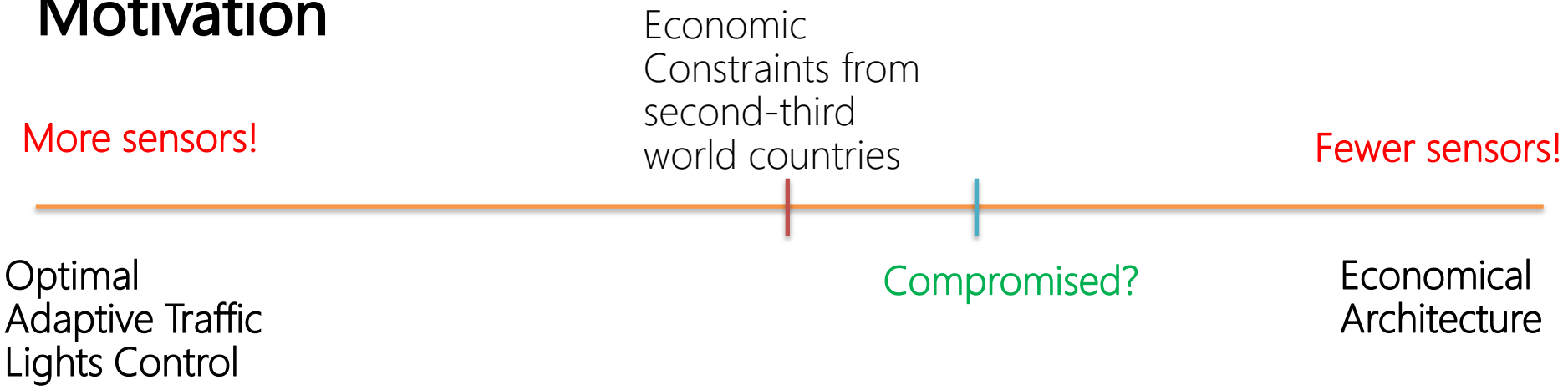
Reinforcement Learning Agent under Partial Observability for Traffic Light Control in Presence of Gridlocks

Thanapapas Horsuwan[1], Chaodit Aswakul[2]

[1] International School of Engineering, Faculty of Engineering, Chulalongkorn University
[2] Wireless Network and Future Internet Research Unit, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University

Motivation

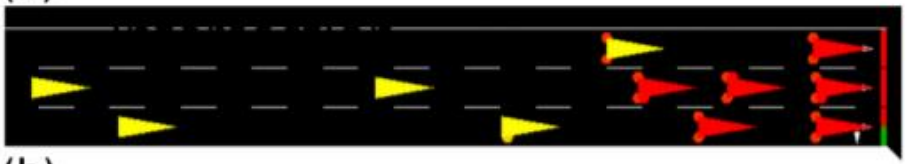


Aim

Further optimize the current traffic control policies without high additional investments

Existing Proposed State Spaces

(a)



(b)

0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	0	1	0	0	0	1	0	1	1
0	1	0	0	0	0	1	0	0	1	0	1

(c)

0	0	0	0	0	0	0	0	0.2	0	0	0
0.8	0	0	0	0.5	0	0	0	0	0	0	0
0	0.6	0	0	0	0	0.4	0	0	0	0	0

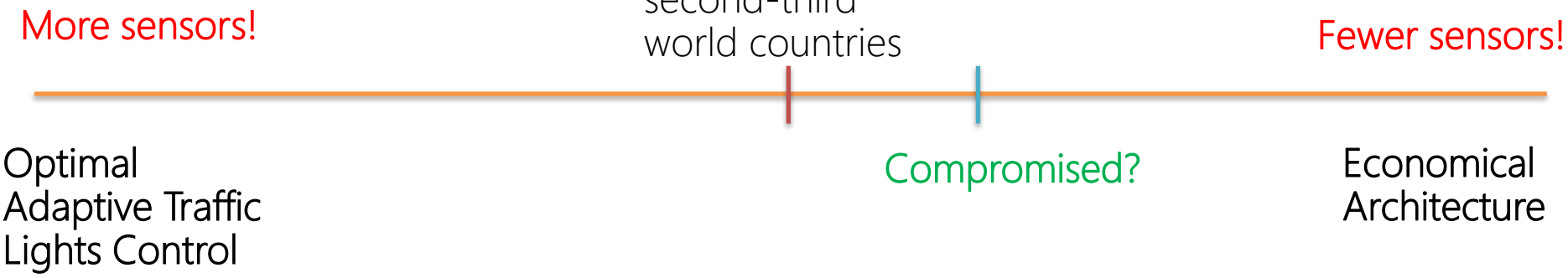
Fig. 1: Example of simulated traffic (a) with corresponding Boolean- (b) and real-valued DTSE vectors (c).

$$S \in (\mathbb{B} \times \mathbb{R})^{\frac{l}{c} \times n} \times P$$

Number of lanes *n*, cell length *c*, lane length *l*
P is the traffic light signal phase

[1] Juntao Gao, Yulong Shen, Jia Liu, Minoru Ito, and Norio Shiratori. Adaptive Traffic Signal Control: Deep Reinforcement Learning Algorithm with Experience Replay and Target Network. arXiv e-prints, page arXiv:1705.02755, May 2017.
[2] Wade Genders and Saiedeh Razavi. Using a Deep Reinforcement Learning Agent for Traffic Signal Control. arXiv e-prints, page arXiv:1611.01142, Nov 2016.

Motivation



Aim

Further optimize the current traffic control policies without high additional investments

Objective

Demonstrate the **controller**'s learning capability in spite of **limited sensory information** in a **gridlock setting**

Controller

Inputs

Reinforcement Learning Agent under Partial Observability for Traffic Light Control in Presence of Gridlocks

Setting

Controller

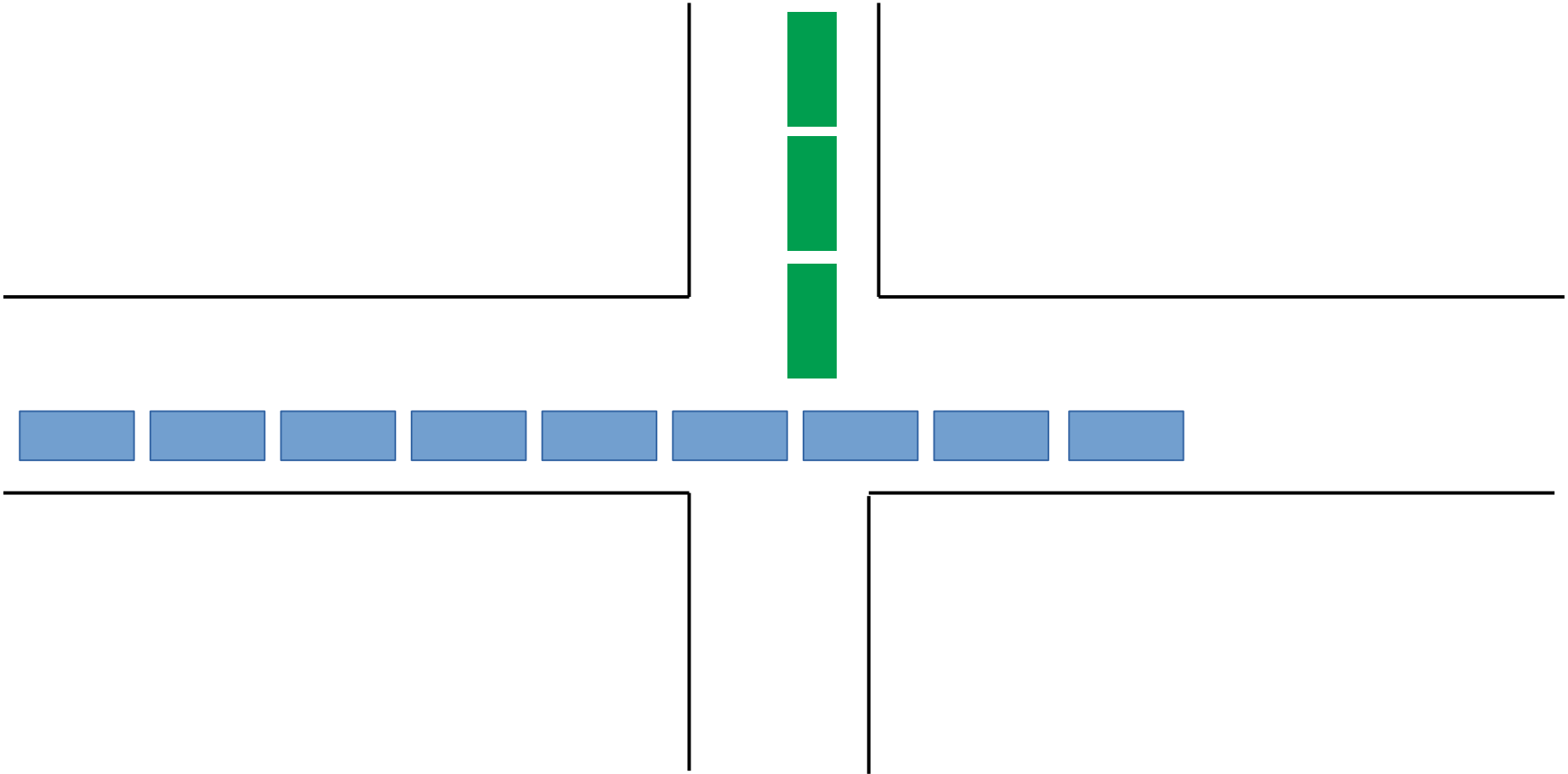
Inputs

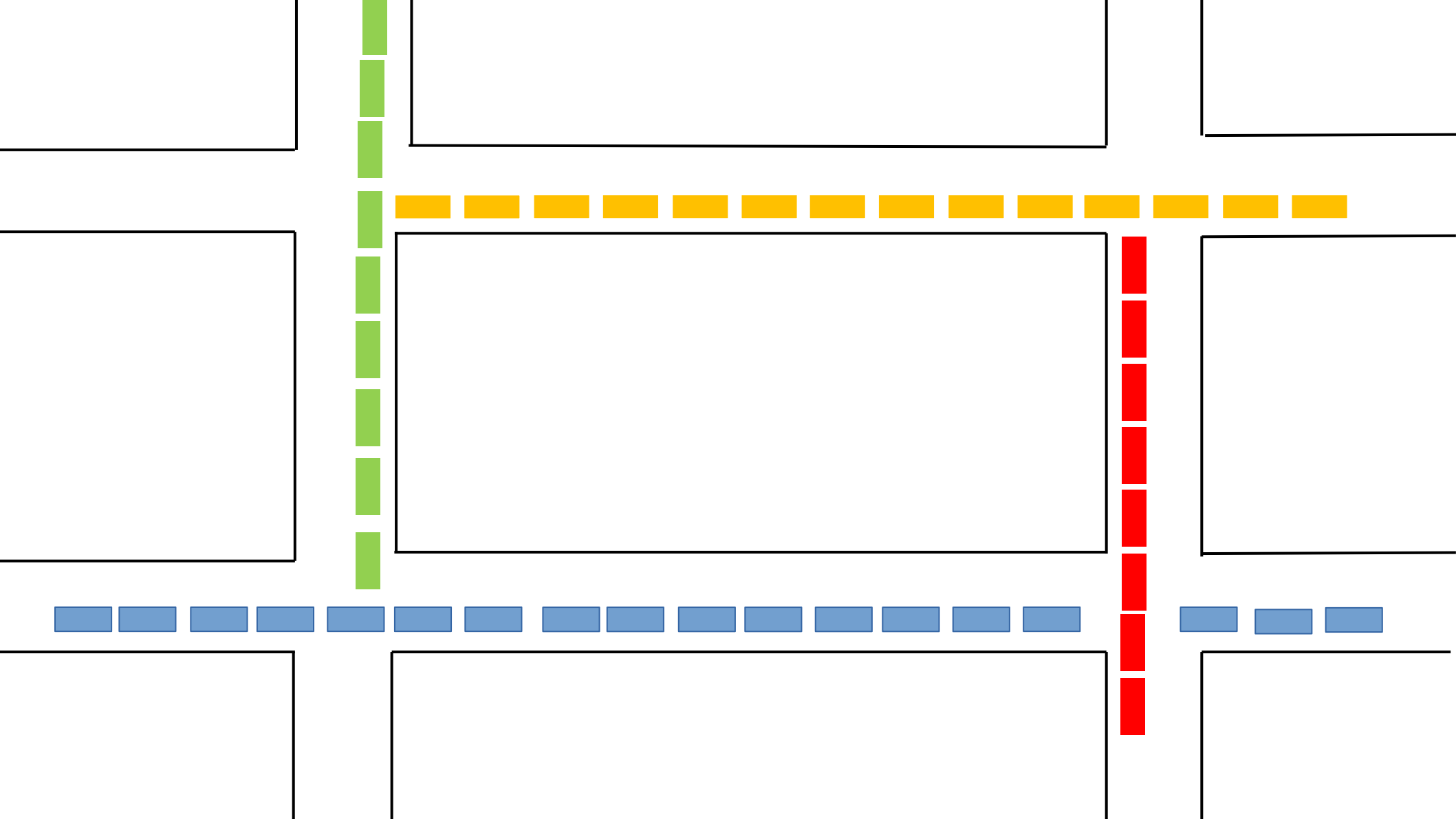
Reinforcement Learning Agent for Traffic Light Control under Partial Observability in Presence of Gridlocks

