TrafficLightRL

Kristian Diana
McMaster University
dianak@mcmaster.ca

Tridib Banik
McMaster University
banikt@mcmaster.ca

Varun Pathak
McMaster University
pathav4@mcmaster.ca

Ryan Li
McMaster University
li3018@mcmaster.ca

Clara Wong
McMaster University
wongc148@mcmaster.ca

***Abstract*—Addressing urban traffic congestion is crucial for environmental sustainability, as inefficient traffic flow leads to increased fuel consumption and greenhouse gas emissions. This paper presents TrafficLightRL, a reinforcement learning (RL)-based traffic light control system designed to reduce vehicle idling times and $C0_2$ emissions. Using SUMO for simulation and the Proximal Policy Optimization (PPO) algorithm from Stable-Baselines3 for RL training, our system dynamically adapts to real-time traffic conditions. Results show that the RL agent reduces idling times by up to 26.4% and stopping probability by up to 9.0% compared to traditional systems, and are analyzed over various traffic densities. This study highlights the potential of RL-driven solutions to enhance urban mobility while mitigating environmental impact. The code and resources for this project are available at: https://github.com/McMasterAI2024-2025/TrafficLightRL.**

I. Introduction

As one of the McMaster AI Society's many projects for the 2024/25 season, this team investigated a reinforcement learning solution that could address a pressing real-world issue. Out of a shared concern for environmental sustainability, as well as inefficiencies in urban mobility, existing traffic light systems were identified as a major contributor to excessive vehicular emissions. Current traffic light control mechanisms rely on fixed-timer schedules or sensor-based adjustments, leading to unnecessary idling and frequent acceleration events—both of which are known to significantly increase carbon dioxide (CO_2) emissions [1]. In response to this challenge, TrafficLightRL was developed as a reinforcement learning (RL)-based system aimed at reducing emissions by optimizing signal timings dynamically. This document discusses the issue addressed by the project, the design and implementation of the RL-based system, the quantifiable environmental benefits observed, and the broader implications for sustainable traffic management.

A. Motivation

Road transport is responsible for nearly 25% of global CO_2 emissions from fuel combustion, with urban congestion playing a significant role in this footprint [2]. One of the primary factors contributing to unnecessary emissions is stop-and-go traffic, where vehicles frequently accelerate from a standstill. Studies have shown that rapid acceleration events can increase fuel consumption and emissions by up to 200% compared to steady-speed travel [1]. Furthermore, in urban areas, drivers spend an average of 54 hours per year idling in traffic, further exacerbating emission levels [3].

Despite advancements in adaptive traffic control, conventional systems still struggle to minimize acceleration and idling simultaneously. This project explores how reinforcement learning can address this gap by dynamically adjusting traffic signal timings based on real-time conditions. By prioritizing reductions in idling duration and acceleration frequency, an RL-based approach has the potential to significantly lower CO_2 emissions in urban environments.

B. Related Works

Traditional traffic light control systems, such as fixed-timer schedules and sensor-based systems, rely on predefined rules that lack adaptability to real-time traffic conditions. While adaptive systems like SCOOT [4] and SCATS [5] offer improvements by dynamically adjusting signal timings, they remain limited by rule-based optimizations that require manual calibration and do not generalize well to varying traffic patterns.

Reinforcement learning has emerged as a promising alternative, offering the ability to learn optimal signal timing policies directly from traffic data. Prior studies have demonstrated that RL-based systems can reduce vehicle stops, travel times, and overall emissions [6]. Unlike traditional adaptive systems, RL approaches continuously refine their control strategies based on observed traffic dynamics, making them well-suited for sustainable urban mobility.

C. Problem Definition

The objective of this project is to develop an RL-based traffic light control system that reduces CO_2 emissions by minimizing vehicle idling and acceleration events. The proposed system optimizes signal timings based on real-time traffic flow, adapting dynamically to different conditions without relying on predefined scheduling rules.

Key evaluation metrics include reductions in total CO_2 emissions, idling duration, and the number of full stops per vehicle. The results aim to highlight the potential of reinforcement learning in mitigating urban traffic's environmental impact and providing a more sustainable solution to this pressing issue.

II. Methodology

This section outlines the design and implementation of the RL-based traffic light control system. It begins by presenting the tools and technologies used, followed by a detailed description of the RL agent's structure and decision-making process.

