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 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 Learnring 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

System Model-Illustration

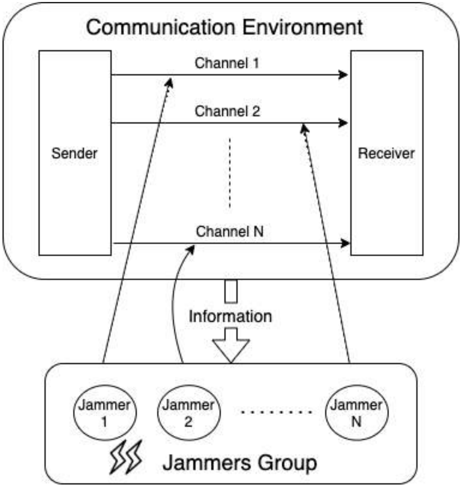

Figure 1: Multi-Jamming Wireless Communication System.

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

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

  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 τ is a pre-defined threshold.
- **Instant Reward:**
  $R^{(t)} = B * \left(\log_2(1 + \text{SNR}^{(t)}) - \log_2(1 + \text{SINR}^{(t)})\right) - \text{Cost}_p * \sum_{i=1}^{N} P_i^{(t)}$, where B is the bandwidth (default $B = 10$ in the simulation), 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.

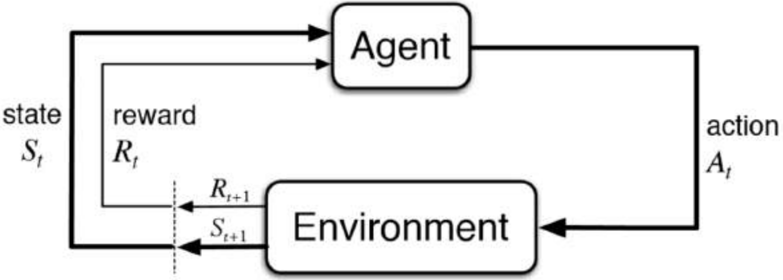

Figure 2: Reinforcement Learning.

Reinforcement Learning

- Key elements of reinforcement learning
  - Environment with internal state $s_t \in \mathcal{S}$
  - Agent's possible action: $a_t \in \mathcal{A}$
  - Agent's policy: $\pi : \mathcal{S} \to \mathcal{A}$
  - State transition: $p : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$
  - Reward function: $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$
- Goal of RL agent is to maximize cumulative rewards (i.e., selecting a policy to maximize the Q-function/action-value function/value function):

$$\max_{\pi} Q^{\pi}(s_t, a_t) = \max_{\pi} \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t R^{(t)} | s_t, a_t; \pi \right).$$

Multi-agent Reinforcement Learning

- 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: $B * \left(\log_2(1+\text{SNR}^{(t)}) - \log_2(1+\text{SINR}^{(t)})\right)$
  - Jamming is not free: Cost for jamming power $Cost_p$
  - $R^{(t)} = B * \left(\log_2(1+\text{SNR}^{(t)}) - \log_2(1+\text{SINR}^{(t)})\right) - \text{Cost}_p * \sum_{i=1}^{N} P_i^{(t)}$
- For each jammer:
  - Individual reward perceived by agent
    $R^{(t)} = B * \left(\log_2(1+\text{SNR}) - \log_2(1+\text{SINR}^{(t)})\right) - \text{Cost}_p * P_i^{(t)}$
  - 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 $\rightarrow$ $(a_t, s_t, r_{t+1}, s_{t+1})$

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, $C_{S,t} = \begin{cases} 5, \text{ if } t\%N \leq 2; \\ 1, \ o.w. \end{cases}$
   - 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 τ is taken as a half value of maximum SINR.
2. Instant Reward, $R^{(t)} = B * \left(\log_2(1 + \text{SNR}^{(t)}) - \log_2(1 + \text{SINR}^{(t)})\right) - \text{Cost}_p * \sum_{i=1}^{N} P_i^{(t)}$, 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_{Rand}
- Greedy Adversary J_{Gre}
- Single-agent jamming J_{Single}
- Multi-agent jamming J_{Multi}
- Multi-agent Greedy RL-agent J_{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), $N = 3$

Single-Channel, Sweep-Type Sender

At each time, the sender picks one channel by $C_S^{(t)} = t\%M$. Note $M = 5$ and constant power $P_S = 5$.

