A Reinforcement Learning Approach to Ship Towing Using Two Tugboats

Name: Mohammad Saifullah Khan
Roll No.: 21169

TEAM:
Rahul Kulkarni
Mohammad Saifullah Khan

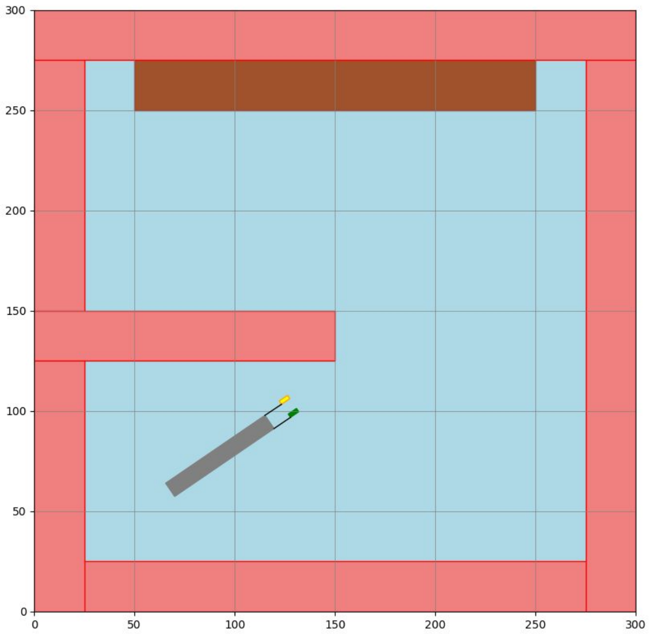
Problem Statement

Towing a ship to target location by two tugboats using Multi-Agent Deep Deterministic Policy Gradient (MADDPG)

Challenges

- Coordination among agents (tugboats).
- High dimensional and continuous action space.
- Scalability and complexity of task
- Reward shaping and credit assignment.
- Adaptation to different ship types and conditions.
- Real world transfer of simulation.

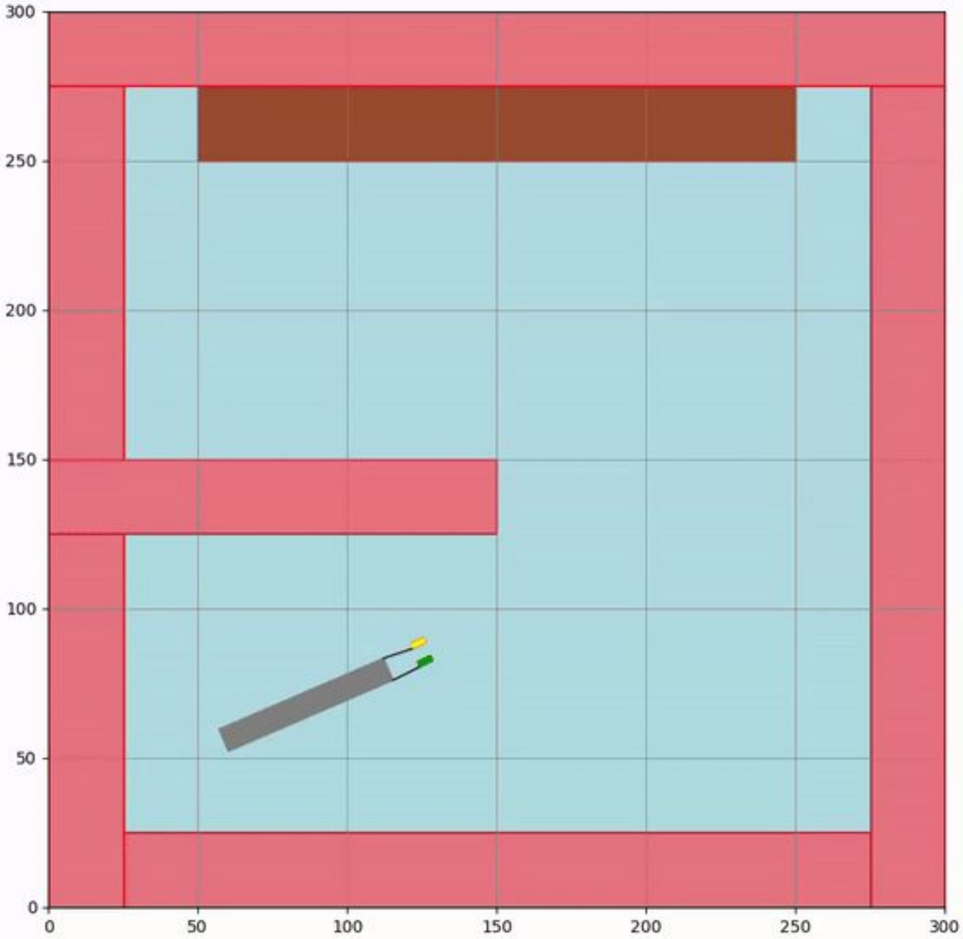
Approach



Built custom environment.

