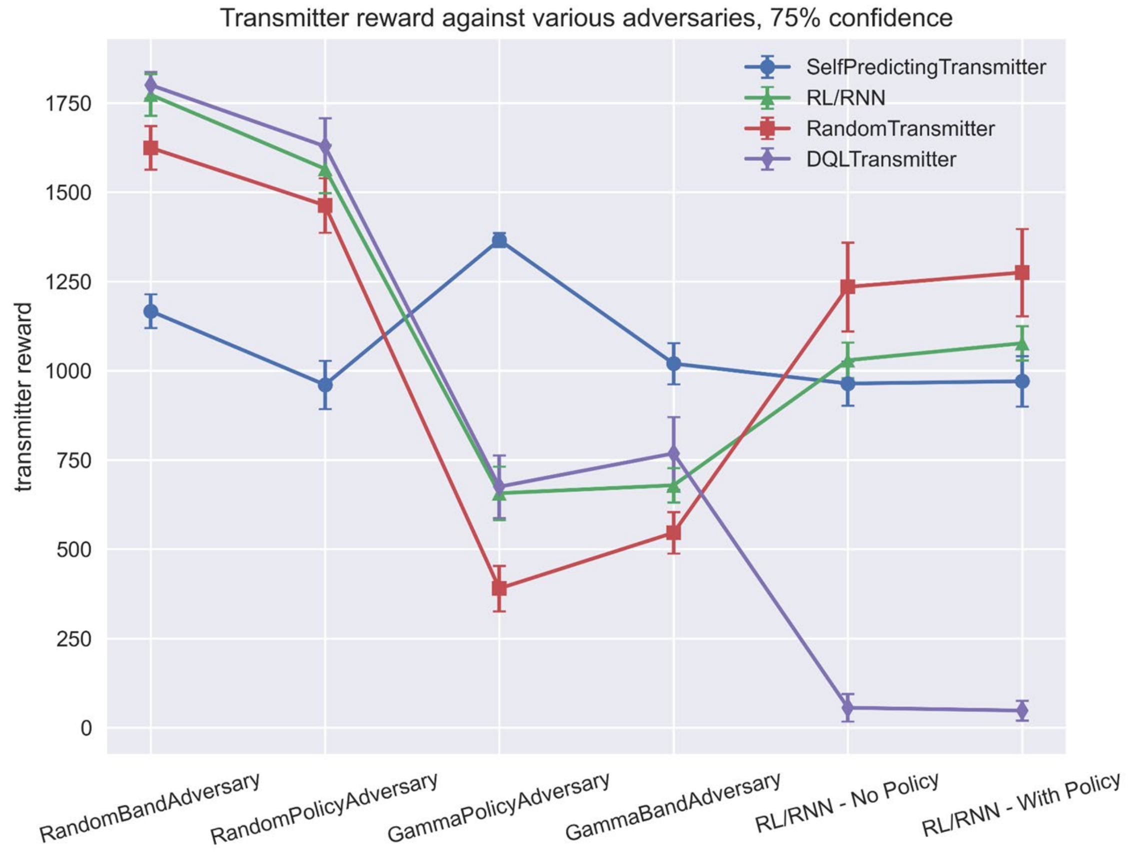
- **Random Band:** chooses bandwidth randomly
- **Random Policy:** chooses policy randomly
- **Gamma Policy:** mathematically based, predicts most likely policy being followed
- **Gamma Band:** improved version of Gamma Policy that instead determines most likely bandwidth
- **RL / RNN - No policy:** RNN using reinforcement learning (transmitter policy choice is unknown)
- **RL / RNN - With policy:** RNN using reinforcement learning (transmitter policy choice is known)

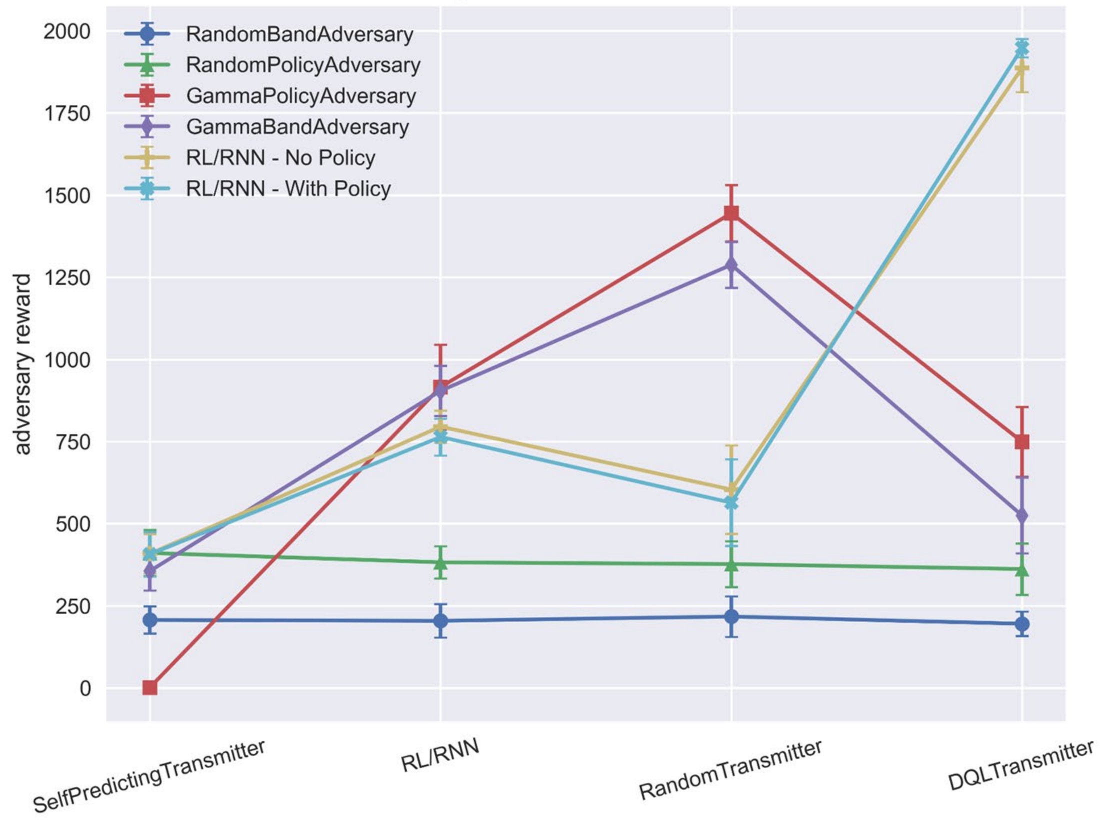
Policy Similarity

The policies (integer sequences) often have inherent overlap, so we developed a metric to quantify the similarity of a set of policies:

$$S = \sum_{t=0}^{T}\sum_{i=0}^{M}\frac{\mathrm{Count}(t,b_i)-1}{T\cdot N} \quad \text{(Note, } b_i \text{ is bandwidth, } t<T \text{ is time)}$$

Results


Transmitter reward against various adversaries, 75% confidence


Adversarial reward against various transmitters, 75% confidence

Conclusions and Next Steps

Against an opponent who plays randomly, the optimal strategy is to use a mathematical formula. While RL efforts show promise against some opponents, more research is needed to improve consistency.

Future directions might include the following:

- Adjust and test hyper-parameters
- Develop receiver and policy-making agents
- Add rate adaptation and live data/signals
- Improve DQN to double or dueling DQN