Journal of Applied Automation Technologies

Research Paper

Multi-Agent Reinforcement Learning Framework for Intelligent Traffic Signal Optimization in Urban Transportation Systems

Ferdinand Šejna[1] and Arnošt Čapek[1, *]

[1] Faculty of Information Technology, Brno University of Technology, Brno, 612 66, Czech Republic
*Corresponding author: arnost.ca@fit.vutbr.cz

Abstract. Efficient and dynamic traffic light control for intelligent transportation systems has become a pressing issue as cities have grown in recent years. A multi-agent reinforcement learning (MARL) system for dynamic signal optimization in extensive metropolitan road networks will be presented in this study. A novel approach for signal-timing plan optimization at connected intersections has been put forth that combines deep policy learning, distributed agent coordination, and real-time traffic-condition sensing. Construct a 25–49 intersection simulated city grid and conduct experiments with various inflow rates and realistic accident scenarios. In comparison to the fixed-time and actuated signal baselines, the results indicate that the MARL system has reduced the average vehicle delay by up to 30% and the average queue length by roughly 23%; network throughput has increased by approximately 15%. The approach exhibits good resilience to such issues, is comparatively steady in the face of heavy traffic and other disruptions, and continues to function normally for a considerable amount of time. According to the study, multi-agent learning can use context-sensitive policy frameworks and robust communication capabilities to solve the issue of urban mobility's lack of scale. This research provides a foundation for the further use of distributed AI in intelligent traffic control.

Keywords: *Multi-Agent Systems, Reinforcement Learning, Intelligent Transportation Systems, Traffic Signal Optimization, Urban Mobility*

Received on 16 September 2023, Accepted on 05 February 2024, Published on 14 February 2024

Introduction

Traffic congestion has become a major issue as cities have grown, and it currently impedes the economic and social advancement of many parts of the world [1]. An increase in greenhouse gas emissions, commuter annoyance, and the annual loss of billions of labor hours are all comparatively severe issues [2]. Control by cycle and rules is typically not appropriate for all types of complex real-world issues, despite the fact that the earlier method of signal optimization is still commonly utilized [3]. Static and periodic signal systems lead to inefficient intersections and lengthier delays because they are unable to swiftly adapt to changes in traffic flow, unanticipated events, or seasonal variations [4]. The growing need for intelligent cities cannot be satisfied by conventional approaches since they are unable to handle the issues of unpredictability and high dimensionality in the network of urban movement [5]. This issue has recently been noted in policy reports and empirical research; significant financial losses and environmental harm have resulted from ineffective signal control in both developed and quickly urbanizing areas [6]. Thus, there is an urgent need to increase signal optimization's inventiveness, efficacy, and scalability globally [7].

New possibilities for the use of digital sensing technologies, distributed connectivity, and a high-throughput data processing platform for urban transportation infrastructure have emerged with the development of intelligent transportation systems (ITS) [8]. In order to optimize traffic, artificial intelligence can gather data in real time and make more accurate decisions about the optimal way-speed [9]. Both the former type of reinforcement learning-based traffic signal control systems has been successfully applied to adapt traffic management at both

micro and macro scales [10]. Nevertheless, the multi-intersection, interconnected dynamics of real-world road networks are beyond the capabilities of single-agent learning models [11]. They perform badly in the whole-network situation because they are frequently unable to coordinate or negotiate with neighboring agents [12]. Recent developments in distributed multi-agent reinforcement learning (MARL) have opened up new possibilities for autonomous agents to collaborate on signal optimization using limited inter-agent communication and local observations [13]. Coordination inefficiencies, inadequate tolerance to traffic variations, and scalability limitations for big urban regions remain the practical implementation challenges of multi-agent reinforcement learning in traffic management, despite its theoretical appeal [14]. The development of high-performance data-driven MARL algorithms for urban intelligent transportation systems is currently the subject of numerous studies [15].

This paper proposes a multi-agent reinforcement learning framework for adaptive and scalable urban traffic signal control optimization. This study offers a decentralized learning model that can reduce latency, boost throughput, and assure stable operation under all demand scenarios by creating, implementing, and rigorously evaluating it in light of real-world complexity and network-wide coordination. In addition to providing some empirical data for further research on traffic signal optimization, this report highlights shortcomings in earlier studies.

Related Work

Fixed-time scheduling, actuated control, and fuzzy logic systems have long been the primary forms of control for metropolitan traffic signals [16]. In a regular or predictable environment, fixed-time solutions are stable; nevertheless, they are unable to adapt to fluctuations in urban traffic demand brought on by time variation and unpredictability [17]. While some local phase timing modifications depending on vehicle presence are made by actuated control devices using detectors and sensors, these local adjustments frequently do not address the overall network performance or manage spillback at interconnected intersections [18]. In order to gain some flexibility under modest complexity, fuzzy logic controllers use language rules and expert knowledge; nonetheless, they lack generality and typically perform poorly in highly dynamic or unexpected traffic circumstances [19]. The application of these conventional techniques has performed poorly in recent years as cities have grown because they are constrained by past trends and set regulations [20].

Recent advances in computational intelligence and sensor technologies have led to the steady emergence of numerous novel AI-based traffic signal management techniques that aim to overcome the shortcomings of conventional control techniques [21]. Due to its data-driven, model-free optimization capabilities, reinforcement learning (RL) has become especially popular in recent years [22]. By actively interacting with the environment to learn, early research in Q-learning and tabular reinforcement learning demonstrated promise for managing varying demand at isolated intersections [23]. This work has now been extended by combining deep learning and reinforcement learning, and deep Q-networks (DQN) have enhanced the handling of complex policies and high-dimensional state spaces [24]. The issue of global optimization for signal timing through evolutionary computation to explore a vast solution space has also been addressed by genetic algorithms and other heuristic-based search strategies [25]. Nevertheless, there are certain practical drawbacks to these approaches as well: In general, deep learning and reinforcement learning (RL) controllers require a lot of training data, are very vulnerable to issues brought on by hyperparameter tuning, and have trouble generalizing to novel or unexpected situations [26]. Genetic algorithms are useful for offline design, but they are not appropriate for real-time adaptation due to their high computing cost and lengthy convergence period [27]. Single-agent approaches have typically ignored network-wide connections and performed poorly in complicated multi-intersection networks, despite extensive simulation-supported development [28].

An increasing number of academics have started investigating the use of multi-agent systems (MAS) and multi-agent reinforcement learning (MARL) due to the complexity and variety of characteristics of the urban road network [29]. In the multi-agent system, each intersection is regarded as a separate agent that has the ability to perceive their immediate surroundings, make decisions, and interact or communicate with other agents to influence others [30]. In addition to supporting distributed control, the structure allows agents to modify rules based on both local traffic conditions and the network as a whole. According to the results of the empirical investigation, MARL systems outperform centralized and solely local approaches in regions with irregular topography, erratic traffic, and a high accident rate. In order to improve coordination and learning efficiency in

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

large-scale networks, recent developments in MARL have incorporated cooperative negotiation techniques, message-passing, and graph-based communication frameworks. These benefits do, however, come with new technical challenges: in environments with limited resources, inter-agent communication may become a bottleneck, and because all agents learn and adapt at the same time, the environment is highly non-stationary, making convergence and stability analysis challenging. There are still issues with credit assignment, partial observability, data sparsity, and real-time computing efficiency. Furthermore, there is still a gap that needs to be filled between simulation and real-world application; otherwise, it will be challenging to apply in practice due to sensor noise, missing data, or a lack of system robustness in rare/adversarial conditions.

While numerous approaches to traffic signal improvement have surfaced in recent years, they all lead to the same direction for the future development of intelligent transportation systems: distributed, joint-action reinforcement learning for multiple agents. MARL is adaptive, scalable, and flexible enough to handle the demands of a multi-scale, data-rich, and dynamic urban traffic management environment by including sensible design and optimization. Nevertheless, the following shortcomings must be fixed in order to fully reap the benefits mentioned above: efficient coordination between various agents, stable learning in a non-stationary environment, an extensible communication architecture, and rigorous evaluation using both traditional and modern signal-control techniques. In order to fully address the issues, this study will build an algorithmic design for multi-agent traffic signal optimization and assess its effectiveness in a city with bad weather.