Next, the simulation environment is discussed alongside real-world considerations to ensure practical applicability. Finally, the evaluation metrics and calculation methods are introduced, providing a foundation for performance assessment in the results section.

A. Tools and Technologies

The project leverages two primary technologies: **SUMO (Simulation of Urban MObility)** and **Stable-Baselines3**. SUMO provides a realistic traffic simulation environment, including real-time visualizations and dynamic traffic scenarios. Stable-Baselines3, a popular reinforcement learning library, is used to train the RL agent. The integration between SUMO and Stable-Baselines3 is facilitated by **TraCI**, an API that enables real-time communication between Python and SUMO. Additionally, **OpenAI Gymnasium** is used to create a consistent interface for the RL agent, abstracting SUMO's functionality into a format compatible with Stable-Baselines3. Figure 1 illustrates the process flow and interactions between these components, which will be elaborated on shortly. Finally, **Matplotlib** is leveraged for analysis of various evaluation metrics.

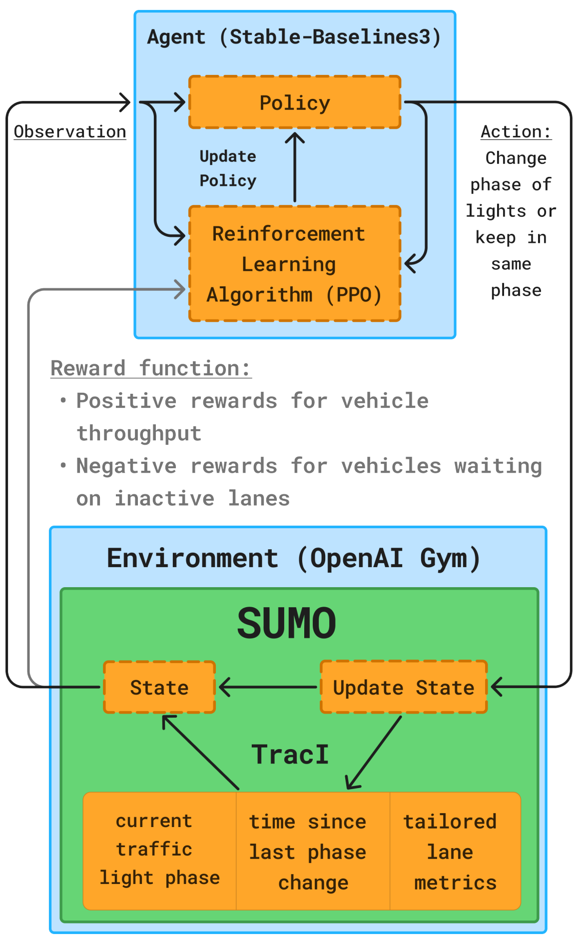

Fig. 1. Process flow diagram illustrating the interaction between SUMO, TraCI API, OpenAI Gymnasium, and Stable-Baselines3.

B. Reinforcement Learning Agent

The RL agent is trained using the **Proximal Policy Optimization (PPO)** algorithm, a state-of-the-art policy gradient method implemented in Stable-Baselines3. PPO was chosen for its stability and efficiency in handling continuous state and action spaces, making it well-suited for dynamic traffic control tasks. The agent interacts with the environment by observing the current state and selecting actions to optimize traffic flow. The key components of the agent are as follows:

- **Observation Space:** The agent observes the current traffic light phase, the time since the last phase change, and lane-specific metrics such as the number of queued vehicles and average idling time.
- **Action Space:** The agent has a discrete action space, where each action corresponds to a specific traffic light phase. Each phase determines which lanes are active (green) and which are inactive (red).
- **Reward Function:** The agent receives positive reinforcement for vehicles passing through the intersection and negative reinforcement for vehicles queued in inactive lanes. This encourages the agent to minimize waiting times and congestion.

