Berdasarkan analisis mendalam terhadap referensi terbaru (2024-2025) yang Anda berikan, saya mengusulkan sebuah framework yang menggabungkan keunggulan **PPO**, **Dueling Network**, dan mekanisme **Safe RL**.
Framework ini dirancang untuk mengatasi kelemahan DQN (kurang stabil di continuous space) dan PPO standar (kadang lambat konvergen), serta secara eksplisit menangani **Critical Port** (Port 4 Healthcare) Anda.

Usulan Framework: **"Safe-Dueling Hybrid PPO (SDH-PPO)"**

Nama framework: **Safe-Dueling Hybrid PPO (SDH-PPO)**.Framework ini menggabungkan tiga komponen utama untuk mencapai stabilitas tinggi dan jaminan keamanan pada Critical Port:

1. **Algoritma Inti: Hybrid PPO (H-PPO)** untuk menangani aksi *hybrid* (diskrit untuk memilih port mana yang diatur, kontinu untuk menentukan besaran bandwidth/rate) 1.
2. **Arsitektur Kritik: Dueling Network** diintegrasikan ke dalam bagian *Critic* dari PPO untuk estimasi nilai state yang lebih presisi dan stabil 2.
3. **Mekanisme Keamanan: Safety Layer (Action Projection)** untuk memproteksi Critical Port dari keputusan agen yang melanggar SLA 3, 4.

1. Metodologi & Algoritma (Mengapa Lebih Stabil?)

Berikut adalah detail teknis mengapa kombinasi ini lebih stabil daripada DQN atau PPO biasa, didukung oleh referensi Anda:

A. Mengapa Hybrid PPO (H-PPO)?

Traffic jaringan seringkali membutuhkan keputusan bertingkat: *Port mana yang harus diubah konfigurasi polisinya?* (Diskrit) dan *Berapa rate barunya?* (Kontinu).

- **Masalah DQN:** DQN hanya bekerja baik pada aksi diskrit. Jika Anda mendiskritisasi bandwidth (misal: naik 1 Mbps, naik 2 Mbps), ruang aksi meledak dan akurasi turun 5.
- **Masalah PPO Standar:** PPO murni biasanya fokus pada satu tipe aksi.
- **Solusi H-PPO:** Referensi **1** dan **6** menunjukkan bahwa **Hybrid PPO** menggunakan *multiple policy heads* (satu untuk probabilitas memilih port, satu untuk distribusi Gaussian nilai bandwidth). Ini memungkinkan agen mengontrol topologi (switch port) dan parameter trafik secara bersamaan tanpa kehilangan presisi.

B. Mengapa Integrasi Dueling Network pada Critic?

Biasanya *Dueling* digunakan pada DQN. Namun, Anda bisa mengusulkan **inovasi** dengan memasang struktur *Dueling* pada jaringan *Critic* di dalam PPO.

- **Konsep:** Memisahkan estimasi *State Value* $V(s)$ dan *Advantage* $A(s,a)$.
- **Keunggulan:** Referensi **2** dan **7** menjelaskan bahwa arsitektur Dueling sangat efektif saat banyak state memiliki nilai yang mirip, terlepas dari aksi apa yang diambil (contoh: saat jaringan sepi, mengubah *policing rate* tidak berdampak banyak). Ini mencegah fluktuasi nilai Q yang drastis, memberikan gradien yang lebih stabil ke *Actor* PPO.

C. Penanganan Critical Port: "Safety Layer"

Ini adalah kontribusi terbesar Anda untuk Port 4 (Healthcare). Jangan hanya mengandalkan *reward penalty* (karena agen RL butuh waktu untuk belajar dari penalti, dan di awal training ia pasti melanggar SLA).

- **Metode:** Gunakan **Safety Layer** berbasis *Constraint Optimization* atau *Projection*.
- **Implementasi:** Setelah agen H-PPO mengeluarkan aksi (misal: kurangi bandwidth Port 4 menjadi 2 Mbps), *Safety Layer* memeriksa apakah aksi ini akan melanggar latensi kritis (berdasarkan prediksi model sederhana atau heuristik). Jika ya, aksi tersebut **diproyeksikan** ke nilai aman terdekat (misal: batas minimum 10 Mbps).
- **Referensi Pendukung:** Konsep **SafeSlice** 3, 4 menggunakan mekanisme serupa di O-RAN untuk menjamin SLA latensi pada slice kritikal tidak pernah dilanggar, bahkan saat fase eksplorasi.

