Learning-based control of AMoD in competitive environments

Joachim Andreasen[1], Frederik Sørensen[1], Asger Tang[1], Carolin Schmidt[1],
Daniele Gammelli[2], Francisco Pereira[1] and Filipe Rodrigues[1,*]
[1]Technical University of Denmark, [2]Stanford University
[*]*Corresponding author*

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 12, 2025

Keywords: AMoD, reinforcement learning, multi-agent systems, dynamic pricing, rebalancing

1 Introduction

Autonomous Mobility-on-Demand (AMoD), where customers request trips from their origin and are assigned an autonomous vehicle from a fleet to take them to their destination, has the potential to play a crucial role in future sustainable transport. AMoD offers passengers a personalized mobility service while eliminating the maintenance and parking costs associated with owning a private vehicle. Due to its high flexibility, AMoD is gaining enormous popularity around the world. However, a core challenge for the AMoD paradigm lies in the spatio-temporal nature of urban mobility, where trip origins and destinations are asymmetrically distributed (e.g., commuting downtown in the morning and vice-versa in the evening), making the overall system imbalanced and sensitive to disturbances. Operators can try to overcome this issue by manually rebalancing vehicles to anticipate future demand, or by developing dynamic pricing strategies to (dis)encourage trips between particular origin-destination pairs to promote a more desirable distribution of the vehicle supply. However, this presents a challenging control problem.

While traditionally, the problems of vehicle rebalancing and dynamic pricing have been tackled either through the lenses of heuristics and optimization (Zardini *et al.*, 2022), the most recent literature focuses on learning-based approaches, mainly due to their scalability and ability to handle dynamic stochastic environments - see Qin *et al.* (2022) for an extended survey. However, existing approaches consider a single-operator scenario. In modern liberal economies, this assumption is highly unrealistic. Therefore, in this work, we consider a multi-operator scenario, which we formulate as a multi-agent reinforcement learning (RL) problem, where each agent centrally controls the vehicles in its own fleet without having knowledge about the competitor's states and actions. To the best of our knowledge, this is the first work to demonstrate that learning-based approaches are robust to the added stochasticity in the environment, being able to rebalance their fleet and dynamically set prices accounting for the interplay with the competitors and to empirically show that the learned policies converge to an equilibrium. Furthermore, we leverage this multi-agent RL setup to empirically study the market dynamics and the achieved equilibrium (e.g., regarding fleet size and the introduction of new competitors).

2 Methodology

Our starting point is the bi-level framework proposed by Gammelli *et al.* (2022) for AMoD rebalancing, where a soft actor-critic (SAC) RL agent decides the desired share of vehicles in each

area (higher-level action - parameterized by a Dirichlet distribution). Then, solving a minimum rebalancing cost flow problem determines the actual vehicle flows between areas (lower-level action). Both the actor and critic use a graph neural network (GNN) architecture, where each area corresponds to a node in the graph, in order to capture the spatial relationships between the states of the different areas. The demand in each area is simulated by a Poisson process with the arrival rate based on real-world data. We extend this centralized control framework to a multi-agent setup by assuming identical (but independent) RL architectures for both agents, without knowledge sharing about states and actions.

The reward of each RL agent corresponds to the operator's profit, defined as the difference between revenue from trips and rebalancing costs. To model user behavior, we incorporate the possibility of trip cancellations, where users may reject trips based on their price. We determine the probability of cancellation using a sigmoidal curve, specifically the Hill equation, which is shaped based on the population-level value-of-time (VOT). Its input is a value-of-a-ride (VOR) variable, representing the ratio of trip price to trip length, with trip length derived from data. We also simulate users' choice of AMoD operators in proportion to the VOR. Finally, we introduce dynamic pricing into the framework by extending the GNN output to include a Gaussian random variable that controls pricing. Concretely, we determine the trip price by a (linear) regression model based on travel time, with the GNN output acting as the regression coefficient.

3 Preliminary experiments

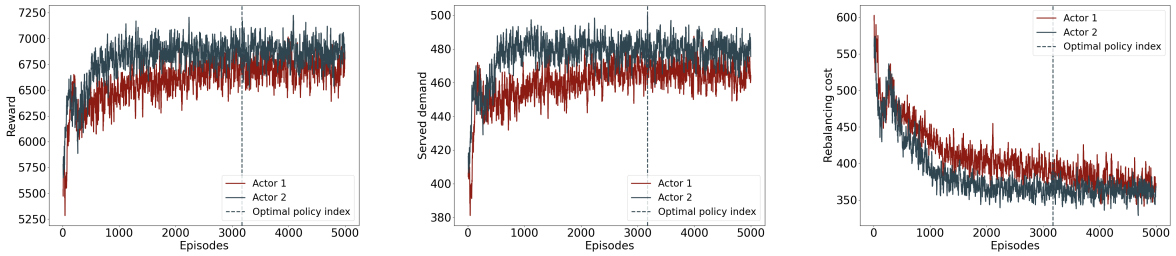
All our preliminary experiments were performed in the New York scenario provided by Gammelli *et al.* (2022). We will extend to other scenarios as part of our future work.

