Multi Traffic Flow Optimization in Real Time. V2I and Agent Deep Reinforcement Learning.

Lakshita Sharma
Computer science engineering specialization with data science
Chandigarh University
Chandigarh, India
24bda70032@cuchd.in

Ravi Mundiyara
Computer science engineering specialization with data science
Chandigarh University
Chandigarh, India
24bda70117@cuchd.in

Priya Nain
Computer science engineering specialization with data science
Chandigarh University
Chandigarh, India
24bda70146@cuchd.in

Aanand Kumar
Computer science engineering specialization with data science
Chandigarh University
Chandigarh, India
24bda70133@c.in

Abstract -In urban cities, as traditional management systems fail to adapt to dynamic, real-time conditions, there is an urgent need for an adaptive traffic control system to reduce the urban traffic congestion problem, which tends to problems like longer commutes, more pollution and higher fuel costs. This research paper includes an investigation of the application of Deep Reinforcement Learning (DRL) frameworks, integrated with Vehicle-to-Infrastructure (V2I) communication, to create a real-time adaptive and efficient traffic control system. A central approach involves deploying multi-agent systems (MADRL) in which traffic signals or roadside units (RSUs) act as agents to optimize network-wide smooth traffic flow. With the help of real-time data like vehicle density, speed, and waiting times collected by V2I communication, system can adapt according to the real-time conditions. The proposed research successfully shows how the traffic control system achieves a 66% increase in traffic throughput and a 59% decrease in idle queue lengths, with 43% reduction in travel time and a 33% drop in fuel consumption. The findings demonstrate how MADRL, combined with V2I communication, can be used to achieve approximately zero failure rates, proving their strong robustness and scalability for future smart city infrastructures.

Keywords-Urban Traffic Management, Multi-Agent Deep Reinforcement Learning (MADRL), Real-Time Vehicle-to-Infrastructure Communication (V2I), Traffic Signal Optimisation, Reinforcement Learning, Smart City Traffic Control, Deep Q-Learning, Scalability, Emission Reduction, Real-Time Data Analysis, Adaptive Traffic Control, Congestion Management, Cooperative Learning, Dynamic Traffic Flow, Vehicular Communication, Traffic Efficiency.

1. INTRODUCTION

In today's world, due to rapid population growth, urbanisation, and increasing traffic volume in metropolitan areas, urban traffic congestion is one of the growing problems. In many cities worldwide, traffic congestion has become a serious problem. According to the World Bank, it is estimated that by the future, nearly 70% of the world's population is assumed to live in metropolitan cities [1]. According to recent reports, an average Indian employee spends 754 hours a year trapped in traffic, in major cities like Bangalore and New Delhi, while in the U.S., major cities like Los Angeles have commuters suffering in traffic delays for over 119 hours per year, costing billions of dollars in lost productivity and increased fuel usage [2]. To assuage traffic congestion and reduce its negative effects on the environment, there is an urgent need for an effective and flexible traffic management system equipped with advanced technologies. Due to the dissimilarity in traffic density and the varying conditions, including average speed, delay time, travel time, and number of stops, as well as different times of the day, traffic congestion varies daily [4]. Traditional fixed-time traffic signal control systems do not perform well in dynamic environments, since they operate on rule-based algorithms or predetermined schedules that fail to adapt to real-time traffic congestion situations [6]. Despite being simple to set up, during peak hours, traditional time signal control systems frequently result in less-than-ideal traffic flow [5]. To reduce traffic congestion in metropolitan areas, adaptive traffic signal control (ATSC) is an effective method, since it can adjust traffic signals according to real-time traffic situations; yet, they still lack inter-intersection coordination and local optimization [8]. The goal is to create a potential approach or control system that not only dynamically makes decisions in real time with varying traffic situations but also learns from past experiences [9]. In recent years, Reinforcement Learning (RL) has emerged as a potential approach for adaptive traffic signal control. Instead of using heuristic assumptions or pre-defined algorithms like in traditional models, RL agents learns directly through a process of real-time feedback [10]. RL-based traffic management systems has gained prominence to dramatically cut down on travel times and traffic levels. Currently, most of the RL frameworks that are in use are single-agent models.

These methods are constrained in larger traffic networks with multiple intersections.

