

Adaptive RL-Based Traffic Signal Control with Dynamic Graph Attention and Temporal Awareness under Disruptions

Wisseem Yousfi

Univ Gustave Eiffel, ENTPE

Rim Slama

`rim.slamasalmi@entpe.fr`

Univ Gustave Eiffel, ENTPE

Pierre-Antoine Laharotte

Univ Gustave Eiffel, ENTPE

Research Article

Keywords: Traffic signal control, Multi-agent reinforcement learning, Spatio-temporal modelling, Robustness and resilience, Traffic disruptions

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11128617/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Adaptive RL-Based Traffic Signal Control with Dynamic Graph Attention and Temporal Awareness under Disruptions

Wissem Yousfi¹, Rim Slama^{1*}, Pierre-Antoine Laharotte¹

¹EMob-Lab, Univ Gustave Eiffel, ENTPE, Lyon, F-69518, France.

*Corresponding author(s). E-mail(s): rim.slamasalmi@entpe.fr;

Abstract

Deep Reinforcement Learning has demonstrated significant potential in optimizing Traffic Signal Control; however, existing Multi-Agent approaches often struggle with the non-stationarity inherent to abnormal traffic conditions. Most models, trained in nominal simulation environments, lack the resilience to withstand unforeseen disturbances such as sensor failures, road blockages, or extreme weather. To address this gap, this paper proposes a robust Spatio-Temporal Graph Reinforcement Learning framework explicitly designed for structural stability. Our architecture facilitates intelligent inter-agent collaboration by selectively weighting relevant neighbors via a Graph Attention Network (GATv2), while a Transformer-based Temporal Interaction Encoder captures long-term historical trends. This dual mechanism enables agents to distinguish between transient noise and persistent failure patterns, facilitating proactive rather than reactive control. Furthermore, we introduce a rigorous stress-testing benchmark comprising stochastic perturbation scenarios; ranging from high-frequency random blockages to critical arterial collapses, to quantify system resilience. Extensive experiments on a stochastic grid network demonstrate that our approach not only achieves superior throughput under high-demand nominal conditions but also exhibits exceptional resilience, recovering stable traffic flow significantly faster than state-of-the-art baselines following severe structural disruptions.

Keywords: Traffic signal control, Multi-agent reinforcement learning, Spatio-temporal modelling, Robustness and resilience, Traffic disruptions

1 Introduction

In the pursuit of smart cities, the structural resilience of intelligent infrastructure remains a critical yet often overlooked vulnerability. Urban traffic networks operate efficiently under nominal conditions but stay inherently fragile, exposed to cascading failures once a perturbation reduces the capacity they were designed around. Adaptive signal control is expected to absorb part of this fragility, and Multi-Agent Reinforcement Learning has become its dominant paradigm (Wei et al. 2019b; Yu et al. 2022), learning decentralized policies that adapt to fluctuating demand far better than fixed-time or actuated heuristics (Koonce et al. 2008). That adaptivity, however, is acquired under a narrow set of operating conditions, and it is precisely where this assumption breaks that the gap between simulation-based success and real-world viability opens (Da et al. 2023). Urban operation is disturbed regularly by events that temporarily alter the very conditions the controller was trained on: heavy rain that lowers speeds network-wide, a demonstration or a public event that reduces the traffic a corridor can carry while part of the demand diverts elsewhere, roadworks that reduce the capacity of an arterial for a period. Because they are trained on stationary demand and nominal capacity, agents develop brittle policies that degrade sharply once such an event begins and that do not necessarily return to nominal behaviour once it ends (Rodrigues and Lima Azevedo 2019; Shi et al. 2024; Nguyen et al. 2025). Withstanding a perturbation and recovering from it, that is, robustness and resilience, therefore becomes a design requirement rather than a secondary property (Garg 2020).

Meeting that requirement is not a matter of optimizing the same objective harder, because a disruption does not merely degrade the traffic state: it changes the information structure of the decision problem. Three questions must then be answered jointly and online. First, *which variations matter*, since a rise in lane density may originate from a stochastic arrival burst, from a platoon released by an upstream green, or from a sustained capacity reduction, and these regimes call for opposite responses. Once a variation is deemed significant, *when to react*, since switching in response to every fluctuation produces unstable phase sequences and wastes capacity in transitions, whereas reacting too late lets a queue reach the upstream intersection and turns a local incident into a spillback chain; symmetrically, keeping a defensive behaviour after the event is cleared prolongs the very recovery it was meant to shorten. Finally, *with whom to cooperate*, since when a corridor loses part of its capacity the intersections carrying useful information for a given agent are no longer those of nominal operation: some neighbours become sources of rerouted demand while others become uninformative because their upstream supply has been throttled.

Of these three questions, the last is the one that has received the most systematic architectural answer. Graph-based methods address it explicitly, from fixed-weight graph convolution (Nishi et al. 2018) to graph attention (Wei et al. 2019b) and, more recently, universal city-scale neighbour encoders (Zeng et al. 2024) and learned collaborator selection beyond geographic adjacency (Ruan et al. 2024). These architectures were designed for nominal efficiency and scale, however, not for reliability when the network they encode changes, and two consequences follow. First, the original graph-attention formulation computes a static neighbour ranking shared across query nodes once trained (Brody et al. 2022), so the coordination pattern that was optimal under

nominal flow is carried unchanged into the disrupted regime. Second, and independently of the weighting scheme, that ranking is derived from an instantaneous spatial snapshot, which lacks the temporal depth to separate transient noise from a persistent capacity reduction. The mechanism that answers the third question therefore gives comparatively little attention to the first two, and an architecture tuned for nominal performance offers no guarantee on robustness.

A few recent methods do confront disruptions directly, and MacLight (Lee et al. 2025) is the most closely related, evaluating its controller under random road blockages. Its design choices, however, limit what an agent can infer about the disruption it faces: the network is compressed into a single latent representation, learned from a global observation and shared by all agents, so an agent cannot tell which part of its own surroundings has changed, nor for how long.

To bridge this gap, this paper introduces a Dynamic Spatio-Temporal Graph Reinforcement Learning framework in which the perturbation acts on the representation itself rather than being absorbed by a coordination scheme fixed at training time. Two blocks are added to the agent’s representation, one for with whom to cooperate, the other for which variations matter and when to react. The first scores each neighbour from the joint state of the pair rather than from a ranking fixed once trained, so that a capacity reduction reshapes the functional neighbourhood without any change to the physical graph: an agent discounts a neighbour whose upstream supply has been throttled and up-weights the one now absorbing the rerouted demand. The second operates on the recent evolution of these spatial states over a finite history window, and is discriminative rather than predictive: a stochastic arrival burst and a sustained capacity reduction produce nearly identical instantaneous queues, and only their persistence separates them (Ni et al. 2022). The policy can then ignore a short-lived fluctuation, commit once a deviation persists, and release the adapted behaviour once the signature of the event has decayed, which directly conditions recovery. Because the temporal block reads the output of the spatial one, the history it accumulates is that of an already contextualized neighbourhood: together with the shared state embedding, this ordering lets the three questions be answered within a single representation.

The main contributions of this work are summarized as follows:

1. **A spatio-temporal MARL controller whose representation tracks the perturbation:** Each agent’s context is re-allocated at every control step toward the neighbours that stay informative under the degraded capacity, and enriched with a persistence signature that separates transient noise from a lasting capacity loss. Coordination and reaction timing therefore adapt online, without retraining or any external incident detector.
2. **A Stochastic Training Environment:** To address the issue of overfitting common in DRL-based traffic control, we introduce a probabilistic traffic generation engine. Instead of relying on static demand files, this environment generates unique, non-stationary traffic patterns for each episode, forcing the agent to learn generalized flow-regulation policies rather than memorizing specific arrival sequences.
3. **A Comprehensive Resilience Stress-Test Scenarios:** We design and propose a rigorous perturbation benchmark specifically for evaluating TSC resilience. This

suite extends beyond standard congestion scenarios to include explicit structural failures, such as high-frequency random blockages, critical arterial collapses, and adverse weather conditions; providing a standardized method for quantifying the robustness and recovery speed of control algorithms.

4. **Superior Efficiency and Structural Stability:** Extensive empirical evaluation demonstrates that our framework significantly outperforms state-of-the-art baselines (e.g., MacLight, MAPPO). Beyond achieving higher network-wide speeds under normal high-demand conditions, our agent exhibits superior elasticity under stress, maintaining lower performance degradation rates and recovering stable traffic flow faster than competitive methods following critical network failures.

The remainder of this paper is organized as follows: Section 2 reviews the related work in DRL-based traffic control and robustness. Section 3 details the problem formulation and the proposed spatio-temporal framework. Section 4 describes the architectural components. Section 5 presents the experimental setup, the stochastic training environment, and the resilience stress-test suite. Section 6 discusses the empirical results and performance analysis. Finally, Section 7 concludes the paper and outlines future research directions.

2 Literature Review

Traffic signal control has progressively evolved from pre-timed plans toward traffic-responsive and learning-based strategies. Fixed-time approaches, including those derived from Webster’s formulation, determine signal plans from historical traffic patterns and therefore offer limited responsiveness to short-term demand variations (Webster 1958). Traffic-actuated controllers and coordinated adaptive systems, such as SCOOT and SCAT, subsequently introduced online adjustments based on detected traffic conditions (Koonce et al. 2008; Robertson and Bretherton 1991; Sims and Dobinson 1980). Although these systems have been successfully deployed, they generally rely on calibrated models, predefined control logic, or centralized optimization procedures whose adaptation to complex and rapidly changing network conditions can be demanding.

Reinforcement Learning (RL) provides a model-free alternative in which signal-control policies are learned through interactions with the traffic environment. Early studies applied value-based methods such as Deep Q-Networks to isolated intersections and demonstrated their potential for reducing delay and congestion (Li et al. 2016). At the network level, however, centralized control becomes difficult to scale because the joint state and action spaces grow rapidly with the number of intersections. Multi-Agent Reinforcement Learning (MARL) addresses this issue by associating each intersection with an individual agent, although decentralized learning introduces partial observability and non-stationarity because the environment observed by one agent depends on the actions and evolving policies of the others (Chu et al. 2019; Li et al. 2021). Independent learners provide a scalable solution but do not explicitly coordinate their decisions, whereas Centralized Training with Decentralized Execution frameworks exploit additional information during training while preserving decentralized operation (Yu et al. 2022).

Transportation-grounded control principles have also been incorporated into RL-based traffic signal control. Max-pressure control selects phases according to imbalances between upstream and downstream traffic states and provides network-stability guarantees under its modelling assumptions (Varaiya 2013). PressLight builds on this principle by introducing pressure-based state and reward representations, thereby connecting reinforcement learning with a traffic-theoretic control objective rather than relying exclusively on heuristic reward design (Wei et al. 2019a). Such approaches underline the importance of aligning learning formulations with the physical mechanisms governing queue propagation and network congestion.

Network-level traffic control also requires explicit representation of interactions between intersections, since queues, platoons, and spillback can propagate over successive links. Graph Neural Networks offer a natural framework for modelling these dependencies by representing intersections as nodes and traffic relationships as edges. Nishi et al. employed graph convolution to aggregate information from connected intersections (Nishi et al. 2018). CoLight later introduced graph attention for network-level cooperation, allowing each agent to assign different importance weights to neighbouring intersections rather than aggregating their information uniformly (Wei et al. 2019b). Graph-based communication has since become a prominent approach for coordinating distributed traffic-signal agents.

Because traffic conditions evolve over time, several studies have combined spatial coordination with temporal representation. STMARL jointly models inter-intersection interactions and temporal traffic evolution for cooperative signal control (Wang et al. 2020). DynSTGAT further integrates historical intersection states into a spatial-temporal graph-attention architecture and employs a Temporal Convolutional Network to extract temporal features before combining them with spatial dependencies (Wu et al. 2021). Transformer-based models have also been explored. TransformerLight formulates traffic signal control as an offline sequence-modelling problem and uses gated Transformer blocks to infer actions from trajectories of states, actions, and rewards (Wu et al. 2023). In a different direction, MacLight compresses global traffic information through a variational autoencoder and combines the resulting representation with a PPO-based controller, with particular emphasis on computational efficiency, convergence stability, and performance across multiple traffic scenarios (Lee et al. 2025). These studies show that graph attention, historical-state encoding, convolutional temporal models, Transformers, and compact latent representations have all been investigated as mechanisms for improving traffic-signal coordination.

