

Intelligent Traffic Signal Control Based on Reinforcement Learning with Edge Computing and Intelligent Reflecting Surface

Guanghua Zhang, Youchen Yue, Jinghao Yang, Jie Tang

College of Electrical and Information Engineering, Northeast Petroleum University, Daqing 163318, China

Abstract: With increasing traffic demand and growing environmental complexity, traditional signal control methods struggle with low efficiency and poor adaptability. This study proposes a reinforcement learning-based framework integrating edge computing and intelligent reflecting surfaces (IRS) to enhance traffic signal control. The framework is validated through simulation and optimized via an improved Proximal Policy Optimization (PPO) algorithm with multi-step returns and entropy-based adaptive regularization, ensuring better convergence and stability. Experiments on a 5×5 intersection network in SUMO, under varying traffic scenarios, including ablation and comparative studies, show that the proposed method significantly reduces average delay and improves throughput, demonstrating its effectiveness and reliability.

Keywords: Proximal Policy Optimization, Edge Computing, Intelligent Reflecting Surfaces.

1. Introduction

In recent years, the rapid urbanization has exacerbated traffic congestion, intersections have emerged as critical bottlenecks. One primary cause is the inefficiency of conventional traffic signal systems [1]. Consequently, there is an urgent demand for a dynamic, adaptive signal control system that ensuring real-time responsiveness, low latency, and wide coverage [2, 3]. Traffic signal control systems have evolved from manual operation to fixed-time and actuated control, and more recently to intelligent control [4]. The introduction of reinforcement learning has enabled signal control systems to adapt to environmental changes. More recent studies integrated six DQN variants, including prioritized experience replay and distributed DQN, improving adaptability and efficiency [5, 6]. However, value-based approaches often suffer from Q-value overestimation in high-dimensional or continuous state spaces, leading to unstable training and poor convergence in dynamic or uncertain environments [7]. The clipping mechanism of PPO may lead to insufficient sampling in some cases, potentially limiting optimal action selection [8-9].

Edge computing (EC) can reduce latency and bandwidth usage by moving computation to the network edge, enabling real-time, low-latency data processing [10]. It allows rapid local processing of massive data streams from traffic sensors, cameras, and other devices, and supports real-time adjustment of traffic signal strategies using deep reinforcement learning, thereby improving responsiveness and accuracy [11].

An intelligent reflecting surface (IRS), as a communication enhancement technology, optimizes signal propagation by controlling reflecting elements, offering advantages in coverage and reliability [12, 13]. By optimizing propagation paths, IRS mitigates signal attenuation and interference, providing reliable data support for deep reinforcement learning-based traffic signal control.

To sum up, this paper proposes an innovative framework for intelligent traffic signal control and optimizes control strategies for intersections with observable traffic states.

2. Problem Statement and System Model

Given the scarcity of publicly available traffic signal control datasets, this study employs simulated road networks, traffic flows, and intersections using SUMO. In particular, device deployment at intersections is modeled via a SUMO-Python co-simulation framework.

A bidirectional six-lane intersection was constructed, where each approach consists of a 180-meter segment with three lanes (Fig.1). Furthermore, a 5×5 grid of interconnected intersections was modeled in SUMO to represent a realistic urban traffic environment (Fig.2).

