Developing Trustworthy Reinforcement Learning Applications for Next-Generation Open Radio Access Networks

Ahmad M. Nagib*‡, Hatem Abou-zeid†, Hossam S. Hassanein*
*School of Computing, Queen's University, Canada, {ahmad, hossam}@cs.queensu.ca
†Department of Electrical and Software Engineering, University of Calgary, Canada, hatem.abouzeid@ucalgary.ca
‡Faculty of Computers and Artificial Intelligence, Cairo University, Egypt

***Abstract*—Artificial intelligence is envisioned to transform the design and operation of 6G networks. Reinforcement learning (RL), in particular, has emerged as a fundamental approach toward this goal with strong support from the industry and the Open Radio Access Network (O-RAN) Alliance. While research efforts have demonstrated the potential of RL, the lack of trustworthiness of RL algorithms remains a challenge to its adoption in real-world networks. In this paper, we propose a trustworthy RL framework that addresses the core challenges experienced by RL-based radio resource management applications deployed in O-RANs. We then demonstrate a case study on O-RAN slicing incorporating several modules of the proposed framework. The experimental results show improvements in the average RL convergence rate, initial reward value, percentage of converged scenarios, and reward variance. Hence, the RL-based algorithms exhibit fast convergence and enhanced generalizability, safety, and robustness.**

***Index Terms*—Radio resource management (RRM), Deep reinforcement learning (DRL), Trustworthy DRL, O-RAN slicing, 6G**

I. Introduction

Machine Learning (ML) techniques, and particularly reinforcement learning (RL), will be among the fundamental ingredients to optimize and control 6G networks. Hence, RL is supported by standard bodies such as the Open Radio Access Network (O-RAN) Alliance [1]. This is motivated by the seamless compatibility of network control operation with the RL feedback loop of observing the system state, taking actions, and receiving corresponding rewards. RL agents have demonstrated strong potential to adapt to a mobile network operator's (MNO) goals and provide a pathway toward self-driving networks [1]–[4]. While there have been some efforts toward adopting RL in O-RANs, the literature fails to systematically design methodologies that tackle several key challenges faced in real-world scenarios.

RL-based radio resource management (RRM) algorithms deployed in live networks encounter generalizability, safety, robustness, and explainability challenges. We discuss such challenges in detail as we propose a comprehensive framework that addresses them in the next section. We also show results from two of the framework's instantiations that employ transfer learning, and time series forecasting techniques in the context of O-RAN slicing. The proposed framework ensures fast convergence of RL-based O-RAN algorithms. It also enhances their generalizability, safety, and robustness compared with the traditional RL framework. The results also indicate that the forecasting models proposed to aid the RL convergence do not have to be ideal.

II. Proposed Framework

We propose four modules to address the practical challenges of deploying RL-based RRM applications in next-generation O-RANs. Such modules, highlighted in Fig. 1, aim at designing algorithms that generalize to different network scenarios, adapt to changes in network conditions, and avoid potential performance instabilities. The decisions taken by such algorithms must also be explainable. The following modules fall under the umbrella of trustworthy RL techniques [5]. They jointly constitute a comprehensive trustworthy RL framework that serves as a guideline for wireless network researchers and practitioners to systematically address the identified challenges.

Generalization. The RL-based RRM algorithms must generalize to network scenarios that were not previously seen. This capability is crucial since the offline simulation environments are usually inaccurate and do not reflect all the situations that could be experienced in O-RAN deployment environments. Hence, we propose a generalization module that can be realized using ML-based techniques such as meta-learning and transfer learning (TL). We demonstrate that module in Sections III and IV using a hybrid TL-aided deep RL (DRL) approach that combines policy reuse and distillation TL methods.

Safety. The RL policies for RRM must integrate safety constraints during training to prevent an RL agent from affecting the end user's quality of experience (QoE). For instance, if an RL-based O-RAN slicing application is deployed, the safety constraints would be avoiding the violation of the defined service level agreement (SLA) for each slice. We, therefore, propose a safety module to address this challenge. In this work, we leverage prior knowledge of the O-RAN slicing problem to design a risk-sensitive RL reward function that includes parameters to reflect the acceptable SLAs for each slice.

Robustness. The RL algorithms adopted in O-RAN must enhance their worst-case performance under network uncertainties. Such uncertainties stem from the gaps between training and real O-RAN environments, and the non-stationary wireless environment. Consequently, we propose a robustness module to handle such environment mismatches. To demonstrate that, we adopt the previously stated hybrid TL-aided approach. We also incorporate a forecasting model to predict the future contribution of slices to the overall traffic demand. Both approaches guide the DRL agent when allocating resources to slices.

