Optimizing Open RAN Energy Consumption through DRL Agents in Diverse Traffic Scenarios

Gerry Agluba
Artificial Intelligence Program
University of the Philippines Diliman
Quezon City, Philippines
gpagluba@up.edu.ph

Prospero Naval, Jr.
Artificial Intelligence Program
University of the Philippines Diliman
Quezon City, Philippines
pcnaval@up.edu.ph

***Abstract*—The rapid evolution of 5G technology has brought transformative capabilities to mobile networks. However, these advancements come with significant energy efficiency challenges due to increased frequencies, denser network topologies, and higher resource demands. This study leverages Open RAN's interoperability and vendor-neutral framework to address these challenges by constructing Deep Reinforcement Learning (DRL) agents for energy optimization in diverse traffic scenarios. Key findings reveal that different agents achieve distinct trade-offs between energy savings and throughput, making them suitable for specific deployment scenarios. For mixed traffic conditions, DQN-1 achieves a 10.71% energy reduction with a 29% throughput trade-off, making it ideal for non-critical and diverse areas. On the other hand, PPO-2 and PPO-3 demonstrate more balanced performance, with energy reductions of 13.34% and 13.26% and throughput trade-offs of 36.08% and 34.51%, respectively, for full-traffic scenarios making them better suited for critical applications on dense urban areas. Additionally, metrics such as Radio Link Failures (RLF) and ON-Cost highlight the robustness of PPO agents in maintaining network stability under bursty and full traffic conditions. While DRL agents show potential for energy optimization, further improvements are needed to outperform the benchmark in balancing throughput and energy consumption. Future work will focus on longer training durations and design refinements to enhance agent performance across diverse network scenarios.**

***Index Terms*—energy saving, reinforcement learning, open-ran**

I. Introduction

The rollout of 5G in wireless technology marks a new milestone in the mobile network industry with a new promise on (a) Enhanced Mobile Broadbance (eMBB), (b) Ultra-Reliable Low Latency Communication (uRLLC) and (c) Massive Machine Type Communcication (mMMTC). Key developments around this new paradigm-shifting technology is the introduction of Open-RAN [1] that champions inter-operability and vendor neutrality through open specifications on software-defined and virtualized network functions. Despite its obvious benefits, 5G technologies comes with new challenges in Energy Efficiency [4] primarily because of the increase frequencies, denser network elements and topologies and general increase in transmission and computing resources.

A survey by Zappone [5] highlights four key broad categories of energy-efficient 5G technologies: **resource allocation**, network planning and deployment, energy harvesting and transfer, and hardware solutions. With the increasing demands on intelligent systems on dynamic environments like that of wireless network, service providers relies on close-loop solutions that can adapt to such changes without degrading promised quality of service (QOS).

This study will be focusing on extending current solutions on near-rt RIC [1] that can control base station cells on-off features on sub-second intervals. In this study we aim to create a DRL [2] agent capable of actions that maximizes energy efficiency while minimizing throughput tradeoff. Additionally, we wanted to implement sets of solution that is diverse to different network traffic scenarios.

A. Objectives

The main objective of this project is to construct a Deep Reinforcement Learning (DRL) agent that can be deployed as a near-RT RIC controller application (xApps) in an Open RAN environment. Additionally, this study aims to

- Train different agents based on latest DRL algorithms (Deep Q-Network [12], Proximal Policy Optimization [13]).
- Evaluate trained agent performance against a baseline, and well-founded heuristic benchmark [6].

B. Scope and Limitation

While the main goal of the study is to ultimately deploy and test the agent in an actual RAN environment. This study was scoped down to using a simulated environment using an open-source RAN simulator (ns3-mmwave-oran) [11]. The simulator in its own is limited to simulate a number of base stations (bs) and number of user equipments (ues).

Moreover, dataset for offline training is scarce in public repository, and thus the agent training (especially in DQN) is only limited to the amount of data observed during offline simulations.