- Observation space of each tugboat: ship position and orientation, position and orientation of own and other agent, distance of ship from target (Euclidean), rope length.
- Action space of each tugboat: velocity in x and y direction.
- Rope (black coloured) considered as a stiff rod for simplicity.
- A force applied on ship due to movement of tugboats.

Agents = tugboats.
Obstacles: red coloured.
Dock: brown coloured.
Water: blue coloured


300
250
200
150
100
50
0
0
50
100
150
200
250
300

Reward structure:

$R_{\text{to\_target}}$ = - ds/(grid_size*0.004)

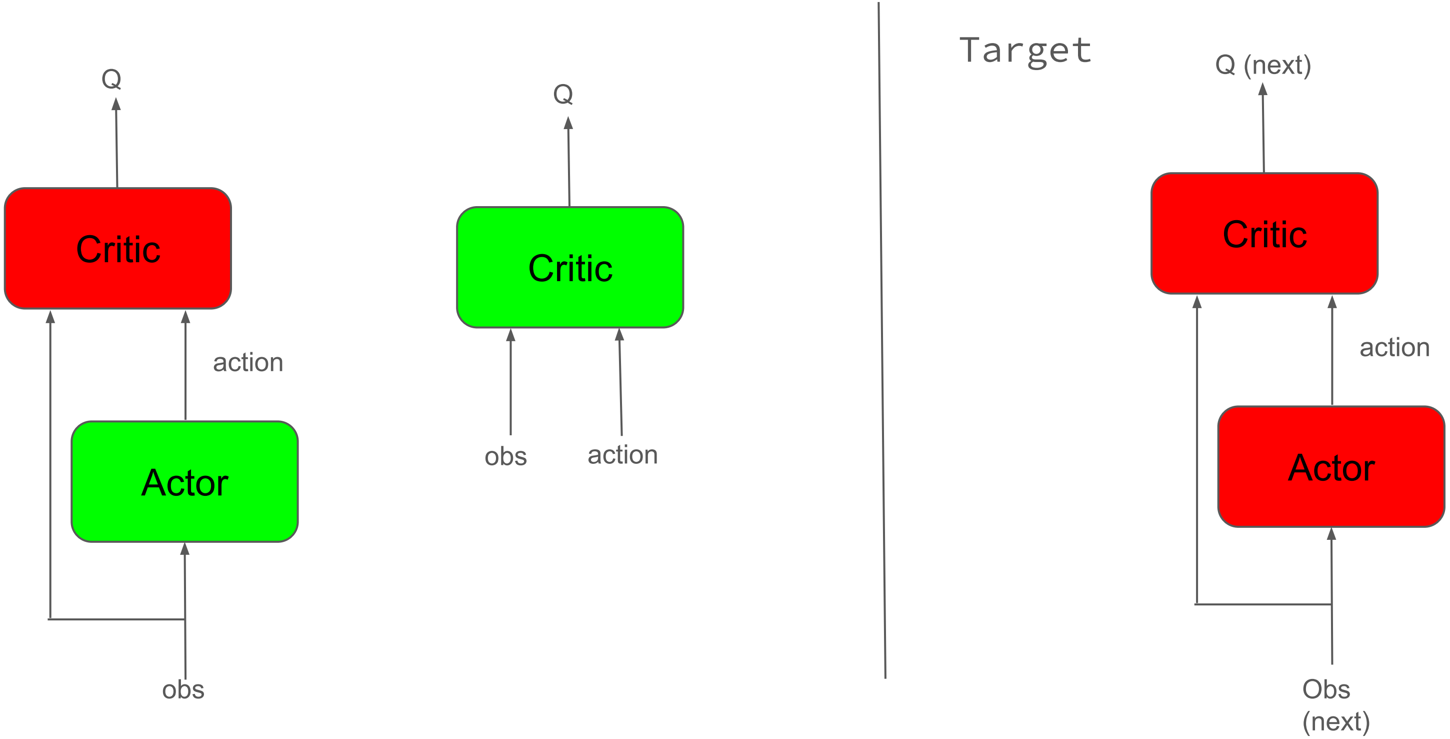
$R_{\text{proximity}}$ = Penalty for coming close to obstacles (for both ship and tugboats)

$R_{\text{in\_target}}$ = 1000

$R_{\text{rope\_length}}$ = penalty for exceeding rope length

$$R_{\text{total}} = R_{\text{to\_target}} + R_{\text{proximity}} + R_{\text{rope\_length}} + R_{\text{in\_target}}$$