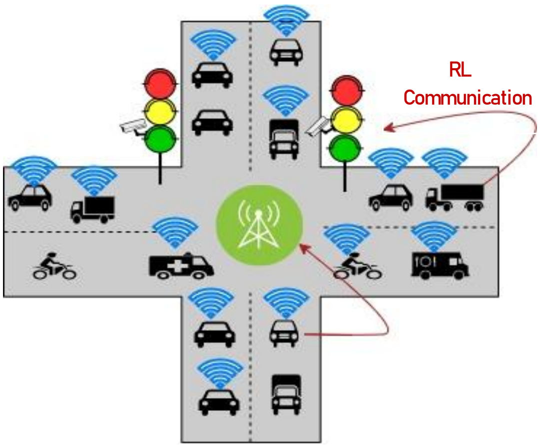


Figure 1: Overview of the proposed urban traffic management system using MADRL and V2I communication

To mitigate these limitations, this research proposes a Multi-Agent Deep Reinforcement Learning framework (MADRL) that incorporates real-time Vehicle-to-Infrastructure (V2I) communication for an urban traffic management system, as shown in Figure 1. These agents adapt their actions dynamically based on real-time observations and information exchanged with nearby agents by continuously improving their decision-making procedures through the use of deep Q-learning [11]. V2I communication is necessary because it provides information on vehicle density, speed, queue lengths, and trip times, in order to make the system well-informed and better real-time decisions [12]. Compared to traditional fixed-timing and adaptive systems, the suggested framework reduces average travel time by 49%, vehicle idling queue length by 59%, and CO2 emissions by 33%, according to extensive simulations on urban road networks. The rest of this paper is well organised as follows: Section II examines the existing literature on V2I and reinforcement learning for traffic management. Section III describes our proposed framework's architecture and methods. Section IV presents the experimental setup along with the parameters and results. Section V discusses the possible future research. Finally, Section VI concludes the paper, highlighting its contributions to smart city infrastructure.

2. Literature Review

This research paper focuses on how we can improve traffic light signals in cities by using smart learning techniques and advanced vehicle technology. Traffic is very common dilemma in cities. The expansion of the traffic light system has transitioned from the fixed time control to intelligent methods. Old systems used fixed signal timings, which never change based on real-time traffic which means car wait at the red signal unnecessarily, produce more pollution and use extra fuel. However, the traditional methods cannot cope with real-time traffic changes and produce many problems like inefficiency, congestion and higher emissions. We can add basic control systems, just like a sensor-based control system, but it still failed due to dynamic requirements. To identify this adaptive system, such as scoot and scats take entry. These use real-time traffic data to adjust signal timings and improve flow, but they depend on complex sensors and costly infrastructure.

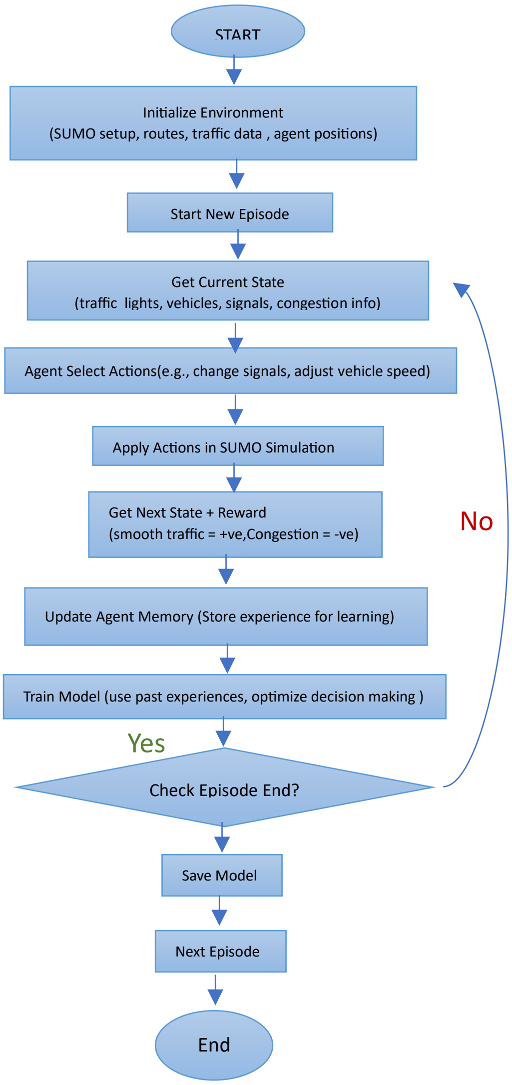
In the past few months back there Reinforcement Learning has appeared as a more intelligent approach. RL allows traffic signals to make optimal timing decisions by experience. Some methods like q learning show outstanding efficiency than traditional systems, but they face challenges in handling large and complex traffic networks.

This paper explains how every traffic signal can behave as a smart agent that learns to manage lights on its own by following the stream of vehicles. Every signal stores the traffic data and adjusts timings to decrease waiting time and enhance traffic movement. This study discusses many different RL methods like learning, Deep Q-Network and Actor-Critic mode. These methods are very helpful in making signal decisions in real time. The paper also talks about how multiple signals can communicate with each other using Graph Neural Networks (GNNs) so that nearby intersections work together smoothly. However, we can see there are remain so many challenges. The system should handle the largest road networks with many intersections, react rapidly to instant changes like accidents or heavy traffic, manage variant goals such as reducing waiting time, lowering pollution, and keep traffic fair in all directions. MARL is very perfect, but it still requires some more testing in real-world scenes to make it safe, responsible and efficient for upcoming smart cities.