C. Simulation Environment

The simulation environment is designed to mimic real-world traffic conditions, which is addressed through various features:

- **Random Vehicle Deployments:** Introducing randomness in the form of vehicle routes allows the agent to generalize effectively to unpredictable traffic patterns. This enhances real-world applicability, as the RL agent will never experience two identical **episodes** during training. To clarify, an **episode** represents the time period required for a specific number of vehicles to pass through the intersection.
- **Variable Traffic Densities:** Adjusting the spawn rate of vehicles enables simulations to model traffic fluctuations, accounting for real-world factors such as **time of day** and **weather conditions**.
- **Standard Traffic Safety Regulations:** The environment enforces standard traffic safety rules, such as requiring the agent to pass through **transition phases** (e.g., green $\rightarrow$ yellow $\rightarrow$ red) before switching traffic lights. Additionally, each phase must be held for a **minimum duration** to prevent chaotic scenarios and to ensure pedestrians have sufficient time to cross safely. These features ensure that the agent adheres to real-world constraints and common safety practices.
- **Real-World Networks:** To enhance realism, we use the **OSM Web Wizard** to export actual geographical location networks into SUMO. This facilitates traffic simulations in real-world environments, such as the road network around McMaster University. Figure 2 shows a side-by-side comparison of the Google Maps view of McMaster University and the corresponding SUMO simulation.

In this paper, the term "traditional system" refers to a fixed-time traffic signal control system implemented in SUMO.

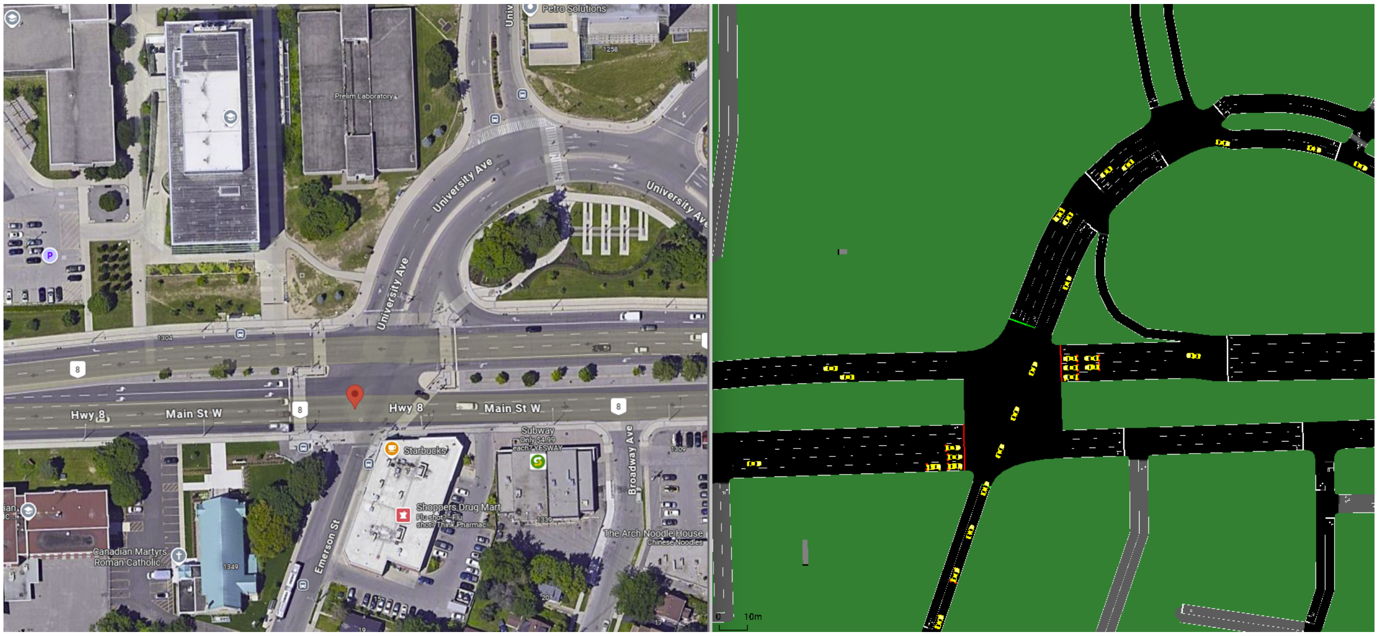

Fig. 2. Side-by-side comparison of the Google Maps view of McMaster University (left) and the corresponding SUMO simulation (right) generated using the OSM Web Wizard.

This system operates on a predetermined cycle, with traffic light phases changing at fixed intervals regardless of real-time traffic conditions. It serves as a baseline for evaluating the performance improvements achieved by the RL-based approach.