Setting

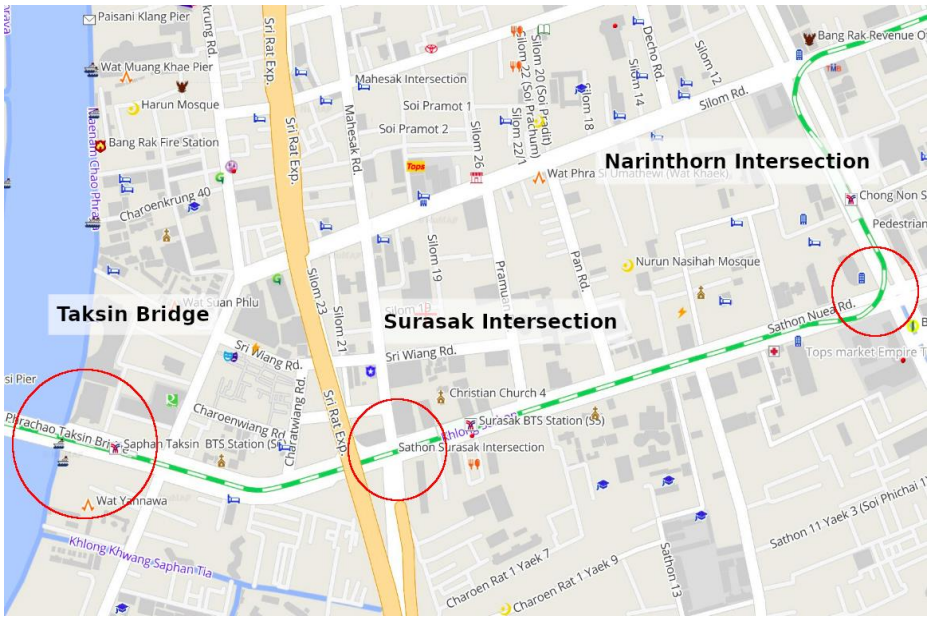
Traffic Gridlocks

A traffic gridlock is a form of congestion state where queue length spillback propagates in a closed loop-resulting in a complete standstill





Chula-Sathorn SUMO Simulator (Chula-SSS)



(a) Longdo Map (with granted usage permission from http://map.longdo.com/

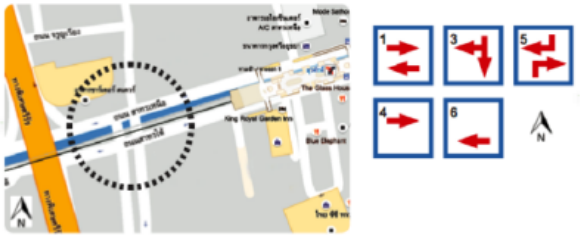
(b) Chula-SSS dataset in SUMO

Figure 1: Comparison between actual map and Chula-SSS dataset in the Sathorn Road Area

Chaodit Aswakul, Sorawee Watarakitpaisarn, Patrachart Komolkiti, Chonti Krisanachantara, and Kittiphan Techakittiroj. Chula-SSS: Developmental Framework for Signal Actuated Logics on SUMO Platform in Over-Saturated Sathorn Road Network Scenario. In *SUMO 2018- Simulating Autonomous and Intermodal Transport Systems*, volume 2 of EPiC Series in Engineering, pages 67–81. EasyChair, 2018

Figure [1]: Queue Length Spillback at Narinthorn Intersection

[1] Piantanongkit, Pon. A Photograph of Narinthorn Intersection with Queue Length Spillback. MotoRival, Bike News, 21 Apr. 2017, https://www.motorival.com/sathorn-model-plan/.



Standard Phase 1-3-1-5

Changing Phase from 1 to 3:

- When the queue of downstream North Sathorn reaches the Sathorn Intersection
- When the queue of downstream South Sathorn reaches the Sathorn Intersection
- Queue of Si Wiang Road reaches Pramuan Road on Bangkok Christian College and Assumption Convent School
- Vehicles on Taksin Bridge 300 meters from Sathorn Intersection is starting to move
- Phase 1 duration more than 120-150 seconds

Changing Phase from 3 to 1:

- Reduced jam length of Si Wiang Road or vehicles are moving on Pramuan Road continuously for 20-30 seconds
- Minimum gap between vehicles that crosses the intersection is too high
- Velocity of the vehicles that crosses the intersection is too low
- Phase 3 duration more than 30-80 seconds

Changing Phase from 1 to 5:

- Queue of Si Wiang Road reaches Pramuan Road on Bangkok Christian College and Assumption Convent School
- Queue of CharoenRat is too long
- Phase 1 duration more than 120-150 seconds

Changing Phase from 5 to 1:

- Reduced jam length of CharoenRat Road
- Phase 5 duration more than 40-50 seconds

Note: Use Phase 1-3-1-3-5 if want to get cars out from Surasak and Pramuan Road. Use Phase 4 (instead of Phase 1) when the head of queue from Sathorn South reaches Sathorn Intersection. Use Phase 6 (instead of Phase 1) when the head of queue from Sathorn North reaches Sathorn Intersection.

Figure 4: Heuristic Signal Actuated Logics at the Morning Rush Hour of Surasak Intersection [3] Version 23/11/2016 Yannawa Police District

Chaodit Aswakul, Sorawee Watarakitpaisarn, Patrachart Komolkiti, Chonti Krisanachantara, and Kittiphan Techakittiroj. Chula-SSS: Developmental Framework for Signal Actuated Logics on SUMO Platform in Over-Saturated Sathorn Road Network Scenario. In *SUMO 2018- Simulating Autonomous and Intermodal Transport Systems*, volume 2 of EPiC Series in Engineering, pages 67–81. EasyChair, 2018

