# Learning the max pressure control for urban traffic networks

Xingmin Wang

Ph.D. student, Civil and Environmental Engineering
University of Michigan, Ann Arbor
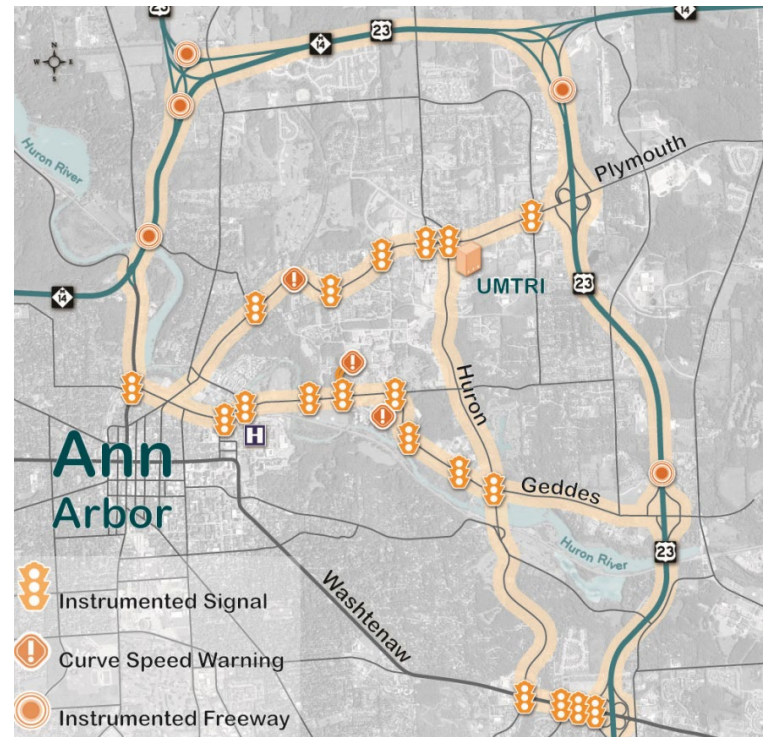
NGTS Seminar, 03/18/2021

# Outlines

- ❑ Traffic signal control: upcoming challenges

- ❑ Introduction to the max pressure control

- ❑ Switching-curve-based max pressure control

- ❑ Max pressure control and reinforcement learning

- ❑ Simulation studies

- ❑ Conclusions

# Outlines

❑ **Traffic signal control: upcoming challenges**

❑ 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

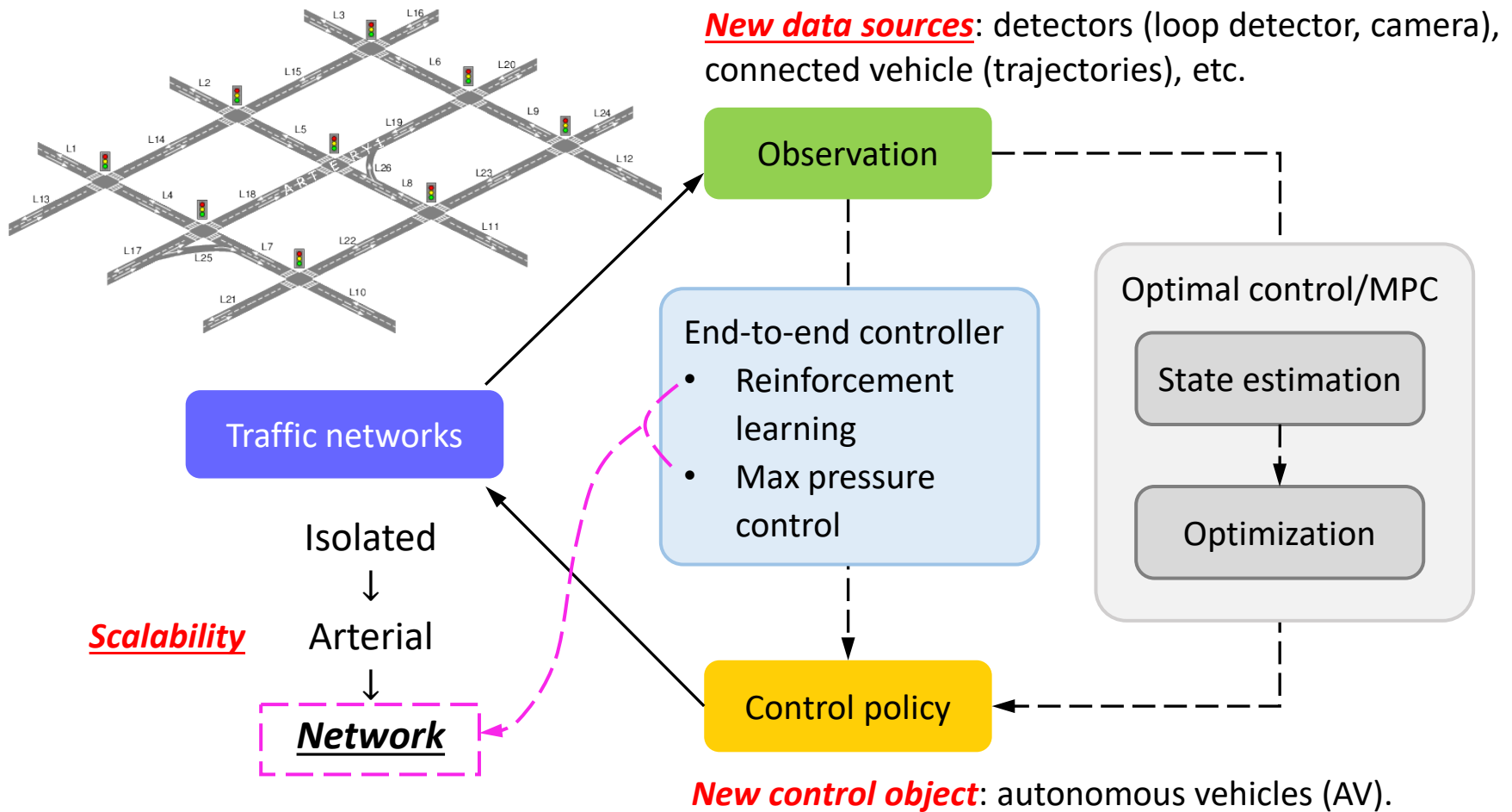  - Vehicle-actuated control

  - Adaptive control



Ann Arbor connected vehicle test environment (AACVTE, live-web)