D. Evaluation Metrics

To evaluate the performance of the RL agent, three key metrics are evaluated and calculated as follows leveraging the TraCI API:

- **Mean Emissions Per Second:** The average emissions per second, measured in mg/s, produced by a single vehicle during the simulation. Emissions are calculated at each timestep of the simulation using SUMO's TraCI API, which provides real-time vehicle-specific emissions data. Specifically, the `traci.vehicle.getCO2Emissions` method returns the CO_2 emissions in milligrams (mg) for each vehicle, assuming a standard emission rate. The emissions are aggregated over the duration the vehicle is on the simulation, then divided by the total time it traveled to obtain the emissions per second value. The choice of evaluating emissions as a rate, rather than accumulated value, allows for broader insights into how emissions are affected beyond a single intersection.
- **Mean Idling Times:** The average time a single vehicle spent waiting at the intersection. This value is calculated for each vehicle of the simulation using the `traci.vehicle.getWaitingTime` method. These values are averaged to provide a cohesive representation rather than focusing on individual vehicles.
- **Stopping Probability:** The likelihood of a vehicle stopping at the intersection, calculated as the ratio of total vehicles required to stop to total vehicles on the simulation. This value is calculated utilizing the mean idling times for each vehicle from earlier, assuming any vehicle that has an idling time above a specific threshold has stopped. A threshold is applied to distinguish true stops from minor slowdowns or coasting. This value is represented as a probability, rather than a count, as variable factors such as traffic density and total number of cars will skew this data.

These metrics are calculated for each episode and compared against traditional traffic light systems to demonstrate the effectiveness of our RL-based approach. These metrics provide a comprehensive evaluation of both traffic efficiency and environmental impact, ensuring the results are applicable beyond a single intersection.

III. Results

This section presents the performance of the RL-based traffic light control system compared to a traditional system. The traditional system, in the context of SUMO, operates on a fixed cycle of pre-determined green and red light intervals, independent of real-time traffic conditions. The system is evaluated based on the three evaluation metrics described previously: **mean emissions per second**, **mean idling time**, and **stopping probability**. The x-axis of all figures represents the vehicle spawn rate, corresponding to the probability of a vehicle being deployed at each time-step. The spawn rate is representative of low, medium, and high traffic densities, illustrated by the blue, green, and red regions respectively.

Each data point is the average of 1000 episodes to ensure statistical reliability and smooth distributions. The spawn rate corresponds to the chance a vehicle is to be deployed at each time-step of the simulation, and each episode deploys 100 vehicles.

A. Effect of Traffic Density on Mean Emissions Per Second

Figure 3 illustrates the effect of traffic density on mean emissions per second, measured in (mg/s).

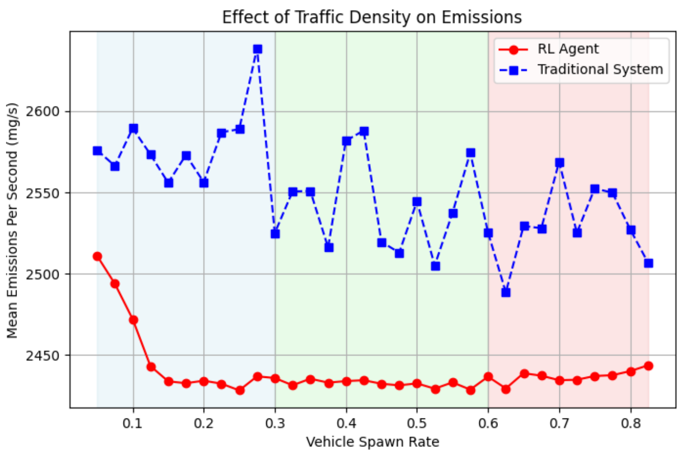

Fig. 3. Effect of traffic density on mean emissions per second. The red line represents the RL agent, while the blue dashed line represents the traditional system.

The RL agent significantly reduces emissions per second compared to the traditional system. In low traffic conditions, emissions per second are reduced by **5.0%**, **4.3%** in medium traffic conditions and **3.7%** in high traffic conditions.

B. Effect of Traffic Density on Mean Idling Times

Figure 4 presents the impact of traffic density on mean idling times, measured in seconds.