2. Rincian Arsitektur Framework untuk Paper

Anda bisa menyusun diagram framework di paper Anda dengan alur berikut:

- **State Observation:** 37 Fitur dari SDN Controller (seperti yang Anda sebutkan: rx_mbps, delay, loss, dll).
- **Feature Extraction (GNN/LSTM):** Gunakan layer LSTM atau GNN (seperti referensi 8, 9) untuk menangkap pola temporal trafik sebelum masuk ke RL. Ini membantu stabilitas karena agen memahami "tren" trafik, bukan hanya nilai sesaat.
- **SDH-PPO Agent (Proposed):**
- **Actor (Hybrid):** Output aksi ganda (Pilih Port & Set Rate).
- **Critic (Dueling):** Mengestimasi nilai state dengan struktur cabang $V(s)$ dan $A(s,a)$.
- **Safety Shield (Critical Port Logic):**
- *Input:* Aksi mentah dari Agent.
- *Logika:* IF Port == P4 AND Predicted_Latency > 10ms THEN Action = Safe_Action.
- *Output:* Aksi tereksekusi ke SDN Controller.
- **Environment (Mininet/Testbed):** Mengembalikan Reward dan Next State.

3. Alternatif Algoritma yang Lebih Stabil dari PPO

Jika Anda ingin alternatif selain PPO modifikasi di atas, referensi Anda menyarankan dua algoritma yang sangat kuat untuk kontrol kontinu yang stabil:

- **TD3 (Twin Delayed DDPG) 10, 11:**
- **Mengapa:** TD3 dirancang khusus untuk mengatasi *overestimation bias* pada DDPG. Ia menggunakan dua *Critic* (Twin Critics) dan mengambil nilai minimum di antara keduanya. Ini seringkali **lebih stabil daripada PPO** dalam skenario di mana kesalahan estimasi kecil bisa fatal (seperti alokasi bandwidth presisi).
- **Kelebihan:** Sangat *sample efficient* (belajar lebih cepat dari PPO).
- **SAC (Soft Actor-Critic) 12, 13:**
- **Mengapa:** SAC memaksimalkan *reward* sekaligus *entropy* (keacakan) dari polisinya. Ini membuat agen lebih tahan banting (robust) terhadap gangguan dan mencegahnya "terjebak" pada strategi jelek terlalu dini.
- **Referensi:** Studi pada **13** menunjukkan SAC dengan *Data Augmentation* sangat efektif untuk alokasi resource kooperatif.

Rekomendasi Pilihan untuk Paper:

Saya sarankan tetap pada **PPO yang dimodifikasi (SDH-PPO)** atau beralih ke **TD3 dengan Safety Layer**.

- Pilih **SDH-PPO** jika Anda ingin menonjolkan kemampuan menangani struktur aksi yang kompleks (memilih port + mengatur rate).
- Pilih **TD3** jika fokus utama Anda adalah *stabilitas murni* dan kecepatan konvergensi pada kontrol continuous.

Referensi Kunci untuk Dikutip (Format APA)

Untuk mendukung klaim metodologi Anda di paper, gunakan sitasi ini:

- **Untuk Hybrid PPO (Methodology):**
- Mengatasi ruang aksi hybrid: *Li, C., et al. (2025). Intelligent decision for joint operations based on improved proximal policy optimization* 14.
- Arsitektur Hybrid: *Luo, Z., et al. (2025). Reinforcement Learning for Traffic Signal Control in Hybrid Action Space* 15.
- **Untuk Dueling Network (Stability Contribution):**
- Keunggulan stabilitas Dueling: *Sittakul, V., et al. (2025). Intelligent congestion control in 5G URLLC Software-Defined Networks... via Reinforced Dueling Deep Q-Networks* 2, 16.
- **Untuk Konsep Safety Layer/Critical Port (Main Contribution):**
- Mekanisme proteksi SLA: *Nagib, A. M., et al. (2025). SafeSlice: Enabling SLA-Compliant O-RAN Slicing via Safe Deep Reinforcement Learning* 3.
- Penanganan prioritas trafik: *Wang, L., et al. (2025). PPO-TSC: A Proximal Policy Optimization based Traffic Signal Control* 8.