C. Contributions of the Study

The main contribution of this study is the **Generation of Datasets** simualted from different traffic conditions and **Train-**

[1] Near real-time Radio Interface Controller
[2] Deep Reinforcement Learning

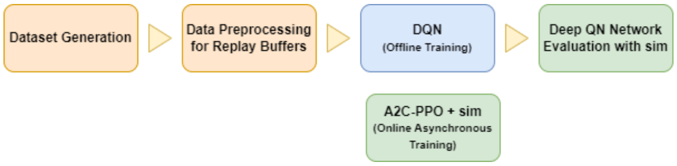

Fig. 1. General RL Pipeline for DQN and PPO trained with ns-oran-sim

ing DRL agents that can work on multiple RAN scenarios and traffic conditions.

II. Related Work

Opportunities arising from the demands of energy-efficient 5G technology has lead to development of different various solutions. Developments in RRC control proposed by [8] , Hetergonous Cache-enabled cellular network schemes [9] have provided significant improvements in energy efficiency.

Additionally, many control solutions have also been deployed on both non-RT and near-RT RIC with the uprise of AI technologies. Salem [6] introduces different sleeping modes that is extended on MDP with methods like Q-Learning. In [7], authors developed an energy-efficient ML pipeline for dynamically deploying closed-loop control solutions for resource allocation and network slicing.

O-RAN is an emerging paradigm that aims to create open and intelligent solutions in RAN systems. Evaluating solutions requires the need to develop various simulation platforms. Lacava et. al. [14] introduced an ns-o-ran simulation - an extension of ns3 network simulator - which provides detailed simulated environment for 5G O-RAN system. Mezzavilla et al [11] introduced similar simulators on 5G mmWave networks. Most modern RL agents relies on gymnasium [3] based environment for both agent training and evaluation. Lacava et al. (2024) developed ns-O-RAN-Gym, a framework designed for online reinforcement learning (RL) training within O-RAN environments. This study highly references the works of Bordin et. al. [10] on the Design and Evaluation of Deep Reinforcement Learning for Energy Saving in Open RAN.

III. Methodology

The general framework for DRL agents (shown in Figure 1) depends on the the algorithm used. DQN agents are first trained offline over sets of processed replay buffers generated from a series of simulations using different initial conditions, network conditions and heuristic policies. Trained DQN are then fine-tuned and evaluated online through the same simulator. PPO agents on the other hand is directly trained and evaluated from the simulator using multiple workers observing different traffic conditions.

A. Simulated Environment

The simulated environment is setup as follow (Figure

- **7 base station** (6 NR only, 1 NR + LTE)[3]

[3]NR - new radio (5G), LTE - long term evolution (4G)

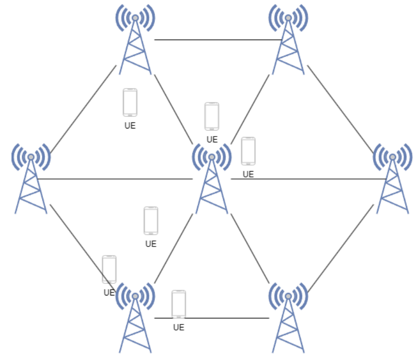

Fig. 2. Simulation Setup

- **21 user equipments (ues)** uniformly distributed around each base station (3 each) with configured random walk dynamics of predefined speed
- Each simulation is configured over different
  - **Traffic Scenarios**: full-buffer traffic, half nodes in full buffer and half nodes in bursty, bursty traffic, 0.25 full buffer, 0.25 bursty 3Mbps, 0.25 bursty 0.75Mbps, 0.25 bursty 0.15Mbps
  - **Data Rates**: high/low
  - RNG (for seeding)

B. State and Action Spaces

Table I details the **State Spaces** $\mathcal{S}$ for the agent totalling to **61 states**.[4]

Description	Count
QoS flow volume to transport blocks for downlink	7 (bs) + 1 (agg)
Enegy Saving State Cost	7 (bs)
QoS PDU volume for downlink flows	7 (bs) + 1 (agg)
Radio link failures	7 (bs) + 1 (agg)
Physical Resource Block (PRB) Percent Usage	7x2 (bs)
Ratio of 64QAM transport blocks	7 (bs) + 1 (agg)
EEE KPI	7 (bs)
Zero Count	1

TABLE I
Network Key Performance Metrics (KPM) as RL States

Additionally, the **Action Space** is set $\mathcal{A} = \{0, 1\}^i$ where is i is the number of base station. Each action indicates **0: ES-ON**, **1: ES-OFF** for each base station - totalling to $2^7 = 128$ **discrete actions**.

