Competitive Pricing in Airline Revenue Management with Multi-agent Reinforcement Learning

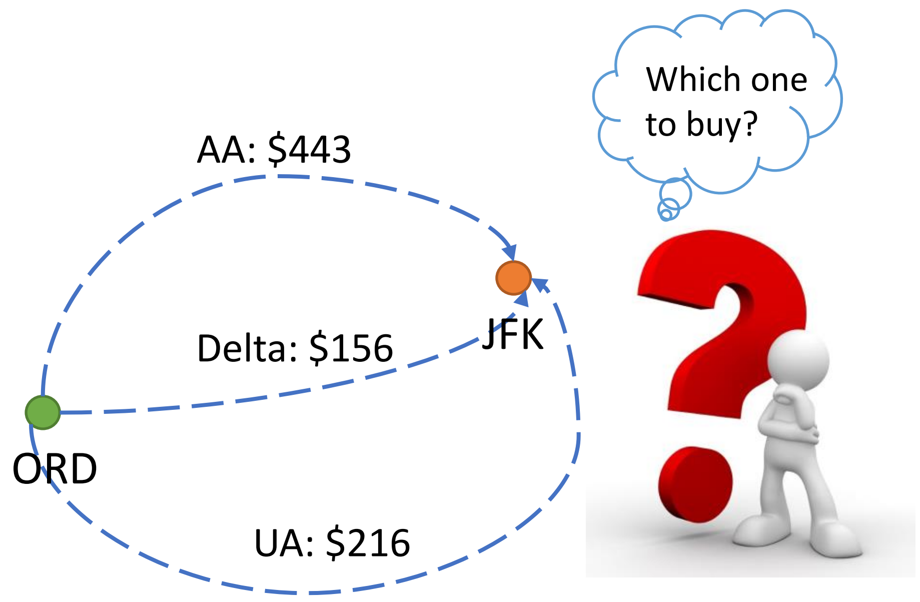
Shulu Chen
George Washington University
INFORMS Annual Meeting
2021 Anaheim, CA

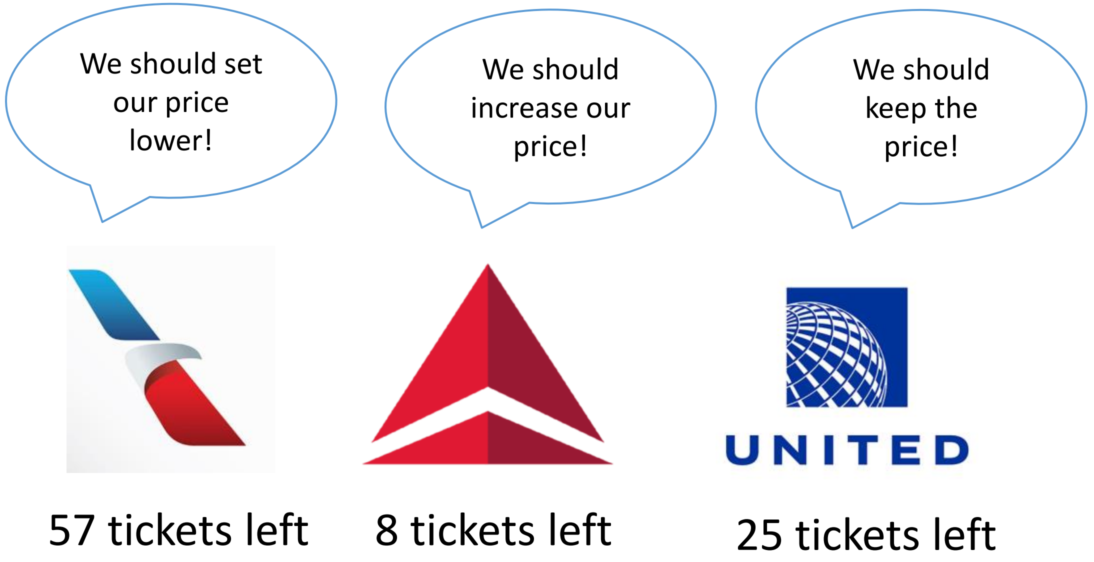
Syed A.M. Shihab
Kent State University

Peng Wei
George Washington University

Motivation

- Traditional method for pricing: based on history data and model, only consider self inventory.
- **New method for pricing**: multi-agent reinforcement learning (MARL), observe competitors' information, make smarter decisions.





