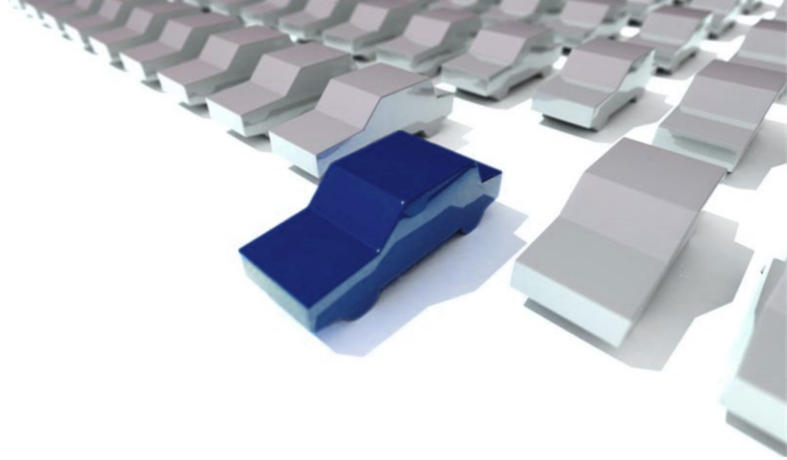


Learning the max pressure control for urban traffic networks considering the phase switching loss

Xingmin Wang

Ph.D. candidate, Civil and Environmental Engineering

University of Michigan, Ann Arbor

INFORMS 2022 Indianapolis

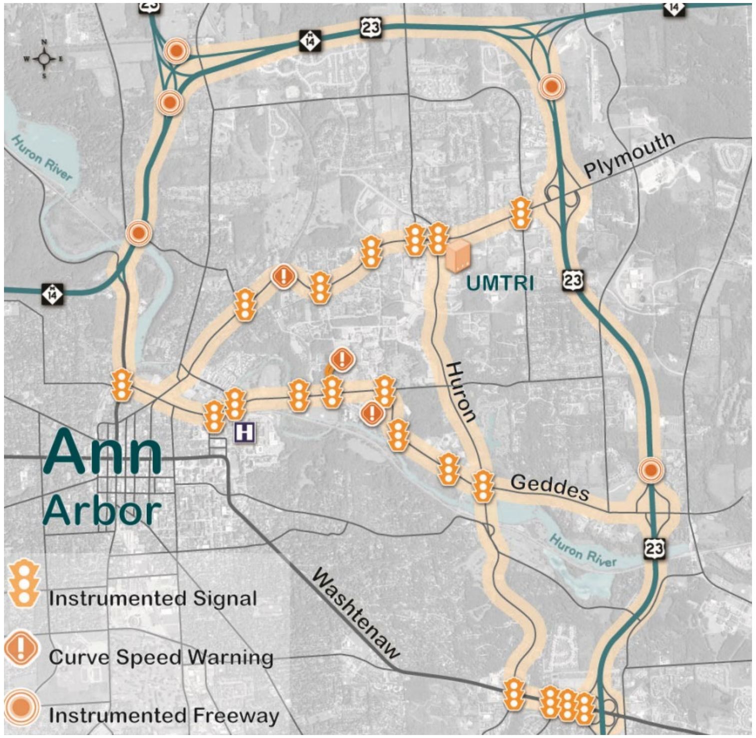
MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Outline

- Background: traffic signal control
- Introduction to the max pressure control
- Switching-curve-based max pressure control
- Max pressure control and reinforcement learning
- Simulation studies
- Conclusions

Background

- Traffic congestions
  - In 2017, traffic congestion caused urban Americans to travel extra 8.8 billion hours and to purchase extra 3.3 billion gallons of fuel*
- Traffic signal control
  - Fixed timing plan
  - Vehicle-actuated control
  - Adaptive control



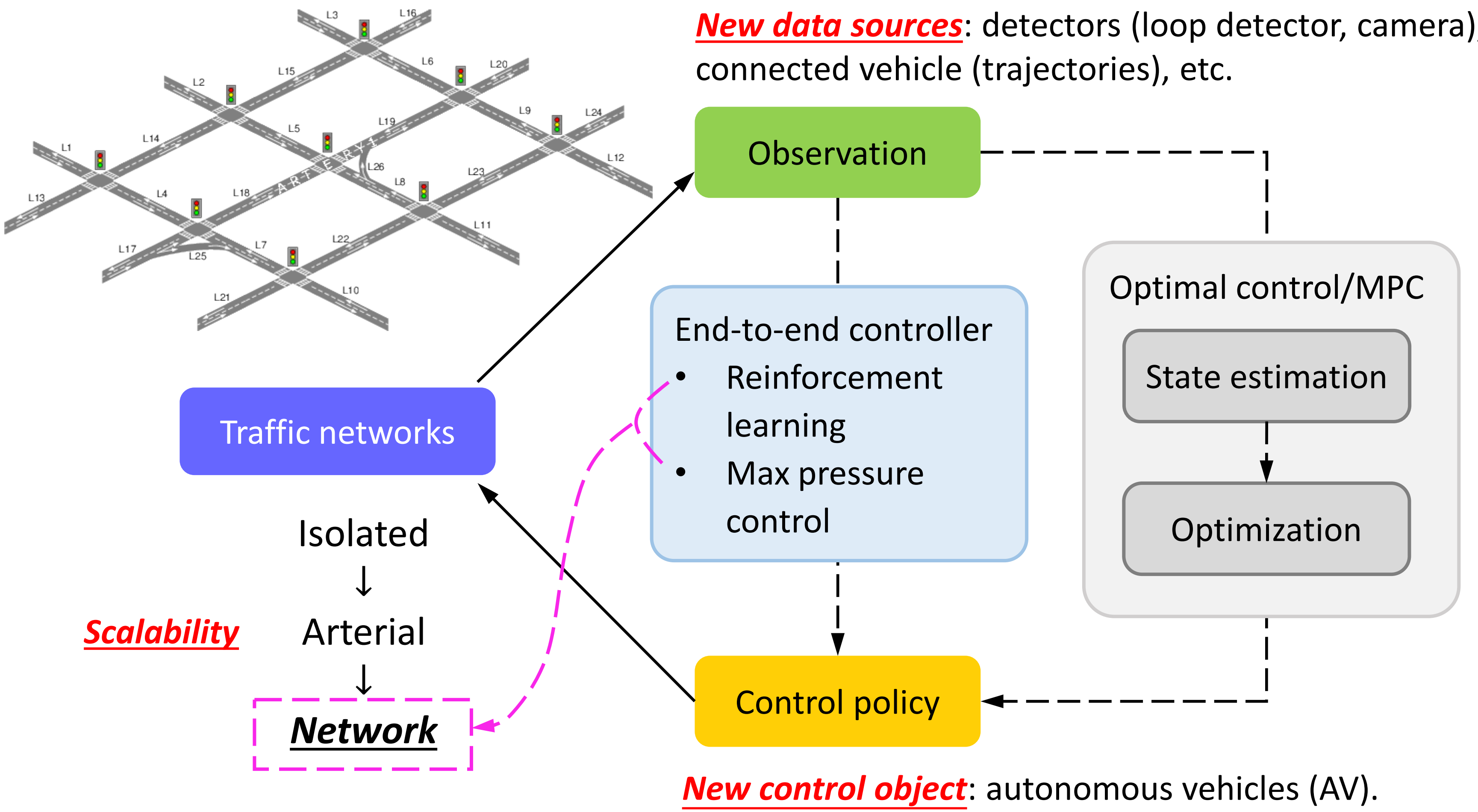
Ann Arbor connected vehicle test environment

- The development of connected and autonomous vehicles (CAV) provides new opportunities for the traffic signal control

*: Lasley, Phil. "2019 Urban mobility report." (2019).