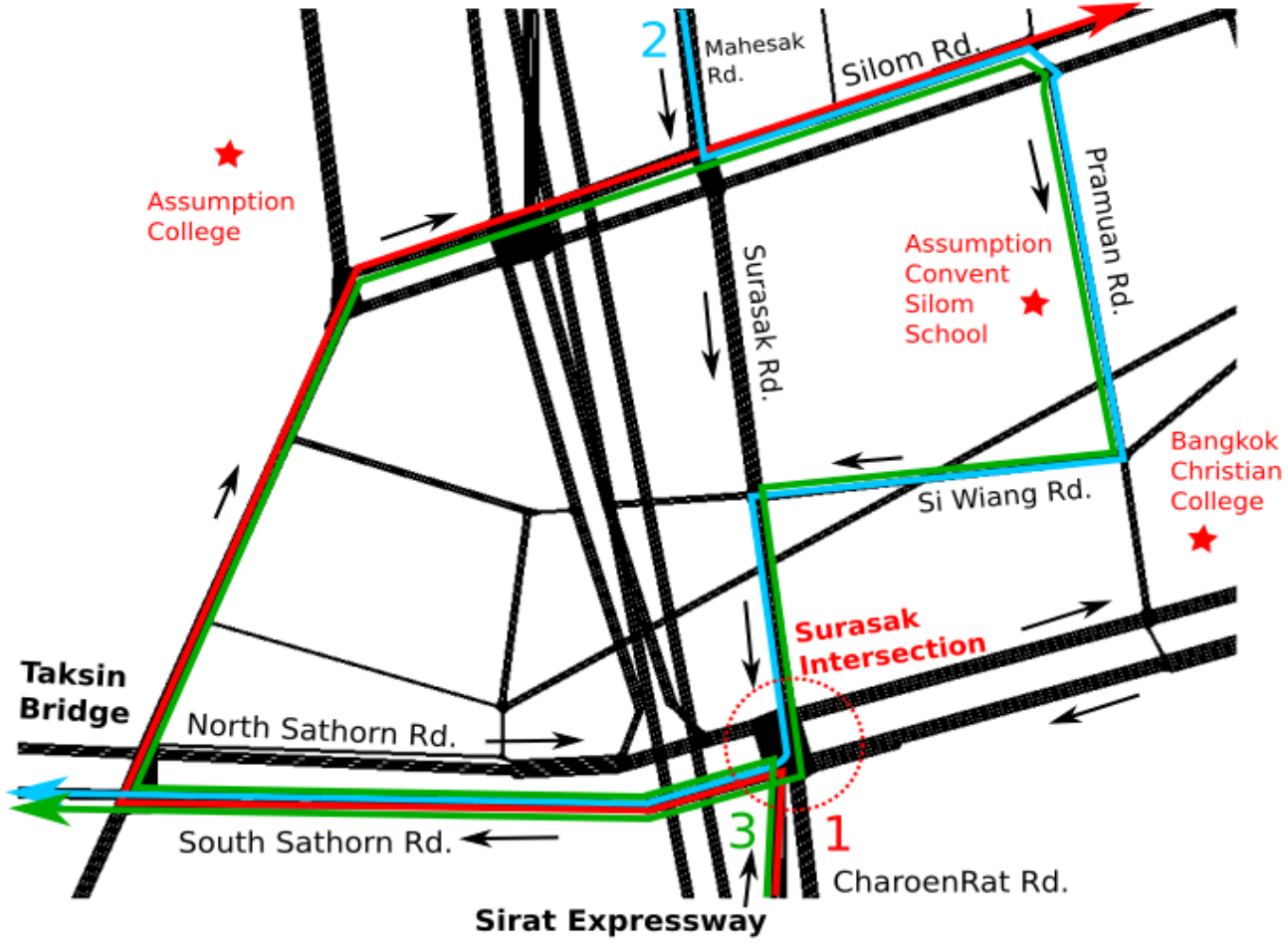


Figure 3: Critical Routes in the Sathorn Network

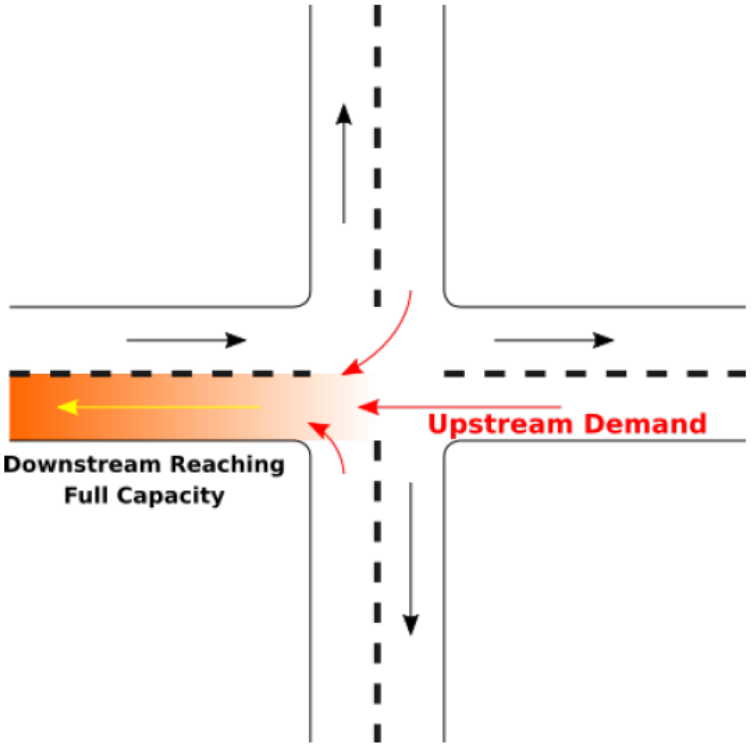
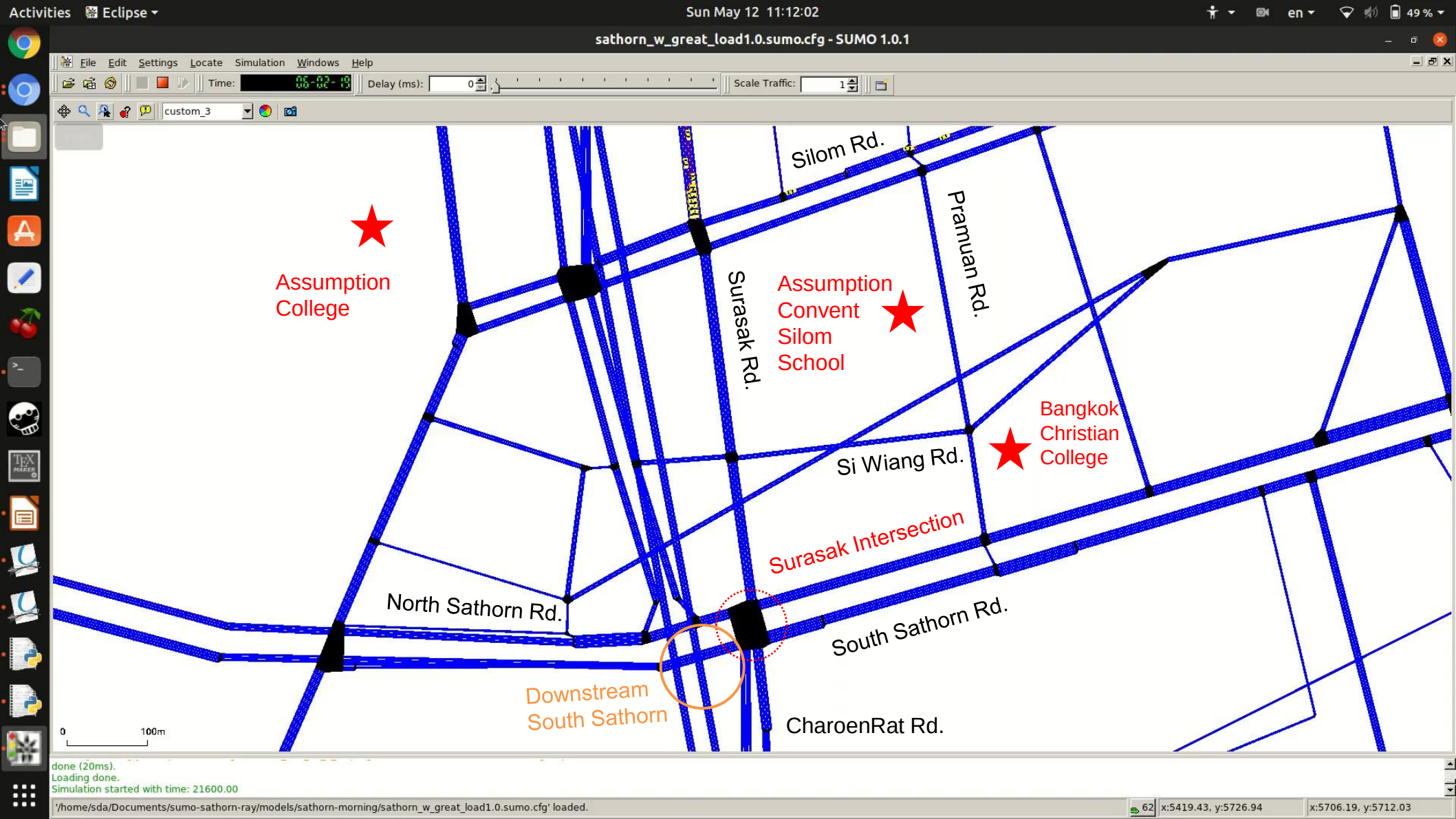


Figure 2: Example scenario for Gridlock


Activities
Eclipse
Sun May 12 11:12:02
en
49 %
sathorn_w_great_load1.0.sumo.cfg - SUMO 1.0.1
File
Edit
Settings
Locate
Simulation
Windows
Help
Time:
Delay (ms):
0
Scale Traffic:
1
custom_3
Silom Rd.
Pramuan Rd.
Surasak Rd.
Assumption College
Assumption Convent Silom School
Bangkok Christian College
Si Wiang Rd.
Surasak Intersection
North Sathorn Rd.
South Sathorn Rd.
Downstream South Sathorn
CharoenRat Rd.
0
100m
done (20ms).
Loading done.
Simulation started with time: 21600.00
'/home/sda/Documents/sumo-sathorn-ray/models/sathorn-morning/sathorn_w_great_load1.0.sumo.cfg' loaded.
62
x:5419.43, y:5726.94
x:5706.19, y:5712.03

Controller

Inputs

Reinforcement Learning Agent for Traffic Light Control under Partial Observability in Presence of Gridlocks

Setting

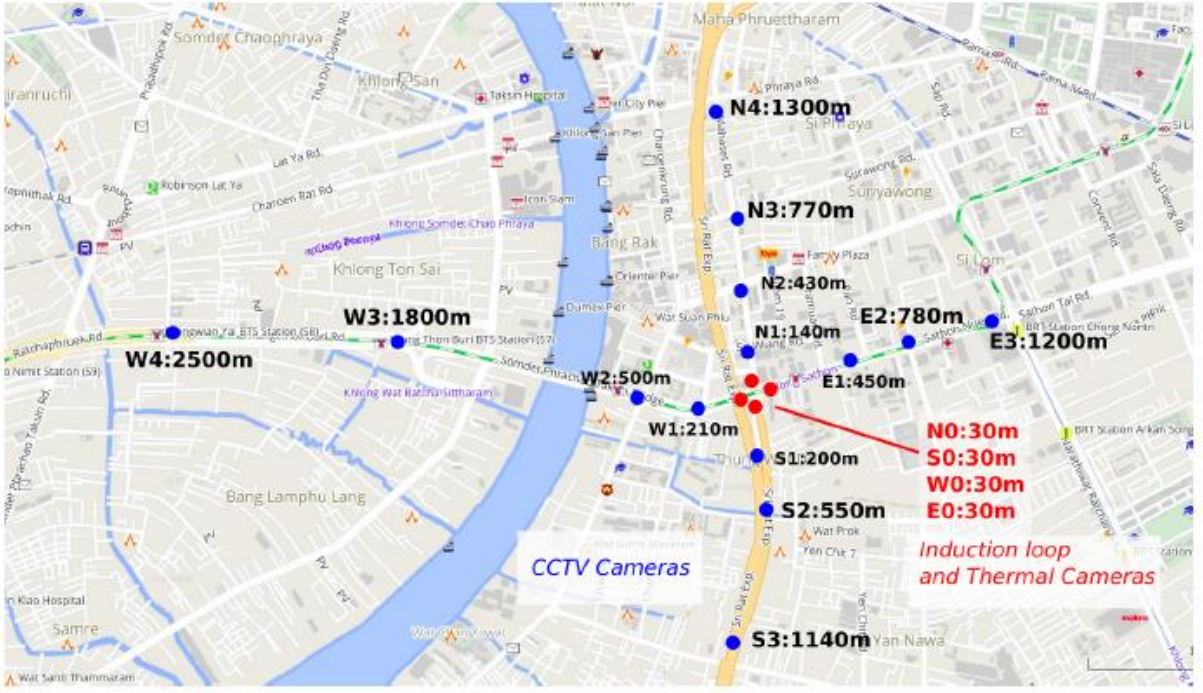


Figure 5: Sensor Configuration with granted usage permission from http://map.longdo.com

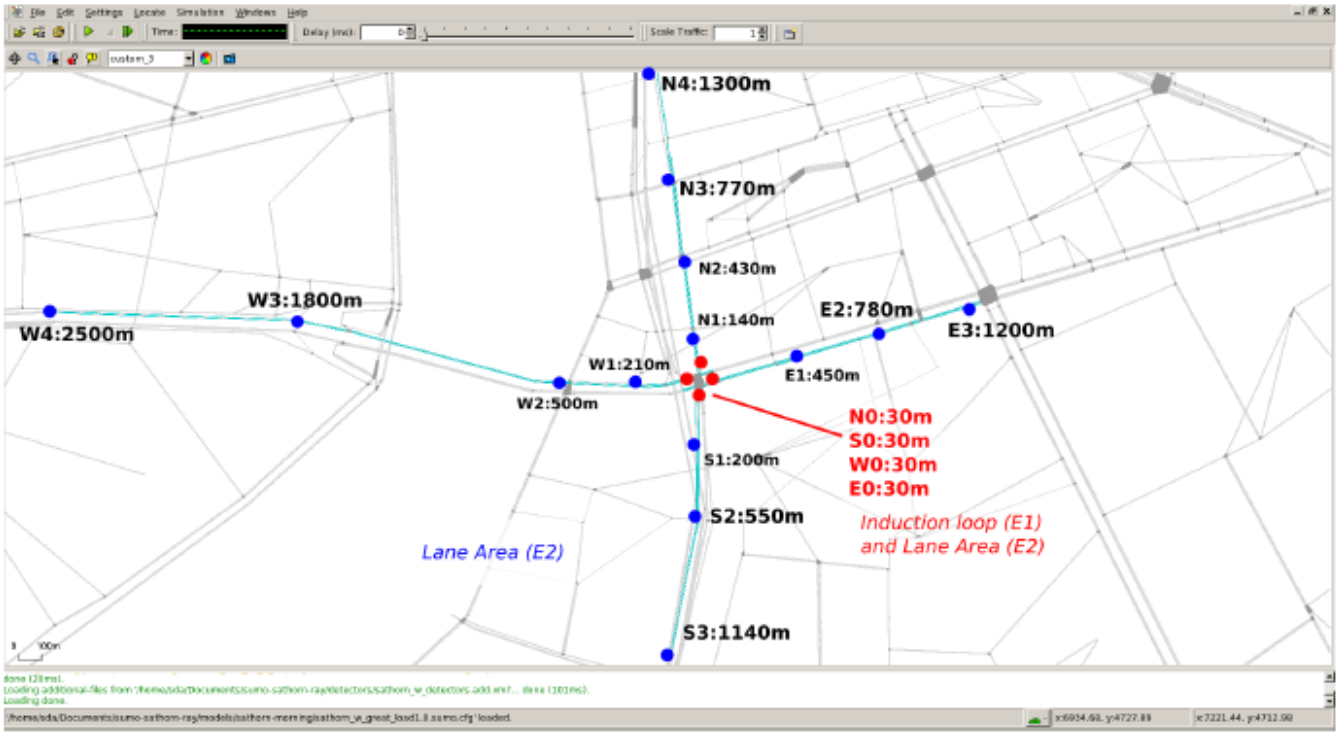


Figure 6: Detector Configuration in SUMO

Controller

Inputs

Reinforcement Learning Agent under Partial Observability
for Traffic Light Control in Presence of Gridlocks

Setting

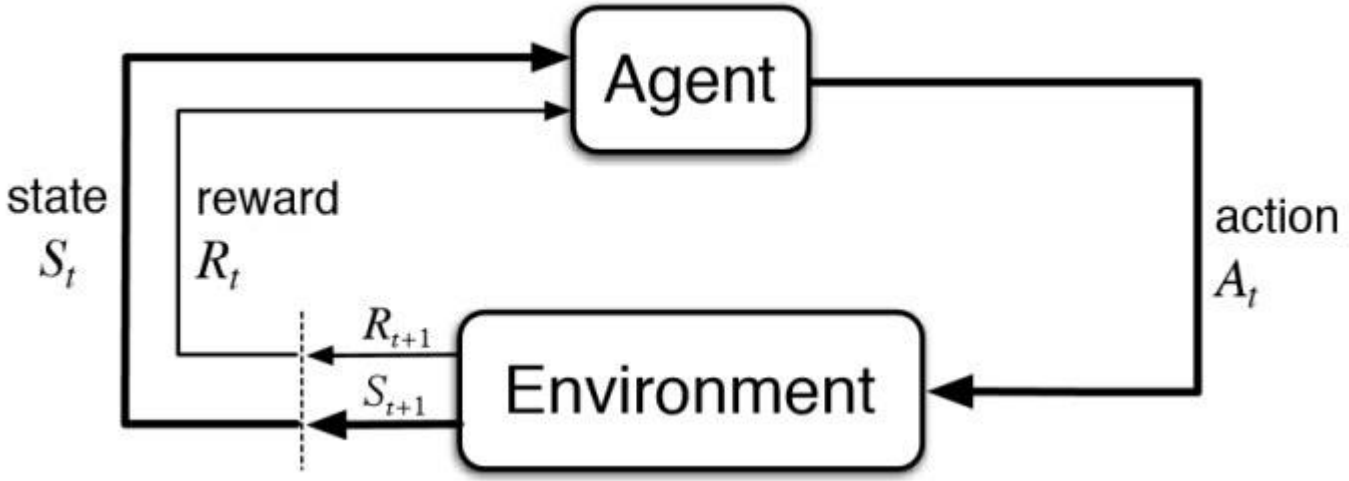


Figure: Reinforcement Learning

$$v_\pi(s) = \mathbb{E}_\pi \left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots \mid S_t = s \right]$$

