Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control

Proceedings of the AAAI Conference on Artificial Intelligence, 2020

Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, Zhenhui Li

발표자: 여지호

Introduction

Traffic signal control using RL: Why?

- **Pre-timed control**
- **Actuated control**
- **Optimization-based control**
- **Recently, reinforcement learning**
  :
  - Traditional transportation modes + New Mobility Models

Research gap

- **City-level traffic signal control**

  : there are three issues

- **Scalability**

  : Handle large-scale traffic network

  : Thousands of traffic lights

  : Manhattan, NYC – 2800 traffic lights

  : Large-scale + global optimization goal

- **Coordination**

- **Data feasibility**

Research gap

- **City-level traffic signal control**
  : there are three issues
- **Scalability**
- **Coordination**
  : 신호연동
  : Optimizing signal timings
- **Data feasibility**
  : Feasible data source
  : Should use real-world data
- **None of the methods are applied to a city-level scenario with thousands of signals.**

Objective

- **Present a decentralized RL model to tackle the city-level traffic signal control problem**
  : satisfies 1) scalability; 2) Coordination; 3) Data feasibility

- **Decentralized RL paradigm**
  : FRAP(Zheng et al., 2019) – Phase competition
  : Add ‘pressure’ concept for coordination

Illustration the concept of pressure

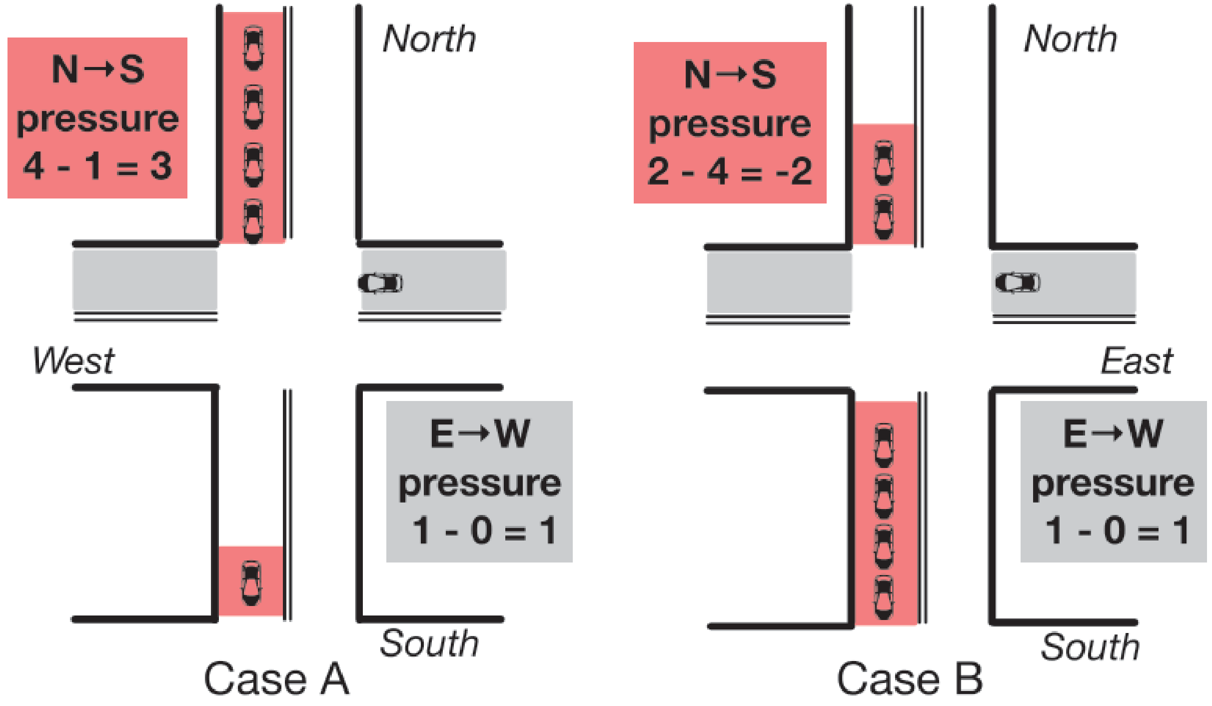


Figure 1: Illustration of max pressure control in two cases (Wei et al. 2019a). In Case A, green signal is set in the North→South direction; in Case B, green signal is set in the East→West direction.