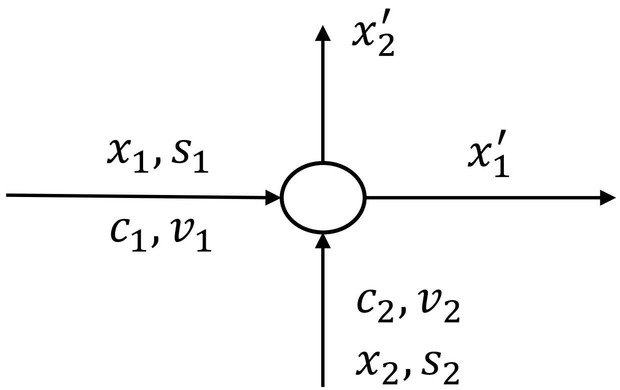
Opportunities and challenges for traffic signal control with CAV

- From the classic control diagram

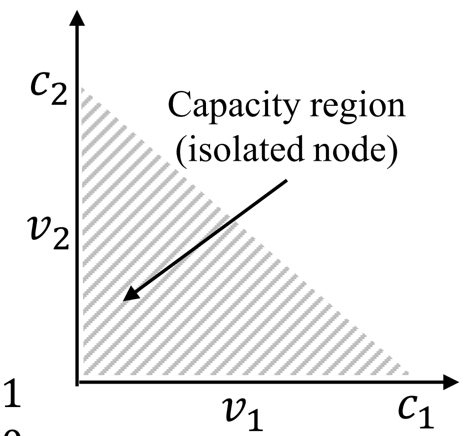


Introduction to the max pressure control

- Max pressure control for general urban traffic networks
  - For each movement, the ***pressure*** is defined as the upstream queue lengths minus the downstream queue lengths times the saturation flow
  - Max pressure control (MP): each intersection always chooses the phase with the largest pressure



MP: $(x_1 - x_1')c_1 > (x_2 - x_2')c_2 \rightarrow \begin{cases} s_1 = 1 \\ s_2 = 0 \end{cases}$



x	Queue length
v	Traffic volume
c	Saturation flow
s	Signal state (0,1)

$s_1 + s_2 \leq 1$

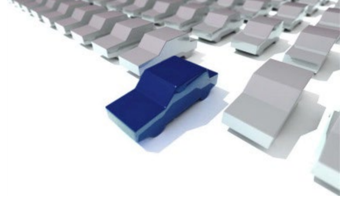
  - It can be proved that, under the *store-and-forward model*, the max pressure control can ***stabilize*** the network queue lengths if the traffic demand is within the network capacity (***throughput-optimal***)

Varaiya, Pravin. "Max pressure control of a network of signalized intersections." *Transportation Research Part C: Emerging Technologies* 36 (2013)

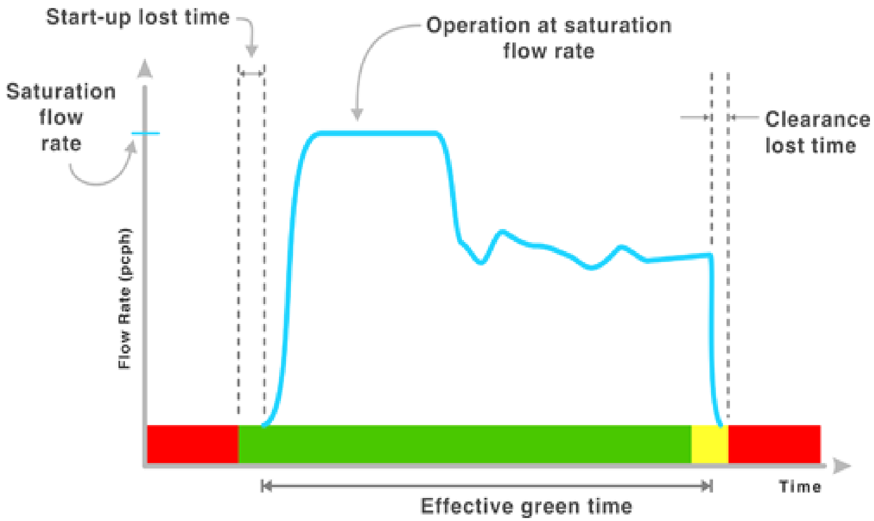
Advantages and limitations

- What we like about the max pressure control
  - ***Distributed***: each intersection decides its own control policy
  - ***End-to-end***: the control policy can be directly generated from the observation
  - Therefore, the max pressure control is easy to implement even for a large-scale traffic network
  - ***Global stability***: it can be proved to stabilize the global network under the store-and-forward model
- What we concern about
  - Most of the concerns are with regard to the assumptions of the store-and-forward model (compared with the real-world traffic)
    - No link travel time → bad coordination among intersections
    - Infinite link capacity → link spill-over or gridlock
    - No ***switching loss*** → over-saturation (undermine the stability)

