PPO-Based Dynamic Resource Allocation with Traffic Classification in B5G O-RAN

Gunasekaran Raja[1], Sanjeev A[2], Krithika Ravishankar,[3]
[1 2 3]Department of Computer Technology, Anna University, MIT Campus, India,
[1]dr.r.gunasekaran@ieee.org, [2]sanjeevananth2211@gmail.com, [3]krithikaravishankar4@gmail.com,

***Abstract*—Modern wireless communication systems must support diverse service classes, ranging from high-throughput multimedia to ultra-reliable low-latency and massive machine-type traffic, under stringent Quality of Service (QoS) constraints. As Beyond 5G (B5G) networks evolve, traditional static resource allocation schemes often result in suboptimal spectrum usage and QoS violations. To address this, we propose PPO-Based Dynamic Resource Allocation with SVM-Aided Traffic Slicing in B5G O-RAN, an AI-native, standards-compliant control framework for Open Radio Access Network (O-RAN) environments. The system integrates a lightweight Support Vector Machine (SVM) classifier to perform real-time traffic slice classification using KPI-driven features, and a Proximal Policy Optimization (PPO)-based reinforcement learning (RL) agent to dynamically allocate Physical Resource Blocks (PRBs) based on slice-specific QoS priorities. Implemented on a fully open-source testbed comprising srsRAN, Open5GS, and the O-RAN Software Community Near-RT RIC, our work demonstrates tangible gains over static baselines. Experimental evaluations report a 25% improvement in enhanced Mobile Broadband (eMBB) throughput, a 28% reduction in Ultra-Reliable Low-Latency Communication (URLLC) latency, a 15% increase in massive Machine-Type Communication (mMTC) access success rate, and over 90% precision in reclaiming idle PRBs. These results validate our project as a practical and scalable solution for real-time, QoS-aware resource orchestration in future B5G networks.**

***Index Terms*—Traffic Classification, Dynamic Resource Allocation, O-RAN, SVM, Reinforcement Learning, Open5GS, srsRAN**

I. Introduction

The evolution from Long-Term Evolution (LTE) to B5G has introduced a paradigm shift in mobile communication, enabling support for diverse traffic classes, including high-throughput media, ultra-reliable low-latency control signals, and energy-efficient IoT data exchange. To address this heterogeneity, the 3GPP classifies services into three categories: eMBB, URLLC, and mMTC, each governed by distinct QoS requirements for throughput, latency, reliability, and energy efficiency [1].

eMBB is designed to deliver high data rates and supports applications such as video streaming, virtual reality, and cloud-based gaming. URLLC, by contrast, is optimized for extremely low-latency, high-reliability communication required by mission-critical services like autonomous driving, remote surgery, or ambulance telemetry. mMTC focuses on supporting massive numbers of low-power, low-cost IoT devices, making it ideal for smart city deployments, environmental sensing, and industrial automation scenarios. [1].

Traditional static resource allocation schemes, ubiquitous in 4G and initial 5G networks, are ill-equipped to accommodate the stringent and heterogeneous QoS requirements of modern mobile systems. Under variable traffic loads, these approaches suffer from spectrum underutilization, elevated latency, and unpredictable service degradation. The O-RAN architecture mitigates these shortcomings by decoupling and standardizing RAN components, thereby enabling vendor-neutral interoperability and fine-grained programmability. At its heart lies the Near-Real-Time RAN Intelligent Controller (Near-RT RIC), which hosts AI/ML-driven xApps capable of orchestrating radio resources with slice-aware, context-adaptive policies [2].

Despite its promise, realizing intelligent control in the Near-RT RIC entails several challenges. First, real-time traffic classification must operate exclusively on abstracted Key Performance Indicators (KPIs), since both privacy regulations and latency budgets constrain deep-packet inspection [3]. Second, wireless traffic exhibits pronounced non-stationarity—characterized by sudden load spikes, silent intervals, and rapid shifts in QoS demands—which complicates both modeling and decision-making. Third, deploying Reinforcement Learning (RL) agents in live RAN environments demands carefully designed exploration strategies and safety guards; without them, suboptimal actions during training can degrade ongoing services.

