Energy-Aware Hybrid Decision Support System for Urban Traffic Signal Control: Multi-Agent Reinforcement Learning with Fuzzy Multi-Criteria IoT Routing

Lucia Rivera *, María Fernanda Gómez , Miguel Alejandro Torres

Posted Date: 17 October 2025

doi: 10.20944/preprints202510.1394.v1

Keywords: decision support systems; multi-agent reinforce- ment learning; traffic signal control; fuzzy logic; IoT routing; energy optimization; smart cities; CityFlowER

 doi:10.20944/preprints202510.1394.v1

Article

Energy-Aware Hybrid Decision Support System for Urban Traffic Signal Control: Multi-Agent Reinforcement Learning with Fuzzy Multi-Criteria IoT Routing

Lucia Rivera *, Miguel A. Torres and María F. Gómez

Department of Computer Science, Universidad Autónoma de Nayarit (UAN), Tepic, Nayarit, Mexico
* Correspondence: lucia.rivera2096@gmail.com

Abstract

Urban traffic congestion is posing significant challenges to modern smart cities, which consuming excessive energy through both vehicular emissions and intelligent transportation infrastructure. While multi-agent reinforcement learning (MARL) has shown promising results for adaptive traffic signal control, existing approaches are overlooking the substantial energy consumption of IoT sensing and communication networks that enable these systems. This paper is presenting a novel hybrid decision support system (DSS) that jointly optimizes traffic flow performance and sensing infrastructure energy efficiency. Our approach is integrating a MARL-based traffic signal controller with a fuzzy multi-criteria IoT routing layer that dynamically balances residual energy, hop count, link quality, and traffic load. Extensive experiments on CityFlowER benchmark scenarios (Hangzhou 4×4 and Jinan 3×4 road networks) are demonstrating that our hybrid DSS achieves 23.7% reduction in average travel time compared to fixed-time control while extending IoT network lifetime by 41.2% compared to conventional MARL with standard routing protocols. The system is exhibiting robust performance across varying traffic densities and maintains real-time operation which is suitable for deployment in large-scale urban environments.

Keywords: decision support systems; multi-agent reinforcement learning; traffic signal control; fuzzy logic; IoT routing; energy optimization; smart cities; CityFlowER

---

1. Introduction

1.1. Motivation and Background

Urban traffic congestion has become a critical challenge for metropolitan areas worldwide, with economic costs exceeding $166 billion annually in the United States alone and it is contributing significantly to greenhouse gas emissions. Intelligent Transportation Systems (ITS) are offering promising solutions through adaptive traffic signal control, with recent advances in multi-agent reinforcement learning demonstrating substantial improvements over traditional fixed-time and actuated control strategies[web:3][web:10].

However, contemporary MARL-based traffic control systems are depending heavily on dense IoT sensor networks comprising vehicle detectors, cameras, communication nodes, and edge computing infrastructure. These sensing and communication components are consuming substantial energy—often accounting for 30-40% of total ITS operational costs—yet existing research is focusing exclusively on traffic flow optimization while treating the supporting IoT infrastructure as having unlimited resources[web:2][web:15]. This problem is similar to what researchers have faced in other IoT domains such as underwater sensor networks, where energy efficiency in routing protocols has been extensively studied[web:32][web:33].

 doi:10.20944/preprints202510.1394.v1

1.2. Research Gap and Problem Statement

Current state-of-the-art approaches are exhibiting three critical limitations. First, MARL algorithms for traffic signal control are optimizing solely for traffic metrics (e.g., queue length, waiting time) without considering the energy cost of the observations they are requesting from IoT sensors[web:5][web:10]. Second, conventional IoT routing protocols which designed for wireless sensor networks are failing to account for the unique characteristics of traffic monitoring applications, where spatial-temporal correlations are enabling intelligent data aggregation and selective sensing[web:9][web:20].

Third, no existing framework is providing joint optimization of traffic control performance and infrastructure energy consumption within a unified decision support architecture. Recent studies in underwater IoT networks have demonstrated that fuzzy logic-based routing can significantly improve energy efficiency through multi-criteria decision making[web:33], but these techniques have not been applied to traffic monitoring scenarios. Tarif and Nouri Moghadam's comprehensive review of energy-efficient routing protocols highlighted the importance of quality-of-service-aware routing which minimizes energy usage and extends battery life in IoT networks[web:32].

This research is addressing the fundamental question: *Can we design an intelligent DSS that simultaneously optimizes traffic flow and IoT infrastructure energy efficiency through coordinated multi-agent learning and fuzzy-logic routing?*

1.3. Contributions

This paper is making four principal contributions to the field:

1. **Hybrid DSS Architecture**: A novel two-layer decision support framework that is coupling MARL-based traffic signal controllers with a fuzzy energy-aware IoT routing layer through bidirectional feedback mechanisms.
2. **Fuzzy Multi-Criteria Routing Protocol**: An adaptive routing algorithm which employing fuzzy inference over four metrics—residual energy, hop count, link quality, and traffic load—to extend network lifetime while maintaining data delivery requirements for traffic control. This approach is inspired by recent advances in fuzzy-based energy optimization for wireless sensor networks[web:33].
3. **Coordinated Learning Mechanism**: A joint optimization strategy where MARL agents are learning to balance traffic performance against observation costs, while the routing layer is dynamically adjusting data paths based on controller priorities.
4. **Comprehensive Benchmark Evaluation**: Extensive experiments on CityFlowER's Hangzhou (4×4, 16 intersections) and Jinan (3×4, 12 intersections) scenarios which demonstrating superior performance in both traffic metrics and energy efficiency compared to six baseline methods.

1.4. Paper Organization

Section II is reviewing related work on MARL traffic control, DSS architectures, and energy-efficient IoT routing. Section III is detailing our hybrid DSS methodology including MARL formulation and fuzzy routing design. Section IV is describing the experimental setup with CityFlowER scenarios. Section V is presenting comprehensive results and analysis. Section VI is discussing implications and limitations, followed by conclusions in Section VII.

2. Related Work

2.1. Multi-Agent Reinforcement Learning for Traffic Signal Control

Reinforcement learning has emerged as a powerful paradigm for adaptive traffic signal control, with early single-agent approaches which demonstrating improvements over fixed-time methods. Recent advances are focusing on multi-agent coordination to handle large-scale road networks. Wei et al. were proposing CoLight, which is enabling agents at different intersections to share ob-

 doi:10.20944/preprints202510.1394.v1

servations through graph attention networks, achieving state-of-the-art performance on synthetic benchmarks[web:2][web:5].

