URLLC Challenges in NTN: An Analysis of O-RAN Split-Function Architectures

Syed Tariq Shah, Rana Muhammad Sohaib, *Member, IEEE,*
Mahmoud A. Shawky, Poonam Yadav, *Member, IEEE,* Ahmad Almogren, *Senior Member, IEE*

***Abstract*—This paper presents a comprehensive analysis of different functional splits (FS) in an Open RAN-based Unmanned Aerial Vehicle (UAV) Non-Terrestrial Networks (NTN) architecture for Ultra-Reliable Low-Latency Communication (URLLC). We systematically analyse and compare the effectiveness of different FSs, focusing on their impact on UAV computational complexity and overall network latency and reliability for URLLC users. Through extensive simulations, we assess the trade-offs between computational complexity, front-haul bandwidth requirements, and network reliability for various FS options, with a particular focus on URLLC traffic. Our results demonstrate that FS-7.2x outperforms other configurations, achieving superior latency performance and enhanced reliability by efficiently balancing onboard processing and front-haul bandwidth utilisation. These findings provide critical insights for network designers in optimising O-RAN configurations for UAV-based NTN, ensuring robust and low-latency communication for next-generation wireless applications.**

***Index Terms*—URLLC, ORAN, DRL, Resource Allocation, Functional Splits**

I. Introduction

OPEN Radio Access Network (O-RAN) is a transformative evolution in next-generation wireless network (NGWN) architecture aimed at disaggregating traditionally integrated Radio Access Network (RAN) components into modular, interoperable elements [1]. Unlike traditional RAN systems, where hardware and software are tightly coupled and typically supplied by a single vendor, O-RAN promotes an open, flexible ecosystem.

O-RAN's architecture is particularly appealing in the context of Non-Terrestrial Networks (NTN), which include the integration of Unmanned Aerial Vehicles (UAVs) as an emerging use case [2]. Leveraging O-RAN principles in UAV-based NTN allows efficient communication across dynamic and distributed network environments. Open interfaces such as A1 and E2, coupled with intelligent controllers like the RAN Intelligent Controller (RIC), facilitate advanced network management. This enables O-RAN-based UAV networks to handle complex, real-time data traffic, ensuring scalable and robust network operations.

S. T. Shah, is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO1 1ZY, UK e-mail: Syed.Shah@essex.ac.uk. R. M. Sohaib and P. Yadav are with the Department of CS, University of York, UK, e-mail: {rana.sohaib,poonam.yadav}@york.ac.uk). M. A. Shawky is with the James Watt School of Engineering, University of Glasgow, G12 8QQ, Glasgow, UK e-mail: mahmoud.shawky@glasgow.ac.uk, and A. Almogren is with the Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia, e-mail: ahalmogren@ksu.edu.sa.

However, integrating UAVs into O-RAN-based NTN environments presents distinct challenges stemming from the disaggregated nature of O-RAN components (e.g., the Open Radio Unit (RU), Open Distributed Unit (DU), and Open Centralised Unit (CU)) and the dynamic aerial environment. The wireless front-haul (FH) links that connect these components in UAV NTNs are inherently less stable than in terrestrial networks (TNs), often constrained by high latency, limited bandwidth (BW), and interference. Additionally, UAVs are subject to strict energy limitations, optimising onboard computational workloads while ensuring ultra-reliable, low-latency communication (URLLC) is critical. These challenges become more pronounced when supporting heterogeneous services, such as enhanced Mobile Broadband (eMBB) and URLLC, which have conflicting quality of service (QoS) demands [2].

Managing these heterogeneous services over an NTN, especially in remote or challenging environments, requires adaptive and intelligent resource allocation mechanisms. Here, the RIC plays a critical role by optimising network resources through real-time analytics, machine learning, and data-driven decision-making [1]. Specifically, the RIC must balance the trade-offs introduced by the choice of functional splits (FS) in the O-RAN architecture, as FS directly affects the processing demands, latency, and BW of UAV-based NTNs. Lower-layer FSs (e.g., FS-7 and FS-8) impose stringent latency and connectivity requirements on the FH, while higher-layer FSs (e.g., FS-2) increase BW demands but allow for greater processing centralisation [3]. To this end, it is evident that selecting the appropriate FS is critical for optimising the overall network performance, particularly in UAV NTN scenarios where maintaining stable and high-performance FH links over wireless mediums is inherently more complex than in TNs.