Single-Channel, Sweep Type Sender - Success Rate

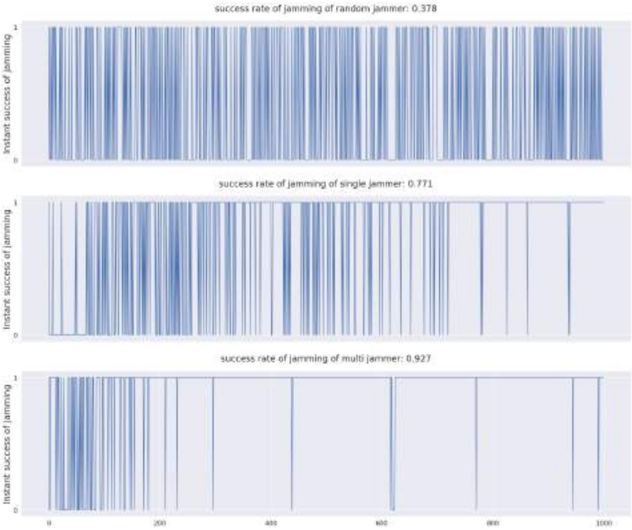

Figure 3: Performance of Jamming vs. Discrete Time Under Sweep Changes of a Single Channel.

Single-Channel, Sweep Type Sender: Instant Rewards

Figure 4: Performance of Jamming vs. Discrete Time Under Sweep Changes of a Single Channel.

Multi-Channel, Sweep Type Sender

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

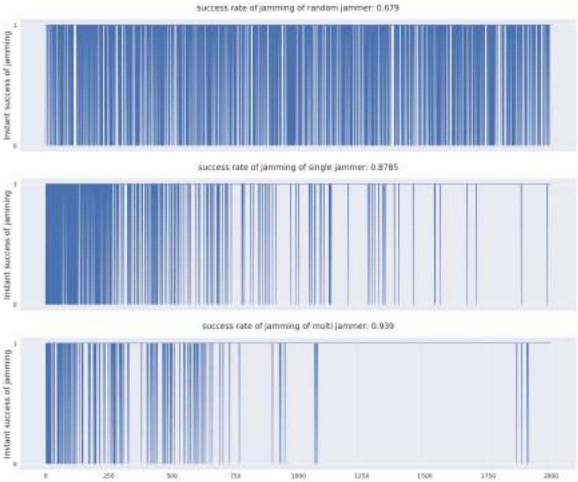

Figure 5: Performance of Jamming vs. Discrete Time Under Sweep Changes of Multi-Channel.

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

	Random	Greedy	Greedy RL	Single RL	**Multi RL**
Sweep-Single	0.378	0.445	0.552	0.771	0.927
Sweep-Multi	0.579	0.358	0.620	0.879	0.939
Pulse-Single	0.374	0.416	0.384	0.768	0.923
Pulse-Multi	0.698	0.431	0.718	0.867	0.959
AR-Single	0.404	0.607	0.694	0.792	0.927
AR-Multi	0.503	0.364	0.393	0.687	0.845

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: ϵ-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

References I

Pirayesh, H. and Zeng, H. (2021).
Jamming Attacks and Anti-Jamming Strategies in Wireless Networks: A Comprehensive Survey.

Xu, J., Lou, H., Zhang, W., and Sang, G. (2020).
An intelligent anti-jamming scheme for cognitive radio based on deep reinforcement learning.
IEEE Access, 8:202563–202572.