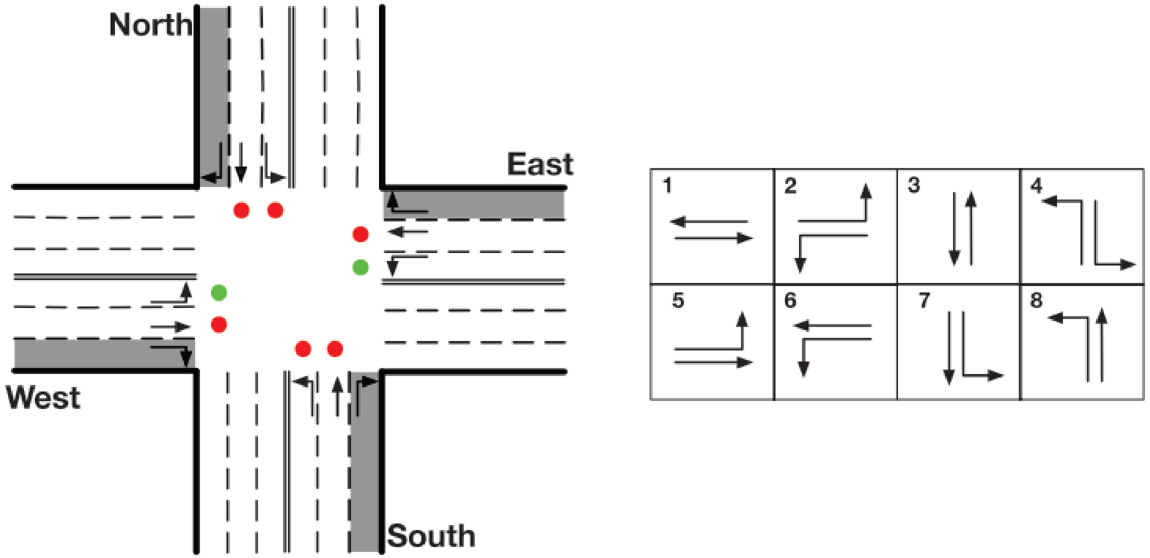
Preliminaries

Definition 1 - Traffic movement

- **Traffic travelling across an intersection from one entering lane to an exiting lane.**
- **We denote a traffic movement from road l to road m as (l,m)**

Definition 2 – Signal Phase

- A set of permissible traffic movements
- S_i denotes the set of all the phases at intersection i



(a) Intersection (b) Eight phases

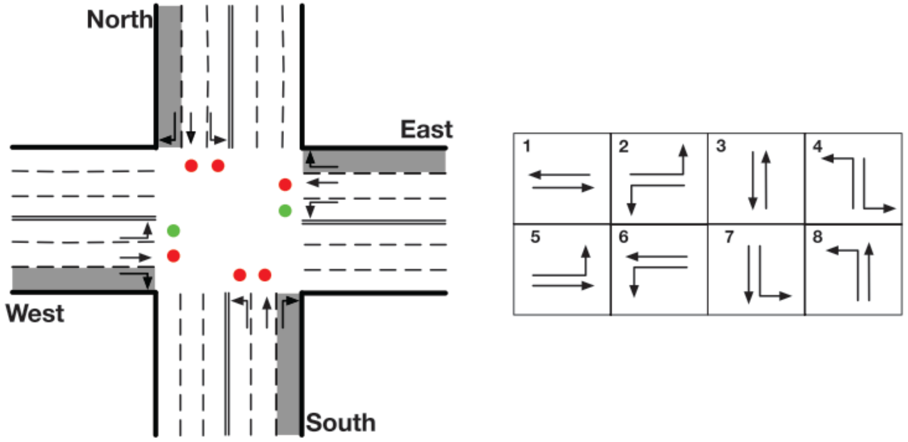
Figure 2: The illustration of an intersection with eight phases. In this case, phase **#2** is set.

Example

- **12 traffic movement**
- **8 signal phases**

Definition 3 – Pressure of each signal phase

For each signal phase s, *there are several permissible traffic movements* (l, m). *Denote by* $x(l, m)$ *the discrepancy of the number of vehicles on lane* l *and lane* m, *for traffic movement* (l, m), *the pressure of a signal phase* $p(s)$ *is simply the total sum of the pressure of its permissable phases* $\sum_{(l,m)} x(l, m), \forall (l, m) \in s$.