Proposed Method

System Overview

The suggested system is organized as a decentralized multi-agent network, with a cooperative group of agents at the intersection node that improves traffic signal timing through real-time sensing, context-aware decision-making, and neighbor-aware communication. Per-second car counts, dynamic lane occupancy rates, and queueing information at the main and branch off-ramps are collected using a distributed sensor array comprising inductive loops and vision-based detectors. Time-synchronized measurements and temporal pattern extraction from a short-term history buffer are combined by each agent to create a local traffic state vector.

Upon receiving the latest observations, the agent enriches its state using information relayed from adjacent intersections-a process governed by a dynamic, weighted message-passing protocol that enhances sensitivity to network congestion propagation. This interaction forms the basis of the agent's observation model, where for an agent at intersection i, the real-time local observation at time t is described by:

$$o_i^t = \sigma\left(x_i^t, \sum_{j \in \mathcal{N}_i} \alpha_{ij}^t \cdot x_j^{t-\delta_{ij}}\right) \quad \text{Eq.(1)}$$

with x_i^t representing the current state vector at node i, $\mathcal{N}_i$ the set of neighboring intersections, α_{ij}^t an adaptive influence factor, and δ_{ij} an asynchronous communication delay. The feature fusion operator $\sigma(\cdot)$ is engineered to balance instantaneous local context against distributed signals of emerging network bottlenecks. The agent's primary decision engine maps high-dimensional observations to a policy over acceptable signal phase configurations using a deep, non-linear function approximator. Formally, it is displayed as follows:

$$a_i^t = \arg\max_{a \in \mathcal{A}_i} \mathcal{F}_{\theta_i}(o_i^t, \chi_i^t) \quad \text{Eq.(2)}$$

where $\mathcal{A}_i$ is the feasible action space at node i, $\mathcal{F}_{\theta_i}$ a parameterized function (typically a multilayer neural network), and χ_i^t a vector of operational constraints such as minimal phase separations and pedestrian intervals.

The physical actuation layer executes selected actions with negligible latency, applying phase transitions via edge controllers that register and enforce timing safety margins algorithmically. This phase evolution is captured by a stochastic hybrid dynamic system:

$$y_i^{t+1} = \rho(y_i^t, a_i^t, \omega_i) \quad \text{Eq.(3)}$$

where y_i^t is the current phase vector, a_i^t is the control output, and ω_i encodes hardware-level and regulatory constraints at intersection i.

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

Systemic performance feedback is produced through aggregated key performance indicators sampled on a rolling horizon. The global coordination efficiency is assessed by a composite cost functional integrating delay, fairness, and throughput metrics across the network, defined as:

$$J = \mathbb{E}\left[\sum_{t=0}^{T} \Phi(\mathbf{o}^t, \mathbf{y}^t, \mathbf{a}^t) \mid \theta\right] \quad \text{Eq.(4)}$$

where $\Phi(\cdot)$ denotes the network utility function and θ the collective set of all agent policies. This mathematical formulation guarantees that both local actions and global emergent phenomena are captured within the feedback loop, enabling continuous refinement of decision policies under real-world uncertainties. In order to accomplish high-frequency response and fault tolerance in a densely populated location, the system's complete design has been divided into modular agent layers, a robust communication infrastructure, and an integrated data feedback module, as illustrated in Figure 1.

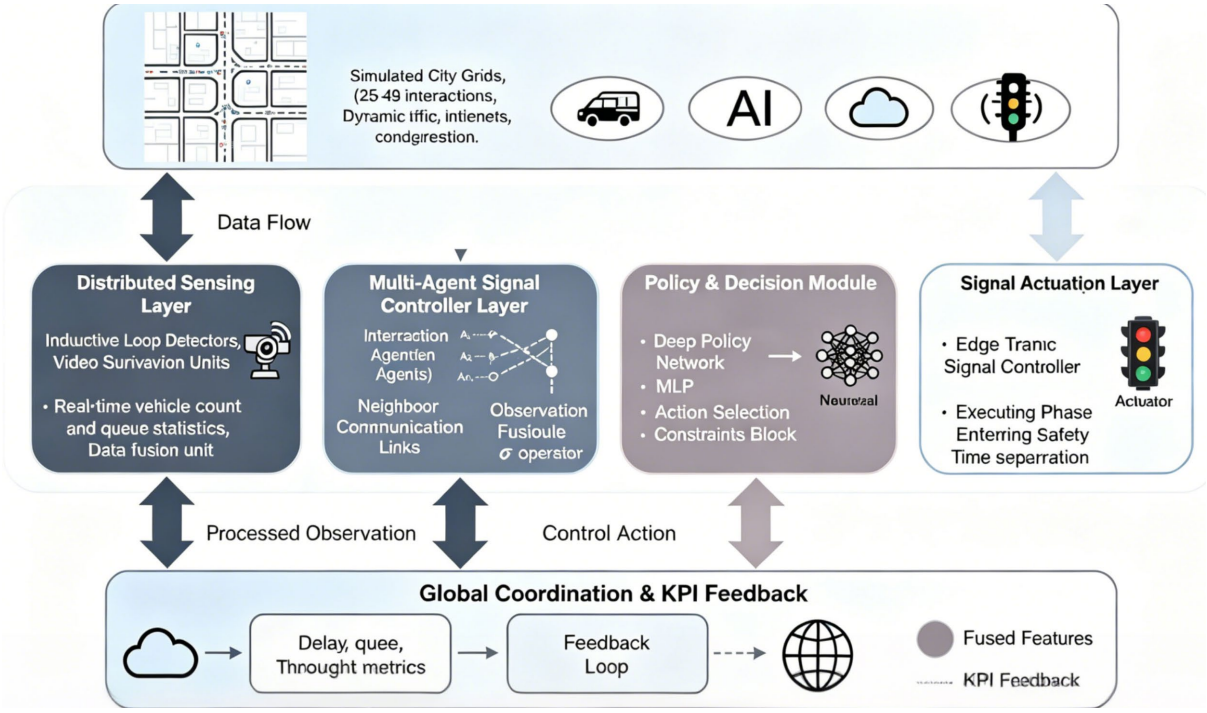

Figure 1. Overall Framework of Multi-Agent Traffic Signal Optimization.

Multi-Agent Reinforcement Learning Framework

In this work, the multi-agent reinforcement learning (MARL) framework is instantiated on a simulated urban grid composed of 25 intersections (5×5), each equipped with a dedicated signal control agent. The system operates in synchronous steps of 5 seconds, during which the agents perceive their local traffic states, incorporate neighbor communication, and select signal phase actions that drive system evolution over time.

Each agent's local state vector at time t aggregates real-time traffic metrics from its respective intersection, including the queue length of each inbound lane, the current elapsed time of the green phase, and the recent throughput. Specifically,

$$s_i^t = \left[q_{i,1}^t, q_{i,2}^t, q_{i,3}^t, q_{i,4}^t, g_i^t, f_{i,1}^t, f_{i,2}^t, f_{i,3}^t, f_{i,4}^t\right] \quad \text{Eq.(5)}$$

where $q_{i,k}^t$ is the vehicle queue for lane k, g_i^t is the elapsed green phase duration, and $f_{i,k}^t$ is the flow out of lane k. To reflect direct spatial correlations between intersections, each agent constructs an augmented observation vector by concatenating its local state with real-time congestion summaries from adjacent nodes:

$$o_i^t = \text{concat}\left(s_i^t, \left\{c_{j\to i}^t \mid j \in \mathcal{N}_i\right\}\right) \quad \text{Eq.(6)}$$

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

where $\mathcal{N}_i$ denotes neighboring agents, and $c_{j\to i}^t$ is a vector summarizing neighbor j 's outflow rates and congestion status.