Given these challenges, efficient Radio Resource Management (RRM) is vital for ensuring robust operation in UAV-based O-RAN NTNs. Traditional RRM approaches often fail to meet the stringent latency, reliability, and throughput demands of URLLC and eMBB services in dynamic and resource-constrained NTN environments. Deep Reinforcement Learning (DRL), a subset of machine learning, has emerged as a promising approach for tackling these challenges [4], [5]. DRL enables real-time optimisation of RRM parameters, such as spectrum utilisation and power allocation, by dynamically adapting to the evolving network conditions. This capability is particularly advantageous in UAV networks, where traffic patterns are highly variable, and latency requirements for URLLC are exceptionally stringent [6].

Despite DRL's potential, significant challenges and research gaps remain, particularly in the context of UAV-based O-RAN NTNs. The interplay between URLLC and eMBB services and the selection of FSs introduces complex trade-offs in latency, BW, and processing overhead. While existing studies have broadly explored DRL applications in TNs, there is a lack of comprehensive investigations into its applicability for UAV NTNs. Moreover, no conclusive studies have evaluated the performance of DRL-based RRM strategies across different FS configurations for URLLC users in UAV NTNs. Addressing these gaps is essential for advancing state-of-the-art communication systems and developing intelligent RRM solutions that meet the unique demands of UAV-based O-RAN NTNs. To the best of our knowledge, no studies have conclusively determined which FS offers the best balance of latency, BW, and processing overhead for DRL-driven URLLC solutions in UAV NTN.

This paper provides a comparative analysis of different FS within O-RAN-based UAV NTN architectures, focusing on URLLC performance. Specifically, a Thompson Sampling (TS)-based DRL approach is proposed for RRM and power allocation for URLLC. The evaluation covers FSs 2, 6, 7.2x, and 8, assessing their impact on transmission error rates and latency outages. The study offers valuable insights into how varying functional splits influence reliability and latency in UAV-based NTN environments.

II. Split Options in ORAN-based UAV NTN

The integration of FSs within O-RAN-based UAV NTNs is crucial for enabling flexible and efficient RAN architectures. Unlike traditional RAN architectures, FSs in O-RAN disaggregate the radio protocol stack into separate units (i.e., CU, DU, RU), allowing for optimised distribution of computational and networking tasks. As shown in Fig. 1a, these FSs allow for flexible distribution of RAN functions, with higher-layer tasks centralised in the CU and lower-layer, real-time tasks handled by the DU and RU. The Physical (PHY) layer in the RAN is divided into Low-PHY (L-PHY) and High-PHY (H-PHY). L-PHY manages real-time operations like FFT/IFFT, cyclic prefix handling, and DAC/ADC conversions for RF signal processing, while H-PHY handles higher-level tasks such as modulation, channel coding, and HARQ combining, which are less time-sensitive but computationally intensive [7]. These FSs enable the separation of control and user plane tasks, making them critical for supporting diverse applications like URLLC and eMBB.

Although FSs have been standardised and evaluated for TNs in 3GPP Release 15 and subsequent releases, their applicability in NTNs has not been fully explored and standardised. FSs for NTNs are expected to be a significant part of the 3GPP NTN Rel-19 framework, which aims to extend these concepts to support non-terrestrial and air-borne platforms such as satellites and UAVs [8]. For TNs, FS-2 has been standardised by 3GPP, FS-6 by the Small Cell Forum, and FS-7 by the O-RAN Alliance. These standards serve as a foundation for understanding how UAV-based NTNs could benefit from distributed RAN functions.