C. Reward Function

The reward function [10] is a function of throughput, energy consumption, coverage and activation cost .

$$R = \max_{a} \sum_{t=t_0}^{\infty} \sum_{i=1}^{N} \omega_1 \rho_i(a_i(t)) - \omega_2 \gamma_i(a_i(t)) - \omega_3 \zeta_i(a_i(t)) - \omega_4 \delta_i(a_i(t)) - \omega_2 BsON(a_i(t))$$

subject to $\sum_{j=1}^{k} \omega_i = 1$ and $a_i(t) \in \mathcal{A}$ where $a_i(t)$ is the action for cell i at time t.

[4]bs - base station, agg - aggregated metric sum

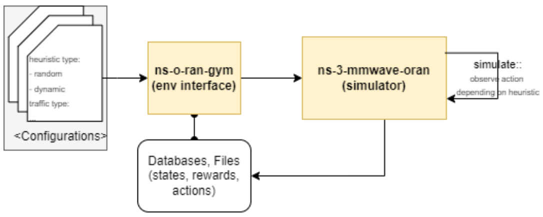

Fig. 3. Generating Data through Simulation

1) ρ_i is defined as cell throughput - number of bytes transmitted at the PDCP layer by cell i
2) γ_i is defined as the cell i energy consumption: $\gamma_i(a_i(t)) = EC_i(a_i(t)) \cdot P_{tx,i}$ where EC_i and $P_{tx,i}$ are the total number of PDU transmitted by cell i and its associated transmit power
3) $BsON(a_i(t))$ is defined as the number of active cells at time t
4) $\zeta_i(\cdot)$ and $\delta_i(\cdot)$ are defined respectively as the UE count in Radio Link Failure (RLF) and the cost to activate a single cell

D. Dataset Generation for Offline Training

80767 observations from 1007 truncated simulated episode were collected. Simulations were run from different traffic conditions, data rates, different kinds of heuristic, RNG, etc. using **ns-3-mmwave-oran** simulator interfaced with gym-capable **ns3-oran-gym**. Generated collections of trajectories are normalized and distributedly stored in a **memory-mapped files** [5] for efficient data access, reduction of RAM usage during training and will serve as persistent store for offline replay buffers.

Aforementioned memmap files can be distributed as a potential benchmark replay buffers for future use cases.

E. Algorithms and Agents

1) Algorithms: The Deep-RL agents will be trained using an off-policy Deep-Q Learning and Proximal Policy Optimization. The following algorithms are among the most widely RL algorithms to date and are among the foundations of most modern RL algorithms.

While there are other newer approach to be considered, the motivation behind a simpler yet effective algorithms roots on the sub-second action interval of Network operations particularly in the Near-RT RIC RAN.

This study references existing implementations of mentioned algorithms from stable-baseline-3 [18] which has a comphrensive implementations[6] of different RL methods. Figures 4 and 5 represents a simple architecture of implemented algorithms.

2) Agent Configurations: Table II summarizes the list of agents trained and evaluated in this study. Most value and policy network used are 2-linear layer dense networks. Three sets of reward function parameters are also experimented upon.

[5]https://discuss.pytorch.org/t/dataloader-and-memmaps/180614
[6]https://github.com/DLR-RM/stable-baselines3

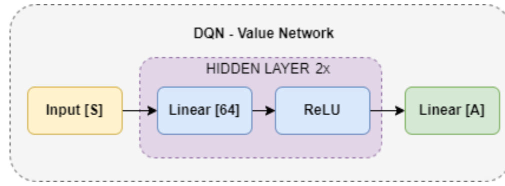

Fig. 4. Stable Baselin3 3 - DQN default architecture

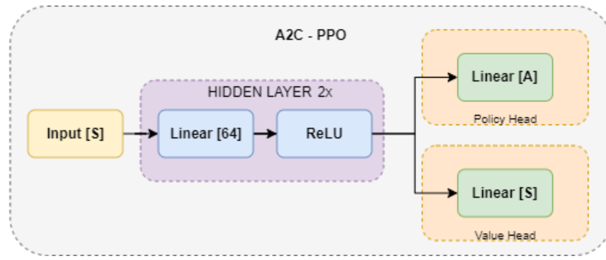

Fig. 5. Stable Baseline - A2C with PPO Head

PPO policy networks are also extended to multi-binary actions instead of discrete actions. Additionally, a deeper network is also evaluated consisting of 5-layer policy and 4-layer value networks. All agents are compared to the **dynamic** sleeping policy described by Salem [6].