Beyond nominal traffic efficiency, an increasing body of work has examined the robustness of RL-based controllers under abnormal or uncertain operating conditions. Aslani et al. evaluated discrete- and continuous-state RL controllers in a microscopic model of downtown Tehran and analysed their behaviour under different traffic disturbances (Aslani et al. 2018). Rodrigues and Azevedo investigated demand surges, incident-induced capacity reductions, and sensor failures, showing that policies trained exclusively under nominal conditions may experience substantial performance deterioration when deployed under unseen disruptions (Rodrigues and Lima Azevedo 2019). Tan et al. addressed uncertainty in traffic-state observations by injecting synthetic perturbations during training, following a principle related to adversarial training (Tan

et al. 2020). Zeinaly et al. considered accident scenarios at a single intersection and enriched the state representation with information describing accident-induced traffic changes (Zeinaly et al. 2023).

The design of evaluation environments is equally important when comparing traffic-signal controllers. The RESCO benchmark demonstrated that the relative performance of RL-based methods can vary substantially with sensing assumptions, intersection layouts, and scenario definitions, highlighting the need for consistent and realistic experimental protocols (Ault and Sharon 2021). More recently, Nguyen et al. introduced T-REX, a SUMO-based framework for evaluating RL-based traffic signal control under incidents while representing probabilistic rerouting, speed adaptation, and contextual lane-changing behaviour (Nguyen et al. 2025). Their comparative analysis shows that no single controller consistently dominates across all robustness dimensions and that exposure to incidents during training does not necessarily guarantee improved deployment performance under incident-driven distribution shifts.

3 Problem Formulation

We formulate network-level traffic signal control as a Partially Observable Markov Decision Process (POMDP) $\langle N, S, O, A, R, P, \gamma \rangle$, where N denotes the set of signalized-intersection agents, S the global state space, O the local observation space, A the action space, R the reward function, P the transition dynamics, and γ the discount factor. At timestep t , each agent i receives a local observation $o_i^t \in O_i$, selects an action $a_i^t \in A_i$, and receives a reward r_i^t . Each agent learns a policy $\pi_i : O_i \rightarrow A_i$ while operating under partial observability of the global traffic state.

3.1 RL-based Traffic Signal Control

The network contains N signalized intersections, each controlled by an autonomous agent. At a four-way intersection, twelve vehicular movements—straight, left-turn, and right-turn movements for each incoming approach—are organized into eight non-conflicting signal phases. Right-turn movements are permitted by default. Figure 1 illustrates the lane configuration and available phases.

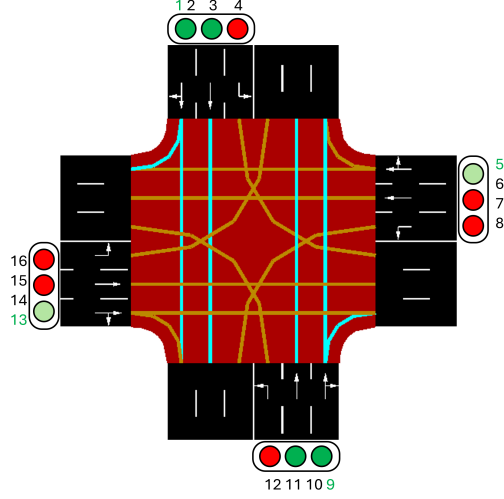


Fig. 1: Four-way intersection with twelve traffic movements and eight non-conflicting signal phases (A–H).

3.1.1 Observation

Following [Alegre et al. \(2021\)](#), the observation of intersection i includes: (i) a one-hot representation of the active phase; (ii) a binary transition-eligibility indicator, activated once the minimum green duration of 10 s is satisfied; (iii) lane density, defined relative to lane capacity; and (iv) queue saturation, corresponding to the normalized density of vehicles travelling below 0.1 m/s.

The observation is therefore

$$o_i^t = [\text{phase}_i^t, \text{minGreen}_i^t, \text{density}_i^t[l], \text{queue}_i^t[l]]_{l \in \mathcal{L}_i^{\text{in}}}, \quad (1)$$

with all components normalized to $[0, 1]$.

3.1.2 Action

The action space consists of the eight feasible signal phases. When the selected action differs from the current phase, a 3 s yellow transition is automatically inserted before activating the new phase.

3.1.3 Reward Function

The control objective is to reduce vehicle delay. We therefore use the Finite Difference of Waiting Time ([Alegre et al. 2021](#)):

$$r_i^t = -(W_i^t - W_i^{t-1}), \quad (2)$$

where W_i^t denotes the cumulative waiting time of vehicles associated with intersection i at timestep t . Positive rewards consequently correspond to reductions in accumulated waiting time between consecutive control steps.

4 Proposed DRL Framework

This section presents the proposed Spatio-Temporal Graph Reinforcement Learning framework, comprising four components: observation embedding, graph-attention-based spatial collaboration, temporal interaction encoding, and a PPO-based traffic signal controller.

4.1 Architecture Overview

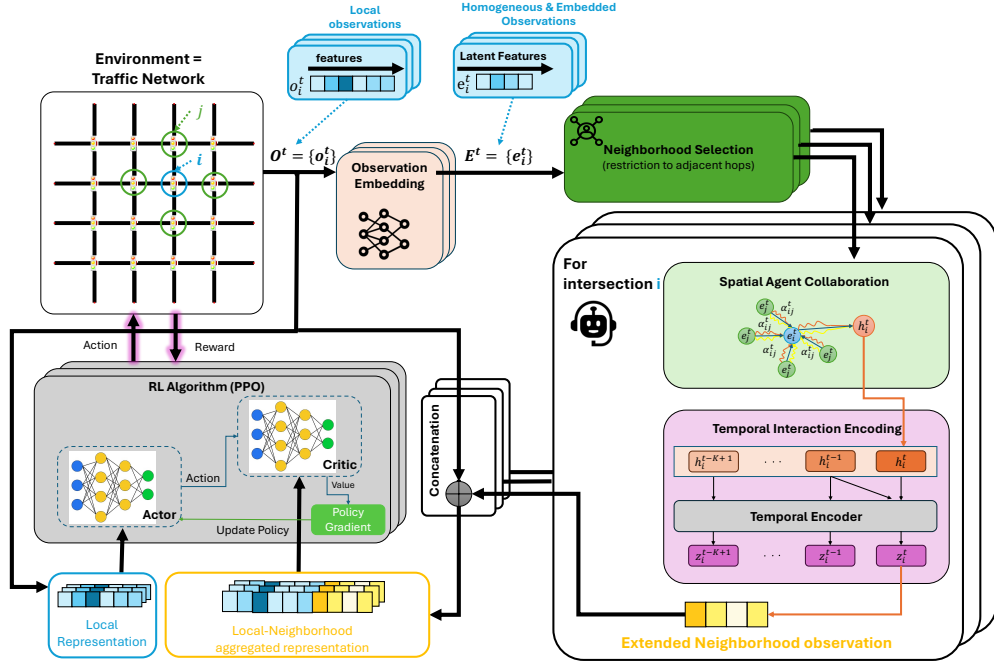


Fig. 2: Overall architecture of the proposed method.

As illustrated in Fig. 2, each signalized intersection is represented as an agent in a graph-based traffic network. At each control step, the local observation is first mapped to a latent embedding, after which GATv2 aggregates information from directly connected intersections using context-dependent attention weights. This adaptive weighting allows the controller to emphasize neighbours whose traffic states become more relevant when congestion propagates or a local capacity reduction occurs. The resulting spatial representations are then processed over a finite history window by a Transformer-based temporal encoder, enabling the controller to distinguish short-lived fluctuations from persistent traffic changes associated with sustained disruptions. By combining these spatial and temporal cues with the current local observation, the PPO controller can adapt its signal decisions to both the immediate and evolving effects of network perturbations.

4.2 Observation Embedding

The local observation o_i^t of agent i contains heterogeneous variables, including the active phase, minimum-green transition indicator, lane densities, and queue-saturation values. To provide the subsequent spatial and temporal modules with a common representation, we project the observation into a latent space:

$$e_i^t \in \mathbb{R}^{d_{\text{model}}}, \quad (3)$$

where $d_{\text{model}} = 32$ in our implementation. The embedding module consists of a two-layer Multi-Layer Perceptron followed by Layer Normalization:

$$e_i^t = \text{LayerNorm} \left(\sigma \left(\sigma \left(o_i^t W_1 + b_1 \right) W_2 + b_2 \right) \right), \quad (4)$$

where W_1 and W_2 are learnable weight matrices, b_1 and b_2 are bias vectors, and $\sigma(\cdot)$ denotes the ReLU activation function. The resulting fixed-dimensional vector e_i^t is used by the spatial collaboration module.

4.3 Spatial Agent Collaboration

Traffic conditions at neighbouring intersections are interdependent because queues and spillback can propagate across connected links. We therefore employ GATv2 (Brody et al. 2022) to aggregate information from directly connected intersections using context-dependent attention weights. Unlike the original GAT formulation (Veličković et al. 2018), GATv2 uses dynamic, query-dependent attention, allowing the relative contribution of neighbour j to depend on the joint states of intersections i and j at time t .

This mechanism is particularly relevant when local capacity reductions alter queue propagation. The physical neighbourhood remains fixed, but the contribution of each directly connected intersection can vary with the current traffic context.

Formally, the traffic network is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. For each connected pair $(i, j) \in \mathcal{E}$, the unnormalized attention score is

$$c_{ij}^t = \mathbf{a}^\top \text{LeakyReLU} \left(W[e_i^t \parallel e_j^t] \right), \quad (5)$$

where W is a learnable weight matrix, $\mathbf{a}$ is the attention vector, and $\parallel$ denotes concatenation. The scores are normalized over the neighbourhood $\mathcal{N}_i$:

$$\alpha_{ij}^t = \frac{\exp(c_{ij}^t)}{\sum_{k \in \mathcal{N}_i} \exp(c_{ik}^t)}. \quad (6)$$

Figure 3 illustrates how the distribution of these attention weights evolves during training.

Neighbour information is aggregated as

$$h_i^t = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^t W e_j^t \right). \quad (7)$$

We employ $M = 4$ attention heads, whose outputs are concatenated to form the spatial representation h_i^t . This representation combines the local state with the

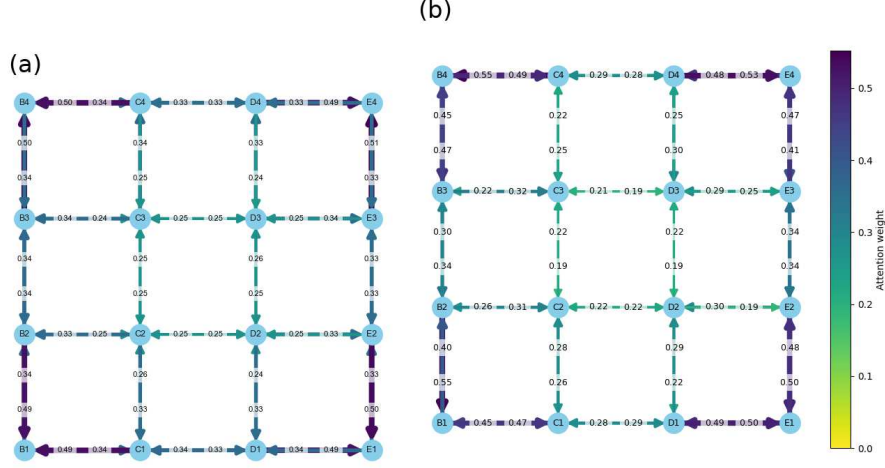


Fig. 3: Evolution of attention coefficients α_{ij} assigned by a GATv2 head: (a) early training phase; (b) advanced training phase.

context-dependent influence of directly connected intersections and may provide information about the propagation of congestion through the immediate neighbourhood.

4.4 Temporal Interaction Encoding

The spatial representation h_i^t describes the traffic context at a single control step, whereas congestion develops over time. Similar instantaneous states can correspond to different traffic evolutions, such as an emerging queue, a persistent congestion pattern, or a queue already dissipating. We therefore process a finite history of spatial representations using a Transformer-based Temporal Interaction Encoder.

For intersection i , the history sequence is defined as

$$H_i^t = \{h_i^{t-K_{\text{hist}}+1}, \dots, h_i^{t-1}, h_i^t\} \in \mathbb{R}^{K_{\text{hist}} \times d_{\text{model}}}, \quad (8)$$

where K_{hist} is the temporal look-back horizon. To preserve the ordering of the sequence, a learnable positional encoding $P \in \mathbb{R}^{K_{\text{hist}} \times d_{\text{model}}}$ is added:

$$X_i^t = H_i^t + P. \quad (9)$$

The resulting sequence is processed by L Transformer Encoder layers using multi-head self-attention to model dependencies among recent spatial states. The output corresponding to the latest timestep is extracted as

$$z_i^t = \text{TransformerEncoder}(X_i^t)_{-1}. \quad (10)$$

The vector z_i^t summarizes the recent evolution of the attention-weighted spatial context and is provided to the PPO controller. Under changing traffic conditions, this representation allows the policy to consider recent temporal evolution in addition to the instantaneous state.

