Lane-Independent Highway Traffic Management for Random Anomalies Using Reinforcement Learning

1st Márk Mitrenga
Systems and Control Laboratory
HUN-REN Institute for Computer Science and Control (SZTAKI)
Budapest, Hungary
mitrenga.mark@sztaki.hun-ren.hu

2nd György Csippán
Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology
and Economics
Budapest, Hungary
gyorgy.csippan@edu.bme.hu

3rd Bálint Kővári
Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology
and Economics
Budapest, Hungary
kovari.balint@kjk.bme.hu

4th Tamás Bécsi
Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology
and Economics
Budapest, Hungary
becsi.tamas@kjk.bme.hu

5th Szilárd Aradi
Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology
and Economics
Budapest, Hungary
aradi.szilard@kjk.bme.hu

Abstract—**Reduced capacity on motorways can easily lead to significant congestion. This congestion is a major contributor to environmental pollution, harming the livability of the peri-urban environment and public health. In this study, we have addressed the congestion caused by lane closures on motorways, one of the many difficulties encountered in the lane closure problem. To overcome this problem, the so-called variable speed limit control, a traffic management system is a helpful tool that improves overall traffic flow characteristics - travel time, waiting time, and queue length - and reduces critical sustainability indicators such as fuel consumption and CO_2 and NO_x emissions. Deep Learning has repeatedly been shown to be an excellent solution to this problem. Hence, this study aims to use Reinforcement Learning to address the traffic management system and to find a general solution to congestion caused by the reduction of highway capacity to apply the model regardless of the number of lanes, improving and surpassing the results achieved in the literature in several aspects.**

Index Terms—**Reinforcement Learning, Variable Speed Limit Control, Intelligent Transportation Systems**

I. INTRODUCTION

In the case of highways, situations often arise — particularly during peak hours, when vehicle traffic is at its heaviest — that make the efficient regulation of traffic flow a significant challenge. The flow of vehicles is governed by physical laws similar to the principles of fluid dynamics. One such principle is the continuity equation, which states that the product of a fluid's flow velocity and the pipe's cross-sectional area is constant, representing the volumetric flow rate. This is expressed by the Equation (1),

$$\dot{Q} = A \cdot v \tag{1}$$

where $\dot{Q}$ is the volumetric flow rate, A is the cross-sectional area and v is the velocity.

Equation (2) - where ρ is the density - describes the mass flow rate ($\dot{m}$), which is worth examining in the context of traffic flow.

$$\dot{m} = \dot{Q} \cdot \rho \tag{2}$$

According to the work of [1], the cross-sectional area A represents the number of lanes in traffic flow. At the same time, v and ρ denote the average speed and density of vehicles, respectively. Regarding mass flow, it can be deduced that traffic flow is also a function of vehicle density. This relationship is illustrated in Fig. 1, where it is evident that the flow reaches its maximum around a critical density point and becomes unstable beyond that. It is, therefore, essential to maintain density around the crucial point. To achieve this, appropriately modifying speed, as the only adjustable component, is key to optimizing traffic flow.

A highly effective traffic management system for influencing speed is the Variable Speed Limit Control (VSLC), excellently suited for addressing congestion caused by reduced highway capacity. The VSLC dynamically adjusts the maximum permissible speed based on traffic conditions to optimize segments, considering static and dynamic objects and their environmental impacts on traffic flow [3]. For example, it facilitates lane changes from a closing lane. Fig. 2 illustrates the system's operation, showing the Variable Message Sign (VMS), a digital traffic sign controlled by the VSLC. Certain lanes on specific road sections are equipped with independently regulated speed limits to influence the active segment's outflow. The continuity equation governs the inflow
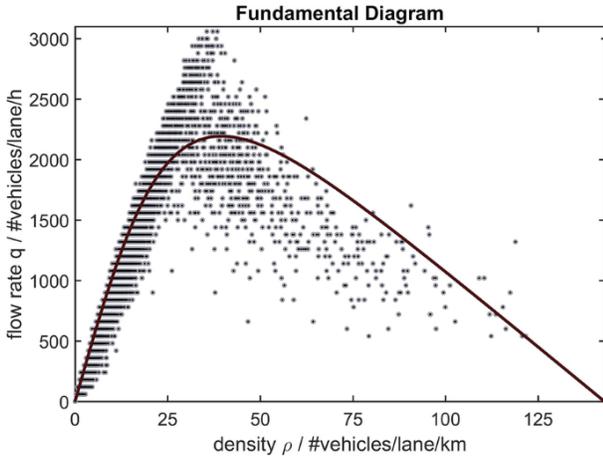
Fig. 1. Fundamental diagram as a function of vehicle flow and density. [2]