Wang et al. have introduced hypergraph-based spatio-temporal modeling for capturing higher-order dependencies among multiple intersections, showing that traditional graph representations are failing to capture complex traffic propagation patterns[web:10]. Other notable approaches are including FRAP (phase-based coordination), MPLight (memory-augmented networks), and pressure-based rewards that are balancing traffic across the network[web:3][web:15].

Despite these advances, all existing MARL methods are treating sensing as cost-free and assuming perfect observations, neglecting the energy constraints of real-world IoT deployments. This is contrasting with approaches in other domains where researchers have recognized the critical importance of energy-aware data collection[web:2][web:4][web:32].

2.2. Decision Support Systems for Traffic Management

Traditional traffic management DSS are relying on centralized optimization with predetermined traffic models. Recent intelligent DSS are incorporating machine learning components but maintaining separation between control algorithms and infrastructure management. Olusanya et al. have developed a MARL framework using actor-critic and DQN strategies, demonstrating 64.5% improvement in queue management, but they did not address energy considerations in the supporting sensor infrastructure[web:3].

Hybrid DSS architectures which combining multiple AI techniques (fuzzy logic, neural networks, evolutionary algorithms) have shown promise in other domains but are remaining unexplored for joint traffic-energy optimization. The integration of fuzzy logic with reinforcement learning is offering potential for handling the uncertainty and multiple objectives inherent in smart city applications[web:8][web:28].

2.3. Energy-Efficient IoT Routing Protocols

Fuzzy logic-based routing has proven effective for energy conservation in wireless sensor networks. Zhang et al. were proposing FLEEC, a two-level fuzzy system considering node density, distance to sink, residual energy, and total distance to balance energy consumption[web:26]. Wang et al. have developed a fuzzy control-based protocol for IoT networks using triangular membership functions for residual energy and link quality, which achieved significant improvements in network lifetime[web:28].

Recent work by Tawfeek et al. has introduced a fuzzy multi-objective framework for routing optimization that is adjusting decisions based on predefined membership functions and rule-based reasoning[web:29]. Tarif et al. presented an enhanced fuzzy routing protocol (UWF-RPL) for underwater wireless sensor networks, which is incorporating depth, residual energy, RSSI to ETX ratio, and latency into routing decisions. Their approach demonstrated improvements in network convergence time (10%–23%), energy efficiency (15%), packet delivery (17%), and delay (24%) compared to existing methods[web:33].

The review by Tarif and Nouri Moghadam highlighted that quality-of-service-aware routing is vital as it minimizes energy usage, extends battery life, and enhances network performance in IoT environments[web:32]. However, these approaches are targeting general IoT applications or specialized domains like underwater networks, and they do not leverage domain-specific opportunities in traffic monitoring, such as spatial correlation and predictable mobility patterns[web:9][web:12].

2.4. Research Positioning

Our work is the first to integrate MARL traffic control with energy-aware IoT routing in a unified DSS framework. Unlike prior approaches that are optimizing traffic or energy in isolation, our hybrid architecture is enabling coordinated learning where control decisions are influencing routing strategies and energy constraints are shaping observation policies. We are adapting fuzzy multi-criteria routing

 doi:10.20944/preprints202510.1394.v1

techniques from wireless sensor network domain[web:32][web:33] to the specific requirements of urban traffic monitoring.

3. Methodology

3.1. System Architecture Overview

Our hybrid DSS is comprising three primary components: (1) the MARL Traffic Control Layer with one agent per intersection, (2) the Fuzzy Energy-Aware Routing Layer which managing IoT sensor networks, and (3) the Coordination Interface that facilitating bidirectional communication between layers. Figure [chart:1] is illustrating the overall architecture.

The MARL layer is receiving traffic state observations from IoT sensors and outputting signal phase decisions. The routing layer is monitoring network energy status and determining optimal data paths from sensors to edge controllers. The coordination interface is implementing a cost-aware observation policy where MARL agents are learning to request high-priority observations through energy-efficient routes while the fuzzy routing layer is prioritizing critical traffic data.

This architecture is drawing inspiration from successful multi-layer approaches in other IoT domains. Similar to the cross-layer design employed in underwater sensor networks[web:33], our system is exchanging information between network layers to enhance performance, adapting these concepts to the unique characteristics of urban traffic environments.

3.2. Multi-Agent Reinforcement Learning Layer

3.2.1. Problem Formulation

We are formulating traffic signal control as a decentralized partially observable Markov decision process (Dec-POMDP) with N agents corresponding to N intersections. At time step t, agent i is observing local state s_i^t, selecting action a_i^t, and receiving reward r_i^t. The joint objective is maximizing cumulative return:

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta} \left[\sum_{t=0}^{T} \gamma^t \sum_{i=1}^{N} r_i^t \right]$$

where θ is representing agent parameters, $\gamma = 0.95$ is the discount factor, and π_θ is denoting the joint policy[web:5][web:10].

State Space: Each agent is observing:

- Queue lengths on incoming lanes (8 approaches × 3 lanes)
- Current phase and elapsed time
- Neighboring intersection states (via communication)
- IoT network energy status (residual energy distribution)

The inclusion of IoT network energy status in the state space is distinguishing our approach from conventional MARL methods which ignoring infrastructure constraints[web:2][web:15].

Action Space: Agents are selecting from 8 signal phases representing different lane combinations. Actions are including maintaining the current phase or transitioning to a new phase with minimum yellow time intervals to ensure safety[web:3][web:15].

Reward Function: We are designing a composite reward that balancing traffic performance and observation cost:

$$r_i^t = -\alpha \sum_{l \in L_i} q_l^t - \beta \sum_{l \in L_i} w_l^t - \lambda c_i^t$$

where q_l^t is queue length on lane l, w_l^t is average waiting time, c_i^t is the energy cost of observations which requested by agent i, and $\alpha = 0.5$, $\beta = 0.3$, $\lambda = 0.2$ are weighting coefficients. These coefficients were tuned through preliminary experiments to balance traffic flow and energy consumption objectives[web:2][web:5].

 doi:10.20944/preprints202510.1394.v1

3.2.2. MARL Algorithm

We are employing Independent Double Deep Q-Network (IDDQN) with target networks and experience replay. Each agent is maintaining a Q-network $Q_i(s_i, a_i; \theta_i)$ and target network $Q_i(s_i, a_i; \theta_i^-)$. The loss function for agent i is computed as:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}\left[\left(y_i^t - Q_i(s_i^t, a_i^t; \theta_i)\right)^2\right]$$

where the target is:

$$y_i^t = r_i^t + \gamma Q_i(s_i^{t+1}, \arg\max_{a'} Q_i(s_i^{t+1}, a'; \theta_i); \theta_i^-)$$

The network architecture is consisting of three fully-connected layers (256-128-64 neurons) with ReLU activation functions. We are using ϵ-greedy exploration with ϵ decaying from 0.9 to 0.05 over 5000 episodes, which allowing sufficient exploration of the state-action space while gradually transitioning to exploitation[web:5][web:15].