4.5 Reinforcement Learning Algorithm (PPO)

The final component maps the spatio-temporal representation to discrete traffic-signal actions. We employ Proximal Policy Optimization (PPO) (Schulman et al. 2017), whose clipped policy-update mechanism limits excessively large policy changes and supports stable optimization in the multi-agent setting.

Training follows the Centralized Training and Decentralized Execution (CTDE) paradigm. During training, the critic provides additional information for policy optimization, whereas during execution only the actor is retained. Each intersection therefore selects its signal phase from locally available observations and the information exchanged with directly connected neighbours.

The actor-critic components are defined as follows:

- **Actor** $\pi_\theta(a_i^t|s_i^t)$: The actor receives $s_i^t = [z_i^t \| o_i^t]$, combining the spatio-temporal representation with the current local observation, and outputs a probability distribution over the available signal phases.
- **Critic** $V_\phi(z_i^t)$: The critic estimates the value of the encoded state during training and provides the baseline used for policy optimization.

PPO optimizes the clipped surrogate objective

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip} \left(r_t(\theta), 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t \right) \right], \quad (11)$$

where

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (12)$$

is the probability ratio between the updated and previous policies, and ϵ denotes the clipping parameter. The advantage estimate is computed using Generalized Advantage Estimation (GAE) (Schulman et al. 2015):

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma \lambda)^l \delta_{t+l}, \quad \delta_t = R_t + \gamma V_\phi(z_i^{t+1}) - V_\phi(z_i^t), \quad (13)$$

where γ is the discount factor and λ controls the bias-variance trade-off. The critic is optimized through the squared value loss

$$L^V(\phi) = \hat{\mathbb{E}}_t \left[\left(V_\phi(z_i^t) - V_t^{\text{target}} \right)^2 \right]. \quad (14)$$

5 Experimental Networks and Perturbation Design

The experimental evaluation is conducted on two complementary traffic networks. First, a controlled synthetic 4×4 grid is used to generate stochastic traffic conditions and to evaluate the controllers under precisely defined structural perturbations. Second, the real-world-derived Hangzhou 4×4 benchmark is used to assess nominal traffic-control performance on a network with heterogeneous road geometry and empirically derived traffic demand. The perturbation analysis is restricted to the controlled

synthetic network to ensure that disruption location, duration, and severity remain consistent across experiments.

This section describes the two network configurations, the stochastic demand-generation process used for the synthetic grid, and the perturbation scenarios considered in the robustness and recovery analyses.

5.1 Experimental Networks

5.1.1 Controlled Synthetic Grid

The first experimental network is a homogeneous 4×4 grid modelled in the SUMO traffic simulator. As illustrated in Fig. 4, it contains 16 signalized intersections arranged in a regular lattice.

The network spans a $1000 \text{ m} \times 1000 \text{ m}$ area, with intersections spaced at regular intervals. The connecting road segments have a uniform length of 186.40 m . Each incoming road contains three lanes:

- **Lane 0:** a shared right-turn and straight-movement lane;
- **Lane 1:** a dedicated straight-movement lane;
- **Lane 2:** a dedicated left-turn lane.

The uniform geometry provides a controlled environment in which stochastic demand patterns and structural disruptions can be introduced without additional effects caused by heterogeneous road characteristics. The speed limit is set to 13.89 m/s , corresponding to approximately 50 km/h .

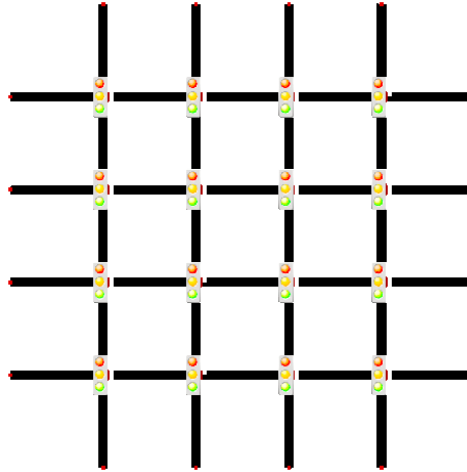


Fig. 4: Controlled synthetic 4×4 grid containing 16 signalized intersections.

5.1.2 Hangzhou Network

The second experimental environment is the real-world-derived Hangzhou 4×4 benchmark, which also contains 16 signalized intersections. In contrast to the controlled synthetic grid, the Hangzhou network presents heterogeneous road lengths, intersection configurations, and traffic-flow distributions derived from observed urban traffic conditions (Lee et al. 2025).

The network is included to examine whether the proposed controller can maintain effective nominal operation beyond the regular geometry and synthetically generated demand of the controlled grid. All Hangzhou experiments use a simulation horizon of 3600 s and the same principal training configuration as the synthetic-network experiments.

Hangzhou is used exclusively for nominal performance evaluation. No structural perturbation is introduced into this network, since the robustness and recovery analyses require a controlled experimental environment in which the characteristics of each disruption can be consistently reproduced. Figure 5 illustrates the Hangzhou network.



Fig. 5: Real-world-derived Hangzhou 4×4 traffic network.

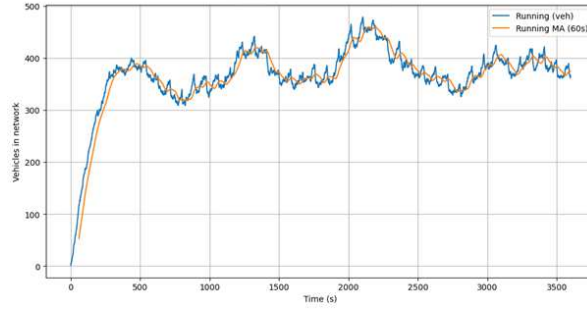
5.2 Stochastic Demand Generation

Traffic-signal-control studies commonly rely on fixed demand files to ensure reproducible comparisons. However, repeatedly exposing a learning agent to the same departure times and origin-destination patterns limits the diversity of traffic states encountered during training. To broaden this distribution, we develop a stochastic demand-generation process that produces distinct spatial and temporal demand realizations across simulation episodes.

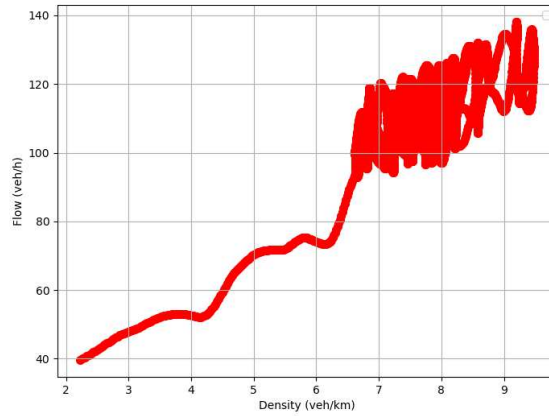
The generation process is governed by a time-varying probability model rather than fixed departure timestamps. A high-demand scenario with a target traffic volume of approximately 6000 vehicles per hour is considered. The simulation horizon, $T = 3600$ s, is divided into intervals of 180 s. Within each interval, vehicle generation is

599 modelled as a sequence of Bernoulli trials, with the probability of generating a vehicle
600 for each origin–destination pair determined by the corresponding demand profile.

601 The probabilistic generation process introduces variability in vehicle-arrival times,
602 origin–destination demand, and the resulting congestion patterns. Fig. 6a illustrates
603 the corresponding network accumulation, which varies over the simulation rather than
604 following a fixed deterministic profile. Fig. 6b presents the associated Macroscopic
605 Fundamental Diagram. The dispersion of observations and the presence of states on the
606 congested branch indicate that the generated demand exposes the controllers to both
607 uncongested and congested operating conditions. The realized inflow in the illustrated
608 episode is approximately 6071 vehicles per hour.



(a) Network accumulation over time.



(b) Macroscopic Fundamental Diagram.

Fig. 6: Analysis of the stochastic traffic environment.

To further analyze the complexity of the generated scenarios, we examined the distribution of traffic density across time and space. Figure 7 presents a heatmap of vehicle occupancy across network edges over the simulation duration.

The heatmap reveals two key characteristics of our stochastic generation:

1. **Spatial Heterogeneity:** Traffic is not uniformly distributed. Certain “arterial” edges consistently bear higher loads due to the weighted OD probabilities, while local streets experience intermittent usage.
2. **Temporal Non-Stationarity:** The traffic intensity shifts dynamically. We observe distinct bands of congestion that form and dissipate unpredictably.

Together, these spatial and temporal variations broaden the range of traffic states presented to the learning agents. They also provide a more demanding evaluation environment than a single fixed and spatially uniform demand realization, while preserving control over the overall demand level and network configuration.

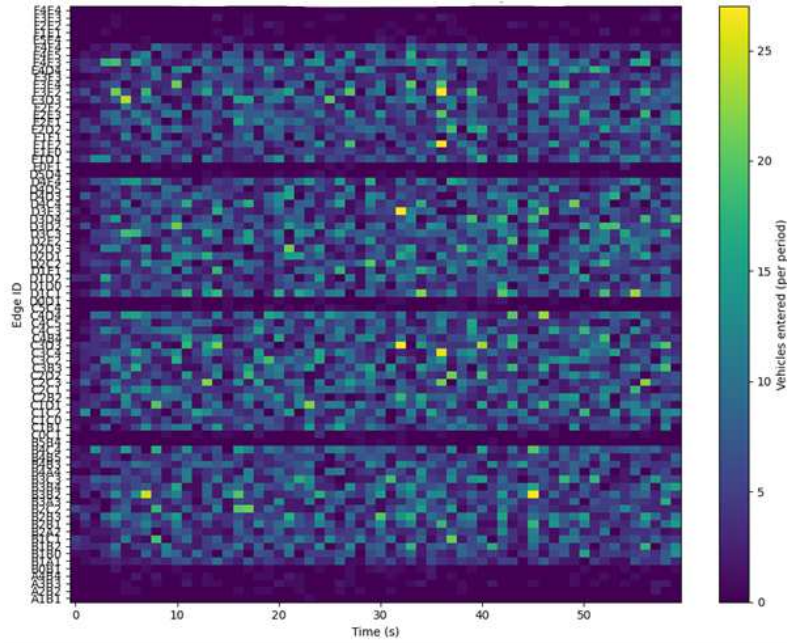


Fig. 7: Spatio-temporal heatmap of traffic density. The varying intensity across different edges (Y-axis) and time steps (X-axis) demonstrates the non-uniform and stochastic nature of the load distribution.

5.3 Perturbation Scenario Modelling

To evaluate the resilience of the proposed RL framework, we extend the simulation environment beyond standard congestion scenarios by introducing explicit stochastic perturbations. These perturbations mimic real-world anomalies such as accidents, maintenance work, or sudden capacity drops, testing the agent’s ability to adapt to structural changes in the network topology.

5.3.1 Traffic Events modelling

Traffic events are modeled as temporary reductions in road capacity, forcing the system to deviate from its nominal steady state. We categorize these events into dynamic random blockages and realistic fixed-location incidents.

Random Road Blockage Scenario

The random blockage scenario simulates multiple, unpredictable incidents occurring simultaneously across the network. This setup tests the agent’s generalization capability when the network topology becomes sparse and time-variant.

- **Mechanism:** The blockage process is governed by a periodic stochastic reset. Within a defined perturbation window ($t \in [t_0, t_1]$), the environment updates the blockage configuration every 300 seconds.
- **Selection Logic:** At each update step, $N = 8$ edges are randomly sampled from a pre-defined “candidate pool” of critical roads (illustrated in Figure 8).
- **Severity:** The selected edges are not completely closed but severely restricted. The maximum allowable speed is reduced to 0.5 m/s (effectively simulating a crawling queue around an accident), and vehicles on these edges are triggered to perform dynamic rerouting.
- **Dynamics:** After 300 seconds, the blocked edges are released, and a new set of 8 random edges is immediately blocked. This creates a highly non-stationary environment where the bottleneck locations shift continuously.