Related Work and Contributions

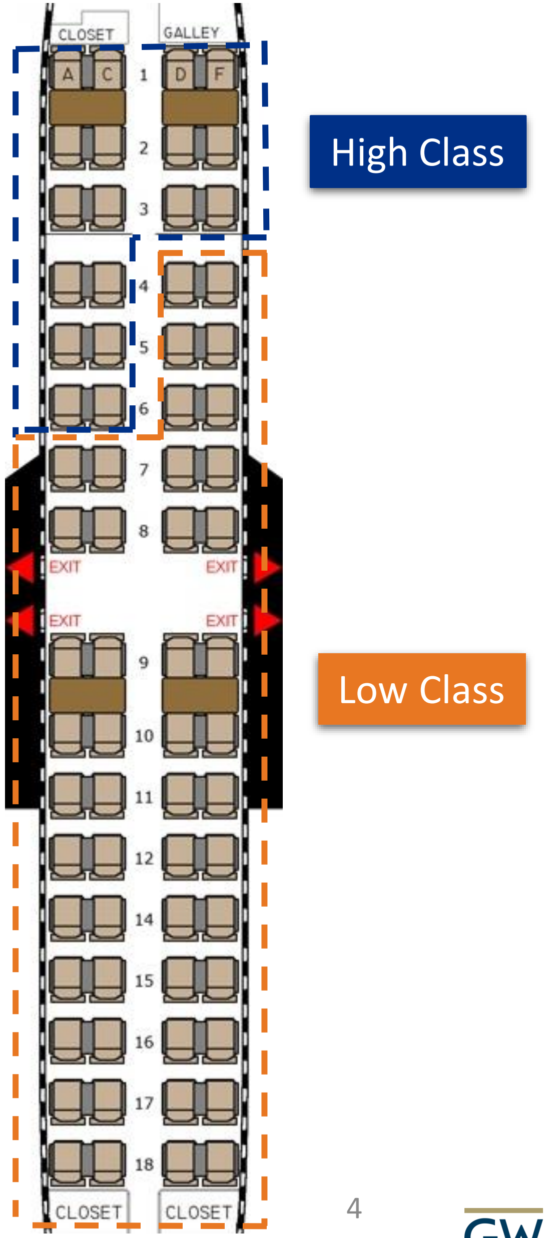
- **Related work:**
  - Belobaba 1987[1] (EMSR)
  - Bondoux et al. 2020 [2] (DQN)
  - Shihab & Wei 2021 [3] (DQN, different formulation)

- **Our contributions:**
  - First work of applying MARL to airline revenue management.
  - Proposed the competitive airline revenue management (CARM) model.
  - High dimensional action space with price points, potentially feasible for continuous dynamic pricing.

[1] Belobaba, P. (1987). *Air travel demand and airline seat inventory management* (Doctoral dissertation, Massachusetts Institute of Technology).
[2] Bondoux, N., Nguyen, A. Q., Fiig, T., & Acuna-Agost, R. (2020). Reinforcement learning applied to airline revenue management. *Journal of Revenue and Pricing Management*, *19*(5), 332-348.
[3] Shihab, S. A., & Wei, P. (2021). A deep reinforcement learning approach to seat inventory control for airline revenue management. *Journal of Revenue and Pricing Management*, 1-17.

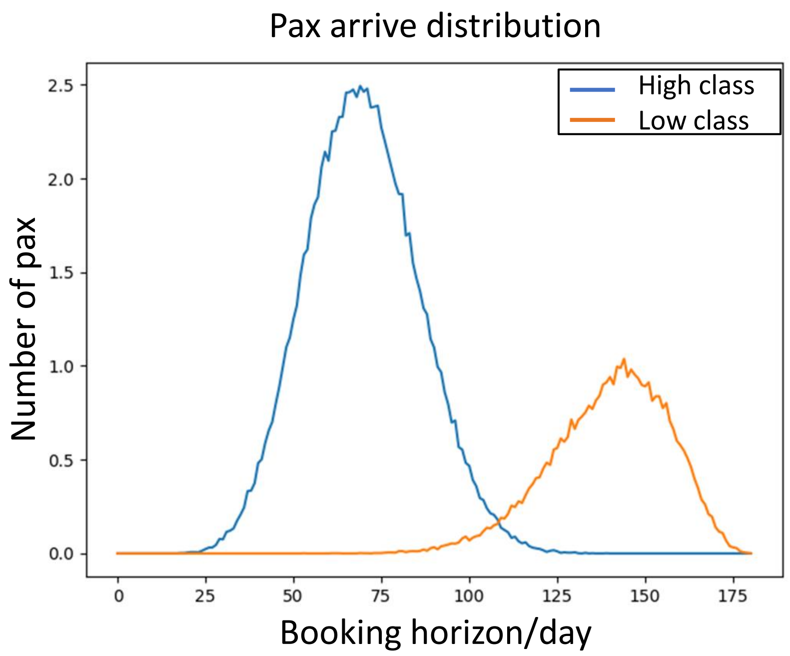
Problem Description & Assumption

- Single flight leg.
- Two agents compete.
- Two different fare classes (low & high) share the cabin seat inventory.
- Each day, each agent offers one price for low fare class and one for high fare class.
- When one agent sells its all tickets, the other monopolizes the market.
- Passengers (pax) are categorized into two corresponding classes.
- Pax only consider the lower price, and 50% probability to consider either flight when given same price.