DDPG



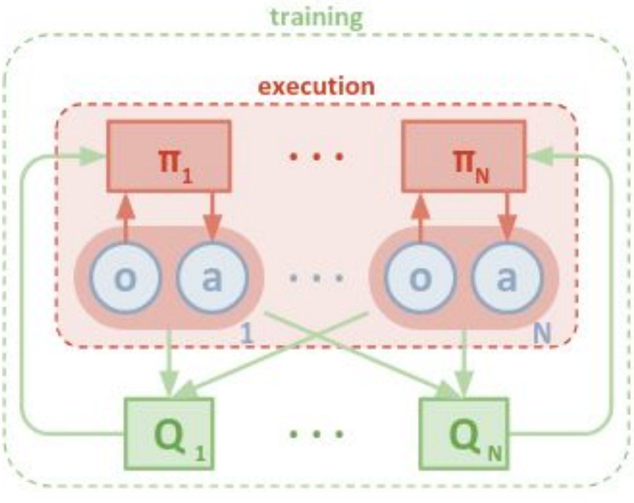
Trained to maximize Q

$$\text{minimize}(Q - |\ \text{reward} + \text{discount} * Q_{next}|)$$

MADDPG

- Soft update of parameters.
  $\theta_{target} \leftarrow \tau\theta_{online} + (1-\tau)\theta_{target}$

- Handles continuous action spaces.
- Sample efficiency
- Training stability.
- Centralized training, decentralized execution.