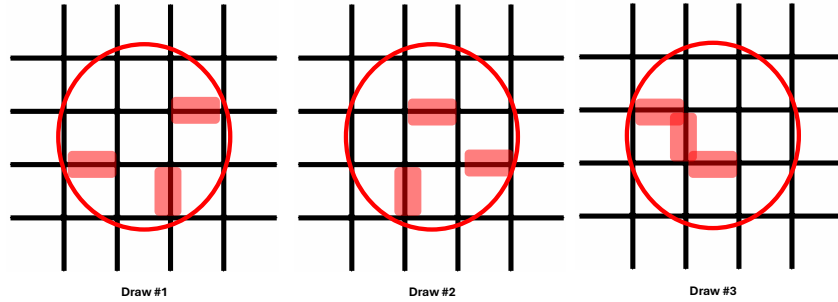


Fig. 8: Schematic of a set of Random Blockage Scenarios. The red circle indicates the “candidate pool” of edges, while the edges highlighted in red match the effectively blocked ones per draw.

Realistic Road Blockage Scenarios

In contrast to the random scenario, realistic blockages simulate sustained structural failures where specific functional zones are rendered inaccessible for the entire duration of the perturbation window ($t \in [t_0, t_1]$). We define three distinct configurations based on common urban incidents:

1. **West-East Corridor Failure (Arterial Maintenance):** This configuration simulates a long-term maintenance corridor or a chain-reaction accident affecting primary horizontal arteries. We block parallel West-to-East edges in the central rows, forcing the agent to reroute cross-town traffic to peripheral corridors (Rows 1 and 4).
2. **South-North Corridor Failure:** Similar to the West-East scenario, this targets the vertical axis by blocking primary South-to-North avenues. This mimics a scenario where longitudinal thoroughfares are closed for utility works, requiring the agent to shift heavy vertical flows to outer edges without causing gridlock at turning points.
3. **Central Square Closure (City Center Event):** This scenario mimics a city-center closure for a major event (e.g., a parade or VIP zone). The blockage targets the geometric center of the grid, creating a “denial-of-service” zone in the central loop. This forces the agent to navigate a “hollowed-out” topology, diverting through-traffic to the perimeter.

Figure 9 illustrates these three blockage configurations.

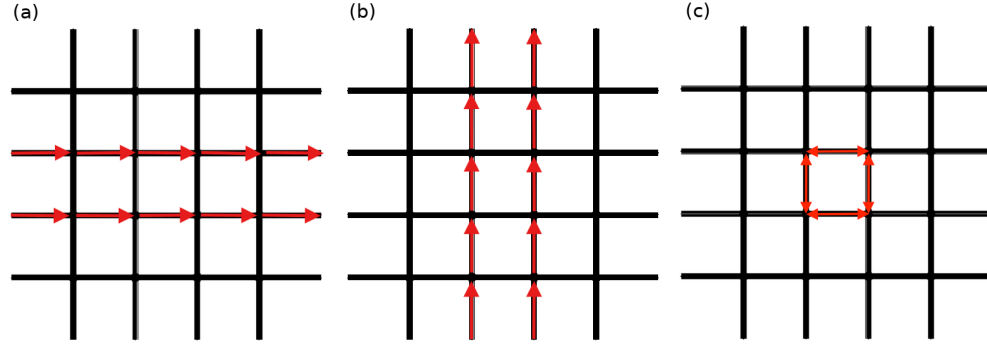


Fig. 9: Realistic blockage scenarios: (a) West-East corridor failure; (b) South-North corridor failure; (c) central square closure. Red arrows indicate the edges subjected to long-term closure during the perturbation window.

5.3.2 Adverse Weather Conditions modeling

Beyond structural capacity restrictions, adverse weather can affect traffic operation through changes in road friction, vehicle speeds, braking behaviour, and inter-vehicle spacing. Representing these interacting effects remains challenging in microscopic

traffic simulation, particularly when detailed weather-specific calibration data are unavailable. We therefore introduce a simplified rainy-weather scenario whose parameter changes are informed by previous work on the representation of road-surface conditions in SUMO (Weber et al. 2019). The objective is not to reproduce the complete physical and behavioural effects of rainfall, but to expose the controllers to a temporary network-wide change in vehicle dynamics.

The perturbation is activated during a predefined interval, $t \in [t_0, t_1]$. Weber et al. (2019) associate wet road surfaces with lower friction coefficients and report that empirical studies commonly observe reductions in traffic speed under rainy conditions. Based on these findings, two microscopic vehicle parameters are modified during the weather interval:

1. **Speed Reduction Factor** ($\beta = 0.80$): Wet road conditions are associated with reduced operating speeds because of lower tire-road friction and more cautious driving behaviour. The literature summarized by Weber et al. (2019) reports a range of speed reductions under rainy conditions. We select a speed factor of 0.80, corresponding to a 20% reduction in desired speed, as a representative value within the reported range.
2. **Increased Minimum Gap** ($g_{\min} = 3.0 \text{ m}$): Rain can also encourage drivers to maintain greater spacing because of reduced friction and increased braking requirements. To approximate this behavioural response, the minimum gap of the car-following model is increased from 2.5 m to 3.0 m. This value constitutes a modelling assumption informed by the expected direction of wet-weather driving behaviour rather than a directly calibrated empirical estimate.

At each simulation step, the environment checks whether the current time lies within the weather interval. For $t \in [t_0, t_1]$, the reduced speed factor and increased minimum gap are applied to all active vehicles. Once the interval ends, the default vehicle parameters are restored. The resulting scenario creates a temporary and non-stationary change in microscopic traffic behaviour while leaving the demand, network topology, and signal-control policies unchanged.

6 Experiments and Results

In this section, we conduct a comprehensive evaluation of the proposed framework to validate its effectiveness in realistic traffic scenarios. We begin by outlining the experimental design, including the comparison baselines, performance metrics, and implementation details. Following this setup, we analyze the system’s performance under normal traffic conditions, benchmarking against state-of-the-art methods. Finally, we extend the evaluation to perturbed environments to assess the robustness and resilience of the proposed approach.

6.1 Experimental Design

To ensure a rigorous and reproducible evaluation of the proposed framework, we establish a comprehensive experimental setup that assesses system performance across multiple critical dimensions. This subsection details the methodological framework,

beginning with the selection of baseline algorithms that serve as benchmarks, ranging from classical control strategies to state-of-the-art deep reinforcement learning methods. We then define a hierarchical set of evaluation metrics, categorizing them into standard efficiency indicators and specialized metrics for quantifying robustness and resilience across various perturbation scenarios. Finally, we specify the implementation details and simulation parameters to ensure the transparency and replicability of the reported results.

6.1.1 Baselines and Comparison Framework

To validate the effectiveness of our proposed approach, we benchmark it against a comprehensive set of baselines representing distinct paradigms in traffic signal control. These range from traditional rule-based strategies widely deployed in real-world infrastructure to advanced state-of-the-art Multi-Agent Reinforcement Learning (MARL) algorithms. The selected comparison methods are detailed as follows:

- **Fixed-Time (FT):** A conventional non-adaptive control strategy where all intersections operate on a pre-defined, constant cycle (Koonce et al. 2008). Phases switch every 45 seconds, regardless of real-time traffic fluctuations, serving as a standard reference for static control performance.
- **Independent Deep Q-Network (IDQN):** A value-based reinforcement learning baseline where each intersection is controlled by an autonomous agent employing Deep Q-Networks (Mnih et al. 2015). Agents learn phase-control policies independently without explicit coordination mechanisms.
- **Independent Proximal Policy Optimization (IPPO):** A strong decentralized baseline where each intersection is modeled as an independent agent trained with PPO (De Witt et al. 2020). Unlike IDQN, IPPO utilizes on-policy learning to optimize policies solely based on local observations, providing a robust reference for decentralized execution without communication.
- **Multi-Agent Proximal Policy Optimization (MAPPO):** A cooperative extension of PPO that follows the Centralized Training and Decentralized Execution (CTDE) paradigm (Yu et al. 2022). While agents execute independently, a centralized critic exploits global information during training to improve coordination, though it often faces scalability challenges in large networks.
- **MacLight:** A recent state-of-the-art MARL approach that integrates variational autoencoders (VAEs) with PPO (Lee et al. 2025). By compressing global traffic features into latent embeddings, MacLight aims to improve stability and convergence, making it the primary competitor to our proposed spatio-temporal framework.

Other well-established graph-based approaches, such as CoLight (Wei et al. 2019b) and DualLight (Lu et al. 2023), could also be considered. However, MacLight (Lee et al. 2025) has already demonstrated superior performance over these methods and, more importantly, is explicitly designed and evaluated for disrupted traffic conditions, including road blockages. We therefore retain MacLight as the representative advanced baseline, avoiding redundant and computationally costly comparisons.

6.1.2 Evaluation Metrics

We adopt a hierarchical evaluation framework that assesses the system’s capabilities across three dimensions: general traffic efficiency, stability under active perturbation (Robustness), and the capacity to recover from perturbations (Resilience).

Figure 10 illustrates the perturbation and recovery phases used to distinguish robustness from resilience.

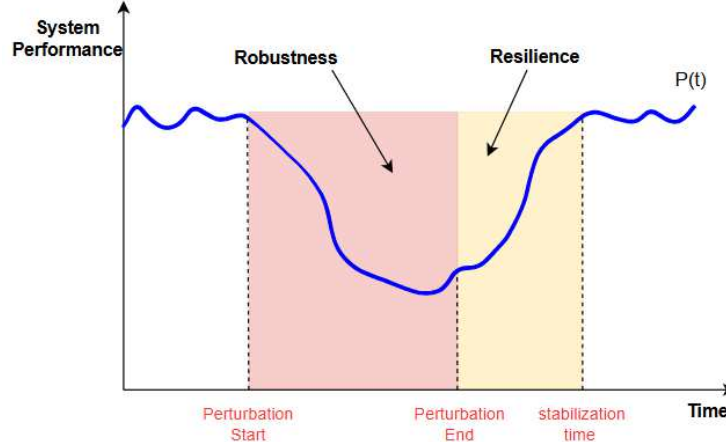


Fig. 10: Conceptual framework of system performance $P(t)$ under disruption. The red shaded region represents the active perturbation window where system *robustness* is evaluated. The yellow shaded region illustrates the recovery phase, characterizing the system’s *resilience* from the end of the perturbation until performance stabilization.

Performance Metrics

To assess general traffic efficiency under normal operating conditions, we utilize three standard indicators widely adopted in the literature (Wei et al. 2019c):

- **Total Waiting Time (s):** The cumulative time that all vehicles spend idling at an intersection. Lower values indicate reduced congestion and smoother traffic flow.
- **Total Queue Length (veh):** The cumulative number of stopped vehicles at an intersection. This metric reflects the extent of congestion and the control strategy’s ability to prevent spillovers.
- **Average Speed (m/s):** The mean velocity of all vehicles in the network. A higher average speed indicates improved mobility and more efficient use of road capacity.

Robustness Metrics

We define *robustness* as the system’s ability to resist performance degradation *during* the active phase of a perturbation. To quantify this resistance, we use

Performance Degradation Rate (PDR) (Biolini 2007):

$$PDR_m = \begin{cases} \frac{P_{\text{perturbed},m} - P_{\text{normal},m}}{P_{\text{normal},m}} \times 100\%, & \text{if metric } m \text{ is minimized,} \\ \frac{P_{\text{normal},m} - P_{\text{perturbed},m}}{P_{\text{normal},m}} \times 100\%, & \text{if metric } m \text{ is maximized.} \end{cases} \quad (15)$$

where $P_{\text{perturbed}}$ and P_{normal} represent the evaluation metric (e.g., Total Waiting Time) observed during the perturbation interval in the perturbed and normal scenarios, respectively. A lower PDR indicates a policy that maintains near-nominal functionality despite environmental stress.

Resilience Metrics

In contrast to robustness, which measures resistance during a shock, *resilience* characterizes the system’s elastic response after the perturbation ceases, quantifying how quickly and effectively it returns to a stable equilibrium. We evaluate this recovery dynamic using three complementary metrics:

- **Area Under the Curve Ratio (ρ_{AUC}):**

$$\rho_{AUC} = \frac{\int_{t_0}^{t_{end}} P_{\text{perturbed}}(t) dt}{\int_{t_0}^{t_{end}} P_{\text{normal}}(t) dt} \quad (16)$$

This ratio compares the cumulative performance profile of the perturbed scenario against the baseline. A value close to 1 implies that the system absorbed the shock with minimal cumulative loss in efficiency.

- **Peak Loss (L_{peak}):**

$$L_{\text{peak}} = \max_{t \in [t_0, t_{end}]} (P_{\text{normal}}(t) - P_{\text{perturbed}}(t)) \quad (17)$$

This metric captures the worst-case instantaneous deviation from the baseline performance, identifying the system’s maximum vulnerability point during the event.