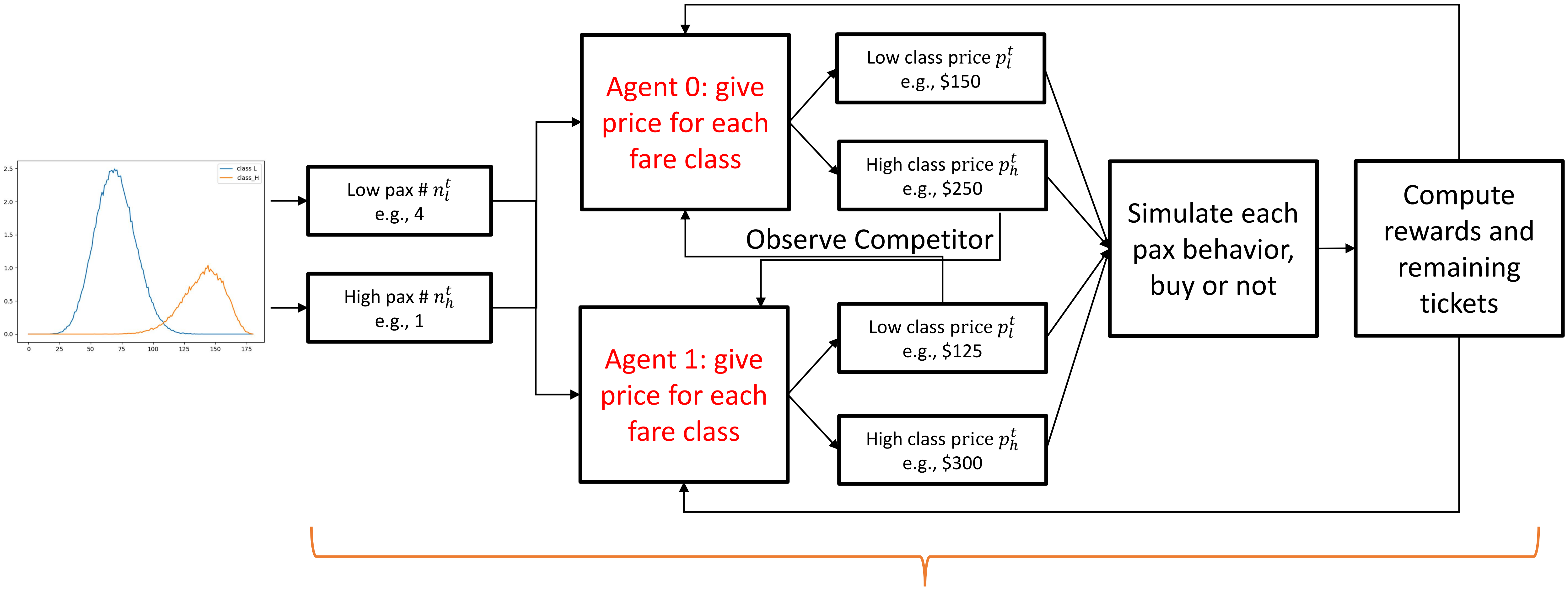
Problem Formulation

- **MARL formulation:**
  - Setting: multi-agent deep deterministic policy gradient (MADDPG) [4].
  - Time step: each single day during the booking horizon.
  - RL ingredients:
    - ➢ $\boldsymbol{State}$: $< b_h, b_l, remaining\_seats, t >$
    - ➢ $\boldsymbol{Action}$: $price\ for\ each\ fare\ class < p_h, p_l >$
    - ➢ $\boldsymbol{Reward}$: $revenue\ earned\ at\ time\ t$: $R^t = p_h^t \times b_h^t + p_l^t \times b_l^t$
- **Passenger behavior simulator:**
  - Pax arrive following the β distribution.
  - Willingness to pay (WTP):
    - Given a price and time, probability of passenger to buy that ticket



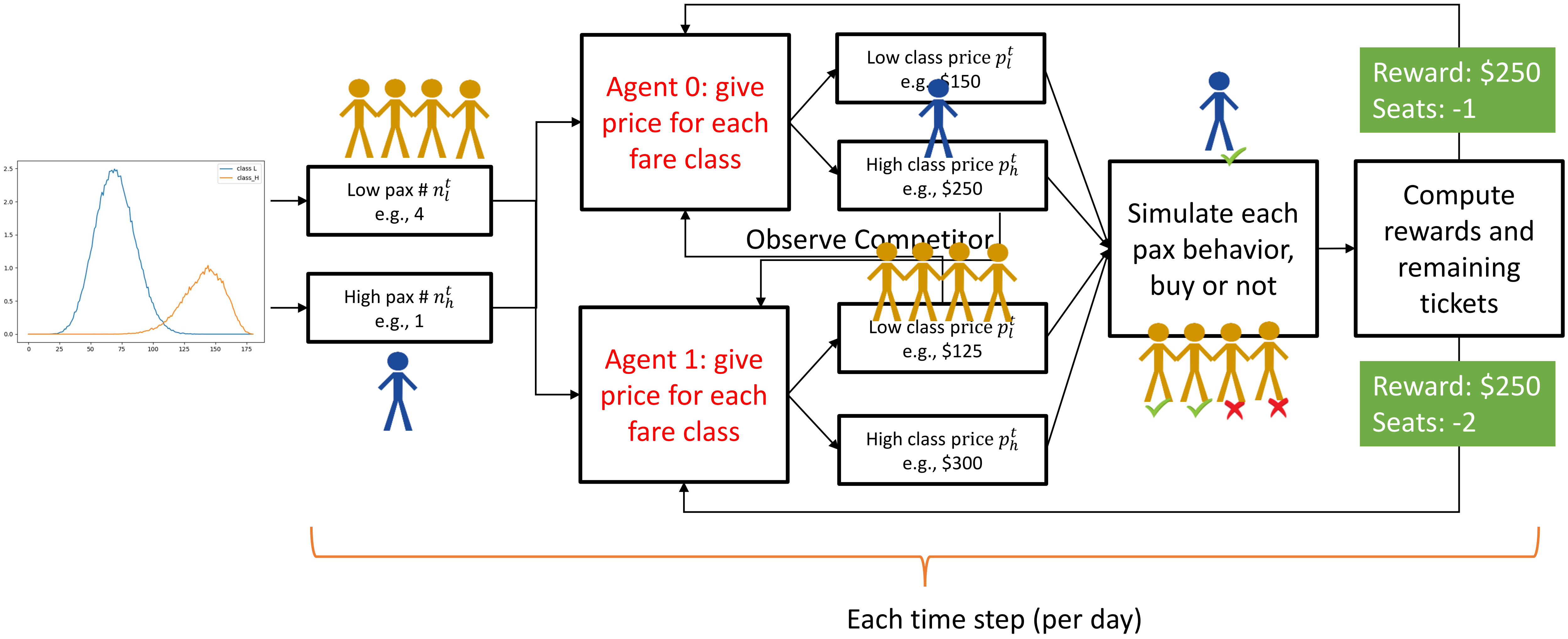
[4] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *arXiv preprint arXiv:1706.02275*