From single-agent to multi-agent setup. We begin by considering the transition from a single-operator scenario to one with two, where each operator has a fleet size of exactly half the fleet size of the single-operator version (374 cars). As the results in Table 1 show, the combined (system) reward of the two agents in the multi-agent setup is slightly lower compared to the single-agent setup, indicating that the shift from a monopolistic market to a competitive market has an effect. This could be due to the increased stochasticity of the environment, which adds complexity for the agents operating in it. They must now account not only for the effects of their own actions but also for the unpredictable behavior of their competitors. Despite this, the system reward remains close to that of the single-agent case (within one std. dev.), indicating that the RL approach is robust to the added stochasticity of the environment. Interestingly, while the system converges to an equilibrium, the two agents converge to different behavior policies, as seen in Figure 1. Agent 1 ends up serving more demand, leading to a higher profit, suggesting distinct strategies by each operator.

Model	**Reward**	**Served demand**	**Cancellations**
Single-agent SAC without cancel.	14471 ± 317	1020 ± 20	-
Multi-agent SAC without cancel.	14184 ± 298	985 ± 26	-
Multi-agent SAC with cancel.	12483 ± 406	978 ± 27	157 ± 8

Table 1 – *The total (system) reward of training SAC in single-agent and multi-agent setups.*

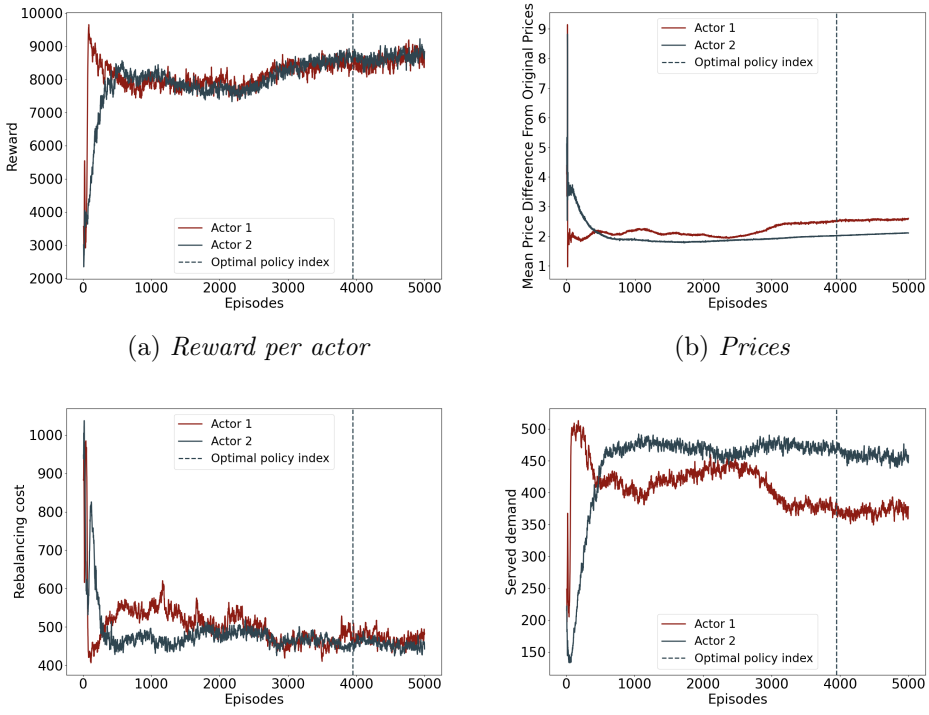
Adding cancellations and dynamic pricing. We now introduce the possibility of cancellations, i.e. users rejecting trips according to their price. As expected, the results in Table 1 show a lower overall system reward in the environment with cancellations, despite the served demand remaining relatively stable. These findings motivate the expansion of the agents' control capabilities to handle both rebalancing and dynamic pricing simultaneously. By doing so, the agent can increase profit in high-demand regions, better mitigate cancellations, and proactively manage manual vehicle rebalancing when necessary. By allowing for dynamic pricing, the system achieves a profit approximately. 40% higher (17445 ± 319) compared to the fixed-price



(a) *Reward for each agent/actor* (b) *Served demand* (c) *Rebalancing cost*

Figure 1 – *Reward, served demand, and rebalancing cost for each actor in the multi-agent setup.*

scenario, by being able to exploit the customers' willingness to pay. Figure 2 shows that again, the system converges to an equilibrium, while the agents learn different behavior policies. The average prices per minute of actors 1 and 2 are \$2.49 and \$2.03, respectively, both higher than the original average price from the data (\$1.54). One actor quickly learns to adopt an undercutting strategy, consistently setting lower prices than the competitor to capture more demand and reduce cancellations, albeit at the cost of lower revenue per trip. Notably, despite these different strategies, both actors achieve comparable profits.