To address this challenge, our proposed contributions are as follows:

- **Lightweight Traffic Classification Using SVM:** A Support Vector Machine (SVM) model classifies traffic into eMBB, URLLC, and mMTC slices based on RIC-exposed KPI features, enabling privacy-preserving and resource-efficient operation.
- **Slice-Aware PRB Allocation via PPO:** A Proximal Policy Optimization (PPO) agent, deployed as an xApp in the Near-RT RIC, dynamically allocates PRBs using slice-specific reward functions derived from 3GPP QoS metrics.
- **Open-Source Testbed Validation with Silence-Aware Optimization:** We validate our system in a fully open-source virtualized testbed (srsRAN, Open5GS, and O-RAN SC RIC) and enhance spectral efficiency through a silence-detection mechanism that reclaims unused PRBs for redistribution among active UEs.

In what follows, Section II reviews related work on

O-RAN architectures, network slicing, traffic classification, and reinforcement-learning-based resource allocation. Section III presents the system architecture and methodology, detailing traffic generation, KPI collection, feature engineering, and SVM-based classification. Section IV describes our dynamic PRB allocation strategy, including the design of slice-specific reward functions and the PPO-based allocation algorithm. Section V reports our experimental setup and results, covering overall PRB utilization, silent-period reclamation, URLLC latency, eMBB throughput, and mMTC access success against static baselines. We conclude in Section VI with a summary of our findings and discuss avenues for future work toward scalable, multi-RIC deployments and proactive, forecast-driven resource orchestration.

II. Background and Related Works

A. O-RAN and Network Slicing in B5G

B5G networks must accommodate a wide spectrum of service classes—including eMBB, URLLC, and mMTC—each imposing unique QoS demands in terms of throughput, latency, reliability, and device density [4]. Traditional static Resource Allocation (RA) mechanisms cannot cope with this heterogeneity, particularly under fluctuating traffic conditions. O-RAN overcomes these limitations through a disaggregated, vendor-neutral design that exposes standardized interfaces for programmability. Central to this paradigm is the Near-RT RIC, which hosts AI/ML-driven xApps to perform adaptive, slice-aware orchestration of radio resources over the E2 interface [5]. Network slicing—another cornerstone of B5G—enables multiple logically isolated virtual networks (eMBB, URLLC, mMTC) to run concurrently on shared physical infrastructure; nevertheless, achieving fair and efficient allocation of PRBs across slices with inherently conflicting performance objectives remains an open research challenge [6].

B. Traffic Classification in O-RAN

Traffic classification underpins QoS-aware scheduling by assigning UE flows to their corresponding slice, yet the constraints of the Near-RT RIC preclude payload inspection and prohibitively large feature vectors [7]. Although deep architectures—such as LSTM, GRU, and Transformer models—can deliver high classification accuracy, their dependence on full-packet data and extensive preprocessing incurs unacceptable latency and privacy risks in real-time xApps [8]. By contrast, lightweight methods like SVMs leverage a limited set of KPI attributes (e.g., throughput ratios, buffer occupancy, and CQI) extracted from E2SM-KPM reports to achieve comparable performance with minimal overhead. Prior work has demonstrated that an SVM classifier can reach approximately 92 percent accuracy using these compact features, making it an ideal candidate for efficient, privacy-compliant deployment within the Near-RT RIC [9].

C. Dynamic PRB Allocation using Reinforcement Learning

Physical Resource Block (PRB) allocation involves real-time spectrum sharing among traffic slices, where traditional

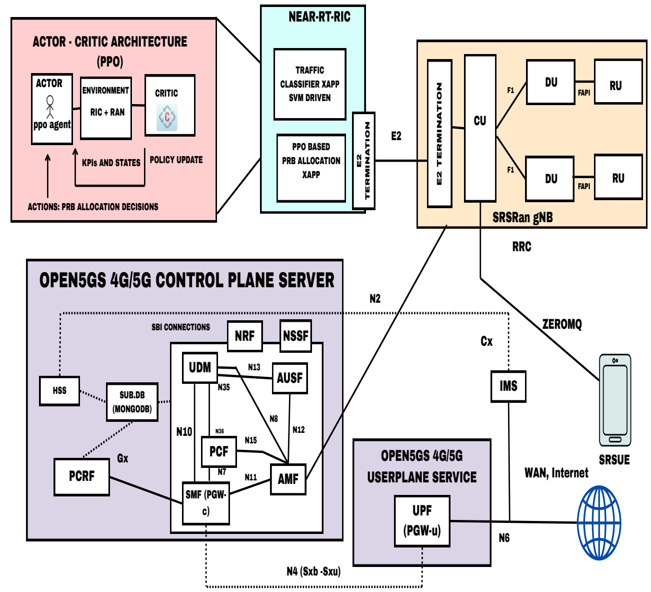

Fig. 1. Proposed Intelligent Resource Allocation Architecture for B5G O-RAN leveraging SRSRAN and Open5GS.

