EcoLight: Reward Shaping in Deep Reinforcement Learning for Environment Friendly Traffic Signal Control

SUMO implementation

THE TEAM

Pedram Agand
Research assistant

Mo Chen
Assistant professor

SFU

Alexey Iskrov
CTO

Alexander Kurtynin
CEO

Breeze Traffic Co.

OUR MOTIVATION

- *124 B$ per year for traffic congestion*
- *1% cost of European union's GDP*
- *3M death due to air pollution, 33% due to traffic*
- *3-4 days/year of time waste in traffic in NY, SF*
- *Better traffic control in china save 15.3% of delay*

3

INDUSTRY OUTLOOK

| Uniform Fixed time	Actuated traffic signal controller	Webster	SOTL	Learning approachs

SUCCESS

Agent design

Environment: intersection

Interpreter

Objective

Action: Green phase

State

Reward: Weighted scheme

Agent: A2C, Sarsa
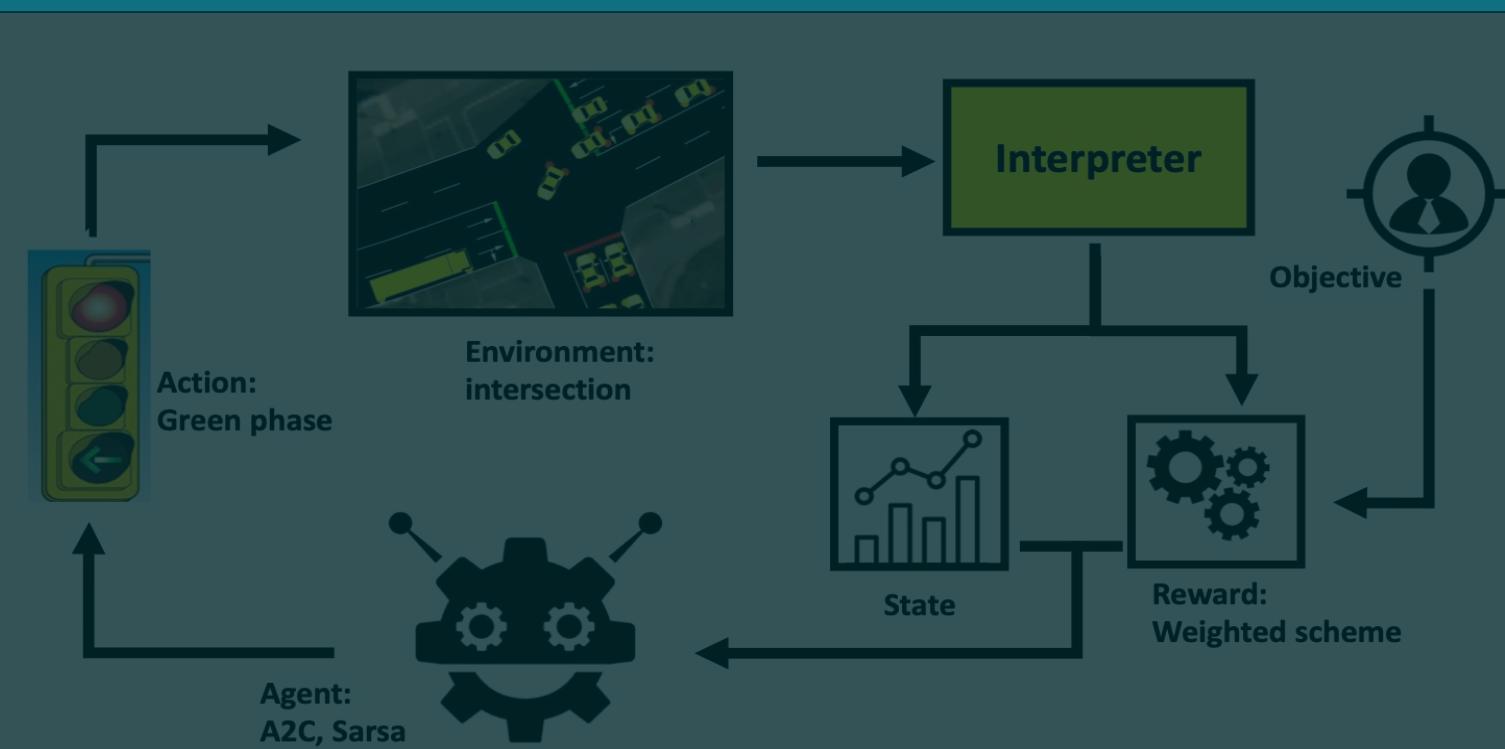
- **State**: *most recent green phase, density, queue length, type of vehicle of incoming lane*