Expected

Reward discounted

Given that state

State Space

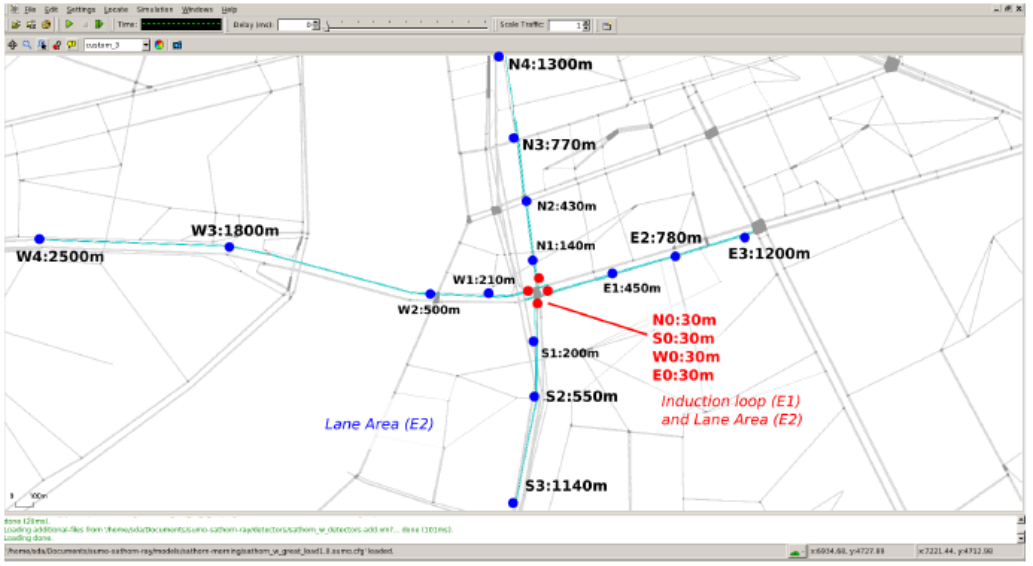


Figure 6: Detector Configuration in SUMO

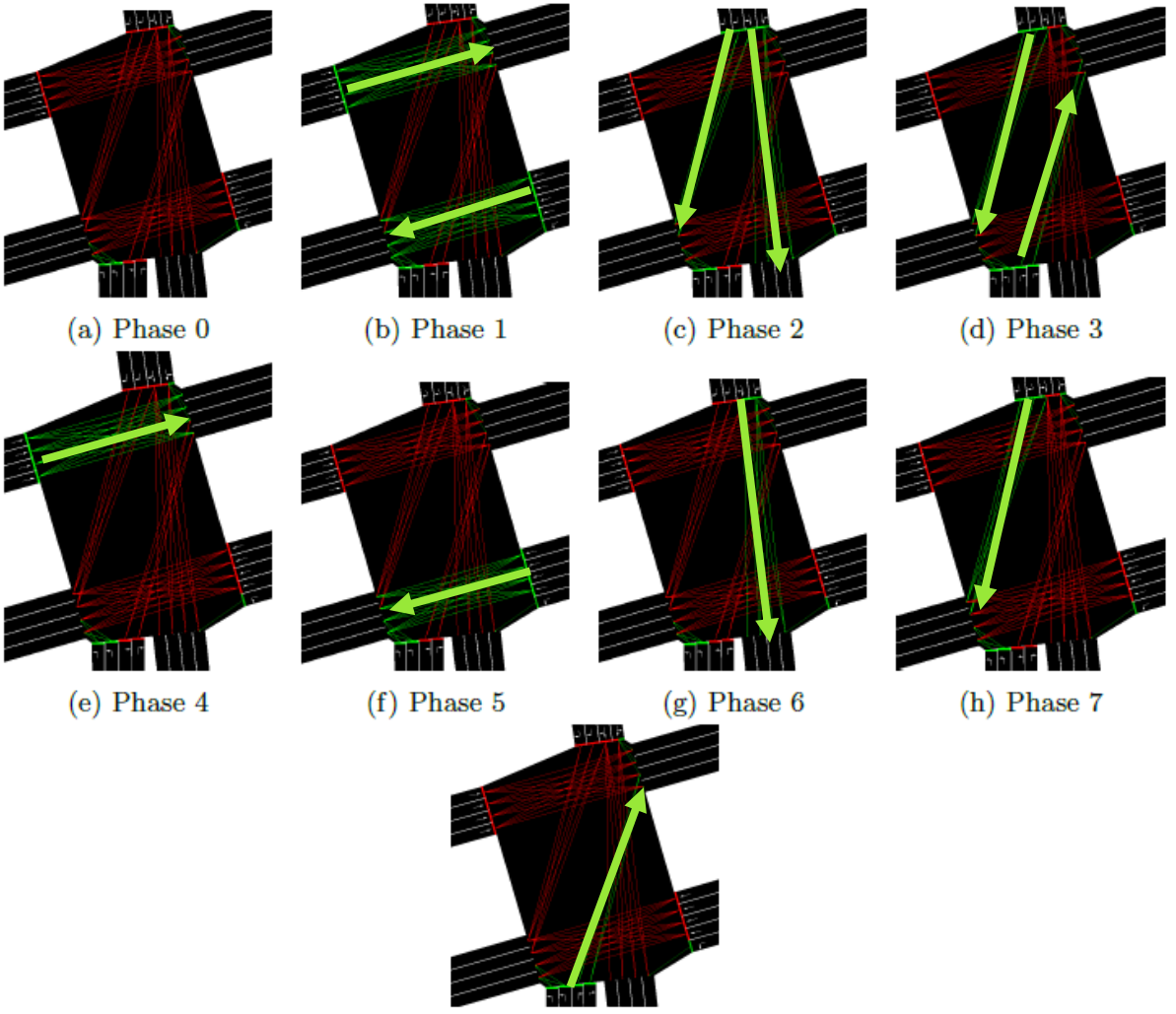
$$S^a \in \mathbb{R}^{21} \times P$$

$$P \in \mathbb{B}^{|\mathcal{A}|}.$$

The State is the Occupancy value of each E2 Detector (cell) and the Traffic Phase *P*

Action Space

$$\mathcal{A} = \{0, ..., 8\}$$



(a) Phase 0 (b) Phase 1 (c) Phase 2 (d) Phase 3

(e) Phase 4 (f) Phase 5 (g) Phase 6 (h) Phase 7

(i) Phase 8

Figure 7: Action Space Consisting of 9 Phases

Reward Function

$$r_{t+1} = \alpha\mu_{t+1} - \beta(\mathcal{O}_{t+1} \cdot C)$$

whereby $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}^+$

μ_{t+1}	Vehicle throughput during time step t to t + 1
$\mathcal{O}_{t+1}$	Observed occupancy in the next time step
C	Maximum cell capacity

α kept constant at 1; β linear sequence from 0.04 to 0.16 in intervals of 0.04

Agent Architecture: Ape-X Deep Q-Network (DQN)

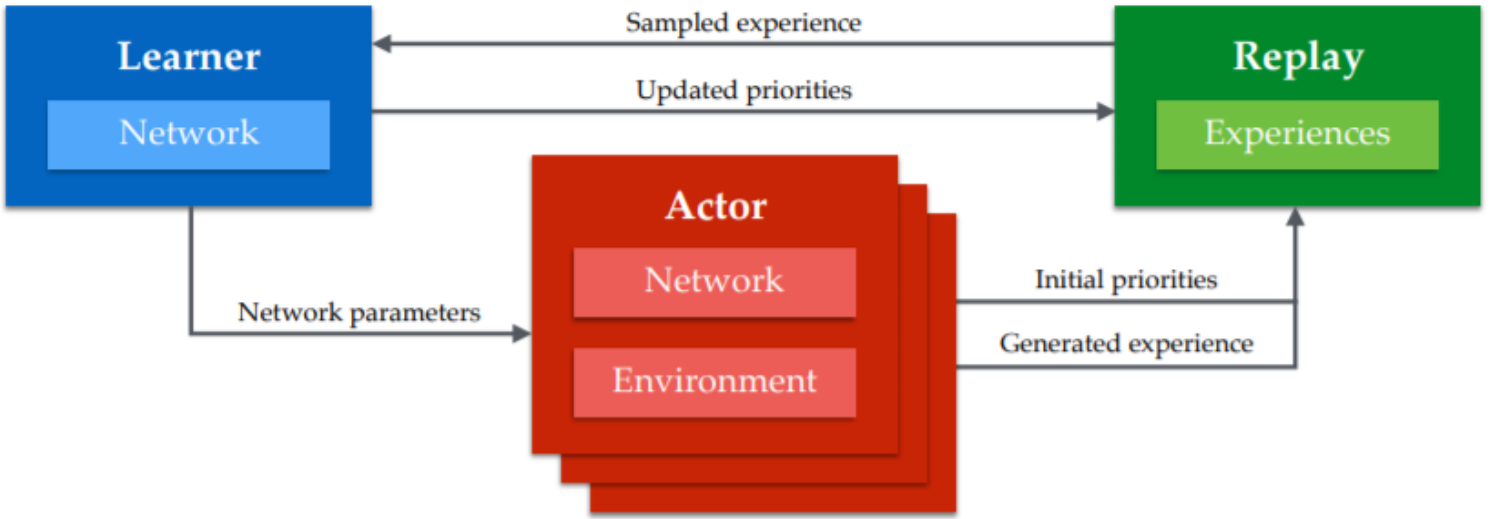


Figure 1: The Ape-X architecture in a nutshell: multiple actors, each with its own instance of the environment, generate experience, add it to a shared experience replay memory, and compute initial priorities for the data. The (single) learner samples from this memory and updates the network and the priorities of the experience in the memory. The actors' networks are periodically updated with the latest network parameters from the learner.

Dan Horgan, John Quan, David Budden, Gabriel Barth-Maron, Matteo Hessel, Hado van Hasselt, and David Silver. Distributed Prioritized Experience Replay. arXiv e-prints, page arXiv:1803.00933, Mar 2018.

Agent Architecture: Ape-X Deep Q-Network (DQN)

$$L_t(\theta_t) = \frac{1}{2}\mathbb{E}_{(s,a,r,s')\sim p_t(D)}\left[\left(r^{(n)} + \gamma^n Q(s^{(n)}, \underset{a}{\operatorname{argmax}}\, Q(s^{(n)}, a; \theta_t), \theta_t^-) - Q(s, a; \theta_t\right)^2\right]$$

where the experience $(s, a, r, s') \sim p_t(D)$ is sampled in a prioritized experience replay manner and p_t denotes the probability distribution of selecting an experience. The superscript (n) denotes the n-step learning.

Dan Horgan, John Quan, David Budden, Gabriel Barth-Maron, Matteo Hessel, Hado van Hasselt, and David Silver. Distributed Prioritized Experience Replay. arXiv e-prints, page arXiv:1803.00933, Mar 2018.

Experimental Setup

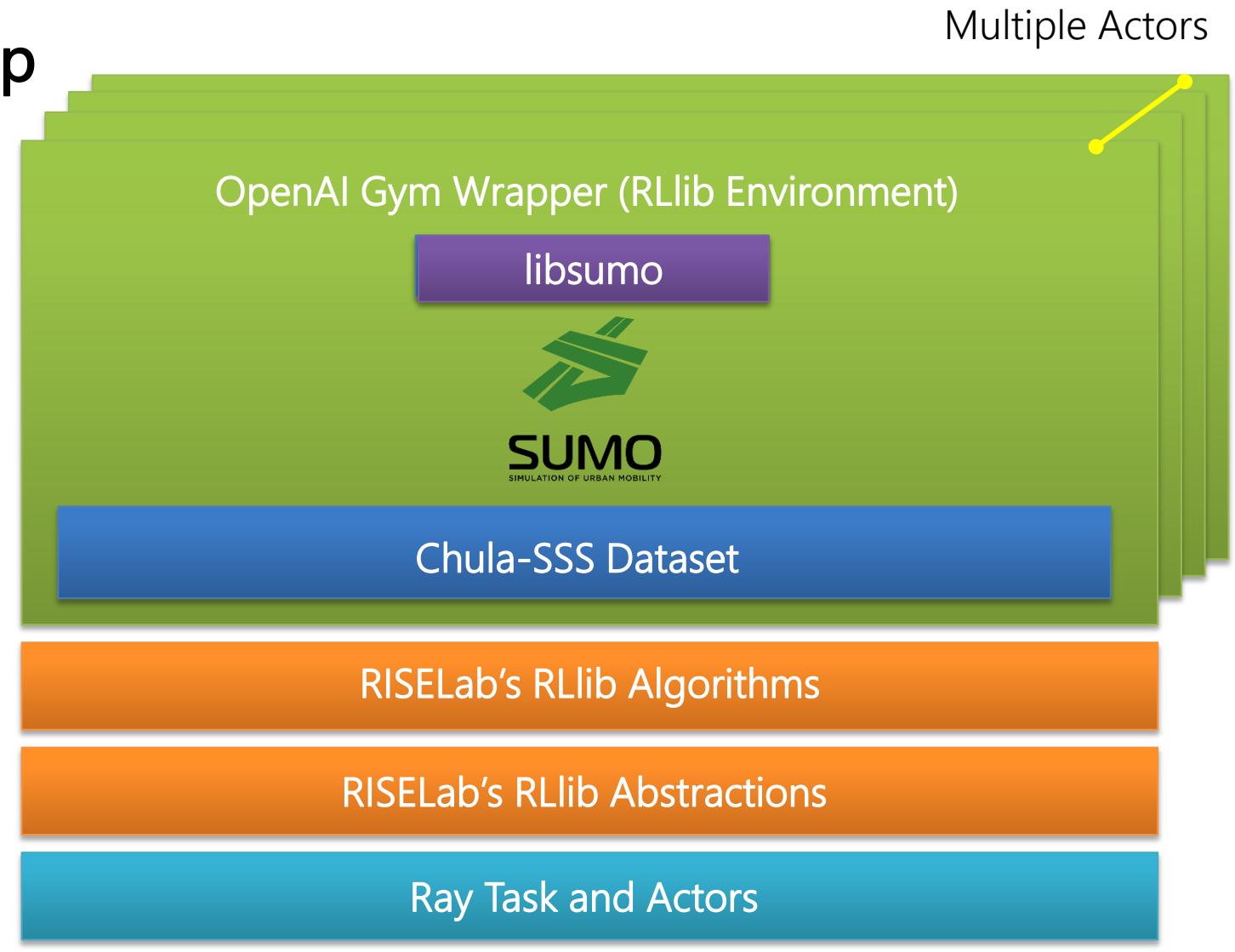
1 Episode: 6 AM to 9 AM
3 hours
10800 Simulation Steps

