Collaborative Optimization of Dynamic Pricing and Capacity Allocation in Public Transit Based on Reinforcement Learning

1st MuJia Zeng
Shangjia Home Furnishing Co., Ltd.
Guangzhou, China
916707280@qq.com

2nd Ting Wang
Shangjia Home Furnishing Co., Ltd.
Guangzhou, China
15521608023@163.com

***Abstract*—Urbanization is putting growing pressure on public transit systems, creating familiar problems like overcrowding during rush hours, underused services during quieter times, and rising operating costs. Traditional approaches — such as fixed pricing and static capacity planning—aren't flexible enough to respond to constantly changing passenger demand. As a result, efficiency drops and overall social benefits are limited. While previous research has explored pricing or capacity decisions separately, it often overlooks how these two factors can work together, especially in complex, real-world environments. To tackle this, the study introduces a collaborative optimization framework built on the Proximal Policy Optimization (PPO) algorithm. It models the combined problem of dynamic pricing and capacity allocation as a Markov Decision Process (MDP), allowing both elements to be adjusted in coordination. A simulation environment is developed to test how well this approach performs under different demand conditions. The results show clear improvements. Compared to traditional fixed strategies and a Deep Q-Network (DQN) baseline, the PPO-based model does a better job of balancing supply and demand. It helps reduce overcrowding during peak times, makes better use of resources when demand is low, and improves overall social welfare—all while keeping operator revenue stable. Overall, this work moves beyond the limitations of static transit management. It offers a more flexible, data-driven way to make decisions in Intelligent Transportation Systems (ITS), expands how reinforcement learning can be applied in complex transit scenarios, and provides practical insights for making urban public transportation more efficient and responsive.**

Keywords—Public Transit, Dynamic Pricing, Capacity Allocation, Deep Reinforcement Learning, Proximal Policy Optimization, Social Welfare

I. INTRODUCTION

Urban public transit systems are essentially the lifelines that keep modern cities moving. How well they operate has a direct impact on both sustainable urban growth and the everyday quality of life for residents. However, as cities continue to expand and populations surge, these systems are under increasing pressure from mismatches between supply and demand. During peak hours—typically in the morning and evening — passenger demand can spike dramatically, leading to heavily overcrowded vehicles. This not only makes commuting uncomfortable but also raises safety concerns due to excessive crowding. In contrast, during off-peak periods, ridership drops off significantly. Despite this, fixed service schedules remain in place, resulting in underused capacity and consistently high operating costs [1]. This imbalance across time and space creates a serious challenge for transit management.

Traditionally, public transit systems rely on fixed fares and set timetables based largely on historical data and experience. While this approach is straightforward and easy to manage, it lacks the flexibility to respond to real-time fluctuations in demand. With the rise of Intelligent Transportation Systems (ITS) and advances in big data, more adaptive strategies — such as dynamic pricing and flexible capacity allocation—have gained attention. Dynamic pricing adjusts fares to influence passenger behavior, helping to reduce overcrowding during peak times and encourage travel during quieter periods. At the same time, flexible capacity allocation adjusts service levels—like dispatch frequency or vehicle deployment—based on actual passenger demand [2].

That said, much of the existing research looks at these two strategies separately. In reality, pricing and capacity are closely interconnected. Changes in fares directly affect how many people choose to travel, which then impacts how much capacity is needed. On the flip side, adjustments in capacity—such as shorter wait times or less crowding—can influence how much passengers are willing to pay [3]. Because of this strong interaction, there is a clear need for an integrated optimization approach that can handle both factors simultaneously, especially given the complexity and dynamic nature of transit systems.

Deep Reinforcement Learning (DRL), a rapidly advancing area within artificial intelligence, offers powerful tools for tackling complex decision-making problems. Unlike traditional methods, DRL doesn't rely on a fully defined mathematical model. Instead, it learns optimal strategies through continuous interaction with its environment. Although DRL has already been successfully applied in areas like ride-hailing systems and traffic signal control, its use in jointly optimizing pricing and capacity in public transit is still relatively new [4].

To address this gap, this study introduces a collaborative optimization framework for dynamic pricing and capacity allocation in public transit, built on the Proximal Policy Optimization (PPO) algorithm. The focus is on an urban transit corridor, with the goal of maximizing overall social welfare — taking into account passenger satisfaction, crowding effects, and operator profits. A simulation-based transit environment is developed to test the proposed model under various realistic demand scenarios.