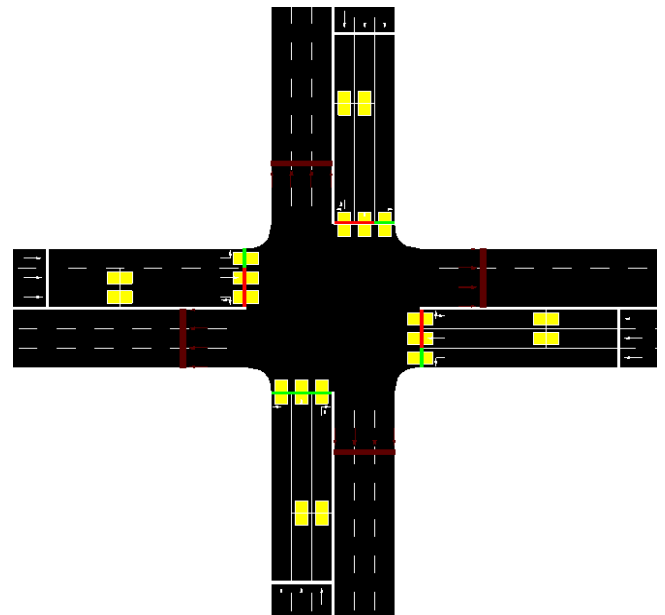


Figure 1. Simulation of a Single-Intersection Lane

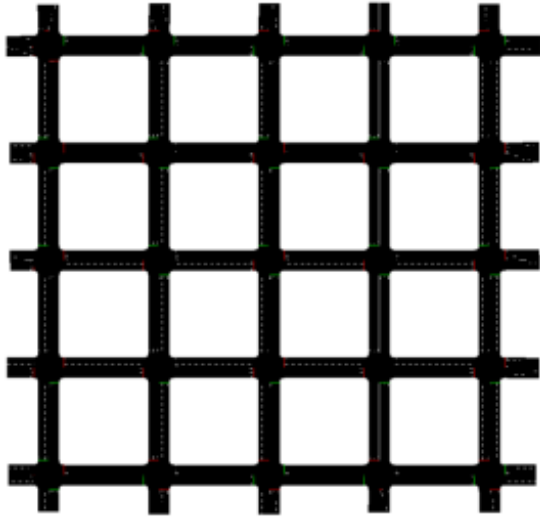


Figure 2. Simulation of a Realistic Road Environment

Analysis of vehicle data from selected intersections shows that, considering intersection layouts and driver routing choices. The simulated traffic flow data are summarized in Table 1.

By offloading computation to the edge, the system enables real-time intersection state perception and timely control decisions. Furthermore, through collaborative optimization with the IRS module, communication link quality and stability are enhanced, forming a closed-loop control system following the ‘perception–decision–execution’ paradigm

Table 1. Simulated Traffic Flow Data for Intersection Movement Patterns

Location	Direction	Traffic Flow	Location	Traffic Direction	Traffic Flow
East Approach	Through	1658	West Approach	Through	1562
	Left Turn	1586		Left Turn	1476
	Right Turn	1308		Right Turn	196
	U-turn	130		U-turn	132
North Approach	Through	1546		South Approach	Through
	Left Turn	1514	Left Turn		1306
	Right Turn	172	Right Turn		1144
	U-turn	128	U-turn		122

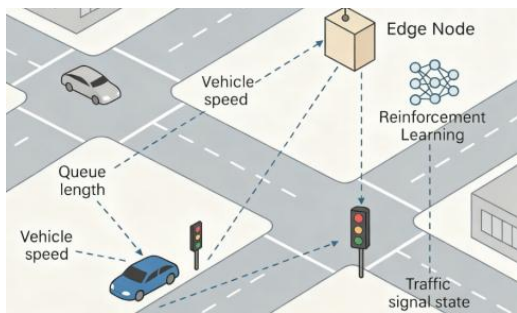


Figure 4. Urban Simulation Environment

By dynamically adjusting the phases of the reflecting elements, the IRS reflects incident signals toward the desired direction, optimizing signal transmission under NLoS conditions. In this system, the IRS is deployed on traffic signals and RSUs at each side of the intersection, covering typical communication blind spots. Its primary functions can be summarized as follows: enhancing vehicle–edge node communication stability and supporting continuous updates of reinforcement learning (Fig.4).

(Fig.3).

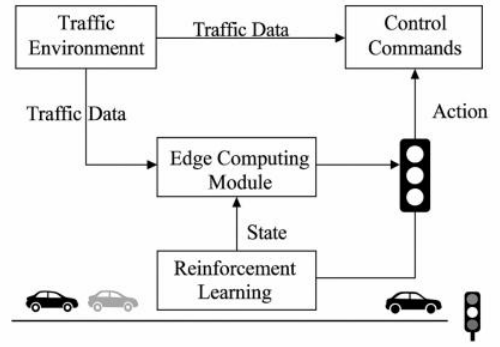


Figure 3. Data Flow Diagram

In this study, the edge intelligence control engine is implemented in Python. It acquires real-time traffic information, including vehicle count, speed, lane occupancy, and queue length, which are preprocessed and normalized into state vectors for the reinforcement learning policy. By monitoring communication quality in real time, the edge nodes dynamically adjust the phases of the reflecting elements θ_n via the IRS control module, ensuring timely and reliable communication between vehicles and RSUs. Decision outputs are transmitted to the traffic signal controllers, and execution results serve as immediate rewards for the reinforcement learning algorithm. This establishes a closed-loop feedback mechanism that continuously optimizes the control policy.