To better understand how FSs impact the performance of UAV-based NTNs, it is essential to explore the various deployment configurations and the appropriate split options for these deployments. Fig. 1 illustrates various deployment options for integrating UAVs into O-RAN-based NTNs. Each subfigure represents a distinct FS configuration that affects how processing tasks are distributed across the UAV and the core network, influencing the system's overall performance in terms of latency, BW, and computational complexity.

- **UAV as a Relay:** In the first scenario, depicted in Fig. 1b, the UAV functions as an RF relay, forwarding analogue signals between eMBB/URLLC users and a ground-based gNB over an RF link. No significant baseband processing is performed onboard, as all RAN functions (RU, DU, CU) are centralised at the terrestrial entity.
- **UAV with RU** In another configuration, shown in Fig. 1c, the UAV hosts the RU, while the DU and CU remain on the ground and are connected via a FH link. The RU's responsibilities can vary based on the UAV's capabilities, FS, and FH conditions. It can handle tasks from basic RF functions to more complex PHY and MAC operations. For this setup, FSs such as FS-8, FS-7.x, and FS-6 can be applied. FS-8 leaves most of the processing centralised at the ground entity, while FS-7.x offloads more tasks to the UAV, offering a balance between FH efficiency and onboard complexity. FS-6 pushes even more responsibility onto the UAV, reducing FH dependency but increasing computational demands. FSs-5 and 4 offer similar advantages to FS-6 but will further increase the onboard complexity.
- **UAV with RU and DU** In this configuration (Fig. 1d), the UAV hosts both the RU and DU, with the CU remaining on the ground, connected via a midhaul link. This shifts substantial baseband processing to the UAV, including H-PHY and MAC operations. FS-2 is ideal for this setup, as it enables most RAN functions (e.g., MAC, RLC, and scheduling) to be processed locally, resulting in significantly reducing midhaul BW demands and improving UAV autonomy. However, this increases onboard computational complexity, requiring enhanced processing resources for real-time tasks like HARQ and scheduling. Additionally, FSs-3 and 1 can also be applied, further lowering midhaul traffic while distributing computational load differently. FS-3 offloads lower-layer MAC and RLC functions to the UAV, introducing moderate computational requirements with reduced real-time processing compared to FS-2. FS-1 centralises only control-plane tasks at the CU, while user-plane data is processed on the UAV, further reducing midhaul BW but adding complexity in managing both planes onboard.
- **UAV with RU, DU, and CU:** The deployment shown in Fig. 1e is the most autonomous, as the UAV hosts the full suite of RAN functions, including the RU, DU, and CU, making it fully autonomous in terms of processing capabilities. This setup eliminates the need for a dedicated FH or midhaul link, as all RAN layers are processed onboard the UAV. While this drastically reduces latency and FH BW requirements, it significantly increases computational and energy demands, limiting

scalability for larger networks or diverse traffic patterns.

A. Performance Implications of FS in UAV-based NTNs

In UAV-based NTNs, FSs impact the computational complexity on the UAV as well as the BW and latency demands on the FH. Lower-layer FSs increase the processing burden on the UAV but reduce FH BW and latency requirements. Conversely, higher-layer FSs place more demands on the FH, requiring higher BW and lower latency while reducing the UAV's computational load. The following sections analyse the major FS options in terms of these key trade-offs.

Option 8 places minimal computational demands on the UAV, as all baseband functions, such as modulation, HARQ, and beamforming, are processed by the DU. However, it imposes significant FH BW requirements because raw IQ samples are transmitted over the FH, resulting in heavy data loads that scale with antenna numbers. While propagation delays are negligible over short UAV heights, the large volume of data can lead to transmission delays and queuing at the DU, making this FS prone to latency issues. The primary contributors to overall latency are the high processing delays in the DU and the large data volume traversing the FH. Furthermore, the reliability of the system is fragile due to the heavy dependence on the FH link, where any instability or congestion can exacerbate delays.

Option 7.1 reduces FH BW compared to FS-8 by transmitting subcarriers after RE mapping. This shifts basic RF functions to the UAV while centralising most real-time operations at the DU. The UAV's computational complexity remains low, as tasks such as scheduling, HARQ, and error correction are still centralised. Regarding overall latency, although transmission delays over the FH are minimal, processing delays at the DU, including queuing and scheduling, are the main contributors. The system's reliability remains tied to the availability of high BW and low-latency FH links.

