Multi-Agent Adversarial Attacks for Multi-Channel Communications

Juncheng Dong, Suya Wu, Mohammadreza Soltani
Vahid Tarokh
Overview

1. Problem and Assumption

2. Introduction to RL and MARL

3. Multi-agent Deep Q-Network (MADQN) Jammers

4. Experimental Results
   Attack for Single-channel transmission
   Attack for Multi-channel transmission

5. Future work
Motivation

- Jamming attacks can be a real threat to assorted communications.
- From jammer’s perspective, a more efficient and powerful jamming system is desired while majority of jamming/anti-jamming publications focus on anti-jamming [Pirayesh and Zeng, 2021].
- From anti-jammer’s perspective, current intelligent anti-jamming framework are not designed to prevent from smart jammer (self-learning jammers) [Xu et al., 2020].
- Study of self-learning jammers leads to better understanding of jammers’ learning behavior, thus possible improved defense mechanism

Objective: A Multi-Jammer System based on Reinforcement Learning that

1. Adapts to unknown environment
2. Learns to improve its jamming success rate
System Model-Assumptions and Notations

- Sender $S$ and Receiver $R$. At each time $t$,
  - $M$ available channels
  - Single-band transmission, the sender $S$ choose current channel $C_s^{(t)}$ to send signals.
  - Multi-band transmission, the sender $S$ choose current channels $C_{S,\ell}^{(t)}$, and corresponding powers $P_{S,\ell}^{(t)}$, $\ell = 1, \ldots, L$, where $L \leq M$.
- Jammers $J_i$, $i = 1, \ldots, N$. At each time $t$,
  - $J_i$ listens to all channels and gains some information
  - $J_i$ takes actions $A_i^{(t)} = [P_i^{(t)}, C_i^{(t)}]$, where $P_i^{(t)}$ and $C_i^{(t)}$ are current power and channel chosen by the jammer $J_i$
Figure 1: Multi-Jamming Wireless Communication System.
System Model-Successful Attack

- Jammer $J_j$ attacks the channel by taking actions $A_i^{(t)} = [P_i^{(t)}, C_i^{(t)}]$. 
- Low signal-to-interference-plus-noise ratio (SINR), where

$$\text{SINR}^{(t)} = \frac{P_S^{(t)} \cdot h_S}{\text{Noises} + \sum_{i=1}^{N} P_i^{(t)} \cdot h_i \cdot I(C_i^{(t)} = C_S^{(t)})}.$$ 

$h_S$ and $h_i$ are power gains from sender and jammer $J_i$ respectively. It’s unrealistic for jammer to know true SINR from receiver, thus we need an estimation of SINR.

- Instant Success, $G^{(t)} = \mathbb{I} \left( \text{SINR}^{(t)} < \tau \right)$, where $\tau$ is a pre-defined threshold.

- Instant Reward: 

$$R^{(t)} = B \ast \left( \log_2(1 + \text{SNR}^{(t)}) - \log_2(1 + \text{SINR}^{(t)}) \right) - \text{Cost}_p \ast \sum_{i=1}^{N} P_i^{(t)},$$

where $B$ is the bandwidth (default $B = 10$ in the simulation), $\text{Cost}_p$ is the cost of unit power of jammers.
Reinforcement Learning

- Reinforcement learning algorithms allows an agent to learn by interacting with the environment to maximize its cumulative received rewards.

Figure 2: Reinforcement Learning.
Reinforcement Learning

- Key elements of reinforcement learning
  - Environment with internal state $s_t \in S$
  - Agent's possible action: $a_t \in A$
  - Agent's policy: $\pi : S \rightarrow A$
  - State transition: $p : S \times A \rightarrow S$
  - Reward function: $R : S \times A \rightarrow \mathbb{R}$

- Goal of RL agent is to maximize cumulative rewards (i.e., selecting a policy to maximize the Q-function/action-value function/value function):

$$\max_{\pi} Q^{\pi}(s_t, a_t) = \max_{\pi} \mathbb{E}\left( \sum_{t=0}^{\infty} \gamma^t R(t) \mid s_t, a_t, \pi \right).$$
Reinforcement learning algorithm is single agent. However, we want to build and study the behavior of a system of multiple collaborative jammers. Multi-agent reinforcement learning algorithm is necessary.

Multi-agent Reinforcement Learning:
- Training: Centralized / Distributed
- Execution: Centralized / Distributed

Centralized Training/Execution requires perfect communication in real time. This is rare and expensive. We choose distributed training/ distributed execution MARL.
Multi-Agent Deep Q-Network (MADQN) Jammers

- Team reward for jammers:
  - Amount of blocked channel: \( B \times \left( \log_2(1 + \text{SNR}^{(t)}) - \log_2(1 + \text{SINR}^{(t)}) \right) \)
  - Jamming is not free: Cost for jamming power \( \text{Cost}_p \)
  - \( R^{(t)} = B \times \left( \log_2(1 + \text{SNR}^{(t)}) - \log_2(1 + \text{SINR}^{(t)}) \right) - \text{Cost}_p \times \sum_{i=1}^{N} P^{(t)}_i \)

- For each jammer:
  - Individual reward perceived by agent
    \( R^{(t)} = B \times \left( \log_2(1 + \text{SNR}) - \log_2(1 + \text{SINR}^{(t)}) \right) - \text{Cost}_p \times P^{(t)}_i \)
  - Deep Q-Network for value function
  - Double Q-Network as fixed target network and actor network for convergence and counteract overestimation problem in initial learning period
  - Prioritized experience replay for faster learning and efficiency of data

Agent's experience at time \( t \) → \((a_t, s_t, r_{t+1}, s_{t+1})\)
Experimental Design

We have tested our model under different scenarios. To avoid being jammed, we assume the sender chooses different strategies to hop across multiple channels.