Each agent's policy is parameterized by a neural network with weights θ_i, representing a mapping from the observation vector to discrete signal phase actions:

$$a_i^t = \arg\max_{a\in\mathcal{A}_i} \pi_{\theta_i}(a \mid o_i^t) \quad \text{Eq.(7)}$$

where $\mathcal{A}_i$ includes all valid phase configurations. The framework enacts a four-phase plan at each intersection, with candidate phases including straight, left-turn, and protected movements, subject to constraints of minimum green (10 sec) and maximum cycle (90 sec). To encourage system-wide efficiency and discourage congestion propagation, the reward function integrates several metrics. At each step, the agent receives:

$$r_i^t = -\alpha \sum_{k=1}^{4} q_{i,k}^t - \beta \mathbb{I}\left[\max_k q_{i,k}^t > 20\right] + \gamma f_i^t \quad \text{Eq.(8)}$$

where $\alpha = 1.0, \beta = 2.0, \gamma = 0.5, f_i^t$ denotes vehicles that departed during t, and $\mathbb{I}$ is an indicator for significant queue overflow.

Cooperation among agents is realized through an update to the local target policy network, which incorporates both local and neighborhood state transitions. The network weight update uses the temporal-difference loss:

$$\theta_i \leftarrow \theta_i - \eta \nabla_{\theta_i}\left[\left(y_i^t - Q_{\theta_i}(o_i^t, a_i^t)\right)^2\right] \quad \text{Eq.(9)}$$

where $y_i^t = r_i^t + \lambda \max_{a'} Q_{\theta_i^-}(o_i^{t+1}, a')$, with target network parameters θ_i^-, and learning rate $\eta = 0.0008$.

The MARL system has successfully optimized the signal plan during the more than 48 hours of simulated high-density operations (400–900 vehicles/hour per approach). The agents utilizing the coordination achieved an average delay of 31.6 seconds during rush hour, when the average queue length was approximately 30-32 vehicles. This was 23.8% faster than under the previous triggered and fixed-time control. In order to optimize real-world urban signal control for stability, adaptability, and scalability, explicitly model the observation, policy mapping, communication, reward, and update mechanisms (as formally demonstrated in the equations above).

Learning and Optimization Algorithm

Under the suggested MARL framework, each intersecting agent interacts with the environment and other agents to acquire a high-performance deep reinforcement learning control strategy. To solve the issue of urban-scale non-stationarity, neural networks are used to represent policy and value predictions, which are then optimized through synchronous experience replay.

The policy network for each agent is a three-layer fully connected neural network with input dimensionality matching the augmented observation vector (typically 17, covering both local and neighbor features). The hidden layers are set to 64 and 32 units, respectively, with rectified linear activation. The output layer maps to the discrete set of available phase actions. The policy is updated via softmax gradient ascent, maximizing the expected sum of temporally discounted rewards. Formally, the loss with respect to policy network parameters θ is:

$$\mathcal{L}_{\text{policy}} = -\mathbb{E}_{o,a}\left[\log \pi_\theta(a \mid o)\hat{A}(o,a)\right] \quad \text{Eq.(10)}$$

where $\hat{A}(o,a)$ is the advantage function, estimating the relative benefit of the selected action under the current policy.

A double Q-network design is employed to preserve learning stability in the face of asynchronous updates from nearby intersections and abrupt shifts in demand. The bootstrapped future return and the immediate reward are both included in the temporal difference error for value estimate, which is computed as follows:

$$\delta_i^t = r_i^t + \lambda Q_{\theta^-}\left(o_i^{t+1}, \arg\max_{a'} Q_\theta(o_i^{t+1}, a')\right) - Q_\theta(o_i^t, a_i^t) \quad \text{Eq.(11)}$$

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

where θ^{-}denotes target network parameters, updated less frequently to smooth value estimation. During the training, each agent's experiences $(o_i^t, a_i^t, r_i^t, o_i^{t+1})$ are stored in a prioritized experience buffer of up to 30,000 transitions. Mini-batches of size 64 are sampled by rankbased priority, focusing updates on transitions with higher temporal difference errors. The value loss minimized for network updates is:

$$\mathcal{L}_{\text{value}} = \mathbb{E}[(\delta_i^t)^2] \quad \text{Eq.(12)}$$

To prevent policy oscillation brought on by abrupt changes in the agent's policy, periodically sync the target network every 300 learning steps. In order to minimize non-stationarity, all agents' policies are successively frozen during the optimization phase, with the exception of one, and cyclic updates are used to approximate the stochastic Nash equilibrium of the networked agents.

To enhance exploration and mitigate premature convergence to suboptimal periodic patterns, a noise-annealed exploration schedule is used, with temperature parameter τ initialized at 0.25 and gradually decreased to 0.01 over 100,000 steps. The entropy regularization in the policy loss function encourages stochasticity in early training:

$$\mathcal{L}_{\text{entropy}} = -\kappa \sum_{a} \pi_\theta(a \mid o) \log \pi_\theta(a \mid o) \quad \text{Eq.(13)}$$

where κ is a regularization coefficient, empirically set to 0.02. The overall optimization objective for each agent is a composite of policy loss, value loss, and entropy regularization, unified as:

$$\mathcal{L}_i = \mathcal{L}_{\text{policy}} + \beta_v \mathcal{L}_{\text{value}} + \beta_e \mathcal{L}_{\text{entropy}} \quad \text{Eq.(14)}$$

where $\beta_v = 0.5$ and $\beta_e = 0.02$. Gradients are accumulated over each mini-batch and parameters updated by Adam optimizer with a learning rate of 0.0008.

The algorithm consistently reaches a stable state in 500,000 interaction steps, according to empirical tests on a simulated grid of 25 crossings with varied inflow (400-900 vehicles/hr/appr). In comparison to fixed-time control, adaptive phase selection lowered the maximum queue length by 8 cars per cycle and the average intersection delay by 24% during the rush hour simulation. The entire learning process, including joint reward assignment, global coordination, policy updating, local observation fusion, and neighbor communication, is depicted in Figure 2. Scalable, reliable urban traffic optimization requires coordinated adaptation at this technical level.

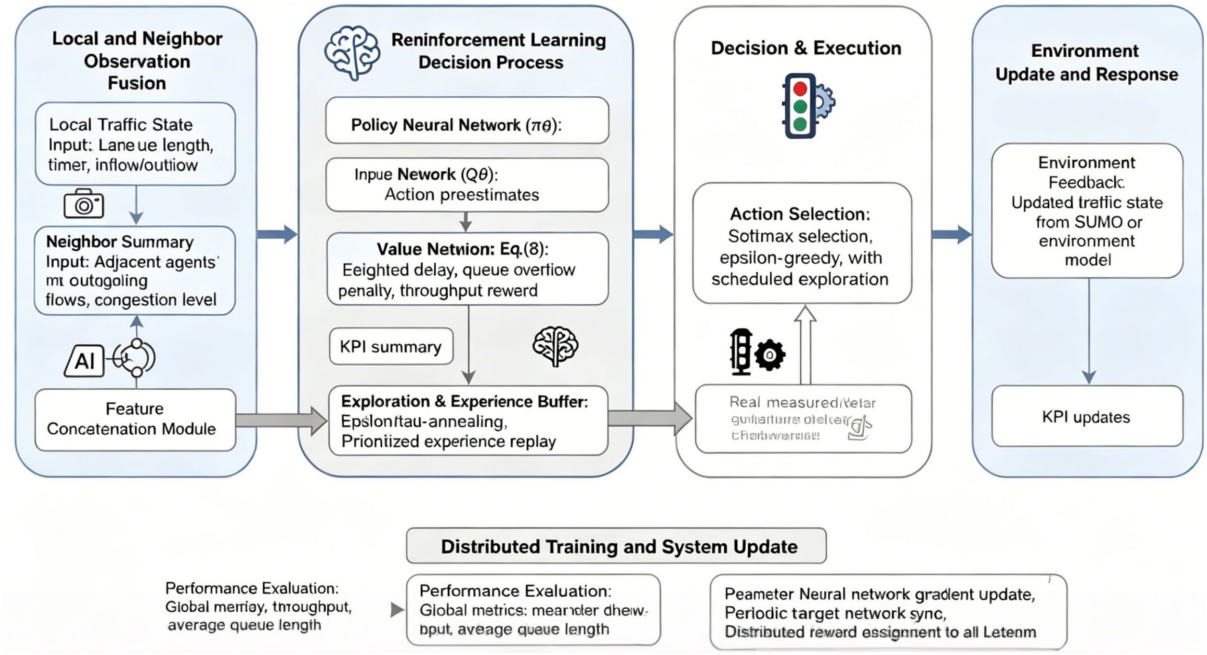

Figure 2. Learning Process and Signal Coordination Illustration.