Option 7.2x offers greater efficiency for UAV-based NTNs by offloading more L-PHY tasks, such as digital beamforming and IFFT, to the UAV. This reduces the amount of data transmitted over the FH but increases the computational complexity on the UAV. Specifically, by moving digital beamforming to the UAV's DU, the bit rate now scales with the number of MIMO layers rather than antenna ports, reducing overall data transmission requirements, though the bit rate remains relatively high and constant. While latency requirements (i.e., up to 0.25 ms) are still significant, they are slightly more relaxed than FS-7.1. Overall latency in FS-7.2x is primarily determined by the UAV's ability to handle the additional processing load and the efficiency of coordination between the UAV and the ground-based DU. Although the FH data load is reduced, delays can still arise from processing overhead at both the UAV and the DU, particularly in real-time operations. Furthermore, reliability improves with the reduced FH data load, but stable FH performance is still essential to maintain seamless coordination.

Option 7.3 improves FH efficiency by offloading additional PHY tasks, such as layer mapping and precoding, to the UAV (RU). Since modulation is performed at the ground-based DU, the bit rate on the FH becomes variable and is significantly lower than in FS-7.2. While the reduced FH data load enhances BW efficiency, the increased computational burden on the UAV requires robust real-time processing capabilities. The overall latency in FS-7.3 is influenced by the processing delays at both the UAV and the DU. Although FH latency requirements remain similar to previous FSs, the system's efficiency depends heavily on the UAV's ability to handle its increased processing load and the coordination between the UAV and the ground-based DU. While the reduced FH dependency improves reliability, the trade-off is the need for more robust onboard processing capabilities to manage real-time operations efficiently.

Option 6 reduces FH BW by transmitting only transport blocks, making it suitable for BW-constrained UAV-based NTNs. However, this comes with increased control-plane overhead, as close coordination between the MAC and PHY layers is needed for scheduling. The UAV assumes additional PHY processing responsibilities while the DU manages scheduling, creating a trade-off between FH BW efficiency and onboard computational demands. Overall latency is influenced by processing delays at both the RU and DU. Ensuring real-time coordination between the UAV and DU is essential to avoid synchronisation issues that could degrade system reliability. Task coordination between MAC and PHY layers directly impacts reliability, and any misalignment can lead to increased latency or degraded performance.

Option 2 or the RLC/PDCP FS reduces the FH BW as the DU handles H-PHY, MAC, and RLC functions co-located with the RU. Only PDCP and network layer functions are processed at the CU, making this FS ideal for BW-constrained midhaul scenarios. This configuration reduces FH load significantly but increases onboard energy consumption due to real-time task handling at the UAV. The trade-off between FH efficiency and UAV complexity is significant. Reduced BW comes at the cost of higher computational load and energy demand on the UAV, which must manage real-time tasks at the DU. Regarding midhaul latency, FS-2 offers relatively relaxed sensitivity (up to 10 ms). However, the overall delay is affected by processing at the DU, mid-haul transmission, and synchronisation between CU and DU/RU.

III. Joint Power and RRM in NTN for URLLC

A. Network Model

We model a network that provides two services, namely eMBB and URLLC, within the O-RAN framework in the context of UAV-NTN. In this architecture, several edge cloud servers are strategically deployed at the near-RT-RIC, establishing connections with a regional cloud server located at the non-RT-RIC. This configuration enables efficient resource allocation and management for the two services, leveraging the capabilities of UAVs for enhanced connectivity and performance in diverse environments. NTNs present a challenging environment due to their long propagation delays, highly dynamic channels, and unpredictable traffic conditions. We consider an NTN with UAV-based access to provide enhanced mobile broadband (eMBB) and URLLC services. The system