- **Action**: *the next green phase*

Reward Shaping

✓ **Queue length**

$$R_q = -\left(\sum_{j \in L_{in}} N_{Hj}\right)^2$$

Weighted version:

$$R_{wq} = -\left(\sum_{j \in L_{in}} N_{wHj}\right)^2, \quad N_{wHj} = \sum_{k=1}^{N_{Hj}} W_k$$

✓ **Waiting time**

$$R_w = 0.01 \sum_{j \in L_{in}} \left(T_{j,t} - T_{j,t-1}\right)$$

Weighted version:

$$R_{ww} = 0.01 \sum_{j \in L_{in}} \left(T_{wj,t} - T_{wj,t-1}\right), \quad T_{wj} = \sum_{k=1}^{N_j} W_k T_{jk}$$
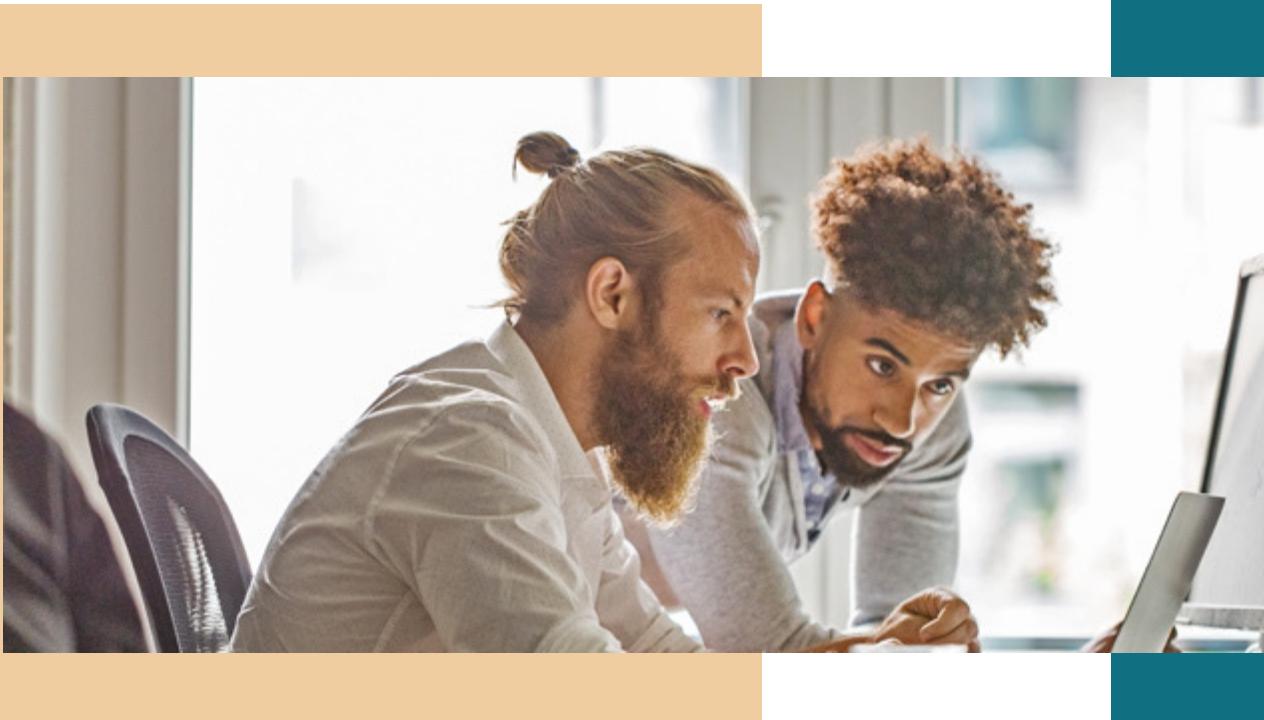
✓ **Pressure**

$$R_p = -\left|\sum_{j \in L_{in}} N_j - \sum_{j \in L_{out}} N_j\right|$$

Weighted version:

$$R_{wp} = -\left|\sum_{j \in L_{in}} N_{wj} - \sum_{j \in L_{out}} N_{wj}\right|, \quad N_{wj} = \sum_{k=1}^{N_j} W_k$$