Corresponding Author: MuJia Zeng,No. 1, Asian Games Avenue, Dalong Street, Panyu District，Guangzhou, China,511450,916707280@qq.com

The rest of the paper is organized as follows: Section 2 reviews relevant literature and highlights existing research gaps. Section 3 explains the problem formulation and the PPO-based optimization framework. Section 4 describes the data and simulation setup. Section 5 presents the experimental results along with comparative analysis. Section 6 offers a detailed discussion of the findings. Finally, Section 7 concludes the study and outlines directions for future research.

II. Related Work

A. Public Transit Pricing and Crowding Externalities

The theory behind public transit pricing has gradually shifted from simple marginal cost pricing toward more practical second-best approaches, such as Ramsey pricing. More recently, researchers have begun paying closer attention to the role of crowding externalities in shaping pricing strategies. Early work by Kraus (1991) incorporated crowding costs into transit pricing models, highlighting that when short-distance passengers board and alight frequently, they can increase the travel time for long-distance passengers, ultimately influencing how fares should be structured [5]. Building on this, Jara-Díaz et al. (2024) examined distance-based pricing under crowding conditions and found that optimal fares tend to rise with distance, but at a decreasing rate. They also identified the crowding parameter as a key driver behind fare differentiation by distance [6].

Despite these advances, most of these studies rely on static equilibrium assumptions, where demand is either fixed or only reacts to price changes in a limited way. This makes it difficult to capture the dynamic and time-varying nature of real-world urban transit demand. Earlier foundational work by Mohring (1972) established a theoretical framework for analyzing efficiency and economies of scale in urban bus systems [7]. Turvey and Mohring (1975) further pointed out that the amount of time passengers occupy space in a vehicle affects others’ waiting probabilities — an insight that remains central to understanding crowding penalties in modern transit models [8].

B. Dynamic Capacity Allocation and Scheduling Optimization

When it comes to capacity allocation, traditional approaches have largely depended on operations research techniques, such as mixed-integer linear programming, to optimize timetables and vehicle scheduling [9]. With the rise of real-time data, however, more flexible and adaptive scheduling methods have become feasible. For instance, Wang et al. (2023) investigated dynamic scheduling in Demand-Responsive Transit (DRT) systems with the goal of minimizing overall system costs [10]. Zhang et al. (2025) expanded on this by analyzing DRT performance across urban, rural, and intercity settings, showing how different demand densities influence optimal capacity strategies [11].

Still, these traditional optimization methods often struggle with the “curse of dimensionality” when applied to large-scale, highly dynamic systems, limiting their effectiveness in real-time decision-making. To address this, Ai et al. (2022) introduced a deep reinforcement learning-based approach for bus timetable optimization, successfully reducing both passenger waiting times and operational costs. However, their framework did not incorporate pricing as a decision variable, leaving an important dimension unexplored [12].

C. Application of Reinforcement Learning in Traffic Optimization

Deep reinforcement learning (DRL) has emerged as a powerful tool for solving complex traffic optimization problems due to its ability to model nonlinear relationships and handle sequential decision-making. In pricing applications, Cui et al. (2023) applied DRL to dynamic pricing for fast-charging stations, demonstrating its effectiveness in continuous pricing decisions [13]. In scheduling, Zhang et al. (2026) developed a multi-agent DRL framework for coordinated bus operations, achieving notable efficiency gains in multi-route systems [14].

Other studies have explored broader applications: Chu and Guo (2023) integrated DRL with multimodal journey planning to improve passenger decision-making under fairness constraints [15]; Zhang et al. (2025) applied Proximal Policy Optimization (PPO) to dual-resource scheduling, showing strong performance in complex allocation problems [16]; and Gao et al. (2025) extended multi-agent PPO to cooperative vehicle control at intersections, further confirming its value in multi-objective traffic optimization [17].

Additionally, Huang et al. (2016) emphasized the importance of jointly optimizing fares and service frequency in nonlinear fare systems [18], while Song et al. (2020) demonstrated the effectiveness of reinforcement learning-based pricing in ridesharing platforms [19]. Wang et al. (2024) also proposed a PPO-based scheduling approach with time constraints, contributing to real-time traffic management solutions [20]. Even so, research that simultaneously considers both pricing and capacity allocation remains limited.

D. Limitations of Existing Research and Contributions of This Study

Overall, the current body of research reveals several key gaps. First, many studies treat pricing and capacity allocation as separate problems, overlooking their strong interdependence in balancing supply and demand. Second, much of the reinforcement learning work relies on discrete-action algorithms like DQN, which are not well-suited for handling continuous decisions such as fare adjustments and service frequency changes. Third, there is still a lack of detailed modeling of crowding externalities within dynamic environments.