3. Reinforcement Learning-Based Intelligent Control Algorithm

To achieve intelligent perception and dynamic control at complex intersections, this study develops a reinforcement learning–based signal control engine. By deploying agents at edge nodes, the system monitors traffic conditions in real time, reducing vehicle delays and queue lengths while improving communication efficiency. An initial model is constructed with DQN [14, 15], followed by PPO, and further enhanced with multi-step returns and adaptive target-entropy regularization (ME-PPO) to achieve a more stable and efficient learning process.

This study abstracts the traffic signal control task as a MDP to enable modeling and problem solving under reinforcement learning. The MDP is defined by a quintuple: (S, A, P, R, λ) ; State Space S , Action Space A , State Transition

Probability $P(s'|s, a)$, Reward Function $R(s, a)$ and Discount Factor λ .

This study develops a dual-network structure for the DQN-based intelligent traffic signal control algorithm, consisting of an Online Q-Network and a Target Q-Network. Both networks share the same topology but differ in parameter update mechanisms, thereby enhancing training stability and mitigating Q-value overestimation.

To address these challenges, this study introduces two key enhancements to PPO: Multi-step Returns (MSR) and Adaptive entropy regularization. The improved PPO,

$$L^{PPO}(\theta) = E_t \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) \hat{A}_t \right) \right] + c_s \cdot H(\pi_\theta) - c_v \cdot L^{value}(\theta) \quad (1)$$

Here, $c_s, c_v > 0$, c_s and c_v denote the weights of the entropy regularization term and the value function term,

$$r_t(\theta) = \frac{\pi_\theta(a_t | s_t)}{\pi_{old}(a_t | s_t)}, \hat{A}_t = \sum_{l=0}^{T-t-1} (\gamma \delta)^l \delta_{t+l}, \delta_t = r_t + \gamma V_\Phi(s_{t+1}) - V_\Phi(s_t) \quad (2)$$

By introducing MSR, the advantage function estimation can be formulated as:

$$A_t^{(n)} = \sum_{k=0}^{n-1} \gamma^k r_{t+k} + \gamma^n V(s_{t+n}) - V(s_t) \quad (3)$$

Here, n denotes the return step size. In this study, $n=5$ is selected as a moderate step length to strike a balance between bias and variance, thereby ensuring

Both stability and convergence speed.

To prevent the agent from prematurely converging to a

$$L^{ME-PPO}(\theta) = E \left[\min \left(r_t(\theta) \hat{A}_t^{(n)}, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) \hat{A}_t^{(n)} \right) \right] + c_s E_t [H(\pi_\theta(\cdot | s_t))] - c_v E_t [H(\pi_\theta(\cdot | s_t))] \quad (5)$$

4. Experiments and Discussion

The experiments were conducted in a SUMO–Python co-simulation environment, with the reinforcement learning component implemented in PyTorch. Three traffic flow scenarios—low, medium, and high—were defined for evaluation.

To evaluate the effectiveness of the proposed ME-PPO algorithm, which utilizes edge computing and IRS for traffic signal control, we compare it with classical timing control, DQN, PPO, and MA2C. We also conduct ablation

hereafter referred to as ME-PPO, is deployed at edge nodes together with IRS-assisted optimization, enabling adaptive real-time traffic signal control at intersections.

In the PPO algorithm, the optimization objective comprises three components: the clipped surrogate policy objective, the value function regression objective, and the entropy regularization term. The optimization problem can be formulated as:

respectively, while ϵ represents the clipping threshold that prevents excessive policy updates.

local optimum due to the limited diversity of reward signals during the learning process, an adaptive entropy regularization term is incorporated into the objective function. Let the target entropy be denoted as:

$$\hat{H} = H(\pi_\theta(\cdot | s_t)) = - \sum_{a \in A} \pi_\theta(a | s_t) \log \pi_\theta(a | s_t) \quad (4)$$

Considering both MSR and adaptive target entropy regularization, the final optimization objective of ME-PPO can be expressed as

experiments to verify the correctness and reliability of the algorithm.