- **Recovery Time (T_{rec}):**

$$T_{\text{rec}} = t_{\text{stabilized}} - t_1 \quad (18)$$

defined as the duration required for the system to return to a stable state after the perturbation ends ($t > t_1$). Here, $t_{\text{stabilized}}$ is the first timestep where performance returns to 90% of the pre-shock nominal average.

6.1.3 Experimental and Implementation Details

All experiments were conducted using the microscopic traffic simulator SUMO (version 1.19), interacting via the LibSUMO interface to maximize computational efficiency during training. The framework was implemented in Python 3.8 and PyTorch 2.0, running on an iMac equipped with a 4.2 GHz Quad-Core Intel Core i7 processor, a

Radeon Pro 580 GPU, and 16 GB of RAM. To ensure statistical reliability and fair comparison, each experimental scenario was executed across five independent random seeds. The final performance results reported in this study represent the average metrics across these five runs, providing a robust measure of the system’s efficiency and stability. To facilitate full reproducibility of the reported results, the specific hyperparameters governing the training process and model architecture are detailed in Table 1.

Table 1: Hyperparameters and Training Configuration.

Parameter	Value	Description
<i>Training Setup</i>		
Episodes	80	Total training episodes
Duration	3600	Episode duration (s)
<i>Learning Rates</i>		
Actor Learning Rate	1×10^{-4}	Policy learning rate
Critic Learning Rate	1×10^{-3}	Value learning rate
GAT & Temporal Encoder LR	1×10^{-3}	Shared learning rate
<i>PPO Hyperparameters</i>		
Discount Factor (γ)	0.99	Discount factor
GAE Parameter (λ)	0.95	GAE smoothing parameter
Clip Range (ϵ)	0.2	Clipping threshold
Update Epochs	10	Updates per batch
<i>Model Architecture</i>		
GAT Attention Heads	4	Attention heads count
Temporal History (K_{hist})	8	Input sequence length

6.2 Performance under Normal Conditions

In this section, we present a comprehensive evaluation of the proposed method under standard traffic conditions, without external perturbations. The experiments were conducted using the simulation scenario described earlier.

We compare the proposed framework with a diverse set of reference controllers, ranging from static traffic-signal control to independent and cooperative reinforcement-learning methods. The evaluation is conducted on two complementary networks. The stochastic synthetic grid provides a controlled environment with variable traffic-demand realizations, whereas the real-world-derived Hangzhou network introduces heterogeneous road geometry and traffic-flow distributions. The Hangzhou experiments are restricted to nominal conditions; all perturbation and recovery analyses are conducted on the controlled synthetic grid.

Table 2 presents the results obtained on the stochastic synthetic grid. The reported values are averaged over five independent evaluation seeds. Lower waiting time and queue length indicate better congestion management, whereas higher average speed indicates improved network mobility.

Table 2: Nominal performance on the stochastic synthetic grid. Results are reported as mean $\pm$ standard deviation over five independent evaluation seeds. Best mean values are highlighted in bold, and second-best mean values are underlined.

Method	Waiting Time (s) $\downarrow$	Queue Length (veh) $\downarrow$	Average Speed (m/s) $\uparrow$
Fixed-Time	54238.8 ± 17354.7	783.4 ± 179.6	2.01 ± 1.21
IDQN	6375121.4 ± 162220.2	2971.6 ± 26.1	0.02 ± 0.01
MAPPO	6461.2 ± 2404.1	200.0 ± 44.9	4.16 ± 0.60
IPPO	<u>643.2 ± 184.8</u>	61.6 ± 12.0	6.70 ± 0.70
MacLight	644.8 ± 244.2	59.2 ± 16.6	<u>6.97 ± 0.82</u>
Ours	639.0 ± 141.4	57.6 ± 12.5	7.24 ± 0.59

The results show substantial differences between the evaluated control paradigms. Fixed-Time control cannot adapt to the variable high-demand conditions and consequently produces long waiting times and large queues. IDQN performs particularly poorly, suggesting that independent value-based learning is unable to manage this demanding network configuration. MAPPO improves network mobility relative to these two references but continues to experience considerably greater congestion than the strongest decentralized and collaborative policy-gradient methods.

IPPO, MacLight, and the proposed controller form the most competitive group. IPPO obtains a mean waiting time of 643.2 s, while MacLight achieves a lower mean queue length and a higher average speed. The proposed controller records the best mean result for all three traffic-efficiency metrics, with a waiting time of 639.0 s, a queue length of 57.6 vehicles, and an average speed of 7.24 m/s. Nevertheless, the differences in mean waiting time among these three methods are relatively small compared with their observed variability. The results should therefore be interpreted as evidence of competitive nominal performance rather than conclusive statistical superiority.

The proposed controller also exhibits the lowest waiting-time standard deviation among IPPO, MacLight, and the proposed method. This indicates more consistent waiting-time outcomes across the evaluated stochastic realizations.

To assess nominal performance under a different network configuration, the controllers are also evaluated on the real-world-derived Hangzhou 4×4 network. In contrast to the regular synthetic grid, the Hangzhou benchmark contains heterogeneous road geometry and empirically derived traffic-flow distributions. It is used to examine whether the controllers remain effective beyond the controlled topology and stochastic demand-generation process of the synthetic environment.

Table 3 reports the waiting time, queue length, and average speed obtained on the Hangzhou network. Lower waiting time and queue length indicate better congestion management, whereas higher average speed indicates improved network mobility.

The relative ranking observed on the Hangzhou network is broadly consistent with that obtained on the synthetic grid. Fixed-Time, MAPPO, and IDQN exhibit substantially greater congestion than the policy-gradient methods. IPPO provides a strong decentralized baseline, while MacLight obtains the second-best mean result across the three reported traffic-efficiency metrics.

Table 3: Nominal performance on the real-world-derived Hangzhou 4×4 network. Results are reported as mean $\pm$ standard deviation over 80 episodes. Best mean values are highlighted in bold, and second-best mean values are underlined.

Method	Waiting Time (s) $\downarrow$	Queue Length (veh) $\downarrow$	Average Speed (m/s) $\uparrow$
Fixed-Time	12311.75 $\pm$ 3795.13	231.75 $\pm$ 53.62	4.63 $\pm$ 1.21
MAPPO	7764.86 $\pm$ 4474.50	191.74 $\pm$ 51.44	4.81 $\pm$ 0.60
IDQN	5399.25 $\pm$ 2521.26	87.02 $\pm$ 28.29	7.15 $\pm$ 0.70
IPPO	1301.20 $\pm$ 1696.11	36.72 $\pm$ 30.96	8.55 $\pm$ 0.79
MacLight	<u>1069.30 $\pm$ 1735.26</u>	<u>35.99 $\pm$ 26.50</u>	<u>8.56 $\pm$ 0.71</u>
Ours	784.97 $\pm$ 783.99	32.28 $\pm$ 23.25	8.65 $\pm$ 0.67

The proposed controller achieves the best mean performance on the Hangzhou network. Compared with MacLight, it reduces mean waiting time from 1069.3 s to 784.97 s, corresponding to a reduction of approximately 26.6%. It also reduces mean queue length by approximately 10.3% and increases average speed by approximately 1.1%.

These findings provide additional evidence that the proposed controller remains effective under the heterogeneous topology and traffic distribution of the Hangzhou benchmark. However, the standard deviations, particularly those associated with waiting time, remain substantial. The reported improvements should therefore be interpreted in terms of mean performance rather than as evidence of statistical significance. Moreover, since Hangzhou is evaluated only under nominal conditions, conclusions regarding robustness and post-disruption recovery remain based on the controlled synthetic-grid experiments.

6.3 Robustness under Perturbations

To rigorously evaluate the robustness of the proposed framework, we subjected the agents to the non-stationary perturbation scenarios defined in the previous section. We focus on the active perturbation window to assess the controllers' ability to maintain traffic performance while the disturbance is present.

Given the extensive benchmarking performed, where MacLight emerged as the most competitive baseline, we restrict our robustness analysis to a direct comparison between our proposed architecture and MacLight. This allows for a focused examination of how the proposed controller behaves relative to the strongest implemented baseline.

Table 4 shows that the proposed controller achieves better mean absolute traffic performance than MacLight across the evaluated perturbation scenarios. It records lower waiting time and queue length and a higher average speed during each active perturbation window. Nevertheless, the magnitude of the relative degradation varies across scenarios, indicating that favourable absolute performance does not always correspond to a smaller change from the controller's own nominal baseline.

Table 4: Comparative analysis of robustness during the perturbation window.
Values are presented as **Mean [PDR%]**.

Scenario	Waiting Time (s)		Queue Length (veh)		Avg Speed (m/s)	
	MacLight	Ours	MacLight	Ours	MacLight	Ours
1. Normal (Baseline)	766.78	677.22	64.86	62.39	6.73	6.80
2. Rainy Weather	743.14	655.34	62.17	60.65	6.21	6.24
<i>PDR (%)</i>	<i>[-3.08%]</i>	<i>[-3.23%]</i>	<i>[-4.16%]</i>	<i>[-2.78%]</i>	<i>[7.66%]</i>	<i>[8.13%]</i>
3. Random Blockage	1412.96	1246.25	94.31	87.62	4.35	4.36
<i>PDR (%)</i>	<i>[84.27%]</i>	<i>[84.02%]</i>	<i>[45.40%]</i>	<i>[40.45%]</i>	<i>[35.33%]</i>	<i>[35.90%]</i>
4. W-E Failure	1579.66	1282.32	92.58	86.71	3.87	3.90
<i>PDR (%)</i>	<i>[106.01%]</i>	<i>[89.35%]</i>	<i>[42.73%]</i>	<i>[38.99%]</i>	<i>[42.53%]</i>	<i>[42.61%]</i>
5. S-N Failure	1389.29	1334.61	87.83	85.58	3.95	3.99
<i>PDR (%)</i>	<i>[81.18%]</i>	<i>[97.07%]</i>	<i>[35.41%]</i>	<i>[37.19%]</i>	<i>[41.27%]</i>	<i>[41.36%]</i>
6. Center Closure	1139.83	934.14	87.04	78.63	5.00	5.20
<i>PDR (%)</i>	<i>[48.65%]</i>	<i>[37.94%]</i>	<i>[34.19%]</i>	<i>[26.04%]</i>	<i>[25.71%]</i>	<i>[23.52%]</i>

The quantitative results presented in Table 4 provide a comprehensive validation of the proposed architecture’s robustness. A primary observation from the absolute performance metrics is that our method demonstrates strict superiority over the MacLight baseline across every tested scenario. Regardless of the perturbation type, whether it is a random disruption, a structural failure, or environmental perturbation, the proposed agent consistently achieves lower average waiting times, reduced queue lengths, and higher network speeds during the perturbation window compared to the baseline. This confirms that the spatio-temporal policies learned by our agent provide a more robust controller ensuring that even under severe stress, the network operates at a higher efficiency level than the state-of-the-art alternative.

Beyond absolute performance, the Performance Degradation Rate (PDR) reveals the agent’s structural adaptability. In scenarios involving severe topological changes, specifically the Random Blockage, West-East Corridor Failure, and Center Closure, our architecture exhibits significantly better relative robustness than MacLight. For instance, during the Center Closure, our method limits the queue length increase to just 26.04%, whereas the baseline suffers a 34.19% surge. Similarly, in the West-East Failure, our agent mitigates the catastrophic delay increase seen in MacLight (106% PDR) to a lower 89% PDR.

The analysis of the South-North (S-N) Failure scenario presents an interesting nuance where our method records a higher PDR (97.07%) compared to MacLight (81.18%). This phenomenon is largely attributable to the “baseline effect” rather than a lack of robustness. Because our agent operates at a significantly higher efficiency during normal conditions (677s waiting time vs. 766s for MacLight), a comparable absolute increase in delay results in a mathematically larger percentage jump. Crucially, despite this higher relative drop, our method’s absolute waiting time during the failure (1334s)

remains lower than MacLight’s (1389s), confirming that the agent retains better flow control even when the margin of improvement narrows under specific constraints.

Finally, under the simplified rainy-weather scenario, both controllers show lower waiting time and queue length but reduced average speed. This result likely reflects the selected parameter changes, as lower desired speeds and larger gaps can reduce downstream arrival rates and intersection pressure. The negative PDR values should therefore be interpreted as an effect of the simplified, literature-informed weather model rather than as evidence of improved robustness under rain.

6.4 Resilience and Recovery Behaviour

While the previous section evaluates robustness during the active perturbation, this section examines post-disruption resilience and recovery. We consider the three complementary indicators defined previously: the Area Under the Curve Ratio (ρ_{AUC}), which captures cumulative performance relative to nominal operation; Peak Loss (L_{peak}), which measures the maximum instantaneous deviation; and Recovery Time (T_{rec}), which characterizes the return toward pre-perturbation performance.