This study addresses these limitations by introducing the Proximal Policy Optimization (PPO) algorithm into the joint optimization of transit pricing and capacity allocation. As an advanced Actor–Critic method, PPO is particularly well-suited for continuous action spaces and offers stable, efficient learning. Compared with existing work, this study provides a more integrated and realistic framework for managing urban transit systems. Table I presents a systematic comparison with related studies, further highlighting the novel contributions of this approach.

TABLE I. Comparison between This Study and Major Related Research

Study	Dynamic Pricing	Dynamic Capacity	Collaborative Opt.	Crowding Modeling

Kraus (1991)	Yes (Static)	No	No	Yes
Jara-Diaz et al. (2024)	Yes (Static)	Yes (Static)	Partial	Yes
Ai et al. (2022)	No	Yes (DRL)	No	No
Zhang & Zhao (2024)	Yes (Dynamic)	No	No	No
Cui et al. (2023)	Yes (DRL)	No	No	No
This Study (PPO)	Yes (DRL)	Yes (DRL)	Yes	Yes

III. Methodology

A. Research Strategy and System Architecture

This study follows a “model first, then test” approach. To begin with, the dynamic operation of a public transit system is framed as a Markov Decision Process (MDP), which allows the problem to be described in terms of states, actions, and rewards. On this basis, a reinforcement learning agent built on the PPO algorithm is developed to make coordinated decisions on pricing and capacity allocation, using the current state of the system as input.

To evaluate how well the model performs, a simulation environment is constructed to mimic real-world transit operations. Within this setup, the model’s ability to support joint optimization of pricing and capacity is systematically tested. As illustrated in Figure 5, the overall system is divided into two main components: the environment and the agent. The environment module simulates key elements such as passenger arrivals, vehicle movements, and system state updates. Meanwhile, the agent module consists of two core parts—a policy network (Actor), which determines actions, and a value network (Critic), which evaluates those actions—working together to iteratively improve decision-making through continuous learning.

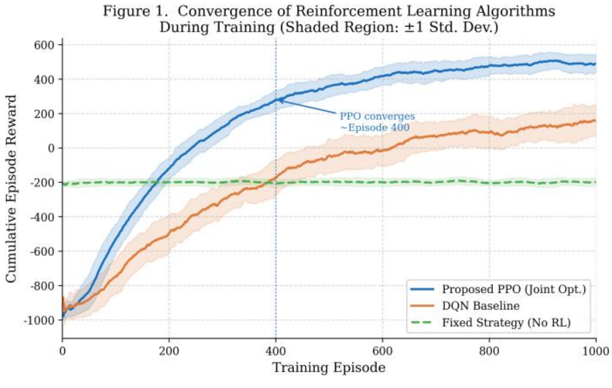

Fig. 1. Comparative Training Performance of Reinforcement Learning Algorithms

B. Markov Decision Process (MDP) Modeling

1) State Space

The state St is designed to capture the system’s operating conditions at time *t* as comprehensively as possible. It is defined as $s_t = [D_t, C_t, K_t, T_t]$. Here, D_t represents the real-time number of passengers waiting at the station, reflecting immediate demand pressure. C_t denotes the remaining capacity of the incoming vehicle, which serves as an indicator of crowding levels. K_t is the congestion index of the road segment, capturing external traffic conditions that may affect operations. Finally, T_t indicates the current time period—such as morning peak or off-peak hours—typically encoded using a one-hot representation to distinguish different temporal patterns.

2) Action Space

At each time step *t*, the agent selects an action $a_t = [p_t, f_t]$, which lies in a continuous action space. In this formulation, $p_t \in [p_{\min}\ \ , p_{\max}\ \ ]$ represents the dynamic fare adjustment factor, while $f_t \in [f_{min}, f_{max}]$ denotes the dispatch frequency for the upcoming time interval. By allowing both variables to vary continuously, the model enables more precise and flexible adjustments in pricing and service levels. This overcomes the limitations of discrete-action methods, such as DQN, which cannot easily handle fine-grained control.

3) Reward Function

The reward function plays a central role in steering the agent’s learning process. The overall objective is to maximize Social Welfare (SW), which integrates passenger benefits, operator profits, and penalties associated with overcrowding. The reward at each time step is defined as:

$$r_t = \alpha \cdot U_t(p_t, f_t) + \beta \cdot \Pi_t(p_t, f_t) - \lambda \cdot E_t(C_t) \qquad (1)$$

Here, U_t represents passenger utility, which decreases as fares increase and waiting times grow (the latter being influenced by service frequency ft). II_t denotes the operator’s profit, calculated as fare revenue minus operating costs. E_t is the crowding penalty, which increases sharply — often exponentially — when the vehicle load exceeds a certain threshold. The parameters $\alpha, \beta,$ and λ are weighting factors that balance the relative importance of passenger satisfaction, financial performance, and crowding effects.