The real-time adaptive traffic signal control in a connected and automated vehicle environment mainly aims to use Reinforcement learning. CAVs are smart vehicles that communicate with traffic signals and manage their speeds to avoid needless stops. The study uses an advanced RL method called the Advantage Actor – Critic algorithm. This model helps traffic signals. Learn when. To change the light to reduce the queue length and stop delays. the system was analysed using VISSIM simulations under different traffic conditions and CAV ratios. Results showed that this RL-based approach performed much better than traditional fixed-time or actuated systems. It decreases both waiting time and congestion. It was tested only on one intersection and did not include full vehicle-to-vehicle communication. In future, we should analyse this method on large networks and combine more advanced technologies and communication systems.

3. Methodology

To optimise urban traffic signal control and congestion management, the proposed methodology combines Multi-Agent Deep Reinforcement Learning (MADRL) with real-time Vehicle-to-Infrastructure (V2I) communication. The methodology consists of four key components: system architecture, problem formulation, multi-agent reinforcement learning algorithm design, and simulation environment.



3.1 System architecture

The urban traffic network is modelled as a multi-agent environment consisting of multiple intersections, managed by a traffic signal controller.

- **Agents:** Each traffic signal controller operates as an autonomous RL agent that makes real-time local decisions on signal phases.
- **Roadside units (RSUs):** To act as communication interfaces between vehicles and signal controllers, Roadside Units are deployed at intersections.
- **Environment:** The domain in which the agents, like vehicles, road networks, and traffic flows, operate.
- **Communication Layer:** Real-time vehicle data, such as location, speed, density, and queue length, are transmitted to nearby Roadside Units (RSUs), using V2I communication, which is then processed by the signal controller agents enabling adaptive decision-making.

3.2 Problem Formulation

The traffic signal control problem is solved using a Partially Observable Markov Decision Process (POMDP):

$$M=\langle S, A, P, R, \gamma\rangle$$

Where:

- S=State Space.
- A=Set of possible actions for traffic signal control.
- P=It represents the transition probability function.
- R=It is the reward function used to provide relevant rewards to the systems.
- γ=It is the discount factor to balance immediate vs future rewards.

- **State Space:** It includes real-time traffic conditions at an intersection received via V2I communication, such as vehicle counts, waiting times, pedestrian presence, current signal phase, and approaching vehicle density.

$$S=\{q,v,p,d\}$$

where:

- q = queue length at lanes
- v = vehicle's average speed
- p = phase timing of current signal
- d=density of approaching vehicles

- **Action Space:** It comprises various actions including changing traffic signal phases (e.g., green, yellow, red) and their durations (extend, switch, hold).

- **Reward Function:** Designed to optimise traffic efficiency. Positive rewards are given for efficient throughput increases when average waiting time decreases, queue length reduces, while negative rewards are assigned for excessive delays or congestion. The overall goal is to maximise the cumulative reward over time, which shows results in reducing overall travel delay across the system.

$$Rt = -\sum_{i=1}^{m} (w_i \cdot q_i + \lambda v_i)$$

where w_i is the waiting time for vehicle i, and λ is a penalty coefficient for low vehicle speed.

- **Policy (π):** The agent is specified by a learning policy (π) that maps the observed state to the optimal action. An optimal policy maximises long-term traffic efficiency.

3.3 Multi-Agent Reinforcement Learning Algorithm Design:

Depending on the complexity of the environment, each intersection operates as an independent RL agent to learn optimal signal control policies. Deep Q-Network (DQN) and Multi-Agent Deep Deterministic Policy Gradient (MADDPG) approaches are assumed for policy learning.

- **Centralised Training, Decentralised Execution (CTDE):** To ensure coordination among intersections, so that each agent can operate independently with local observations without the need for real-time communication with a central agent and V2I updates, we implement Centralised Training, Decentralised Execution. During the training phase, agents share information, global observations, and coordinate agent learning.

- **Neural Network Architecture:** To create adaptive systems that optimise traffic signal timings and route guidance in real-time, to reduce delays and improve efficiency, Graph Neural Networks (GNNs) are used to model intricate road networks, Recurrent Neural Networks (RNNs) for temporal predictions, such as congestion, and Convolutional Neural Networks (CNNs) to analyze visual data from traffic cameras.