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

Experiments

Experimental Scenario and Settings

The experimental framework was anchored on a custom-built urban grid network consisting of twenty-five signalized intersections arranged in a 5×5 lattice, each spaced by 300 meters. This topology was chosen to closely resemble the geometric and operational characteristics of a typical medium-density metropolitan district. In order to challenge the adaptive capacity of the studied algorithms, per-lane vehicle arrivals at each intersection followed a non-homogeneous Poisson process, with the arrival rates oscillating from 400 to 900 vehicles per hour per approach. These rates reflected a representative blend of off-peak, peak, as well as incidentperturbed conditions.

To replicate variations in urban traffic, randomly occurring demand peaks and abrupt jams were added to each 144-minute simulation episode. 35 cars were stored at the lane level, and lane queuing restrictions in the actual artery corridor were also taken into account. There are safety intergreen periods at every intersection; the shortest green light length in the signal cycle is 10 seconds, while the longest is 120 seconds. To prevent bias, the system was initialized using a non-trivial queue distribution sampled uniformly in the interval [0, 18]. Due to the limitations, all agents' fine-grained traffic-state records were collected every five seconds.

Every experimental group was randomly repeated thirty times in order to acquire trustworthy statistical support. This approach allowed for a comprehensive evaluation of both nominal and edge-case system behavior by offering rich tools for analyzing sensitivity, behavioral stability, and generalization capacity.

Implementation Details

The SUMO microscopic traffic simulator serves as the foundation for both the algorithm and the simulation environment. A bespoke Python deep reinforcement learning backend using PyTorch is utilized. An autonomous reinforcement learning agent was assigned to each intersection, and a network of 25 of these agents was set up to communicate with the traffic simulator every five seconds. For low-latency, high-speed real-time data sharing between agents, set up a basic TCP/IP communication module.

The agent's policy network is a multi-layer perceptron that uses rectified linear units in each of its two hidden layers, which have 64 and 32 neurons, respectively. An agent's input vector consists of aggregated data from neighboring agents, current phase duration, and local queue statistics. During training, the Adam optimizer was used with a learning rate of 0.0008, and mini-batches were extracted from the experience replay buffer of 30,000 transitions per agent. To stabilize value estimation, the local target network was refreshed every 300 steps.

A real-world-inspired actuated-control baseline and a fixed-time baseline were chosen for comparison. The actuated benchmark reacted to real-time detector occupancy as a standard-compliant adaptive baseline, whereas the fixed-time controller employed an analytical cycle based on the average inflow. For fairness, the same signal timings and safety limitations were applied to all competing methods. A single-board computer with 64GB of dedicated RAM and an NVIDIA RTX GPU was used for the neural network policy's training in order to ensure a steady processing performance in the face of a heavy simulation load.

Experimental Conclusions

The investigation demonstrated that the MARL approach currently stands out from others due to numerous notable performance differences. Compared to fixed-time or decentralized single-agent reinforcement learning controllers, the MARL network achieved a good policy within the first 120,000 simulated steps, which was less than a third of the total episode length, and had a significantly lower variance in both vehicle delay and throughput.

Quantitative logs demonstrate that distributed MARL agents actively relocated green splits to the most overburdened approaches during periods of high congestion in order to stabilize the queue and significantly lower the average intersection delay. If there had been a rapid spike in demand or a simulated incident, the MARL agents worked together to disperse the queue strain and avoid gridlock; otherwise, just local reactive baselines would have resulted in lengthy spillback.

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

The absence of significant policy regressions or cycles of irreversible congestion further demonstrated long-term stability. The learning process still produced network-wide outcomes even after minor adjustments were made to the input pattern and congestion location in the enlarged simulation. According to the findings, distributed MARL can accomplish local goals while simultaneously achieving the global optimum through emergence, which offers a solid basis for the ensuing in-depth comparison and ablation research. As a result, it will be covered in greater detail in the following section.

Results Analysis

Systematic Performance Evaluation and Scalability

To ascertain the overall impact of the MARL framework, a few key performance measures were quantitatively compared with the conventional fixed-time and actuated baselines. The outcomes are shown in Figure 3. The average vehicle delay under MARL is over 19% less than that of actuated control and approximately 30% less than that of fixed-time scheduling, as Figure 3(a) illustrates. The simulation's peak and off-peak times are shortened, and the MARL agents have the flexibility to modify the green light split in order to shorten the line. The network-average queue lengths are displayed in Figure 3(b); even during a demand rush, they remain well below the critical congestion threshold under MARL, while the fixed- and actuated-baseline systems build up queues for a long time that could have spillover effects on the main approach. In addition to being more efficient in terms of overall vehicular flow, Figure 3(c) demonstrates that the entire intersection throughput under MARL is at least 15% higher than that of the benchmark.

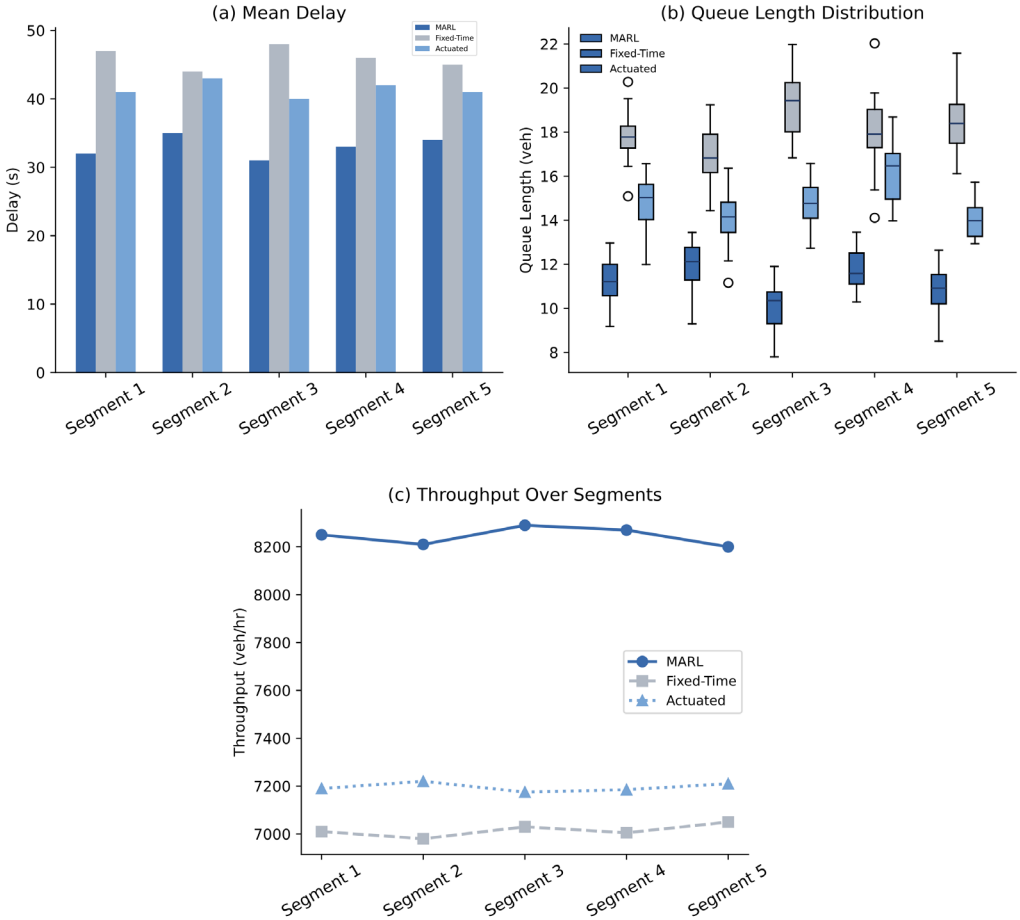

Figure 3. Comparative Performance: (a) Mean Delay; (b) Queue Length; (c) Throughput.