- 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



**New data sources**: detectors (loop detector, camera), connected vehicle (trajectories), etc.

Observation

End-to-end controller
- Reinforcement learning
- Max pressure control

Optimal control/MPC
- State estimation
- Optimization

Traffic networks

Isolated
↓
**Scalability** Arterial
↓
**Network**

Control policy

**New control object**: autonomous vehicles (AV).

# Outlines

❑ Traffic signal control: upcoming challenges

❑ **Introduction to the max pressure control**

❑ Switching-curve-based max pressure control

❑ Max pressure control and reinforcement learning

❑ Simulation studies

❑ Conclusions

# 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: $(x_1 - x_1')c_1 > (x_2 - x_2')c_2 \rightarrow \begin{cases} s_1 = 1 \\ s_2 = 0 \end{cases}$
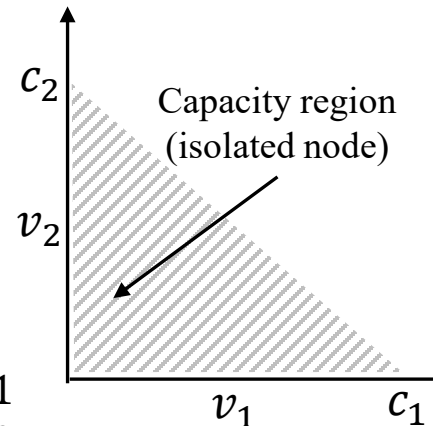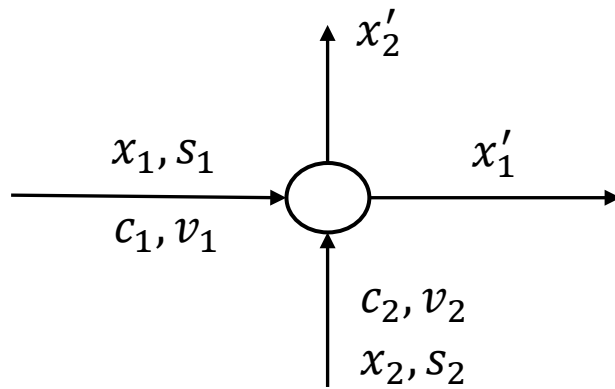
| | |
|---|---|
| $x$ | Queue length |
| $v$ | Traffic volume |
| $c$ | Saturation flow |
| $s$ | Signal state (0,1) |

$s_1 + s_2 \leq 1$

- 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)

**MICHIGAN ENGINEERING**
UNIVERSITY OF MICHIGAN

# Store-and-forward network model

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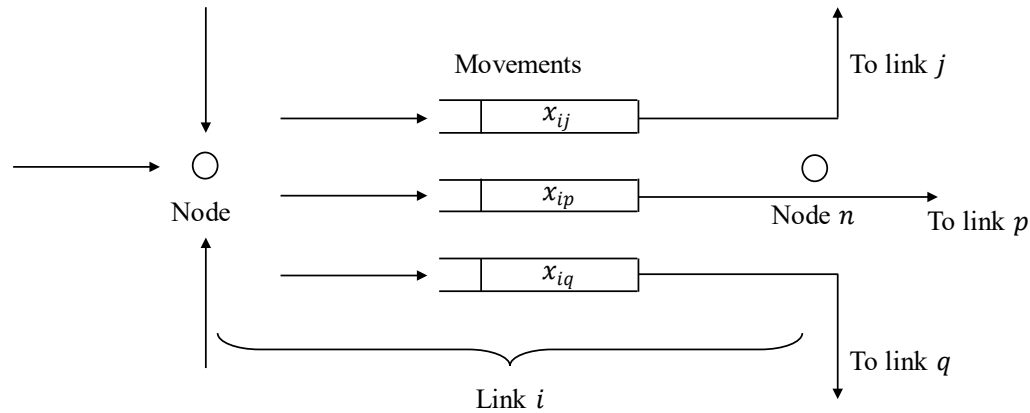


$$x_{ij}(t+1) = x_{ij}(t) + a_{ij}(t) + \sum_k r_{ij}(t) \min\{x_{ki}(t), c_{ki}s_{ki}(t)\} - \min\{x_{ij}(t), c_{ij}s_{ij}(t)\}, \forall m = (i,j) \in \mathcal{M}^o$$

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

| | | | |
|---|---|---|---|
| $x$ | 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

  o 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)

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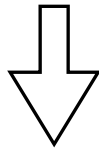
    Substitute the dynamics equation

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

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

  o Minimize the upper bound → max pressure control

    $\min B - \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{Cs}(t) \to \max \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{Cs}(t)$