Competitive Airline Revenue Management (CARM)



Each time step (per day)

Competitive Airline Revenue Management (CARM)



Experiments

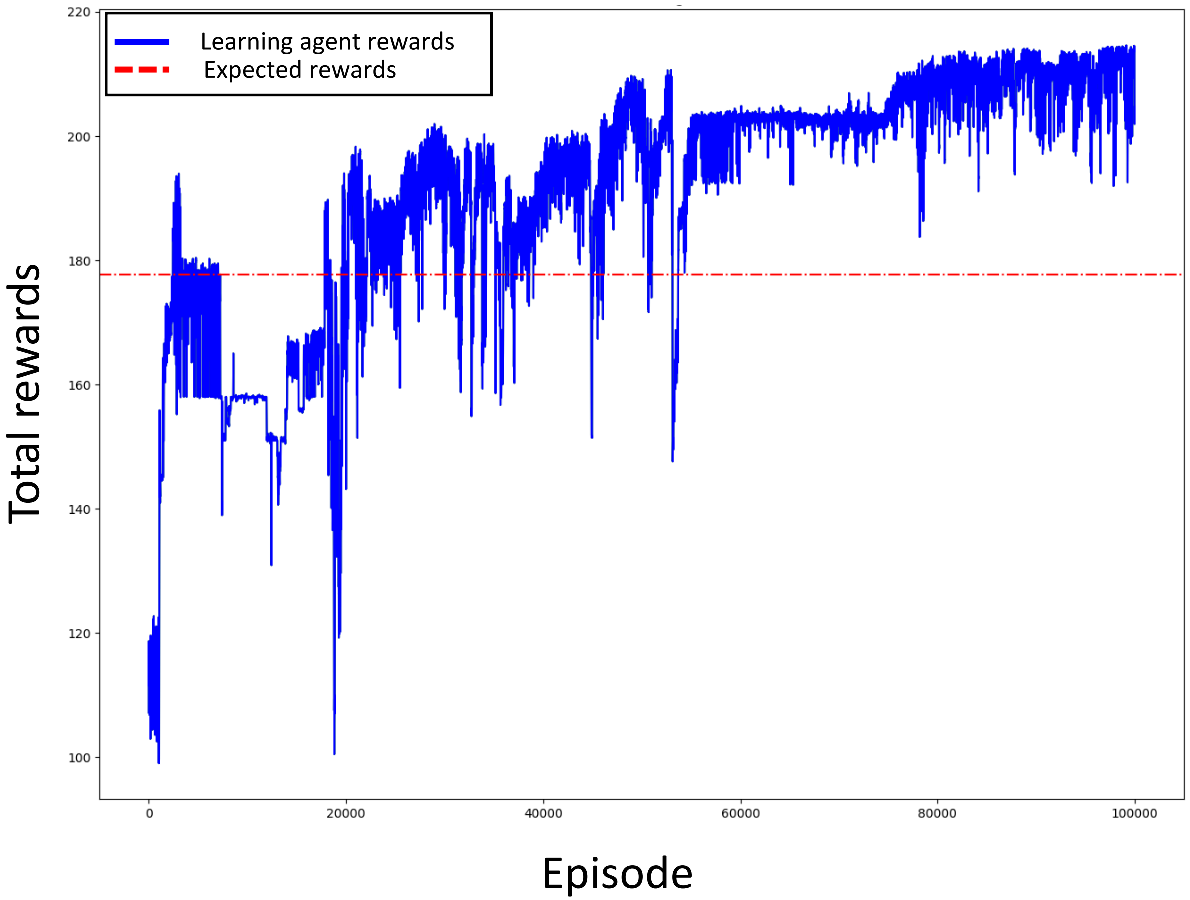
1. Single learning agent: learning agent vs. pre-defined policy agent
2. Single learning agent: learning agent vs. price matching agent
3. Multiple learning agents: two learning agents compete, no WTP
4. Multiple learning agents: cooperative or competitive?

