

Section II: Methodology

Role of Student vs. Mentor

I decided the direction of the project, read the research papers noted in the project proposal, and conducted the research. During the months August to October, Dr. Kevin Crowthers was my adviser, which was when the work on this project was started. At the end of October, Mr. Nicholas Medeiros became my advisor.

Equipment and Materials

To complete the research, Python version 3.9.13 was used. The major libraries used in the code are PettingZoo 1.22.0, specifically the Simple and Simple Crypto environment (which handled the game environment), and torch 1.13.0, specifically the torch.nn class which was used to construct the neural networks. The foundation for the code used in this project was based on the paper for implementing PPO (Proximal Policy Optimization) (Terry et al., 2020). A simple multilayer perceptron model was created to test the black box AI. Unfortunately, this test was unfinished. Continued work is necessary to finish this paper.

As an intermediate step in this research project, another game environment, called Simple, was used to check if the logic behind the code worked. In this environment, an agent takes its relative position and velocity (both x and y components) to a goal as inputs, and outputs an action from a predetermined set of actions (accelerate up, down, left, or right).

Technique 1 – Simple Crypto Environment

The technique was supposed to be to compare the performances of four teams of AI against a very large black box AI. The goal of this was to simulate what ideal play would look like against any strategy. Since the Simple Crypto environment is focused on communication, the idea was to test how well a white box and black box AI are able to communicate by comparing them to a team of white box AI and a team of black box AI.

This test was not finished in the allotted time period and it shall continue to be pursued. There were many changes that needed to be made to the implementation of PPO, and the code is currently unable to run. This code will likely be able to work in the near future, as the Simple environment currently works.

Technique 2 – Simple Environment

The simple environment from the PettingZoo library was used to test that the logic in the code was correct. This environment was run for 5000 episodes with a pytorch sequential neural network. The network had 4 inputs, 8 intermediate layers with 1,000 neurons, and 1 layer with 10,000 neurons in that order. The actor had 4 outputs for each of the actions and the critic had one output. The logic used to update the functions is based on the paper from “PettingZoo: Gym for Multi-Agent Reinforcement Learning” (Terry et. al, 2020).

There were two policies tested, one with the black box AI and one which was random. Each was run for 5,000 episodes, and the average episodic return was calculated for each episode.