    *Max pressure control*

- Proof of the stability: Foster-Lyapunov theorem

# Summary of the original max pressure control

- 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

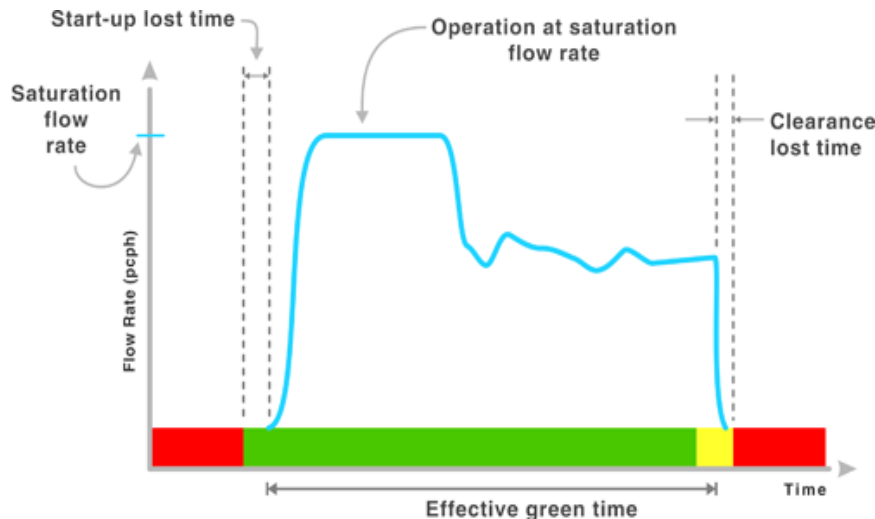    - No ***switching loss*** → over-saturation (undermine the stability)

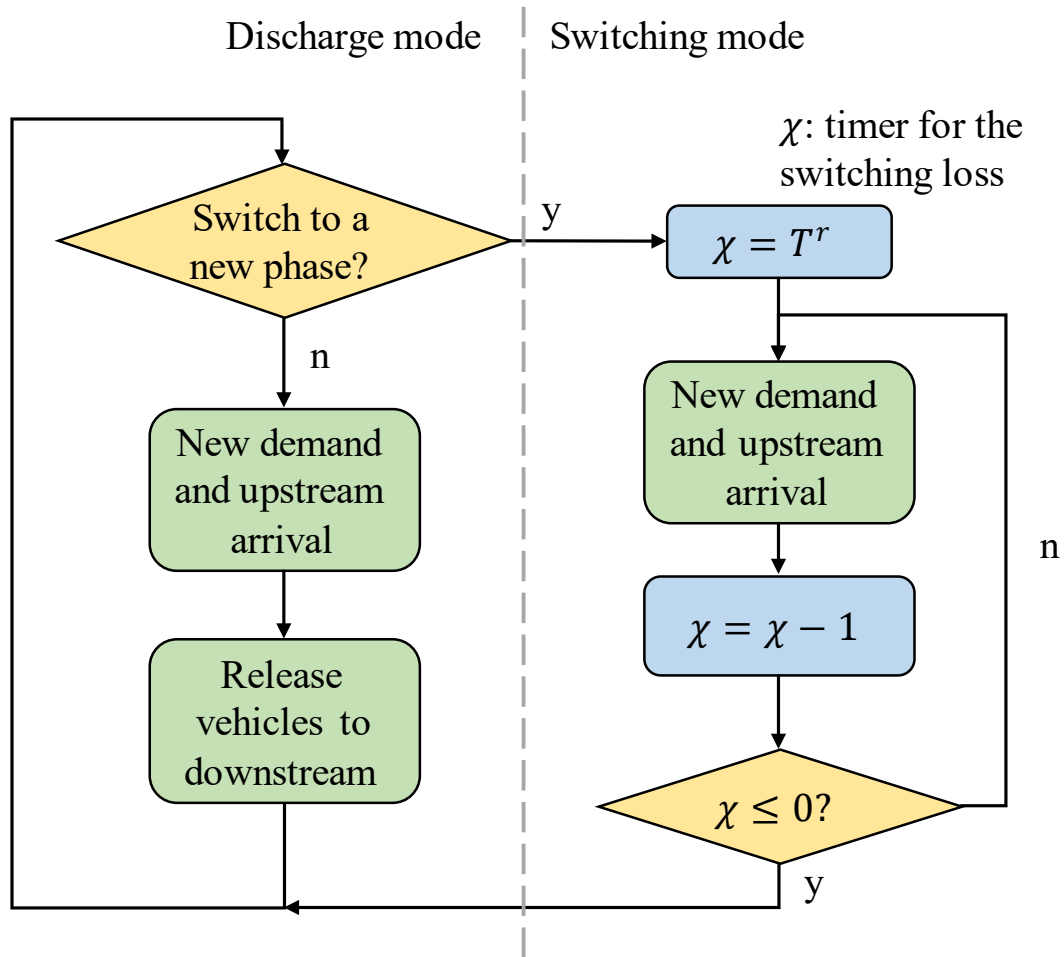# Outlines

# 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

Discharge mode | Switching mode

$\chi$: timer for the switching loss

Switch to a new phase? → y → $\chi = T^r$

n

New demand and upstream arrival

Release vehicles to downstream

New demand and upstream arrival

$\chi = \chi - 1$

$\chi \leq 0?$

n

y

- 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 $x$. We need to augment the signal state to maintain the Markovian property of the dynamic system

