WORCESTER POLYTECHNIC INSTITUTE

Team 7 Final Report: Learning a Protocol for Minimum Probability of Detection Wireless Transmissions: A DQN Experiment

Mac Carr, Bryan Nguon, Kyle McClintick

December 11, 2019

1 Definitions and Problem Statement

A common goal of military communications is to achieve the lowest probability of detection and the highest throughput possible, where detection of the signal by an adversary means leaked information or exposure to jamming, and higher throughput means urgent information is delivered quickly. Typically, detection is avoided by using Direct Sequence Spread Spectrum (DSSS), or operating below the noise floor but still managing to recover the signal by correlation with an encryption key (Pseudo random Noise or PN sequence made to look like noise) shared by receiver and transmitter. However, there are techniques available to estimate these keys.

The motivation of this project is as follows: can a wireless transmission be taught:

- An intelligent jammer's detection rule
- Spectrum traffic patterns

such that throughput can be maximized by avoiding as much interference as possible while also avoiding detection? This learning is anticipated to cause the transmitter to 'hide' amongst other signals in the wireless spectrum, allowing acceptable reductions to throughput to avoid high power jamming. The strength of this technique is that the agent or transmitter has no knowledge of how the adversary detects them, it learns by exploring and making mistakes, where detection means high penalties to rewards.

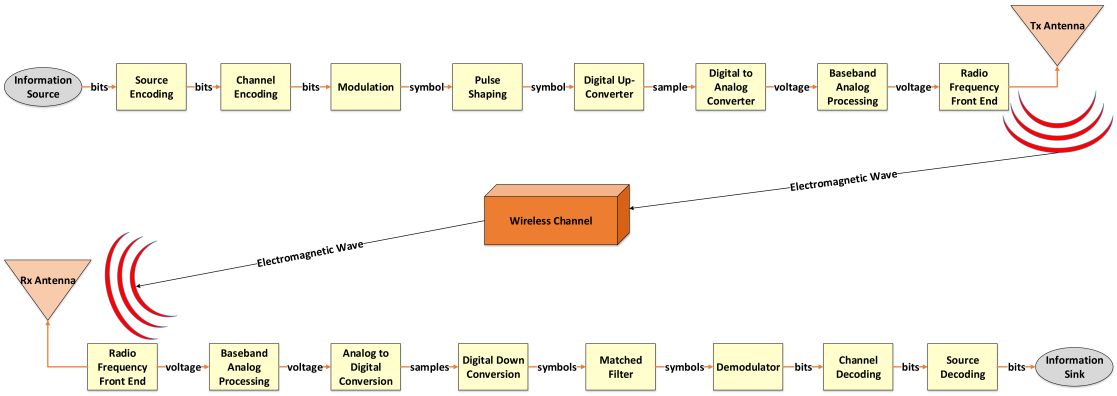

Figure 1.1: The sequential model for estimating wirelessly transmitted binary data across a noisy wireless channel. Our Python simulation performs each of these tasks except for source encoding, introducing many realistic aspects of a real wireless experiment.

How do we define our action, state, and rewards? Rather than throughput, we compare a similar metric, Bit Error Rate (BER), or the percentage of bits sent that are correctly estimated by the receiver. Higher rewards must have higher goodness, and BER is typically very small even when performing badly $BER \in \{0, 0.5\}$, so an inverse log transform is applied. The action space allows the transmitter to choose both its power and carrier frequency. Much like talking in different rooms, signals are typically organized by carrier frequency to allow multiple conversations at once without disruption. Allowing power choice gives the agent another tool to

avoid transmitting too strongly and being detected by the adversary, in addition to purposely choosing an frequency occupied by other signals.

$$BER = \frac{correctbits}{incorrectbits} \tag{1.1}$$

$$reward = \log(1/(BER + \epsilon)) \tag{1.2}$$

$$f_c \in \{f_s/10, f_s/2\} \tag{1.3}$$

$$power \in \{0.1, 1.0\} \tag{1.4}$$

$$action \in \{(power, f_c)\} \tag{1.5}$$

$$state \in \mathbb{R}^{(BS,1,257,311)} \tag{1.6}$$

2 Non-Adversarial Case

Wireless traffic of a wide variety causes unintended interference. In this case, actions do not effect the game state, as there is no adversary that responds to our actions. $N = 10,000$ bits are sent per 'packet', making up a single time step. Other signals in the spectrum can be defined as wide-band signals, tones, or hopping tones.

EXAMPLE SIGNALS HERE, TRIPLE PLOT

Wide-band signals are generated by random noise filtered by a Butterworth Band Pass Filter (BPF) of random cutoff and order. Random noise is used as we do not care about the specific bits sent by these signals. Tones occupy a random frequency from the set of frequencies our agent is given, and hopping tones occupy a random number of those frequencies (at least two of them) and switch between those frequencies at a random interval of time. All of these behaviors are set at the start of an episode, during which 100 packets are transmitted.

The DQN used was the same as in project three, however our hyper parameters vary greatly for some key reasons. Firstly, while the training is episodic in nature, the distributions driving spectrum behavior is stationary, such that actions taken at one time step do not impact reward observed at another time step. For this reason, we set our discount factor $\gamma = 0$ to avoid teaching the DQN that transmitting on a low interference frequency at some time step and getting a good reward somehow causes a bad reward at a later time step when transmitting on a high interference frequency. Additionally, since this is such a simple pattern to learn, we dramatically decreased the number of episodes to 50, and experienced rapid overfitting in our reward curve until learning rate was reduced $\alpha = 1 \times 10^{-6}$ and dropout was introduced to the fully connected layers $p = 0.5$.

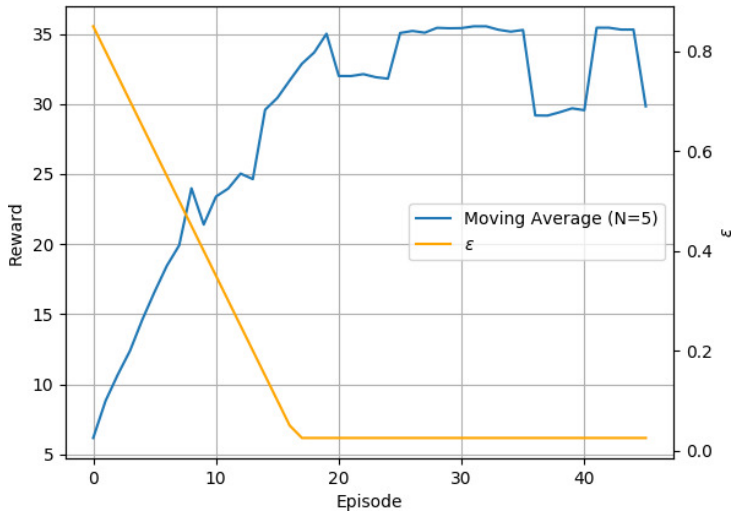

Figure 2.1: Training reaches a maximum reward very quickly as exploration ceases in this simple problem. Testing average over 100 episodes gave 31.30 reward ($\sim 2.5 \times 10^{-14}$ BER)

3 Adversarial Case

Signal to Interference and Noise Ratio (SINR) is computed as in-band signal power divided by the sum of noise and interference power: $P_s/(P_i + P_n)$. The adversary is assumed to be able to detect our transmission if this ratio exceeds $0dB$ or a one to one ratio.