Multi-scale model-free perimeter control and local signal control in urban networks

D. Zhou[a,*], V. V. Gayah[a]

[a] Department of Civil and Environmental Engineering, The Pennsylvania State University, USA
dongqinzhou1110@gmail.com, gayah@engr.psu.edu
[*] 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 5, 2024

Keywords: Perimeter control, traffic signal control, joint control, deep reinforcement learning

1 INTRODUCTION

Perimeter control, based on aggregate dynamics modeling using network Macroscopic Fundamental Diagrams (MFDs), has been shown effective in congestion mitigation and throughput maximization for urban networks comprised of a single or multiple homogeneous regions. However, in dense urban areas, local pockets of congestion might form, resulting in traffic heterogeneity, which diminishes the effectiveness of perimeter control. To this end, an integrated framework that regulates both the inter-regional exchange flows (viz., perimeter control) and intra-regional traffic signals is proposed, wherein the upper-level perimeter control helps maintain regional accumulations around the critical levels while the lower-level signal control combats local congestion to improve traffic homogeneity.

Early endeavors in such frameworks often require the exchange of information between the levels to coordinate control objectives, thus demanding extensive communication infrastructure. To relieve such requirements, frameworks with independent controllers are receiving increasing interest. In (Keyvan-Ekbatani et al., 2019), a modified SCATS strategy and a volume-based approach for local signal control are combined with a proportional-integral (PI) type perimeter controller. In (Tsitsokas et al., 2023), the PI-type regulator is used with the max pressure (MP) controller, while in (Su et al., 2023) the perimeter control problem is solved with reinforcement learning (RL). Following this line of research, this work studies the joint perimeter and signal control problem, where both levels are controlled by RL agents. While RL has been applied to signal control extensively and is also gaining momentum in perimeter control applications, its effectiveness hasn't been investigated for the joint control problem. This work thus extends the frameworks in (Su et al., 2023; Zhou and Gayah, 2024) to consider RL for lower-level signal control. A multi-timescale multi-agent training approach is presented, with its effectiveness evaluated in simulated single-region networks. The experiment results show the presented approach is highly comparable (and often times superior) to a baseline scheme comprised of upper-level Bang-Bang control and lower-level MP controller (Varaiya, 2013).

2 A MULTI-TIMESCALE RL APPROACH

The joint perimeter and signal control problem is formulated as a Markov decision process, where the environment represents a simulated single-region network (see Figure 1). An upper-level agent (dubbed U-RL) selects actions at regular intervals of ΔT that determine the green times at perimeter intersections. Similarly, a lower-level agent (dubbed L-RL) selects actions every $\Delta t \leq \Delta T$ to set the signal timings for signal control intersections. Concretely, at each interval ΔT, the U-RL takes as input the accumulation, regional speed and flow, and standard deviation of lane-level vehicle counts. It then selects among $\{0,0.2,\cdots,0.8,1.0\}$ as the ratio of green times allocated to entering vehicles at perimeter intersections. At the end of interval ΔT, the U-RL receives a reward from the environment,

which is the traveled distance of all vehicles to encourage higher traffic throughput. The Double DQN algorithm and a distributed learning architecture are adopted to facilitate training for U-RL, and to consider the delayed impacts of perimeter control, multi-step return is used.

Similarly, the L-RL takes actions every Δt but does so for each intra-regional intersection. For this, the parameter sharing technique is utilized to reduce training time since many intersections exist. Specifically, each intersection is controlled by an agent, and all these agents share the same model weights. Each of them receives individualized information and selects individualized actions as well as obtains individualized rewards. Afterwards, all state-action-reward transitions are pooled together to train the L-RL, and the updated model weights are shared by all agents again in the subsequent action-taking processes. Here, the input includes the average number of vehicles (weighted by turn ratios) of the four downstream approaches, the grouped upstream vehicle counts, the current phase, and the regional accumulation. The action specifies which phase to choose for each intersection, and the reward is the (normalized) number of discharged vehicles to encourage higher traffic production.