rule-based schedulers (e.g., proportional fair, max-min fairness) struggle under bursty traffic and silence periods [10]. Reinforcement Learning (RL) offers a dynamic alternative, with frameworks like REAL [11] and others [12] leveraging PPO and DQN for closed-loop, slice-specific PRB allocation in Near-RT RIC settings [13].

However, RL-based RAN control faces challenges: high-dimensional state spaces (scaling with UEs and slices) hinder convergence; unsafe exploration risks QoS violations; and non-stationary channel conditions require continual policy adaptation [14]. To address these, recent approaches incorporate traffic forecasting [15] and adversarial bandit strategies [16] to improve robustness and anticipate worst-case dynamics [17].

D. Research Gap and Motivation

Most existing solutions treat slice identification and resource scheduling as separate tasks, limiting applicability in dynamic O-RAN settings and overlooking inefficiencies like PRB underutilization during silent periods in low-activity slices. Our framework addresses this by unifying KPI-driven SVM classification with PPO-based PRB allocation, using slice-specific reward functions and a silence-aware reclamation mechanism. Deployed on open-source B5G components (srsRAN, Open5GS, srsUE, and the O-RAN SC RIC), it enhances spectrum efficiency, ensures QoS, and offers a scalable path toward AI-driven RAN control in B5G/6G networks.

III. System Architecture and Methodology

A. Traffic Generation

Traffic generation is performed by a ZeroMQ-based broker that emulates the Radio Frequency interface between the UE and gNB, with Python scripts producing mixed TCP and UDP flows. To emulate realistic packet-size variability, Gaussian noise is applied to the payload length, and a roulette-wheel selection mechanism limits the number of concurrently active UEs to three. This approach prevents buffer saturation and naturally models the silent intervals characteristic of bursty wireless traffic. Each UE is assigned a unique traffic profile, comprising smoothed KPI-driven traffic patterns, an associated QoS slice identifier (eMBB, URLLC, or mMTC), and continuously updated performance metrics that feed directly into downstream classification and allocation modules.

B. KPI Collection

KPI data is collected via the E2SM-KPM service (Style 2) over the E2 interface between the gNB and Near-RT RIC. A dynamic weighted-average filter smooths raw values, adapting for responsiveness during activity and stability during idle periods. The resulting time series, forms the state input for the SVM classification and PPO-based PRB allocation xApps.

IV. Proposed Methodology

A. Traffic Classification

1) Feature Engineering: The feature engineering stage converts each UE's smoothed KPI time series into a set of statistical descriptors over a fixed-length sliding window. Metrics such as moving averages, standard deviations, per-interval throughput deltas, and buffer-occupancy gradients capture temporal dynamics and transient behavior. These windowed features not only improve class separability for the SVM classifier but also enrich the state representation provided to the subsequent reinforcement-learning agent for PRB allocation.

2) QoS Classification: Using the engineered feature vectors, we deploy a multi-class Support Vector Machine (SVM) to assign each UE to its appropriate slice—eMBB, URLLC, or mMTC—solely based on KPI telemetry available via E2SM-KPM. To avoid instability during session start-up or silent periods, we incorporate a delayed decision mechanism that defers classification until the feature history exceeds a minimum window length, thereby ensuring robust and reliable slice identification under varying traffic conditions, aligning with prior SVM-based classification works. The model employs delayed prediction for stable early learning, uses normalized and smoothed throughput ratios alongside QoS metrics as features, and supports semi-online learning with periodic updates.

B. Dynamic PRB Allocation

1) Design of Reward Functions: To achieve differentiated optimization for the heterogeneous QoS requirements of URLLC, eMBB, and mMTC traffic, distinct reward functions are formulated [18], [19], [20]. Each reward is explicitly mapped to measurable KPIs referenced in 3GPP Release 19 specifications.