Phase switching loss



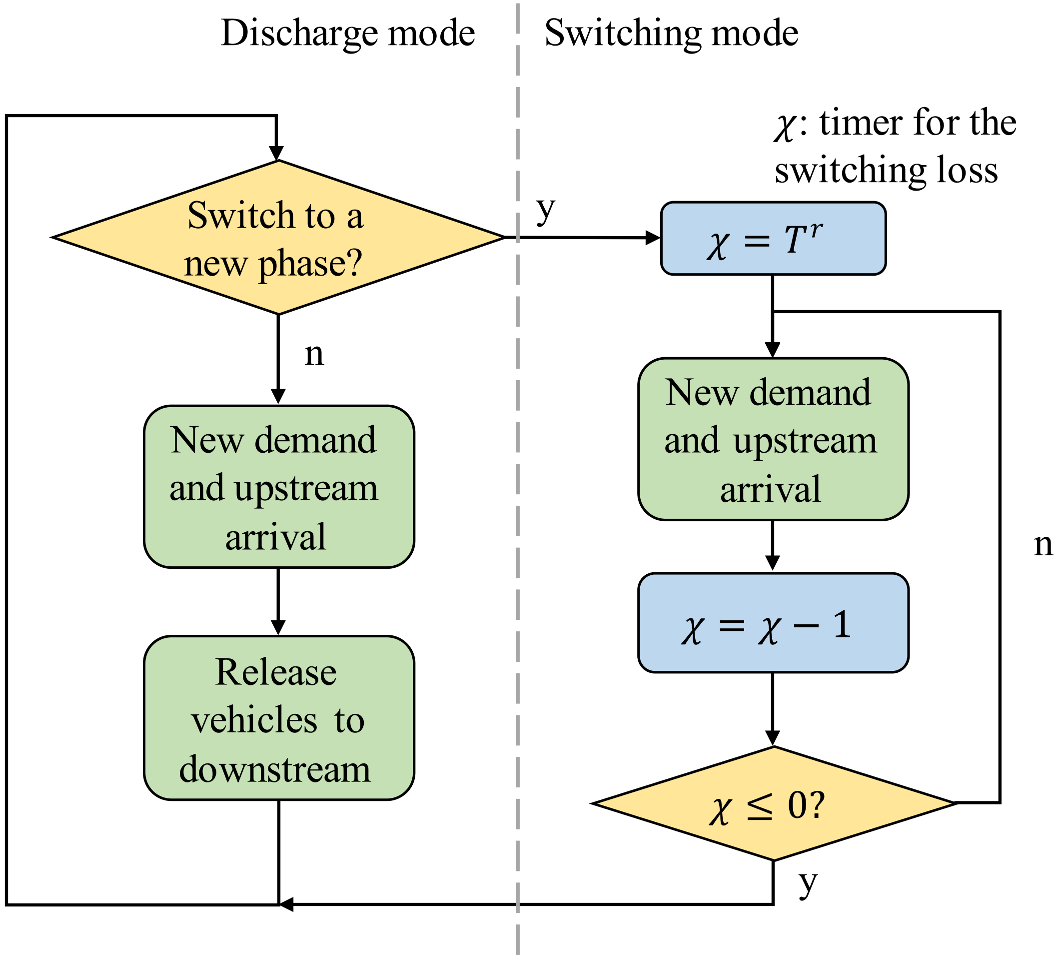
- Phase switching loss
  - Start-up loss and clearance time
  - Higher traffic volume → larger cycle length
- Max pressure control and phase switching loss
  - There is no cyclic structure in the max pressure control but the phase switching frequency reflects the cycle length



- To ensure the network stability, the switching frequency of the max pressure control should decrease with the increase of the traffic volumes.
- However, the conventional max pressure control does not consider this factor

Modeling the phase switching loss in the store-and-forward model

- With the phase switching loss, the vehicle can only pass the intersections during the discharge mode

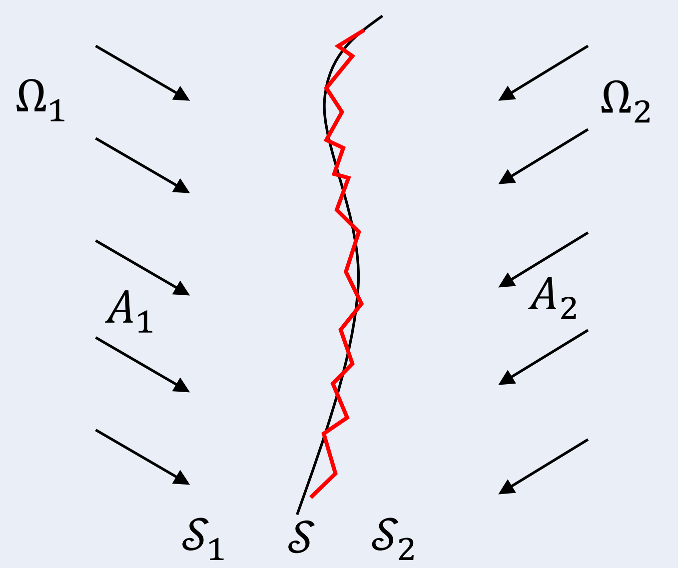


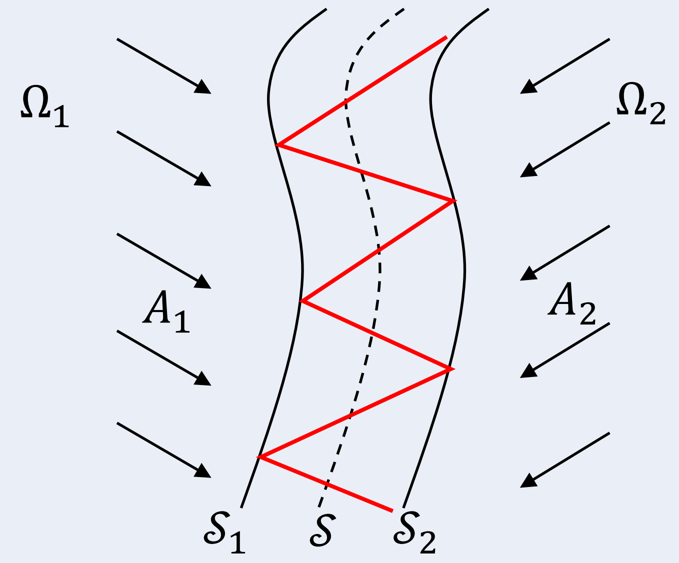
- The original max pressure control is no longer throughput-optimal with this new system dynamics
- The system is no longer a Markov chain only with the state representation $\boldsymbol{x}$. We need to augment the signal state to maintain the Markovian property of the dynamic system

Control policy design considering the switching loss: hysteresis switching

- State-dependent switching and hysteresis switching

Without switching loss	With switching loss
State-dependent switching	Hysteresis switching