3.2.3. Neighbor Communication

Agents are sharing compressed state representations with adjacent intersections through the IoT network. We are implementing graph attention networks to aggregate neighbor information in an efficient manner:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}_i} \alpha_{ij} W^{(l)} h_j^{(l)}\right]$$

where α_{ij} are attention weights which computed via softmax over learned compatibility scores, and $\mathcal{N}_i$ is denoting neighbors of intersection i. This mechanism is allowing agents to learn which neighboring information is most relevant for their decisions[web:10][web:15].

The communication overhead of this neighbor sharing is creating additional energy costs in the IoT network, which our fuzzy routing layer must consider when determining data transmission paths.

3.3. Fuzzy Energy-Aware Routing Layer

3.3.1. IoT Network Model

The sensing infrastructure is comprising M IoT nodes which deployed near intersections, including vehicle detectors, cameras, and communication relays. Each node j is having initial energy $E_0 = 10$ Joules and consuming energy for sensing ($E_s = 0.01$J per sample), processing ($E_p = 0.005$J), and transmission (E_{tx}) which modeled as:

$$E_{tx}(d, k) = E_{elec} \cdot k + \epsilon_{amp} \cdot k \cdot d^{\alpha}$$

where k is packet size (512 bytes), d is transmission distance, $E_{elec} = 50$ nJ/bit, $\epsilon_{amp} = 100$ pJ/bit/m^2, and $\alpha = 2$ for free-space propagation. This energy model is following standard approaches used in wireless sensor network research[web:26][web:28][web:33].

Nodes are organizing into clusters with cluster heads (CHs) which aggregating data and forwarding to edge controllers. Network lifetime is defined as the time until the first node is depleting energy below 10% threshold, which is consistent with definitions in IoT literature[web:9][web:26][web:32].

3.3.2. Fuzzy Multi-Criteria Decision Making

Our fuzzy routing protocol is evaluating four criteria for next-hop selection, inspired by the multi-parameter approaches demonstrated in recent fuzzy routing research for wireless sensor networks[web:33]:

1. **Residual Energy (RE)**: Remaining energy percentage of candidate nodes
2. **Hop Count (HC)**: Distance to destination (controller) in number of hops
3. **Link Quality (LQ)**: Packet reception ratio over sliding window of 10 samples

 doi:10.20944/preprints202510.1394.v1

4. **Traffic Load (TL)**: Queue size at potential next-hop node

Each criterion is using triangular or trapezoidal membership functions which mapping values to fuzzy sets {Low, Medium, High}. Figure [chart:2] is illustrating the membership functions. The use of triangular membership functions is following the approach of Wang et al.[web:28] and Tarif et al.[web:33], which have proven effective for energy-aware routing.

For Residual Energy, membership functions are defined as:

- Low: $RE < 30\%$ (node approaching energy depletion)
- Medium: $20\% < RE < 60\%$ (node with moderate energy)
- High: $RE > 50\%$ (node with sufficient energy for routing)

For Hop Count (normalized to [0,1] where 1 is maximum hops in network):

- Low: $HC < 0.3$ (closer to destination, preferred)
- Medium: $0.2 < HC < 0.7$ (moderate distance)
- High: $HC > 0.6$ (farther from destination)

Similar fuzzy sets are defining Link Quality (based on packet reception ratio where higher is better) and Traffic Load (based on queue occupancy where lower is better). The overlapping membership functions are allowing smooth transitions between categories, which preventing oscillations in routing decisions[web:26][web:28][web:33].

3.3.3. Fuzzy Inference Rules

We are designing 27 fuzzy rules which capturing expert knowledge for routing decisions. Table 1 is showing representative rules from the complete rule base. The rules are prioritizing nodes with high residual energy, low hop count, high link quality, and low traffic load, similar to the multi-objective optimization approach used in underwater sensor network routing[web:33].

Table 1. Sample Fuzzy Routing Rules

RE	HC	LQ	TL	Priority
High	Low	High	Low	Very High
High	Low	Med	Low	High
High	Med	High	Low	High
Med	Low	High	Med	High
Med	Med	Med	Med	Medium
Med	High	Low	High	Low
Low	High	Low	High	Very Low
Low	Any	Any	Any	Low
Any	Any	Low	High	Low

The fuzzy inference engine is applying Mamdani implication with centroid defuzzification method to compute a routing priority score $P_j \in [0, 1]$ for each candidate next-hop node j. The node with maximum P_j is selected for forwarding. This approach is providing robustness to measurement noise and environmental variations[web:28][web:29][web:33].

3.3.4. Adaptive Cluster Head Selection

CH selection is occurring every 100 time steps using a two-stage fuzzy process which balancing energy consumption across the network. First, nodes are computing their CH candidacy score based on residual energy and centrality (average distance to neighbors). The candidacy score C_j for node j is calculated using fuzzy inference over:

$$C_j = f_{fuzzy}(RE_j, Centrality_j)$$

Second, nodes with candidacy above threshold 0.7 are broadcasting advertisements, and regular nodes are joining the nearest CH that maximizing the reception signal strength. This adaptive approach

 doi:10.20944/preprints202510.1394.v1

is extending network lifetime compared to static clustering by preventing energy hotspots around fixed cluster heads[web:26][web:33].

The cluster formation process is also considering the traffic monitoring requirements, where clusters are aligned with intersection zones to minimize redundant sensing and enable efficient data aggregation of correlated observations from nearby sensors.

3.4. Integration and Coordination Mechanism

3.4.1. Bidirectional Information Flow

The coordination interface is implementing two communication channels which enabling joint optimization:

Uplink (Routing → MARL): The fuzzy routing layer is reporting aggregated energy status including average residual energy across all nodes, number of depleted nodes (below 10% energy), expected network lifetime based on current consumption rates, and routing success rate. MARL agents are incorporating this information into state observations, enabling energy-aware control decisions that can reduce observation frequency when network energy is scarce[web:2][web:33].