Agent	Description	Architecture	Reward Parameters (ω_i)
DQN-1	Base DQN	2 Layer	0.51, 0.19, 0.2, 0.1
DQN-2	DQN (reward variant 1)	2 Layer	0.4, 0.4, 0.1, 0.1
PPO-1	Base PPO	2 Layer	0.51, 0.19, 0.2, 0.1
PPO-2	PPO (reward variant 1)	2 Layer	0.4, 0.4, 0.1, 0.1
PPO-3	PPO (reward variant 2)	2 Layer	0.75, 0.15, 0.05, 0.05
PPO-MB-1	PPO (multi-binary)	2 Layer	0.51, 0.19, 0.2, 0.1
PPO-MB-2	PPO (deep) (multi-binary)	5 Layer policy 4 Layer value	0.4, 0.4, 0.1, 0.1

TABLE II
AGENT CONFIGURATIONS

3) Training Hyperparameters: The following are the list of hyperparameters used for DQN (Table IV) and PPO (Table III) variants used when training different DRL agents.

Hyperparameter	**Value**
policy	"MlpPolicy"
batch_size	32
learning_rate	3e-4
n_steps (steps before update)	32
n_epochs (number of epochs when updating)	10
gamma	0.99
gae_lambda	0.95
clip_range	0.2
ent_coef	0.01
vf_coef	0.5
max_grad_norm	0.5

TABLE III
PPO HYPERPARAMETERS

Hyperparameter	Value
policy	”MlpPolicy”
batch_size	32
learning_rate	3e-4
learning_starts	1000(offline), +1 (online)

TABLE IV
DQN HYPERPARAMETERS

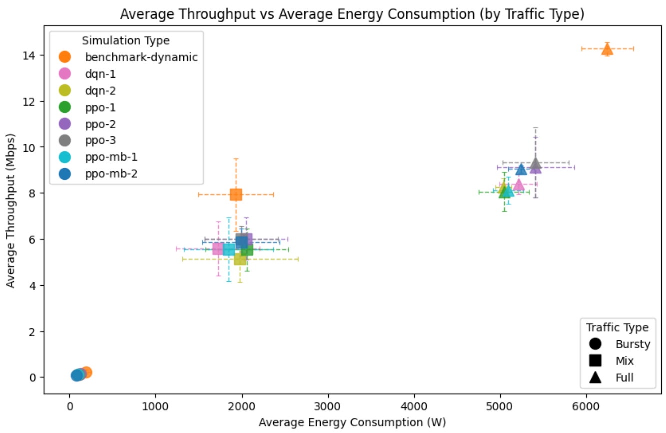

Fig. 6. Aggregated Performance Measure per Traffic Type

IV. Results and Discussion

A. Agent Performance: Throughput and Energy Saving Trade-offs on Different Traffic Scenarios

Tables V, VI and VII details the performance of each agents and its relative comparisons with the benchmark policy. **Most agents** have **aggressive** energy saving policy for network conditions with **burst traffic**. **DQN** have the lowest consumption (-10.71% relative energy saving with 29% throughput tradeoff) in **mix-traffic network scenarios** while the rest of the agents

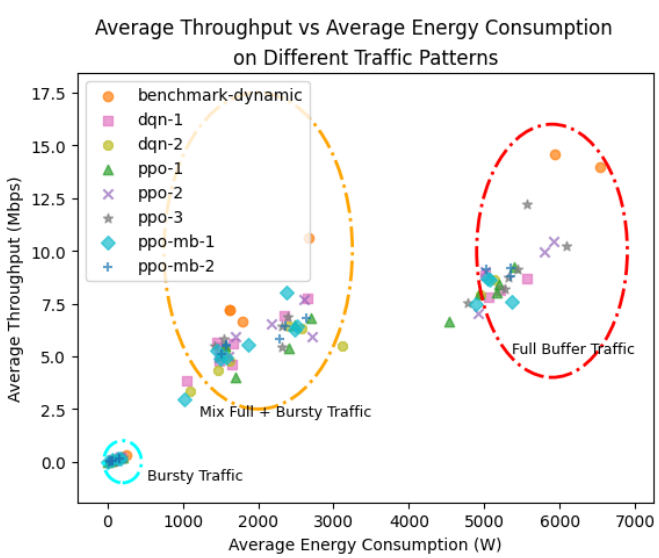

Fig. 7. Performance Measure of all evaluated simulations clustered by Traffic Type