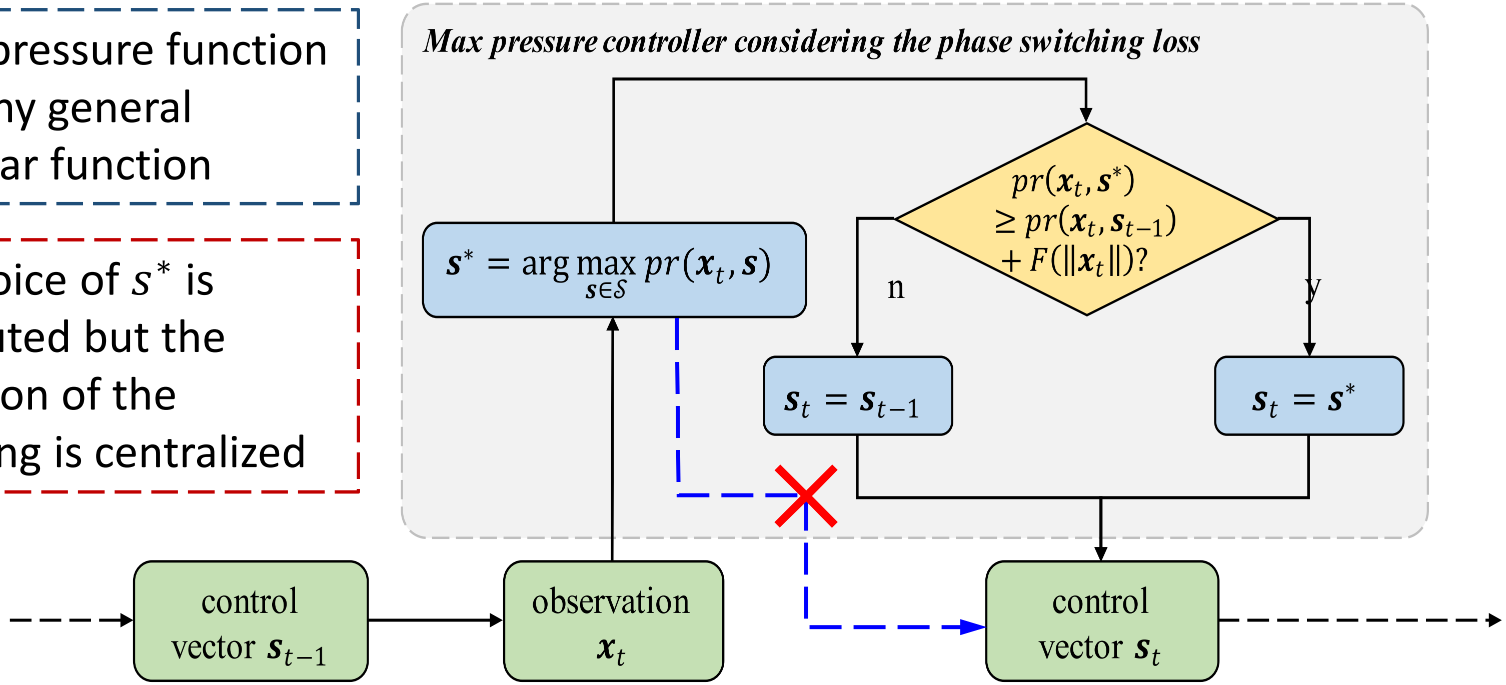
The zigzag of the trajectories without the switching loss would disappear if the system is continuous and there is no disturbance

Proposed switching-curve-based max pressure

- Switching-Curve-based Max Pressure control (SCMP)
  - Intuition (*hysteresis switching*): instead of choosing the policy with the largest pressure for each time step, we only switch to a new control policy when the pressure is larger than the current control policy by a certain threshold

- $pr(\cdot)$: pressure function
- $F(\cdot)$: any general sublinear function

- The choice of s^* is distributed but the activation of the switching is centralized



Extension of SCMP

- Extend SCMP for practical implementation
  - Although SCMP is proved to be a throughput-optimal policy, it is a centralized control policy derived based on a simplified point-queue network model
  - To further adapt SCMP to the real-world implementation, we extend SCMP to ***Extended-SCMP*** in two aspects: *distributed approximation* and use the *position-weighted pressure*
- Distributed switching
  - The switching rule in SCMP is centralized for the convenience of the proof

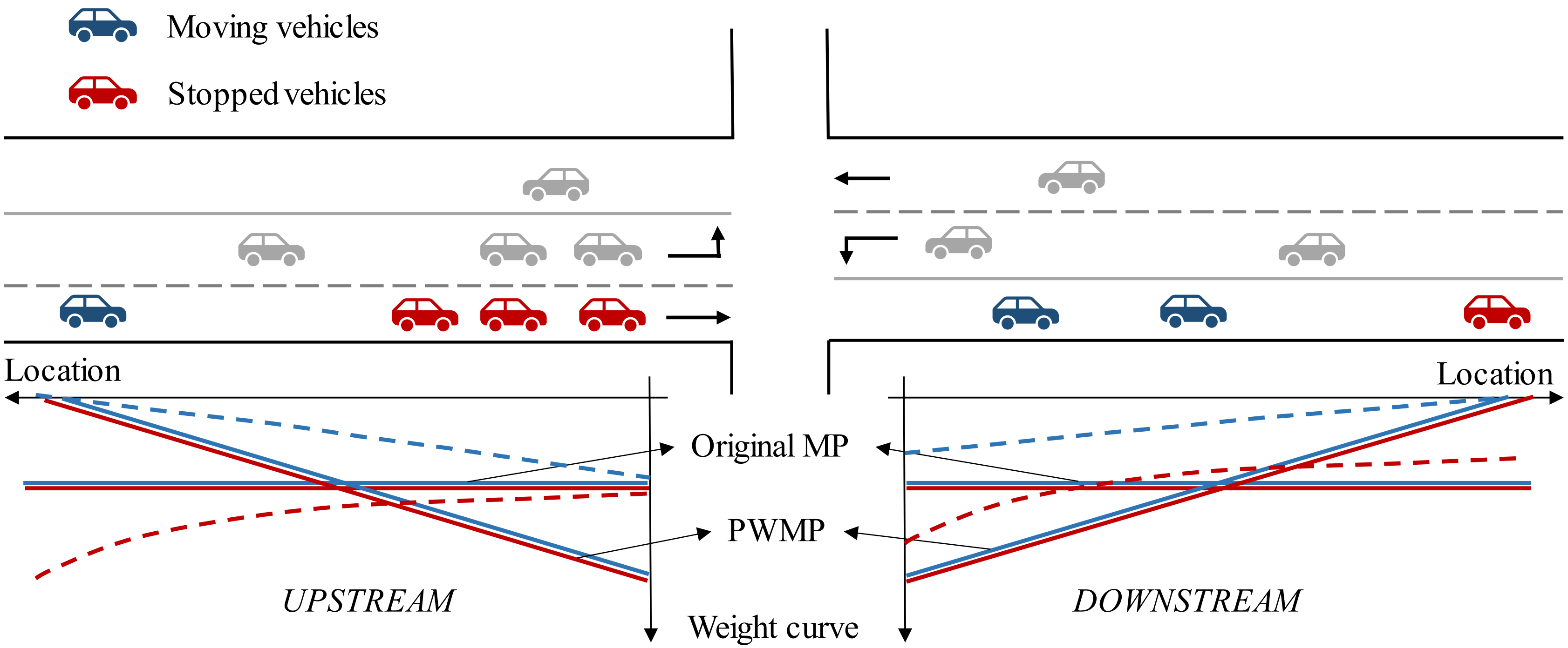
SCMP: $$\psi(t) = \max_{s \in S} \mathrm{pr}\left(\boldsymbol{x}_t, \boldsymbol{s}\right) - \mathrm{pr}\left(\boldsymbol{x}_t, \boldsymbol{s}_{t-1}\right) - F\left(\|\boldsymbol{x}_t\|\right) \geq 0$$

- Distributed approximation in which each intersection decides its own decision

ESCMP: $$\psi^n(t) = \max_{s^n \in S^n} \mathrm{pr}\left(\boldsymbol{x}_t^n, \boldsymbol{s}^n\right) - \mathrm{pr}\left(\boldsymbol{x}_t^n, \boldsymbol{s}_{t-1}^n\right) - F\left(\|\boldsymbol{x}_t^n\|\right) \geq 0$$