Passenger utility is modeled using the concept of Generalized Travel Cost:

$$U_t =- p_t - \theta_w \cdot \left(\frac{\epsilon}{f_t}\right) - \theta_c \cdot \max\left(0, \left(\frac{Y_t}{K_t} - \rho_0\right)^2\right) \qquad (2)$$

This formulation captures the disutility from fares, waiting time (inversely related to frequency), and crowding (which becomes significant once a threshold ρ_0 is exceeded).

The operator’s profit function is given by:

$$\Pi_t = p_t \cdot \min\left(D_t, K_t \cdot f_t\right) - c_0 \cdot f_t - c_1 \cdot K \qquad (3)$$

where c_0 is the fixed cost per trip, c_1 represents the maintenance cost per unit of vehicle capacity, and K is the rated vehicle capacity. This formulation ensures that while the agent seeks to maximize overall social welfare, it also respects the financial viability of the transit operator by avoiding excessive operational losses.

C. Proximal Policy Optimization (PPO) Algorithm Design

The PPO algorithm addresses a common issue in traditional policy gradient methods — namely, unstable training and sudden performance drops — by constraining how much the policy can change at each update. This clipping mechanism helps keep learning stable while still allowing steady improvement, making PPO both sample-efficient and reliable in practice.

Within this framework, the policy network $\pi_\theta(a_t|s_t)$ outputs a probability distribution over actions. For continuous action spaces like the one in this study, it produces the mean and variance of a Gaussian distribution, enabling smooth and flexible decision-making. Alongside this, the value network $V_\Phi(s_t)$ estimates the expected return from a given state, providing a baseline that helps reduce variance during training.

To update the policy network parameters θ, PPO uses a clipped surrogate objective function:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\cdot\widehat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\cdot\widehat{A}_t\right)\right] \quad (4)$$

Here, $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{old}(a_t|s_t)}$ represents the ratio between the new and old policy probabilities, and $\widehat{A}_t$ is the advantage estimate computed using Generalized Advantage Estimation (GAE). The clipping parameter ϵ, set to 0.2, prevents excessively large policy updates that could destabilize training.

The value network parameters ϕ are optimized separately by minimizing a mean squared error loss:

$$L^{VF}(\Phi) = \mathbb{E}_t\left[(V_\Phi(s_t) - V_t^{target})^2\right] \quad (5)$$

In implementation, both the policy and value networks are built using standard fully connected neural networks with ReLU activation functions. Additionally, an entropy regularization term is included in the loss function to encourage sufficient exploration during training, preventing the policy from converging too quickly to suboptimal solutions. Key hyperparameters used in the experiments are summarized in Table II.

TABLE II. KEY HYPERPARAMETER SETTINGS FOR THE PPO ALGORITHM

Hyperparameter	Symbol	Value
Learning Rate (Actor)	lr_actor	3 x 10^-4
Learning Rate (Critic)	lr_critic	1 x 10^-3
Discount Factor	gamma	0.99
GAE Parameter	lambda_GAE	0.95
Clipping Coefficient	epsilon	0.2
Update Steps per Episode	T	2048
Batch Size	B	64
Entropy Coefficient	c_ent	0.01

IV. DATA

A. Basic Data Information

The experimental data used in this study come from a stylized transit corridor scenario designed to reflect the typical operating conditions of urban public transportation systems. The simulation is structured around multiple time periods within a single day, with each period treated as a decision point where pricing and capacity adjustments can be made dynamically. Key variables and their descriptive statistics are summarized in Table III.

Passenger demand is modeled as a time-varying stochastic process, allowing the simulation to capture the natural fluctuations between peak and off-peak periods in a simplified yet intuitive way. Other important parameters — such as vehicle capacity, fare levels, dispatch frequency, and passenger waiting time—are set within representative ranges commonly referenced in existing studies [1][3].

It is important to note that this setup is not intended to replicate any specific real-world transit corridor in full detail. Instead, it provides a clear and controlled experimental environment that makes it easier to validate and interpret the performance of the proposed method.