into the next cell [4]. The primary objective of this solution is not to expedite the resolution of existing traffic jams but to proactively prevent congestion from occurring.

II. RELATED WORK

During peak hours, traffic congestion often occurs on highways at sections affected by capacity reduction. Instead of increasing capacity, a more cost-effective and environmentally friendly solution is to apply Intelligent Transportation Systems (ITS) for traffic management.

Variable Speed Limit Control (VSLC) offers an excellent implementation solution. As mentioned, this system regulates highway traffic by dynamically adjusting speed limits based on current traffic conditions. Standard VSLC approaches rely on online feedback control and the fundamental flow-density relationship mapped to speed values [5].

The system selects the appropriate speed based on a specific situation, making Reinforcement Learning (RL) techniques a promising alternative for configuring VSLC. These methods can learn and respond to various traffic scenarios without requiring an explicit model of highway dynamics, as highlighted in [6], [7], and [8]. Thus, by applying RL, we avoid building
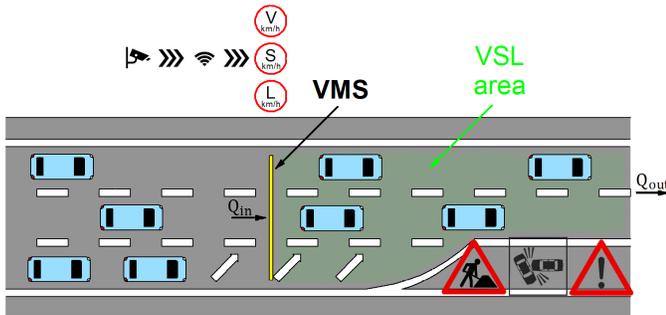

Fig. 2. VSLC operation to handle random anomalies.

complex mathematical models and instead leverage the agent's learning capability.

The input current situation corresponds to the environment-defined state, while the output-defined speed limit corresponds to the action, which are the core components of RL [9]. This highlights RL as a viable solution, with its scalability and generalization enabling exceptional performance in sequential decision-making, allowing agents to reassess states and make rapid decisions effectively [10] [11].

It has been demonstrated multiple times that using a Deep Q-Network (DQN) provides an effective solution to this problem [12] [13].

III. CONTRIBUTION

The innovation of this research lies in integrating state-of-the-art Reinforcement Learning (RL) techniques into traffic management through Multi-Agent Reinforcement Learning (MARL). This approach offers a scalable, adaptive, and generalizable solution capable of dynamically addressing congestion and improving highway capacity utilization. It enables real-time management of traffic scenarios, minimizes the negative impacts of random anomalies, and provides a universally applicable method without requiring modifications to the state representation, action space, or reward function.

A key advantage of RL that plays a significant role in achieving this goal is its ability to evaluate and learn from traffic data in real-time. This allows for rapid adaptation to new traffic conditions and quick responses to continuously changing traffic dynamics.

The method's robustness is demonstrated by its generalizability, achieving excellent results regardless of the number of lanes. Consequently, this method provides an outstanding solution for optimizing highway traffic flow.

To validate the method, tests were conducted in three different environments: 3, 5, and 7-lane highway sections where lane closures occurred due to random anomalies. The Simulation of Urban MObility (SUMO) system served as an excellent platform for these simulations, offering customization and enabling the evaluation of the solution's efficiency in various traffic scenarios. This contributes to significant advancements in traffic management and congestion prevention.