6.4.1 Scenario 1: Rainy Weather

This scenario evaluates the controllers under a temporary network-wide modification of desired vehicle speed and minimum spacing. These changes approximate selected effects of wet-road conditions but do not constitute a fully calibrated rain model.

Table 5 reports the resilience metrics for this scenario, while Fig. 11 shows the corresponding temporal dynamics for the proposed method.

Table 5: Resilience Metrics Comparison: Rainy Weather ($t \in [1200, 2400]$ s). Best values are highlighted in bold.

Traffic Metric	Algorithm	ρ_{AUC}	L_{peak}	T_{rec}
Waiting Time	MacLight	0.987	0.210	0.0
	Ours	0.997	0.202	0.0
Queue Length	MacLight	0.996	0.142	0.0
	Ours	0.997	0.139	4.0
Avg Speed	MacLight	0.764	0.336	9.0
	Ours	0.758	0.384	8.0

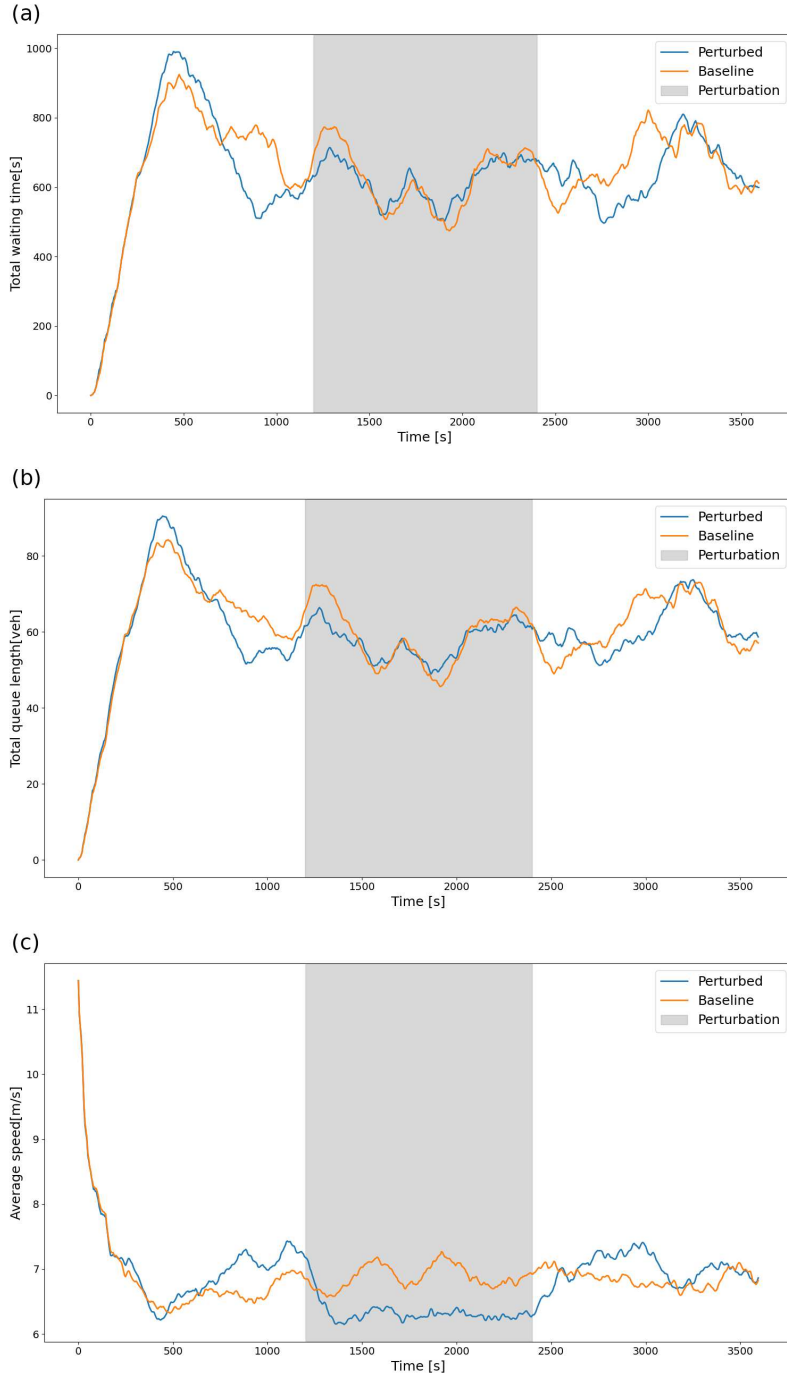


Fig. 11: Temporal dynamics under rainy weather ($t \in [1200, 2400]$ s) for the proposed method: (a) total waiting time; (b) total queue length; (c) average speed. The shaded region indicates the active perturbation window.

The rainy-weather results are mixed. The proposed controller obtains slightly higher AUC ratios and lower peak losses for waiting time and queue length, whereas MacLight performs better for average-speed AUC and peak loss. Recovery times remain short for both controllers. Given the simplified representation of rain through desired-speed and minimum-gap modifications, these differences should be interpreted as sensitivity results rather than evidence of resilience under fully realistic adverse-weather conditions.

6.4.2 Scenario 2: Random Road Blockage

This scenario introduces repeated structural changes by randomly restricting a new set of edges every 300 seconds during the perturbation interval.

Table 6 reports the resilience metrics for this scenario, while Fig. 12 shows the corresponding temporal dynamics for the proposed method.

Table 6: Resilience Metrics Comparison: Random Road Blockage ($t \in [1200, 2400]$ s). Best values are highlighted in bold.

Traffic Metric	Algorithm	ρ_{AUC}	L_{peak}	T_{rec}
Waiting Time	MacLight	0.800	0.350	329.0
	Ours	0.833	0.404	245.0
Queue Length	MacLight	0.859	0.272	277.0
	Ours	0.890	0.271	219.0
Avg Speed	MacLight	0.409	0.385	235.0
	Ours	0.496	0.449	212.0

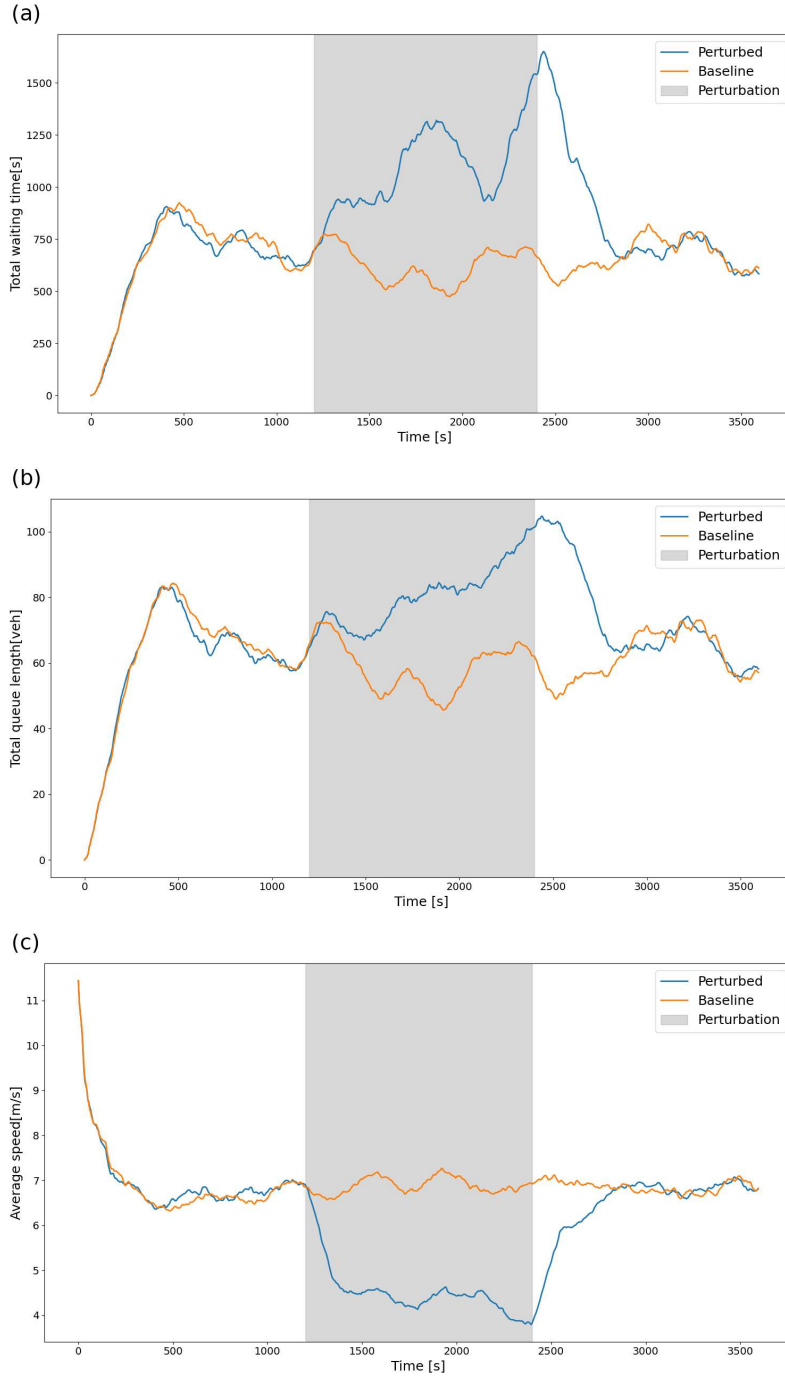


Fig. 12: Temporal dynamics under random road blockage ($t \in [1200, 2400]$ s) for the proposed method: (a) total waiting time; (b) total queue length; (c) average speed. The shaded region indicates the active perturbation window.

Under random blockages, the proposed controller achieves higher AUC ratios and shorter recovery times across all three traffic metrics, indicating better cumulative performance and faster recovery. MacLight nevertheless exhibits lower peak losses for waiting time and average speed, while the queue-length peak losses are nearly identical. The scenario therefore highlights a trade-off between worst-point degradation and cumulative recovery behaviour.

6.4.3 Scenario 3: West-East Corridor Failure

This scenario represents a sustained capacity restriction along the horizontal axis of the network, requiring traffic to redistribute toward alternative routes.

Table 7 reports the resilience metrics for this scenario, while Fig. 13 shows the corresponding temporal dynamics for the proposed method.

Table 7: Resilience Metrics Comparison: West-East Corridor Failure ($t \in [1200, 2400]$ s). Best values are highlighted in bold.

Traffic Metric	Algorithm	ρ_{AUC}	L_{peak}	T_{rec}
Waiting Time	MacLight	0.797	0.383	284.0
	Ours	0.820	0.384	297.0
Queue Length	MacLight	0.877	0.253	287.0
	Ours	0.888	0.261	278.0
Avg Speed	MacLight	0.412	0.385	282.0
	Ours	0.443	0.455	245.0