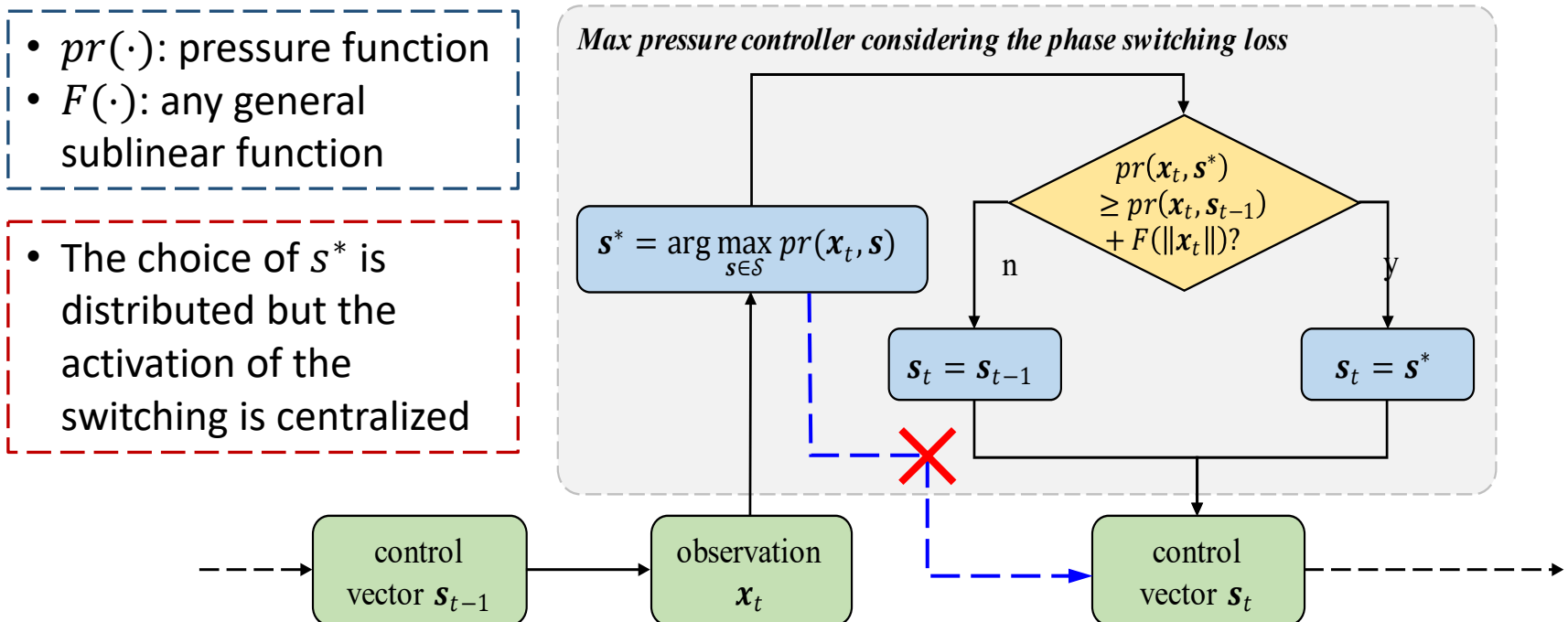# Control policy design considering the switching loss: hysteresis switching

- State-dependent switching and hysteresis switching

| Without switching loss | With switching loss |
|---|---|
| Memoryless | Memory system |
| State-dependent switching | Hysteresis switching |
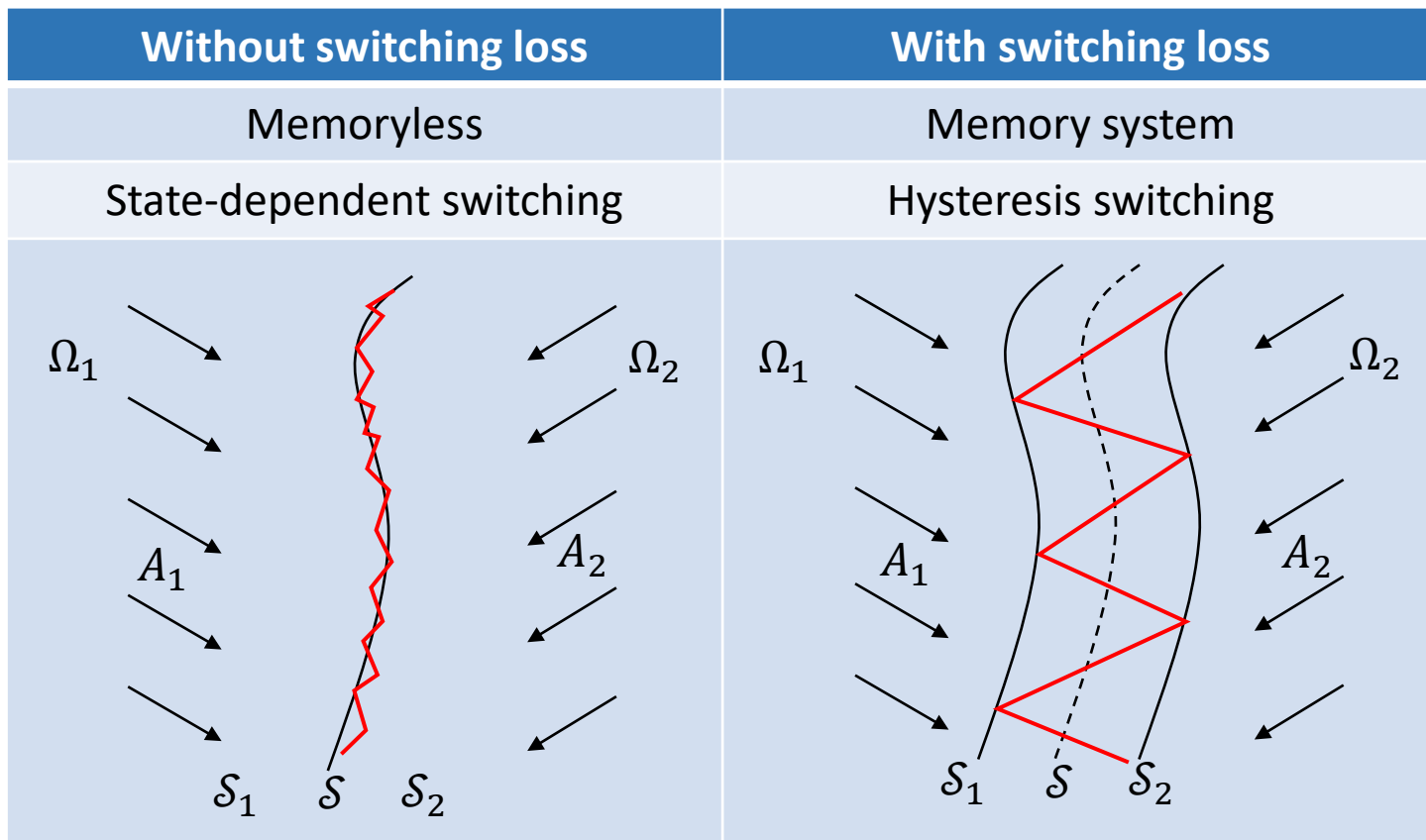