IV. METHODOLOGY

A. Motorway Control System

The Motorway Control System (MCS) is a rule-based VSL algorithm that selects the appropriate speed limit from predefined values based on the given conditions to ensure continuous traffic flow. Traffic conditions are obtained using simple sensors, and the system sets the maximum permissible speed on the segments in km/h according to Equation (3),

$$speed\ limit = \begin{cases} 130 & if\ average\ speed \geq 40 \\ [110; 90; 70] & if\ average\ speed < 40 \end{cases} \quad (3)$$

where, if the average speed of the test segment exceeds 40 km/h, the maximum speed is set to 130 km/h; otherwise, a three-step speed limit is applied.

In the system's basic structure, the slowest lane, defined as the one with the lowest average speed, determines the speed limit for all lanes on a segment-by-segment basis.

B. Mainstream Traffic Flow Control

The Mainstream Traffic Flow Control (MTFC) algorithm helps maintain optimal traffic conditions, preventing congestion and ensuring smooth traffic flow. The MTFC determines the value of the variable speed limits at time t as a ratio of the original speed limit, according to Equation (4),

$$v(t) = v(t-1) + K_I' \cdot e_0(t) \qquad (4)$$

where $v(t)$ and $v(t-1)$ represent the new and previous speed limits, K_I' denotes the integral gain, and $e_0(t)$ is the occupancy error, defined as the difference between the critical occupancy $\hat{o}_{out}$ and the occupancy measured at the bottleneck $\tilde{o}_{out}$.

The algorithm's parameters are derived from well-established studies, with speed limits ranging between $[20\ km/h; 130\ km/h]$. The system integrates four sensors around the bottleneck area and uses the measured maximum occupancy as input.

C. Reinforcement Learning

Reinforcement Learning (RL) effectively addresses complex decision-making problems across various fields. Its application is becoming increasingly popular in different fields [14], including robotics, autonomous vehicle development, traffic system management, healthcare, and even financial markets.

The foundation of RL is the interaction between an agent and its environment, described by the Markov Decision Process (MDP). The environment provides the agent with the current state s and the available actions a. During decision-making, the agent selects an action, after which the environment transitions to a new state s' and provides a reward $R_a(s, s')$ to evaluate the agent's decision. This reward can be positive ("reward") or negative ("punishment"). The agent's goal is to learn to act in a way that maximizes future rewards over a defined time horizon. The MDP's four abstractions (S, A, R_a, P_a) describe the agent's control problem and enable it to develop the desired behavior while interacting with the environment in discrete time steps.

The MDP abstractions are defined as follows: The state space S and the action space A contain the possible states s_i and actions a_i. $R_a(s, s')$ is the immediate reward received after transitioning from state s to state s' due to action a. $P_a(s, s')$ represents the probability of transitioning from state s to state s' as a result of action a.

The agents were trained over 3000 episodes using a uniform replay buffer of 1 million transitions with a batch size of 512.

D. Multi Agent Reinforcement Learning

A multi-agent system can be defined as a group of autonomous, interacting entities that share a common environment, sense it using sensors, and act upon it using actuators.

Multi-Agent Reinforcement Learning (MARL) extends the traditional single-agent RL paradigm by involving multiple agents that interact, learn, and make decisions within a shared environment to optimize their objectives. Each agent is tasked with finding its optimal policy; however, in the current environment, all agents aim to determine the same policy, as described by Equation (5).

$$\pi = \{\pi_1, \pi_2, ..., \pi_n\} \qquad (5)$$

Here, π represents the set of policies, with each π_i in the set corresponding to an individual agent's policy, which in this case are identical. The MARL agents dynamically update their policies based on real-time feedback with the use of back-propagation, enabling adaptation to changing traffic conditions and anomalies. MARL offers a robust approach to developing cooperative strategies among agents, thereby enhancing overall decision-making performance.

V. ENVIRONMENT

During training, testing, and evaluation, we utilized the environment created with the Simulation of Urban MObility (SUMO) [15], [16] software package. SUMO is an open-source, widely recognized traffic simulator in the literature. It is dynamically extensible, highly customizable, and supports the creation of various scenarios with its robust tool set. SUMO enables the modeling of diverse scenarios, such as traffic congestion or accident situations, allowing the agent to learn in different environments, avoiding overfitting, and becoming more robust in optimal decision-making.