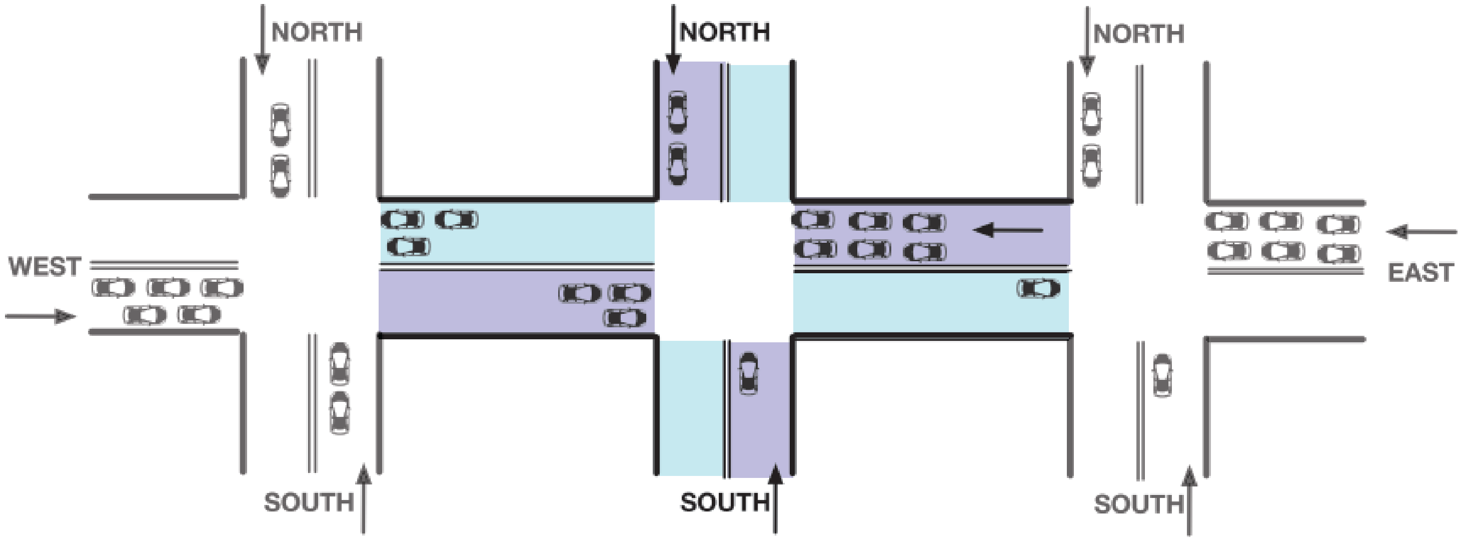


(a) Intersection (b) Eight phases

Figure 2: The illustration of an intersection with eight phases. In this case, phase **#2** is set.

Definition 4 – Pressure of an intersection

- **The pressure of an intersection**
  : Difference between
    the sum of the queuing vehicles on all the entering lanes and
    the sum of the queuing vehicles on all the exiting lanes



Pressure = |#queueing cars on entering lanes - #queueing cars on exiting lanes|
= | 3 + 2 + 6 + 1 - 3 - 0 - 1 - 0 |
= 8

Problem 1 – Multi-intersection traffic signal control

- Each intersection is controlled by an RL agent
- At time step t, agent i views part of the environment as its observation o_i^t
- Given the traffic situation and current traffic signal phase, the goal of the agent is to take an optimal action a (which phase to set)

Method

MPLight : Deep Q-Network

- **Pressure-based control law**
- **For large-scale network signal control,**

  we leverage parameter sharing among the agents

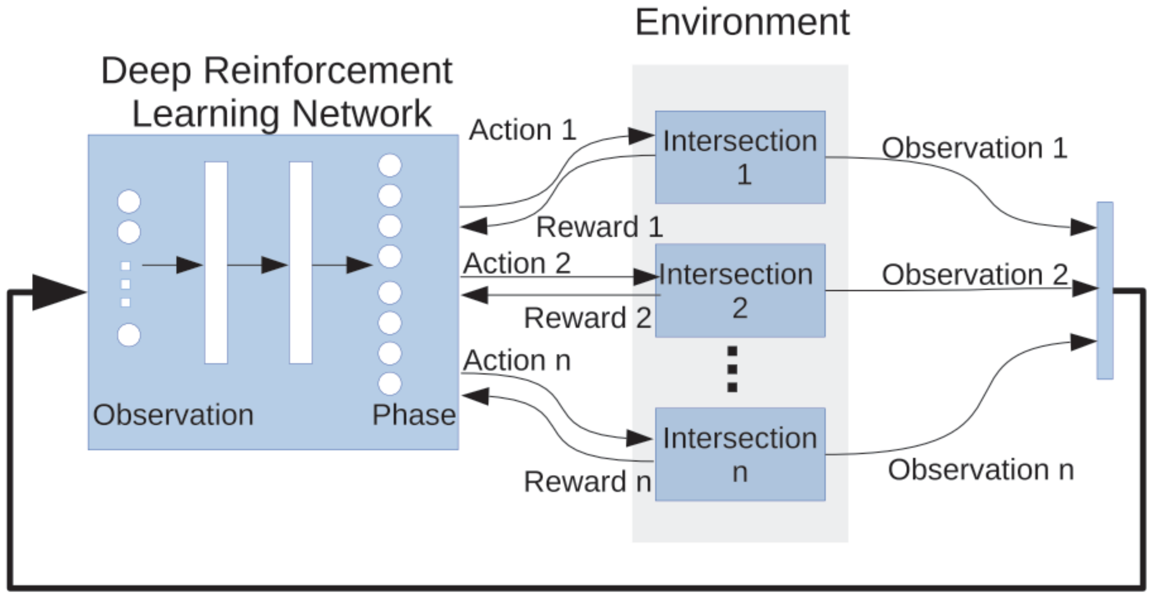


Figure 4: The framework of MPLight for multi-intersection signal control.

Pressure-based coordination

- **Pressure**
  : difference between upstream and downstream queue length
- **By minimizing the pressure,**
  : balance the distribution of the vehicles
  : maximize the system throughput

Pressure-based coordination

- **Max Pressure Control Law**
  : Max pressure control law select the phase with maximum pressure
- **Design an RL agent, PressLight**
  : using the **pressure-based reward** for long-term optimization.

