SISTEM KONTROL SINYAL LALU LINTAS ROBAS DENGAN *FRAMEWORK* RADIAL MENGGUNAKAN *DEEP Q NETWORK* (DQN) DAN *PROXIMAL POLICY OPTIMIZATION* (PPO)

Aditya Girza Utama (23823002) | Pembimbing : Prof. Ir. Endra Joelianto. Ph.D & Faqihza Mukhlish, S. T., M. T., Ph.D

Pendahuluan

Kemacetan & Tantangan Manajemen Lalu Lintas

Pesatnya perkembangan teknologi & ilmu pengetahuan memberikan dampak yang signifikan pada sistem transportasi
➔ **meningkatnya volume kendaraan**

Masalah muncul saat pengaturan lampu lalu lintas tidak menyesuaikan dengan jumlah kendaraan
➔ **kemacetan & waktu tunggu lebih lama**

Manajemen Lalu Lintas Modern

➔Menciptakan sirkulasi lalu lintas aman, efektif, mengurangi kemacetan & kecelakaan.
- **Teknologi modern**
  ⟶ Machine learning & AI digunakan untuk membantu manajemen.
- **Tantangan**
  ⟶ Prediksi buruk, manajemen arus kurang optimal, kesulitan manajemen darurat

***Reinforcement Learning* untuk Kontrol Lalu Lintas**

digunakan untuk mengatur sinyal lalu lintas secara adaptif dengan cara belajar dari interaksi lingkungan, sehingga dapat mengoptimalkan arus kendaraan dan mengurangi kemacetan secara *real-time*.

Keamanan *Reinforcement Learning*
- ➔ *Reinforecement Learning* sudah sukses di banyak bidang, tetapi tetap rentan terhadap gangguan/*adversarial perturbations*.
- ➔ *Deep RL agent* : perlu algoritma pelatihan yang kuat agar tahan terhadap serangan & *noise*.
- ➔ Diperlukan pengembangan algoritma robas RL (DQN, PPO) menggunakan metode RADIAL-RL

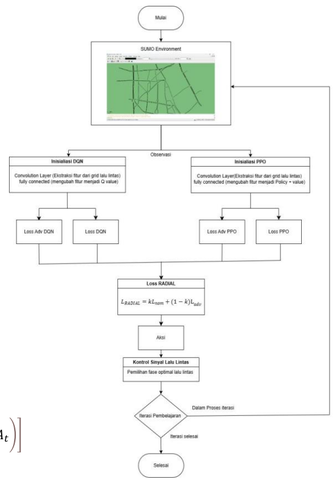
Tujuan Penelitian

Mengembangkan dan menguji sistem Robas pada algoritma *Deep Q Network* (DQN) dan *Proximal Policy Optimization* (PPO) untuk sistem lalu lintas agar dapat meningkatkan efisiensi, efektivitas dan tahan terhadap serangan *adverserial* dalam mengendalikan arus lalu lintas dengan menggunakan metode (Robas *Adversarial Loss*) RADIAL-RL khususnya di kota Jakarta

Metodologi

Arsitektur Sistem Kontrol

Diagram menunjukkan alur kontrol lalu lintas dengan agen DQN dan PPO di lingkungan SUMO. Data observasi diproses menjadi aksi melalui pembelajaran dengan loss RADIAL, yang menggabungkan *loss* normal dan *adversarial* untuk ketahanan terhadap serangan. Aksi digunakan untuk memilih fase sinyal lalu lintas secara optimal, dan proses berulang hingga iterasi pembelajaran selesai.



Formula yang digunakan pada penelitian ini :

➔ *Loss* DQN

$$l\,(\theta_{act}) = E_{(s,a,s',r)}\left[\left(r + \gamma \max_{a'} Q_{trg}(s', a'; \theta_{trg}) - Q_{act}(s, a; \theta)\right)^2\right]$$

➔ *Loss Adverserial* DQN

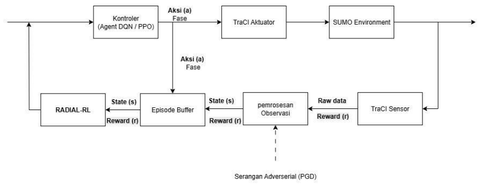
$$L_{adv}(\theta, \epsilon) = E_{(s,a,s',r)}\left[\sum_y Qdiff(s, y)\,.\,Ovl(s', y', \epsilon)\right]$$