1 Epoch: 4 Episodes
43200 Simulation Steps

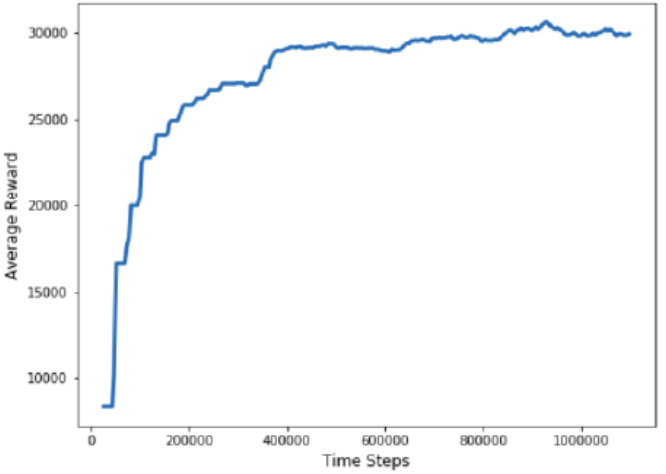
Training:

250 Epochs:
10.8M Simulated Steps
~125 Simulated Days

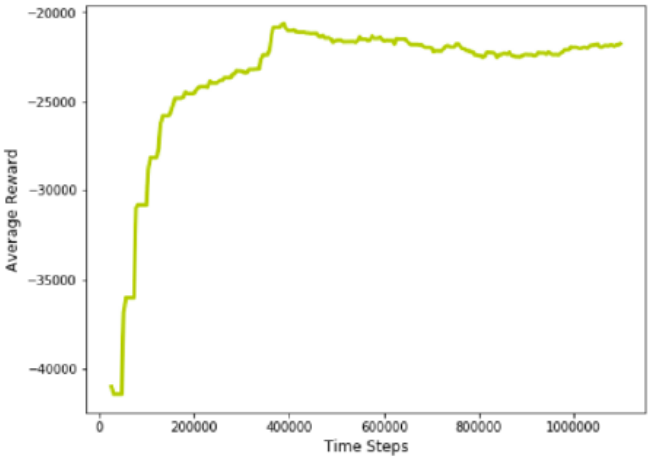
Agent Step:
10 Simulation Steps



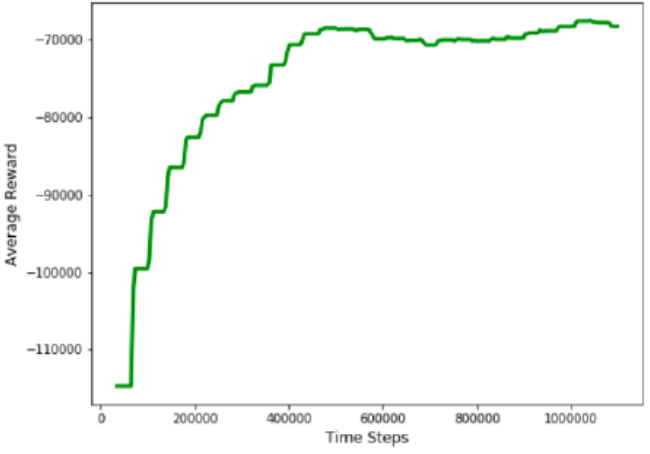
Results



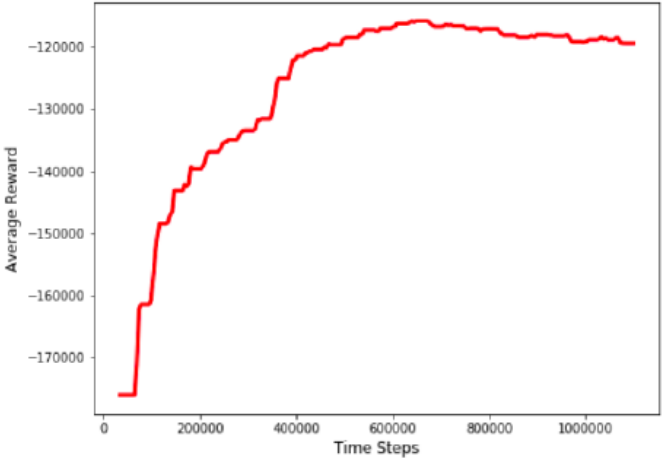
(a) Reward α=1 β=0.04



(b) Reward α=1 β=0.08



(c) Reward α=1 β=0.12



(d) Reward α=1 β=0.16

Figure 8: Reward of Different Parameters

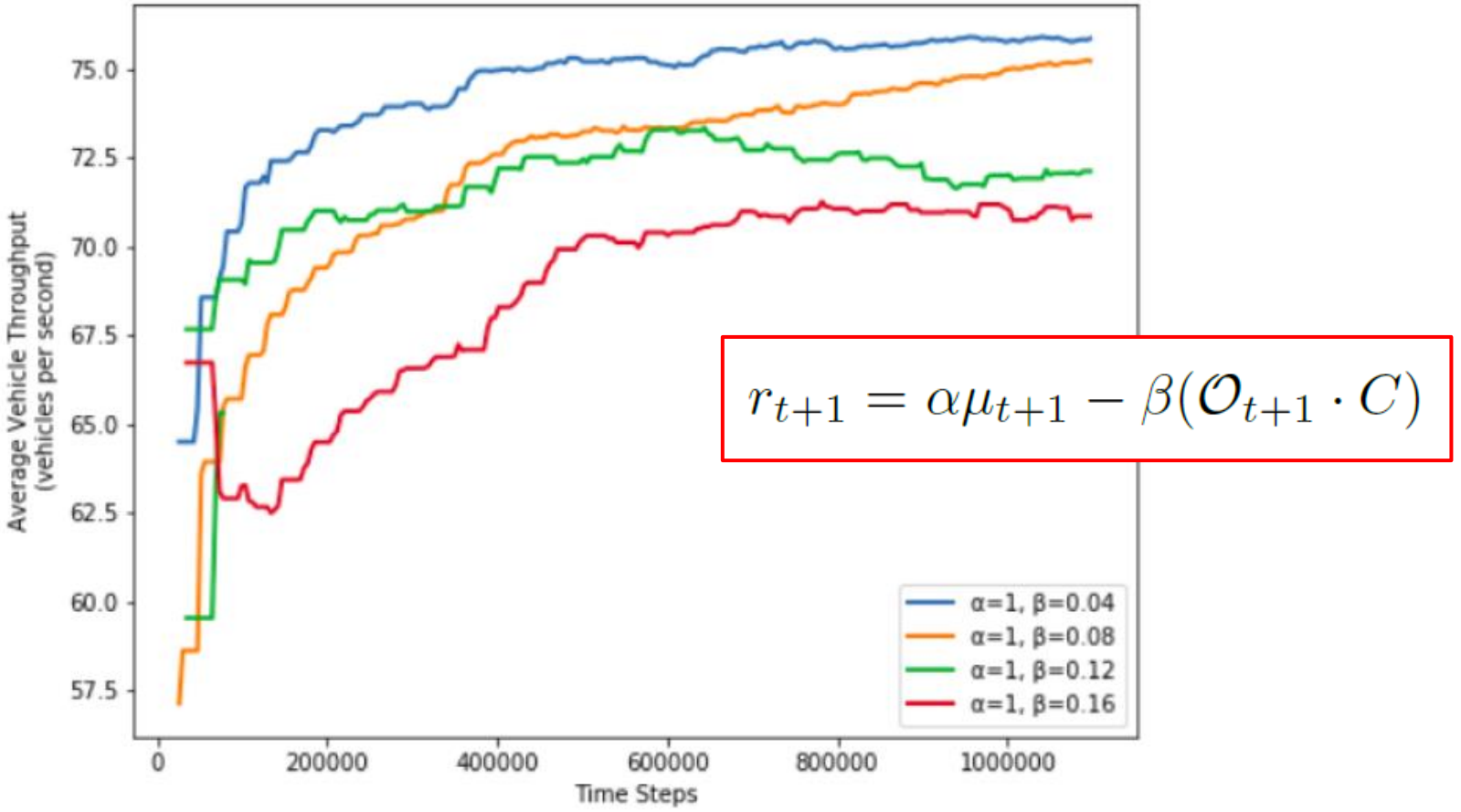
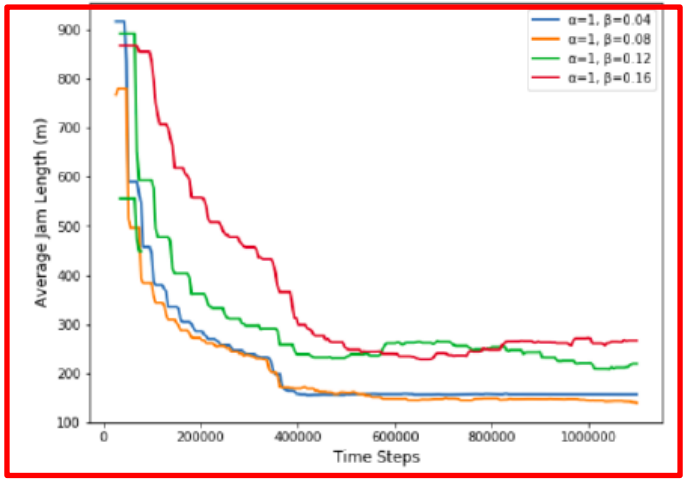
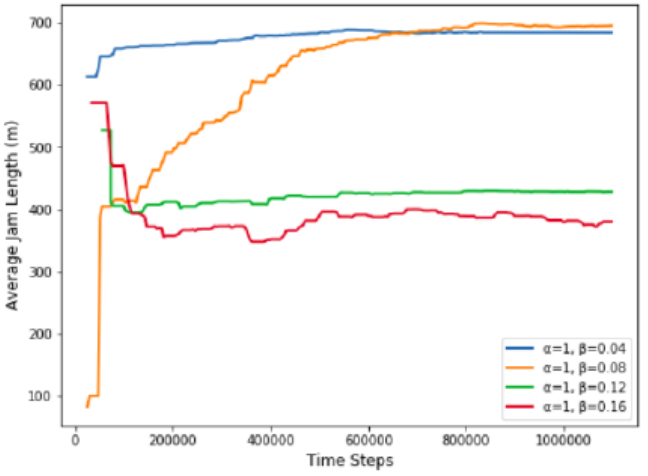