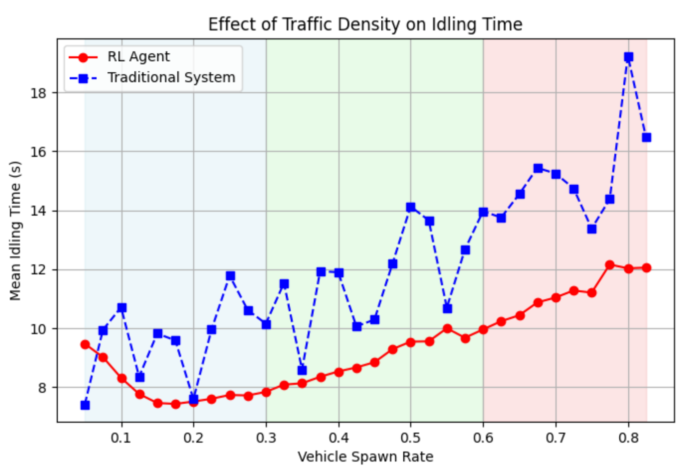

Fig. 4. Effect of traffic density on mean idling times. The red line represents the RL agent, while the blue dashed line represents the traditional system.

The RL agent consistently outperforms the traditional system by reducing idling times. In low traffic conditions idling times decrease by **16.4%**, while in medium traffic conditions they decrease by **22.7%**. Even in high traffic conditions the RL agent maintains its advantage with a **26.4%** reduction in idling times.

C. Effect of Traffic Density on Stopping Probability

Figure 5 shows the effect of traffic density on stopping probability, defined as the proportion of vehicles required to stop at the intersection.

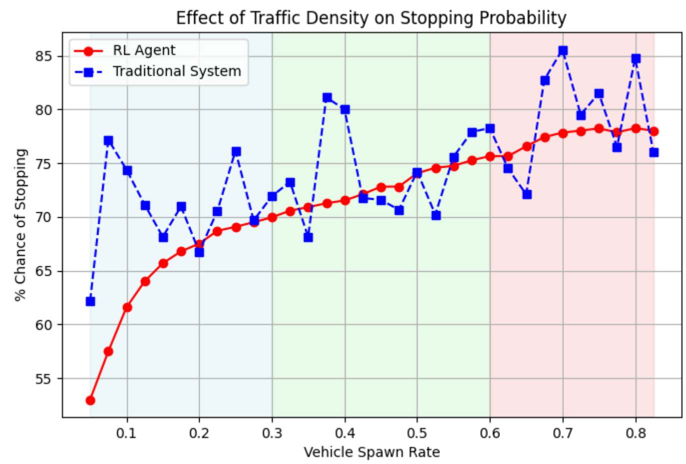

Fig. 5. Effect of traffic density on stopping probability. The red line represents the RL agent, while the blue dashed line represents the traditional system.

The RL agent also reduces stopping probability. In low traffic conditions, stopping probability decreases by **9.0%**, while in medium and high traffic conditions, it is reduced by **1.8%** and **2.3%**, respectively.

IV. CONCLUSION

In this project, an RL-based traffic light control system was developed using SUMO and Stable-Baselines3 to address the inefficiencies of traditional traffic management systems. Our findings highlight that the RL agent substantially reduces CO_2 emissions across varying traffic densities by minimizing idling time and decreasing the frequency of complete stops. These efficiency gains directly contribute to a more environmentally sustainable urban traffic management system.

A. Discussion of Emissions Reduction

The RL agent's ability to reduce emissions stems from two key factors: minimizing idle time, which mitigates prolonged low-efficiency fuel consumption, and reducing stops, which limits high-emission acceleration events after red lights.

- **Minimization of Idling Time:** Idling, while producing lower emissions per second compared to acceleration, contributes substantially to total emissions over time due to its prolonged nature. Although idling times saw the greatest reduction in high-traffic scenarios, emissions

improvements were less pronounced, highlighting the diminishing returns of idling reductions compared to stop reductions
- **Reduction in Stops:** Acceleration from a complete stop is a significant contributor to vehicle emissions. By optimizing traffic flow and reducing the frequency of stops, the RL agent minimizes the instances of high-emission acceleration events. The RL agent significantly reduces stops, particularly in low-traffic conditions, where a **9.0%** decrease in stopping probability resulted in the greatest emissions reduction. This suggests that minimizing unnecessary stops is a key factor in emissions control.