Algorithm 1 Max Pressure Control

for *each intersection* i **do**
  for *each phase* s **do**
    calculate $p(s)$
  end
  next phase $\leftarrow \arg\max\{p(s)|s \in S_i\}$
end

DQN Agent

- Setting the reward of our RL agents to be the same as max pressure control objective
- Each local agent is maximizing its own cumulative reward

DQN Agent - Observation

- **Each agent observes part of the system state**
- **For standard intersection with 12 traffic movements**

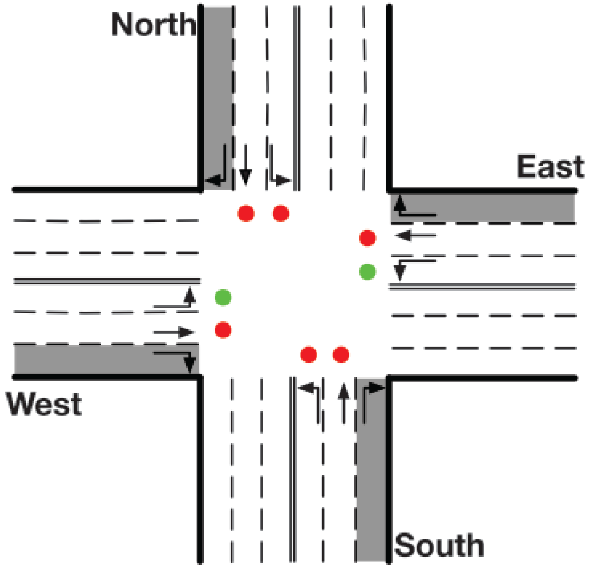
  : 1) current phase p

  : 2) Pressure of the 12 traffic movements

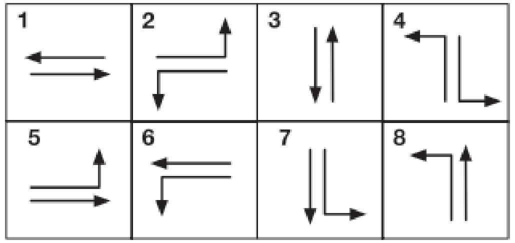
  : fewer than 12 movements, the vector is zero-padded

DQN Agent - Action

- At time t, each agent chooses a phase p as its action a_t
- Agents choose from a full set of **eight candidate** phases



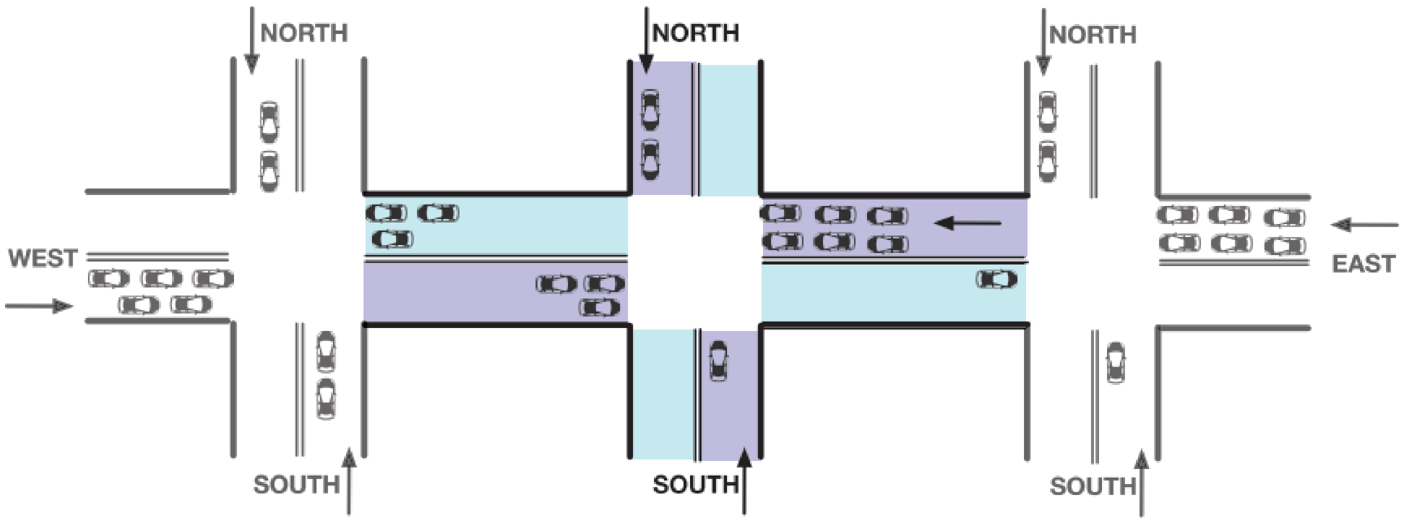
(a) Intersection



(b) Eight phases

DQN Agent – Reward

- **Reward r_i for agent i as the pressure on the intersection**
- **Pressure**
  : The queueing vehicles on all the entering lanes
  – Sum of the queueing vehicles on all the exiting lanes



Pressure = |#queueing cars on entering lanes - #queueing cars on exiting lanes|
= | 3 + 2 + 6 + 1 - 3 - 0 - 1 - 0 |
= 8

DQN Agent – FRAP Base Model

- **FRAP architecture as our base model**
  : design a network architecture for learning the **phase competition**
  in traffic signal control problem.