A few further remarks are provided here for the multi-timescale multi-agent training approach. First, the two agents interact with the environment simultaneously, but at different timescales (ΔT and Δt). As such, a state-action-reward transition is obtained every Δt for L-RL but every ΔT for U-RL. Also, following the popular independent learning paradigm, the two agents are trained separately, thus reliving the need for increasing communication infrastructure. Further, with the distributed learning architecture, both L-RL and U-RL are updated only once per iteration. However, such updates can utilize all transitions collected during the iteration which helps improve convergence for the agents.

To demonstrate the presented method, a variety of joint control schemes are utilized for comparison. At the upper level, the Bang-Bang (BB) policy is used which builds upon MFD-based modeling and alternates its action by comparing the regional accumulation to the critical value. At the lower level, the MP controller (Varaiya, 2013) is adopted. Further, a non-adaptive signal control scheme (fixed time, FT) is used together with the no control (NC) policy at the upper level to benchmark the lower-bound performances. The full list of baselines includes NC+FT, NC+MP, BB+FT, and BB+MP.

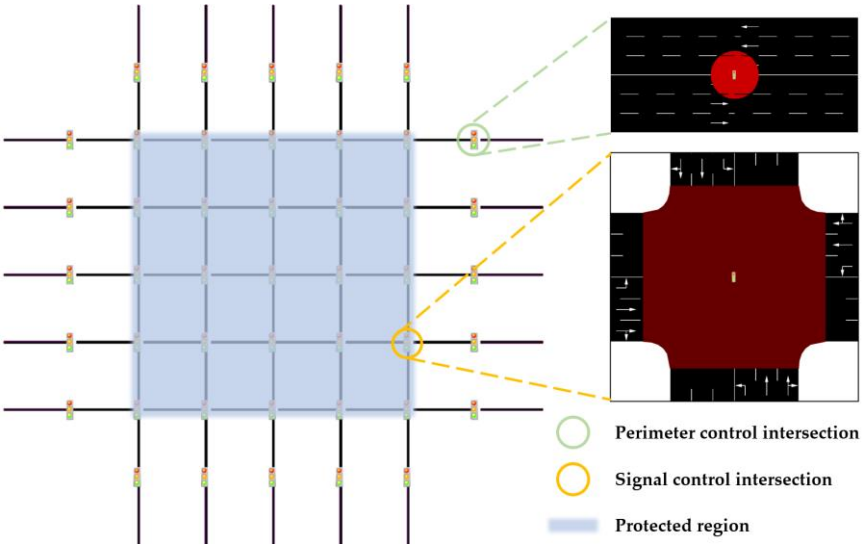


Figure 1. The simulated single-region urban network.

3 EXPERIMENTS

3.1 Single-region urban network setup

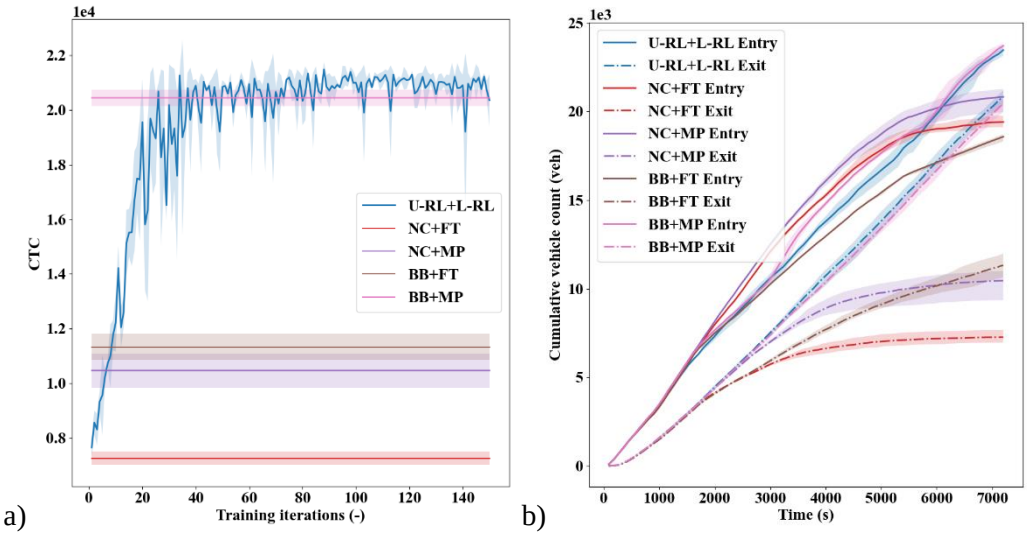
Figure 1 shows the simulated single-region network. Each link is 500m long with three lanes in each direction. The free flow speed is 50 km/h and the saturation flow 1800 veh/h/lane. All intersections