URLLC Reward Function: Effective Usage Computation:

Algorithm 1: Adaptive KPI Smoothing and Normalized State Generation

Input: List of UEs $\mathrm{UE}_{\mathrm{List}}$, Reporting interval T, Decay factor λ, Influence floor ε

Output: Smoothed KPI time series s_t, Updated capacity estimates $C_{\max,t}$

Initialize data structures for each UE in $\mathrm{UE}_{\mathrm{List}}$;
foreach *UE in* UE_{List} **do**
  foreach *KPI metric* **do**
    Subscribe via E2SM-KPM Service Style 2 over E2 interface;
    Receive initial KPI v_0;
    Set $s_0 \leftarrow v_0$;
    Set $C_{\max,0} \leftarrow v_0$;
  end
end
while *true* **do**
  foreach *received KPI value* v_t *for a UE* **do**
    Update maximum capacity estimate: $C_{\max,t} \leftarrow \max(\lambda \cdot C_{\max,t-1}, v_t)$;
    Compute normalized activity: $a_t \leftarrow \min\left(\frac{v_t}{C_{\max,t}}, 1.0\right)$;
    Compute adaptive smoothing influence: $\alpha_t \leftarrow \varepsilon + (1-\varepsilon) \cdot a_t$;
    Apply exponential smoothing: $s_t \leftarrow \alpha_t \cdot v_t + (1-\alpha_t) \cdot s_{t-1}$;
    Save s_t and $C_{\max,t}$ to memory or database;
    if *logging enabled* **then**
      Log α_t, a_t for diagnostics;
    end
  end
end

$$\mathrm{EffUsage} = \frac{\mathrm{TransmittedVolume\ (UL\ or\ DL)}}{\mathrm{AvgMcsEff} \times \mathrm{BWPPRB} \times \mathrm{AllocPRBs} \times \mathrm{T}}; \tag{1}$$

where EffUsage quantifies the ratio of actually transmitted data to the maximum achievable capacity, serving as a measure of PRB utilization efficiency. Formally, TransmittedVolume (UL/DL) is the total number of bits (or bytes) sent in the uplink or downlink during the reporting window. AvgMcsEff denotes the mean Modulation and Coding Scheme efficiency over that interval, capturing the average spectral efficiency of the link. BWPPRB is the per-PRB bandwidth (in Hertz), and AllocPRBs is the total number of PRBs assigned to the UE in the same scheduling period. Finally, T specifies the duration of the reporting window. Together, these parameters yield a normalized metric that measures the effectiveness of PRB allocation in contributing to actual data throughput.

Overall Reward Function:

$$R_{\mathrm{URLLC}} = w_r \cdot R_{\mathrm{reliability}} + w_l \cdot R_{\mathrm{latency}} - w_c \cdot C_{\mathrm{ovp}} \tag{2}$$

Algorithm 2: PPO-Based Online PRB Allocation and Training

Initialize: Actor parameters θ, critic parameters ϕ, replay buffer $\mathcal{B}$, PPO hyperparameters (lr, clip ϵ, γ);
Input: KPI feature set $\{s_i\}$, UE slice labels $\{\ell_i\}$, reporting interval T
Output: Trained policy $\pi_\theta(a \mid s)$
while *system is operational* **do**
  for *each scheduling interval* t **do**
    for *each UE* i **do**
      Observe s_i^t;
      Classify slice $\ell_i^t \in \{\text{eMBB}, \text{URLLC}, \text{mMTC}\}$;
      $\tilde{s}_i^t \leftarrow [s_i^t, \ell_i^t]$;
      $a_i^t \sim \pi_\theta(\cdot \mid \tilde{s}_i^t)$; apply a_i^t;
    end
    wait T;
    for *each UE* i **do**
      Observe s_i^{t+1};
      $r_i^t \leftarrow R_{\ell_i^t}(s_i^t, a_i^t, s_i^{t+1})$;
      Store $(\tilde{s}_i^t, a_i^t, r_i^t, s_i^{t+1})$ in $\mathcal{B}$;
    end
    if $|\mathcal{B}| \geq N_{\min}$ **then**
      Sample batch $\{(s, a, r, s')\}$ from $\mathcal{B}$;
      for *each* (s, a, r, s') **do**
        $\delta \leftarrow r + \gamma V_\phi(s') - V_\phi(s)$;
        Accumulate actor gradient via PPO surrogate loss using δ;
        Accumulate critic gradient to minimize δ^2;
      end
      Update θ, ϕ with accumulated gradients; (Optional) trim $\mathcal{B}$;
    end
  end
end
Return π_θ;