- **FRAP has two following advantages**
  (1) superior performance
  (2) faster training process compared with other sota signal control methods

DQN Agent – FRAP explanation - 1

- **FRAP** (**F**lipping and **R**otation and considers **A**ll **P**hase configurations)
  http://www.personal.psu.edu/~gjz5038/paper/cikm2019_frap/cikm2019_frap_paper.pdf

  : The difficulty is mainly due to the explosion of state space
  : A considerable portion of state-action pairs are unnecessary to explore

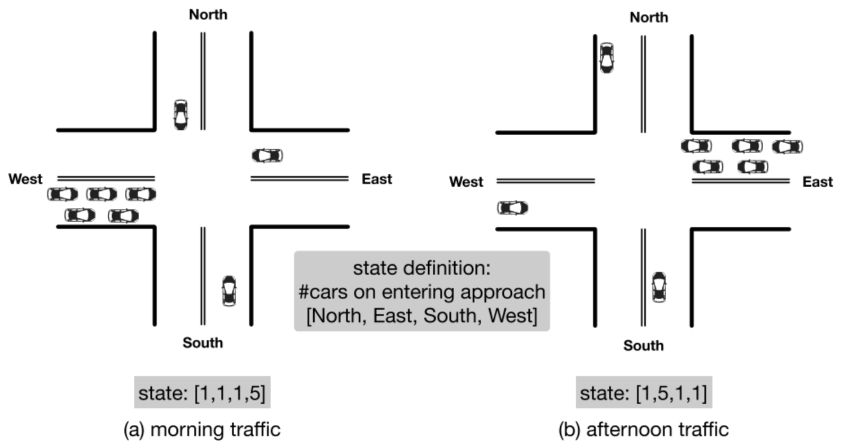


Figure 1: Traffic (a) and (b) are approximately flipped cases of each other.

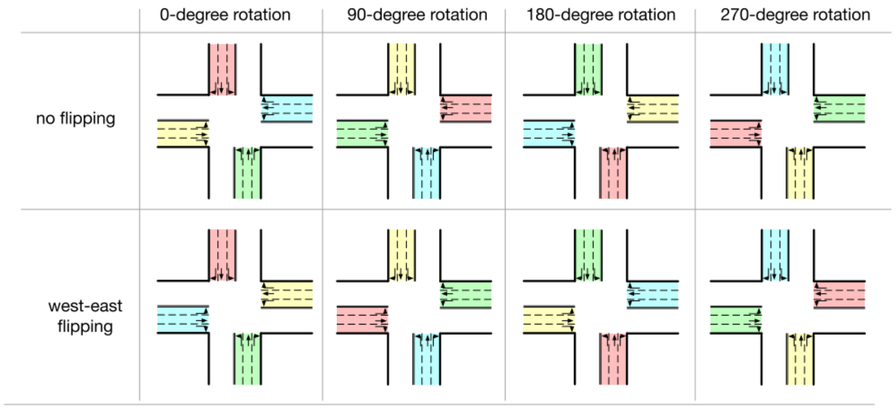


Figure 2: All the variations based on rotation and flipping of the left-most case. Ideally, a RL model should handle all these cases equally well.

DQN Agent – FRAP explanation - 2

- **When two traffic signals conflict, priority should be given to one with larger traffic movement**

DQN Agent – Deep Q-learning

- **We use Deep Q-Network (DQN) to solve the multi-intersection signal control problem.**
- **DQN takes the state features on the traffic movements as input and predicts the score (i.e., Q value) for each action candidate (i.e., phase)**

DQN Agent – Parameter sharing

- **Parameters of the network are shared among all the agents.**
- **The single PressLight model receives observations from different intersections to predict the corresponding actions and learns from environment rewards for parameter update.**