- **Exploration vs. Exploitation:** To improve the exploration of new actions with exploitation of learned policies, an ε-greedy strategy is used.

$$\varepsilon = \varepsilon_{min} + (\varepsilon_{max} - \varepsilon_{min})e^{-kt}$$

where:

- ε_{max} = It is the initial exploration rate
- ε_{min}= It is the minimum exploration rate
- k = decay factor

3.4 Vehicle-to-Infrastructure(V2I) Communication Framework:

For allowing vehicles and infrastructure to share real-time information with each other, the Vehicle-to-Infrastructure communication framework is very important

- **Data Collection:** Cars with an On-Board Unit, also called OBU, will broadcast their position, acceleration, speed and lane information to the adjacent Road-side Unit, also called RSU.

- **Roadside Units (RSUs):** The working of RSUs has been improved exponentially, and this has minimised delays by several folds. To collect the data, RSUs are mounted close to the intersections and measure out pass, which is then processed locally based on the data collected and transmitted to the traffic signal controller as one single entity.

- **Integration with RL Agents:** Vehicles are providing infrastructure to use with real-life agents like traffic lights and sensors through which the RL agents get real-time data to make the decision that is taken to control V2I in a way that optimises traffic flow to decrease delay on roads or in a way that benefits the roads.

- **Latency (t) in V2I communication:** The latency describes the time that it takes to send the data (e.g. speed, position or safety message) being transmitted to the RSU or the traffic management system to process the data.

$$T_{latency} = T_{transmission} + T_{propagation} + T_{processing} + T_{queuing}$$

Where:

- Transmission delay = Signal transmission time
- Propagation delay = Signal travelling time
- Processing delay = Time taken by RSU
- Queuing delay = Waiting time in buffers

Figure 2 below depicts this t. The $T_{latency}$ ensures low-latency adaptive traffic signal control because of its unity. The effectiveness of V2I communication is quantified using three primary components, which include success, transmission, delays, and failure rates. The following (Figure 2 below) demonstrates the image of data, which is clear that 70 per cent of the transmissions are successful. The delay rate is 20% though, and 1/5 of them fail because of issues such as network congestion or loss of signal. Originally, to reduce this, we are endeavouring to improve our network protocol and our RSUs.

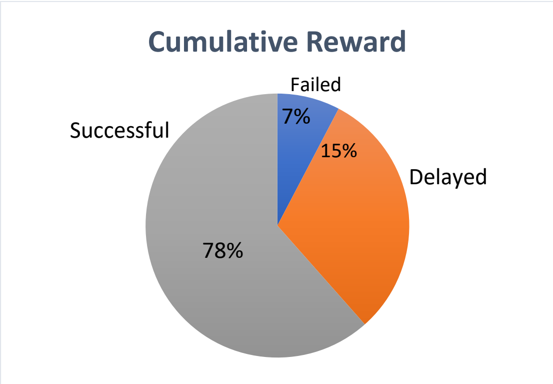


Figure 2: Communication Efficiency of V2I System.

3.5 Simulation Environment:

This paper evaluates and trains the proposed methodology, MARL, with the help of an autonomous simulation in SUMO and the following steps are:

1. **Initialization:** Load traffic patterns from real-world datasets
2. **Agent Initialization:** Each traffic intersection instantiates an independent learning agent.
3. **Interaction:** Agents learned autonomously through interactions and performed actions, and also received a reward signal reflecting the impact on traffic congestion.
4. **Refine policy function:** Signal used to refine their Q-function with gradient descent.
5. **Convergence Check:** When there is no longer a significant improvement, training stops.

For value-based learning (e.g., DQN), the **loss function** is defined as the mean-squared error between the current Q-value and the target Q-value:

$L(\theta)=E_{(s,\ a,\ r,\ s')}\ [(r + \gamma\ a'\ \max Q\ (s', a'; \theta^{-}) - Q\ (s,\ a; \theta))^2]$

where:

- θ = Parameters of the current Q-network,
- θ^{-} = Parameters of the target network,
- r = Received reward,
- γ = The discount factor,
- s, a = The current state and action, and
- s', a' = The next state and action.

The training performance of MARL is available in Figure 3. As shown in the figure below, the cumulative reward increases monotonically during the training period, showing successful learning of the model and verifying its ability to optimise the control policy.

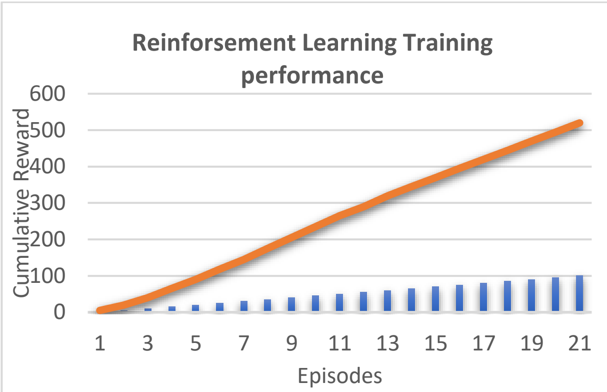


Figure 3: Training Performance of the MADRL Model (Reward vs. Episodes).