The Traffic Control Interface (TraCI) is essential to control the simulation process in real-time, particularly for managing agent-environment interactions. Traci allows the manipulation of the environment using various programming languages. The Traci API provides functionalities such as modifying vehicle speeds. Using these API calls, the agent can intervene in real-time traffic management, influencing the simulation while receiving real-time data on vehicle speeds and positions. Due to this, vehicle speeds are remotely controlled in real time via the TraCI API in SUMO, allowing agents to send speed commands ensuring adaptive management.

The simulation environments consist of 8 one-way road segments, each $300\ m$ long and featuring 3, 5, or 7 lanes. The starting points of the first segment's lanes randomly introduce vehicles into the simulation, ensuring diverse traffic scenarios to enhance the robustness of results. The first two segments ($600\ m$) are designated for observation purposes, while the 3rd and 4th segments are VSLC zones without anomalies. The 5th to 8th segments are also VSLC zones but include random anomalies, leading to potential bottlenecks.

To optimize traffic flow, the speed limits for each lane are independently adjustable. These values range from a minimum of $20\ km/h$ to a maximum of $130\ km/h$ in $10\ km/h$ increments. The time interval between speed limit adjustments is restricted to a minimum of 15 seconds. These discrete transitions help to take into account the human reaction time and the time needed to perform the speed changes. Using continuous models, it would not be possible to adapt to a continuously changing speed limit at the traffic signals.

A. State Representation

In this study, the state representation of a single observation section (one lane of one segment) is described by a 5×5 matrix. The analyzed state is always located at the center of the matrix, leaving two units in each direction to populate the matrix.

Based on Fig. 3, the state matrix for the cell filled with yellow vehicles takes the form specified in Equation (6),

$$state_i = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & \rho_i & \rho_i & 0 \\ -1 & -1 & \rho_i & \rho_i & \rho_i \\ -1 & -1 & \rho_i & \rho_i & \rho_i \\ -1 & -1 & \rho_i & \rho_i & 0 \end{bmatrix} \qquad (6)$$

where the value of saturated cells is given by the occupancy metric ρ_i. This metric is calculated as the ratio of the number of vehicles in the section to the length of the section. Consequently, the value of empty cells is 0. The cells of the state representation matrix take a value of -1 if no segment or lane exists to be examined at the given distance from the center of the analyzed state.

B. Action Space

Several studies in this field utilize high-dimensional vectors to handle the wide range of permissible speed limit values ([17], [18]). In contrast, this research reduces the action space to a three-dimensional vector that defines speed limit changes. Equation (7) presents the action space vector,

$$action = \begin{bmatrix} +10 \ km/h \\ 0 \ km/h \\ -10 \ km/h \end{bmatrix} \qquad (7)$$

illustrating that the possible actions include changing the speed limit by $-10 \ km/h$ or $+10 \ km/h$, or maintaining the current speed limit.

C. Reward Function

The reward function for the agents is designed to minimize the waiting time across the entire observed traffic network, as shown in Equation (8).

$$reward_{t+1} = \frac{1}{\omega_t + \epsilon} \qquad (8)$$

The reward at time step $t+1$ is determined by the waiting time ω_t during the previous time interval Δt, with the inclusion of a small constant ϵ to prevent division by zero.
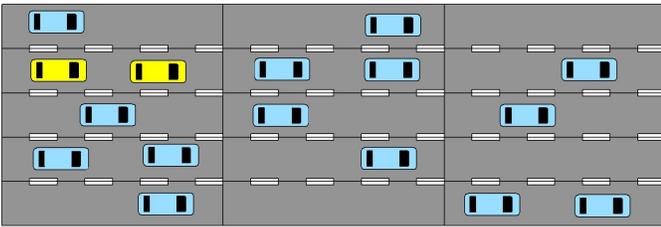

Fig. 3. Schematic illustration of the state representation matrix

VI. RESULTS

The continuous and smooth traffic flow is critically important from environmental and safety perspectives. As such, proper traffic management is indispensable for preventing congestion. Therefore, it is essential to develop a system that delivers optimal performance in resolving traffic-related issues while also being generalizable to handle various random anomalies.