Weight selection

Constant

Optimize a constant multiplier based of the vehicle emission class type

Normalized emission

Adaptive weight for all vehicle in a lane based of their normalized emission
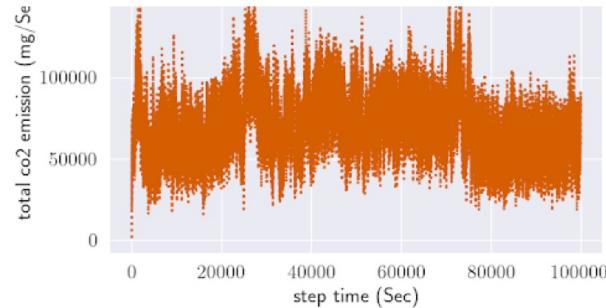
$$W_j = \frac{\mathcal{E}_j - \bar{\mathcal{E}}}{\mathcal{E}_{max} N_j}$$
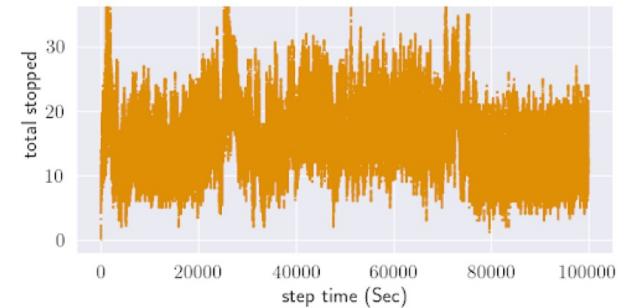
Adaptive weights

*Adaptive weight for each vehicle based of its current emission*
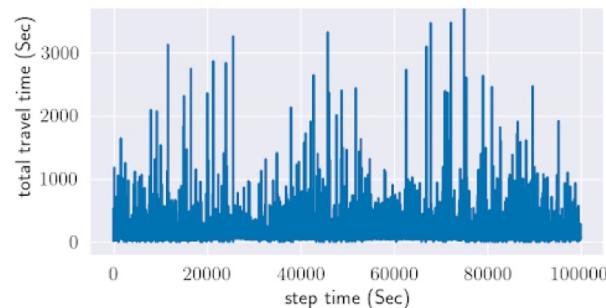
Experiment
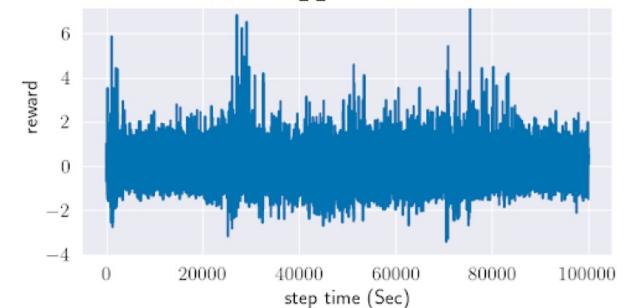
(a) CO2 emission