Where $R_{\text{reliability}} = 1 - \text{PacketDropRate (UL/DL)}$ represents the inverse of the packet drop rate, ensuring high reliability is rewarded.

The latency reward R_{latency} is defined piecewise as:

$$R_{\text{latency}} = \begin{cases} 1, & l \leq L_{\text{target}} \\ \frac{L_{\max} - l}{L_{\max} - L_{\text{target}}}, & L_{\text{target}} < l < L_{\max} \\ 0, & l \geq L_{\max} \end{cases}$$

This function penalizes latency values exceeding the target L_{target}, with a gradual decay until the maximum acceptable latency $L_{\max}$ is reached.

Finally, the overprovisioning cost is defined as:

$$C_{\text{ovp}} = \text{AllocatedPRBs} - \text{EffectiveUsage} \times \text{AllocatedPRBs}$$

TABLE I
SLICE-WISE GAINS: PPO-BASED VS STATIC ALLOCATION

Slice	Metric	Static	PPO	Gain
eMBB	Throughput (Mbps)	$\sim$45	$\sim$58	+25–30%
URLLC	Latency (ms)	$>$10	$<$7	28%
mMTC	Access Success (%)	72–75	85–87	+15%
All	PRB Utilization (%)	$\sim$65	$>$85	+30–38%
Idle	Reclaimed PRBs (%)	$<$50	$>$90	+40%

which captures the fraction of PRBs that were allocated but not effectively utilized.

URLLC demands extreme reliability and low latency simultaneously. The first component drives the allocation policy toward improved reliability, the second rewards low-latency transmissions, and the third prevents overprovisioning by rewarding efficient resource usage.

eMBB Reward Function

Effective Usage Computation:

$$\text{Rut} = \frac{\text{TransmittedVolume (UL or DL)}}{\text{AvgMcsEff} \times \text{BWPPRB} \times \text{AllocPRBs} \times T} \tag{3}$$

where Rut denotes the resource utilization ratio, representing how effectively the allocated PRBs are used. *TransmittedVolume (UL or DL)* refers to the total volume of data transmitted in either uplink or downlink, measured in bits or bytes. *AvgMcsEff* stands for the average Modulation and Coding Scheme (MCS) efficiency, indicating the spectral efficiency achieved during transmission. *BWPPRB* is the bandwidth assigned per Physical Resource Block, expressed in Hertz. *AllocPRBs* represents the number of PRBs allocated during the scheduling window, and T denotes the duration of the reporting interval over which the transmission metrics are observed.

Adaptive Utilization Term:

$$t = \min\left(1, \frac{\text{TransmittedVolume}}{\text{MaxVolume}}\right), \tag{4}$$

$$R_{\text{adaptive\_utilization}} = (1 - t) \cdot \text{Rut} + t \cdot \text{Rut}^2, \tag{5}$$

Overall Reward Function:

$$R_{\text{eMBB}} = w_1 \cdot R_{\text{throughput}} + w_2 \cdot R_{\text{adaptive\_utilization}}. \tag{6}$$

mMTC Reward Function: Overall Reward Function:

$$R_{\text{mMTC}} = w_1 \cdot R_{\text{acc\_succ}} + w_2 \cdot R_{\text{ue\_dens}} + w_3 \cdot R_{\text{eff}} - w_4 \cdot C_{\text{undpov}}, \tag{7}$$

where the metric $R_{\text{acc\_succ}} = 1 - \frac{\text{Connection Setup Failures}}{\text{Connection Attempts}}$ quantifies the access success rate, rewarding scenarios with minimal connection failures. Similarly, $R_{\text{ue\_dens}} = \frac{\text{Connected UEs}}{\text{Attempted UEs}}$ reflects the network's ability to sustain high user equipment (UE) density by measuring the ratio of successfully connected devices to the total attempting access. The term $R_{\text{eff}} = \text{EffectiveUsage}$, as defined earlier, captures how efficiently allocated PRBs are utilized. Finally, $C_{\text{undpov}} = \max(0, \text{Threshold} - \text{EffectiveUsage})$ introduces a penalty for under-provisioning, ensuring that the allocation meets a minimum efficiency threshold.