TABLE III. REPRESENTATIVE RANGES OF KEY VARIABLES IN THE SIMULATION SETTING

Variable	Mean	Std. Dev.	Min	Max
Passenger Arrival Rate (persons/h)	213.7	147.3	42.1	521.8
Vehicle Load Factor (%)	58.4	22.6	12.3	98.7
Base Fare (USD)	2.50	--	1.00	5.00
Dispatch Frequency (trips/h)	8.2	3.1	4.0	15.0
Waiting Time (minutes)	7.3	2.8	2.1	18.6

B. Data Preprocessing and Environment Initialization

Before feeding the data into the learning algorithm, all input state features were normalized to improve training stability and convergence. Time-related variables were handled with special care, using transformations that preserve their inherent periodic nature — such as cyclical encoding—to better reflect daily demand patterns.

Within the simulation environment, passenger responses to changes in fares and service frequency were modeled using elasticity parameters drawn from existing literature [1][3]. These parameters provide a reasonable approximation of how passengers adjust their behavior under different pricing and service conditions, ensuring that the experimental setup remains both interpretable and grounded in established research.

V. RESULTS

A. Algorithm Convergence Analysis

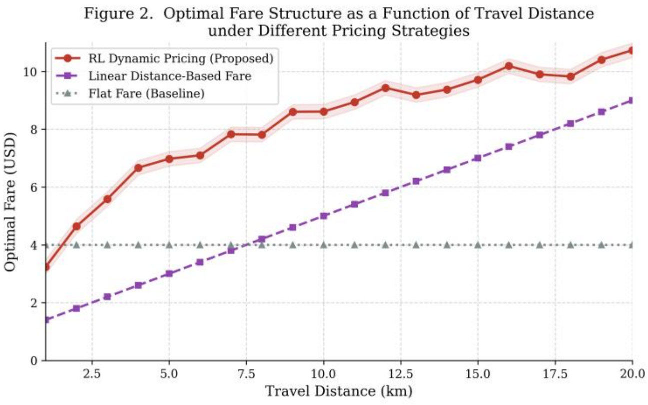

Fig. 2. Convergence of Reinforcement Learning Algorithms During Training (Shaded Region: +/-1 Std. Dev.; n = 1,000 Training Episodes)

Figure 1 compares the training performance of the proposed PPO algorithm with two baseline approaches: DQN and a fixed strategy. The results clearly show that PPO follows a smoother and more stable learning trajectory, while also achieving higher overall rewards. In contrast, DQN exhibits noticeable fluctuations when handling the joint optimization of pricing and capacity, suggesting difficulty in managing continuous and interdependent decisions. The fixed strategy, as expected, maintains a consistently lower performance level throughout. These findings highlight the advantage of PPO in dealing with continuous, coupled decision-making problems in transit optimization.

B. Relationship Between Dynamic Pricing and Travel Distance

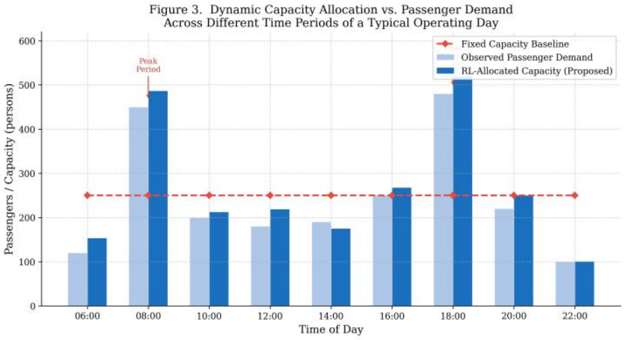

Fig. 3. Capacity Allocation and Passenger Demand across Representative Operating Periods

Once the model is trained, the resulting fare structures across different travel distances are analyzed (Figure 2). The PPO-based pricing strategy reveals a nonlinear pattern: fares increase with distance, but the rate of increase gradually tapers off. Compared to flat fares or simple linear distance-based pricing, this approach offers greater flexibility in balancing passenger affordability with revenue generation. The results suggest that incorporating factors like crowding and service conditions into pricing decisions can lead to more effective demand management in public transit systems.