To address this, the methods are tested on highways with 3, 5, and 7 lanes. By comparing the performance of different techniques, it can be determined whether the method implemented with Multi-Agent Reinforcement Learning (MARL) is sufficiently robust to generalize the task and handle random anomalies that arise on highways with varying lane capacities.

Fig. 4 clearly illustrates that traffic managed by our proposed RL-based method achieved more favorable results across all examined aspects than the commonly used MTFC and MCS algorithms in the literature, which have real-world applications and are therefore suitable for comparison with MARL, in all aspects considered. The robustness of MARL was particularly evident in addressing issues caused by random anomalies in all three simulated environments—highways with 3, 5, and 7 lanes. This method successfully reduced critical traffic parameters, ensuring continuous flow. Additionally, we significantly decreased CO_2 and NO_x emissions and fuel consumption (which are emission estimates based on SUMO's built-in emission models, taking into account vehicle speed and acceleration), contributing to the development of sustainable transportation systems.

The results also demonstrate that the method provides a cost-effective, energy-efficient solution for ensuring optimal traffic flow without increasing capacity. The robustness of the RL-based method proved particularly outstanding in managing traffic fluctuations and anomalies, offering a long-term, effective solution for stabilizing dynamically changing traffic conditions, such as sudden congestion or accidents.

The six criteria considered in the evaluation and their corresponding values are presented in table I. As clearly shown in the diagrams, MARL outperformed the examined methods in every evaluated aspect.

The relative efficiency gains achieved by MARL are represented in the last row, which reveals a significant reduction in waiting time and in queue length ($56 - 63$ %) compared to previous methods. This ensures that traffic flow remains dynamic, with vehicles able to move continuously without being forced to stop. Along with reducing the risks associated with sudden starts and stops, the approach also achieved notable reductions in emissions and fuel consumption ($11 - 13$ %). The overall travel time was also reduced by $4 - 5$ %.

These results demonstrate that this solution is universally applicable for optimizing highway traffic, regardless of highway capacity, whether it involves n-lane sections. The flexibility of the method enables effective use under various traffic loads, helping to mitigate congestion, shorten vehicle transit times, and reduce fuel consumption and environmental impact.
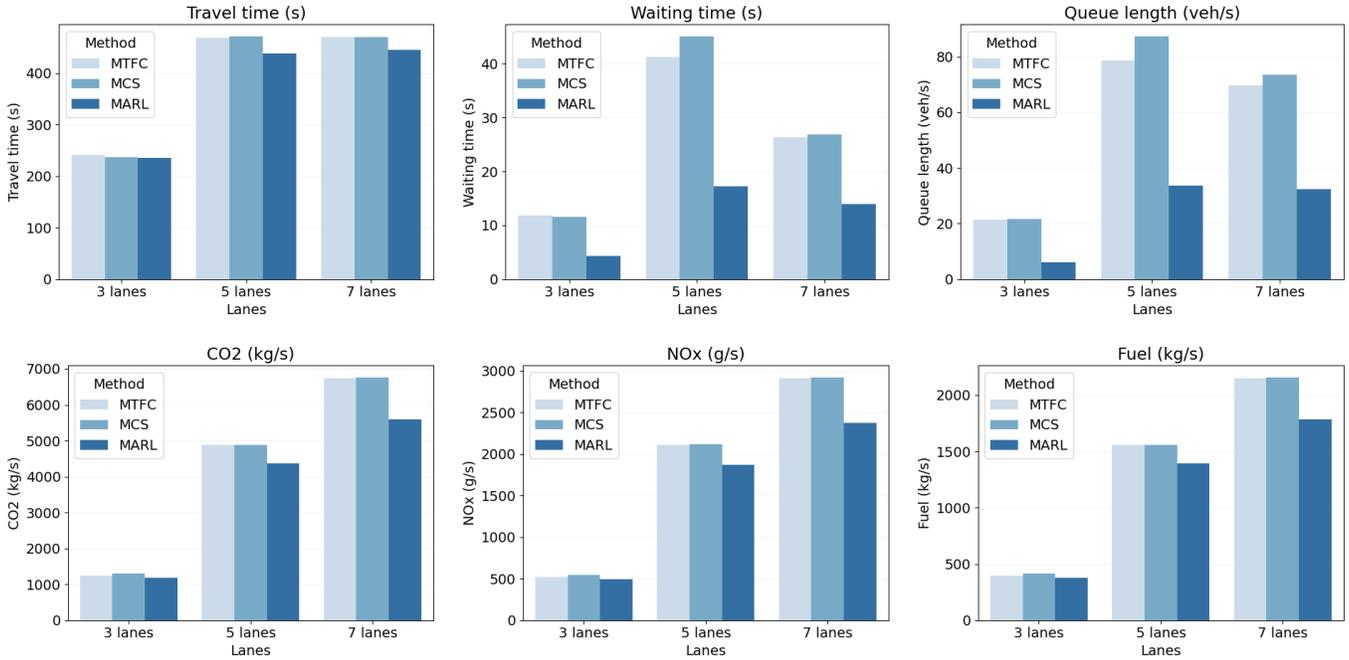
Fig. 4. The effects of the analyzed methods on traffic parameters for highways of differing lane capacities.