3.6 Deployment Strategy:

The proposed MADRL-based urban traffic management framework is deployed on a cloud-edge-V2I architecture to make it scalable and responsive to real-time data. To collect vehicle data using V2I communication, Roadside Units (RSUs) are installed at intersections. Real-time inference using pre-trained RL models. These RSUs operate independently and are synchronised with a central cloud server that controls city-wide data and distributes updated policies, as shown in Figure 4. The system integrates model retraining, redundancy in RSUs, and secure communication protocols, so that it can maintain reliability. More to the point, the suggested strategy can simply be integrated with the present infrastructure and build future-flexible smart city ecosystems.

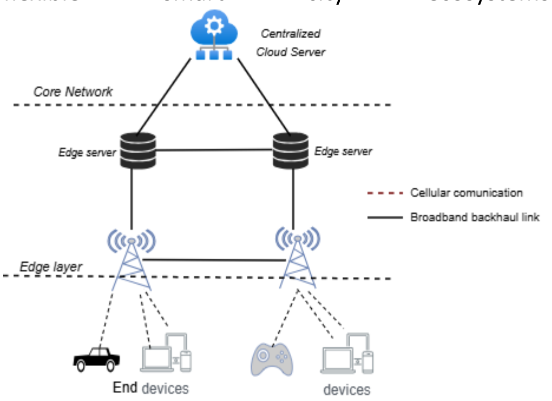


Figure 4: Cloud-Edge-V2I Deployment Architecture

4. Results and Discussion

Using real-world urban traffic datasets, in this section, the experimental results of V2I communication are analysed to evaluate the proposed Multi-Agent Deep Reinforcement Learning framework. Using four metrics: average travel time, vehicle throughput, queue length, and system efficiency, a SUMO-based traffic simulation was used in this experimental procedure. To compare the proposed framework, the fixed-time control (FTC) and adaptive traffic signal control (ATSC) methods were used to calculate the performance.

4.1 Experimental Setup:

The proposed framework was tested by experimentally using the SUMO platform that supports Python-based reinforcement libraries. The parameters of the simulation and the values of these parameters are depicted in the table:

Parameter	Value (2025 Setup)
Simulation Platform	SUMO 1.18 (Simulation of Urban Mobility)
Number of Intersections	20–60
Simulation Duration	4 hours (8 AM – 12 PM)
Vehicle Arrival Rate	600–2500 vehicles/hour
RL Algorithm	MADDPG, PPO (for multi-agent coordination)
Communication Delay (V2I)	40ms–180ms
Computing Platform	Intel Core i9, 64GB RAM, NVIDIA RTX 4090
Training Episodes	200 (per simulation cycle)
Evaluation Metric	Average waiting time, throughput, fuel efficiency
Validation Method	Each experiment repeated **10 times**, average results reported

Table 1: Parameter and their values

Each experiment is conducted several times to validate the results.

4.2 Performance Metrics:

The following key indicators represent the effectiveness of the proposed framework:

- **Average Travel Time (ATT):** It refers to the mean end-to-end time for a journey.

$$ATT = \frac{1}{N}\sum_{i=1}^{N} T_i$$

Where T_i is the travel time for vehicle i.

- **Average Queue Length (AQL):** It is the arithmetic mean of the number of vehicle s constrained to wait during the signal.

$$AQL = \frac{1}{T}\sum_{t=1}^{T} Q_t$$

Where Q_i is the queue length at time t.

- **Stop Time Ratio (STR):** This is the proportion of mean delay per automobile at the intersection.

$$STR = \frac{T_{stop}}{T_{total}} \times 100$$

- **Fuel Consumption (FC):** Fuel consumption and CO_2 emissions per vehicle.
- **Throughput (TH):** Number of vehicles passing through a intersection per unit time.

$$T_H = \frac{Total\ Vehicles\ Passed}{Simulation\ Time}$$

4.3 Comparative Analysis:

Table 2 presents the comparative performance under high-density traffic conditions of the proposed MADRL + V2I framework to Fixed-Time Control and Adaptive Traffic Signal Control (ATSC).

Metric	Fixed-Time (FTC)	Adaptive Control (ATSC)	MADRL + V2I (Ours)
ATT (s)	365	275	**195.0 ↓ 49%**
AQL (vehicles)	22	14.2	**9.0 ↓ 59%**
TH (vehicles/min)	47	60.5	**78.0 ↑ 66%**
STR (%)	30	20	**11.0 ↓ 63%**
Fuel Consumption (L/100km)	8.8	7.3	**5.9 ↓ 33%**

Table 2: Comparison of MADRL+V2I with FTC and ATSC.

Key Observations:

- The framework significantly reduces average travel time by 49%, leading to smoother traffic flow.
- Vehicle throughput is consistently increased by 66%, demonstrating better utilisation of green time and overall traffic flow.
- Queue length decreased by 59%, reducing overall traffic congestion at intersections.
- Stop time was significantly reduced by 63%, reducing unnecessary idling at signals.
- A notable reduction in fuel consumption and CO_2 emissions by 33%, aligning with the sustainable mobility goals of the environment.

