Modeling paradigm for adaptive decentralized traffic control via a rollout reinforcement learning approach

Zhongyang Lu, Andy H.F. Chow*

Department of Systems Engineering, City University of Hong Kong, Hong Kong

*Corresponding author: andychow@cityu.edu.hk

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

October 30, 2024

Keywords: Decentralized adaptive traffic control, rollout reinforcement learning, Markov decision process, spatial queue modeling, model-plant mismatch

1 INTRODUCTION

Most previous RL-based traffic signal controllers primarily adopt a model-free approach, which does not rely on an underlying traffic model but directly derives control policies from observed traffic states. The absence of an underlying traffic model within the RL-based control framework can lead to system instability under demand uncertainties and disruptions that are not well-trained. Therefore, it is important to incorporate a tractable traffic model with purely data-driven RL-based algorithms to enhance traffic control systems (Chow *et al.*, 2019). Meanwhile, to mitigate the increased computational time associated with sophisticated modeling design, it is necessary to develop a modeling paradigm that can adapt to arbitrary traffic models without requiring sophisticated calibration and address deviations arising from model-plant mismatch.

This paper presents a modeling paradigm for decentralized adaptive network traffic control using rollout reinforcement learning (Sutton, 2018). The optimal control problem is formulated to minimize network delays. Unlike model-free RL methods, rollout reinforcement learning integrates a model-based component that produces short-term traffic state estimates by explicitly modeling traffic flows within a predefined planning horizon. To address the computational complexity in policy evaluation, the model-based simulation is further broken down into a decentralized approach, allowing local intersections to simulate asynchronously.

2 DYNAMIC NETWORK TRAFFIC MODEL

The stochastic transition function P is represented by the network traffic model, derives the traffic state $\boldsymbol{x}$ by considering the control policy $\boldsymbol{\mu}$ and the stochastic traffic demand $\boldsymbol{\Pi}_t$, as follows:

$$\boldsymbol{x}_{k+1} \sim P(\boldsymbol{x}_k, \boldsymbol{\mu}_k | \boldsymbol{\Pi}_t) \tag{1}$$

where k denotes the decision stage. The traffic state $\boldsymbol{x}_k$ can be determined by traffic managers based on the available data sources. The control policy $\boldsymbol{\mu}_k$ specifies the signal aspects that regulate traffic flows at each signalized junction for every decision interval k.

In the node model, node n connects a group of entrance links in $\boldsymbol{I}_n$ with a set of exit links in $\boldsymbol{J}_n$. The turning ratio $\theta^n_{ij}(t)$ determines the proportion of vehicles flowing from entrance link

$i \in \boldsymbol{I}_n$ to exit link $j \in \boldsymbol{J}_n$ through node n at time t. The traffic flow $\nu_{ij}(t)$ from entrance link i to exit link j through node n at time t can be determined as follows:

$$\nu_{ij}(t) = \theta_{ij}(t)\psi_i(t)\mu_{ij}^n(t) \tag{2}$$

where $\psi_i(t)$ denotes the sending flow on link i that is aiming to pass through node n at time t. The control signal $\mu_{ij}^n(t)$ is a binary variable that controls the traffic flow from link i to link j through node n during time interval t. Correspondingly, the total received flow $\phi_j(t)$ of link j through node n can be derived as:

$$\phi_j(t) = \sum_{\forall i \in I_n} \nu_{ij}(t) \tag{3}$$

We introduce the point queue model and the spatial queue model to illustrate traffic flow propagation rules between links. The primary difference between these two models is that the spatial queue model accounts for the physical length of vehicles, whereas the point queue model does not (Zhang *et al.*, 2013). With the determined outflow $\psi_i(t)$ of link i, the number of queuing vehicles can be derived as follows:

$$\lambda_i(t+1) = \lambda_i(t) + \phi_i(t-\tau_i) - \psi_i(t) \tag{4}$$

where τ_i denote the free-flow travel time of link i, the number of queuing vehicles $\lambda_i(t)$ is determined by the inflow $\phi_i(t-\tau_i)$ at time $t-\tau_i$ and the outflow $\psi_i(t)$ at time t in link i. We can derive the overall network delay at time t calculated by aggregating the delays from all nodes:

$$\sigma_t = \sum_{\forall n \in \mathcal{N}} \sum_{\forall i \in I_n} \lambda_i(t)\Delta t \tag{5}$$