are signalized, and the perimeter intersections have a shared cycle length of 90s ($\Delta T = 90$s). The minimum and maximum green times are respectively 5s and 87s. The intra-regional intersections either adopt a FT signal plan or are controlled by the MP policy or L-RL. The FT plan, MP policy, and L-RL all share the same set of phases (to make sure the simulation results are comparable), but their sequence order and duration may differ. The simulation step is set to 1s and $\Delta t = 10$s.

The origins are evenly distributed across the entire network whereas the destinations are only located inside the protected region. A strong directional demand is assumed from outside of the region, which lasts for 90 minutes as followed by a recovery period of 30 minutes. The simulated vehicles are routed using the stochastic C-logit model, and a subset (60%) can perform adaptive rerouting at regular intervals of 3 minutes. Initially, strong demands were assumed to fill up the network and to obtain the MFD plot, and the critical accumulation (used by BB policy) is determined as 3000 veh.

3.2 Experiment results

The objective of the joint control framework is to maximize network throughput, i.e., cumulative trip completion (CTC), and the learning curves express the evolution of CTC over training iterations; see Figure 2(a). The baselines have constant CTCs as they are not learning-based methods, and the bands reflect the randomness in the simulation. As can be seen, the joint control agent (U-RL+L-RL) can learn effectively to realize performances directly comparable BB+MP. This is notable since the BB controller is optimal for single-region perimeter control whereas the MP policy is an established signal controller with proven ability of throughput maximization. Figure 2(b) presents the cumulative count curves by each method, where "Exit" indicates the cumulative trip completion and "Entry" the cumulative number of vehicles from either demand generation or arrival from outside of the region. As shown, without perimeter control, using the MP policy yields much higher cumulative vehicle entry and trip completion than FT. Yet, the overall trip completion by MP alone is rather modest as it cannot handle oversaturated conditions well; however, this can be remedied by using BB policy at the upper level, since it can delay vehicle entry, thus allowing for more trip completion and vehicle entry later on. The curves by BB+MP and U-RL+L-RL exhibit a high level of similarity, and by comparing the differences between (as well as the differences of areas under) the entry and exit curves, one can conclude U-RL+L-RL realizes the steadiest accumulation and smallest total travel time among all control methods, which showcases its competitiveness.



a) b)

Figure 2. Learning curve (a) and cumulative count curves (b).

To evaluate the robustness of U-RL+L-RL, measurement noise of accumulation is considered, in the form of a mean-zero normal distribution $\mathbb{N}(0, \delta^2)$. The BB policy is prone to such noise as it acts upon the accumulation, contrary to NC methods. Here, the noise level δ ranges from 25 to 300, and the realized CTC values by each method are shown in Figure 3(a), where the error bars indicate 95%

confidence interval. As shown, BB+FT has decreasing CTCs that can even be smaller than NC+MP, whereas both BB+MP and U-RL+L-RL are extremely robust against the noise. Importantly, note U-RL+L-RL is impacted by the noise to a greater degree, as both agents take accumulation information (in comparison, only the BB policy, not MP, is prone to the noise) and such inaccuracy is in effect at a higher frequency (every 10s for L-RL) than BB+MP (every 90s). This contrast thus highlights the learning robustness of U-RL+L-RL. Furthermore, notice the lane-level vehicle counts are used by L-RL, so additional errors on these counts (also in the form of a normal distribution) are examined as well. In particular, this error impacts vehicle counts on all lanes, so its magnitude is smaller than measurement noise. We consider combinations of noise in accumulation (from 100 to 300) and errors in lane-level vehicle counts (3 and 6), and two example learning curves are shown in Figure 3(b) for combinations 100/3 and 300/6. These curves again illustrate the robustness of U-RL+L-RL.