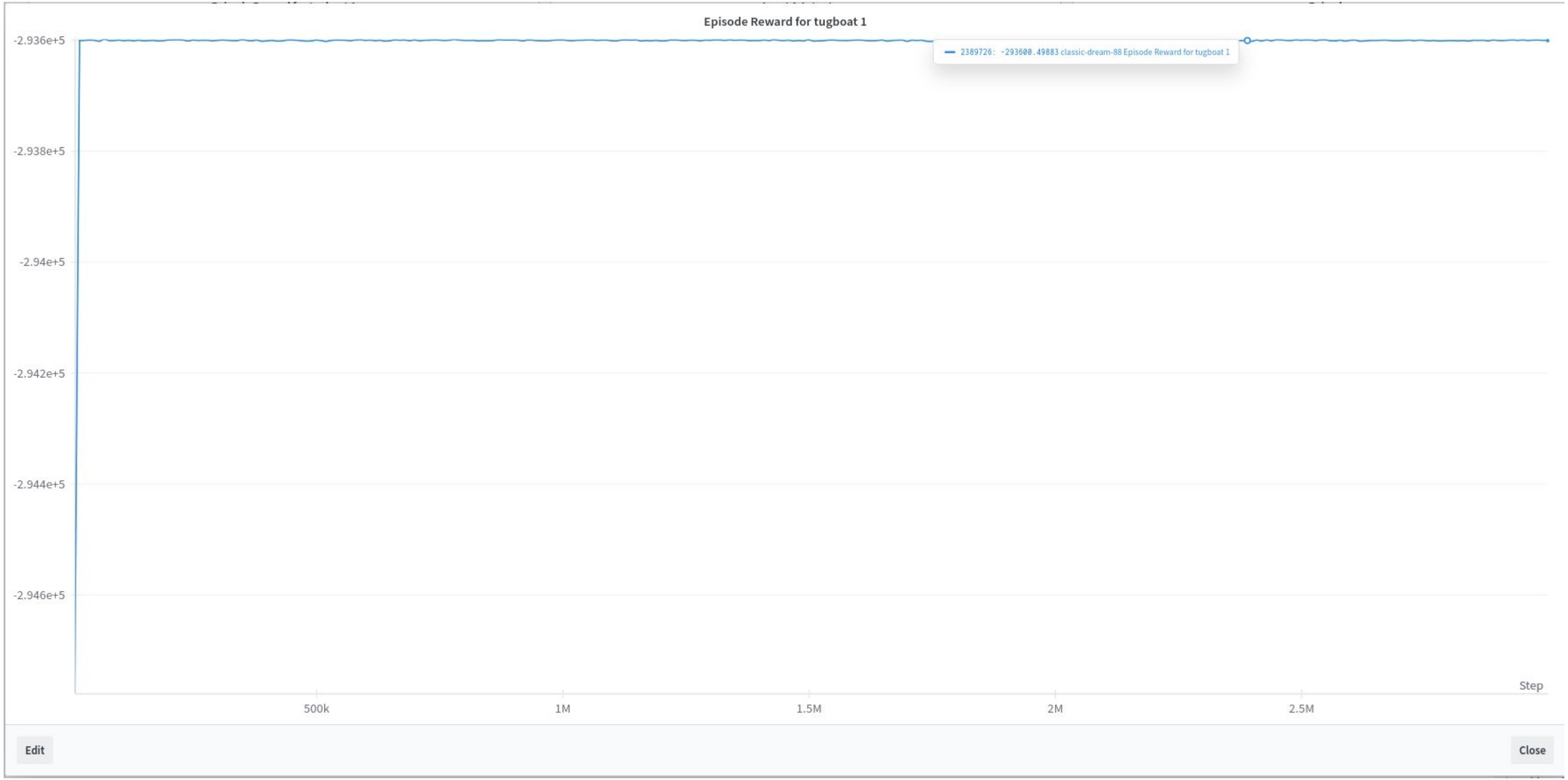


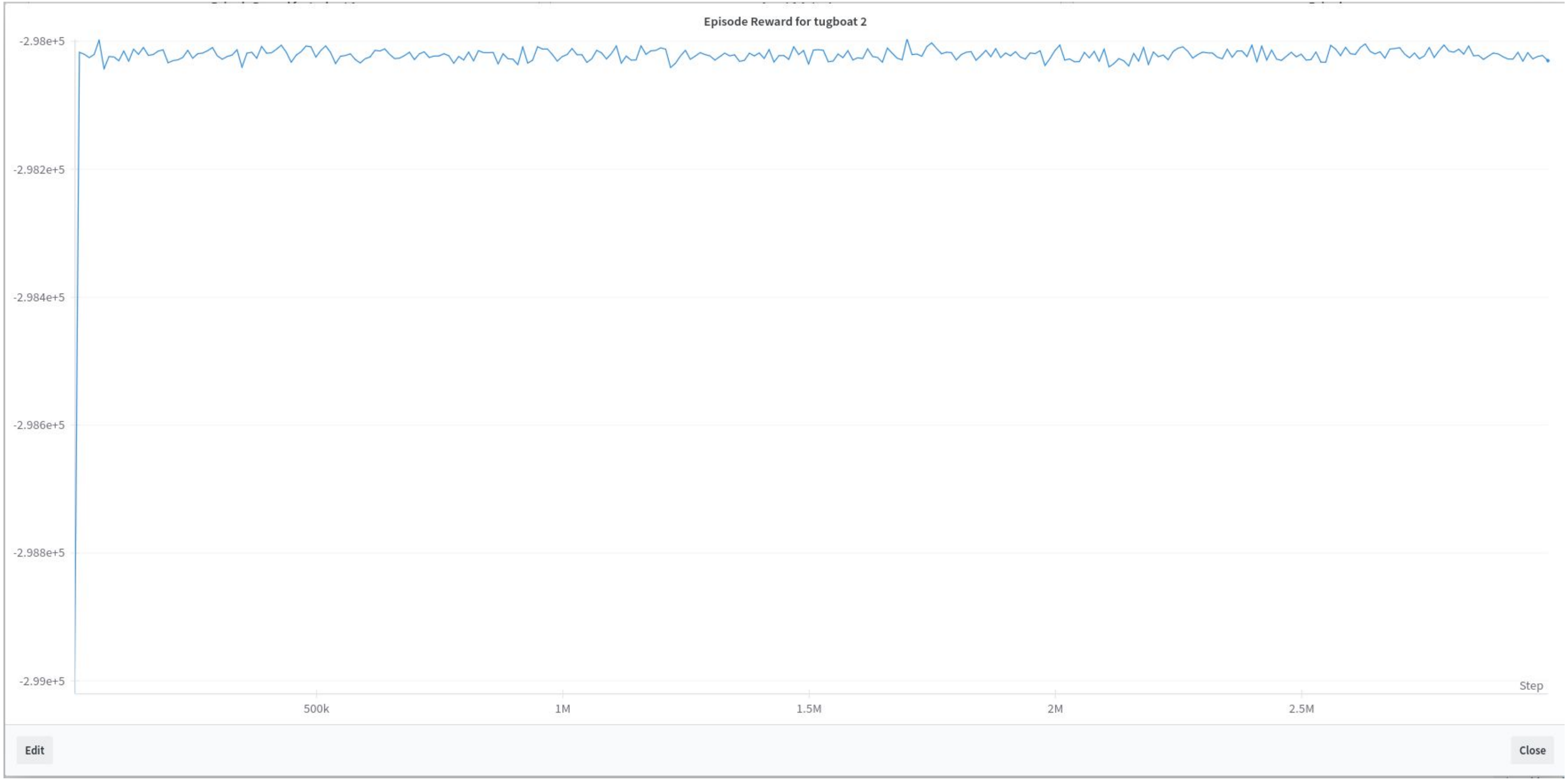
Source: https://arxiv.org/pdf/1706.02275