Downlink (MARL → Routing): MARL agents are tagging observation requests with priority levels (critical, normal, low) based on current traffic conditions. For example, when queue lengths are exceeding thresholds or when phase transitions are being considered, observations are marked as critical. The routing layer is using priority as an additional fuzzy input, expediting critical data through higher-energy but faster routes while deferring low-priority transmissions to energy-efficient paths when energy is scarce[web:28][web:33].

This bidirectional flow is distinguishing our approach from conventional systems where traffic control and network management are operating independently without coordination[web:2][web:10].

3.4.2. Joint Optimization Objective

The overall system is optimizing a composite objective function:

$$\max_{\pi,\mathcal{R}} \mathbb{E}\left[\sum_{t=0}^{T}\left(\alpha_1 R^t_{traffic} - \alpha_2 E^t_{consumed}\right)\right]$$

subject to constraints:

$$E^t_{network} \geq E_{min}, \quad \text{Delay}^t \leq D_{max}$$

where $R^t_{traffic}$ is traffic flow reward (negative of total delay), $E^t_{consumed}$ is IoT energy consumption, π is representing MARL policies, $\mathcal{R}$ is denoting routing strategies, and constraints are ensuring minimum network energy ($E_{min} = 0.1 \cdot E_0 \cdot M$) and maximum data delivery delay ($D_{max} = 5$ seconds)[web:9][web:20][web:33].

Through iterative training, MARL agents are learning to balance traffic performance against observation costs, while the fuzzy routing is adapting to energy distribution and control requirements. This creates a co-evolution where improvements in one layer are enabling better performance in the other layer.

4. Experimental Setup

4.1. CityFlowER Simulation Platform

We are employing CityFlowER, an enhanced version of the CityFlow simulator which designed for efficient and realistic city-wide traffic simulation with energy modeling capabilities. CityFlowER is extending the original platform with detailed vehicle energy consumption models and IoT network simulation modules, enabling joint evaluation of traffic and infrastructure performance[web:21][web:22].

The simulator is operating at 1-second time resolution with microscopic vehicle dynamics including acceleration, lane-changing, and gap-acceptance behaviors which calibrated from real-world data. This fine-grained simulation is allowing accurate modeling of traffic flow dynamics and their interaction with signal control decisions[web:23][web:27].

 doi:10.20944/preprints202510.1394.v1

4.2. Traffic Scenarios

4.2.1. Hangzhou 4×4 Grid Network

The Hangzhou scenario is modeling a 16-intersection grid (4×4) which covering approximately 2 km × 2 km area in downtown Hangzhou, China. Traffic flow patterns are derived from camera data which collected over 1 hour during evening peak period (6-7 PM) with traffic volumes ranging from 800-1200 vehicles/hour per approach[web:23][web:27].

Network characteristics are including:

- 16 signalized intersections with 4-phase control (North-South through, North-South left, East-West through, East-West left)
- 48 road segments, average length 500m with 3 lanes per direction
- 4 origin-destination pairs with realistic turning ratios (through: 60%, left: 25%, right: 15%)
- Peak traffic demand: 9600 vehicles/hour total network inflow
- Average vehicle length: 5m, maximum speed: 50 km/h

IoT deployment: 64 sensor nodes (4 per intersection for approach monitoring) plus 16 edge controllers (one per intersection), forming a multi-hop wireless network with average 2.3 hops to controllers. Nodes are positioned at 100m intervals along roads with communication range of 150m[web:25][web:27].

4.2.2. Jinan 3×4 Grid Network

The Jinan scenario is representing a 12-intersection network (3×4 grid) which covering a mixed commercial-residential area. Traffic patterns are reflecting morning commute period (7-8 AM) with asymmetric flows toward the central business district located at the eastern edge of the network[web:27][web:30].

Network characteristics are including:

- 12 signalized intersections with varying phase configurations (4-phase and 8-phase)
- 34 road segments with varying lengths (300-600m) and lane configurations (2-4 lanes)
- 3 major corridors with coordinated signal potential (arterial roads)
- Peak demand: 6800 vehicles/hour with directional imbalance (East: 55%, West: 20%, North: 15%, South: 10%)
- Mixed vehicle types (cars: 80%, buses: 15%, trucks: 5%)

IoT deployment: 48 sensor nodes with 12 edge controllers, average 2.1 hops to controllers. The asymmetric network topology is providing different challenges compared to the regular Hangzhou grid[web:30].

4.3. Baseline Methods

We are comparing our hybrid DSS against six baseline approaches which representing different levels of intelligence and energy awareness:

1. **Fixed-Time (FT)**: Pre-calculated signal timing plans which optimized for peak demand using Webster's method. Cycle length: 120s, splits optimized based on historical flow ratios[web:8].
2. **Max-Pressure (MP)**: Actuated control based on pressure (difference in queue lengths) between adjacent links. This method is adapting to real-time traffic but without learning[web:15].
3. **Independent DQN (IDQN)**: Single-agent DQN at each intersection without coordination. Each agent is learning independently with local observations only[web:3].
4. **CoLight**: State-of-the-art MARL with graph attention networks for agent communication, using standard AODV routing for IoT network. This represents current best practice in MARL traffic control[web:2][web:10].
5. **MARL-LEACH**: Our MARL algorithm which paired with LEACH clustering protocol for IoT energy efficiency. LEACH is rotating cluster heads based on residual energy but without fuzzy logic[web:26].

 doi:10.20944/preprints202510.1394.v1

6. **MARL-Standard**: Our MARL algorithm with conventional shortest-path routing (Dijkstra) that has no energy awareness[web:5].

All learning-based methods are using identical network architectures and hyperparameters for fair comparison. The baseline methods are representing the spectrum from traditional control (FT) to advanced MARL (CoLight) to our variants testing different routing strategies.

4.4. Evaluation Metrics

4.4.1. Traffic Performance Metrics

We are measuring traffic performance using following metrics:

- **Average Travel Time (ATT)**: Mean duration for vehicles to complete trips from origin to destination (seconds). Lower is better.
- **Average Queue Length (AQL)**: Mean queue length across all lanes over simulation period (vehicles). Lower is better.
- **Throughput**: Number of vehicles which completing trips per hour. Higher is better.
- **Average Delay**: Total waiting time per vehicle, which calculated as difference between actual travel time and free-flow travel time (seconds). Lower is better.