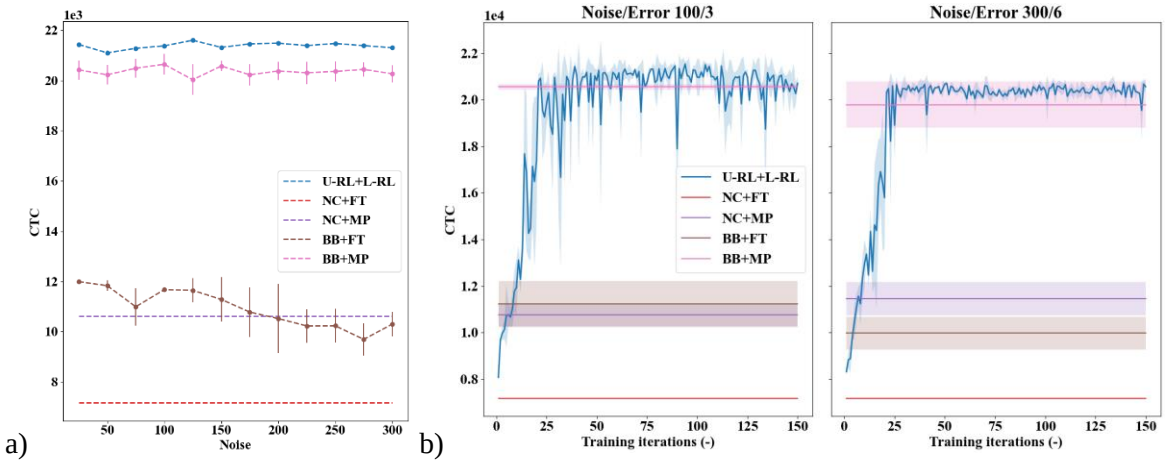


Figure 3. Realized CTCs against measurement noise (a) and example learning curves (b).

4 CONCLUSION

This paper presents a multi-timescale reinforcement learning approach for the joint perimeter and signal control problem in urban networks. Using established techniques like parameter sharing and independent learning, the method exhibits excellent control effectiveness and robustness. This joint control framework holds promise for efficient urban traffic management and contributes ultimately to emerging intelligent transportation systems, with a learning-based design (not built upon heuristic rules) and comprehensive (featuring perimeter and signal control) yet fully implementable policies.

REFERENCES

Keyvan-Ekbatani, M., Gao, X., Gayah, V. V., Knoop, V.L., 2019. Traffic-responsive signals combined with perimeter control: investigating the benefits. Transp. B Transp. Dyn. 7, 1402–1425. https://doi.org/10.1080/21680566.2019.1630688

Su, Z.C., Chow, A.H.F., Fang, C.L., Liang, E.M., Zhong, R.X., 2023. Hierarchical control for stochastic network traffic with reinforcement learning. Transp. Res. Part B Methodol. 167, 196–216. https://doi.org/10.1016/J.TRB.2022.12.001

Tsitsokas, D., Kouvelas, A., Geroliminis, N., 2023. Two-layer adaptive signal control framework for large-scale dynamically-congested networks: Combining efficient Max Pressure with Perimeter Control. Transp. Res. Part C Emerg. Technol. 152, 104128. https://doi.org/10.1016/J.TRC.2023.104128

Varaiya, P., 2013. Max pressure control of a network of signalized intersections. Transp. Res. Part C Emerg. Technol. 36, 177–195. https://doi.org/10.1016/j.trc.2013.08.014

Zhou, D., Gayah, V. V., 2024. Evaluating the Effectiveness and Transferability of a Data-Driven Two-Region Perimeter Control Method Using Microsimulation. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/03611981241230313