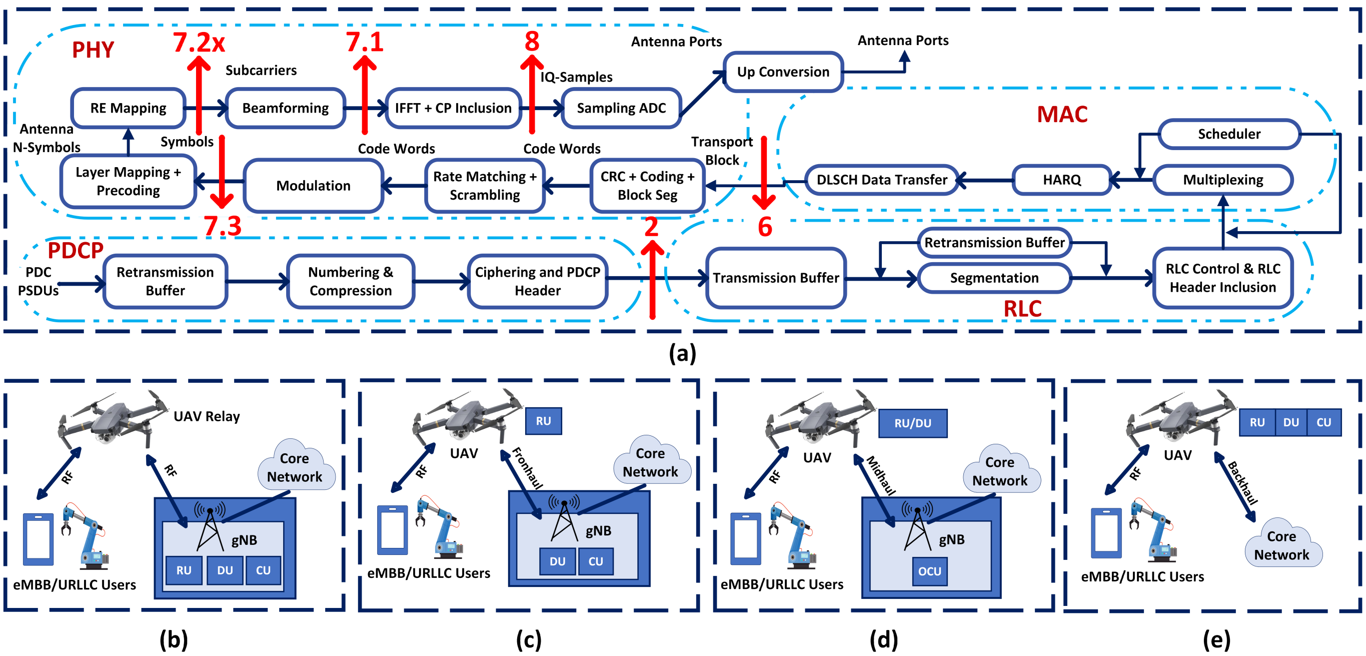


Fig. 1: Deployment options, (a) Functional Split Options (b) UAV act as a relay (c) UAV act as an NT RU (d) UAV act as an NT RU and DU (e) UAV act as an NT gNB

must efficiently allocate resources across users with different service requirements while adhering to stringent URLLC constraints. The eMBB focuses on a high data rate, requiring efficient BW utilisation with relatively lenient latency and reliability constraints.

However, URLLC targets ultra-low latency ($\leq 1ms$) and extreme reliability ($\geq$ 99.999%), which is critical for autonomous driving and telemedicine applications. We use Numerology 1 for eMBB traffic and Numerology 2 for URLLC. Numerology 1 has a sub-carrier spacing of 30 KHz, which is beneficial for high throughput, while Numerology 2 offers a sub-carrier spacing of 60 KHz, enabling faster and more reliable transmissions for URLLC. The system uses a dynamic frame structure, where eMBB slots are punctured to accommodate urgent URLLC traffic. This flexibility allows the system to meet the stringent URLLC latency requirements while maximising throughput for eMBB users. The challenge lies in puncturing the eMBB slots without significantly affecting their throughput while ensuring that URLLC packets are delivered within their latency bounds. This paper outlines a comprehensive system model that integrates RAN intelligent controller (RIC) and O-RAN functionalities to optimise the performance of both eMBB and URLLC services. The RIC leverages O-RAN FS, enabling flexible resource allocation and improved system efficiency.