Extension of SCMP (cont'd)

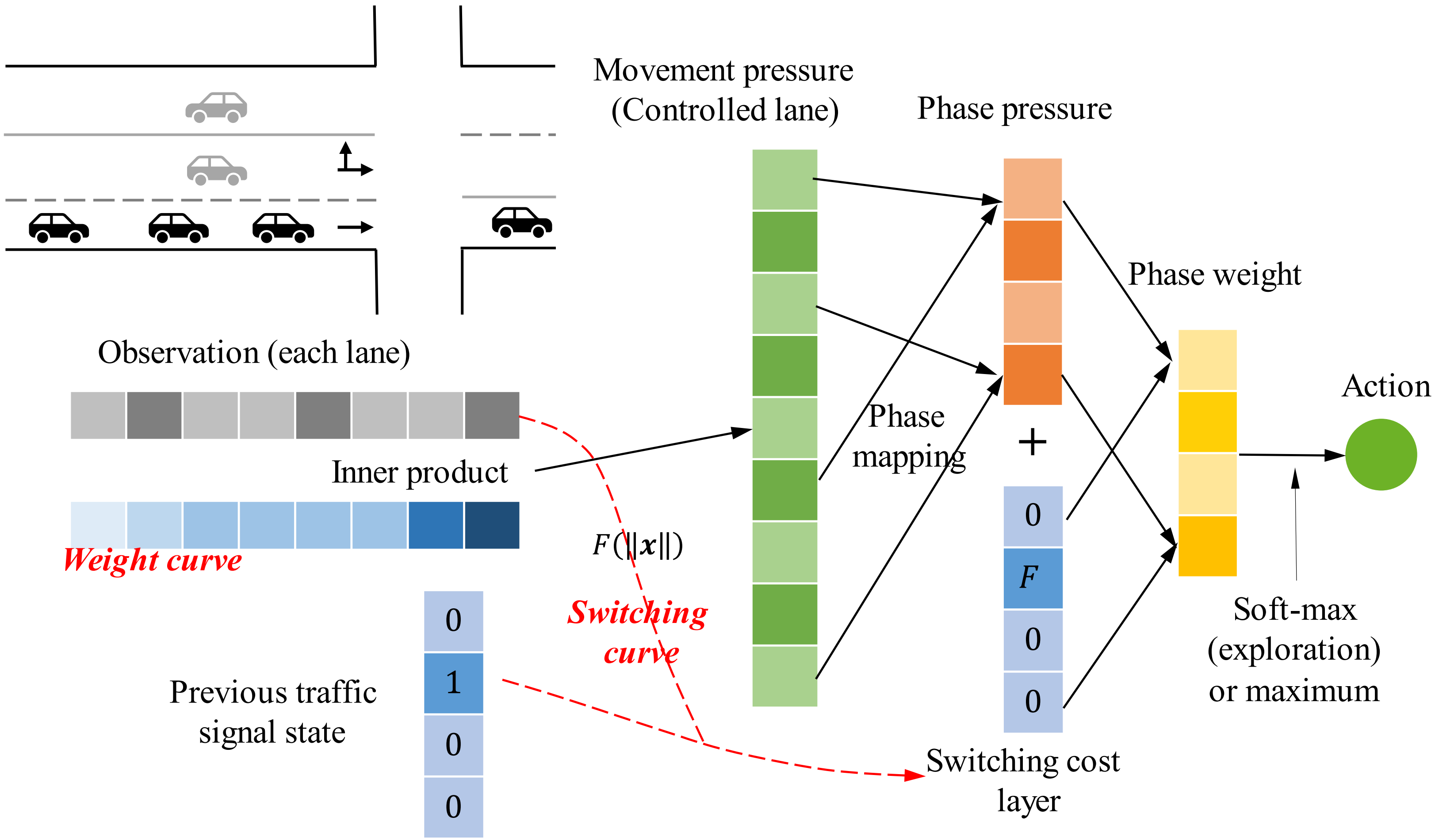
- Position-weighted pressure
  - Intuition: vehicle at different locations along the road might exert different influence on the signalized intersection



Li, Li, and Saif Eddin Jabari. "Position weighted backpressure intersection control for urban networks." *Transportation Research Part B: Methodological* 128 (2019): 435-461.

Overall max pressure policy network

- Max pressure control policy network with the switching curve and the weight curve (ESCMP)



Parameter optimization of the max pressure control using the reinforcement learning

- Parameter optimization using reinforcement learning
  - Based on the proposed max pressure policy network with the ***switching curve*** and ***the weight curve***, we can use the reinforcement learning to optimize these two parametric curves to get a better system performance
- Introduction to the policy-gradient reinforcement learning
  - System state trajectory: $\tau = [s_0, a_0, s_1, a_1, \dots]$, reward $r(\tau)$
  - ***Parametric policy*** $a_t \sim \pi_\theta(\cdot \mid s_t)$

$$\theta^* = \arg\max_\theta \mathbb{E}_{\tau \sim \pi_\theta} r(\tau) = \arg\max_\theta J(\theta)$$

  - Policy gradient optimization

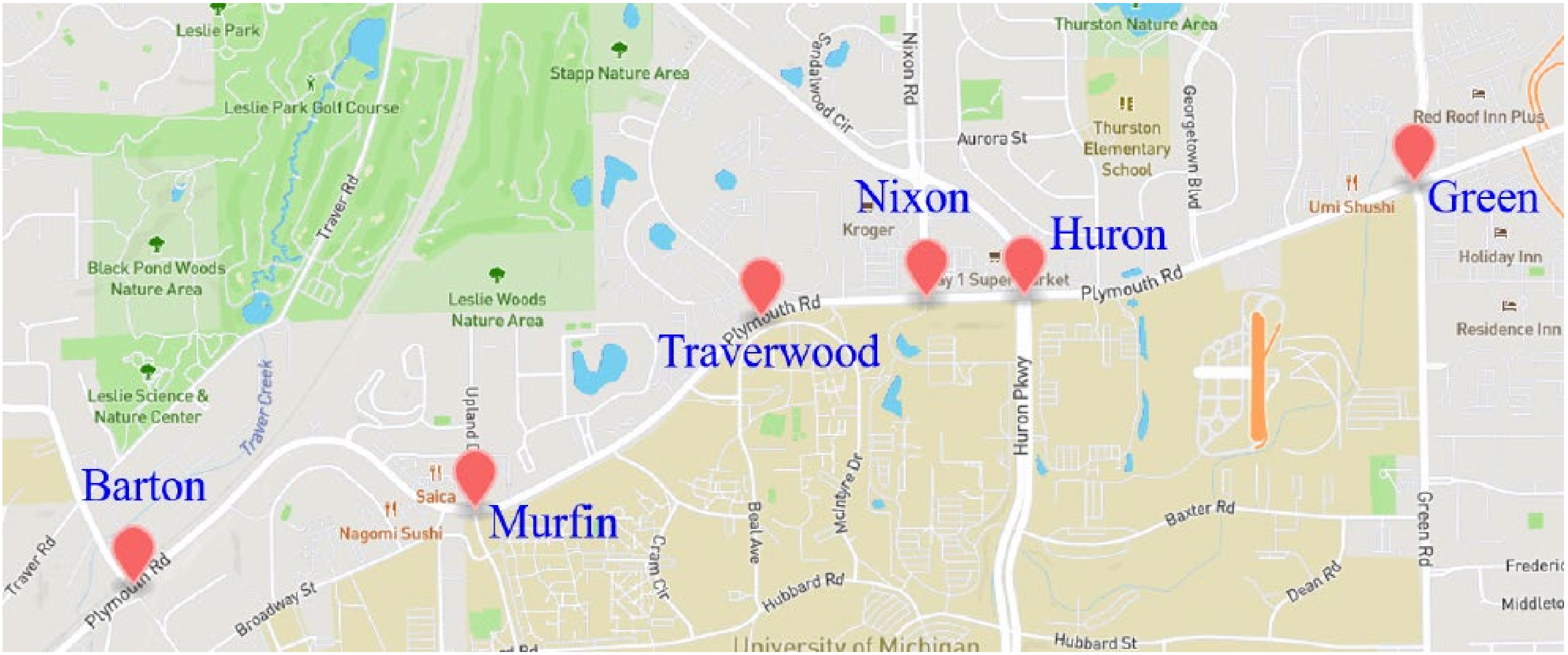
$$\theta_{k+1} = \theta_k + \alpha \nabla_\theta J(\theta) \qquad \nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}\left[r(\tau) \sum_t \pi_\theta(a_t \mid s_t) \right]$$