To assess the individual and combined effects of edge computing and IRS on ME-PPO performance, ablation studies were conducted under four system configurations: A—without Edge or IRS; B—with edge computing only; C—with IRS only; D—with both Edge and IRS enabled. All other variables were held constant. The experimental results are as follows (Fig.5):

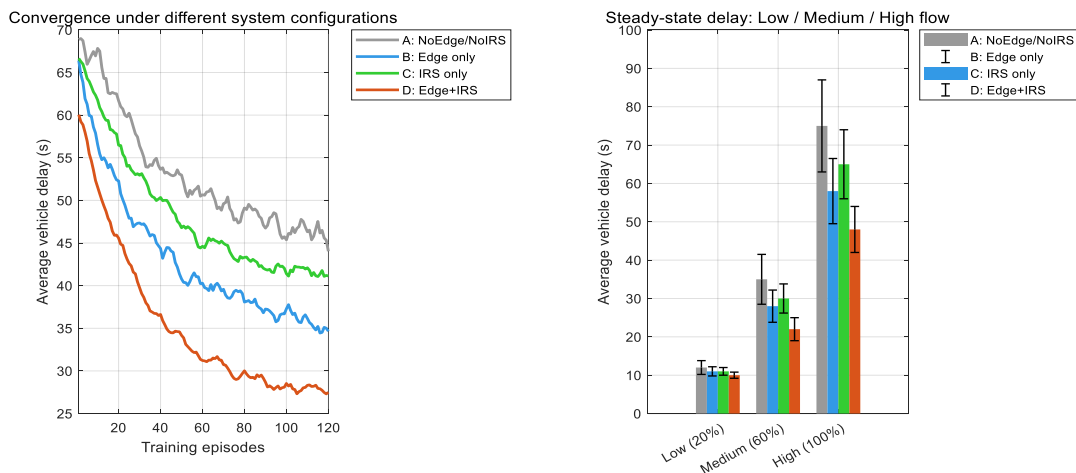


Figure 5. Comparison of Ablation Study Results

Compared with configuration A, B significantly accelerated convergence and reduced the final delay, while C

primarily reduced delay variability by enhancing communication reliability. Configuration D achieved both

low delay and high stability. Under low traffic conditions, differences among the four configurations were negligible; under medium and high traffic, B markedly improved delay and sample of D was reduced by approximately 28% relative

to A and the standard deviation of delay decreased by about 15%, indicating that the combined operation of the system modules not only enhances performance but also improves robustness.

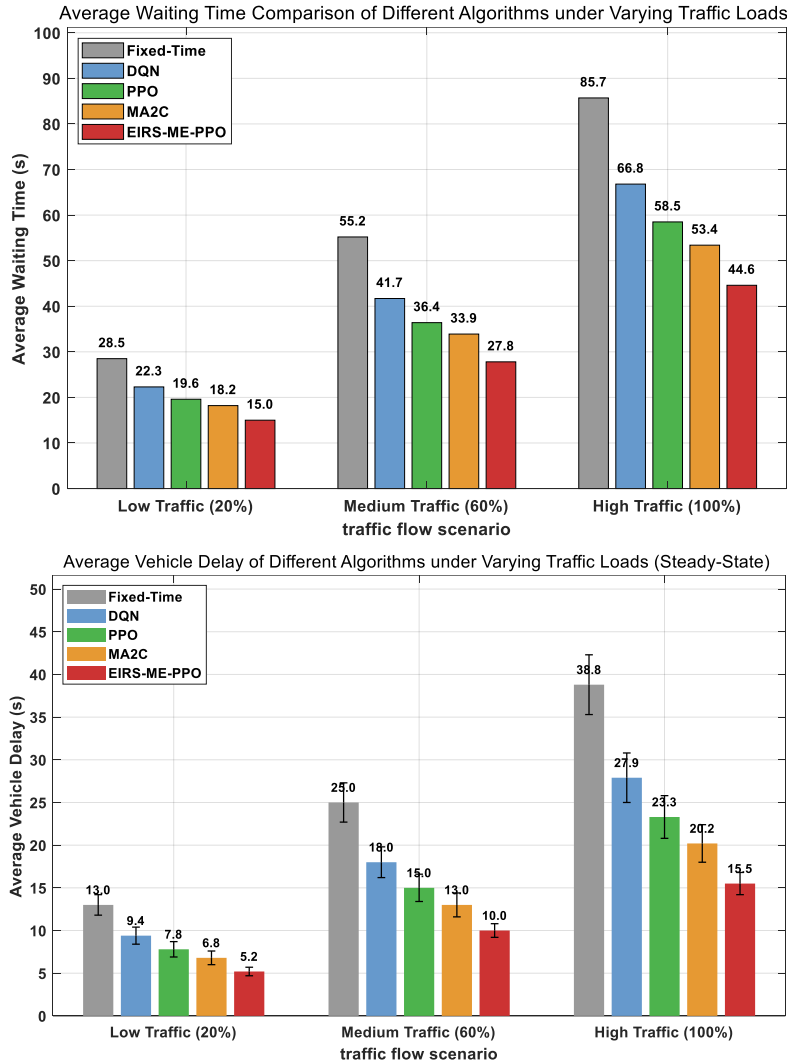


Figure 6. Average Waiting Time (Right) and Average Vehicle Delay (Left) under Three Traffic Levels