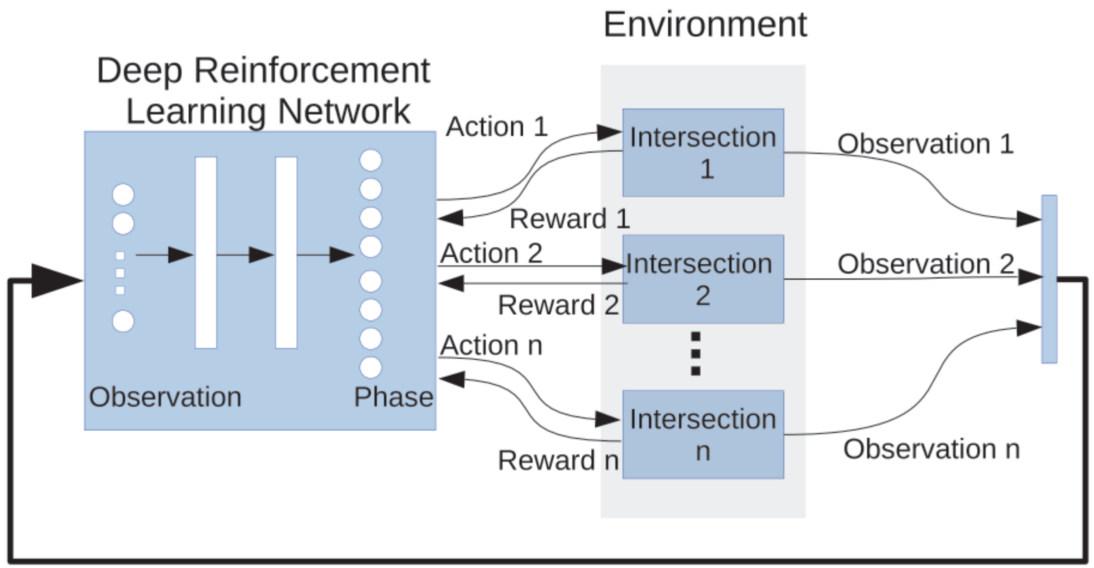


Figure 4: The framework of MPLight for multi-intersection signal control.

Experiment

Datasets – synthetic data

- **Both synthetic and real-world datasets are used**
  : Synthetic data on a 4 × 4 network
  : The turning ratios at the intersection are set as 10% (left), 60%(straight)

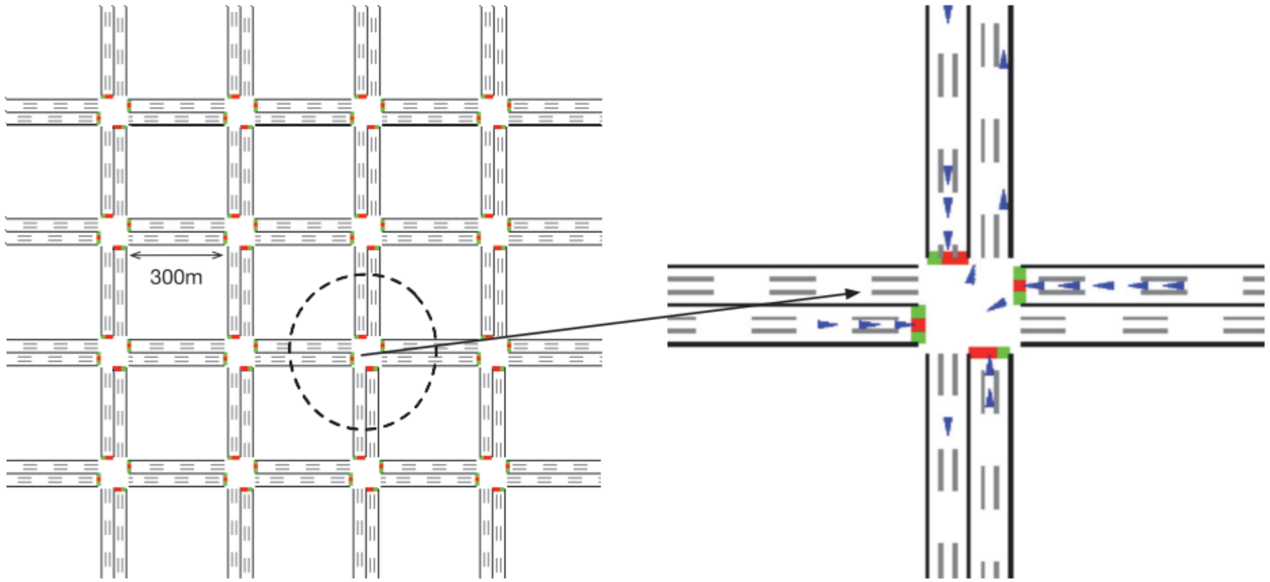


Figure 5: 4 × 4 road network.