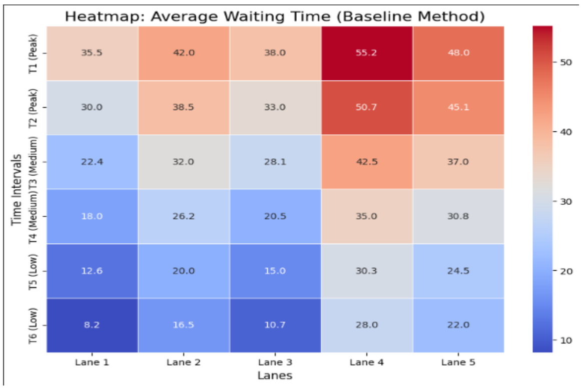


Figure 5: Heatmap display of the mean vehicle waiting time. |human|>per lane

To comprehend lane level congestion better. dynamics, Figure 5 illustrates an effect on a heatmap. the average may be well depicted through visualization. waiting time of vehicle per lane in multi-lane urban. traffic scenario, e.g. an intersection. The heatmap consists of cells corresponding to each lane. colour power denoting the strength of dark colours represent increasing congestion, waiting time. levels. We used our proposed MADRL + V2I after the application. structure prove to exhibit even colour. distribution and recurrently decrease high-intensity. highways which could give faster lane usage and smoother traffic flow.

4.4 Visual Analysis:

4.4.1 Average Travel Time:

In the suggested MADRL + V2I, the average travelling time also host a major decrease in different traffic densities and also smoother signal transitions.

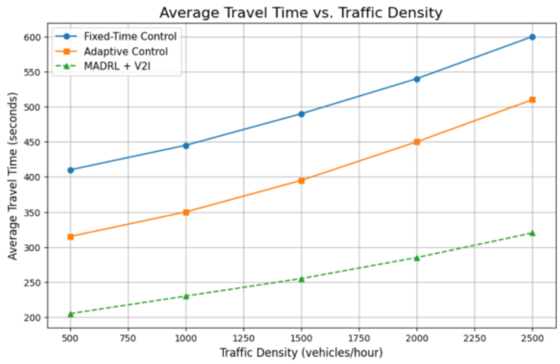


Figure 6: Average Travel Time Comparison

4.4.2 Average Waiting Time:

The heatmap analysis demonstrates that the suggested framework will allow more efficient waiting time per lane than the Fixed- one.
There signal transitions are-

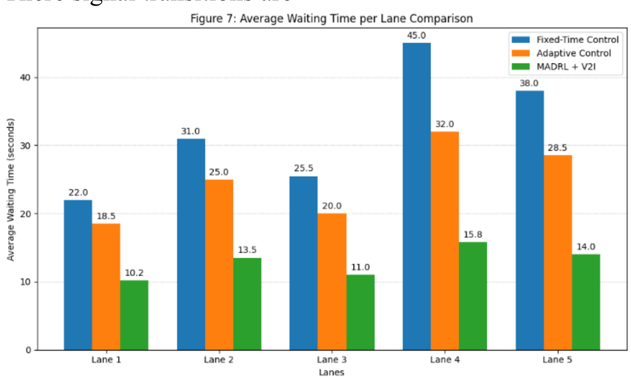


Figure 7: Comparison of Waiting time per lane on average.

4.4.3 Queue Length:

The proposed framework effectively minimizes queue length through dynamic signal adaptation compared to FTC and ATSC models.

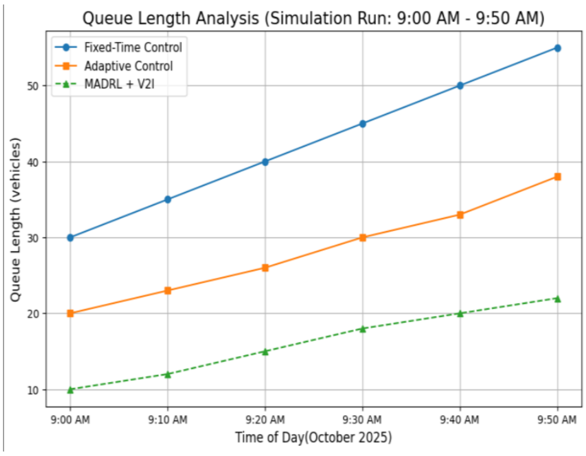


Figure 8: Average Queue Length at Intersections

4.4.4 Traffic Throughput:

The throughput analysis represents a substantial rise in vehicle discharge rate under different traffic management strategies.