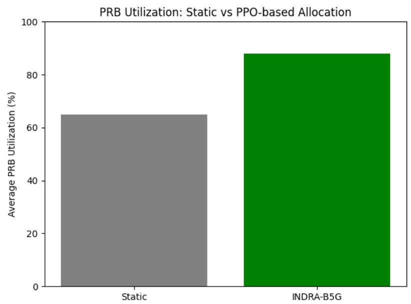

Fig. 2. PRB Utilization: Static vs PPO-based Allocation

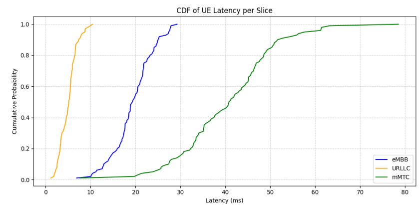

Fig. 3. UE Latency CDF per Slice

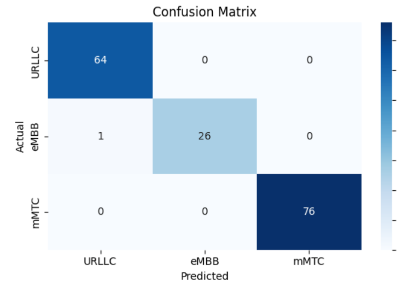

Fig. 4. SVM Slice Classification Accuracy

C. PPO-Based Online PRB Allocation for eMBB, URLLC, and mMTC

The algorithm shown in Algorithm 2 dynamically allocates PRBs to UEs while simultaneously updating the PPO agent during real-time operation based on KPIs and classified QoS slice types (eMBB, URLLC, mMTC) inspired by DRL usage in latency-constrained vehicular environments [21] and leveraging the hierarchical DRL policies.

V. Evaluation and Results

To evaluate the system's performance, we conducted comprehensive simulations on a virtualized 5G testbed that integrates srsRAN (gNB/UE), Open5GS (5GC), and the O-RAN Software Community's Near-RT RIC. Our goal was to assess the effectiveness of the proposed xApps for traffic classification and resource allocation in comparison to traditional static allocation policies.

A. Overall PRB Utilization

Figure 2 illustrates the average PRB utilization achieved under both static allocation and the PPO-based approach. The latter consistently achieves over 85% utilization across scenarios, significantly outperforming the static baseline (65%). This improvement highlights the PPO agent's capability to dynamically optimize PRB usage by responding to slice-specific demand fluctuations.

B. Silent Period PRB Reclamation

To minimize resource underutilization, our system implements a silence-aware reclamation mechanism. As shown in Figure 7, the PPO agent detects inactive UEs and reclaims unused PRBs with high efficiency, achieving over 90% utilization in several time windows. This dynamic adjustment leads to improved spectral efficiency and fairness.

C. URLLC Latency Comparison

Figure 6 compares URLLC latency under static and PPO-based PRB scheduling. The PPO agent ensures latencies consistently stay below 7 ms, whereas the static baseline often exceeds 10 ms. This improvement is crucial for mission-critical applications that require ultra-reliable, low-latency communication.

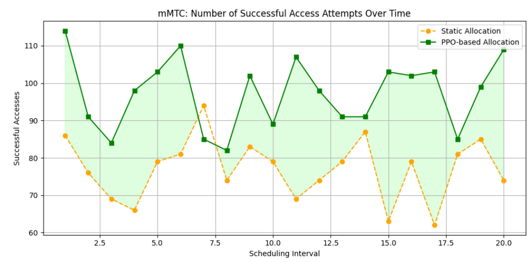

Fig. 5. mMTC: Number of Successful Access Attempts Over Time

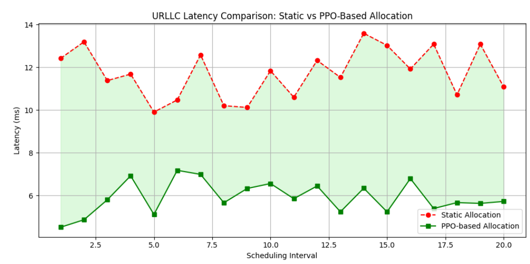

Fig. 6. URLLC Latency Comparison: Static vs PPO-Based Allocation

D. eMBB Throughput Comparison

As depicted in Figure 8, PPO-based scheduling yields higher throughput for eMBB UEs. The average throughput improves by approximately 25–30%, confirming the agent's ability to prioritize bandwidth-intensive services under congestion.