Comment: the zigzag of the trajectories without the switching loss would disappear if the system is continuous and there is no disturbance

# 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

*Max pressure controller considering the phase switching loss*

$$s^* = \arg\max_{s \in \mathcal{S}} pr(\boldsymbol{x}_t, \boldsymbol{s})$$

$$pr(\boldsymbol{x}_t, \boldsymbol{s}^*) \geq pr(\boldsymbol{x}_t, \boldsymbol{s}_{t-1}) + F(\|\boldsymbol{x}_t\|)?$$

n

y

$$\boldsymbol{s}_t = \boldsymbol{s}_{t-1}$$

$$\boldsymbol{s}_t = \boldsymbol{s}^*$$

control vector $\boldsymbol{s}_{t-1}$

observation $\boldsymbol{x}_t$

control vector $\boldsymbol{s}_t$
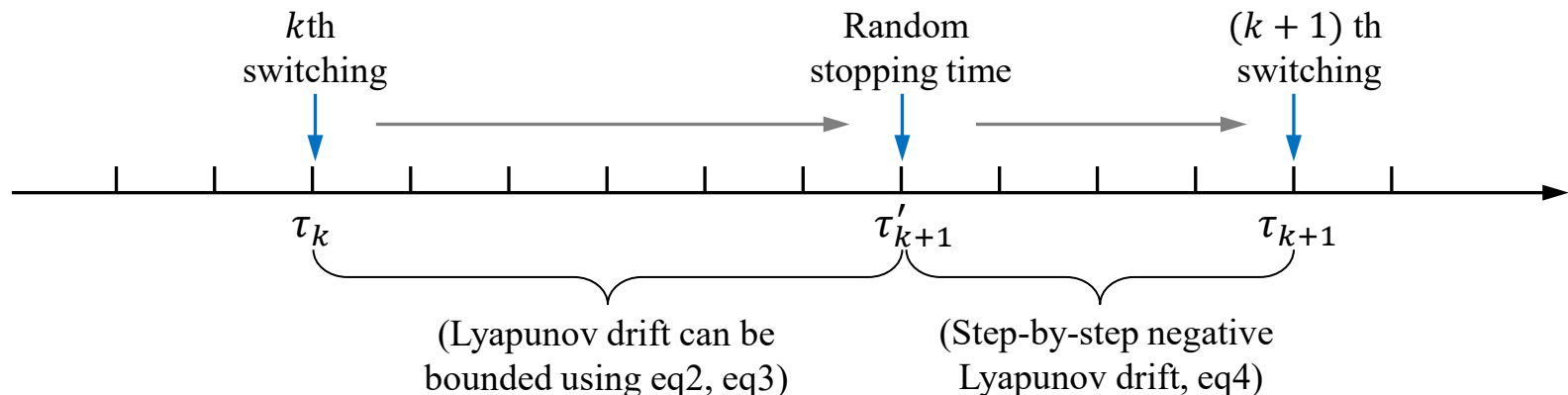
# Stability of SCMP

- 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)

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

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

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

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



$k$th
switching

Random
stopping time

$(k+1)$ th
switching

$\tau_k$

$\tau'_{k+1}$

$\tau_{k+1}$

(Lyapunov drift can be
bounded using eq2, eq3)

(Step-by-step negative
Lyapunov drift, eq4)