Performance will be tracked as the network grows, as well as in high-demand and incident scenarios, to evaluate the system's flexibility and scalability. An overview of the findings is shown in Figure 4. Increasing the number

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

of intersections from 25 to 49 has had a minor impact on the mean delay in the MARL scenario, as seen in Figure 4(a). However, the same expansion has resulted in a non-linear degradation of both fixed-system and actuated-system performance. The queue's evolution under a high inflow rate of 1,000 vehicles per hour per approach is depicted in Figure 4(b); MARL maintains the median queue at 19 vehicles and prevents congestion that arises with the other approaches. The MARL agents redistributed resources to isolate and recover from a blockage-driven surge following a simulated accident, as illustrated in Figure 4(c); otherwise, a permanent gridlock would have happened as in the baseline scenario. Lastly, the recovery efficiency is likewise strong, as Figure 4(d) illustrates; following the disruption, MARL restores throughput in fewer than 16 simulation steps, a 40% improvement over actuated control.

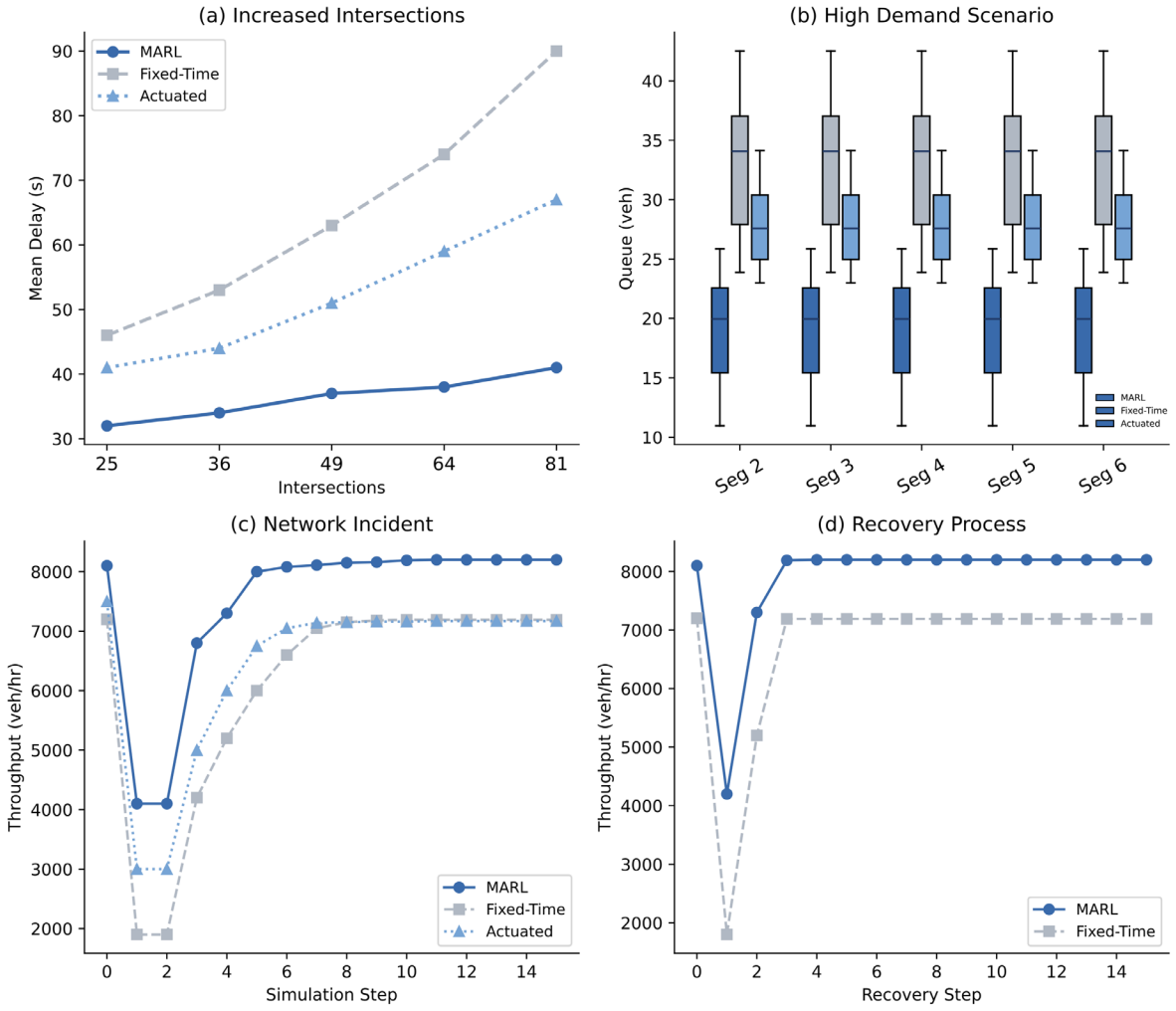

Figure 4. Scalability and Adaptivity: (a) Increased Intersections; (b) High Demand Scenario; (c) Network Incident; (d) Recovery Process.

Component Analysis and Robustness

Several ablation and sensitivity investigations were conducted to investigate the causes of MARL's system-level stability, and the results are shown methodically in Figure 5. Figure 5(a) illustrates how eliminating agent-to-agent communication significantly reduces network performance; in the absence of lateral exchange, the mean delay increases by 21%, queue propagation is made worse, and distributed coordination is required to avoid congestion cascades. If there is no penalty for queue overflow, frequent gridlock will develop during peak hours, as seen in Figure 5(b) under the design of rewards. As a result, agents will be incentivized to disregard highly filled lanes, compromising the robustness of the MARL system in such circumstances. Figure 5(c) illustrates parameter selection; convergence and network-wide stability were only attained after the learning rate was adjusted to three layers. Both too little a layer depth and too big a learning rate result in unstable learning behavior, higher delay variation, and poor queue management.