Why utilizing RL further?

- The throughput-optimal policy only ensures the system stability (bounded total queue lengths); it does not ensure the optimal total system delay
- The max pressure control is derived based on a simplified point-queue model, RL can help to tune the parameters based on a more realistic simulation environment

Simulation studies

- Setup of the simulation studies
  - Simulation platform: SUMO
  - Chosen network topology: Plymouth Rd., Ann Arbor
  - Traffic demand: calibrated during the peak hour

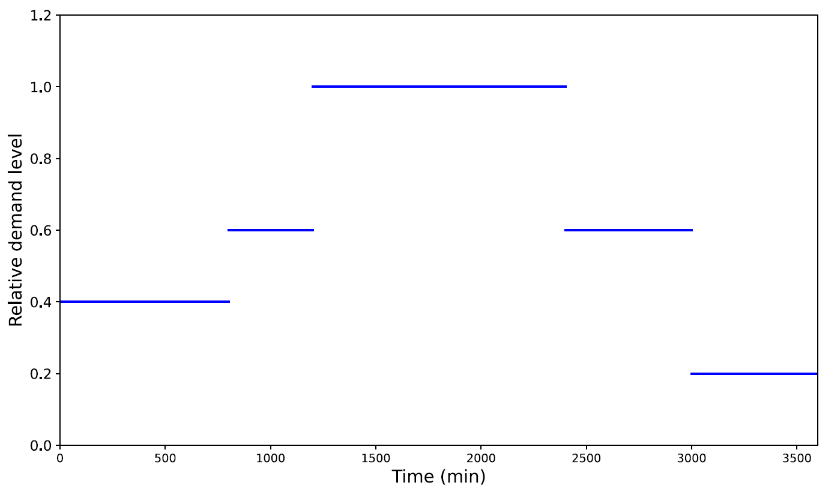


Performance evaluation: experimental settings

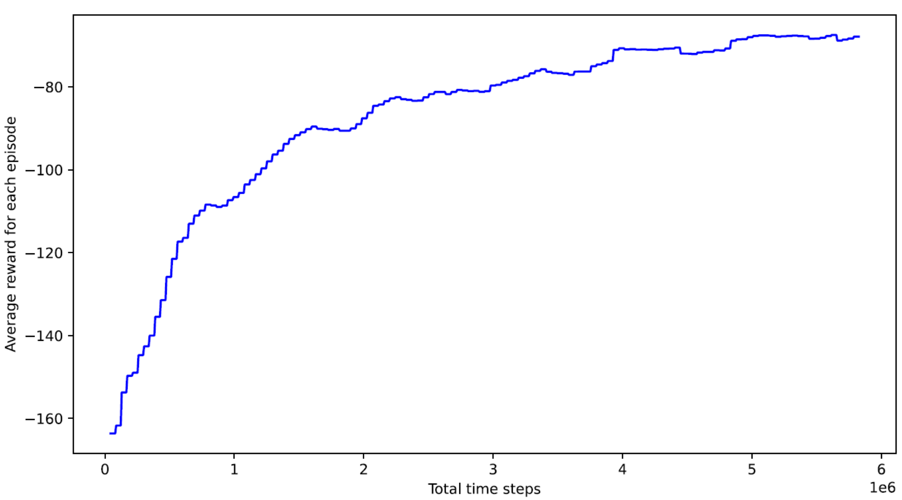
- Three controllers are tested using the simulation platform
  - PWBP (position-weighted back pressure control): benchmark max pressure controller does not consider the phase switching loss
  - ESCMP (Extended-SCMP): extended switching-curve-based max pressure control
  - LESCMP (Learned-ESCMP): ESCMP in which the parameters are optimized using the policy-gradient RL algorithms
- Two test scenarios
  - Time-varying demand scenario
  - Stationary demand scenario: using different demand levels
- Two metrics: network stop delay and throughput

Max pressure control parameters optimization using the reinforcement learning

- Configuration for the reinforcement learning
  - Algorithms: proximal policy optimization (PPO)
  - ***Policy network***: the proposed max pressure policy network with the switching curve and weight curve
  - Value network (critic): fully-connected neural network
  - State: number of vehicle in each cell and the current signal state
  - Reward (cost): network stop delay