➔ *Loss* PPO

$$l(\theta) = E_{(s_t,a_t,r_t)}\left[-min\left(\frac{\pi(a_t|s_t\,\theta;)}{\pi(a_t|s_t\,\theta_{old};)}A_t, clip\left(\frac{\pi(a_t|s_t\,\theta;)}{\pi(a_t|s_t\,\theta_{old};)}, 1-\eta, 1+\eta\right)A_t\right)\right]$$

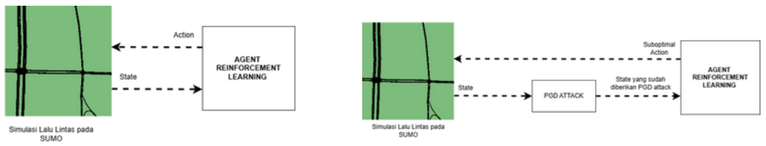
➔ *Loss Adverserial* PPO

$$l_{adv} = E_{(s_t,a_t,r_t)}\left[-min\left(\frac{\tilde{\pi}(a_t|s_t\,\theta;)}{\pi(a_t|s_t\,\theta_{old};)}A_t, clip\left(\frac{\tilde{\pi}(a_t|s_t\,\theta;)}{\pi(a_t|s_t\,\theta_{old};)}, 1-\eta, 1+\eta\right)A_t\right)\right]$$



Sistem ini terdiri dari agen DQN/PPO yang mengontrol fase lalu lintas berdasarkan data observasi dari SUMO. SUMO menghasilkan data kepadatan lalu lintas yang dibaca oleh TraCI sensor, lalu diterjemahkan kembali ke dalam aksi oleh agen. Serangan PGD ditambahkan untuk menguji ketahanan agen, dan metode RADIAL-RL digunakan agar agen tetap optimal meskipun ada gangguan.

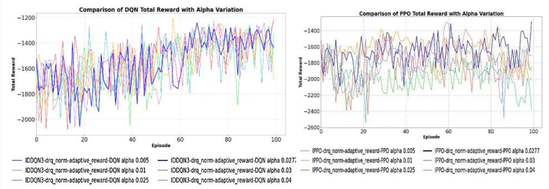
Skema Pengujian



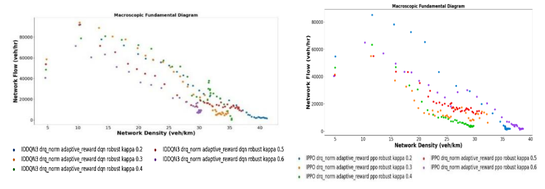
No	Model	Tipe Pelatihan	Robustifikasi (kappa (κ))	PGD *Attack* (epsilon (ϵ))	Tujuan Pengujian
1 2	DQN PPO	Tanpa Robas	-	-	Evaluasi performa awal
3 4	DQN PPO	Tanpa Robas	-	0.05	Uji pengaruh Serangan
5 6	DQN PPO	Robas	0.2, 0.3, 0.4, 05, 0.6	-	Uji pengaruh robas
7 8	DQN PPO	Robas	0.2, 0.3, 0.4, 05, 0.6	0.05,0.15,0.25,0.35, 0.45	Evaluasi robustness dan resistance

Hasil dan Analisa

Parameter Testing

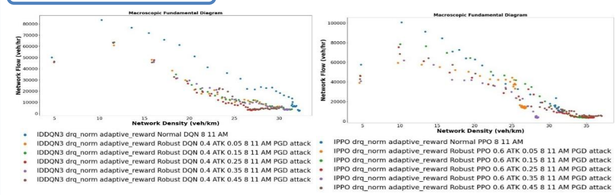


➔kondisi perlakuan kontrol dengan menggunakan nilai alpha 0.0277 (garis warna biru tebal) menunjukan hasil terbaik, untuk DQN maupun PPO



➔hasil pengujian terhadap beberapa nilai kappa untuk DQN dan PPO, menunjukan bahwa performa terbaik adalah nilai 0,4 (titik warna hijau) untuk DQN dan 0,6 (titik warna ungu) untuk PPO.