B. Open Challenges

The primary challenge addressed in this paper is the simultaneous provisioning of eMBB and URLLC services within an NTN environment, ensuring that the distinct QoS requirements for each are met effectively. The immediate scheduling of URLLC users can affect the rate of eMBB users. To address this, a puncturing decision variable is introduced, which allows the system to puncture the eMBB transmission to accommodate URLLC users [9].

To maintain the quality of service for both eMBB and URLLC users, we must address the outage probabilities (OP) associated with each service type. For URLLC, the reliability must be extremely high, often quantified as a packet error rate of 10^{-5}. The URLLC xApp scheduler on RIC manages the OP by dynamically adjusting the HARQ processes based on real-time conditions to ensure that URLLC users receive reliable service. Latency is another critical factor, particularly for URLLC, whose goal is to keep end-to-end latency below 1 ms. The network must minimise the total transmission delay while accommodating the HARQ round-trip time. The main objectives can be summarised as :

- Maximise eMBB Throughput: The primary goal is to maximise the data rate for eMBB users while ensuring that the prioritisation of URLLC traffic does not significantly degrade their service. This requires efficient resource allocation strategies adapting to varying network conditions and user demands.
- Meet URLLC QoS Requirements: URLLC users require ultra-reliable communication with minimal latency. To satisfy these stringent requirements, we must ensure that the OP for both reliability and latency are kept within acceptable bounds. Specifically, the reliability of URLLC must remain above a threshold (e.g., 99.999%), and latency must be maintained below 1 ms.
- Dynamic Resource Management: Effective resource management is essential for balancing the competing needs of eMBB and URLLC users. This involves dynamically allocating resources such as power and BW based on

real-time traffic conditions and user requirements.

- Implement O-RAN FS: We aim to create a flexible network architecture that allows seamless communication between different network components by leveraging O-RAN FS. This flexibility is essential for rapidly adapting to changing network conditions and user demands.

C. URLLC frame structure

The traditional LTE frame structures are unsuitable for URLLC due to excessive user plane latency, often exceeding 1 ms because of retransmission delays and HARQ round-trip times (RTT). To address this, 5G introduces wider subcarrier spacings (e.g., 60 kHz) and mini-slot designs, which are essential for NTN applications where large propagation delays and communication interruptions occur due to atmospheric conditions or handovers. For URLLC in NTNs, a 60 kHz subcarrier spacing ensures ultra-low latency and high reliability. Each OFDM symbol duration is 17.85 µs, leading to a mini-slot duration of 35.71 µs (2 symbols × 17.85 µs). This enables multiple mini-slots within a single transmission time interval (TTI), allowing for rapid HARQ feedback within or immediately after a mini-slot. Considering HARQ retransmissions, the maximum allowable latency is 0.21 ms (6 × 0.0357 ms), effectively meeting URLLC's strict latency requirements [5].

IV. DRL-based Solution in NTN and O-RAN

To effectively tackle the challenges presented by URLLC in NTN, it is crucial to customise the DRL framework to align with the unique characteristics of these applications. URLLC's stringent requirements for ultra-low latency ($\leq 1ms$) and extreme reliability ($\leq 99.999\%$) in the dynamic NTN environment demand intelligent resource allocation mechanisms. This involves defining states, actions, and rewards that encapsulate the intricacies of URLLC performance metrics while adapting to fluctuating channel conditions.

In the context of decision-making problems addressed by DRL, an ϵ-greedy approach is often employed to select actions. Here, the parameter ϵ represents the probability of the agent choosing a random (exploration) action rather than the one currently deemed optimal (exploitation). When ϵ is small, the agent primarily exploits known actions, while a larger ϵ encourages exploration through random selections. Although this method can yield effective results, it may not always guide the agent toward a globally optimal solution, as the agent may become trapped in a sub-optimal solution by repeatedly choosing the action with the highest estimated reward.