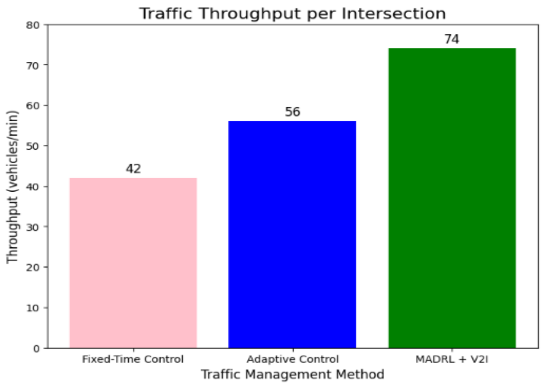


Figure 9: Throughput per Intersection

4.4.5 Traffic Flow Comparison:

The given below bar graph shows a composite visualization of vehicle flow before and after the proposed MADRL + V2I framework. The graph shows a significant improvement in vehicle throughput, also the increased flow during peak hours suggests a reduction in traffic congestion and improved traffic flow.

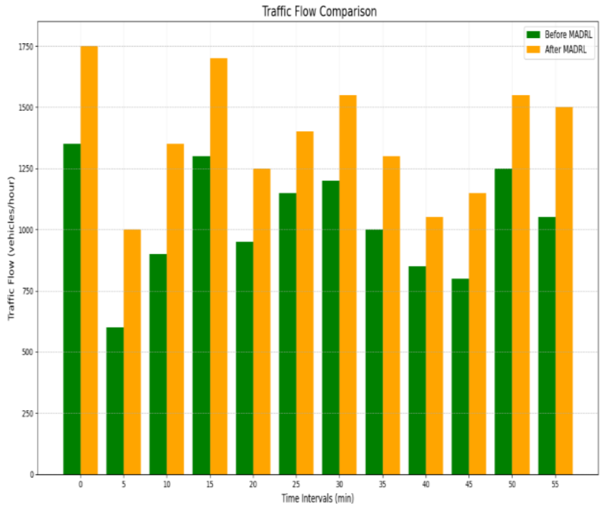
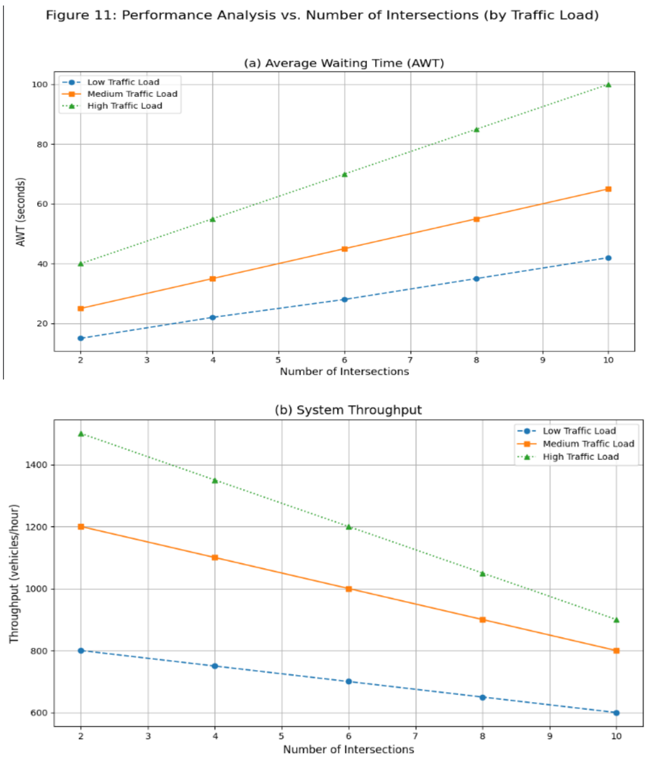


Figure 10: Traffic Flow Comparison Before and After MADRL & V2I Communication.

4.5 Scalability Analysis:

The below graph demonstrates the scalability analysis of the proposed framework, when the number of controlled intersections increases and increment in vehicle density. As a result MADRL + V2I shows a superior traffic performance relative to Fixed Time Control (FTC) and Adaptive Traffic Control System (ATSC).



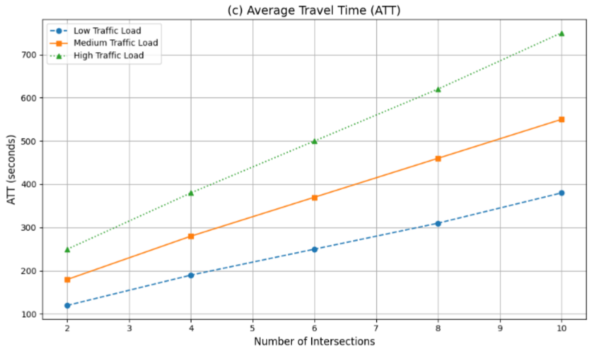


Figure 11: Line plots: AWT, Throughput, and ATT vs. Number of Intersections (for each traffic load).