3 DECENTRALIZED SIGNAL CONTROL

The optimal control framework aims to derive the optimal signal control policy $\boldsymbol{\mu}_{k_0}^*$ for the entire network that minimizes overall network delay at each decision stage k_0 as follows:

$$\min_{\boldsymbol{\mu}_{k_0} \in \mathcal{A}} Z(\boldsymbol{x}_{k_0}) = \mathbb{E}_{\boldsymbol{x}_{k+1} \sim P(\boldsymbol{x}_k, \boldsymbol{\mu}_k | \boldsymbol{\Pi}_t)} \lim_{K \to \infty} \left\{ \sum_{k=k_0}^{k_0+K} \gamma^k \sigma_k \right\} \tag{6}$$

where $\boldsymbol{x}_{k_0}$ represents the initial condition at k_0, and K denotes the control horizon of the MDP. σ_k denotes the traffic delay at each stage k. γ denotes the discount factor that facilitates the convergence of the decision process. We establish an acyclic signal plan by updating the control policy $\boldsymbol{\mu}_k$ at each stage k based on the traffic state $\boldsymbol{x}_k$. The action space $\mathcal{A}$ of the control policy $\boldsymbol{\mu}_{k_0}^*$ is subject to the all-red time.

The computational effort in seeking the optimal signal $\boldsymbol{\mu}_{k_0}^*$ for stage k_0 will grow exponentially as the road network expands. A common approach is to approximate the unknown value function with a surrogate model $\tilde{Z}$ parameterized by $\boldsymbol{w}$. However, given the constantly changing conditions in real traffic, a surrogate model that heavily relies on training with historical data may struggle to accurately predict future cost-to-go based on real-time traffic states. Hence, we propose a decentralized signal control via rollout RL-based approach, where explicit traffic models are incorporated to derive short-term traffic states and costs based on model-based simulations. We define $\mu_{k(n)}$ as the local control setting at each node n during stage k. Building upon the principle of decentralization discussed in Su *et al.* (2021), the optimal control policy $\mu_{k_0(n_0)}^*$ for node n_0 can be computed as follows:

$$\begin{aligned}\boldsymbol{\mu}_{k_0(n_0)}^* =& \arg \min_{\boldsymbol{\mu}_{k_0(n_0)} \in \mathcal{A}_{n_0}} \mathbb{E}_{\boldsymbol{x}_{k+1} \sim P(\boldsymbol{x}_k, \boldsymbol{\mu}_k | \boldsymbol{\pi}_t)} \\ & \left[\sum_{k=k_0}^{k_0+K'-1} \gamma^{k-k_0} \hat{\sigma}_{k,n_0} + \sum_{k=k_0}^{k_0+K'-1} \sum_{\forall n \neq n_0} \gamma^{k-k_0} \hat{\sigma}_{k,n} + \gamma^{K'} \tilde{Z}(\hat{\boldsymbol{x}}_{k_0+K'}, \boldsymbol{w}) \right]\end{aligned} \tag{7}$$

where the first term inside the expectation denotes the minimum traffic delay for the local node n_0 within the planning horizon K', evaluated across all potential control policies $\mu_k(n_0)$ within the action space $\mathcal{A}_{n_0}$. The second term represents the traffic delays for the remaining nodes within the network, assumed to be unaffected by $\mu_{k_0}(n_0)$ with their control policies fixed. The surrogate function $\tilde{Z}$ approximates the network delay beyond stage $k_0 + K'$ based on $\hat{\boldsymbol{x}}_{k_0+K'}$, which ensures that global system costs are considered when evaluating the local control policies $\mu^*_{k_0(n)}$. By asynchronously rolling out the model-based component at each node iteratively, the computational complexity can be reduced significantly compared to that of evaluating the control policies by simulating the entire road network from a centralized perspective (Chow *et al.*, 2019). In this study, we use an artificial neural network (ANN) to parameterize the surrogate function $\tilde{Z}$. A temporal difference (TD) learning method is utilized to specify the set of parameters $\boldsymbol{w}$ defining the surrogate $\tilde{Z}$ before applying to real-world application (Bertsekas, 2019).

4 NUMERICAL EXPERIMENTS

We assess the efficacy of the designed control paradigm using a grid network with nine nodes. The external environment is simulated through the SUMO interface. The proposed controllers that utilize the point queue model and the spatial queue model are denoted as “PQ+ANN" and “SQ+ANN", respectively. We adopt the standard Deep Q-Network (DQN) method as a benchmark, which represents a purely data-driven RL-based controller.