A. Thompson Sampling (TS) for URLLC

In previous studies, TS has demonstrated its effectiveness in the intelligent resource scheduling of URLLC users [9]. In NTN contexts where uncertainties in the communication environment are prevalent, TS is particularly valuable. For resource allocation tasks, TS dynamically adapts to fluctuations in channel conditions by continually fine-tuning its probability distributions.

Using Bayesian inference, TS begins by modelling the uncertainty associated with each action's true underlying reward distribution. This is particularly useful in NTN environments, where channel conditions are highly variable, and accurate resource allocation is critical. The ability of TS to adapt its probability distributions in real-time ensures efficient handling of URLLC's stringent requirements, such as minimising latency outages and maintaining transmission reliability under fluctuating conditions.

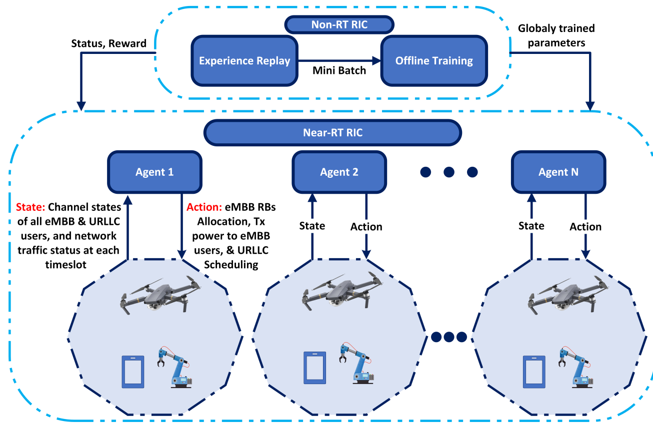


Fig. 2: Distributed DRL Framework for RRM in NTN

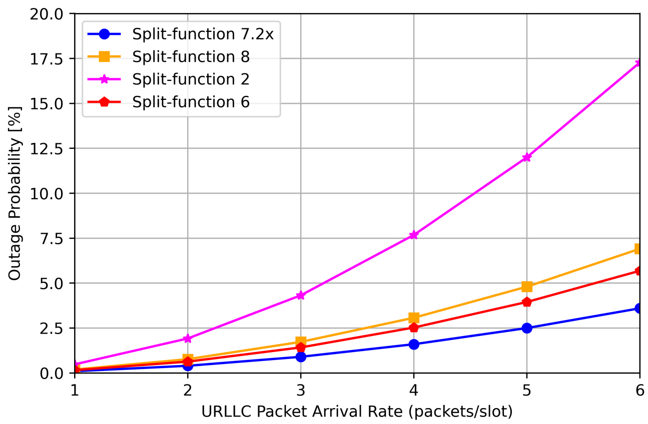
B. TS-based Twin-Delayed Deep Deterministic Policy Gradient (TD3) Solution

The TD3 algorithm builds upon the Deep Deterministic Policy Gradient (DDPG) algorithm, specifically addressing its inherent overestimation bias that can compromise policy performance [10]. Introducing twin critics within TD3 enhances stability during the learning process, which is particularly essential in the dynamic environments of URLLC. By embedding TS within the TD3 architecture, we introduce a probabilistic exploration mechanism that complements the deterministic policy updates of TD3. Twin critics ensure stability in policy learning by mitigating overestimation biases, which is crucial for maintaining consistent performance in the dynamic and uncertain environments of URLLC applications. This integration enables the framework to dynamically allocate spectrum, power, and scheduling resources while adhering to strict URLLC latency and reliability thresholds.