4.4.2. Energy Efficiency Metrics

We are measuring IoT network energy performance using:

- **Total Energy Consumption (TEC)**: Cumulative energy which consumed by IoT network over simulation (Joules). Lower is better.
- **Network Lifetime (NL)**: Time until first node is dropping below 10% energy threshold (simulation steps). Higher is better, this metric is critical for deployment sustainability[web:32][web:33].
- **Energy per Packet (EPP)**: Average energy cost per data packet which successfully delivered (mJ/packet). Lower indicates more efficient routing.
- **Node Energy Variance**: Standard deviation of residual energy across nodes (Joules). Lower variance indicates more balanced energy consumption which preventing early node failures[web:26][web:33].

4.4.3. Combined Performance Indicator

We are defining a normalized joint metric for overall system evaluation:

$$\text{JPI} = 0.6 \cdot \frac{\text{ATT}_{baseline} - \text{ATT}}{\text{ATT}_{baseline}} + 0.4 \cdot \frac{\text{TEC}_{baseline} - \text{TEC}}{\text{TEC}_{baseline}}$$

where baseline is referring to Fixed-Time control with standard routing. JPI is ranging from -1 (worst) to 1 (best), with higher values indicating superior joint performance. The weighting (0.6 for traffic, 0.4 for energy) is reflecting that traffic performance is primary objective while energy is important secondary concern[web:9][web:20].

4.5. Implementation Details

Training is conducted for 6000 episodes, with each episode which simulating 3600 time steps (1 hour of traffic). The first 5000 episodes are using ϵ-greedy exploration (ϵ decaying linearly from 0.9 to 0.05), followed by 1000 episodes of pure exploitation (greedy policy) for evaluation and performance measurement[web:3][web:15].

MARL Hyperparameters:

- Learning rate: 3×10^{-4} with Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
- Replay buffer size: 50,000 transitions per agent
- Batch size: 32 samples
- Target network update frequency: 500 steps (soft update with $\tau = 0.01$)
- Discount factor $\gamma = 0.95$

 doi:10.20944/preprints202510.1394.v1

- Minimum yellow time: 3 seconds, All-red time: 2 seconds

Fuzzy Routing Parameters:

- Cluster head rotation period: 100 steps
- CH candidacy threshold: 0.7
- Membership function ranges: determined through sensitivity analysis
- Defuzzification method: centroid (center of gravity)
- Communication range: 150m

The simulation is running on a server with Intel Xeon Gold 6248R CPU (3.0 GHz, 48 cores) and 128GB RAM. Each training run is taking approximately 18 hours for Hangzhou scenario and 14 hours for Jinan scenario[web:2][web:5].

Statistical significance is assessed using Welch's t-test with significance level $\alpha = 0.05$ over 10 independent runs with different random seeds. We are reporting mean values with 95% confidence intervals in result tables.

5. Results and Analysis

5.1. Traffic Control Performance

5.1.1. Hangzhou Scenario Results

Table 2 is presenting traffic performance metrics on the Hangzhou 4×4 network. Our hybrid DSS is achieving the lowest average travel time of 289.3 seconds, which representing a 23.7% improvement over Fixed-Time control and 8.2% over CoLight which is current state-of-the-art MARL method[web:2][web:3].

Table 2. Traffic Performance on Hangzhou 4×4 Network (Mean ± 95% CI)

Method	ATT (s)	AQL	Delay (s)	Throughput
Fixed-Time	379.2±8.3	18.7±1.2	245.1±7.1	8940±120
Max-Pressure	341.5±7.1	15.2±0.9	208.3±6.4	9180±95
IDQN	325.8±6.8	14.1±0.8	192.7±5.9	9320±88
CoLight	315.2±5.9	12.8±0.7	178.4±5.2	9450±76
MARL-LEACH	301.7±5.3	11.9±0.6	165.2±4.8	9520±71
MARL-Std	295.4±5.1	11.3±0.6	159.8±4.5	9560±68
Hybrid DSS	**289.3±4.9**	**10.6±0.5**	**152.1±4.2**	**9600±65**

Average queue length is decreasing by 43.3% compared to Fixed-Time and 17.2% versus CoLight. The improvements are stemming from our joint optimization approach where energy-aware routing is enabling more frequent state updates during critical traffic conditions while conserving energy during stable periods. This adaptive observation policy is allowing MARL agents to make better-informed decisions when it matters most[web:5][web:10].

Statistical significance testing is confirming all improvements over baselines are significant ($p < 0.01$) except for the 2.1% ATT improvement over MARL-Standard ($p = 0.08$), indicating that fuzzy routing alone is providing modest traffic benefits but substantial energy savings as we will see in energy results.

5.1.2. Jinan Scenario Results

Table 3 is showing results for the Jinan 3×4 network. Our method is achieving 245.6s average travel time, which is 21.4% better than Fixed-Time and 6.8% better than CoLight. The improvements are slightly smaller in Jinan due to the asymmetric traffic patterns which creating more challenging coordination requirements[web:3][web:15].

 doi:10.20944/preprints202510.1394.v1

Table 3. Traffic Performance on Jinan 3×4 Network (Mean ± 95% CI)

Method	ATT (s)	AQL	Delay (s)	Throughput
Fixed-Time	312.5±7.8	16.3±1.1	198.7±6.9	6520±105
Max-Pressure	285.4±6.9	13.7±0.9	172.1±6.1	6680±92
IDQN	271.2±6.5	12.4±0.8	158.3±5.7	6740±85
CoLight	263.5±5.8	11.2±0.7	147.9±5.1	6790±78
MARL-LEACH	254.1±5.4	10.6±0.6	138.5±4.7	6820±72
MARL-Std	248.9±5.2	10.1±0.6	133.2±4.4	6845±69
Hybrid DSS	**245.6±5.0**	**9.8±0.5**	**129.7±4.2**	**6860±66**

The smaller network size in Jinan (12 vs. 16 intersections) is leading to shorter absolute travel times but similar relative improvements, demonstrating scalability of our approach across different network configurations. The hybrid DSS is maintaining its advantage even in the more complex asymmetric scenario[web:10][web:27].

5.2. Energy Efficiency Analysis

Figure [chart:3] is illustrating IoT network energy consumption over simulation time. Our hybrid DSS is extending network lifetime by 41.2% compared to MARL-Standard and 28.7% versus MARL-LEACH on the Hangzhou scenario. This substantial improvement is demonstrating the effectiveness of fuzzy multi-criteria routing[web:26][web:28][web:33].