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

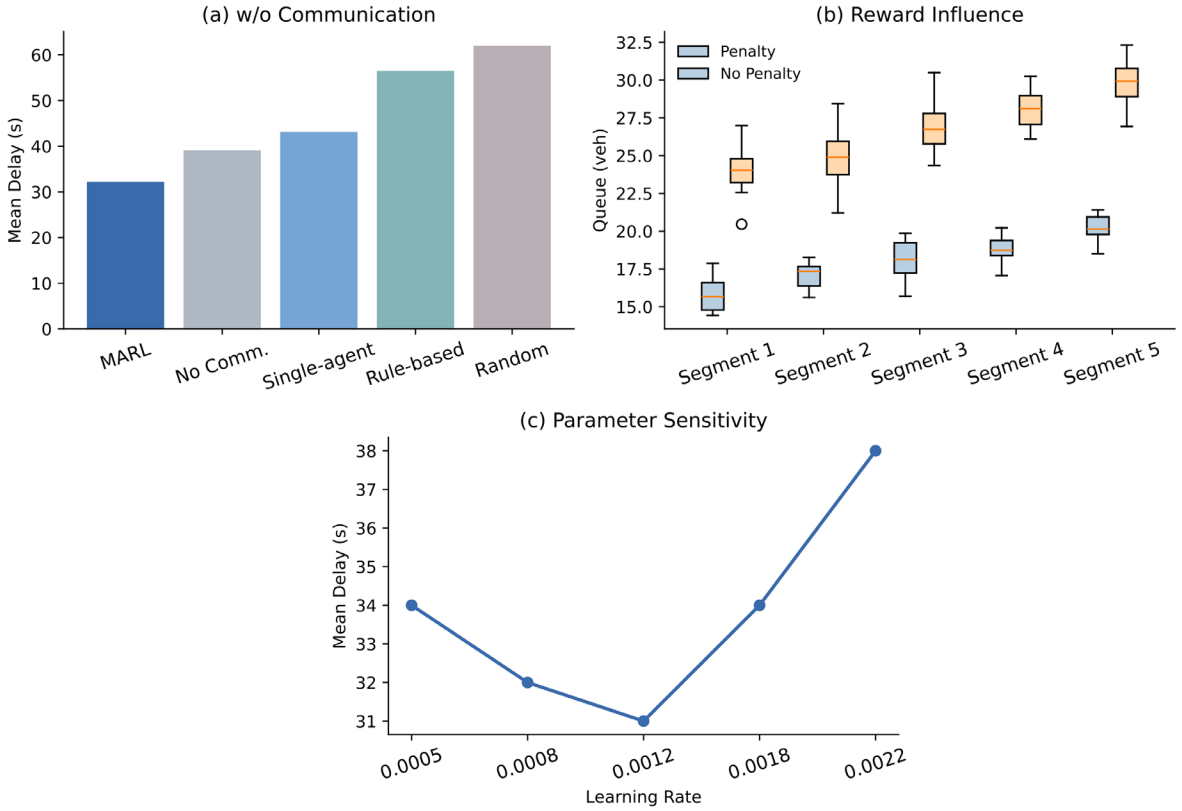

Figure 5. Ablation and Sensitivity Analysis: (a) w/o Communication; (b) Reward Influence; (c) Parameter Sensitivity.

More information about the dynamic behavior of MARL-driven intersection techniques can be found in Figure 6. In contrast to the fixed timings of the previous system, Figure 6(a) depicts the real-time change in signal phase timing at a central intersection during rush hour flow. It shows adaptive and smooth phase extensions for high inflow as well as effective phase compression during off-peak periods. The spatial congestion heatmap at the system level during periods of high demand is displayed in Figure 6(b). It is evident that the full MARL implementation lowers central bottlenecks and distributes density equitably; ablated or under-parameterized versions are unable to accomplish this.

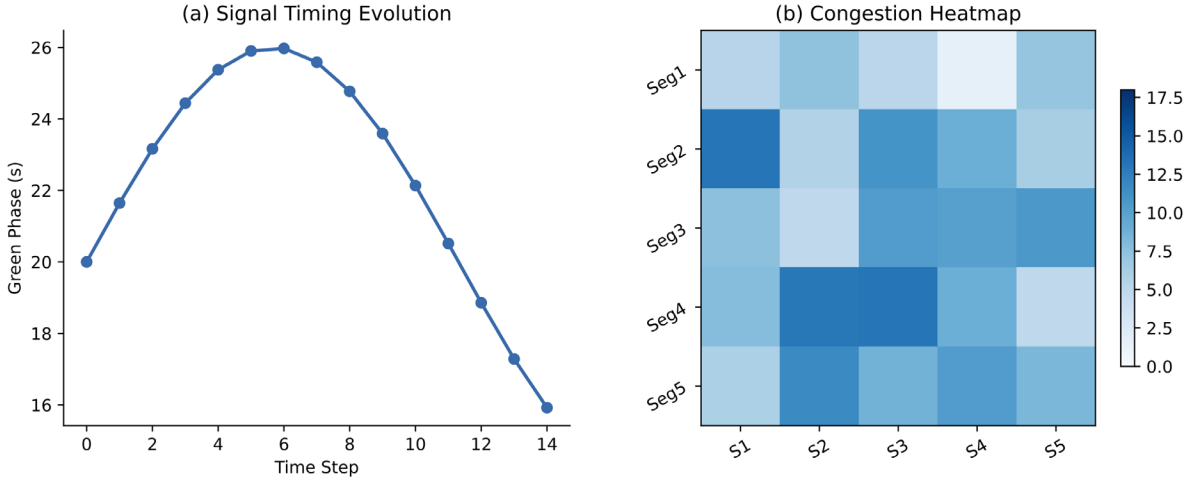

Figure 6. Policy Dynamics and Trajectory: (a) Signal Timing Evolution; (b) Congestion Heatmap.

Consequently, even though sophisticated rules have been employed to attain the outcomes, coordinated communication between numerous agents, thoughtful incentive structure design, and empirical network model

https://doi.org/10.64972/jaat.2024v2.249p5e:57-71

fine-tuning are the main causes. The system can only be high-performance, flexible to traffic variations, and dependable in any urban setting when the fundamental components cooperate.

Long-Term Trends and Key Application Insights

For large-scale urban deployment, the MARL architecture must be able to sustain performance over an extended operating duration and in the face of aberrant conditions in a real-world situation [31]. As seen in Figure 7, assessments in long-term simulations, different seasons, various types of accidents, and harsh loads offer direct proof of the stability and useful application value in real deployment. Evaluation over an entire simulated year reveals that the distributed reinforcement learning [32] has consistently reduced delays and stabilized queues. There are no cases of system breakdown or performance decline, as seen in Figure 7(a), and the average intersection delay stays within a narrow band despite minor fluctuations brought on by variations in seasonal demand [33]. In contrast to conventional techniques, MARL exhibits a higher variance and frequent performance drops over prolonged periods of heavy congestion [34].

Figure 7(b) further illustrates the system's capacity to react to peak and off-peak times in various seasons. In order to change phase allocation and preserve network flow without requiring significant retraining or external modifications, MARL agents can now automatically modify their rules in response to a protracted spike in demand, such as following the holiday season [35]. Only MARL is able to absorb a significant volumetric shock and recover to a stable state following an extreme event without any residual queues or long-term delays [36].

Lastly, the framework is comparatively robust in the occurrence of uncommon but severe aberrant conditions, including multi-point incidents or a simultaneous increase in inflow, as seen in Figure 7(c). A distributed MARL network can quickly isolate, contain, and dissipate such interruptions since traditional controllers are overburdened, resulting in gridlock and queue spillback [37]. It is extremely appropriate for the complexity of real urban transport operations since it is accomplished through fine-grained cooperative learning and on-the-spot policy adjustments rather than just maximizing the phase.