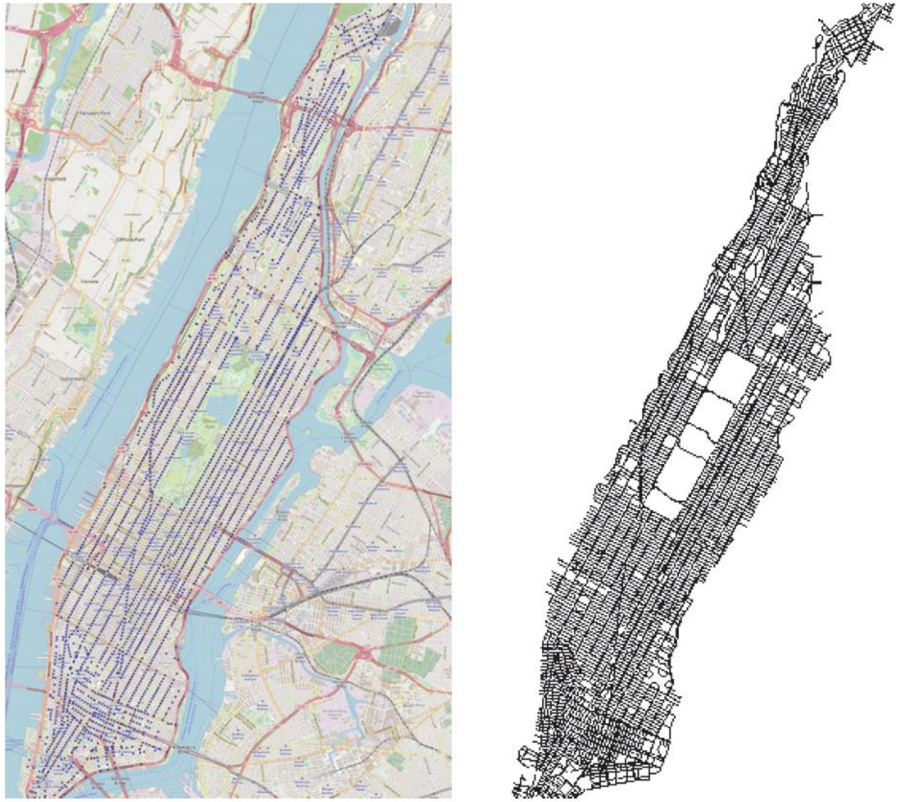
Datasets – real-world data

- **Both synthetic and real-world datasets are used**

  : Manhattan, New York City
  from Open Street Map

  : Traffic flow generated
  from the open-source taxi trip data

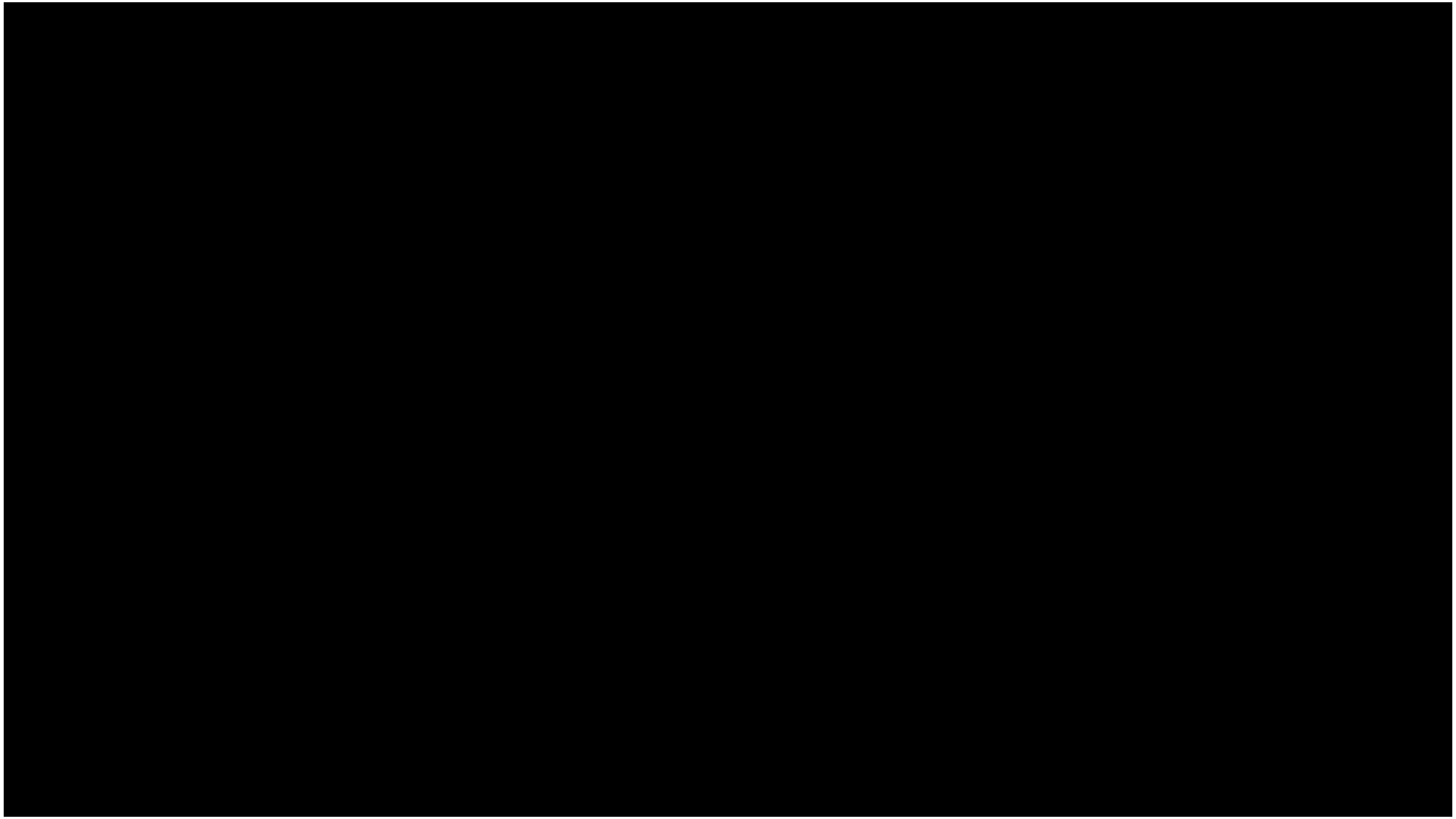
  : Manhattan dataset contains
  signalized 2510 traffic lights



(a) Manhattan (b) Manhattan in simulator

Figure 6: Road network of Manhattan in our experiments.