We conduct a comparative analysis of traffic performance under scenarios with and without the data-driven value function approximation as shown in Figure 1. Without the ANN-based surrogate model, signal plans are determined by evaluating all possible control policies within the planning horizon K', without considering future costs beyond this horizon. As the demand increases, a noticeable gap in traffic delay emerges between the pure point queue model and the pure spatial queue model. This can be attributed to the spatial queue model’s consideration of physical queues and road capacity, which enables it to reflect real-world traffic conditions under oversaturated conditions more accurately. We further compare the proposed PQ+ANN and SQ+ANN controllers. We can note that the controllers with the surrogate function outperform their counterparts under all demand settings. This is because the ANN approximator acts as a central coordinator, enhancing coordination between neighboring junctions. Moreover, it is notable that the gap between the point queue model and the spatial queue model is minimized by incorporating the ANN approximator. This demonstrates that the proposed rollout RL controller can effectively mitigate the model-plant mismatch induced by modeling errors and further validates the efficacy of the proposed modeling paradigm. To further analyze the network performance of different controllers under oversaturated conditions, we present the ratio of links with queue spillback for various controllers under 120% of nominal demand, as shown in Figure 2. Compared to the DQN controller, the proposed PQ+ANN and SQ+ANN controllers signifi-

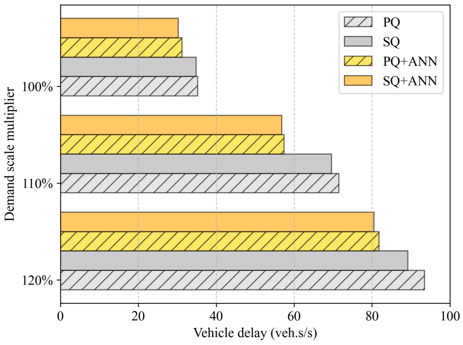

Figure 1 – *Comparison of solution methods with and without ANN approximation.*

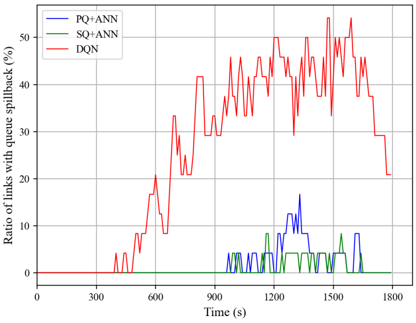

Figure 2 – *Ratio of links with queue spillback*

cantly reduce the occurrence of queue spillbacks, thereby improving network performance. This suggests that the surrogate function can act as a central coordinator, enhancing coordination between neighboring junctions and thereby mitigating long queue accumulations. Moreover, the ratio in the SQ+ANN controller is half of the PQ+ANN controller. This is because the spatial queue model accounts for vehicle length, allowing the physical queue to provide more reliable traffic state estimates for the surrogate function. To conclude, the modeling paradigm can effectively reduce queue spillbacks with the support of the model-based rollout component, without requiring sophisticated model calibrations.

5 DISCUSSION

In this paper, we develop a novel modeling paradigm for RL-based adaptive network traffic control that integrates a model-based rollout component with a data-driven surrogate function. The adoption of point queue and spatial queue models allows for explicit representation of network flow propagations, enabling more reliable estimations of traffic states and system costs through short-term rollout simulations. The decentralized mechanism introduced in our framework ensures efficient optimization of signal plans at individual intersections in an asynchronous manner.

The effectiveness of the proposed control paradigm is validated through experiments on two urban networks. The results demonstrate that the rollout RL-based framework significantly reduces network delays and their associated variability compared to a purely data-driven DQN algorithm. This suggests that the proposed rollout RL approach is robust across different traffic models without requiring complex model calibration. Specifically, signal coordinations can be achieved by the acyclic signal plans, which facilitate the traffic flow efficiency and therefore reduce the possibility for queue spillbacks.

References

Bertsekas, Dimitri. 2019. *Reinforcement learning and optimal control.* Vol. 1. Athena Scientific.

Chow, Andy HF, Sha, Rui, & Li, Ying. 2019. Adaptive control strategies for urban network traffic via a decentralized approach with user-optimal routing. *IEEE Transactions on Intelligent Transportation Systems*, **21**(4), 1697–1704.

Su, ZC, Chow, Andy HF, & Zhong, RX. 2021. Adaptive network traffic control with an integrated model-based and data-driven approach and a decentralised solution method. *Transportation Research Part C: Emerging Technologies*, **128**, 103154.

Sutton, Richard S. 2018. Reinforcement learning: an introduction. *A Bradford Book.*

Zhang, HM, Nie, Yu, & Qian, Zhen. 2013. Modelling network flow with and without link interactions: the cases of point queue, spatial queue and cell transmission model. *Transportmetrica B: Transport Dynamics*, **1**(1), 33–51.