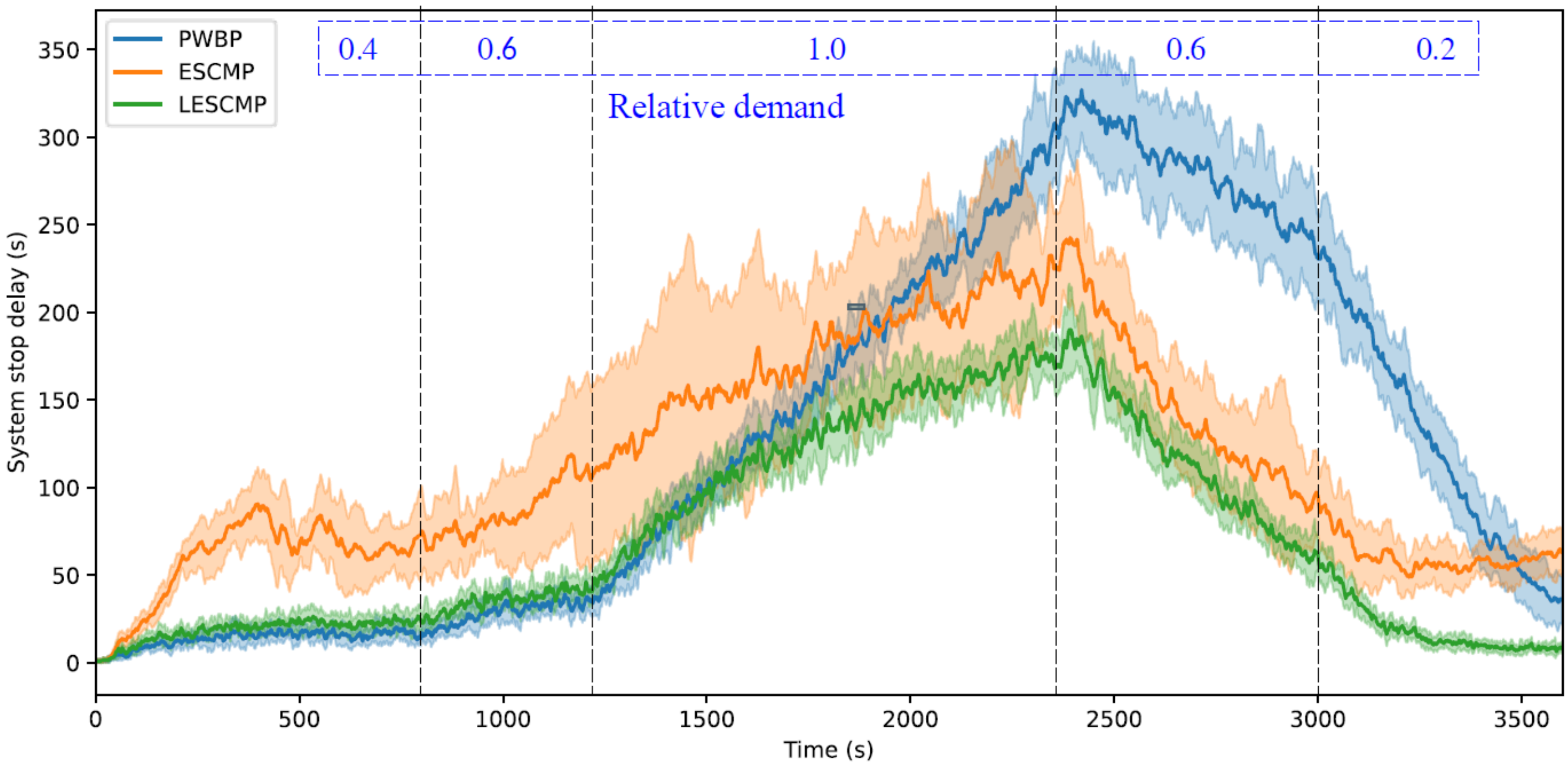


Input demand profile



Training curve

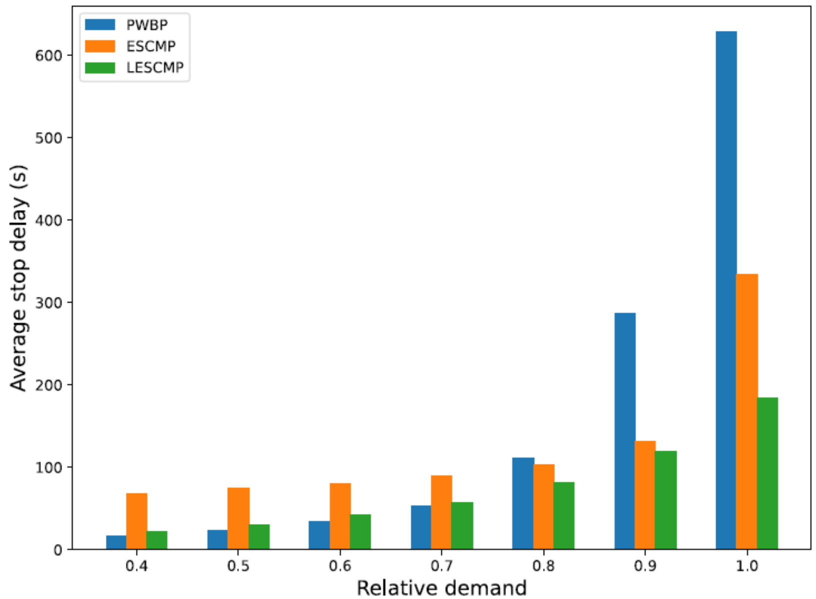
Performance evaluation: time-varying demand



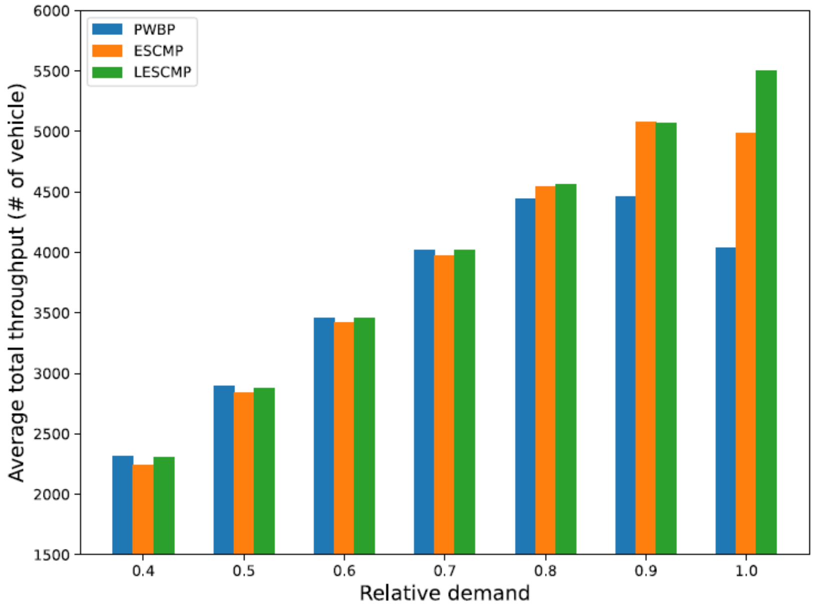
Control policy	Total delay (h)	Delay std (h)	Total throughput (veh)	Throughput std (veh)
PWBP	129.15	8.8	3664.5	25.02
ESCMP	115.79	17.44	3659.9	10.74
LESCMP	72.15	2.63	3714.7	9.2

Performance evaluation: stationary demand

- Summary for the numerical results
  - PWBP has a good low-demand performance but a bad high-demand performance
  - ESCMP and LESCMP performs much better than PWBP under the high-demand scenario
  - LESCMP performs well in all levels of demand



(a) Average total delay



(b) Average total throughput

Summary

- We propose a switching-curve-based max pressure (SCMP) control that is ***proved to be throughput-optimal*** over the store-and-forward model with phase switching loss
- We extend SCMP by utilizing a distributed switching approximation and the position-weighted pressure (Extended-SCMP)
- With the max pressure policy network (ESCMP), the policy-gradient reinforcement learning is further utilized to optimize the parameters in the controller
- The simulation studies show that both ESCMP and LESCMP have better high-demand performance than the conventional max pressure control and LESCMP performs well in all demand levels