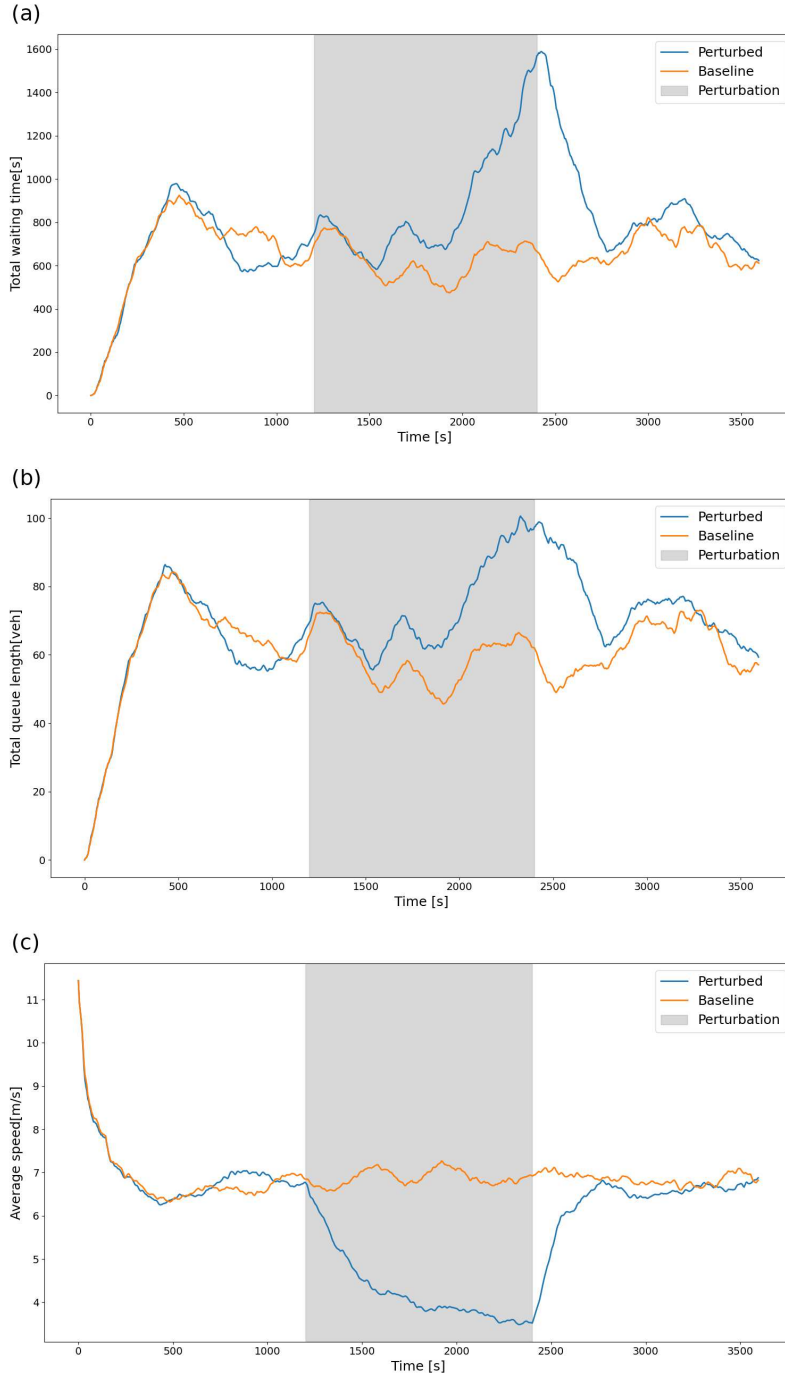


Fig. 13: Temporal dynamics under West-East corridor failure ($t \in [1200, 2400]$ s) for the proposed method: (a) total waiting time; (b) total queue length; (c) average speed. The shaded region indicates the active perturbation window.

In the West-East corridor failure, the proposed controller achieves higher AUC ratios across all three metrics, indicating better cumulative performance over the disruption and recovery period. MacLight records slightly lower peak losses and a shorter waiting-time recovery, whereas the proposed controller recovers queue length and average speed more rapidly. The relative advantage therefore depends on the resilience dimension considered.

6.4.4 Scenario 4: South-North Corridor Failure

This scenario restricts vertical network capacity and creates sustained bottlenecks for South-North traffic.

Table 8 reports the resilience metrics for this scenario, while Fig. 14 shows the corresponding temporal dynamics for the proposed method.

Table 8: Resilience Metrics Comparison:
South-North Corridor Failure ($t \in [1200, 2400]$ s).
Best values are highlighted in bold.

Traffic Metric	Algorithm	ρ_{AUC}	L_{peak}	T_{rec}
Waiting Time	MacLight	0.808	0.388	308.0
	Ours	0.810	0.406	267.0
Queue Length	MacLight	0.877	0.244	303.0
	Ours	0.890	0.255	238.0
Avg Speed	MacLight	0.421	0.390	283.0
	Ours	0.479	0.455	238.0

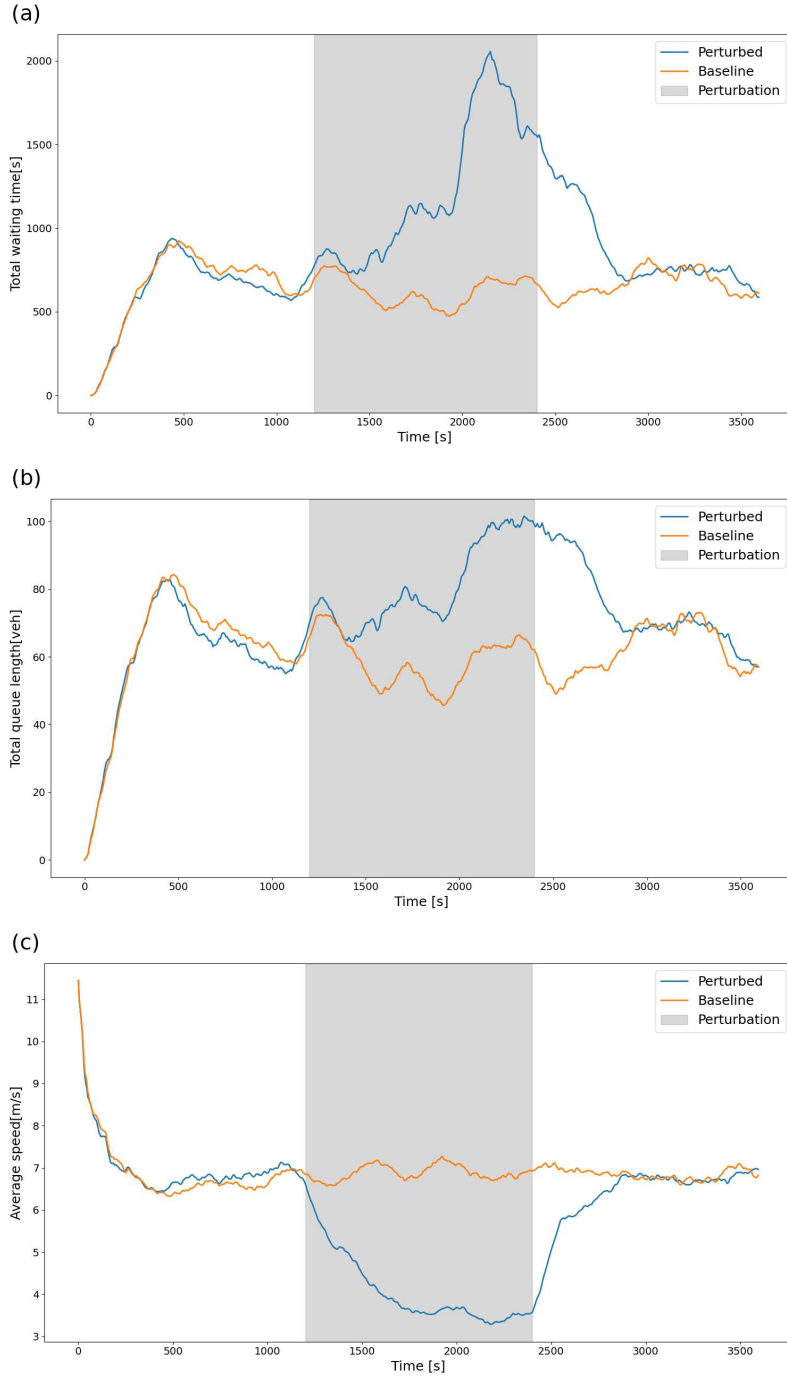


Fig. 14: Temporal dynamics under South-North corridor failure ($t \in [1200, 2400]$ s) for the proposed method: (a) total waiting time; (b) total queue length; (c) average speed. The shaded region indicates the active perturbation window.

For the South-North corridor failure, the proposed controller achieves higher AUC ratios and shorter recovery times across all three metrics. The largest recovery advance is observed for queue length, which returns to the specified threshold 65 seconds earlier than with MacLight. MacLight, however, records lower peak losses for all three metrics, indicating smaller worst-point deviations. The results therefore suggest better cumulative and recovery performance for the proposed controller despite somewhat larger instantaneous losses.

6.4.5 Scenario 5: Center Square Closure

This scenario restricts the central part of the network, forcing through-traffic to redistribute toward peripheral routes.

Table 9 reports the resilience metrics for this scenario, while Fig. 15 shows the corresponding temporal dynamics for the proposed method.

Table 9: Resilience Metrics Comparison: Center Square Closure ($t \in [1200, 2400]$ s). Best values are highlighted in bold.

Traffic Metric	Algorithm	ρ_{AUC}	L_{peak}	T_{rec}
Waiting Time	MacLight	0.856	0.304	238.0
	Ours	0.893	0.311	156.0
Queue Length	MacLight	0.892	0.224	249.0
	Ours	0.920	0.219	154.0
Avg Speed	MacLight	0.453	0.384	206.0
	Ours	0.491	0.446	125.0

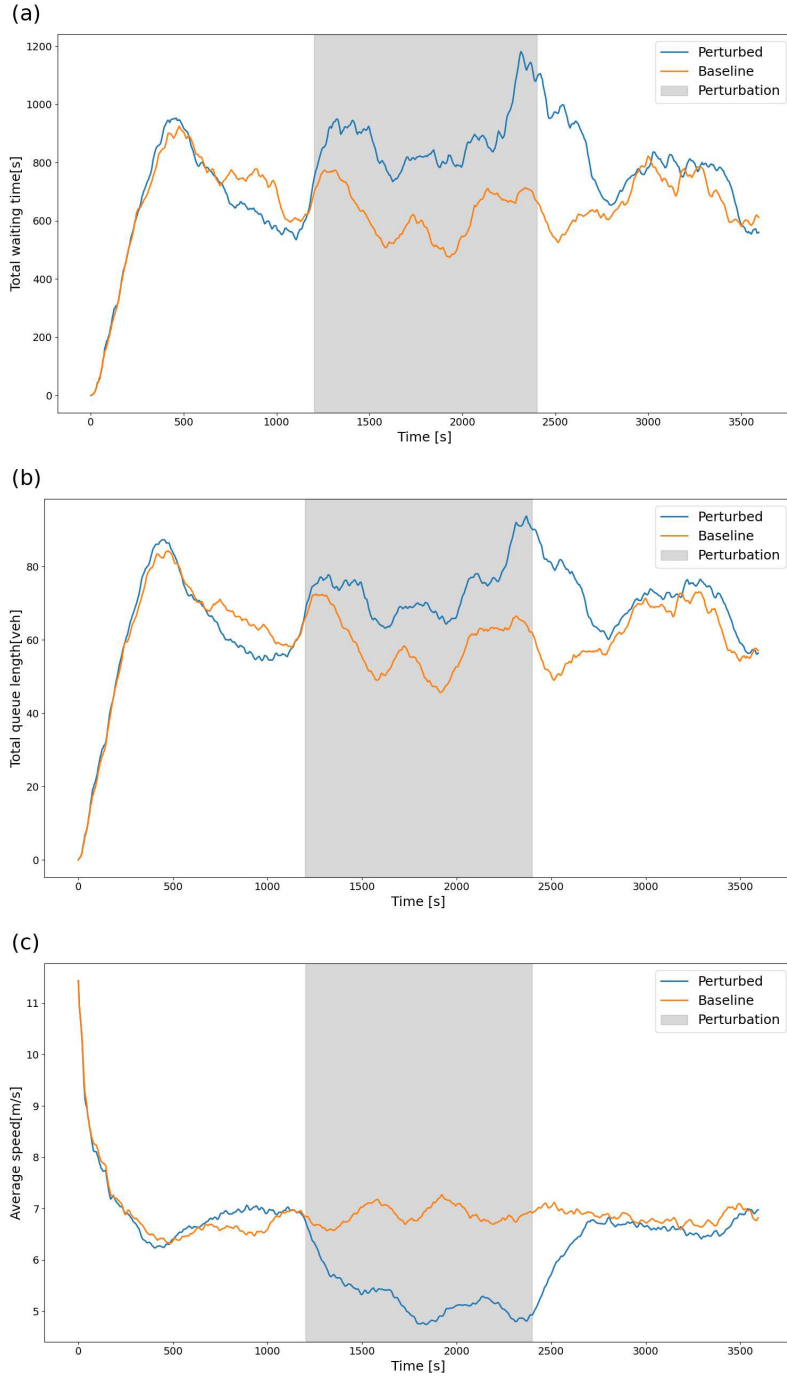


Fig. 15: Temporal dynamics under center square closure ($t \in [1200, 2400]$ s) for the proposed method: (a) total waiting time; (b) total queue length; (c) average speed. The shaded region indicates the active perturbation window.

1611 In the center-closure scenario, the proposed controller achieves higher AUC ratios
1612 and shorter recovery times across all three metrics. Recovery is 82 seconds faster for
1613 waiting time, 95 seconds faster for queue length, and 81 seconds faster for average
1614 speed. Peak-loss results are mixed: the proposed method performs slightly better for
1615 queue length, whereas MacLight records lower peak losses for waiting time and average
1616 speed. These results again show that cumulative performance, maximum degradation,
1617 and recovery speed capture complementary aspects of resilience.