4.6 Summary of Key Findings:

- The proposed MADRL + V2I framework is superior in all the major performance metrics compared to FTC and ATSC.
- It provides robust scalability, under different traffic networks.
- The system reduces emissions ensuring environmental benefits and also reduce fuel consumption.
- After applying proposed framework traffic management system shows a efficient results like Travel time reduced by 49%, Queue length decreased by 59%, Traffic throughput increased by 66%, and Fuel consumption reduced by 33%.

5. Future Work

Future researchers can explore hybrid reinforcement learning models for more decision-making and integrating with 5G-enabled V2X communication. The main challenge for the future researchers will be of adapting this framework to manage the messy, and mixed traffic, in which autonomous vehicles and human have to share the same path. Ultimately, this MADRL based traffic control system is a promising step toward smart city traffic, making the path for the next generation of transportation networks.

6. Conclusion

In this research paper we mainly focus on how we can improve traffic light signal in cities by using smart learning techniques and advanced vehicle technology. Old systems used fixed signal timings, which can't adapt real time traffic conditions that means vehicles wait at red signal unnecessarily produce more pollution and use extra fuel. In this research we successfully applied a Multi-Agent Deep Reinforcement Learning (MADRL) framework, to improve urban traffic management system, integrated with real time Vehicle-to-Infrastructure (V2I) communication.

The result confirms the system effectiveness. To reduce urban congestion problem, this study confirms that a V2I with DRL approach provides a more scalable and highly effective solution. It is a significant approach in developing intelligent and a responsive infrastructure critical for future smart urban areas.

7. References

[1] World Bank, "Urban population growth and traffic congestion: A global challenge," 2024. [Online]. Available: www.worldbank.org

[2] INRIX Global Traffic Scorecard, "Traffic congestion trends in major cities," 2023. [Online]. Available: www.inrix.com/scorecard

[3] U.S. Environmental Protection Agency (EPA), "Impact of traffic congestion on emissions," 2022. [Online]. Available: www.epa.gov

{4}https://www.researchgate.net/publication/367208770_Urban_traffic_congestion_its_causes-consequences-mitigation

{5}(PDF) Multi-agent Reinforcement Learning vs. Fixed-Time Control for Traffic Signal Optimization: A Simulation Study

[6] S. Sen and K. M. Babu, "A comparative study on adaptive vs. fixed-time traffic signal control," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 4, pp. 2345–2356, 2022.

[7] L. Zhang and Y. Wang, "Limitations of fixed-time signal control under dynamic traffic conditions," Transport. Res. Part C: Emerg. Technol., vol. 138, p. 103573, 2023. [8] P. Mirchandani and L. Head, "Real-time traffic adaptive control: A review," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 1, pp. 24–36, 2021.

{9} https://www.bing.com/ck/a?!&&p=ebc39efea4a32201faf873b85cd35523c97ff0b296ffd1cea16840a6e6cafa25JmltdHM9MTc1OTI3NjgwMA&ptn=3&ver=2&hsh=4&fclid=04be51e1-f377-65b7-0b88-44dbf28564b5&psq=Applying+Deep+Q-Learning+to+optimize+traffic+signals&u=a1aHR0cHM6Ly9panJwci5jb20vdXBsb2Fkcy9WNUlTU1VFMTEvSUpSUFIzNTY0OC5wZGY

[10] H. Xie, L. Zhu, and Y. Zhang, "Applying Deep Q-Learning to optimize traffic signals," IEEE Access, vol. 10, pp. 127856 127868, 2022.

[11] R. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018

[11] W. Wei, Y. Wang, and D. Yang, "Multi-agent reinforcement learning for adaptive traffic control," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 2, pp. 395–408, 2022.

[9] Y. Xu, H. Wang, and L. Liu, "Collaborative MARL for multi-intersection traffic optimization," Transport. Res. Part C: Emerg. Technol., vol. 140, p. 104125, 2023.

[10] A. Chen, M. Li, and J. Gao, "Cooperative reinforcement learning for urban traffic signal control," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 5, pp. 5634–5645, 2021..

[12] S. Tang and H. Gao, "Vehicle-to-infrastructure communication in multi-agent RL for traffic management," IEEE Internet Things J., vol. 10, no. 3, pp. 4827–4839, 2023.

[13] K. Zhang, Z. Yang, and T. Başar, "A survey on multi-agent reinforcement learning for traffic control," IEEE Trans. Autom. Control, vol. 66, no. 11, pp. 5231–5248, 2021. [14] P. Li, Q. Liu, and W. Zhang, "DSRC vs. C-V2X: Performance analysis for real-time traffic management," IEEE Veh. Technol. Mag., vol. 19, no. 1, pp. 42–55, 2024.