5. Conclusion

This work addresses insufficient traffic signal control efficiency in complex urban environments by proposing a reinforcement learning optimization framework integrating edge computing and intelligent reflective surfaces (IRS). Within the reinforcement learning module, multi-step returns and adaptive target entropy regularization are incorporated, further enhancing algorithm stability and exploration. To evaluate the approach, we conducted ablation studies and baseline comparisons on a co-simulation platform developed using SUMO and Python. Experimental results demonstrate that the proposed method achieves superior performance in traffic signal control, effectively reducing average delay and vehicle waiting time, improving overall intersection throughput, and validating the framework’s effectiveness.

References

- [1] Xu Dongwei, Zhou Lei, Wang Da, et al. Review of Urban Traffic Signal Control Based on Deep Reinforcement Learning [J]. *Journal of Transportation Engineering and Information*, 2022, 20(01): 15-30.
- [2] Liu Zhimin, Ye Baolin, Zhu Yaodong, et al. Traffic Signal Control Methods Based on Deep Reinforcement Learning [J]. *Journal of Zhejiang University: Engineering Edition*, 2022, 56(6): 8.
- [3] Haydari A, Yilmaz Y. Deep Reinforcement Learning for Intelligent Transportation Systems: Survey [J]. *IEEE Transactions on Intelligent Transportation Systems*. 2022, 23(1):11-32.
- [4] Liang, X., Du, X., Wang, G., Han, Z.: A deep reinforcement learning network for traffic light cycle control. *IEEE Trans. Veh. Technol.* 68, 1243–1253 (2019).
- [5] Alemzadeh, S., Moslemi, R., Sharma, R., Mesbahi, M.: Adaptive traffic control with deep reinforcement learning towards state-of-the-art and beyond. *arXiv:2007.10960* (2020).
- [6] Vidali, A., Crociani, L., Vizzari, G., Bandini, S.: A deep reinforcement learning approach adaptive traffic lights management. In: *WOA* (2019).
- [7] Mousavi SS, Schukat M, Howley E. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems* 2017, 11(7): 417-423. DOI:10.1049/iet-its.2017.0153.
- [8] Ma ZB, Cui TC, Deng WX, et al. Adaptive optimization of traffic signal timing via deep reinforcement learning. *Journal*

- of Advanced Transportation, 2021, 2021:6616702. DOI:10.1155/2021/6616702.
- [9] Huang LB, Qu XH. Improving traffic signal control operations using proximal policy optimization. IET Intelligent Transport Systems, 2023, 17(3): 592-605. DOI:10.1049/itr2.12286.
- [10] Chen Y, Chang Z, Min G, et al. Joint Optimization of Sensing and Computation for Status Update in Mobile Edge Computing Systems [J]. IEEE Transactions on Wireless Communications, 2023, 22(11): 8230-8243.
- [11] Cheng Yaodong, Tian Runxin. A Deep Reinforcement Learning-Based Computation Offloading Strategy for Edge Computing [J]. Wireless Interconnection Technology, 2024, 21(13): 19-21.
- [12] Deng RQ, Di BY, Zhang HL, et al. Reconfigurable Holographic Surface: Holographic Beamforming for Metasurface-Aided Wireless Communications [J]. IEEE Transactions on Vehicular Technology, 2021, 70(6): 6255-6259.
- [13] Zhu YS, Mao BM, Kato N. Intelligent Reflecting Surface in 6G Vehicular Communications: A Survey [J]. IEEE Open Journal of Vehicular Technology, 2022, 3: 266-277.
- [14] Gao, J., Shen, Y., Liu, J., Ito, M., Hiratori, N.: Adaptive traffic signal control deep reinforcement learning algorithm with experience replay and target network. arXiv: 1705.02755(2017).
- [15] Deep Reinforcement Q-Learning for Intelligent Traffic Signal Control with Partial Detection. International Journal of Intelligent Transportation Systems Research (2023) 21:192–206 arXiv:2109.14337.