1618

1619 7 Conclusion

1620

1621 This study investigates the behaviour of Multi-Agent Reinforcement Learning con-
1622 trollers beyond nominal traffic conditions. A controlled perturbation-based evaluation
1623 protocol was developed to examine traffic performance during stochastic demand
1624 variations, random and structured road-capacity restrictions, and post-disruption
1625 recovery. The proposed controller combines attention-weighted information from
1626 directly connected intersections with a temporal representation of recent traffic
1627 evolution.

1628 The experimental evaluation was conducted on a controlled stochastic grid and
1629 the real-world-derived Hangzhou network. Under nominal conditions, the proposed
1630 method achieved the best mean waiting time, queue length, and average speed on both
1631 networks, although the differences from the strongest baselines were limited in some
1632 cases and accompanied by substantial variability. The Hangzhou results provide addi-
1633 tional evidence that the controller remains effective under a heterogeneous topology
1634 and traffic-demand configuration.

1635 Under structural perturbations, the controller maintained favourable absolute traf-
1636 fic conditions and achieved lower degradation or faster recovery than MacLight in
1637 several scenarios, including recovery improvements of up to 95 seconds for selected
1638 metrics. These results show the importance of evaluating both the traffic conditions
1639 maintained during a disruption and the subsequent recovery process, rather than rely-
1640 ing exclusively on nominal efficiency or a single robustness indicator. They should
1641 nevertheless be interpreted as scenario-dependent empirical findings rather than as
1642 evidence that the architecture inherently guarantees resilience.

1643 The adverse-weather experiment was treated as a simplified, literature-informed
1644 sensitivity scenario. Although the selected changes in desired speed and minimum
1645 spacing reflect expected wet-road behaviour, rainfall affects several interacting physical
1646 and behavioural mechanisms that are difficult to reproduce through these parameters
1647 alone. The resulting improvements in some congestion metrics should therefore not be
1648 interpreted as evidence of increased robustness under rain.

1649 Future work will extend the evaluation to larger and more heterogeneous urban
1650 networks, investigate additional incident and sensing disturbances, and calibrate
1651 weather-related vehicle behaviour using observed traffic data. Further evaluation
1652 across a broader range of controllers and network configurations will also be necessary
1653 to assess the generality of the reported robustness and recovery patterns.

1654

1655 **Acknowledgements.** [TO COMPLETE: funding body written in full and grant
1656 number, or remove this heading.]

Statements and Declarations	1657
• Funding: [TO COMPLETE, e.g. “This work was supported by ... (Grant No. ...)” or “No funding was received for conducting this study.”]	1658
• Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.	1659
• Ethics approval and consent to participate: Not applicable.	1660
• Consent for publication: Not applicable.	1661
• Data availability: The Hangzhou 4×4 network and demand data are publicly available benchmark data (Lee et al. 2025). The synthetic 4×4 grid, the stochastic demand generator and the perturbation scenarios are available at [TO COMPLETE: repository URL / DOI].	1662
• Materials availability: Not applicable.	1663
• Code availability: [TO COMPLETE: repository URL, or “available from the corresponding author upon reasonable request”].	1664
• Author contribution: [TO CONFIRM] Conceptualization: Wissem Yousfi, Rim Slama, Pierre-Antoine Laharotte; Methodology, Software, Formal analysis, Visualization: Wissem Yousfi; Writing – original draft: Wissem Yousfi; Writing – review and editing: Rim Slama, Pierre-Antoine Laharotte; Supervision: Rim Slama, Pierre-Antoine Laharotte.	1665
	1666
	1667
	1668
	1669
	1670
	1671
	1672
	1673
	1674
	1675
	1676
	1677
References	1678
Alegre LN, Bazzan AL, da Silva BC (2021) Quantifying the impact of non-stationarity in reinforcement learning-based traffic signal control. <i>PeerJ Computer Science</i> 7:e575	1679
Aslani M, Seipel S, Mesgari MS, et al (2018) Traffic signal optimization through discrete and continuous reinforcement learning with robustness analysis in downtown Tehran. <i>Advanced Engineering Informatics</i> 38:639–655. https://doi.org/10.1016/j.aei.2018.08.002	1680
Ault J, Sharon G (2021) Reinforcement learning benchmarks for traffic signal control. In: <i>Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track</i>	1681
Birolini A (2007) <i>Reliability engineering: theory and practice</i> . Springer Science & Business Media	1682
Brody S, Alon U, Yahav E (2022) How attentive are graph attention networks? In: <i>International Conference on Learning Representations</i>	1683
Chu T, Wang J, Codecà L, et al (2019) Multi-agent deep reinforcement learning for large-scale traffic signal control. <i>IEEE Transactions on Intelligent Transportation Systems</i> 21(3):1086–1095	1684
Da L, Mei H, Sharma R, et al (2023) Sim2Real transfer for traffic signal control. In: <i>2023 IEEE 19th International Conference on Automation Science and Engineering</i>	1685
	1686
	1687
	1688
	1689
	1690
	1691
	1692
	1693
	1694
	1695
	1696
	1697
	1698
	1699
	1700
	1701
	1702

1703 (CASE), IEEE, pp 1–2
 1704
 1705 De Witt CS, Gupta T, Makoviichuk D, et al (2020) Is independent learning all you need
 1706 in the StarCraft multi-agent challenge? arXiv preprint [https://doi.org/10.48550/
 1707 arXiv.2011.09533](https://doi.org/10.48550/arXiv.2011.09533)
 1708
 1709 Garg D (2020) Autonomous traffic signal control using deep reinforcement learning
 1710
 1711 Koonce P, et al (2008) Traffic signal timing manual. Tech. rep., United States. Federal
 1712 Highway Administration
 1713
 1714 Lee S, Lyu H, Gong Y, et al (2025) MacLight: Multi-scene aggregation convolutional
 1715 learning for traffic signal control. In: Proceedings of the 24th International Confer-
 1716 ence on Autonomous Agents and Multiagent Systems. International Foundation for
 1717 Autonomous Agents and Multiagent Systems, pp 1263–1271
 1718
 1719 Li L, Lv Y, Wang FY (2016) Traffic signal timing via deep reinforcement learning.
 1720 IEEE/CAA Journal of Automatica Sinica 3(3):247–254
 1721
 1722 Li Z, Yu H, Zhang G, et al (2021) Network-wide traffic signal control optimization
 1723 using a multi-agent deep reinforcement learning. Transportation Research Part C:
 1724 Emerging Technologies 125:103059
 1725
 1726 Lu J, Ruan J, Jiang H, et al (2023) DuaLight: Enhancing traffic signal control by
 1727 leveraging scenario-specific and scenario-shared knowledge. arXiv preprint <https://doi.org/10.48550/arXiv.2312.14532>
 1728
 1729 Mnih V, Kavukcuoglu K, Silver D, et al (2015) Human-level control through deep
 1730 reinforcement learning. Nature 518(7540):529–533
 1731
 1732 Nguyen DVA, Azevedo CL, Toledo T, et al (2025) Robustness of reinforcement
 1733 learning-based traffic signal control under incidents: A comparative study. arXiv
 1734 preprint <https://doi.org/10.48550/arXiv.2506.13836>
 1735
 1736 Ni T, Eysenbach B, Salakhutdinov R (2022) Recurrent model-free RL can be a strong
 1737 baseline for many POMDPs. In: International Conference on Machine Learning,
 1738 PMLR, pp 16691–16723
 1739
 1740 Nishi T, Otaki K, Hayakawa K, et al (2018) Traffic signal control based on reinforce-
 1741 ment learning with graph convolutional neural nets. In: 2018 21st International
 1742 Conference on Intelligent Transportation Systems (ITSC), IEEE, pp 877–883
 1743
 1744 Robertson DI, Bretherton RD (1991) Optimizing networks of traffic signals in real
 1745 time: the SCOOT method. IEEE Transactions on Vehicular Technology 40(1):11–15
 1746
 1747 Rodrigues F, Lima Azevedo CM (2019) Towards robust deep reinforcement learning
 1748 for traffic signal control: Demand surges, incidents and sensor failures. In: 2019
 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, pp 3559–3566,

https://doi.org/10.1109/ITSC.2019.8917451	1749
Ruan J, Li Z, Wei H, et al (2024) CoSLight: Co-optimizing collaborator selection and decision-making to enhance traffic signal control. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, pp 2500–2511	1750 1751 1752 1753 1754
Schulman J, Moritz P, Levine S, et al (2015) High-dimensional continuous control using generalized advantage estimation. arXiv preprint https://doi.org/10.48550/arXiv.1506.02438	1755 1756 1757 1758
Schulman J, Wolski F, Dhariwal P, et al (2017) Proximal policy optimization algorithms. arXiv preprint https://doi.org/10.48550/arXiv.1707.06347	1759 1760 1761
Shi T, Devailly FX, Larocque D, et al (2024) Improving the generalizability and robustness of large-scale traffic signal control. IEEE Open Journal of Intelligent Transportation Systems 5:2–15. https://doi.org/10.1109/OJITS.2023.3331689	1762 1763 1764 1765
Sims AG, Dobinson KW (1980) The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits. IEEE Transactions on Vehicular Technology 29(2):130–137. https://doi.org/10.1109/T-VT.1980.23833	1766 1767 1768 1769
Tan KL, Sharma A, Sarkar S (2020) Robust deep reinforcement learning for traffic signal control. Journal of Big Data Analytics in Transportation 2(3):263–274. https://doi.org/10.1007/s42421-020-00029-6	1770 1771 1772 1773
Varaiya P (2013) Max pressure control of a network of signalized intersections. Transportation Research Part C: Emerging Technologies 36:177–195. https://doi.org/10.1016/j.trc.2013.08.014	1774 1775 1776 1777
Veličković P, Cucurull G, Casanova A, et al (2018) Graph attention networks. In: International Conference on Learning Representations	1778 1779 1780
Wang Y, Xu T, Niu X, et al (2020) STMARL: A spatio-temporal multi-agent reinforcement learning approach for cooperative traffic light control. IEEE Transactions on Mobile Computing 21(6):2228–2242	1781 1782 1783
Weber TM, Driesch P, Schramm D (2019) Introducing road surface conditions into a microscopic traffic simulation. In: SUMO, pp 172–186	1784 1785 1786
Webster FV (1958) Traffic signal settings. Road Research Technical Paper 39, Road Research Laboratory, HMSO, London	1787 1788 1789
Wei H, Chen C, Zheng G, et al (2019a) PressLight: Learning max pressure control to coordinate traffic signals in arterial network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 1290–1298, https://doi.org/10.1145/3292500.3330949	1790 1791 1792 1793 1794

1795 Wei H, Xu N, Zhang H, et al (2019b) CoLight: Learning network-level cooperation for
1796 traffic signal control. In: Proceedings of the 28th ACM International Conference on
1797 Information and Knowledge Management. ACM, pp 1913–1922, [https://doi.org/10.](https://doi.org/10.1145/3357384.3357902)
1798 [1145/3357384.3357902](https://doi.org/10.1145/3357384.3357902)
1799
1800 Wei H, Zheng G, Gayah V, et al (2019c) A survey on traffic signal control methods.
1801 arXiv preprint <https://doi.org/10.48550/arXiv.1904.08117>
1802
1803 Wu L, Wang M, Wu D, et al (2021) DynSTGAT: Dynamic spatial-temporal graph
1804 attention network for traffic signal control. In: Proceedings of the 30th ACM
1805 International Conference on Information and Knowledge Management. ACM, pp
1806 2150–2159, <https://doi.org/10.1145/3459637.3482254>
1807
1808 Wu Q, Li M, Shen J, et al (2023) TransformerLight: A novel sequence modeling based
1809 traffic signaling mechanism via gated transformer. In: Proceedings of the 29th ACM
1810 SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, pp 2639–
1811 2647, <https://doi.org/10.1145/3580305.3599530>
1812
1813 Yu C, Velu A, Vinitzky E, et al (2022) The surprising effectiveness of PPO in coop-
1814 erative multi-agent games. Advances in Neural Information Processing Systems
1815 35:24611–24624
1816
1817 Zeinaly Z, Sojoodi M, Bolouki S (2023) A resilient intelligent traffic signal con-
1818 trol scheme for accident scenario at intersections via deep reinforcement learning.
1819 Sustainability 15(2):1329. <https://doi.org/10.3390/su15021329>
1820
1821 Zeng J, Yu C, Yang X, et al (2024) CityLight: A universal model for coordinated
1822 traffic signal control in city-scale heterogeneous intersections. arXiv preprint [https:](https://doi.org/10.48550/arXiv.2406.02126)
1823 [//doi.org/10.48550/arXiv.2406.02126](https://doi.org/10.48550/arXiv.2406.02126)
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840