Experiment 1 -
Learning Agent vs. Pre-defined Policy Agent

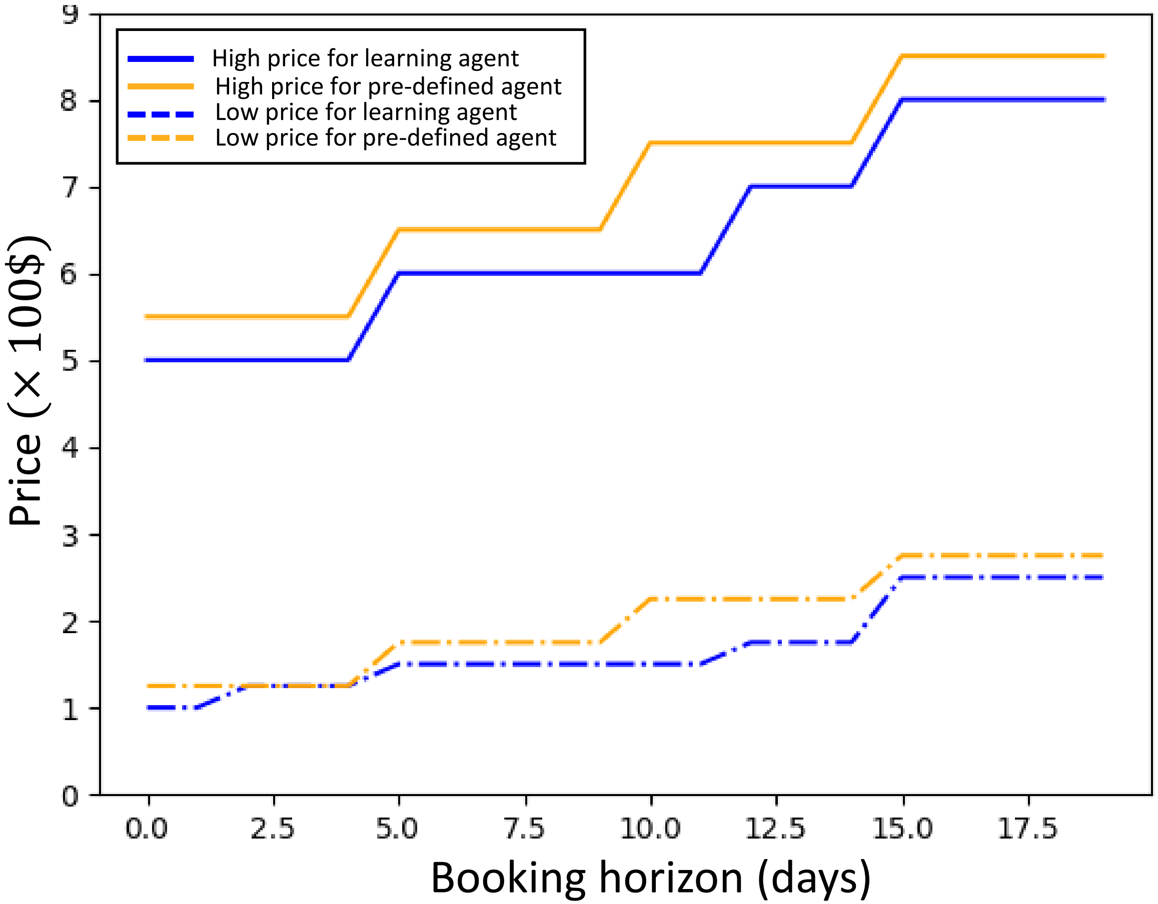
- **Summary:**
  - One learning agent uses deep deterministic policy gradient (DDPG) algorithm, the other agent maintains a pre-defined pricing policy (step up over time).
- **Setting:**
  - Booking horizon: 20 days
  - Seats per flight: 40
  - Low class pax: 40; high class pax: 20
  - Total seats (80)> total pax (60)
  - Action space:
    - ➢ High class price points: [$500, $550, $600, $650, $700, $750, $800, $850]
    - ➢ Low class price points: [$100, $125, $150, $175, $200, $225, $250, $275]