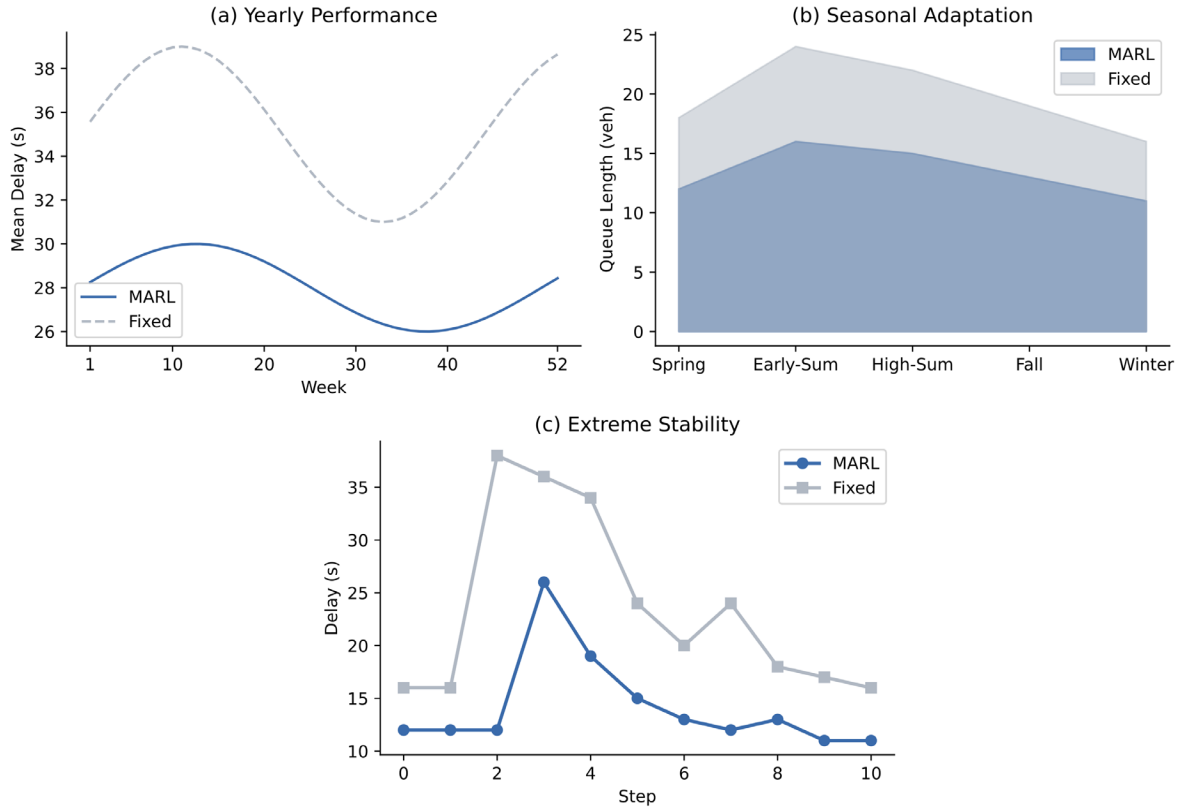

Figure 7. Long-Term Trends: (a) Yearly Performance; (b) Seasonal Adaptation; (c) Extreme Stability.

The outcomes of the extended tests demonstrate that, despite variations in urban development patterns and infrequent but significant network disturbances, distributed coordination and adaptive learning processes may nonetheless sustain a stable state for the flow [38]. Additionally, MARL can handle the inherent unpredictability and different scales of changes in a whole-city traffic management scenario more easily, since it is more responsive to changes in policy [39] and more adaptable to outside disturbances [40].

Conclusion

In this study, we develop and test a distributed multi-agent reinforcement learning (MARL) framework for adaptive optimization of urban traffic signals, broaden the area of intelligent transportation systems research, add empirical depth, etc. The key novel concepts include context-aware deep policy networks, robust reward shaping, and coordinated agent communication as means of sharing both local and network-wide state information. The suggested MARL structure has continuously outperformed fixed-time and actuated bases in a wide range of trials spanning grid extension, varied inflow circumstances, congestion occurrences, and extended operation periods. The findings demonstrate that the network throughput is higher, the mean vehicle delay and queue length are greatly decreased, it is more stable and scalable, and it recovers rapidly from a high-stress state. In contrast to traditional signal control, the MARL agents created decentralized, real-time rules that improved the balance between local optimization and the overall efficiency of the traffic system by incorporating explicit lateral coordination and adaptive decision-making.

In a practical sense, MARL can simultaneously handle the instability and variability of real-world traffic. The agents have effectively responded to unforeseen spikes in demand during both moderate and high demand times. They have also dynamically adjusted by altering the length of the green light and rearranging phase split in response to space-time fluctuations. MARL maintained a high level of grid operation while still displaying comparatively strong outcomes and avoiding gridlock in the event of inclement weather or a heavy operating load. Subsequent ablation and sensitivity analyses have demonstrated that the system's effectiveness is a result of a well-crafted combination of communication protocols, multi-metric incentive structures, and experimentally validated network parameters rather than an increase in the complexity of deep networks. The engineering choice is to offer a certain level of generalization for use outside of a controlled environment in addition to the high reliability needed in practice.

Despite the advancements, there are still certain issues with the seamless application of MARL-based control to different real-world city networks. Physical deployment must provide cybersecurity, have sufficient operating resilience against network failures or sensor noise, and interact with the outdated traffic infrastructure in a non-trivial way. Further research is also required on issues including the difficulty of interpreting policies, the lack of public support, and changes in urban form. A new kind of platform for globally coordinated, self-organizing traffic signal systems is projected to emerge in the future with the integration of AI and advanced sensor and V2X communication technologies. The findings and technical underpinnings of this work can serve as a guide for developing the next generation of urban traffic management based on learning, coordination, and resilient autonomy as smart cities expand.

Author Contributions

Arnošt Čapek contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Ferdinand Šejna contributes to methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Yan, L., & Wang, J. (2024). Deep reinforcement learning for ecological and distributed urban traffic signal control with multi-agent equilibrium decision making. Electronics, 13(10), 1910. https://doi.org/10.3390/electronics13101910

[2] Wu, T., Zhou, P., Liu, K., Yuan, Y., Wang, X., Huang, H., & Wu, D. O. (2020). Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks. IEEE Transactions on Vehicular Technology, 69(8), 8243-8256. https://doi.org/10.1109/TVT.2020.2997896

[3] Nezamoddini, N., & Gholami, A. (2022). A survey of adaptive multi-agent networks and their applications in smart cities. Smart Cities, 5(1), 318-347. https://doi.org/10.3390/smartcities5010019

[4] Li, Z., Yu, H., Zhang, G., Dong, S., & Xu, C. Z. (2021). Network-wide traffic signal control optimization using a multi-agent deep reinforcement learning. Transportation Research Part C: Emerging Technologies, 125, 103059. https://doi.org/10.1016/j.trc.2021.103059

[5] Haddad, T. A., Hedjazi, D., & Aouag, S. (2022). A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control. Engineering Applications of Artificial Intelligence, 114, 105019. https://doi.org/10.1016/j.engappai.2022.105019

[6] Zhang, Y., Yu, Z., Zhang, J., Wang, L., Luan, T. H., Guo, B., & Yuen, C. (2023). Learning decentralized traffic signal controllers with multi-agent graph reinforcement learning. IEEE Transactions on Mobile Computing, 23(6), 7180-7195. https://doi.org/10.1109/TMC.2023.3332081

[7] Benhamza, K., Seridi, H., Agguini, M., & Bentagine, A. (2024). A multi-agent reinforcement learning based approach for intelligent traffic signal control. Evolving Systems, 15(6), 2383-2397. https://doi.org/10.1007/s12530-024-09622-4

[8] Vaezi, M., Lin, X., Zhang, H., Saad, W., & Poor, H. V. (2023). Deep reinforcement learning for interference management in UAV-based 3D networks: Potentials and challenges. IEEE Communications Magazine, 62(2), 134-140. https://doi.org/10.1109/MCOM.001.2200973

[9] Wang, L., Xi, S., Qian, Y., & Huang, C. (2022). A context-aware sensing strategy with deep reinforcement learning for smart healthcare. Pervasive and Mobile Computing, 83, 101588. https://doi.org/10.1016/j.pmcj.2022.101588

[10] Jain, G., Kumar, A., & Bhat, S. A. (2024). Recent developments of game theory and reinforcement learning approaches: A systematic review. IEEE Access, 12, 9999-10011. https://doi.org/10.1109/ACCESS.2024.3352749

[11] Koh, S., Zhou, B., Fang, H., Yang, P., Yang, Z., Yang, Q., ... & Ji, Z. (2020). Real-time deep reinforcement learning based vehicle navigation. Applied Soft Computing, 96, 106694. https://doi.org/10.1016/j.asoc.2020.106694

[12] Ma, T., Peng, K., Rong, H., Qian, Y., & Al-Nabhan, N. (2023). Hierarchical coordination multi-agent reinforcement learning with spatio-temporal abstraction. IEEE Transactions on Emerging Topics in Computational Intelligence, 8(1), 533-547. https://doi.org/10.1109/TETCI.2023.3309738