Figure 9: Average Vehicle Throughput

Decreasing Trend

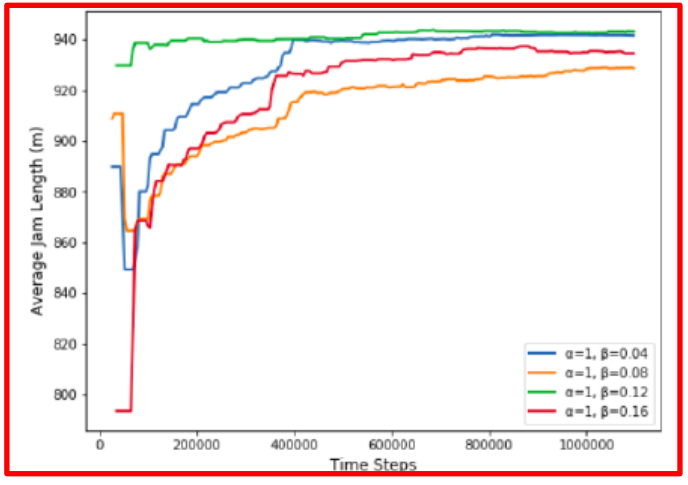


(a) Surasak Jam Length

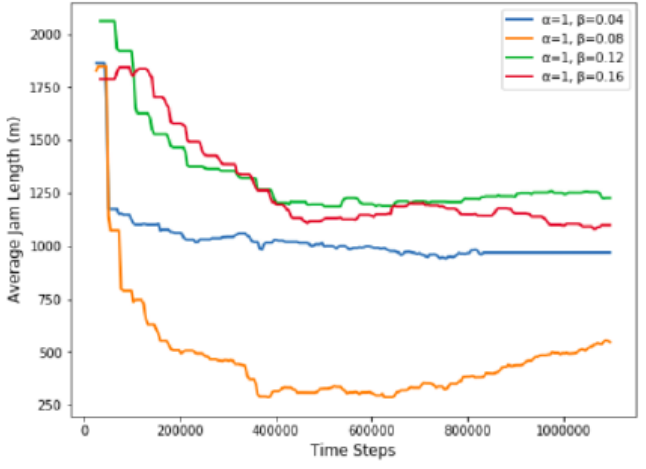


(b) CharoenRat Jam Length

Increasing Trend

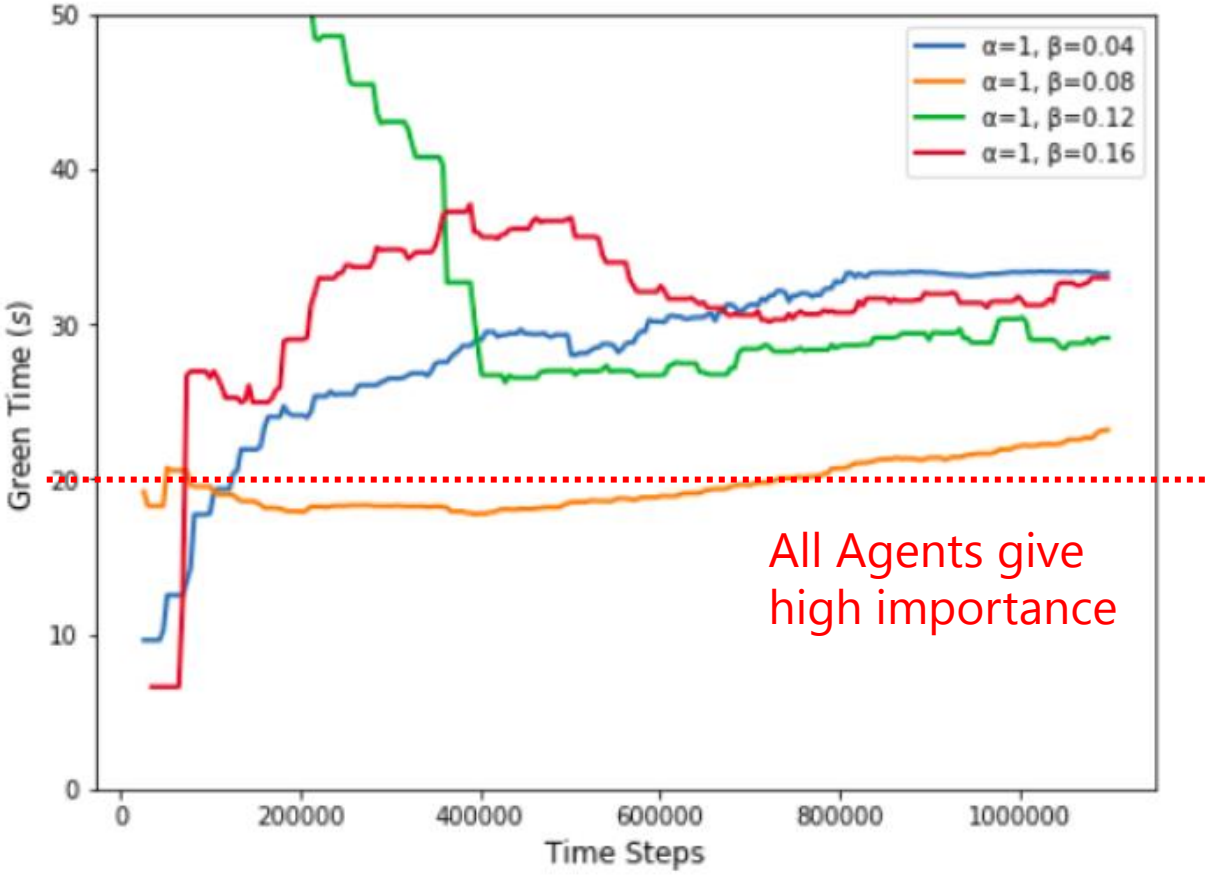


(c) South Sathorn Jam Length



(d) North Sathorn Jam Length

Figure 11: Jam Length of the Approaching Lanes



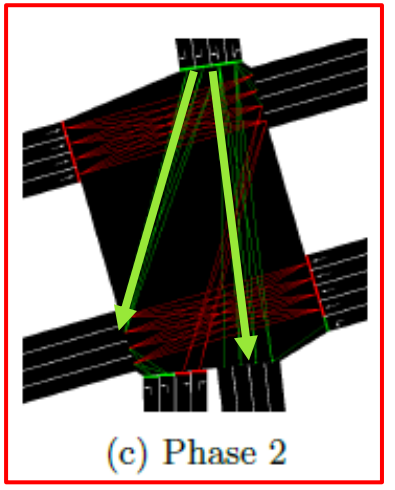
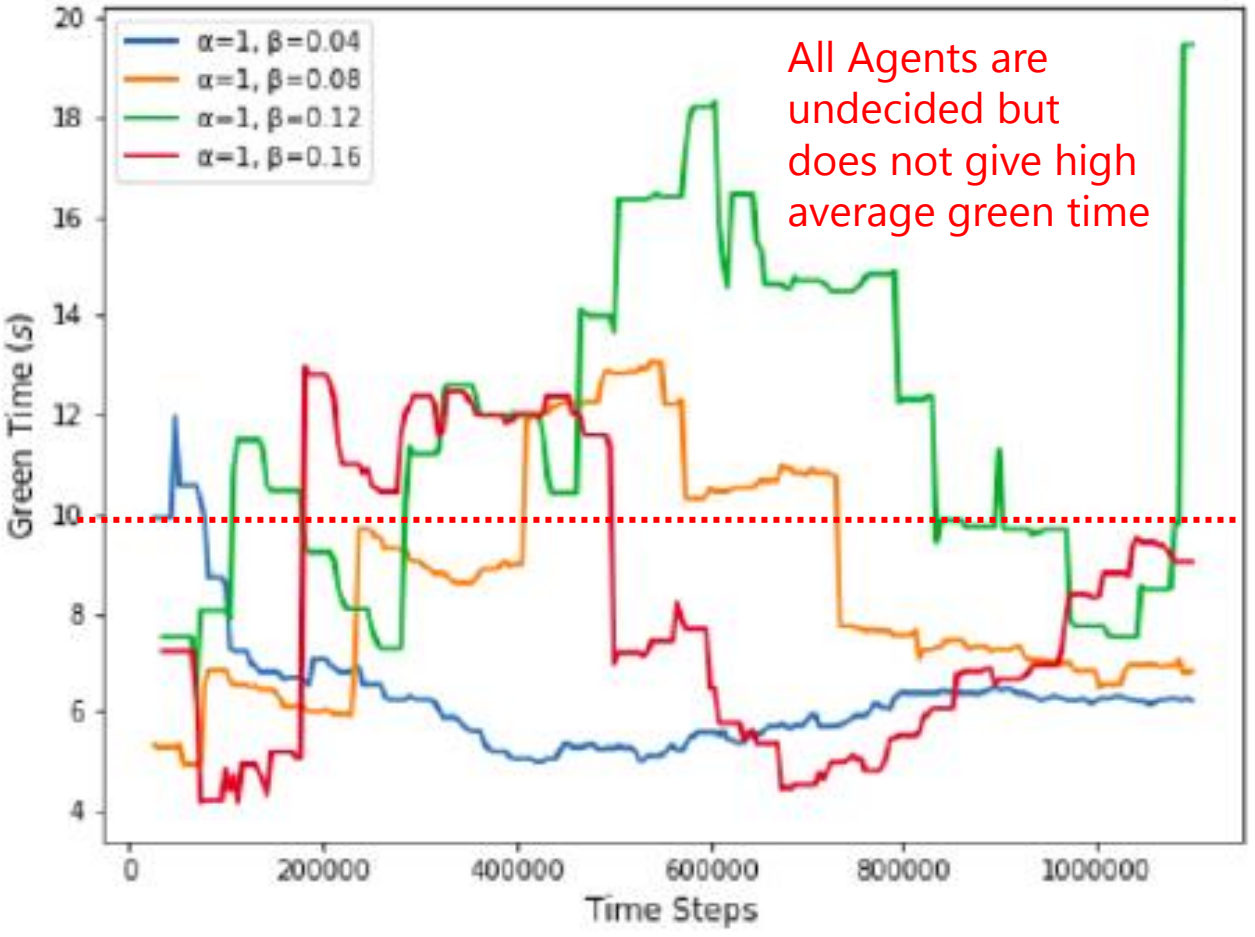


Figure: Average Green Time of Phase 2



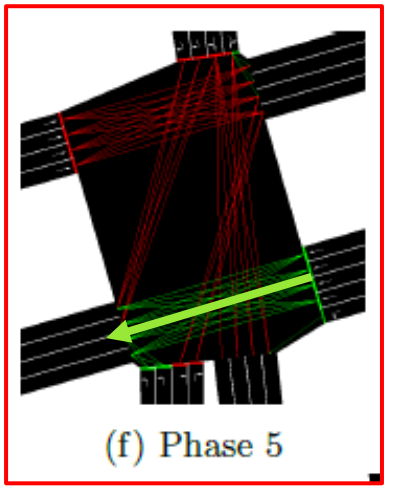
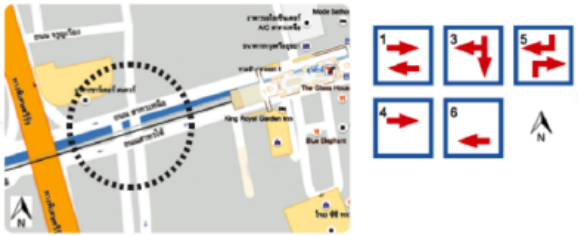


Figure: Average Green Time of Phase 5

Phase 3 and 5:
Surasak Upstream

Phase 1:
N. and S. Sathorn Upstream



Standard Phase 1-3-1-5

Changing Phase from 1 to 3:

- When the queue of downstream North Sathorn reaches the Sathorn Intersection
- When the queue of downstream South Sathorn reaches the Sathorn Intersection
- Queue of Si Wiang Road reaches Pramuan Road on Bangkok Christian College and Assumption Convent School
- Vehicles on Taksin Bridge 300 meters from Sathorn Intersection is starting to move
- Phase 1 duration more than 120-150 seconds

Changing Phase from 3 to 1:

- Reduced jam length of Si Wiang Road or vehicles are moving on Pramuan Road continuously for 20-30 seconds
- Minimum gap between vehicles that crosses the intersection is too high
- Velocity of the vehicles that crosses the intersection is too low
- Phase 3 duration more than 30-80 seconds

Changing Phase from 1 to 5:

- Queue of Si Wiang Road reaches Pramuan Road on Bangkok Christian College and Assumption Convent School
- Queue of CharoenRat is too long
- Phase 1 duration more than 120-150 seconds

Changing Phase from 5 to 1:

- Reduced jam length of CharoenRat Road
- Phase 5 duration more than 40-50 seconds

Note: Use Phase 1-3-1-3-5 if want to get cars out from Surasak and Pramuan Road. Use Phase 4 (instead of Phase 1) when the head of queue from Sathorn South reaches Sathorn Intersection. Use Phase 6 (instead of Phase 1) when the head of queue from Sathorn North reaches Sathorn Intersection.