Experiment 1 - Result

Learning curve for the learning agent



Final policy of both agents

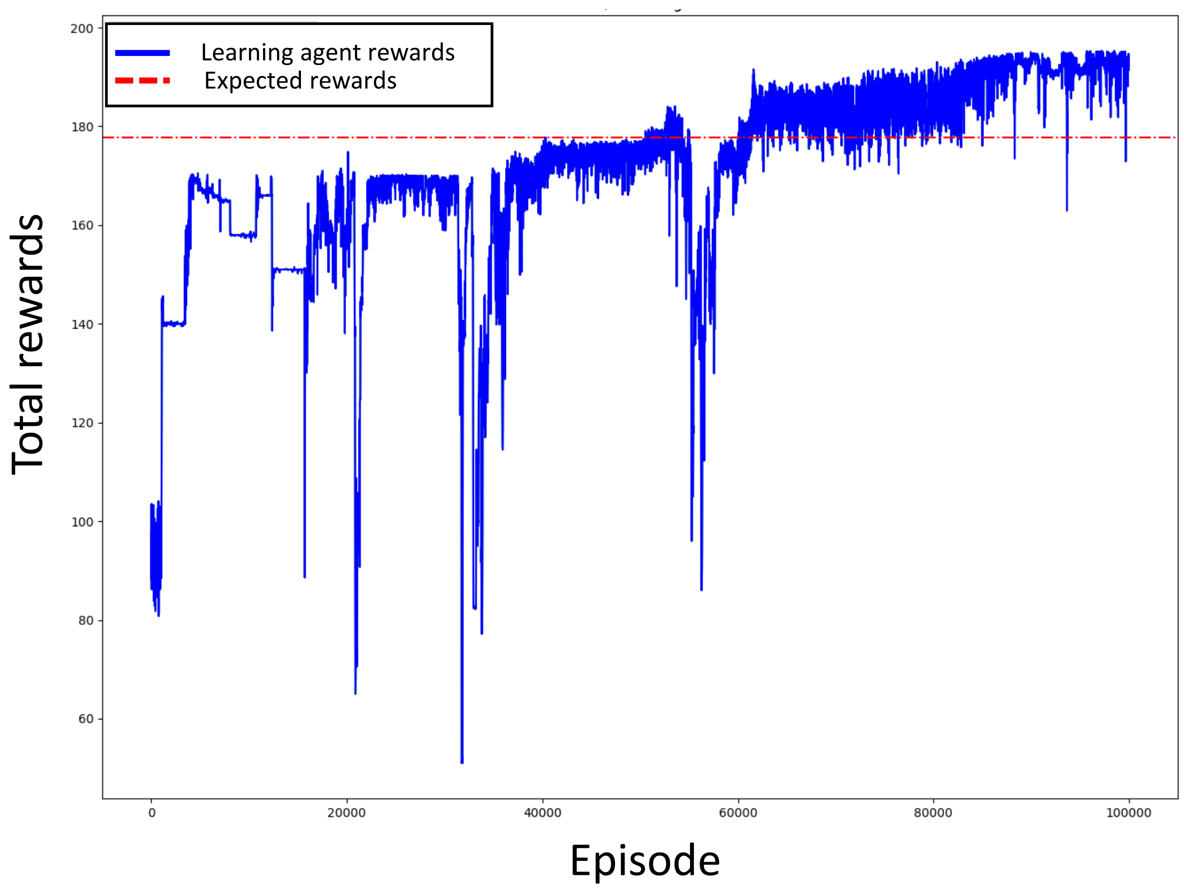


Experiment 2 –
Learning Agent vs. Price Matching Agent

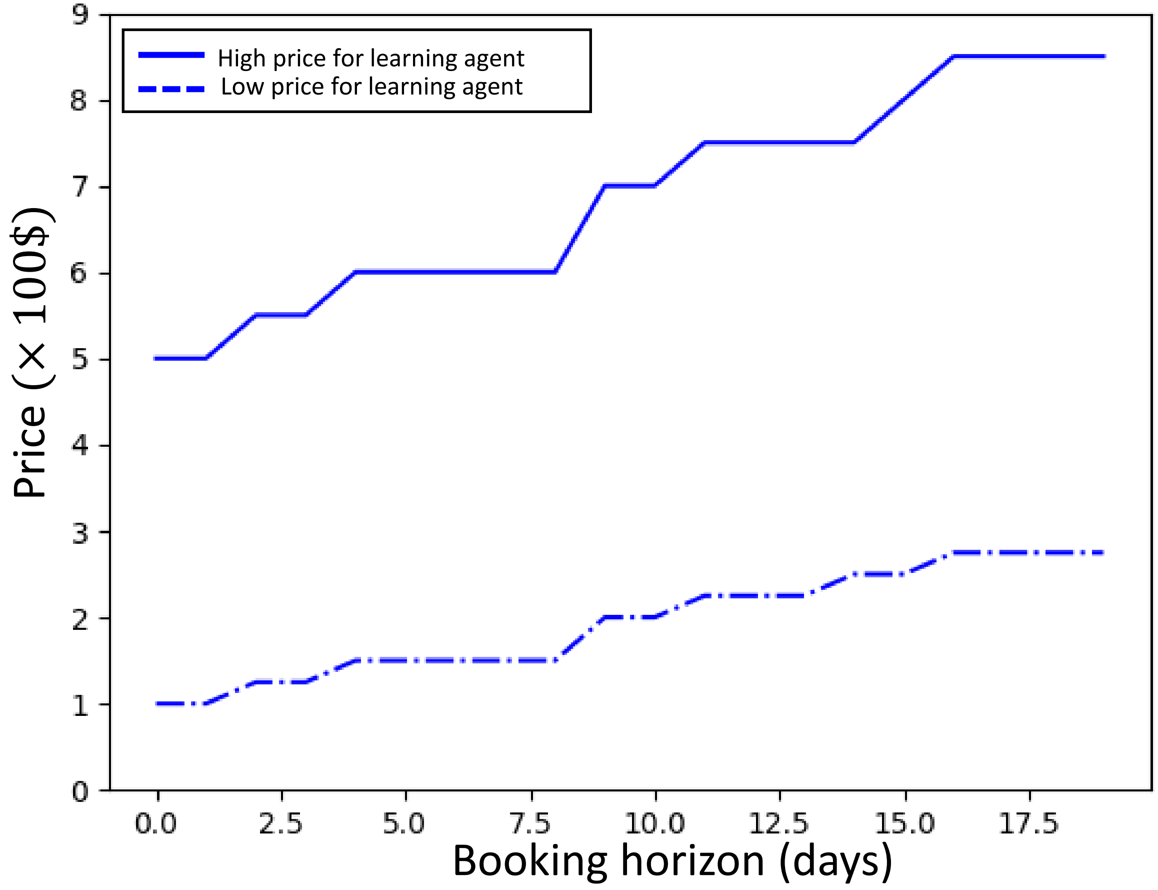
- **Summary:**
  - One learning agent uses DDPG algorithm, and the other agent observes and chooses the same price as the learning one.
- **Setting:**
  - Booking horizon: 20 days
  - Seats per flight: 40
  - Low class pax: 40; high class pax: 20
  - Total seats (80)$>$ total pax (60)
  - Action space:
    - ➢ High class price points: [$500, $550, $600, $650, $700, $750, $800, $850]
    - ➢ Low class price points: [$100, $125, $150, $175, $200, $225, $250, $275]