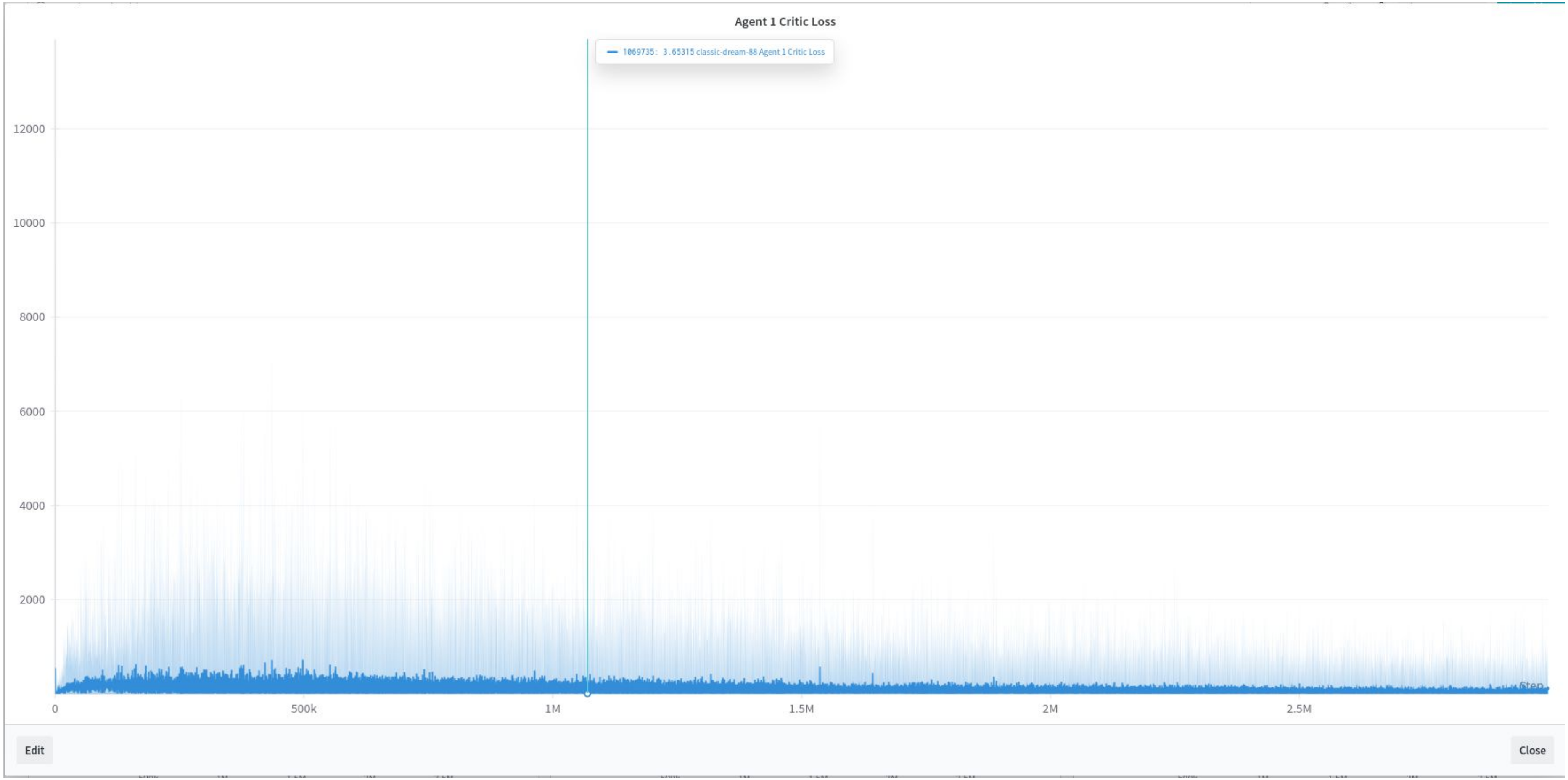
Parameters

- Gamma = 0.99
- Adam optimizer, learning rate = 0.003.
- Tau = 0.01
- Training episodes = 300, each with 2500 steps.