E. mMTC Access Success

Figure 5 highlights the access success rate for mMTC devices. The PPO model consistently achieves more than 15% successful access attempts compared to the static control, demonstrating better support for high access density and efficient contention resolution during peak device activity.

VI. Conclusion and Future Work

This paper introduces a modular, O-RAN-compliant framework that combines lightweight SVM-based traffic classification with PPO-driven reinforcement learning to achieve real-time, slice-aware PRB allocation in B5G networks. Our experimental evaluation shows that it significantly enhances spectrum utilization, reduces URLLC latency, increases eMBB

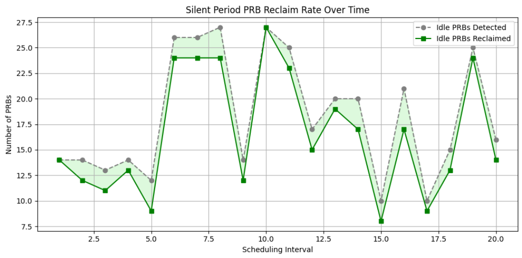

Fig. 7. Silent Period PRB Reclaim Rate Over Time

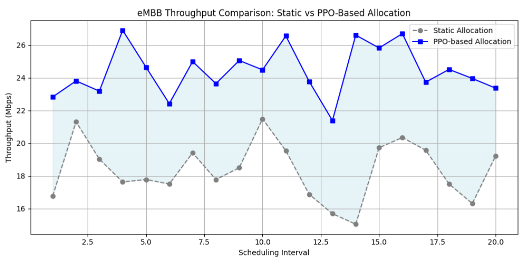

Fig. 8. eMBB Throughput Comparison: Static vs PPO-Based Allocation

throughput, and improves mMTC access success compared to static allocation baselines. Looking ahead, we plan to enhance its scalability and intelligence by deploying xApps across distributed RIC instances using Kubernetes, integrating probabilistic traffic forecasting models to predict slice demand and proactively adjust resource allocation [22], and exploring federated reinforcement-learning techniques to coordinate policies across multiple RICs while maintaining data privacy. These advances will further strengthen adaptability, scalability, and QoS compliance in complex multi-slice B5G and future 6G deployments.

These extensions aim to further enhance adaptability, scalability, and QoS compliance in complex multi-slice 5G and 6G deployments, facilitating self-learning and autonomous edge orchestration.

Acknowledgments

This work was supported by the Telecom Centres of Excellence India (TCOE), Government of India, under the Telecom Technology Development Fund (TTDF) Scheme (File No. TCOE/TTDF/2024/60).

References

[1] 3GPP, "Service requirements for the 5g system," 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 22.261, 2017. [Online]. Available: https://www.3gpp.org/DynaReport/22261.htm

[2] O.-R. Alliance, "O-ran architecture," O-RAN Alliance, Tech. Rep., 2020, accessed: 2025-04-29. [Online]. Available: https://www.o-ran.org/resources

[3] M. Hoffmann, S. Janji, A. Samorzewski, Łukasz Kułacz, C. Adamczyk, M. Dryjański, P. Kryszkiewicz, A. Kliks, and H. Bogucka, "Open ran xapps design and evaluation: Lessons learnt and identified challenges," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 473–486, Feb 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1109/JSAC.2023.3336190

[4] B. Adhikari, M. Jaseemuddin, and A. Anpalagan, "Resource allocation for co-existence of embb and urllc services in 6g wireless networks: A survey," *IEEE Access*, vol. 12, pp. 552–581, 2024.

[5] J. F. Santos, A. Huff, D. Campos, K. V. Cardoso, C. B. Both, and L. A. DaSilva, "Managing o-ran networks: xapp development from zero to hero," *IEEE Communications Surveys & Tutorials*, 2025.