Experiment 2 - Result

Learning curve for the learning agent



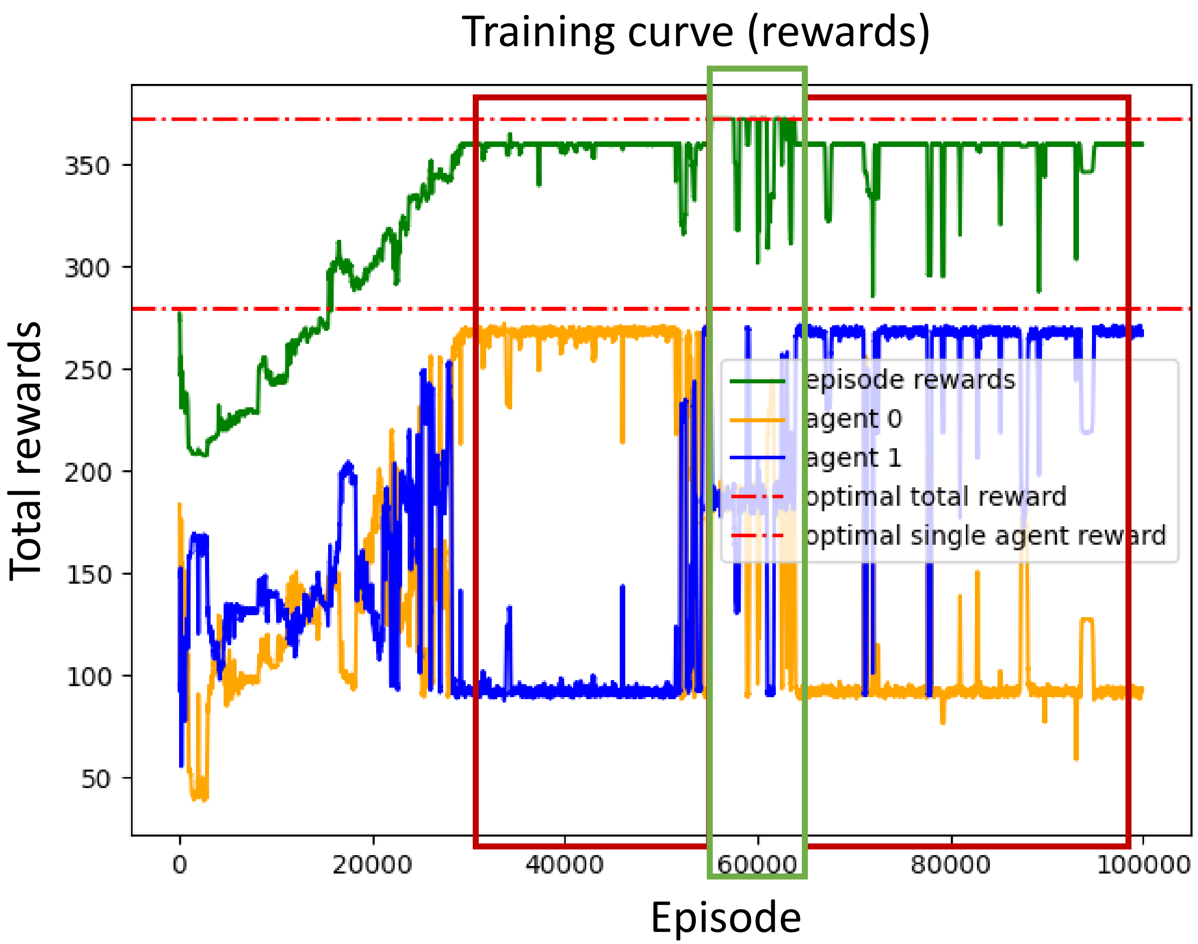
Final policy of the agents



Experiment 3 -
Two Learning Agents Compete, No WTP

- **Summary:**
  - Both agents are learning by using MADDPG algorithm
  - No WTP - reduce the uncertainty
- **Setting:**
  - Booking horizon: 20 days
  - Seats per flight: 40
  - Low class pax: 40; high class pax: 20
  - Total seats (80)> total pax (60)
  - Action space:
    - ➢ High class price points: [$500, $550, $600, $650, $700, $750, $800, $850]
    - ➢ Low class price points: [$100, $125, $150, $175, $200, $225, $250, $275]

Experiment 3 - Result



Scenario 1: One agent achieves much better revenue than the other, and the total revenue is suboptimal.

Scenario 2: Agents have same policy, each agent achieves lower revenue, but their total revenue is optimal.

Experiment 4 –
Cooperative or Competitive?