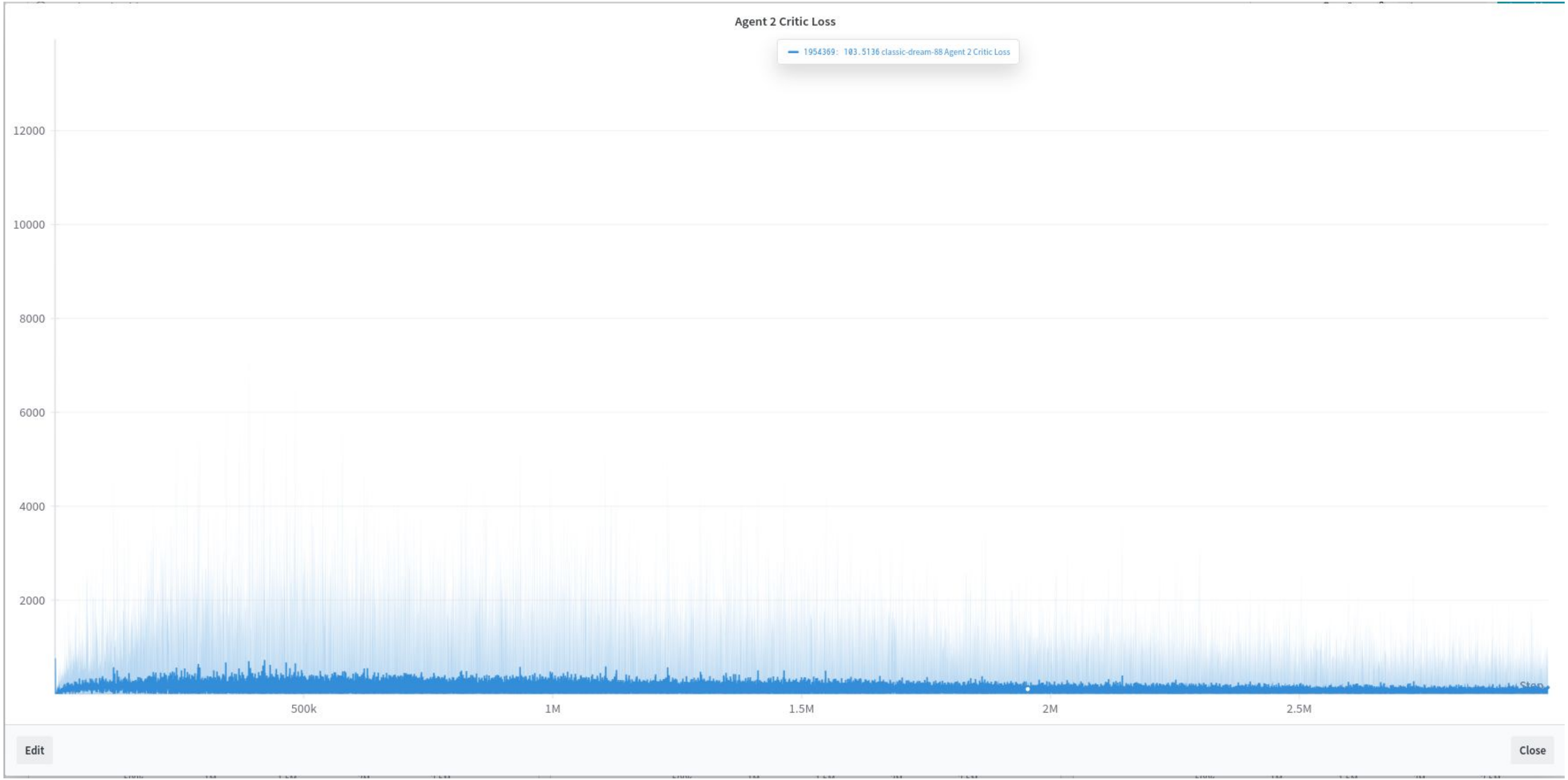
Results

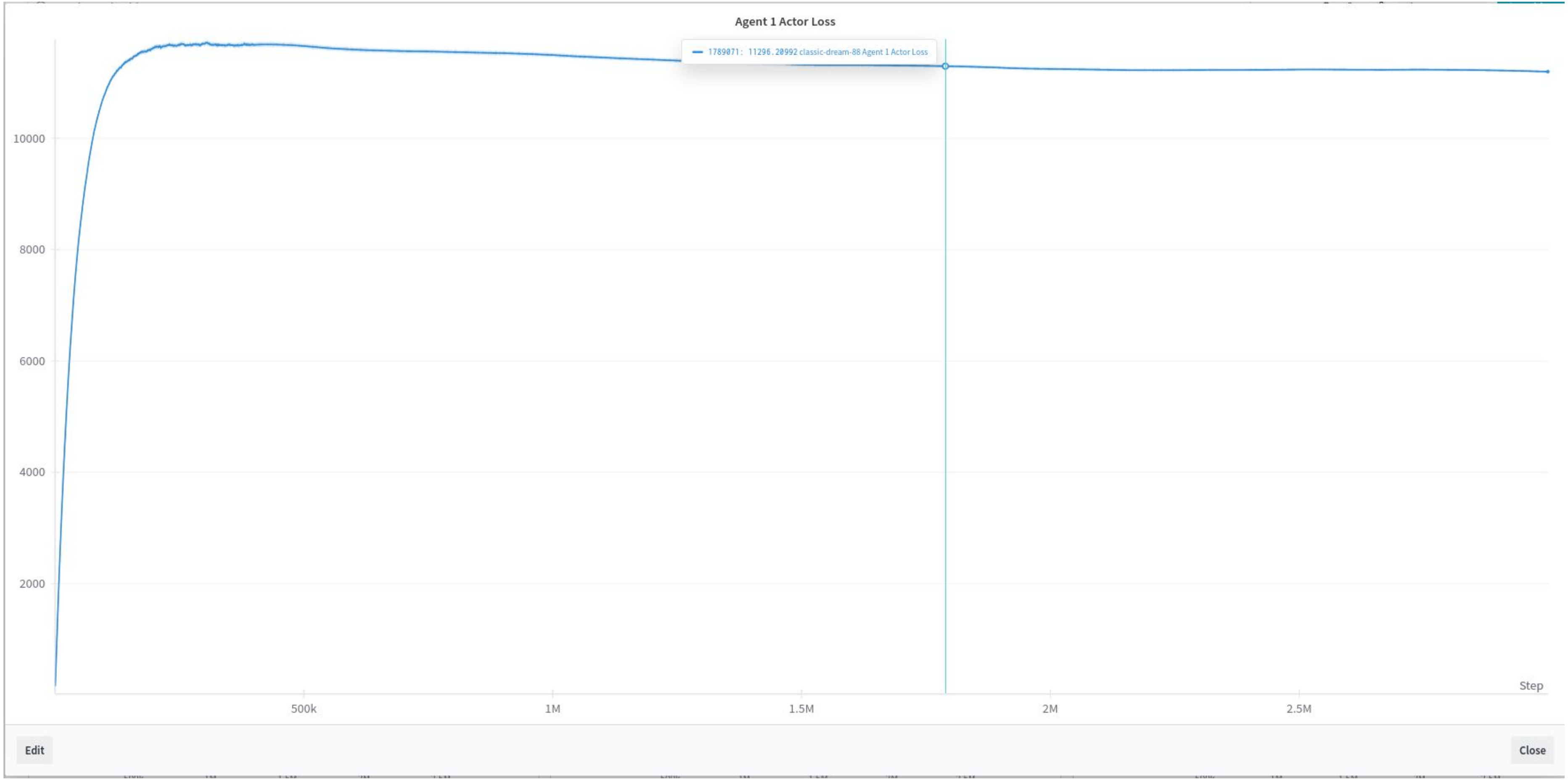

Episode Reward for tugboat 1
-2.936e+5
-2.938e+5
-2.94e+5
-2.942e+5
-2.944e+5
-2.946e+5
2389726: -293600.49883 classic-dream-88 Episode Reward for tugboat 1
500k
1M
1.5M
2M
2.5M
Step
Edit
Close