Explainability. MNOs must understand the reasoning behind

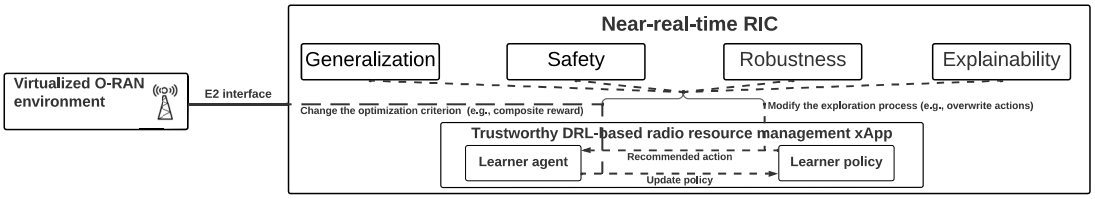


Fig. 1: The proposed trustworthy framework for RL-based applications in next-generation O-RANs.

the RL agents' decisions. The dynamic nature of RL and the uncertainties present in O-RAN environments complicate the task of offering comprehensive and verifiable explanations. Thus, we propose an explainability module to provide semantic and logical explanations for RL-based RRM actions. This can be in the form of attention mechanisms based on Shapley additive explanations (SHAP) as in [4]. Such mechanisms assess the importance of a network state in the decision-making process and steers the RL agent towards executing interpretable actions via explainable RL (XRL)-enhanced reward shaping.

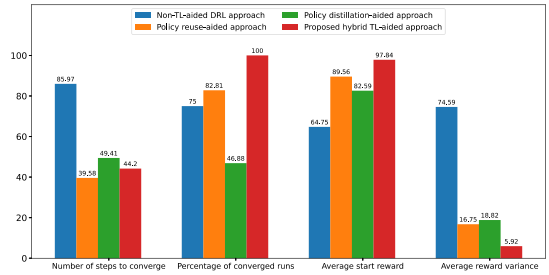
III. Experiment Setup

We manifest two instantiations of the proposed framework to demonstrate three of its modules as part of our case studies on O-RAN slicing in [2], [3]. The first instantiation uses a hybrid TL-aided approach and the second uses a forecasting-aided approach to guide the DRL convergence. The simulation is designed to reflect extreme situations in which the available resources are configured to be less than the actual demand. We compare the convergence performance of the proposed approaches against multiple baselines including the traditional RL framework. We use live VR gaming data as an example of realistic immersive service patterns in 6G. Moreover, we combine such patterns with video and voice over new radio (VoNR) traffic to reflect 3 different slice types in our experiments.

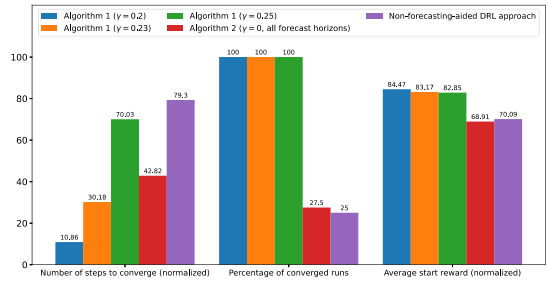
IV. Numerical Results

We measure the average initial normalized reward, variance in the reward, number of steps to converge to the best reward, and percentage of converged simulation runs for each approach as shown in Fig 2. This evaluates whether an approach starts with a good reward value, the change in reward values afterward, the speed of convergence, and the ability to finally converge to the optimal policy respectively. The proposed hybrid TL-aided approach inherits the best of both policy reuse and distillation approaches. Consequently, as illustrated in Fig. 2a, it has the highest initial reward value and percentage of converged runs with, at least, 7.7% and 20.7% improvements respectively over the policy reuse approach. It also yields the lowest variance in reward values per run with, at least, a 64.6% decrease in variance when compared with policy reuse. It does so while still having a competitive performance in terms of the number of steps to converge.

Moreover, the proposed forecasting-aided approach (Algorithm 1) has the fastest convergence rate as seen in Fig 2b. It noticeably outperforms the traditional DRL baseline and maintains a higher initial reward value, even when forecasting error is high. Furthermore, 100% of the conducted scenarios converge to the optimal resource allocation configuration given that the error's standard deviation is 0.25 or lower. This shows that our framework is robust against forecasting errors.



(a)



(b)

Fig. 2: DRL convergence performance of the proposed approaches: a) hybrid TL-aided DRL; b) forecasting-aided DRL.

V. Conclusion

In this work, we discuss the core challenges that hinder the wide adoption of RL in next-generation O-RANs. We propose a comprehensive framework that proves to maintain fast convergence and enhance the trustworthiness of RL-based RRM. We will explore combining the proposed techniques with other approaches such as constrained and explainable DRL as a major step toward trustworthy RL-based RRM in O-RANs.

References

[1] M. Tsampazi, S. D'Oro, M. Polese, L. Bonati, G. Poitau, M. Healy, and T. Melodia, "A comparative analysis of deep reinforcement learning-based xapps in o-ran," in *IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 1638–1643.

[2] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based o-ran slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024.

[3] A. M. Nagib, H. Abou-zeid, and H. S. Hassanein, "How does forecasting affect the convergence of drl techniques in o-ran slicing?" in *IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 2644–2649.

[4] F. Rezazadeh, H. Chergui, L. Alonso, and C. Verikoukis, "Sliceops: Explainable mlops for streamlined automation-native 6g networks," *IEEE Wireless Communications*, pp. 1–7, 2024.

[5] M. Xu, Z. Liu, P. Huang, W. Ding, Z. Cen, B. Li, and D. Zhao, "Trustworthy reinforcement learning against intrinsic vulnerabilities: Robustness, safety, and generalizability," *arXiv preprint arXiv:2209.08025*, 2022.