(b) Stopped time

(c) travel time
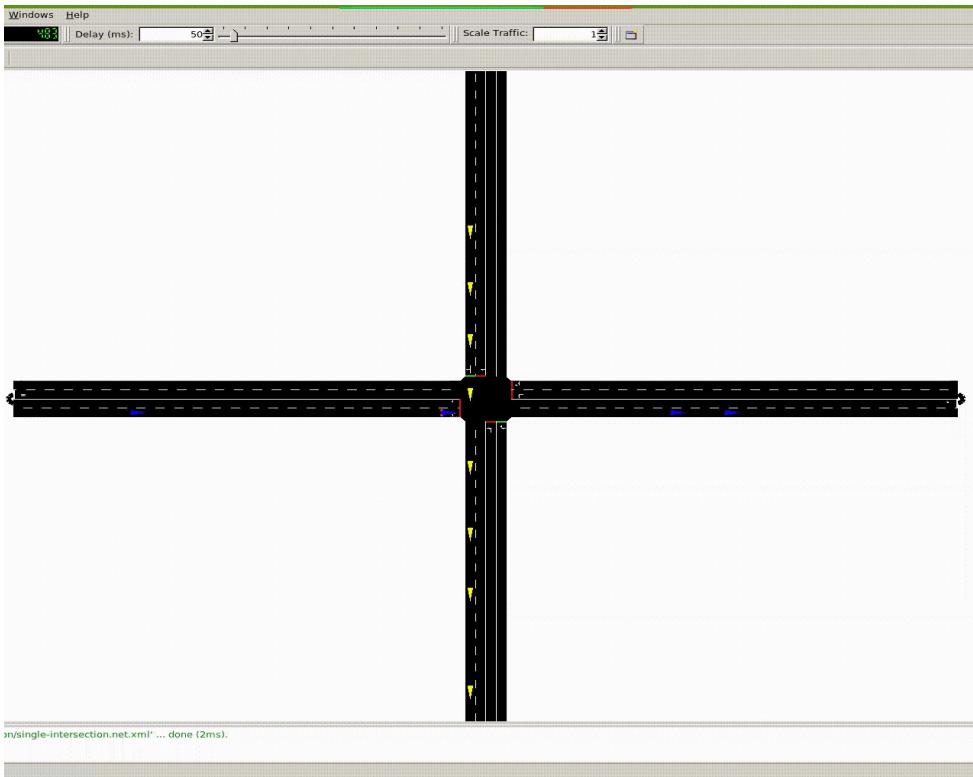
(d) Reward
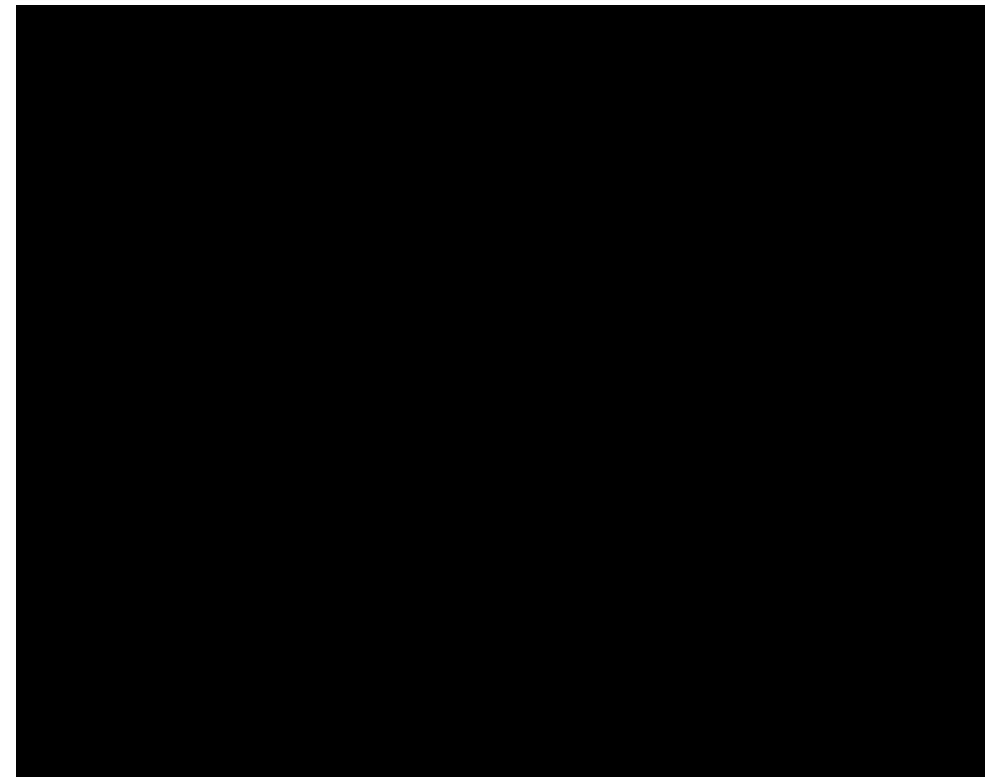
Metric	Type	Fixed time	Waiting time		Queue length		Pressure	
			a2c	sarsa	a2c	sarsa	a2c	sarsa
Travel	not weighted	226.34	162.40	125.67	224.11	157.38	248.43	210.06
	weighted		153.64	110.91	229.43	164.34	262.48	236.36
CO2	not weighted	149.76	113.48	84.11	145.45	111.26	135.85	128.35