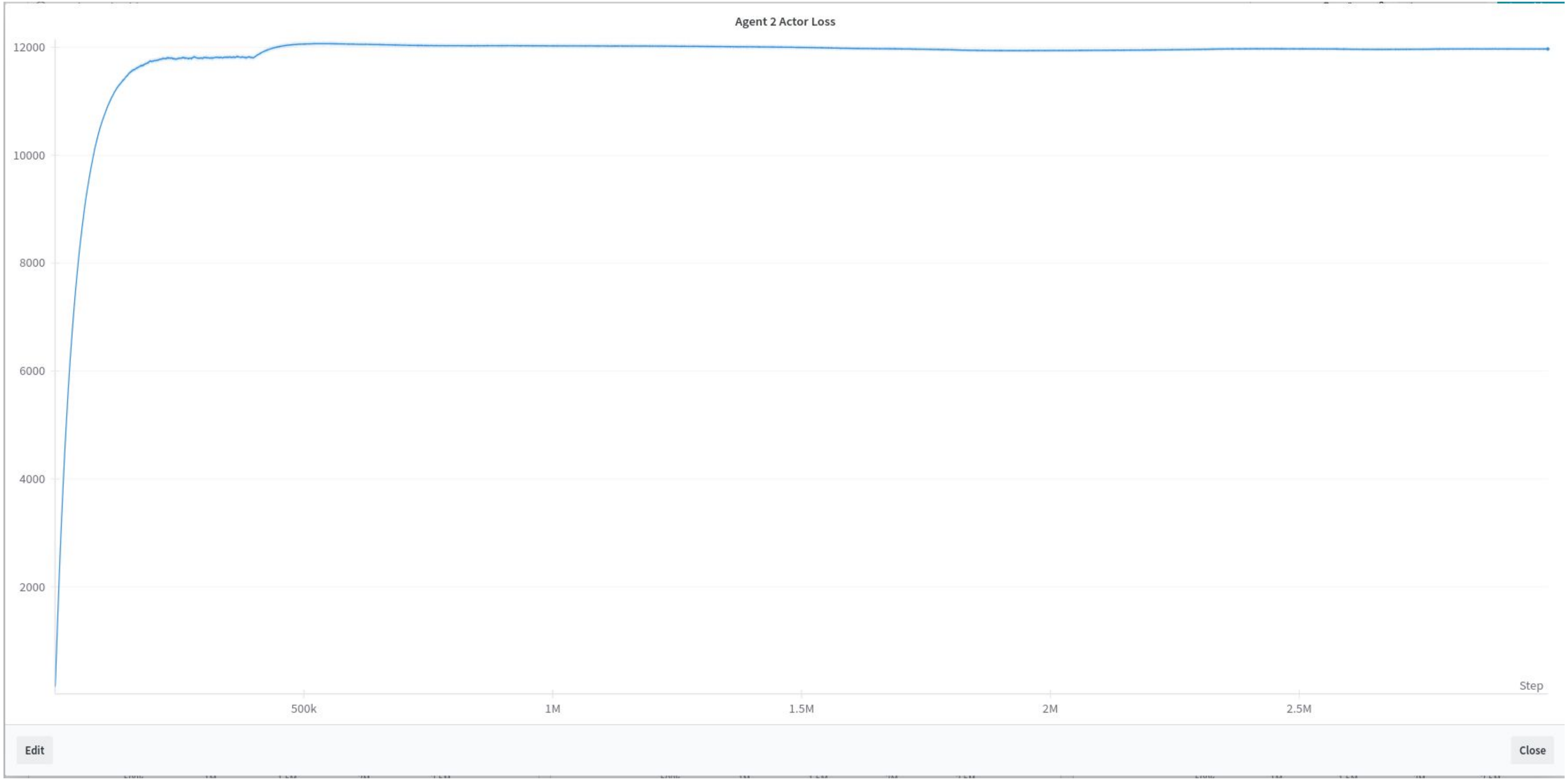

Episode Reward for tugboat 2
-2.98e+5
-2.982e+5
-2.984e+5
-2.986e+5
-2.988e+5
-2.99e+5
500k
1M
1.5M
2M
2.5M
Step
Edit
Close



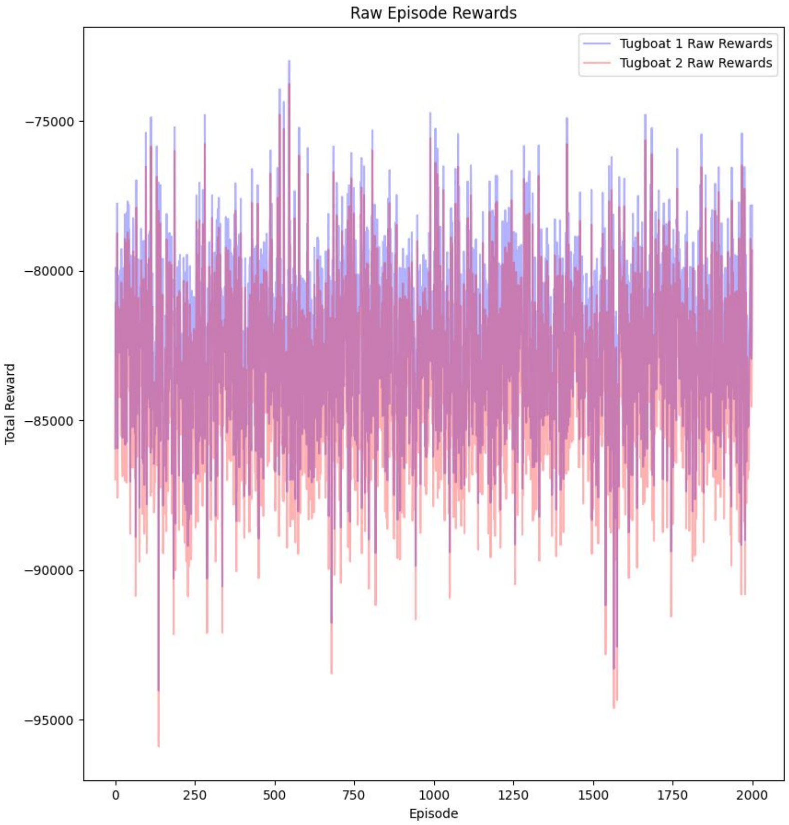
Agent 1 Critic Loss
1069735: 3.65315 classic-dream-88 Agent 1 Critic Loss
12000
10000
8000
6000
4000
2000
0
500k
1M
1.5M
2M
2.5M
Step
Edit
Close


Agent 2 Critic Loss
1954369: 103.5136 classic-dream-88 Agent 2 Critic Loss
12000
10000
8000
6000
4000
2000
Step
500k
1M
1.5M
2M
2.5M
Edit
Close


Agent 1 Actor Loss
1789071: 11296.20992 classic-dream-88 Agent 1 Actor Loss
10000
8000
6000
4000
2000
Step
500k
1M
1.5M
2M
2.5M
Edit
Close


Agent 2 Actor Loss
12000
10000
8000
6000
4000
2000
500k
1M
1.5M
2M
2.5M
Step
Edit
Close

For random steps



Future Work

- Improve reward structure.
- Random spawn in the grid.
- Add dynamic obstacles.
- Managing traffic of multiple tugboat - ship combinations.