TABLE I
TRAFFIC-RELATED DATA FOR HIGHWAYS WITH DIFFERENT LANE CAPACITIES, MANAGED BY THE EVALUATED METHODS

ine ine Number of lanes	Method	Travel time [s]	Waiting time [s]	Queue length [veh/s]	CO_2 [kg/s]	NO_x [g/s]	Fuel [kg/s]
	MCS	237,3	11,6	21,6	1294,7	546,6	413,0
3 lane	MTFC	241,1	11,8	21,5	1242,7	522,1	396,4
	MARL	235,9	4,3	6,1	1186,4	495,4	378,4
ine 5 lane	MCS	470,7	45,0	87,2	4891,8	2114,3	1560,3
	MTFC	469,0	41,2	78,6	4883,8	2110,8	1557,8
	MARL	437,5	17,3	33,7	4366,2	1869,5	1392,7
ine 7 lane	MCS	470,2	26,9	73,6	6754,2	2917,0	2154,3
	MTFC	469,8	26,3	69,8	6738,5	2909,1	2149,3
	MARL	445,3	14,0	32,4	5589,4	2374,6	1782,8
ine ine Relative efficiency gain ine ine	MTFC	4,69%	56,18%	60,83%	10,73%	11,64%	10,73%
	MCS	4,31%	57,55%	63,10%	12,12%	13,18%	12,12%

VII. Conclusion and Future Work

Efficient traffic management on highways poses significant challenges, particularly in the presence of random traffic anomalies such as sudden braking or the resulting accidents that lead to capacity reductions. Traditional traffic management solutions often fail to dynamically address unexpected situations and struggle to adapt to ever-changing traffic conditions. To track sudden, unpredictable changes in traffic, optimize traffic flow, and reduce environmental impact, the goal is to develop an ITS capable of dynamically adjusting speed limits to ensure smooth traffic flow, even in the face of random anomalies.

With the introduction of the MARL algorithm described in this paper, the traffic control system dynamically adjusts the allowable speed limits for each lane and segment of the highway, adapting to the specific situation. This is made possible by MARL's ability to intervene in real-time, respond to the current state, adjust to various anomalies, and change highway traffic conditions. The algorithm focuses on preventing traffic congestion and effectively addresses bottleneck-related issues without relying on complex mathematical models.

The results demonstrate that the MARL-based traffic management solution can be applied to highways with any lane capacity and outperforms traditional algorithms in handling random traffic anomalies. This approach ensures continuous traffic flow. In addition, the algorithm has successfully reduced travel time, waiting time, and queue length while significantly lowering harmful emissions (CO_2, NO_x) and fuel consumption, providing a more sustainable and efficient solution.