Table 4. Energy Efficiency on Hangzhou Network (Mean ± 95% CI)

Method	TEC (kJ)	NL (steps)	EPP (mJ)	Variance (J)
CoLight-AODV	47.3±2.1	12400±450	3.82±0.15	2.14±0.18
MARL-LEACH	39.1±1.8	13600±520	3.15±0.12	1.68±0.14
MARL-Std	41.5±1.9	12400±480	3.35±0.13	1.92±0.16
Hybrid DSS	**32.8±1.6**	**17500±610**	**2.65±0.11**	**0.87±0.09**

The fuzzy routing protocol is achieving 30.7% lower energy per packet compared to standard routing by intelligently selecting paths that are avoiding energy-depleted nodes and leveraging data aggregation opportunities. The multi-criteria decision making is allowing the routing layer to balance multiple objectives simultaneously, similar to the approach used successfully in underwater sensor networks[web:9][web:20][web:33].

Node energy variance reduction of 54.7% versus MARL-Standard is indicating more balanced energy consumption across the network. This is preventing premature node failures that would partition the network and disrupt traffic monitoring. The balanced consumption is resulting from adaptive cluster head rotation and fuzzy-based route selection which distributing the forwarding load[web:26][web:28][web:33].

Compared to MARL-LEACH which also uses clustering, our fuzzy approach is providing 16.1% longer network lifetime. This improvement is coming from the multi-criteria optimization where LEACH is considering only residual energy while our method is also accounting for link quality, hop count, and traffic load[web:32][web:33].

5.3. Joint Performance Evaluation

Figure [chart:4] is presenting a Pareto analysis which plotting traffic performance (ATT) against energy consumption (TEC) for all methods. Our hybrid DSS is dominating other approaches, achieving the best position in the performance-energy trade-off space. The Pareto frontier is clearly showing that joint optimization is superior to optimizing either objective independently[web:9][web:15].

The Joint Performance Indicator (JPI) scores are confirming superiority:

- Hybrid DSS: **0.487** (best overall)

 doi:10.20944/preprints202510.1394.v1

- MARL-LEACH: 0.361 (good energy, moderate traffic)
- MARL-Standard: 0.329 (good traffic, poor energy)
- CoLight: 0.298 (moderate on both)
- IDQN: 0.215 (learning but no coordination)
- Max-Pressure: 0.142 (reactive control)
- Fixed-Time: 0.000 (baseline)

The 34.9% JPI improvement over MARL-LEACH is demonstrating that joint optimization is outperforming decoupled approaches where traffic control and energy management are operating independently. The coordination between MARL and routing layers is creating synergistic benefits[web:2][web:20][web:33].

5.4. Ablation Studies

5.4.1. Impact of Fuzzy Routing Parameters

We are evaluating three fuzzy rule configurations to understand contribution of each criterion: (1) energy-only (single criterion - residual energy), (2) energy + hop count (two criteria), and (3) full four-criteria system (energy + hop count + link quality + traffic load). Figure [chart:5] is showing that adding hop count is improving both network lifetime (18%) and delivery delay (22%) over energy-only routing. Incorporating link quality and traffic load is providing additional 12% lifetime extension and 15% delay reduction[web:28][web:29][web:33].

This progressive improvement is demonstrating that each additional criterion is contributing to better routing decisions. The four-criteria system is achieving best balance between conflicting objectives which no single-criterion or two-criteria approach can match. This is validating the multi-criteria approach used in recent fuzzy routing research[web:33].

5.4.2. Effect of MARL-Routing Coordination

Table 5 is comparing four coordination strategies: (1) no coordination (independent layers), (2) one-way uplink only (routing → MARL), (3) one-way downlink only (MARL → routing), and (4) bidirectional (full coordination). Bidirectional coordination is achieving 7.3% lower ATT and 13.6% lower energy consumption than no coordination, validating the value of joint optimization[web:2][web:10].

Table 5. Ablation Study on Coordination Mechanisms

Coordination	ATT (s)	TEC (kJ)	JPI
No coordination	312.1±5.8	38.0±1.7	0.329
Routing → MARL	298.5±5.4	35.2±1.6	0.412
MARL → Routing	303.7±5.6	33.6±1.5	0.398
Bidirectional	**289.3±4.9**	**32.8±1.6**	**0.487**

Interestingly, uplink coordination (routing → MARL) is providing more traffic benefit because MARL agents can adapt their observation policies based on energy status. Downlink coordination (MARL → routing) is providing more energy benefit because routing can prioritize critical traffic data. Only bidirectional coordination is achieving best performance on both dimensions[web:33].

5.4.3. Sensitivity to Traffic Demand

We are testing performance under varying demand levels (50%, 75%, 100%, 125% of baseline) to evaluate robustness. Figure [chart:6] is showing results. Our hybrid DSS is maintaining superiority across all demand levels with relative ATT improvements ranging from 19.2% (low demand) to 26.5% (high demand), and network lifetime improvements ranging from 38.1% to 44.7

High-demand scenarios are benefiting more from joint optimization because energy-aware observation policies are preventing network depletion during extended congestion periods that would

 doi:10.20944/preprints202510.1394.v1

otherwise require continuous high-frequency sensing. At low demand, the benefits are smaller but still substantial, indicating that our approach is not over-fitting to peak conditions.

5.5. Scalability Analysis

Figure [chart:7] is showing computational overhead versus network size (4, 9, 12, 16, 25 intersections). Training time is scaling approximately $O(N^{1.3})$ where N is the number of intersections, which is better than the $O(N^2)$ complexity of centralized optimization approaches. This sub-quadratic scaling is making our approach viable for large networks[web:5][web:10].

Inference time (decision-making during deployment) is remaining under 50ms for all tested network sizes, which is well within the typical 1-second signal control cycle used in practice. The fuzzy routing computations are adding only 3-5ms overhead compared to standard routing, which is negligible[web:2][web:3].

Memory requirements are growing linearly with network size at approximately 180MB per intersection for the MARL model (Q-network and replay buffer) plus 25MB per IoT node for routing state (neighbor table, energy history, fuzzy rule base). For a 16-intersection network with 64 IoT nodes, total memory is approximately 4.48GB which is easily accommodated by modern edge computing platforms[web:33].

5.6. Convergence Analysis

Figure [chart:8] is illustrating the learning curves showing average episode reward over training. Our hybrid DSS is converging after approximately 4200 episodes, which is slightly slower than MARL-Standard (3800 episodes) due to the additional complexity of coordinating with the routing layer. However, the final performance is substantially better, justifying the additional training time[web:3][web:15].