Figure 4: Heuristic Signal Actuated Logics at the Morning Rush Hour of Surasak Intersection [3] Version 23/11/2016 Yannawa Police District

Heuristic Signal Actuated Logics

Phase 1→3, Phase 1→5

Triggering Event:
Si Wiang Queue Length

Phase 3→1, Phase 5→1

Triggering Event:
Never about South Sathorn Queue Length

Chaodit Aswakul, Sorawee Watarakitpaisarn, Patrachart Komolkiti, Chonti Krisanachantara, and Kittiphan Techakittiroj. Chula-SSS: Developmental Framework for Signal Actuated Logics on SUMO Platform in Over-Saturated Sathorn Road Network Scenario. In *SUMO 2018- Simulating Autonomous and Intermodal Transport Systems*, volume 2 of EPiC Series in Engineering, pages 67–81. EasyChair, 2018

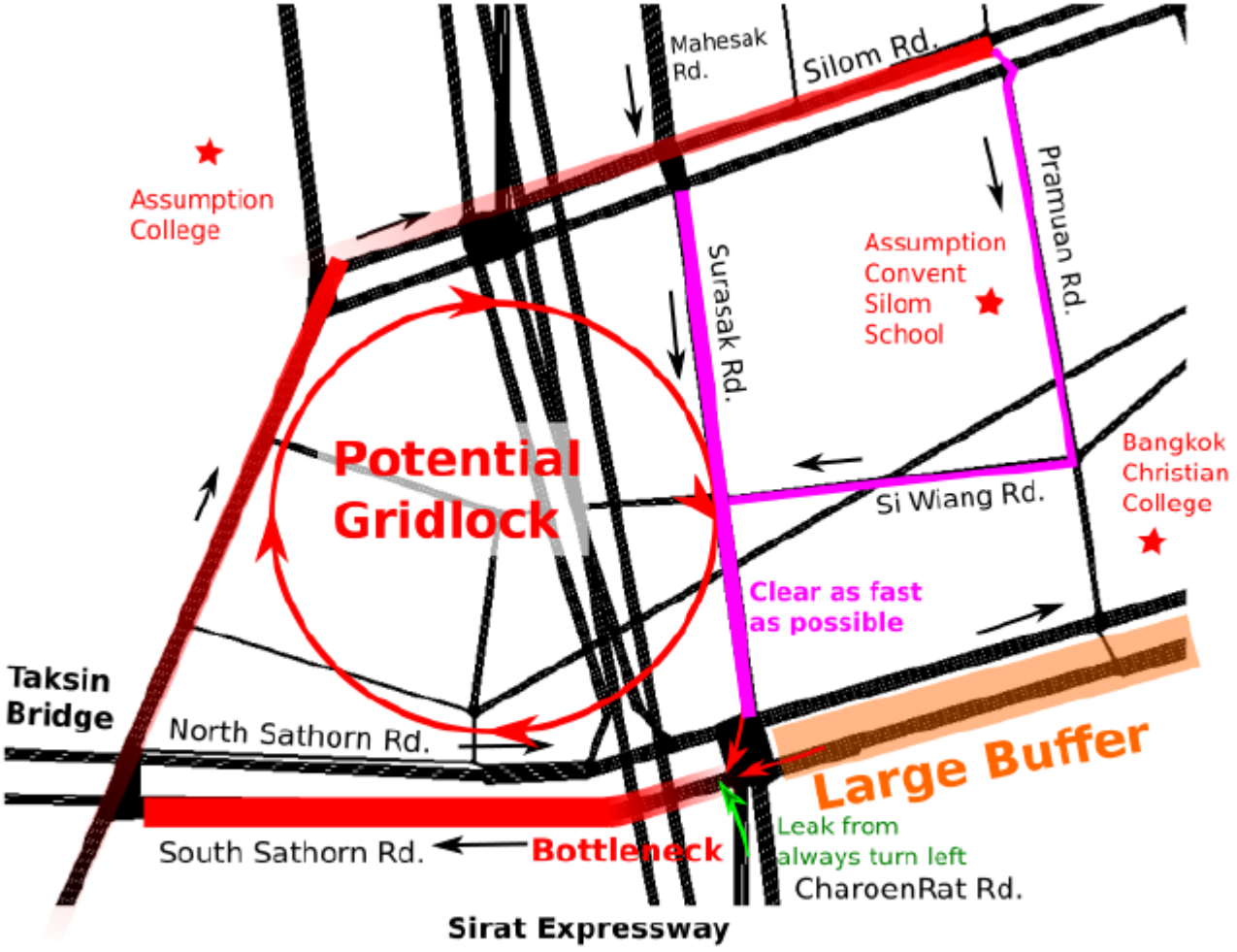
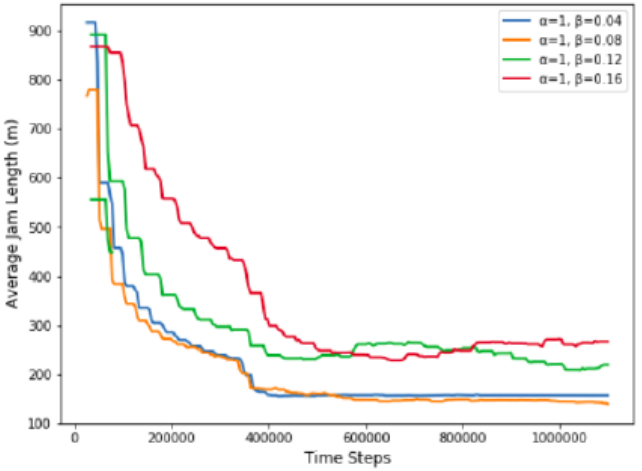
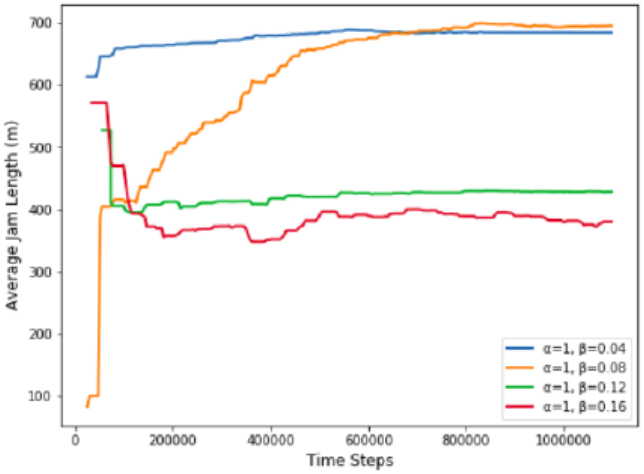


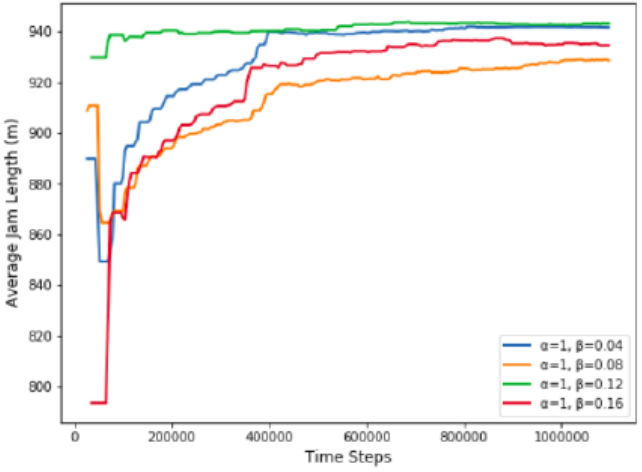
Figure 10: Potential Visual Policy of the Sathorn Road Network