(a) *Reward per actor* (b) *Prices*

(c) *Rebalancing cost* (d) *Served demand*

Figure 2 – *Average rewards, prices, rebalancing costs, and served demand over the training.*

Varying fleet sizes in a multi-agent setup. A critical factor contributing to the efficiency and profit of an AMoD system is fleet size. Therefore, we explore its impact on the learned behavior policies. As shown in Table 2, a decrease in fleet size leads to an increase in price. This outcome is expected, as the agent raises prices to shift some of the unmet demand —caused by limited capacity— into demand lost due to cancellations, thus balancing capacity constraints with profitability. Interestingly, we observe a decrease in profit when the fleet size exceeds 374 cars, suggesting that the additional vehicles merely introduce higher operational costs. For reference,

we perform a similar sensitivity analysis on the fleet size in the monopolistic (single-agent) setup. The results in Table 2 show a similar relationship between fleet size and the average price set by the policy. However, a key insight is that the multi-agent scenario results in lower overall prices, making the system more favorable for users. This highlights the competitive nature of the multi-operator scenario where each actor reduced prices to capture market share. The combination of lower prices and reduced profits in the multi-agent setup clearly illustrates the difference between monopolistic and competitive markets. In the monopolistic setting, prices are mainly driven by the supply-demand ratio, resulting in a supply-focused model. In contrast, the multi-agent environment provides a more complex dynamic where the actors do not only have to take supply and demand into account but also the competitive nature of the system.

	Environment with 2 competitors			**Monopolistic setup**	
Fleet size	**Avg. price actor 1**	**Avg. price actor 2**	**Best reward**	**Avg. price**	**Best reward**
100	3.3452 ±0.0406	3.8928 ±0.0137	9408 ± 536	3.99 ± 0.06	10249 ± 579
187	2.9118 ±0.0527	1.9239 ±0.0054	10756 ± 483	3.78 ± 0.09	14062 ± 651
200	2.3317 ±0.0460	2.4960 ±0.0645	13691 ± 371	3.14 ± 0.07	14730 ± 563
300	2.3071 ±0.0508	2.5242 ±0.0677	16621 ± 415	2.99 ± 0.04	15627 ± 579
374	2.4921 ±0.0550	2.0306 ±0.0131	$\mathbf{17445 \pm 319}$	2.40 ± 0.02	$\mathbf{18926 \pm 866}$
400	1.9691 ±0.0659	1.6678 ±0.0211	15014 ± 511	2.19 ± 0.02	17737 ± 660

Table 2 – *Average prices and rewards (profit) variation based on the fleet size.*

Analysis of new competitors in an established monopolistic market. Lastly, we analyze how an already established (pre-trained) monopolistic RL agent reacts to the introduction of a new competitor in the market. To do so, we further train this agent, which had previously learned a monopolistic policy, in an environment with an additional competitor of equal fleet size. As illustrated in Figure 3, while the pre-trained agent initially performs better, the new (untrained) competitor quickly adapts by undercutting prices and capturing a larger share of the demand. The established agent is slow to adjust its (monopolistic) pricing strategy, allowing the competitor to outperform it.

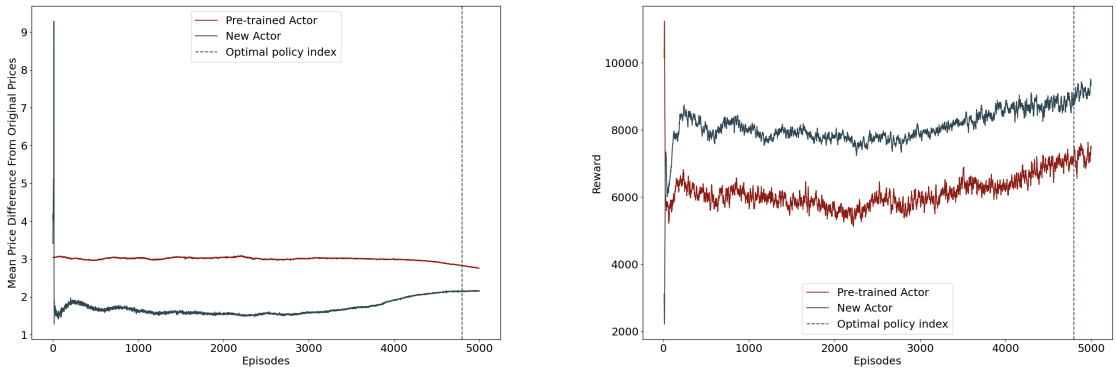


Figure 3 – *Price (left) and reward (right) evolution over the training episodes.*

References

Gammelli, D., Yang, K., Harrison, J., Rodrigues, F., Pereira, F., & Pavone, M. 2022. Graph meta-reinforcement learning for transferable autonomous mobility-on-demand. *Pages 2913–2923 of: Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining.*

Qin, Z., Zhu, H., & Ye, J. 2022. Reinforcement learning for ridesharing: An extended survey. *Transportation Research Part C: Emerging Technologies*, **144**, 103852.

Zardini, G., Lanzetti, N., Pavone, M., & Frazzoli, E. 2022. Analysis and control of autonomous mobility-on-demand systems. *Annual Review of Control, Robotics, and Autonomous Systems*, **5**(1), 633–658.