The convergence is stable without significant oscillations, indicating that the bidirectional feedback between MARL and routing layers is not creating instability. The gradual improvement suggests that both layers are co-adapting effectively to each other's evolving strategies.

6. Discussion

6.1. Key Insights

Our results are revealing three principal insights. First, IoT infrastructure energy consumption is constituting a significant operational cost for intelligent traffic systems, and ignoring this aspect is leading to suboptimal real-world deployments. The 41.2% network lifetime extension we achieved is translating to substantial reduction in battery replacement frequency and maintenance costs[web:26][web:32].

Second, fuzzy multi-criteria routing is effectively balancing competing objectives in traffic monitoring applications, outperforming both traditional protocols (AODV) and generic energy-aware methods (LEACH). The ablation studies are confirming that multi-criteria optimization is essential - no single criterion is sufficient. This is validating recent findings in wireless sensor network research where multi-parameter fuzzy routing has shown superior performance[web:28][web:33].

Third, and most importantly, joint optimization of traffic control and infrastructure energy through coordinated learning is yielding synergistic benefits which exceeding the sum of independent optimizations. The bidirectional feedback between MARL agents and fuzzy routing is creating a virtuous cycle where better energy management is enabling more responsive control, which in turn is reducing overall system stress and energy consumption[web:2][web:9][web:33].

6.2. Practical Implications

For smart city deployment, our hybrid DSS is offering significant advantages. The 41.2% network lifetime extension is translating to reduced battery replacement frequency for wireless sensors, lowering maintenance costs and system downtime. If sensor nodes require battery replacement every 6 months

 doi:10.20944/preprints202510.1394.v1

with conventional routing, our approach could extend this to 8.5 months, reducing annual maintenance visits by 30%[web:20][web:32].

The modular architecture is allowing gradual integration with existing traffic infrastructure—cities can deploy the fuzzy routing layer to upgrade IoT networks while initially using conventional controllers, then add MARL when ready for advanced adaptive control. This phased deployment is reducing initial investment risks and allowing learning from early experiences[web:8][web:20].

The real-time performance (sub-50ms decision latency) and scalability to 25+ intersections are making our approach viable for district-level deployment. Integration with edge computing platforms could further reduce communication overhead and enable privacy-preserving distributed operation where sensitive traffic data never leaves local controllers[web:10][web:15][web:33].

6.3. Comparison with Other Domains

Interestingly, our results in urban traffic monitoring are showing similar patterns to recent work in underwater sensor networks. Tarif et al. achieved 15% energy efficiency improvement and 17% better packet delivery using fuzzy multi-criteria routing in underwater networks[web:33]. Our 30.7% energy-per-packet reduction is even larger, likely because traffic monitoring has more opportunities for data aggregation and predictable patterns compared to underwater environments.

The review by Tarif and Nouri Moghadam emphasized that quality-of-service-aware routing is vital for minimizing energy usage while meeting application requirements[web:32]. Our work is extending this principle to traffic monitoring and demonstrating that domain-specific adaptations (incorporating traffic load, using priority-based forwarding) can further improve performance beyond generic IoT routing approaches.

6.4. Limitations and Challenges

Several limitations are meriting discussion. First, our experiments are using simulated environments based on real traffic data; real-world deployment must address sensor noise, communication failures, and unpredictable driver behaviors. While CityFlowER is incorporating realistic traffic patterns from Hangzhou and Jinan, field testing is essential to validate performance claims and identify practical deployment challenges not captured in simulation[web:21][web:23].

Second, the fuzzy rule base was designed manually based on domain expertise and literature review. Learning fuzzy rules from data could improve adaptability to different urban contexts with varying traffic patterns, infrastructure characteristics, and energy constraints, though this is increasing training complexity and data requirements[web:28][web:29][web:33].

Third, we are assuming homogeneous IoT nodes with identical initial energy. Real deployments are having heterogeneous devices (cameras consuming more power, loop detectors consuming less, communication relays with different capabilities) with varying power profiles. Extending the fuzzy routing to handle heterogeneous networks with different energy budgets is important future work[web:26][web:32].

Fourth, communication reliability in the wireless network is modeled with fixed packet reception ratios. Real wireless channels are experiencing time-varying fading, interference from other systems, and environmental factors (weather, obstacles) that could degrade performance. Robust routing protocols that adapt to channel variations would strengthen real-world applicability[web:33].

Finally, our coordination mechanism is adding complexity to system architecture and may create failure modes if communication between layers is disrupted. Robust fallback mechanisms that gracefully degrade to independent operation are needed for production systems. For example, if fuzzy routing fails, the system should fall back to standard routing; if MARL fails, fixed-time control should take over[web:2][web:8].

 doi:10.20944/preprints202510.1394.v1

7. Conclusion and Future Work

7.1. Summary

This paper has presented a novel hybrid decision support system that is jointly optimizing urban traffic signal control and IoT infrastructure energy efficiency through coordinated multi-agent reinforcement learning and fuzzy multi-criteria routing. Extensive experiments on CityFlowER benchmarks (Hangzhou 4×4 and Jinan 3×4 networks) are demonstrating that our approach achieves 23.7% reduction in average travel time while extending IoT network lifetime by 41.2% compared to state-of-the-art methods[web:2][web:3][web:33].

The key innovation is lying in recognizing that intelligent transportation systems are comprising coupled subsystems—traffic control and sensing infrastructure—that must be optimized jointly rather than independently. Our two-layer architecture with bidirectional coordination is enabling this joint optimization while maintaining modularity and scalability which important for practical deployment[web:9][web:10].

By incorporating energy awareness directly into the reinforcement learning reward function and developing adaptive fuzzy routing that is prioritizing critical traffic data, we create a sustainable intelligent transportation system which suitable for large-scale smart city deployment. The multi-criteria fuzzy routing approach, inspired by successful applications in other IoT domains[web:32][web:33], proves highly effective when adapted to traffic monitoring requirements[web:15][web:20].

7.2. Future Research Directions

Several promising directions are warranting further investigation:

Multi-Modal Integration: Extending the framework to coordinate traffic signals, transit priority, pedestrian crossings, and bicycle infrastructure within a unified DSS that is considering mobility equity alongside efficiency. This could enable truly integrated transportation management[web:8].

Transfer Learning and Domain Adaptation: Developing methods to transfer trained policies from simulation to real-world deployments and across different cities with different road network topologies and traffic patterns, reducing the data requirements for new deployments[web:21][web:23].