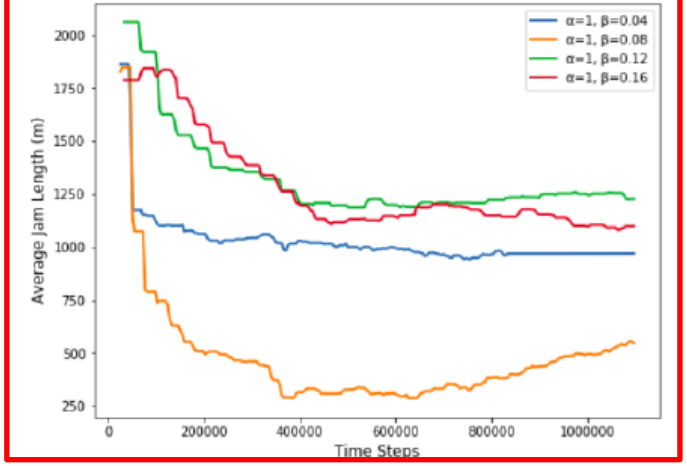
(a) Surasak Jam Length



(b) CharoenRat Jam Length



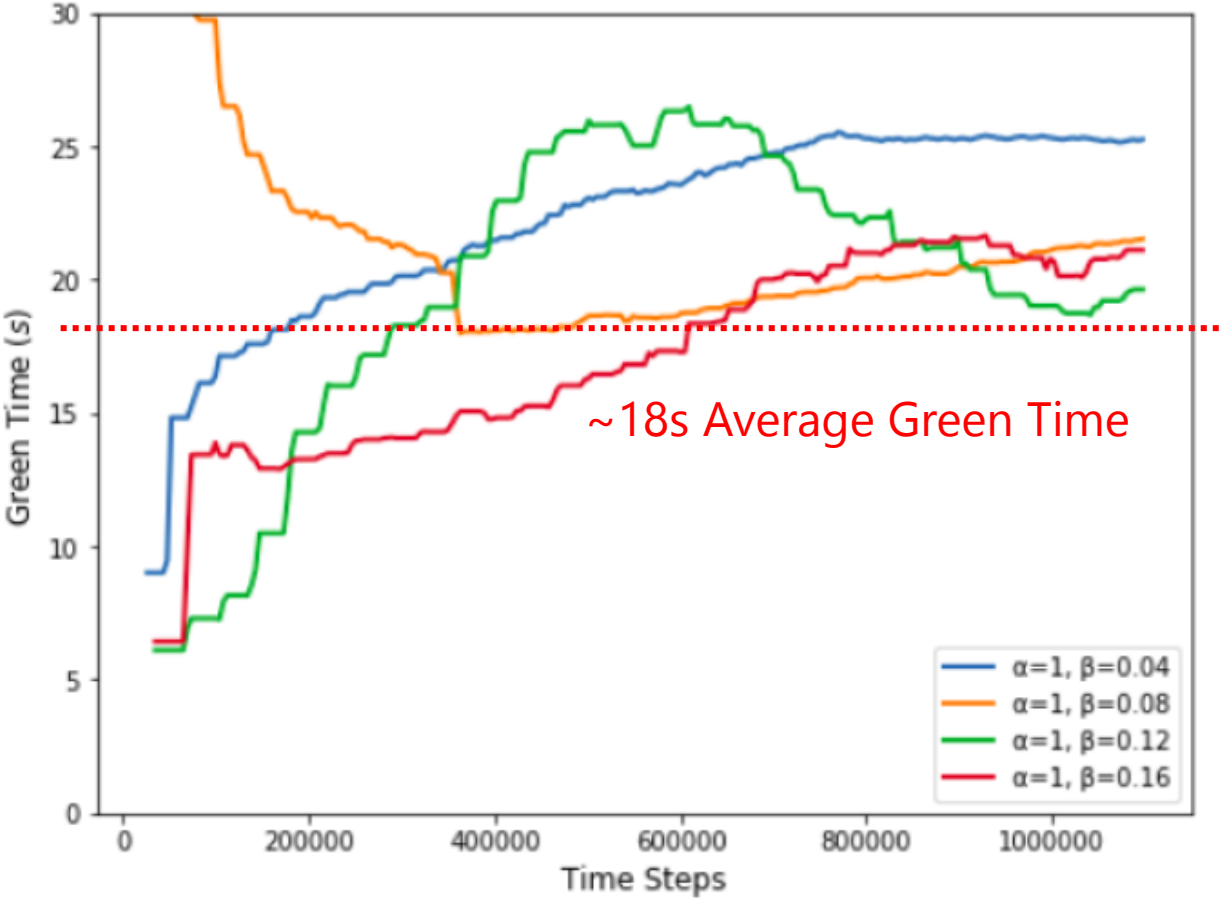
(c) South Sathorn Jam Length



(d) North Sathorn Jam Length

Slightly Decreasing Trend

Figure 11: Jam Length of the Approaching Lanes



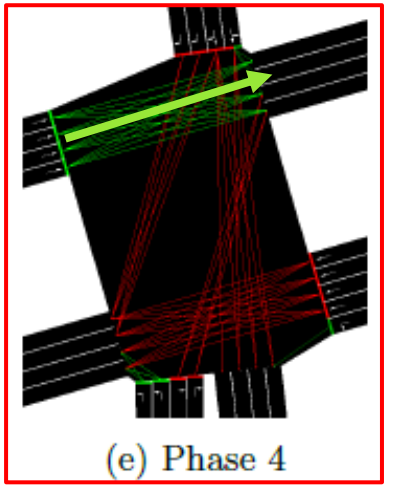
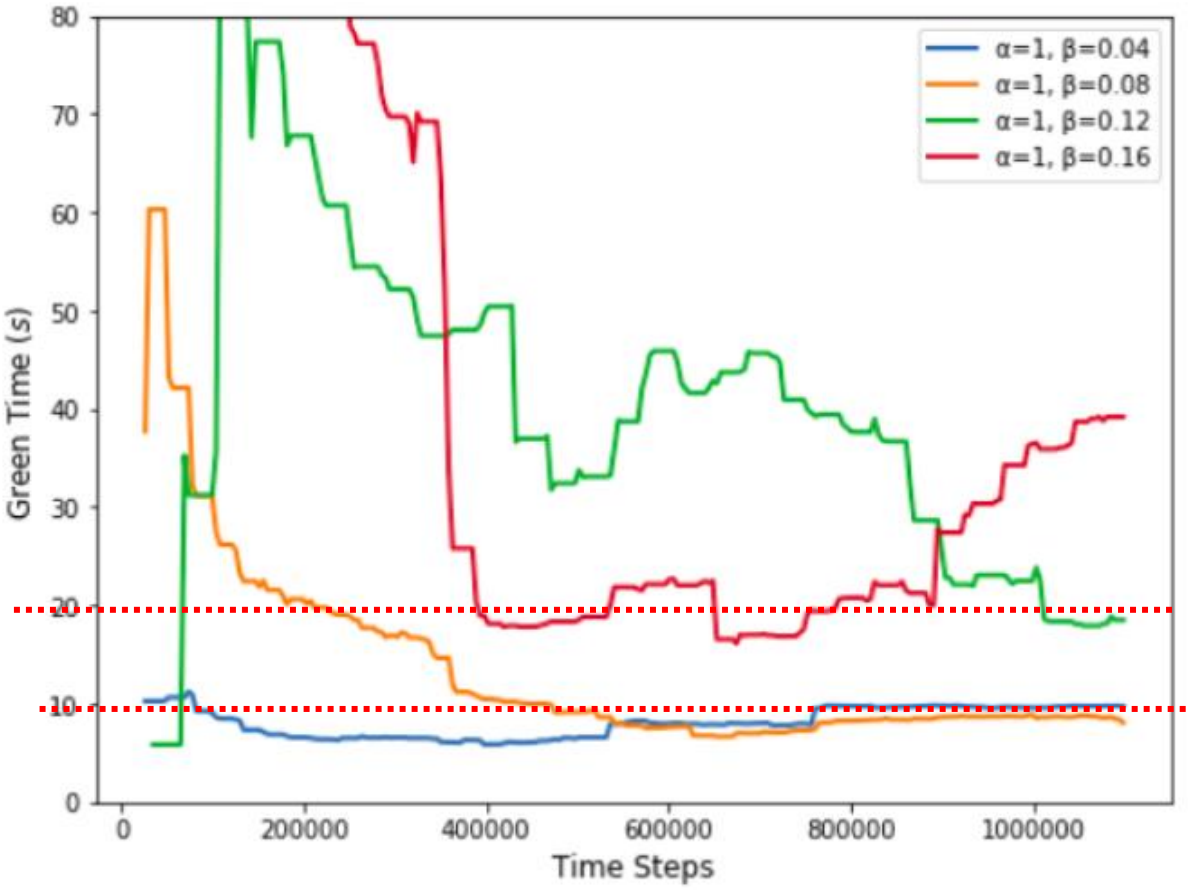
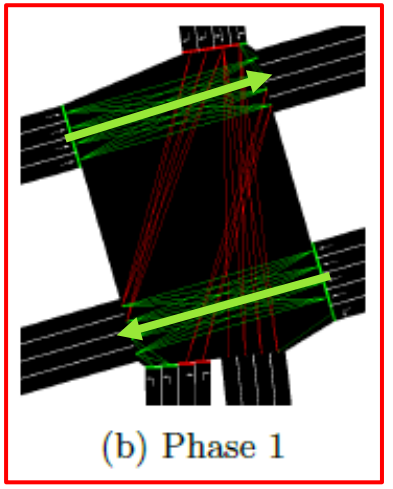


Figure: Average Green Time of Phase 4

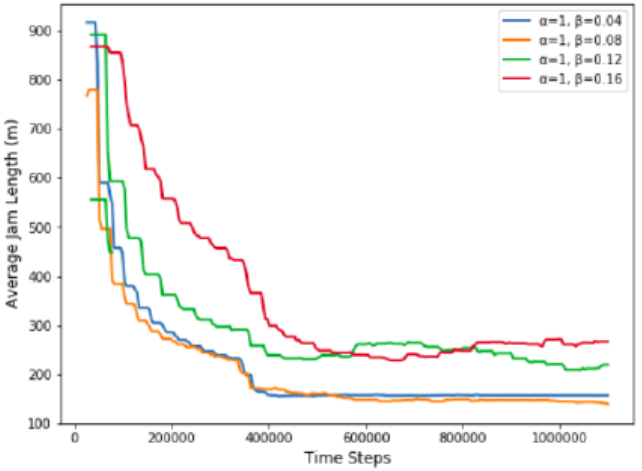




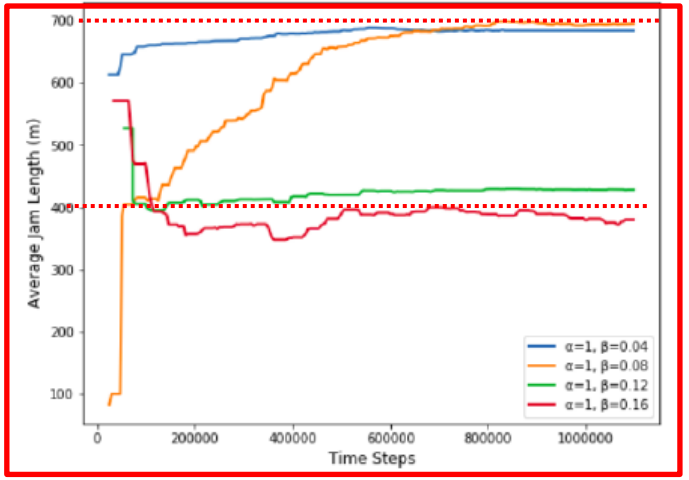
>20s Average Green Time

~10s Average Green Time

Figure: Average Green Time of Phase 1



(a) Surasak Jam Length

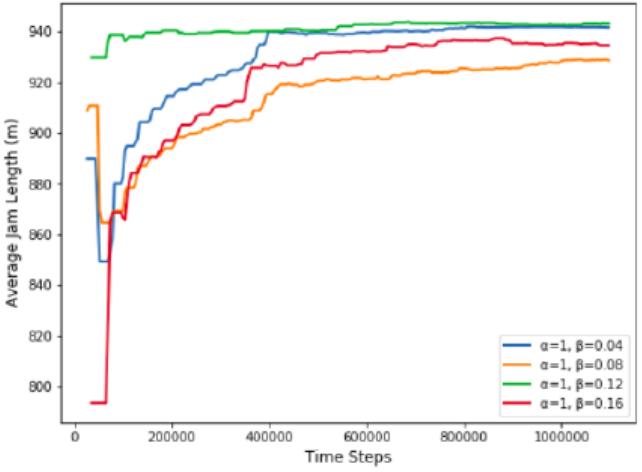


(b) CharoenRat Jam Length

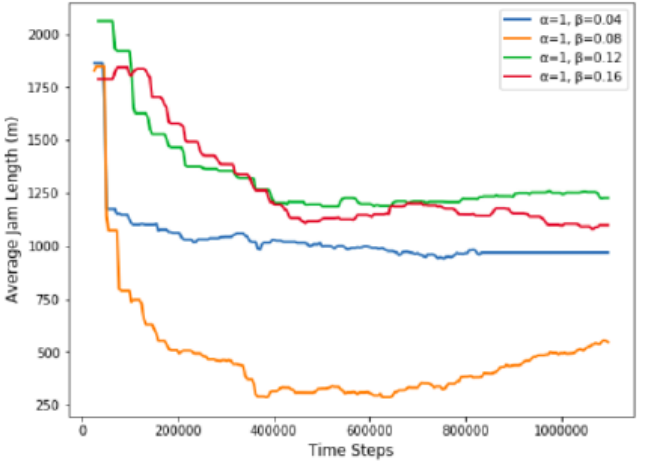
~700m

~400m

2 Different Convergence



(c) South Sathorn Jam Length



(d) North Sathorn Jam Length

Figure 11: Jam Length of the Approaching Lanes

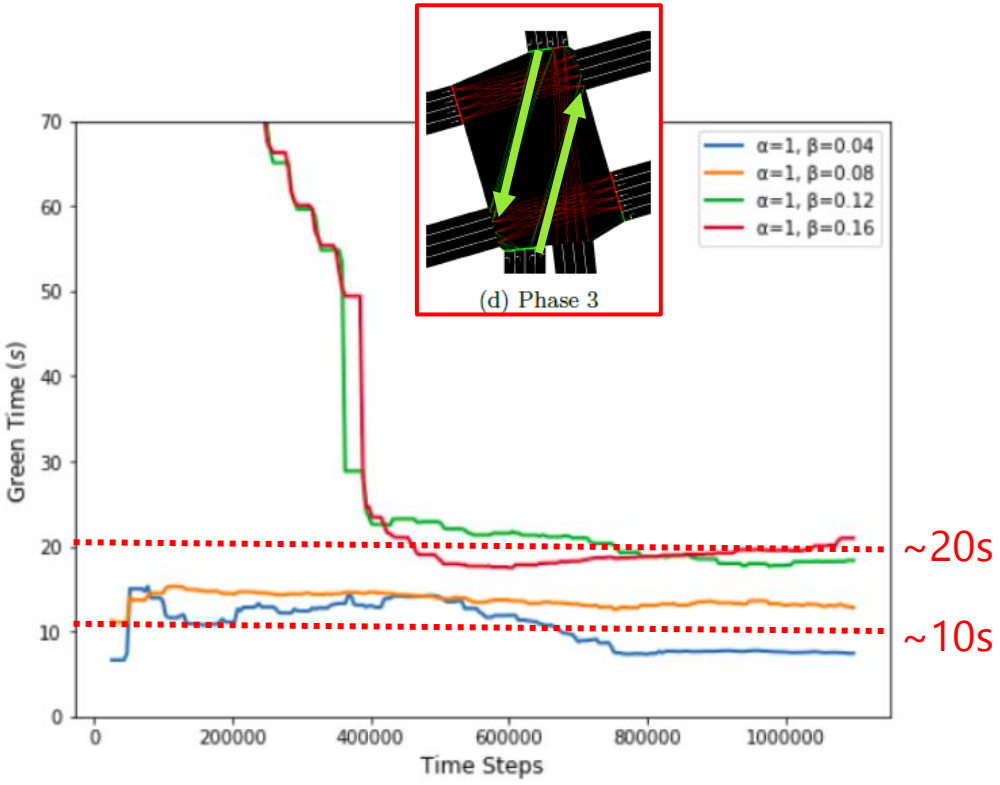


Figure: Average Green Time of Phase 3

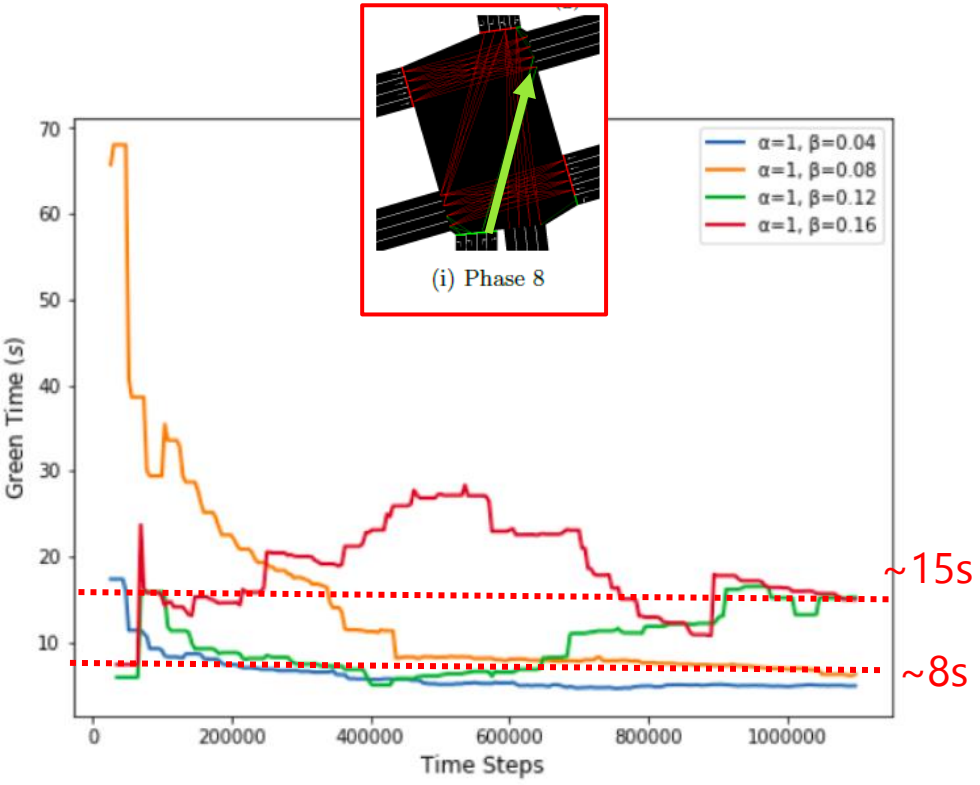


Figure: Average Green Time of Phase 8

Conclusion and Future Works

- Ablation studies on hyperparameters
- Varying load of the extended flows
- Varying agent's discount factor
- Varying types and number of sensory inputs
- Exploring on different rewards