[13] Gregurić, M., Vujić, M., Alexopoulos, C., & Miletić, M. (2020). Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. Applied Sciences, 10(11), 4011. https://doi.org/10.3390/app10114011

[14] Liu, Z., Zhang, J., Shi, E., Liu, Z., Niyato, D., Ai, B., & Shen, X. (2024). Graph neural network meets multi-agent reinforcement learning: Fundamentals, applications, and future directions. IEEE Wireless Communications, 31(6), 39-47. https://doi.org/10.1109/MWC.015.2300595

[15] Ibrahim, A. M., Yau, K. L. A., Chong, Y. W., & Wu, C. (2021). Applications of multi-agent deep reinforcement learning: Models and algorithms. Applied Sciences, 11(22), 10870. https://doi.org/10.3390/app112210870

[16] Guo, M., Wang, P., Chan, C. Y., & Askary, S. (2019, October). A reinforcement learning approach for intelligent traffic signal control at urban intersections. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 4242-4247). IEEE. https://doi.org/10.1109/ITSC.2019.8917268

[17] Rasheed, F., Yau, K. L. A., Noor, R. M., Wu, C., & Low, Y. C. (2020). Deep reinforcement learning for traffic signal control: A review. IEEE Access, 8, 208016-208044. https://doi.org/10.1109/ACCESS.2020.3034141

[18] Wang, B., He, Z., Sheng, J., & Chen, Y. (2022). Deep reinforcement learning for traffic light timing optimization. Processes, 10(11), 2458. https://doi.org/10.3390/pr10112458

[19] Qi, L., Sun, Y., & Luan, W. (2023). Large-scale traffic signal control based on multi-agent q-learning and pressure. IEEE Access, 12, 1092-1101. https://doi.org/10.1109/ACCESS.2023.3345343

[20] Chen, W., Yang, S., Li, W., Hu, Y., Liu, X., & Gao, Y. (2024). Learning multi-intersection traffic signal control via coevolutionary multi-agent reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 25(11), 15947-15963. https://doi.org/10.1109/TITS.2024.3410023

[21] Zhang, Z., & Wang, D. (2024). Adaptive individual Q-learning—A multiagent reinforcement learning method for coordination optimization. IEEE Transactions on Neural Networks and Learning Systems, 36(4), 7739-7750. https://doi.org/10.1109/TNNLS.2024.3385097

[22] Li, T., Zhu, K., Luong, N. C., Niyato, D., Wu, Q., Zhang, Y., & Chen, B. (2022). Applications of multi-agent reinforcement learning in future internet: A comprehensive survey. IEEE Communications Surveys & Tutorials, 24(2), 1240-1279. https://doi.org/10.1109/COMST.2022.3160697

[23] Chu, T., Wang, J., Codecà, L., & Li, Z. (2019). Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE transactions on intelligent transportation systems, 21(3), 1086-1095. https://doi.org/10.1109/TITS.2019.2901791

[24] Damadam, S., Zourbakhsh, M., Javidan, R., & Faroughi, A. (2022). An intelligent IoT based traffic light management system: Deep reinforcement learning. Smart Cities, 5(4), 1293-1311. https://doi.org/10.3390/smartcities5040066

[25] Jiang, S., Huang, Y., Jafari, M., & Jalayer, M. (2021). A distributed multi-agent reinforcement learning with graph decomposition approach for large-scale adaptive traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 23(9), 14689-14701. https://doi.org/10.1109/TITS.2021.3131596

[26] Yuan, S., Xu, S., & Zheng, S. (2022, January). Deep reinforcement learning based green wave speed guidance for human-driven connected vehicles at signalized intersections. In 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 331-339). IEEE. https://doi.org/10.1109/ICMTMA54903.2022.00070

[27] Kumar, N., Rahman, S. S., & Dhakad, N. (2020). Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system. IEEE Transactions on Intelligent Transportation Systems, 22(8), 4919-4928. https://doi.org/10.1109/TITS.2020.2984033

[28] Wu, Q., Wu, J., Shen, J., Du, B., Telikani, A., Fahmideh, M., & Liang, C. (2022). Distributed agent-based deep reinforcement learning for large scale traffic signal control. Knowledge-based systems, 241, 108304. https://doi.org/10.1016/j.knosys.2022.108304

[29] Shi, H., Liu, B., Wang, E., Han, W., Wang, J., Cui, S., & Wu, L. (2024). Cooperative multi-agent reinforcement learning framework for edge intelligence-empowered traffic light control. IEEE Transactions on Consumer Electronics, 70(4), 7373-7384. https://doi.org/10.1109/TCE.2024.3416822

[30] Gao, X., Li, X., Liu, Q., Li, Z., Yang, F., & Luan, T. (2022). Multi-agent decision-making modes in uncertain interactive traffic scenarios via graph convolution-based deep reinforcement learning. Sensors, 22(12), 4586. https://doi.org/10.3390/s22124586

[31] Liu, J., Qin, S., Su, M., Luo, Y., Wang, Y., & Yang, S. (2023). Multiple intersections traffic signal control based on cooperative multi-agent reinforcement learning. Information Sciences, 647, 119484. https://doi.org/10.1016/j.ins.2023.119484

[32] Zhou, B., Zhou, Q., Hu, S., Ma, D., Jin, S., & Lee, D. H. (2024). Cooperative traffic signal control using a distributed agent-based deep reinforcement learning with incentive communication. IEEE Transactions on Intelligent Transportation Systems, 25(8), 10147-10160. https://doi.org/10.1109/TITS.2024.3352730

[33] Fang, J., You, Y., Xu, M., Wang, J., & Cai, S. (2023). Multi-objective traffic signal control using network-wide agent coordinated reinforcement learning. Expert Systems with Applications, 229, 120535. https://doi.org/10.1016/j.eswa.2023.120535

[34] Ge, H., Gao, D., Sun, L., Hou, Y., Yu, C., Wang, Y., & Tan, G. (2021). Multi-agent transfer reinforcement learning with multi-view encoder for adaptive traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 23(8), 12572-12587. https://doi.org/10.1109/TITS.2021.3115240

[35] Stepanov, E. P., Smeliansky, R. L., Plakunov, A. V., Borisov, A. V., Zhu, X., Pei, J., & Yao, Z. (2024). On fair traffic allocation and efficient utilization of network resources based on MARL. Computer Networks, 250, 110540. https://doi.org/10.1016/j.comnet.2024.110540

[36] Zhou, X., Liang, W., Yan, K., Li, W., Wang, K. I. K., Ma, J., & Jin, Q. (2022). Edge-enabled two-stage scheduling based on deep reinforcement learning for internet of everything. IEEE Internet of Things Journal, 10(4), 3295-3304. https://doi.org/10.1109/JIOT.2022.3179231

[37] Zhang, Z., Tian, W., & Liao, Z. (2023). Towards coordinated and robust real-time control: a decentralized approach for combined sewer overflow and urban flooding reduction based on multi-agent reinforcement learning. Water Research, 229, 119498. https://doi.org/10.1016/j.watres.2022.119498

[38] Yang, J., Zhang, J., & Wang, H. (2020). Urban traffic control in software defined internet of things via a multi-agent deep reinforcement learning approach. IEEE Transactions on Intelligent Transportation Systems, 22(6), 3742-3754. https://doi.org/10.1109/TITS.2020.3023788

[39] Tan, T., Bao, F., Deng, Y., Jin, A., Dai, Q., & Wang, J. (2019). Cooperative deep reinforcement learning for large-scale traffic grid signal control. IEEE transactions on cybernetics, 50(6), 2687-2700. https://doi.org/10.1109/TCYB.2019.2904742

[40] Ghanadbashi, S., & Golpayegani, F. (2022). Using ontology to guide reinforcement learning agents in unseen situations: A traffic signal control system case study. Applied Intelligence, 52(2), 1808-1824. https://doi.org/10.1007/s10489-021-02449-5