Based on the results of this research, the presented algorithm could also be applied to other traffic management problems, such as addressing challenges in urban traffic or regulating traffic on regional roads. However, this solution may be limited by drivers' non-compliant behavior, which can be eliminated by the introduction of autonomous vehicles. Future research directions include incorporating different weather conditions and human rule-breaking behavior into the models for a more accurate representation of reality and exploring the integration of autonomous vehicles and intelligent infrastructure to optimize traffic flow further.

References

[1] M. Treiber and A. Kesting, "Traffic flow dynamics," *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*, pp. 983–1000, 2013.

[2] N. Khoudari and B. Seibold, "Multiscale properties of traffic flow: the macroscopic impact of traffic waves," in *Multiplicity of Time Scales in Complex Systems: Challenges for Sciences and Communication I*. Springer, 2022, pp. 267–296.

[3] E. F. Grumert, A. Tapani, and X. Ma, "Characteristics of variable speed limit systems," *European transport research review*, vol. 10, pp. 1–12, 2018.

[4] B. Khondaker and L. Kattan, "Variable speed limit: an overview," *Transportation Letters*, vol. 7, no. 5, pp. 264–278, 2015.

[5] K. Kušić, N. Korent, M. Gregurić, and E. Ivanjko, "Comparison of two controllers for variable speed limit control," in *2016 International Symposium ELMAR*. IEEE, 2016, pp. 101–106.

[6] K. Kušić, E. Ivanjko, M. Gregurić, and M. Miletić, "An overview of reinforcement learning methods for variable speed limit control," *Applied Sciences*, vol. 10, no. 14, p. 4917, 2020.

[7] F. Zhu and S. V. Ukkusuri, "Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach," *Transportation research part C: emerging technologies*, vol. 41, pp. 30–47, 2014.

[8] K. Kušić, E. Ivanjko, and M. Gregurić, "A comparison of different state representations for reinforcement learning based variable speed limit control," in *2018 26th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2018, pp. 1–6.

[9] B. Pelenczei, I. G. Knáb, B. Kővári, T. Bécsi, and L. Palkovics, "Adaptive highway traffic management: A reinforcement learning approach for variable speed limit control with random anomalies," 2024, preprint available on ResearchGate.

[10] S. Swapno, S. Nobel, P. Meena, V. Meena, A. T. Azar, Z. Haider, and M. Tounsi, "A reinforcement learning approach for reducing traffic congestion using deep q learning," *Scientific Reports*, vol. 14, no. 1, pp. 1–20, 2024.

[11] H. Zhao, C. Dong, J. Cao, and Q. Chen, "A survey on deep reinforcement learning approaches for traffic signal control," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108100, 2024.

[12] B. Kővári, I. Knáb, and T. Bécsi, "Variable speed limit control for highway scenarios a multi-agent reinforcement learning based appraoch," EasyChair, Tech. Rep., 2024.

[13] M. Gregurić, K. Kušić, F. Vrbanić, and E. Ivanjko, "Variable speed limit control based on deep reinforcement learning: A possible implementation," in *2020 International Symposium ELMAR*. IEEE, 2020, pp. 67–72.

[14] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[15] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

[16] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility: an overview," in *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

[17] Z. Li, C. Xu, Y. Guo, P. Liu, and Z. Pu, "Reinforcement learning-based variable speed limits control to reduce crash risks near traffic oscillations on freeways," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 4, pp. 64–70, 2020.

[18] K. Kušić, I. Dusparic, M. Guériau, M. Gregurić, and E. Ivanjko, "Extended variable speed limit control using multi-agent reinforcement learning," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.

[19] M. Héder, E. Rigó, D. Medgyesi, R. Lovas, S. Tenczer, F. Török, A. Farkas, M. Emődi, J. Kadlecsik, G. Mező, Á. Pintér, and P. Kacsuk, "The past, present and future of the ELKH cloud," *Információs Társadalom*, vol. 22, no. 2, p. 128, aug 2022. [Online]. Available: