Multi-region perimeter control in complex urban networks: A reinforcement learning approach

E. Kampitakis[a*], E. I. Vlahogianni[a]

[a] National Technical University of Athens - School of Civil Engineering, Athens, Greece
ekampitakis@central.ntua.gr, elenivl@central.ntua.gr
* Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 29, 2024

Keywords: perimeter control, reinforcement learning, traffic control, queue balancing

1 INTRODUCTION

As urban centers become more densely populated, the rise in vehicular traffic has overwhelmed existing transportation infrastructure. To mitigate the adverse effects of traffic congestion, various traffic control methods have been developed and implemented. Perimeter control is a promising direction which aims to regulate vehicle inflows at the boundaries of protected areas to prevent congestion within. Most studies in literature have developed MFD-enabled perimeter control strategies showcasing excellent performance in single and multi-region systems. More recently, the success of AI control methods has led the researchers to utilize deep reinforcement learning (DRL) methods. In general, there have been limited studies introducing RL-based perimeter control methods that are trained through direct interaction with a microsimulation environment, which is considered the most accurate representation of real-world traffic dynamics. In this work, we formulate the multi-region perimeter control task as a model-free DRL problem trained directly on a microsimulation environment and a 2-stage method is proposed that combines DRL with classic optimization to account for the queue formation on the boundaries of the multi-region system.

2 METHODOLOGY

In this study, a multi-region system is assumed, i.e., an urban traffic network that is partitioned in homogeneous subregions. The goal is to regulate the incoming flow at the boundaries of each subregion (external and internal) by altering traffic signal programs in order to retain the number of vehicles to a point where the network works at capacity. Although the existence of an MFD is assumed, the developed method is a model-free one, i.e., no MFD dynamics are available, and subregions' critical accumulation is unknown to the agent. In stage I, a centralized RL agent interacts with the microscopic simulator and by getting feedback about the network's conditions accordingly decides the number of vehicles that should enter through the predefined boundaries resulting from a partitioning algorithm. The RL agent's observation includes information regarding the accumulation per subregion as well as the number of waiting vehicles, the measured outflow and inflow per boundary. The action of the agent is the number of vehicles that are allowed to enter per boundary during the next control step (90s– cycle duration) bounded by a minimum and a maximum inflow. The agent's reward is the trip completion rate of the whole network, as its ultimate objective is to minimize the total travel time at a network level. As our RL training algorithm, we utilized the Proximal Policy Optimization (Schulman et al., 2017) algorithm which has demonstrated great performance in various continuous control tasks.

As a gating strategy is applied on a subregion's boundary, queues are formed upstream the gated links, which if treated uniformly, can result in wasted green time and spillovers, followed by unnecessary delays on vehicles either waiting to enter or exit the boundary. To mitigate the negative effects of queueing, a queue balancing strategy is developed (stage II) which acts in combination with the RL agent. The objective of the strategy is to minimize queues with respect to RL agent's actions while enhancing the stability of the system (i.e. by minimizing the difference between consecutive green time settings).

(a) (b)

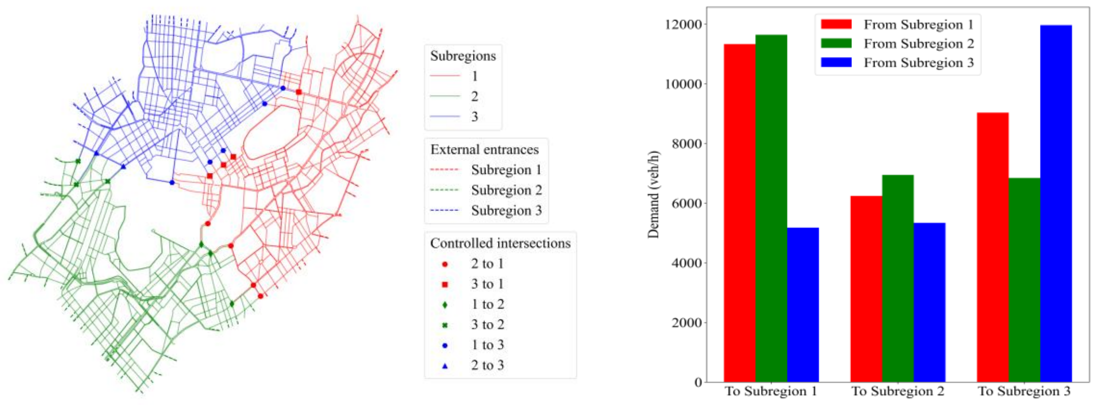

Figure 1 – *(a) The partitioned network of Athens' city center; (b) Subregion-level peak hour demand.*

For this study, the Athens city center network (~24km^2) was created in Simulation of Urban Mobility (SUMO) traffic simulator (Krajzewicz et al., 2012). The network is partitioned into three homogeneous subregions (9 boundaries) as is depicted in Figure 1(a) by utilizing the methodology proposed in (Saeedmanesh & Geroliminis, 2016). Moreover, we assume that we can control all external entrances including those that are not controlled by a traffic light ensuring in that case that we allow at least 20% of the real saturation flow. The generated demand (Figure 1(b)) consists of a 15-minute warm-up period followed by a 1-hour constant peak demand (morning peak hour). The demand after that period of 1.25 hours (~78000 trips) becomes zero. The total simulation time is 4.3 hours, which coincides with the time needed for all vehicles to complete their trip.

3 RESULTS

To verify the efficiency of the proposed method 10 replications with different random seeds were run for each of the following control scenarios:

- **NPC**: no perimeter control is applied
- **PI**: multivariable Proportional–Integral (PI) feedback regulator (Kouvelas et al., 2017)
- **RL**: our approach (converged after 200 episodes – simulation runs)

Regarding the setup of the PI controller, after an initial manual tuning, it was further calibrated using the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm. For the distribution of the ordered flows per boundary the same queue balancing framework was used as the RL controller (stage II).

The network exhibits the worst performance when no perimeter control (NPC) is applied, as is presented in Table 1. The proposed RL-based method reached a mean improvement of 20.54% in VHT compared to NPC also improving the results of the PI controller by ~8%.

Table 1 - *Vehicle Hours Travelled, Total Depart Delay and Speed comparison for NPC, PI and RL.*

Control scenarios	Vehicle Hours Travelled	Percentage Difference (%)	Total Depart Delay (hours)	Percentage Difference (%)	Speed (km/h)	Percentage Difference (%)
NPC	81961	-	28479	-	7.73	-
PI	70915	-13.48	38175	34.05	10.14	31.2
RL	65129	-20.54	37079	30.20	11.11	43.7

In Figure 2, the network-level MFD and in Figure 3 the MFDs per subregion for the onload of congestion are illustrated for all control scenarios. Both control methods are successful in keeping the accumulation around the critical value, avoiding the saturated branch of the MFD, with the RL method consistently keeping the trip completion rate slightly higher compared to the PI controller.

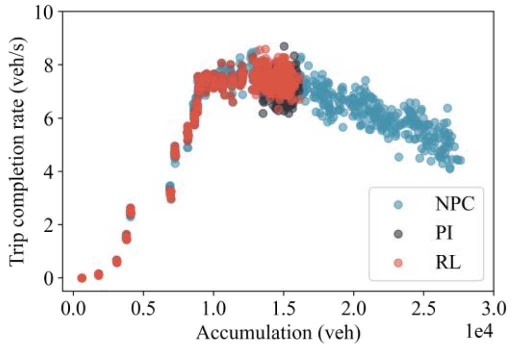

Figure 2 – *MFD (network-level) for the NPC, PI and RL scenarios for the onload of congestion.*

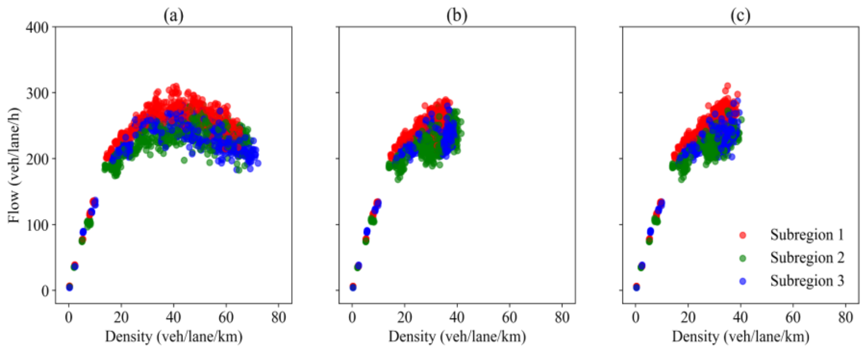

Figure 3 - *MFDs (subregion-level) for NPC (a), PI (b) and RL (c) scenarios for the onload of congestion.*

In Figure 4, the accumulation measurements per control step (90s) for all scenarios are illustrated. An interesting observation is that until the 50th timestep which coincides with the end of the onload period the pattern for PI and RL controller is almost the same. After that point in time the accumulation trajectories are completely different, with the RL controller possessing lower accumulation values but delaying emptying the network compared to PI.

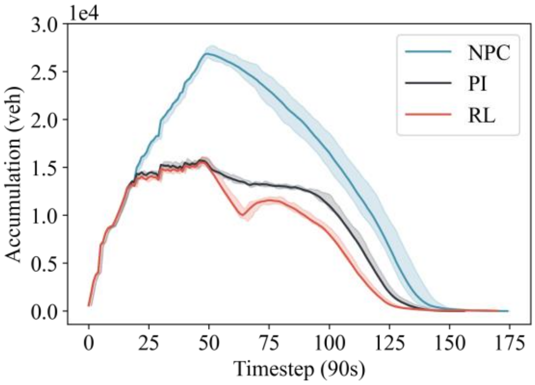

Figure 4 - *Accumulation trajectory for all control scenarios*

4 DISCUSSION

Perimeter control has emerged as a promising traffic management strategy to mitigate the adverse effects of traffic congestion, especially under oversaturated conditions. Existing studies have mostly focused on MFD-based modeling approaches or approaches that require knowledge of the MFD. Furthermore, most model-free approaches found in the literature are using macroscopic traffic dynamics which puts limitations as a result of neglecting important idiosyncrasies of real-world traffic dynamics. This study presented a 2-stage model-free RL approach for multi-region perimeter control which is trained on a microscopic level. The proposed method efficiently learns to maximize the network's trip completion rate and clearly perceives the existence of the subregions' MFDs, achieving an improvement of 20.54% in VHT compared to NPC and improving the results of the PI controller by ~8%.

ACKNOWLEDGMENTS
This work is part of the DIT4TraM project which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 953783.

5 REFERENCES

Kouvelas, A., Saeedmanesh, M., & Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B: Methodological*, *96*, 26–45. https://doi.org/10.1016/j.trb.2016.10.011

Krajzewicz, D., Erdmann, J., Behrisch, M., & Bieker, L. (2012). *Recent Development and Applications of SUMO - Simulation of Urban MObility*.

Saeedmanesh, M., & Geroliminis, N. (2016). Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transportation Research Part B: Methodological*, *91*, 250–269. https://doi.org/10.1016/j.trb.2016.05.008

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal Policy Optimization Algorithms* (arXiv:1707.06347). arXiv. http://arxiv.org/abs/1707.06347