Uji Coba Sistem



perbandingan performa antara Robas DQN dan Robas PPO dengan pemberian serangan yang beragam. Dari setiap variasi nilai serangan yaitu, 0.05, 0.15, 0.25, 0.35 dan 0.45 semua menunjukan bahwa performa Robas PPO berada diatas performa Robas DQN, terutama untuk angka dari *throughput*.

Penilaian	Tingkat Serangan	Robas PPO	Robas DQN	Selisih nilai DQN dan PPO	Lebih unggul
Throughput (veh)	0.05	1,066,368	513,384	552,984	PPO
	0.15	720,948	475,956	244,992	PPO
	0.25	599,976	419,664	180,312	PPO
	0.35	584,376	497,784	86,592	PPO
	0.45	612,288	504,216	108,072	PPO
Max flow (veh/hr)	0.05	59,436	60,840	- 1,404	DQN
	0.15	78,480	63,648	14,832	PPO
	0.25	75,360	63,432	11,928	PPO
	0.35	61,932	63,036	-1,104	DQN
	0.45	65,580	63,456	2,124	PPO
Peak flow (veh/hr)	0.05	1,224	1,212	12	PPO
	0.15	1,200	1,248	-48	DQN
	0.25	1,200	1,236	-36	DQN
	0.35	1,176	1,176	0	-
	0.45	1,188	1,248	-60	DQN

Ketahanan Sistem

Kondisi	SERANGAN					RATA-RATA
	ATK 0.05	ATK 0.15	ATK 0.25	ATK 0.35	ATK 0.45	
Rata-rata reward tanpa serangan (dari 100 data)	-6876,88	-6876,88	-6876,88	-6876,88	-6876,88	-6876,88
Rata-rata reward dengan serangan (dari 100 data)	-6710,88	-6611,31	-6526,28	-6594,64	-6538,32	-6596,29
Penurunan performa akibat serangan	2,43%	3,86%	5,11%	4,10%	4,92%	4,08%
Persentase Ketahanan sistem	97,57%	96,14%	94,89%	95,90%	95,08%	95,92%

Kondisi	SERANGAN					RATA-RATA
	ATK 0.05	ATK 0.15	ATK 0.25	ATK 0.35	ATK 0.45	
Rata-rata reward tanpa serangan (dari 100 data)	-7747,68	-7747,68	-7747,68	-7747,68	-7747,68	-7747,68
Rata-rata reward dengan serangan (dari 100 data)	-7875,12	-7616,71	-7455,98	-7704,97	-7676,57	-7665,87
Penurunan performa akibat serangan	-1,64%	1,69%	3,76%	0,55%	0,92%	1,38%
Persentase Ketahanan sistem	101,64%	98,31%	96,24%	99,45%	99,08%	98,94%

Perbandingan ketahanan sistem metode Robas DQN dan Robas PPO yang diberi serangan, dapat dilihat pada algoritma DQN, rata-rata penurunan performa sebesar 4.08%, dengan tingkat ketahanan sistem mencapai 95.92%. Sementara , algoritma PPO menunjukkan rata-rata penurunan performa sebesar 1.38% dan ketahanan sistem sebesar 98.94%. Penurunan *reward* pada kedua algoritma menunjukan bahwa serangan *adversarial* berhasil mempengaruhi kebijakan pengambilan keputusan yang telah dipelajari. Meskipun demikian, ketahanan di atas 90% menunjukkan bahwa baik DQN maupun PPO memiliki kemampuan untuk mempertahankan stabilitas kinerja secara umum dalam menghadapi gangguan input bersifat *adversarial*.

Kesimpulan

1. Nilai parameter alpha (α) optimal untuk DQN dan PPO adalah 0.0277.
2. Nilai kappa (κ) optimal untuk DQN adalah 0.4 dan PPO adalah 0.6.
3. PPO normal memiliki performa lebih baik dari pada DQN normal, dibuktikan dengan *throughput* PPO normal sebesar 1.096.680, sedangkan DQN normal sebesar 849.096.
4. PPO Robas dengan serangan memiliki performa lebih baik dari pada DQN.
5. Persentase ketahanan sistem (*Robustness*) DQN adalah 95.92% dan PPO 98.94%, sehingga layak digunakan pada sistem pembelajaran penguatan di lingkungan yang rentan terhadap gangguan eksternal