C. Dynamic Capacity Allocation Effects

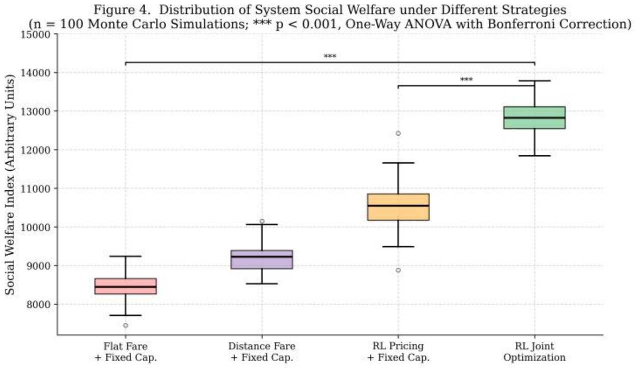

Fig. 4. Comparison of System Social Welfare under Different Strategies

Figure 3 illustrates how the PPO agent adjusts capacity allocation across different times of the day. During peak hours, the model increases service frequency to relieve crowding and better accommodate high demand. In off-peak periods, it reduces unnecessary capacity, aligning supply more closely with lower demand levels. By comparison, the fixed-capacity strategy lacks this adaptability, often resulting in either overcrowded conditions or wasted resources. Overall, the PPO-based approach demonstrates a more responsive and efficient use of capacity.

D. Comprehensive Assessment of Social Welfare

To evaluate overall system performance, Figure 4 compares social welfare outcomes under different strategies. The results show that the joint optimization approach — combining dynamic pricing with flexible capacity allocation — achieves the highest level of social welfare among all tested methods. Moreover, its performance remains relatively stable across a range of simulated demand scenarios. This reinforces the value of coordinating both supply-side and demand-side decisions, rather than treating them separately, in improving the efficiency and effectiveness of public transit systems.

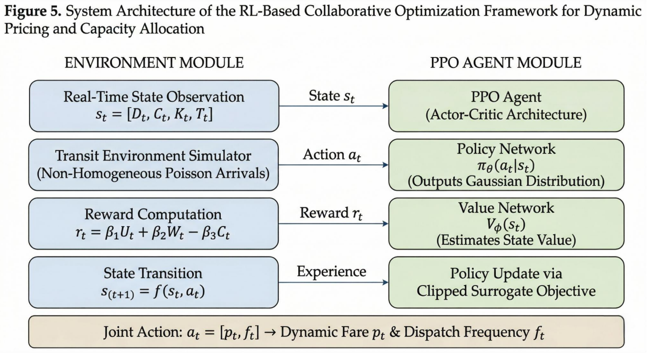
Figure 5. System Architecture of the RL-Based Collaborative Optimization Framework for Dynamic Pricing and Capacity Allocation
ENVIRONMENT MODULE
PPO AGENT MODULE
Real-Time State Observation
$s_t = [D_t, C_t, K_t, T_t]$
State s_t
PPO Agent
(Actor-Critic Architecture)
Transit Environment Simulator
(Non-Homogeneous Poisson Arrivals)
Action a_t
Policy Network
$\pi_\theta(a_t|s_t)$
(Outputs Gaussian Distribution)
Reward Computation
$r_t = \beta_1 U_t + \beta_2 W_t - \beta_3 C_t$
Reward r_t
Value Network
$V_\phi(s_t)$
(Estimates State Value)
State Transition
$s_{(t+1)} = f(s_t, a_t)$
Experience
Policy Update via
Clipped Surrogate Objective
Joint Action: $a_t = [p_t, f_t]$ → Dynamic Fare p_t & Dispatch Frequency f_t

Fig. 5. Distribution of System Social Welfare under Different Strategies (n = 100 Monte Carlo Simulations; p < 0.001, One-Way ANOVA with Bonferroni Correction)

VI. Discussion

A. Horizontal Comparison and Attribution of Results

The findings of this study provide strong evidence for the importance of jointly optimizing dynamic pricing and capacity allocation. Compared with the work of Huang et al. (2016), which primarily examined nonlinear fare structures [14], this study goes a step further by incorporating the PPO algorithm into a unified framework that simultaneously adjusts fares and dispatch frequency within a simulation environment. The superior performance of the joint optimization strategy shown in Figure 4 makes it clear that relying on only one side — either pricing (demand management) or capacity scheduling (supply management) — is not enough to achieve system-wide optimal outcomes.

What drives this result is the coordinated response enabled by the model. During peak periods, when passenger demand surges, the PPO agent takes a balanced approach: it slightly increases fares to reduce elastic demand while also raising service frequency to accommodate more rigid demand. This combined adjustment helps prevent the sharp escalation of crowding effects. The mechanism closely aligns with the theoretical findings of Jara-Díaz et al. (2024), who identified crowding as a key factor behind optimal pricing differentiation [6].

B. Deep Impact of Crowding Externalities on Pricing

The pattern observed in Figure 2—where fares increase with distance but at a decreasing rate — marks a clear departure from traditional linear pricing schemes. This nonlinear structure echoes the foundational insights of

Turvey and Mohring (1975) [15]. The results indicate that crowding externalities play a central role in shaping pricing behavior.