1. Single-Band Transmission

- Sweep Type, $C_S^{(t)} = t\%N$
- Pulse Type, $C_{S,t} = \begin{cases} 5, & \text{if } t\%N \leq 2; \\ 1, & \text{otherwise}. \end{cases}$
- Autoregressive Type,
  \[
  C_{S,t} = \begin{cases} 
  C_{S,t-1} + i\%N, & \text{if } C_{S,t-1}\%2 = 0 \\
  C_{S,t-1} - i\%N, & \text{if } C_{S,t-1}\%2 = 1 \\
  X_t \in \{1, N\}, & \text{if } C_{S,t-1} > N, \text{where } p(X_t = 1) = 0.1 \text{ and } p(X_t = N) = 0.9 \\
  X_t \in \{1, N\}, & \text{if } C_{S,t-1} < 1, \text{where } p(X_t = 1) = 0.9 \text{ and } p(X_t = N) = 0.1 
  \end{cases}
  \]
- Random Type, $C_S^{(t)} = \text{Uniform}(1, \ldots, N)$

2. Multi-Band Transmission - Sweep, Pulse and Autoregressive Types
Experimental Design

We consider two evaluation metrics:

1. Instant Success Rate, \( G(t) = \mathbb{I}\left(\text{SINR}(t) < \tau\right) \), where \( \tau \) is taken as a half value of maximum SINR.

2. Instant Reward, \( R(t) = B \times \left(\log_2(1 + \text{SNR}(t)) - \log_2(1 + \text{SINR}(t))\right) - \text{Cost}_p \times \sum_{i=1}^{N} P_i(t) \), where \( B = 10 \) and \( \text{Cost}_p > 0 \) denotes the cost of power by each jammer.

We compare the performance of five different type of adversaries:

- Random jamming \( J_{\text{Rand}} \)
- Greedy Adversary \( J_{\text{Gre}} \)
- Single-agent jamming \( J_{\text{Single}} \)
- Multi-agent jamming \( J_{\text{Multi}} \)
- Multi-agent Greedy RL-agent \( J_{\text{GreRL}} \)
Experimental Design

- Power of sender $P_S$
- Power sets of jammers, $P_J = [0, 1, 3, 5]$
- Number of available channels, $M = 5$
- Number of used channels for mult-channel case, $L = 2$
- Number of jamming agents (adversaries), $N = 3$
At each time, the sender picks one channel by \( C_S(t) = t\%M \). Note \( M = 5 \) and constant power \( P_S = 5 \).
Figure 3: Performance of Jamming vs. Discrete Time Under Sweep Changes of a Single Channel.
Figure 4: Performance of Jamming vs. Discrete Time Under Sweep Changes of a Single Channel.
At each time $t$, the sender picks channels $[C_{S,1}^{(t)}, C_{S,2}^{(t)}]$ by $C_{S,\ell}^{(t)} = (t + \ell)\%M$. Note that constant power $P_S = [1, 5]$ and the number of total available channels $M = 5$. 
Multi-Channel, Sweep Type Sender: Success Rates

Figure 5: Performance of Jamming vs. Discrete Time Under Sweep Changes of Multi-Channel.
Multi-Channel, Sweep Type Sender: Instant Rewards

Figure 6: Performance of Jamming vs. Discrete Time Under Sweep Changes of Multi-Channel.
Single-Channel, Pure Random Type Sender

Figure 7: Performance of Jamming vs. Discrete Time Under Random Changes of a Single Channel.
## Performance Overview: Averaged Success Rates

<table>
<thead>
<tr>
<th></th>
<th>Random</th>
<th>Greedy</th>
<th>Greedy RL</th>
<th>Single RL</th>
<th>Multi RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweep-Single</td>
<td>0.378</td>
<td>0.445</td>
<td>0.552</td>
<td>0.771</td>
<td>0.927</td>
</tr>
<tr>
<td>Sweep-Multi</td>
<td>0.579</td>
<td>0.358</td>
<td>0.620</td>
<td>0.879</td>
<td>0.939</td>
</tr>
<tr>
<td>Pulse-Single</td>
<td>0.374</td>
<td>0.416</td>
<td>0.384</td>
<td>0.768</td>
<td>0.923</td>
</tr>
<tr>
<td>Pulse-Multi</td>
<td>0.698</td>
<td>0.431</td>
<td>0.718</td>
<td>0.867</td>
<td>0.959</td>
</tr>
<tr>
<td>AR-Single</td>
<td>0.404</td>
<td>0.607</td>
<td>0.694</td>
<td>0.792</td>
<td>0.927</td>
</tr>
<tr>
<td>AR-Multi</td>
<td>0.503</td>
<td>0.364</td>
<td>0.393</td>
<td>0.687</td>
<td>0.845</td>
</tr>
</tbody>
</table>

Table 1: Success Jamming Rate for Various Jammers Under Assorted Communication Scenarios.

- **Greedy**: Record average reward of its actions and choose the action with the highest history reward (Variation of Multi-Armed Bandit problem)
- **Greedy RL**: $\epsilon$-greedy RL agent with $\epsilon = 0$ (Skip the exploration part in exploration/exploitation dilemma)
Experimental Results

- In different scenarios, multi-agent jamming outperforms single-agent jamming, and gain much in multi-channel cases.
- With low cost of unit jamming power, the multi-agent jamming benefits more advantages than single-agent jamming.
- More realistic simulations need to be considered.
Future Work

- Estimation of SNR and SINR under realistic cases
- Multi-agent jamming that each jammer can communicates with each other
  - Jammers can choose their actions based on communicating with each other in a given jammer communication network
  - Jammers can jam in more than one channel
- Centralized multi-agent jamming