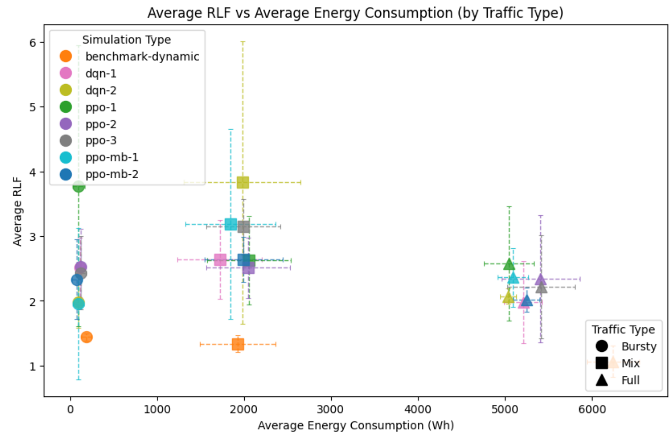

Fig. 8. Caption

fails with most of them have even higher energy consumption than the benchmark. Generally, most agents shows promise in energy saving. **PPO-2, PPO-3 and PPO-MB-2 are generally more conservative** when it comes to energy saving but also has less tradeoff with throughput.

Figure 6 shows a scatter plot of the mean performance (energy saving vs throughput) with associated standard error while figure 7 shows the scatter plot of all evaluated simulations.

B. Agent Performance: Radio Link Failures and ON-Cost on Different Traffic Scenarios on Different Traffic Scenarios

Radio Link Failures measures the number of user equipments that loses connection with the network due to poor signal quality, interference, or mobility issues. **DQN and PPO-MB variants perform better** over PPO variants on**RLF metrics** on **both burst and full traffic**. PPO-1 have shown the worst RLF performance across all different traffic scenarios. On the other hand PPO-2 and PPO-3 have **consistent sub 2 performance ON Cost** indicating the rarer base-ON scenarios. Tables VIII, IX and X details the stability of agents using RLF and ON Cost metrics over bursty, mix and full-traffic scenarios respectively while figures **??** and **??** shows their relationship with energy consumption.

V. Conclusion and Future Work

Different reinforcement learning agents demonstrate varying suitability depending on the criticality of the scenario. Deep Q-Network. Deep Q-Network (DQN) can be deployed in mixed-traffic environments, leveraging both offline and online training for broader experience exploration. Aggressive agents are well-suited for non-critical RAN use cases such as deployment in rural areas. Conversely, conservative agents, are preferable for more critical application including enhanced mobile broadband (eMBB) and ultra-reliable low-latency communications (URLLC). While these agents exhibit potential energy potential capabilities, none surpass the benchmark tradeoff between throughput and energy consumption [6]. To further optimize future performance, future designs should

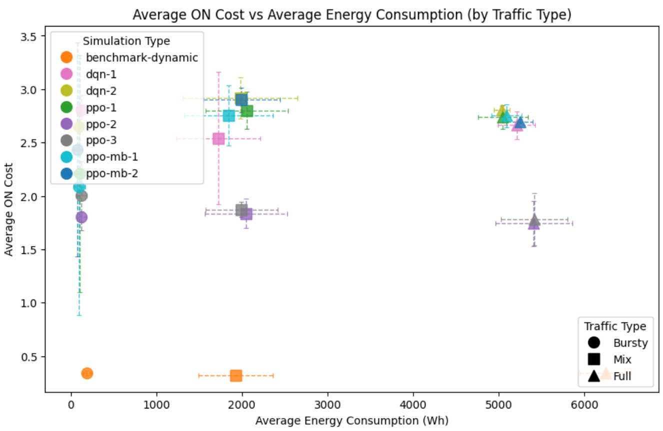

Fig. 9. Caption

undergo reconsideration with extended training periods to enhance throughput while maintaining energy efficiency.