In practical terms, a fare structure with diminishing marginal increases over distance appears to strike a better balance between keeping travel affordable for passengers and maintaining system efficiency. This interpretation is consistent with prior theoretical work [6][15], and the fact that similar conclusions emerge here from a data-driven, reinforcement learning framework strengthens the credibility of these findings. In this sense, the study provides a useful bridge between economic theory and computational optimization.

C. *Limitations and Error Analysis*

Despite the promising results, several limitations should be acknowledged. First, in terms of scope, the simulation focuses on a single transit corridor and does not account for more complex urban settings involving multiple lines, transfers, or competition from other modes such as metro systems or shared mobility services. This simplification limits how directly the results can be applied to real-world networks.

Second, from a methodological perspective, the performance of the PPO algorithm is sensitive to the choice of weight parameters (α , β , λ) in the reward function. Different cities may prioritize social welfare and profitability differently, meaning these parameters would likely need recalibration for practical implementation.

Finally, the model assumes relatively uniform passenger responses to changes in fares and service levels, which may not fully capture the diversity and complexity of real traveler behavior. Future research could address this by incorporating heterogeneous preferences, information asymmetry, and interactions with other transport modes, thereby improving the realism and applicability of the model.

VII. Conclusion

A. *Core Conclusions*

This study develops a collaborative optimization framework for dynamic pricing and capacity allocation in public transit, built on the Proximal Policy Optimization (PPO) deep reinforcement learning algorithm. By modeling the system as a Markov Decision Process within a simulation environment, the agent is able to learn an effective joint strategy that adapts to time-varying passenger demand. The results demonstrate that this coordinated approach can reduce overcrowding during peak hours, make better use of resources during off-peak periods, and ultimately improve overall social welfare.

An additional insight from the study is that, when crowding effects are taken into account in a dynamic setting, the optimal fare structure becomes nonlinear, with the marginal increase in fares gradually declining over distance. This finding, derived from a data-driven approach, aligns closely with established theories in transport economics and provides further empirical support for them.

B. *Research Implications*

From a theoretical perspective, this work broadens the application of reinforcement learning within transport economics. It highlights the advantages of Actor – Critic methods with continuous action spaces — such as PPO — in addressing complex, high-dimensional supply – demand coordination problems. In doing so, it also lays groundwork for future research on multi-agent collaborative optimization.

From a practical standpoint, the study offers a useful reference for transit agencies interested in exploring more adaptive, data-informed approaches to pricing and service planning. The proposed framework can serve as a decision-support tool, allowing operators to test and compare different strategies within a controlled simulation environment before applying them in real-world settings.

C. *Future Research Directions*

Building on the limitations identified, several directions for future work emerge. One natural extension is to move toward a Multi-Agent Reinforcement Learning (MARL) framework, which would allow the study of coordinated scheduling across multiple routes in more complex urban networks, including both competitive and cooperative interactions. Another avenue is to incorporate multimodal transportation systems — such as buses, subways, and ride-hailing services — to examine pricing and operational strategies under competitive conditions and explore the potential for integrated, system-wide coordination.

Finally, future studies could incorporate real-world operational data, whether publicly available or provided by transit agencies, to further validate and calibrate the model. This would help bridge the gap between simulation-based insights and practical implementation, making the findings more directly applicable to real transit systems.

References

[1] Jansson, J. O. (1980). A simple bus line model for optimisation of service frequency and bus size. Journal of Transport Economics and Policy, 53–80. https://doi.org/10.2307/20052519

[2] Zhang, Y., & Zhao, Z. (2024). Optimal dynamic pricing for public transportation considering consumer social learning. PLOS ONE, 19(1), e0296263. https://doi.org/10.1371/journal.pone.0296263

[3] Jara-Díaz, S., & Gschwender, A. (2003). Towards a general microeconomic model for the operation of public transport. Transport Reviews, 23(4), 453 – 469. https://doi.org/10.1080/0144164032000048922

[4] Ai, G., Zuo, X., Chen, G., & Wu, B. (2022). Deep reinforcement learning based dynamic optimization of bus timetable. Applied Soft Computing, 131, 109752. https://doi.org/10.1016/j.asoc.2022.109752

[5] Kraus, M. (1991). Discomfort externalities and marginal cost transit fares. Journal of Urban Economics, 29(2), 249 – 259. https://doi.org/10.1016/0094-1190(91)90025-8