[6] K. Suh, S. Kim, Y. Ahn, S. Kim, H. Ju, and B. Shim, "Deep reinforcement learning-based network slicing for beyond 5g," *IEEE Access*, vol. 10, pp. 7384–7395, 2022.

[7] J. Groen, Z. Yang, D. Muruganandham, M. Belgiovine, L. Ying, and K. Chowdhury, "From classification to optimization: Slicing and resource management with tractor," 2025. [Online]. Available: https://arxiv.org/abs/2312.07896

[8] K. R. T, V. C, J. S, and R. B, "Deep learning-assisted application traffic classification for 5g/b5g network slicing," in *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India, 2024, pp. 1–6.

[9] J. E. Preciado-Velasco, J. D. Gonzalez-Franco, C. E. Anias-Calderon, J. I. Nieto-Hipolito, and R. Rivera-Rodriguez, "5g/b5g service classification using supervised learning," *Applied Sciences*, vol. 11, no. 11, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/11/4942

[10] A. S. Sastry and S. Venkatesan, "A novel approach to the resource allocation for the cell edge user in 5g multiuser mimo network," *Journal of Communications*, vol. 17, no. 1, pp. 38–46, 2022. [Online]. Available: https://www.academia.edu/92270653/A_Novel_Approach_to_the_Resource_Allocation_for_the_Cell_Edge_Users_in_5G

[11] R. Barker, A. Ebrahimi Dorcheh, T. Seyfi, and F. Afghah, "Real: Reinforcement learning-enabled xapps for experimental closed-loop optimization in o-ran with osc ric and srsran," *arXiv preprint arXiv:2502.00715*, 2025, https://arxiv.org/abs/2502.00715.

[12] R. V. Pamba, R. Bhandari, A. Asha, and A. S. Bist, "An optimal resource allocation in 5g environment using novel deep learning approach," *Journal of Mobile Multimedia*, vol. 19, no. 5, pp. 1331–1356, 2023.

[13] X. Chen, Y. Yang, and H. Zhang, "Optimizing network slicing with deep reinforcement learning in o-ran architecture," *Proceedings of IEEE WCNC*, pp. 1–6, 2022.

[14] A. Asheralieva and Y. Miyanaga, "An autonomous resource management framework based on deep reinforcement learning for 5g heterogeneous systems," *Proceedings of IEEE ICC*, pp. 1–7, 2019.

[15] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "How does forecasting affect the convergence of drl techniques in o-ran slicing?" 2023. [Online]. Available: https://arxiv.org/abs/2309.00489

[16] C.-H. Lai, L.-H. Shen, and K.-T. Feng, "Intelligent load balancing and resource allocation in o-ran: A multi-agent multi-armed bandit approach," 2023. [Online]. Available: https://arxiv.org/abs/2303.14355

[17] R. Li, Z. Zhao, X. Zhou, and H. Zhang, "Deep reinforcement learning for resource management in network slicing," *Proceedings of IEEE INFOCOM*, pp. 1–9, 2018.

[18] A. K. Verma, V. K. Singh, M. R. Khan, and P. K. Sethy, "Q-learning based heterogeneous network selection decision algorithm for ultra reliable and low latency communication services," *Wireless Networks*, 2024. [Online]. Available: https://doi.org/10.1007/s11276-025-03931-5

[19] R. M. Sohaib, O. Onireti, Y. Sambo, R. Swash, S. Ansari, and M. A. Imran, "Intelligent resource management for embb and urllc in 5g and beyond wireless networks," *IEEE Access*, vol. 11, pp. 79 434–79 450, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3288698

[20] A. A. E. Hassan, A. E. Mehdi, and M. Saber, "Nb-iot and lte-m towards massive mtc: Complete performance evaluation for 5g mmtc," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 308–320, 2021.

[21] L. Li, L. Song, and Y. Li, "Deep reinforcement learning for resource allocation in vehicular networks," *Proceedings of IEEE ICC*, pp. 1–6, 2018.

[22] A. Alabbasi and M. Krunz, "Traffic prediction for 5g network slicing using machine learning: A review and case study," *Proceedings of IEEE GLOBECOM*, pp. 1–6, 2020.