REFERENCES

[1] M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376-1411, Secondquarter 2023, doi: 10.1109/COMST.2023.3239220.

[2] https://www.ericsson.com/en/openness-innovation/open-ran-explained

[3] Gymnasium: An API standard for reinforcement learning with a diverse collection of reference environments. https://gymnasium.farama.org/

[4] Shurdi, Olimpjon & Ruci, Luan & Biberaj, Aleksandër & Mesi, Genci. (2021). 5G Energy Efficiency Overview. European Scientific Journal ESJ. 17. 10.19044/esj.2021.v17n3p315.

[5] Zappone, Alessio. (2016). A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. IEEE Journal on Selected Areas in Communications. 34. 10.1109/JSAC.2016.2550338.

[6] Fatma Ezzahra Salem. Management of advanced sleep modes for energy-efficient 5G networks. Networking and Internet Architecture [cs.NI]. Institut Polytechnique de Paris, 2019.

[7] T. M. Ho, K. -K. Nguyen and M. Cheriet, "Energy Efficiency Learning Closed-Loop Controls in O-RAN 5G Network," GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 2748-2753, doi: 10.1109/GLOBECOM54140.2023.10437790

[8] S. Ryoo, J. Jung and R. Ahn, "Energy efficiency enhancement with RRC connection control for 5G new RAT," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018, pp. 1-6, doi: 10.1109/WCNC.2018.837711

[9] Zhang, J., Zhang, X., Imran, M., Evans, B., and Wang, W. (2017) Energy Efficiency Analysis of Heterogeneous Cache-enabled 5G Hyper Cellular Networks. In: IEEE Globecom 2016, Washington, DC, USA, 04-08 Dec 2016, ISBN 9781509013289 (doi:10.1109/GLOCOM.2016.7841790)

[10] M. Bordin et al., "Design and Evaluation of Deep Reinforcement Learning for Energy Saving in Open RAN," 2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2025, pp. 1-6, doi: 10.1109/CCNC54725.2025.10976108.

[11] M. Mezzavilla et al. End-to-End Simulation of 5G mmWave Networks (2018). IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2237-2263, thirdquarter 2018, doi: 10.1109/COMST.2018.2828880.

[12] V. Mnigh, K.Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller. PLaying Atari with Deep Reinforcement Learning (2013), https://arxiv.org/abs/1312.5602

[13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford and O. Klimov. Proximal Policy Optimization Algorithms (2017),https://arxiv.org/abs/1312.5602

[14] A. Lacava, M. Bordin, M. Polese, R. Sivaraj, T. Zugno, F. Cuomo, and T. Melodia. "ns-O-RAN: Simulating O-RAN 5G Systems in ns-3", Proceedings of the 2023 Workshop on ns-3 (2023), DOI:10.1145/3592149.3592161

[15] Lacava, Andrea and Pietrosanti, Tommaso and Polese, Michele and Cuomo, Francesca and Melodia, Tommaso. Enabling Online Reinforcement Learning Training for Open RAN (2024). pp. 577-582 2024 IFIP Networking Conference (IFIP Networking).

[16] Pritz, Ma and Leung. Jointly-Learned State-Action Embedding for Efficient Reinforcement Learning. CIKM '21: Proceedings of the 30th ACM International Conference on Information & Knowledge Management pp. 1447 - 1456 https://doi.org/10.1145/3459637.3482357

[17] Fujimoto, et.al. For SALE State-Action Representation Learning for Deep Reinforcement Learning (2023). https://arxiv.org/abs/2306.02451

[18] Raffin, Hill, Gleave and Kanervisto. Stable-Baselines3: Reliable Reinforcement Learning Implementations (2021). http://jmlr.org/papers/v22/20-1364.html.