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

- Distributed approximation in which each intersection decides its own decision

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

# Extension of SCMP (cont'd)

- Position-weighted pressure

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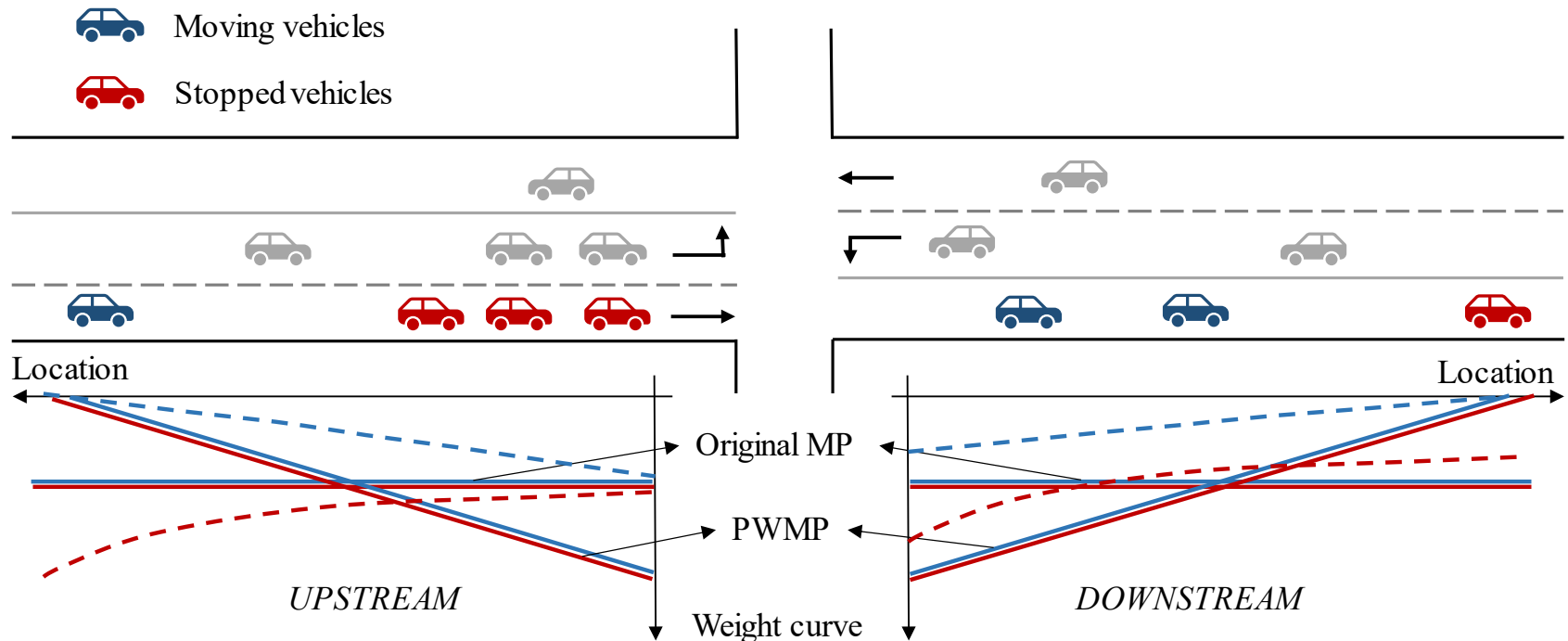


Li, Li, and Saif Eddin Jabari. "Position weighted backpressure intersection control for urban networks." *Transportation Research Part B: Methodological* 128 (2019): 435-461.
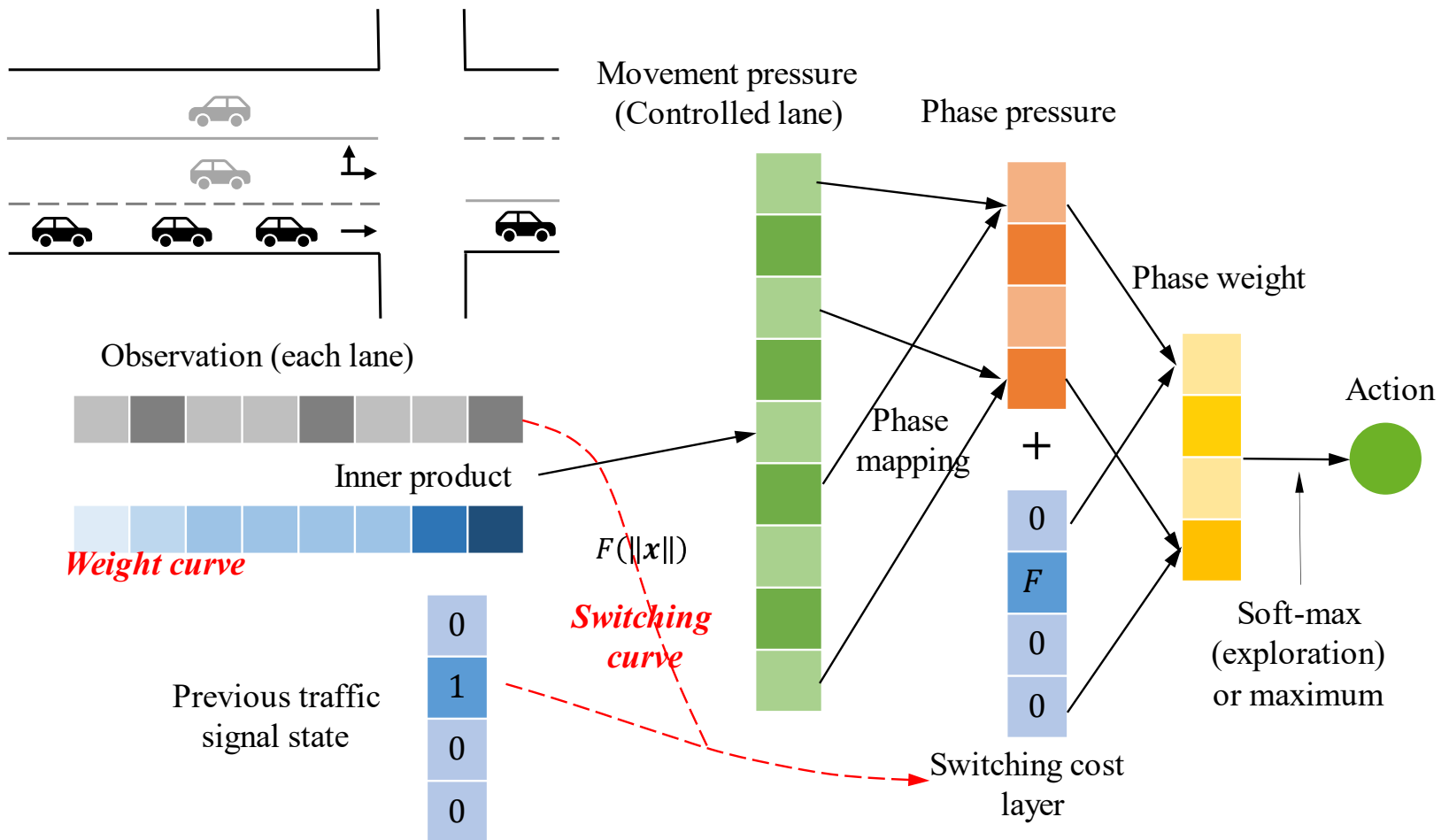
# Outlines

# 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: $\tau = [s_0, a_0, s_1, a_1, \ldots]$, 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

- It would be quite difficult to directly analyze the delay performance or replace the point-queue model with a more realistic traffic model using the analytical methods (stability analysis, etc.)