Simulation demo



Evaluation metrics

- **Travel time**
  : Average travel time of all vehicles in the system

- **Throughput**
  : The number of trips completed by vehicles

Performance Comparison

Table 2: Performance comparison of different methods evaluated in the four configurations of synthetic traffic data. For average travel time, the lower the better while for throughput, the higher the better.

Model	Travel Time				Throughput			
	Config 1	Config 2	Config 3	Config 4	Config 1	Config 2	Config 3	Config 4
FixedTime	573.13	564.02	536.04	563.06	3555	3477	3898	3556
MaxPressure	361.17	402.72	360.05	406.45	4702	4324	4814	4386
GRL	735.38	758.58	771.05	721.37	3122	2792	2962	2991
GCN	516.65	523.79	646.24	585.91	4275	4151	3660	3695
NeighborRL	690.87	687.27	781.24	791.44	3504	3255	2863	2537
PressLight	354.94	353.46	348.21	398.85	4887	4742	5129	5009
FRAP	340.44	298.55	361.36	598.52	5097	5113	5483	4475
MPLight	**309.33**	**262.50**	**281.34**	**353.13**	**5219**	**5213**	**5652**	**5060**

Scalability Analysis

- GRL and NeighborRL
  : unable to scale to large networks due to high complexity and computational costs.

Table 3: Performance of different methods on Manhattan, a large-scale road network with 2510 traffic signals.

Model	Travel Time	Throughput
FixedTime	974.23	1940
MaxPressure	497.76	2143
GRL	-*	-*
GCN	653.45	5045
NeighborRL	-*	-*
PressLight	600.42	3447
FRAP	512.70	6346
MPLight	**472.51**	**6932**

*No result

as GRL and NeighborRL can not scale up to thousands of intersections in New York's road network.

Ablation Study – pressure design

- **Impact of Pressure-based Design**
  : unable to scale to large networks due to high complexity and computational costs.

Table 4: Performance of different RL-based methods with and without "pressure" on Manhattan network.

Model	Travel Time
GCN	653.45
GCN + pressure	**646.47**
PressLight- pressure	654.04
PressLight	**600.42**
FRAP	512.70
FRAP + pressure	**472.51**

Ablation Study – parameter sharing

- **Impact of Parameter sharing**
  : unable to scale to large networks due to high complexity and computational costs.

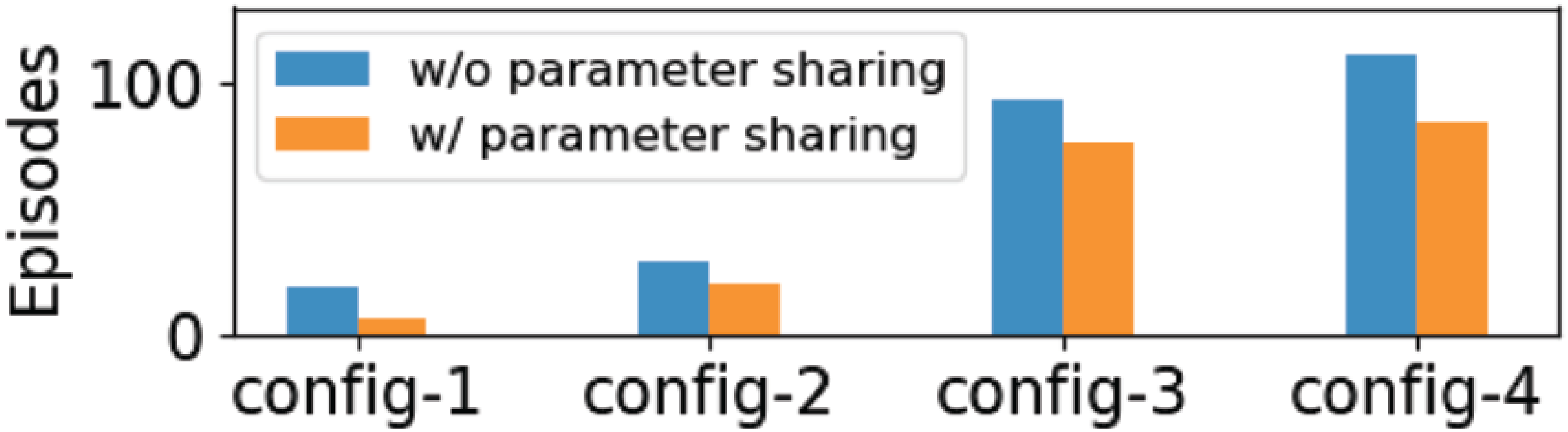


Figure 7: Number of episodes for models to converge.

Additional information

- **Reinforcement learning for traffic signal control**
  : https://traffic-signal-control.github.io/

- **Cityflow**
  : https://arxiv.org/abs/1905.05217
  : Twenty times faster than SUMO

- **Tutorial**
  :https://docs.google.com/presentation/d/12cqabQ_V5Q9Y2DpQOdpsHyrR6MIxy1CJlPmUE3Ojr8o/edit

감사합니다

Q & A