APPENDIX
AGENT PERFORMANCE: THROUGHPUT AND ENERGY CONSUMPTION

Agent	Throughput (Mbps)	Energy Consumption (W)
benchmark-dynamic	0.22 (0.00%)	190.37 (0.00%)
dqn-1	**0.13** (-41.92%)	118.05 (-37.98%)
dqn-2	0.11 (-52.19%)	**97.40** (-48.83%)
ppo-1	0.11 (-53.17%)	100.02 (-47.47%)
ppo-2	**0.13** (-40.08%)	125.60 (-34.00%)
ppo-3	0.12 (-44.70%)	120.80 (-36.52%)
ppo-mb-1	0.10 (-55.61%)	**96.16** (-49.47%)
ppo-mb-2	0.08 (-64.33%)	**76.06** (-60.04%)

TABLE V
THROUGHPUT AND ENERGY CONSUMPTION UNDER BURSTY TRAFFIC

Agent	Throughput (Mbps)	Energy Consumption (W)
benchmark-dynamic	7.93 (0.00%)	1930.07 (0.00%)
dqn-1	5.57 (-29.71%)	**1723.37** (-10.71%)
dqn-2	5.13 (-35.32%)	1978.67 (+2.52%)
ppo-1	5.54 (-30.14%)	2058.31 (+6.64%)
ppo-2	**6.01** (-24.23%)	2049.69 (+6.20%)
ppo-3	**6.01** (-24.27%)	1994.74 (+3.35%)
ppo-mb-1	5.55 (-30.00%)	**1847.37** (-4.29%)
ppo-mb-2	**5.87** (-25.99%)	1993.08 (+3.27%)

TABLE VI
THROUGHPUT AND ENERGY CONSUMPTION UNDER MIXED TRAFFIC

Agent	Throughput (Mbps)	Energy Consumption (W)
benchmark-dynamic	14.26 (0.00%)	6243.78 (0.00%)
dqn-1	8.38 (-41.21%)	5210.75 (-16.56%)
dqn-2	8.27 (-41.95%)	**5037.69** (-19.30%)
ppo-1	8.06 (-43.46%)	**5046.26** (-19.20%)
ppo-2	**9.11** (-36.08%)	5410.07 (-13.34%)
ppo-3	**9.33** (-34.51%)	5413.99 (-13.26%)
ppo-mb-1	8.11 (-43.09%)	**5093.54** (-18.43%)
ppo-mb-2	**9.03** (-36.64%)	5245.27 (-15.99%)

TABLE VII
THROUGHPUT AND ENERGY CONSUMPTION UNDER FULL TRAFFIC

APPENDIX
AGENT PERFORMANCE: RADIO LINK FAILURE AND ON-COST

Agent	RLF	ON Cost
benchmark-dynamic	1.44 (0.00)	0.34 (0.00)
dqn-1	2.52 (+1.08)	2.80 (+2.46)
dqn-2	**1.97** (+0.53)	2.65 (+2.31)
ppo-1	3.77 (+2.33)	2.21 (+1.87)
ppo-2	2.52 (+1.08)	**1.80** (+1.46)
ppo-3	2.43 (+0.99)	**2.00** (+1.66)
ppo-mb-1	**1.96** (+0.52)	**2.09** (+1.75)
ppo-mb-2	**2.33** (+0.89)	2.44 (+2.10)

TABLE VIII
RLF AND ON COST UNDER BURSTY TRAFFIC

Agent	RLF	ON Cost
benchmark-dynamic	1.34 (0.00)	0.32 (0.00)
dqn-1	**2.64** (+1.30)	**2.54** (+2.22)
dqn-2	3.83 (+2.49)	2.92 (+2.60)
ppo-1	**2.63** (+1.29)	2.80 (+2.48)
ppo-2	**2.51** (+1.17)	**1.83** (+1.51)
ppo-3	3.15 (+1.81)	**1.87** (+1.55)
ppo-mb-1	3.19 (+1.85)	2.75 (+2.43)
ppo-mb-2	**2.64** (+1.30)	2.90 (+2.58)

TABLE IX
RLF AND ON COST UNDER MIXED TRAFFIC

Agent	RLF	ON Cost
benchmark-dynamic	1.07 (0.00)	0.34 (0.00)
dqn-1	**1.99** (+0.92)	**2.66** (+2.32)
dqn-2	**2.07** (+1.00)	2.81 (+2.47)
ppo-1	2.58 (+1.51)	2.74 (+2.40)
ppo-2	2.34 (+1.27)	**1.74** (+1.40)
ppo-3	2.22 (+1.15)	**1.78** (+1.44)
ppo-mb-1	2.36 (+1.29)	2.75 (+2.41)
ppo-mb-2	**2.01** (+0.94)	2.70 (+2.36)

TABLE X
RLF AND ON COST UNDER FULL TRAFFIC