- The proposed policy-gradient RL with the max pressure policy network combines the strengths of the analytical methods and the data-driven RL methods

# Outlines

❑ Traffic signal control: upcoming challenges

❑ Introduction to the max pressure control

❑ Switching-curve-based max pressure control

❑ Max pressure control and reinforcement learning

❑ **Simulation studies**

❑ Conclusions

# 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
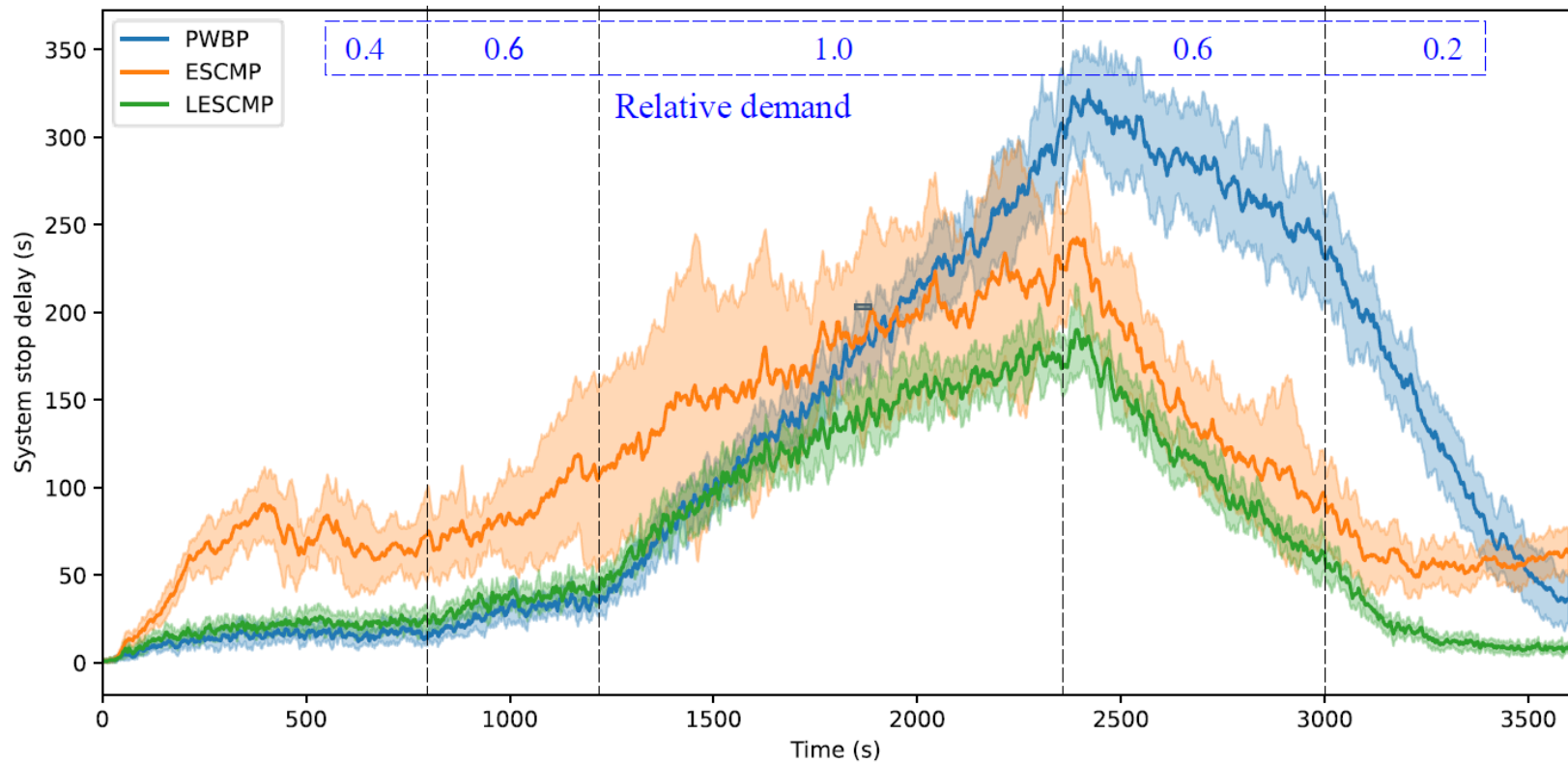
  - Reward (cost): network stop delay
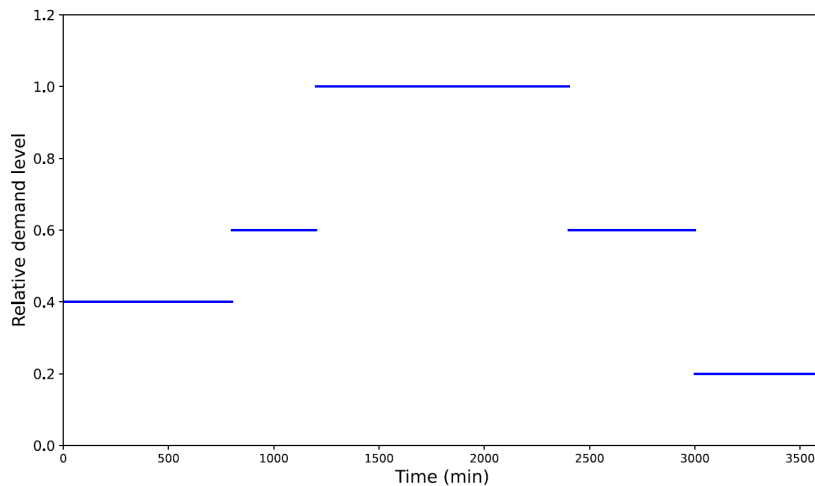