	weighted		101.29	69.98	123.43	84.96	140.19	119.79
Wait	not weighted	15337	2371	1091	5365	7442	5025	15665
	weighted		2117	788.06	4878	5138	6544	11109
Stop	not weighted	32.70	24.24	17.55	31.62	22.08	31.95	30.16
	weighted		23.14	15.57	30.27	23.80	33.76	35.77
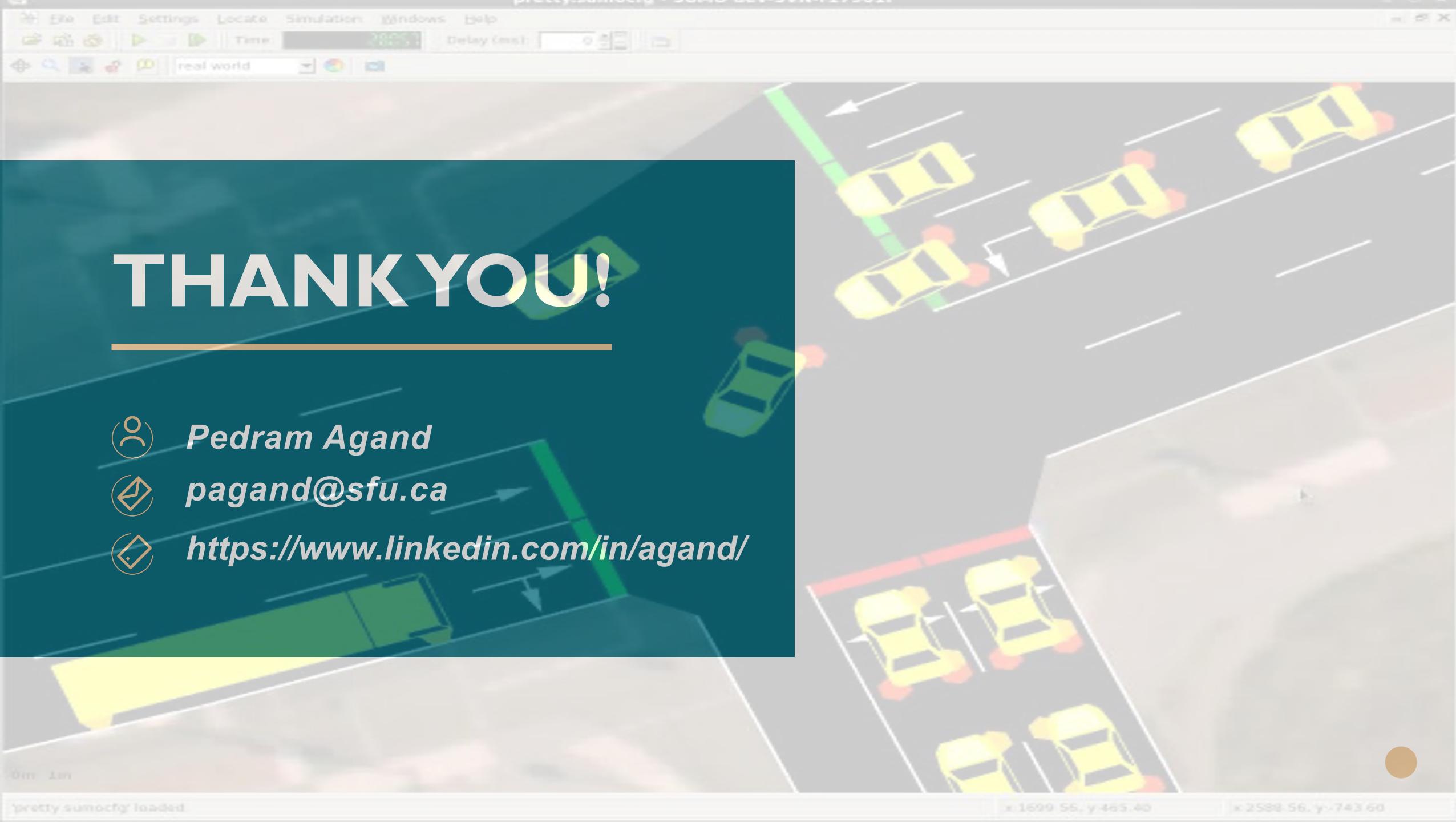
Comparison:

Weighted waiting time (Sarsa):

Fixed time:

THANK YOU!

👤 *Pedram Agand*

✉ *pagand@sfu.ca*

🔖 *https://www.linkedin.com/in/agand/*