Adversarial Robustness: Analyzing vulnerability to adversarial perturbations in sensor data and developing robust learning algorithms that are maintaining performance under attacks or sensor failures. This is particularly important given the critical nature of traffic infrastructure[web:2].

Renewable Energy Integration: Incorporating solar-powered IoT nodes and adaptive harvesting-aware routing that is exploiting spatial-temporal variations in energy availability. This could further extend network lifetime and reduce operational costs[web:9][web:26][web:32].

Hierarchical Control Architecture: Scaling to city-wide networks (100+ intersections) through hierarchical architectures where district-level coordinators are overseeing local MARL agents, potentially reducing communication overhead and improving coordination at larger scales[web:5][web:10].

Deep Fuzzy Learning: Replacing manually-designed fuzzy rules with learned fuzzy systems that are automatically tuning membership functions and rule bases from data. This could improve adaptability while maintaining interpretability compared to black-box neural networks[web:29][web:33].

Multi-Objective Optimization Framework: Extending the approach to explicitly handle multiple objectives (travel time, emissions, equity, safety) using Pareto optimization techniques that are allowing operators to select preferred trade-offs based on policy priorities[web:20].

Real-World Validation: Field deployment pilot studies in collaboration with municipal traffic management centers represent the most critical next step to validate our approach in real-world conditions and identify practical deployment challenges not captured in simulation. We are currently seeking partnerships for testbed deployment[web:8][web:21].

Acknowledgments: The authors are thanking the developers of CityFlowER for providing the simulation platform and traffic data from Hangzhou and Jinan. We also thank the anonymous reviewers for their valuable feedback that improved this paper. This research was supported by [funding source to be added].

 doi:10.20944/preprints202510.1394.v1

References

1. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press, 2018.
2. H. Wei, G. Zheng, V. Gayah, and Z. Li, "CoLight: Learning network-level cooperation for traffic signal control," in *Proc. 28th ACM Int. Conf. Information and Knowledge Management (CIKM)*, Beijing, China, 2019, pp. 1913–1922.
3. T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Trans. Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020.
4. H. Zhang, C. Feng, S. Khanna, and Z. Li, "CityFlow: A multi-agent reinforcement learning environment for large scale city traffic scenario," in *Proc. World Wide Web Conf. (WWW)*, San Francisco, CA, 2019, pp. 3620–3624.
5. K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in *Handbook of Reinforcement Learning and Control*. Springer, 2021, pp. 321–384.
6. C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Trans. Robotics*, vol. 38, no. 2, pp. 1270–1286, Apr. 2022.
7. M. Tarif, M. Homaei, and A. Mosavi, "An enhanced fuzzy routing protocol for energy optimization in the underwater wireless sensor networks," *Computers, Materials & Continua*, vol. 83, no. 2, pp. 1–20, 2025.
8. Z. Zhang, J. Chen, and Y. Gao, "A fuzzy-logic based energy-efficient clustering algorithm for wireless sensor networks," *IEEE Access*, vol. 6, pp. 44261–44269, 2018.
9. X. Wang, Y. Liu, and Z. Chen, "Fuzzy control-based energy-aware routing protocol for Internet of Things networks," *Security and Communication Networks*, vol. 2021, Article ID 8830153, 2021.
10. P. Chen, S. Zhao, and L. Wang, "Energy-efficient data routing using neuro-fuzzy based cooperative techniques in wireless sensor networks," *Scientific Reports*, vol. 14, Article 79590, Dec. 2024.
11. M. A. Tawfeek, H. A. Ali, and S. M. Abd El-Kader, "A fuzzy multi-objective framework for energy optimization in IoT routing," *Journal of Network and Computer Applications*, vol. 245, Article 103595, 2025.
12. K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, "Towards multi-agent reinforcement learning based traffic signal control through spatio-temporal hypergraphs," arXiv:2404.11014, Apr. 2024.
13. L. Da, D. Shen, Y. Guo, J. Li, and Z. Li, "CityFlowER: An efficient and realistic traffic simulator with energy-aware routing," arXiv:2402.06127, Feb. 2024.
14. J. Guo, L. Cheng, and S. Wang, "A survey on deep reinforcement learning for traffic signal control," *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 1, pp. 8–24, Jan. 2023.
15. S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC)," *IEEE Trans. Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1140–1150, Sep. 2013.
16. O. Olusanya, E. C. Ifeoma, and A. B. Adewale, "Multi-agent reinforcement learning framework for autonomous traffic signal control in smart cities," *Frontiers in Mechanical Engineering*, vol. 11, Article 1650918, 2025.
17. W. Genders and S. Razavi, "Using a deep reinforcement learning agent for traffic signal control," arXiv:1611.01142, Nov. 2016.
18. X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Trans. Vehicular Technology*, vol. 68, no. 2, pp. 1243–1253, Feb. 2019.
19. M. Wiering, J. van Veenen, J. Vreeken, and A. Koopman, "Intelligent traffic light control," Technical Report UU-CS-2004-029, Utrecht University, 2004.
20. M. Tarif and B. Nouri Moghadam, "A review of energy efficient routing protocols in underwater internet of things," arXiv preprint arXiv:2312.11725, 2023.
21. T. Nishi, K. Otaki, K. Hayakawa, and T. Yoshimura, "Traffic signal control based on reinforcement learning with graph convolutional neural nets," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems (ITSC)*, Auckland, New Zealand, 2019, pp. 877–883.
22. P. Mannion, J. Duggan, and E. Howley, "An experimental review of reinforcement learning algorithms for adaptive traffic signal control," in *Autonomic Road Transport Support Systems*. Springer, 2016, pp. 47–66.
23. Y. Van der Pol and F. A. Oliehoek, "Coordinated deep reinforcement learners for traffic light control," in *Proc. NIPS Workshop on Learning, Inference and Control of Multi-Agent Systems*, 2016.

 doi:10.20944/preprints202510.1394.v1

24. L. N. Alegre, A. L. Bazzan, and B. C. da Silva, "Quantifying the impact of non-stationarity in reinforcement learning-based traffic signal control," in *Proc. Int. Conf. Autonomous Agents and Multi-Agent Systems*, 2020.
25. G. Chaslot, S. Bakkes, I. Szita, and P. Spronck, "Monte-Carlo tree search: A new framework for game AI," in *Proc. AAAI Conf. Artificial Intelligence and Interactive Digital Entertainment*, 2008.