Input demand profile
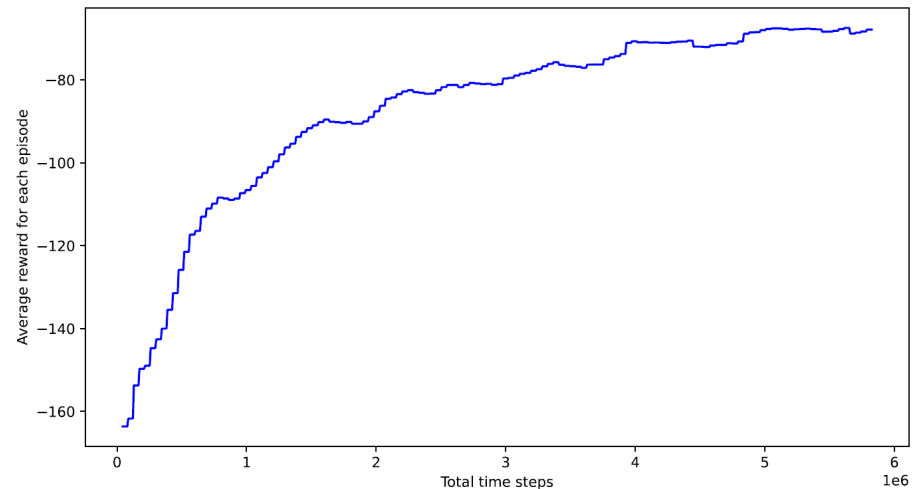
Training curve

| 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
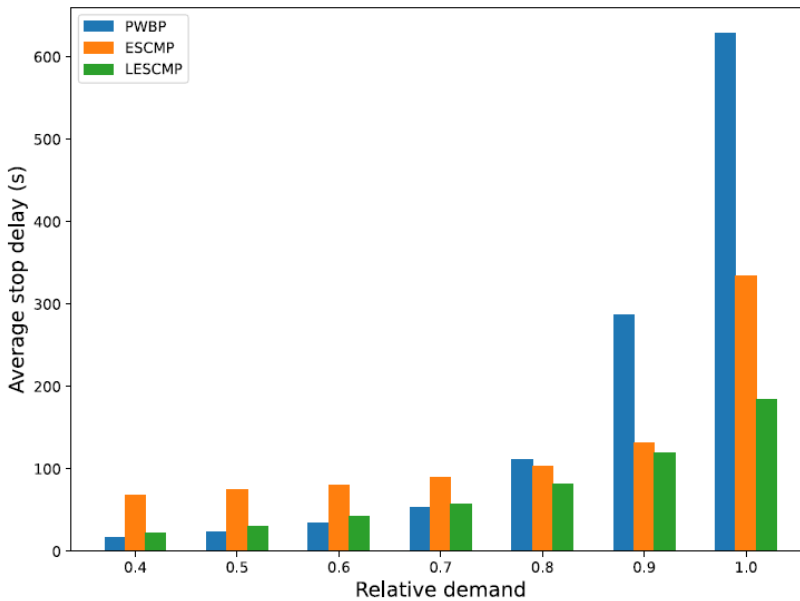


(a) Average total delay
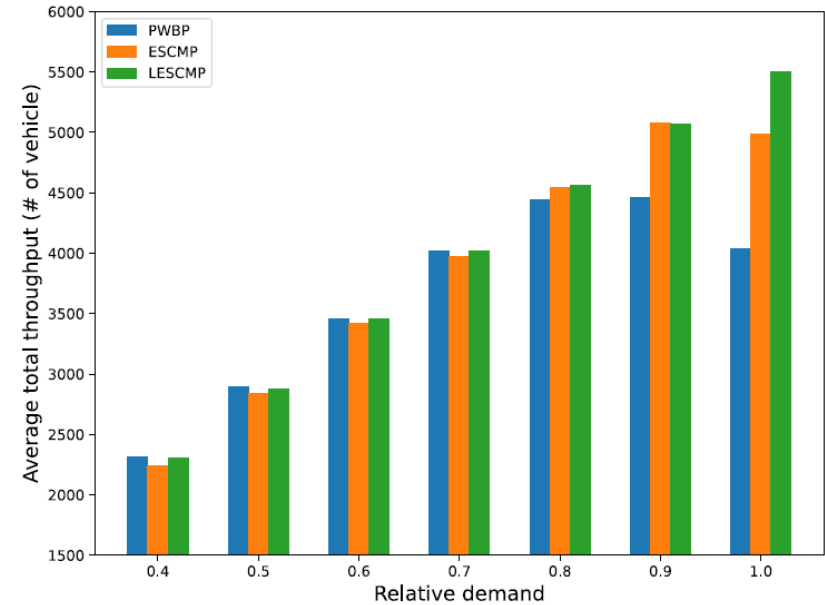


(b) Average total throughput

# Outlines

❑ Traffic signal control: upcoming challenges

❑ Introduction to the max pressure control

❑ Switching-curve-based max pressure control

❑ Max pressure control and reinforcement learning

❑ Simulation studies

❑ **Conclusions**

# 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



- Collaborators
    - Dr. Henry Liu, Dr. Yafeng Yin, and Dr. Yiheng Feng
    - Plymouth Rd. environment is provided by Dr. Tian Mi

# Thank you!

Contact: [xingminw@umich.edu](mailto:xingminw@umich.edu)