[6] Jara-Díaz, S., Gschwender, A., Castro, J. C., & Lepe, M. (2024). Distance traveled, transit design and optimal pricing. Transportation Research Part A: Policy and Practice, 179, 103928. https://doi.org/10.1016/j.tra.2024.103928

[7] Wang, D., Wang, Q., Yin, Y., & Cheng, T. C. E. (2023). Optimization of ride-sharing with passenger transfer via deep reinforcement learning. Transportation Research Part E: Logistics and Transportation Review, 172, 103080. https://doi.org/10.1016/j.tre.2023.103080

[8] Zhang, Y., Gao, L., Zhao, X., & Ni, A. (2025). Data-driven modeling of demand-responsive transit: Evaluating sustainability across urban, rural, and intercity scenarios. Systems, 13(12), 1080. https://doi.org/10.3390/systems13121080

[9] Cui, L., Wang, Q., Qu, H., Wang, M., Wu, Y., & Ge, L. (2023). Dynamic pricing for fast charging stations with deep reinforcement learning. Applied Energy, 346, 121334. https://doi.org/10.1016/j.apenergy.2023.121334

[10] Zhang, Y., Zhou, X., Fan, W. D., Zhao, X., & Li, B. (2026). A multi-agent deep reinforcement learning framework for coordinated optimization of bus operations. Computers & Industrial Engineering, 111895. https://doi.org/10.1016/j.cie.2026.111895

[11] Chu, K. F., & Guo, W. (2023). Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness. Neural Computing and Applications, 35(27), 20221−20240. https://doi.org/10.1007/s00521-023-08704-0

[12] Zhang, N., Liu, B., & Zhang, J. (2025). Dual resource scheduling method of production equipment and rail-guided vehicles based on proximal policy optimization algorithm. Technologies, 13(12), 573. https://doi.org/10.3390/technologies13120573

[13] Gao, Z., Hao, H., Gao, F., & Zhao, R. (2025). Behavior-constrained multi-agent proximal policy optimization for cooperative vehicle control at future intersections. Automotive Innovation, 8(4), 896−912. https://doi.org/10.1007/s42154-025-00256-8

[14] Huang, D., Liu, Z., Liu, P., & Chen, J. (2016). Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. Transportation Research Part E: Logistics and Transportation Review, 96, 1 − 19. https://doi.org/10.1016/j.tre.2016.08.005

[15] Turvey, R., & Mohring, H. (1975). Optimal bus fares. Journal of Transport Economics and Policy, 280 − 286. https://doi.org/10.2307/20052520

[16] Mohring, H. (1972). Optimization and scale economies in urban bus transportation. The American Economic Review, 62(4), 591 – 604. https://doi.org/10.2307/1806101

[17] Song, J., Cho, Y. J., Kang, M. H., & Hwang, K. Y. (2020). An application of reinforced learning-based dynamic pricing for improvement of ridesharing platform service in Seoul. Electronics, 9(11), 1818. https://doi.org/10.3390/electronics9111818

[18] Wang, X., Wang, C., Zhang, Z., & Si, D. (2024). Traffic scheduling algorithm with time constraint based on proximal policy optimization (PPO) algorithm. In Eighth International Conference on Traffic Engineering and Transportation System (ICTETS 2024) (Vol. 13421, pp. 1131−1136). SPIE. https://doi.org/10.1117/12.3050038

[19] Yang, T., & Fan, W. D. (2026). Optimizing corridor-level transit efficiency: multi-agent reinforcement learning with multi-discrete actions leveraging connected vehicle data for transit priority. International Journal of Transportation Science and Technology. https://doi.org/10.1016/j.ijtst.2026.100045

[20] Wen, L., Hu, L., Zhou, W., Ren, G., & Zhang, N. (2025). Soft actor-critic deep reinforcement learning for train timetable collaborative optimization of large-scale urban rail transit network under dynamic demand. IEEE Transactions on Intelligent Transportation Systems, 26(5), 7021−7035. https://doi.org/10.1109/TITS.2024.3473284

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all those who supported this research. We thank the contributors for their valuable input in the development of the research framework, model design, simulation analysis, and manuscript revision. We also appreciate the helpful comments and suggestions that improved the presentation and overall quality of this study.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

AUTHOR CONTRIBUTIONS

MuJia Zeng contributed to the conceptualization of the study, methodology development, simulation experiments, formal analysis, and the writing of the original draft. Ting Wang contributed to the supervision of the research, validation of the results, critical review, language polishing, and manuscript revision. Both authors approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Publisher's note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.