These factors highlight the RL agent's ability to address both the high-emission events (acceleration from stops) and the sustained emissions (idling) that characterize inefficient traffic systems. By dynamically adapting to real-time traffic conditions, the RL agent provides a more environmentally sustainable solution compared to traditional fixed-timer based systems. The results suggest that further emphasis on decreasing high-emission events, such as acceleration from stops, is most significant in relation to reduction of $C0_2$ emissions.

B. Limitations

Although the RL agent was trained across various traffic densities, the training distribution was linearly incremented, causing the agent to encounter medium traffic densities most frequently. As a result, the agent is more adept at optimizing performance in these conditions but may be less effective in handling extreme congestion or sparse traffic. A more balanced training approach—incorporating additional timesteps for low- and high-traffic scenarios—could improve generalizability across a wider range of conditions.

While SUMO's emissions model provides a realistic approximation of vehicle emissions, it relies on generalized vehicle dynamics. It does not fully account for real-world factors such as variations in vehicle types, fuel efficiency, or environmental influences like weather conditions. However, because the model prioritizes acceleration and idling behavior—the primary contributors to emissions in urban traffic—it remains a useful tool for comparative analysis of traffic control strategies.

Additionally, our study is confined to a controlled simulation environment, which, while useful for experimentation, does not capture network-wide congestion effects or interactions with external infrastructure. This limitation may lead to discrepancies between simulated and real-world performance, as the learned policy may not generalize effectively to larger urban networks. Future iterations should incorporate broader network effects to assess scalability and adaptability in dynamic traffic ecosystems.

C. Future Considerations

Several enhancements could improve the RL agent's real-world applicability. Future implementations could incorporate additional considerations such as pedestrian right-of-way, emergency vehicle prioritization, and adaptive responses to weather conditions. These factors play a crucial role in urban traffic systems and would enhance the agent's ability to operate effectively in diverse environments.

Another major challenge lies in scaling this approach to coordinate multiple intersections. As the number of controlled intersections increases, the complexity of synchronizing signals grows exponentially, necessitating more advanced agent communication strategies and higher computational resources. Multi-agent reinforcement learning (MARL) techniques, such as decentralized policies with shared learning objectives, could be explored to tackle this scalability issue.

Additionally, the rise of electric vehicles (EVs) presents a new avenue for optimization. Since EVs have different acceleration profiles and do not produce emissions while idling, an RL-based system tailored for mixed traffic compositions—including both traditional internal combustion engine (ICE) vehicles and EVs—could further enhance sustainability outcomes. Incorporating real-time EV-specific traffic data would refine emissions predictions and improve energy efficiency across urban road networks.

While this study demonstrates promising results in optimizing traffic flow at a single intersection, testing the approach across a variety of urban environments is essential for broader applicability. Our current implementation focuses primarily on Ontario university campuses; expanding to diverse real-world locations with varied infrastructure and traffic patterns would provide further validation and refinement of the system.

REFERENCES

[1] M. Barth and K. Boriboonsomsin, "Energy and emissions impacts of a freeway-based dynamic eco-driving system," *Transportation Research Part D: Transport and Environment*, vol. 14, no. 6, pp. 400–410, 2009.

[2] I. E. Agency, "Co2 emissions from fuel combustion 2022," https://www.iea.org/reports/co2-emissions-in-2022, 2022, [Online; accessed 2022].

[3] INRIX, "Global traffic scorecard," https://inrix.com/scorecard/, 2022, [Online; accessed 2022].

[4] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, "Scoot—a traffic responsive method of coordinating signals," Transport and Road Research Laboratory, Technical Report, 1981.

[5] P. Lowrie, "Scats: The sydney coordinated adaptive traffic system - principles, methodology, algorithms," *Traffic Engineering & Control*, vol. 33, no. 6, pp. 278–281, 1992.

[6] W. Wu, Y. Zhang, R. Jiang, H. Zhang, K. Gao, and D. Zhao, "A comprehensive survey on traffic signal control methods," *Transportation Research Part C: Emerging Technologies*, vol. 123, p. 102974, 2020.