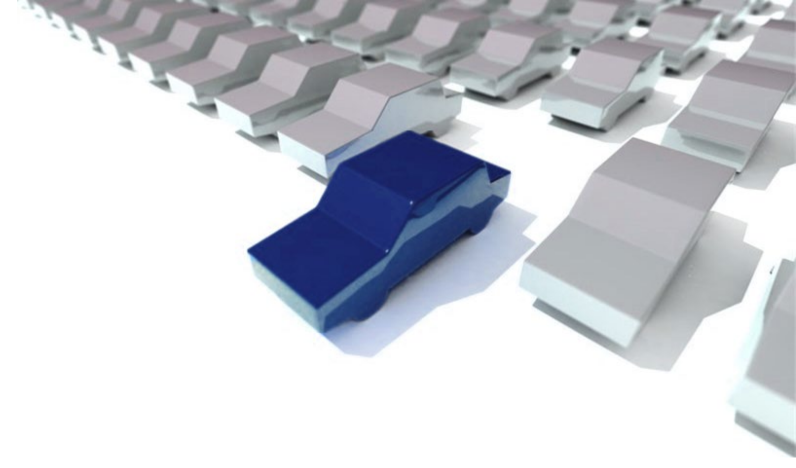
- Practical significance: the proposed control policy suits well for the real-world implementation especially for the ***large-scale network*** since it is ***distributed*** among intersections and directly generates the control policy from the observation (***end-to-end***)

Acknowledgement

- Research project sponsored by

CENTER FOR CONNECTED AND
AUTOMATED TRANSPORTATION

- Collaborators
  - Dr. Henry Liu, Dr. Yafeng Yin, and Dr. Yiheng Feng
- Presented work published in:
  - **Wang, X.,** Yin, Y., Feng, Y. and Liu, H.X., 2022. Learning the max pressure control for urban traffic networks considering the phase switching loss. *Transportation Research Part C: Emerging Technologies*, *140*, p.103670.

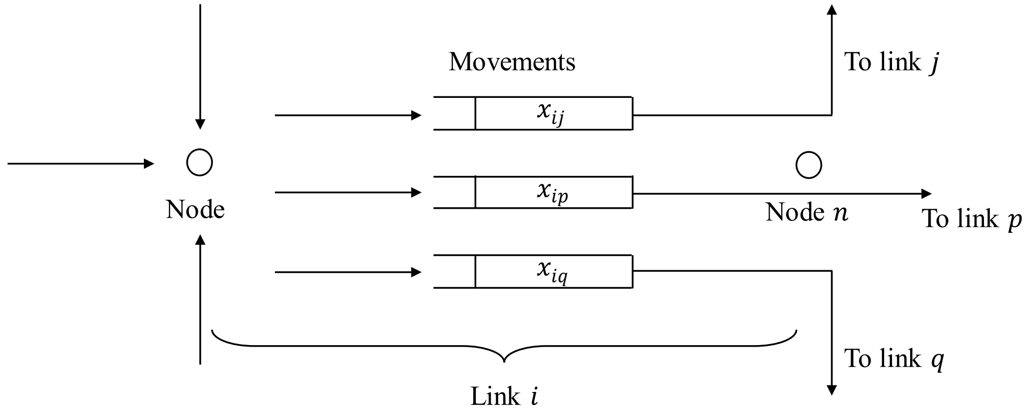


Thank you!

Contact: xingminw@umich.edu

Store-and-forward network model

- The max pressure control is derived based on the store-and-forward model (point-queue network model)



$$x_{ij}(t+1) = x_{ij}(t) + a_{ij}(t) + \sum_k r_{ij}(t)\min\{x_{ki}(t), c_{ki}s_{ki}(t)\} - \min\{x_{ij}(t), c_{ij}s_{ij}(t)\}, \forall m = (i,j) \in \mathcal{M}^o$$

$$\boldsymbol{x}(t+1) = \boldsymbol{x}(t) + \boldsymbol{a}(t) - (\boldsymbol{I} - \boldsymbol{R}) \cdot \min\{\boldsymbol{x}(t), \boldsymbol{C}\boldsymbol{s}(t)\},$$

x	Queue length	c	Saturation flow
a	Exogenous arrival	s	Signal state (0,1)
r	Turning ratio		

Stability of the max pressure control

- The max pressure control is essentially the *Lyapunov drift minimization policy* under the store-and-forward model

  - Lyapunov function: $V(t) = \frac{1}{2}\boldsymbol{x}(t)^T\boldsymbol{x}(t) = \frac{1}{2}\sum_{m\in\mathcal{M}} x_m^2$

    (Bounded Lyapunov function → bounded total queue lengths)

  - Lyapunov drift: $\Delta(t) = \mathbb{E}[V(t+1) - V(t) \mid \boldsymbol{x}(t)]$

    Substitute the dynamics equation

    $$\boldsymbol{x}(t+1) = \boldsymbol{x}(t) + \boldsymbol{a}(t) - (\boldsymbol{I} - \boldsymbol{R}) \cdot \min\{\boldsymbol{x}(t), \boldsymbol{Cs}(t)\}$$

  - Upper bound of the drift: $\Delta(t) \le B - \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{Cs}(t)$

  - Minimize the upper bound → max pressure control

    $$\min B - \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{Cs}(t) \rightarrow \underbrace{\max \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{Cs}(t)}_{\text{Max pressure control}}$$

- Proof of the stability: Foster-Lyapunov theorem

Stability of SCMP

- Sufficient conditions for network stability (main theorem)
  - Given a policy that always chooses the max pressure policy whenever the switching is activated, the network will be stable if the following conditions are satisfied (demand ***strictly*** within the capacity, otherwise no controller can stabilize the network)

$$\tau_{k+1} \geq \tau'_{k+1}; \qquad \text{(eq 1)}$$

$$\mathbb{E}\left[\left(\tau'_{k+1} - \tau_k\right) \middle| S_{\tau_k}\right] \geq c_1(1 - \delta'(\|\boldsymbol{x}(\tau_k)\|))F(\|\boldsymbol{x}(\tau_k)\|); \qquad \text{(eq 2)}$$

$$\mathbb{E}\left[\left(\tau'_{k+1} - \tau_k\right)^2 \middle| S_{\tau_k}\right] \leq T_r^2 + c_2\left(F(\|\boldsymbol{x}(\tau_k)\|)\right)^2; \qquad \text{(eq 3)}$$

$$\mathbb{E}\left[L(\boldsymbol{x}(t+1)) - L(\boldsymbol{x}(t)) \middle| S_t\right] \leq -\epsilon\|\boldsymbol{w}(\boldsymbol{x}(t))\|, \quad \forall \boldsymbol{x}(t) \in C^o, t \in \left\{\tau'_{k+1}, \tau'_{k+1}+1, ..., \tau_{k+1}\right\}; \qquad \text{(eq 4)}$$