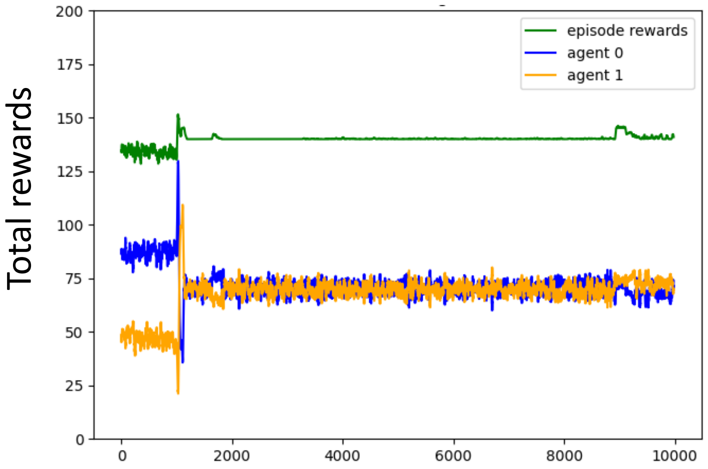
- Apply WTP in the environment.
- Target of the agent: total rewards (cooperative) vs. self rewards (competitive).
- **Prisoner's dilemma** in airline revenue management:
  - If the game is played exactly N times and both players know this, the only possible Nash equilibrium is to always defect (low price).
  - More supply than demand leads to fierce competition.

Agent 0 / Agent 1	Lower price	Higher price
Lower price	L/2,L/2	0,L
Higher price	L,0	H/2,H/2

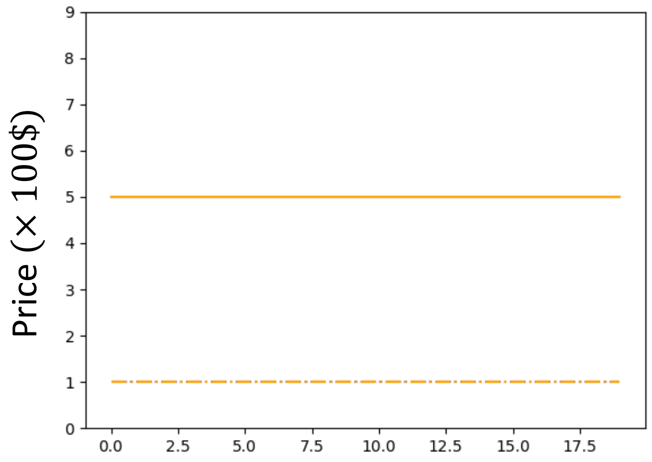
Experiment 4 - Result

Competitive case:
- Agents aim to increase their self rewards
- Both agents choose the selfish policy

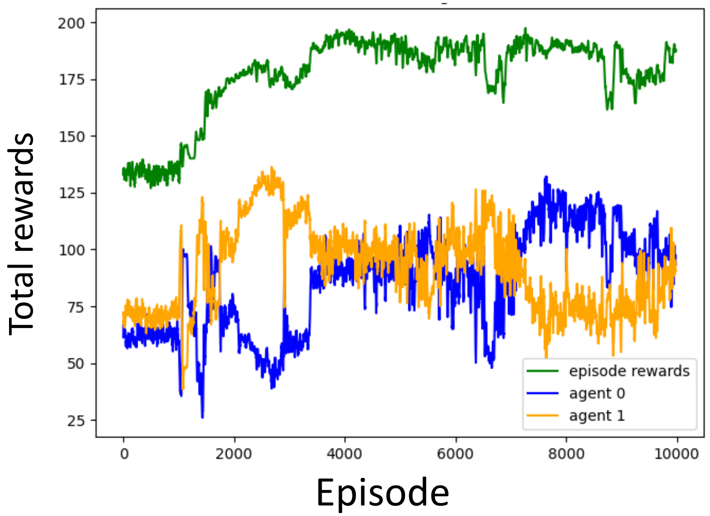
Learning curve for the learning agent

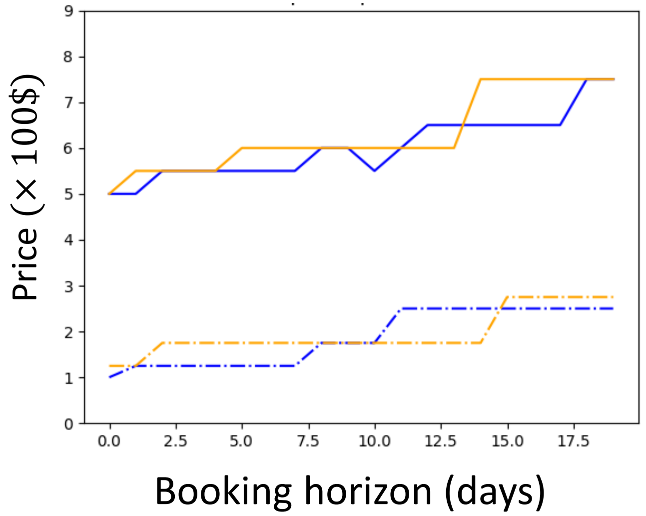


Final policy of the agents



Cooperative case:
- Agents aim to increase their total rewards
- Pricing policy is much better





Challenges and Future Work

- **Learning instability during training:**
  - Too much uncertainty (WTP, pax arrival distribution).
  - The dynamic of the competitor's action/policy.
- **Game theory issue:**
  - Which case is more realistic? Cooperative or competitive.
  - Investigate the game mechanism to design a better agent.
- **Partial observation during execution:**
  - An agent needs to predict its competitor's action.

Thanks for your attention.
Comments & questions?

Shulu Chen
George Washington University
shulu@gwu.edu