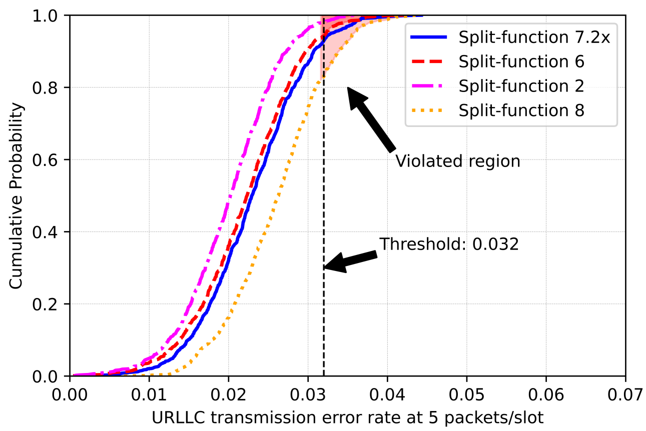
As illustrated in Figure 2, the distributed framework designates each UAV as a DRL agent operating within the near-RT RIC. This setup allows each UAV to make autonomous, real-time decisions based on localised data, significantly reducing the dependency on the central non-RT RIC for immediate decisions. Offline training conducted at the central server ensures that agents can learn from aggregated data on all UAVs, improving generalisation and speeding up convergence.

V. DRL-Powered RRM: A Case Study

In this section, we validate the effectiveness of the proposed Multi-Agent TD3-based resource allocation approach for multi-UAV networks through comprehensive simulations. The multi-UAV network is deployed within a disc radius of 500 m. Ground users are randomly and uniformly distributed



(a) Percentage of users not meeting a 1 ms latency requirement at the 10^{-4} percentile



(b) CDF of URLLC transmission error probability at 5 packets/slot

Fig. 3: URLLC latency and reliability analysis for different FSs.

throughout this area, allowing us to assess the performance of our algorithm under various traffic conditions. All UAVs are assumed to operate at a fixed altitude of 200m and are connected with a gNB and a core network. We analyse the performance of different FSs with varying eMBB and URLLC requests. The proposed DRL approach is trained using diverse communication configurations, including varying URLLC arrival rates.

In Fig. 4, we investigate the URLLC latency and reliability obtained by the proposed DRL scheme and compare the performance with the different FSs of O-RAN in NTN. Fig. 3a shows that even as the URLLC packet arrival rate increases, the FS-7.2x and FS-6 configurations effectively maintain OP well below the 10^{-4} threshold for all tested packet arrival rates. This is likely due to improved synchronisation and resource allocation between the CU and DU, reducing processing delays and more optimal use of available radio resources. Additionally, 7.2x likely benefits from more parallelism in handling data flows, minimising queueing and processing latency. This characteristic is critical for applications demanding near-zero latency, such as autonomous vehicles and remote surgery, where failure to meet the latency requirement could lead to catastrophic outcomes. In contrast, FS-2 shows a sharp increase in OP as the packet arrival rate increases, surpassing the 10^{-4} threshold at relatively low arrival rates, particularly evident at 4 packets per slot, where it approaches approximately 15% of UEs not satisfying the latency requirement. This is primarily due to its increased complexity, which is likely to incur a higher processing overhead, leading to delays in packet scheduling and transmission. This behaviour underscores FS-2's inadequacy in high-demand scenarios, illustrating its inability to support low-latency applications. FS-8, while slightly better than FS-2, still exhibits a worrying trend as it approaches the critical threshold, especially as the traffic load increases. Thus, the results reinforce the importance of selecting a robust FS configuration to ensure compliance with the 10^{-4} percentile requirement. The performance of FS-7.2x stands out as it provides a resilient solution capable of handling increasing traffic without compromising latency guarantees.

The reliability results depicted in Fig. 3b highlight the URLLC transmission error rates at five packets per slot across different FS configurations, focusing on maintaining performance under the stringent threshold of 0.032. This value reflects an acceptable reliability limit, informed by previous work and relevant performance benchmarks for URLLC in NTN and O-RAN environments [5]. From the graph, it is evident that FSs 7.2x, 6, and 2 remain within acceptable limits, while FS-8 exhibits significant deviation, exceeding the error rate threshold. FS-7.2x, represented by the solid blue line, consistently demonstrates the best performance, with its cumulative probability curve staying well below the threshold, showcasing its superior reliability for URLLC traffic. This indicates that FS-7.2x is the most suitable candidate for ensuring low-latency, ultra-reliable communication, as URLLC standards require. FSs-6 and 2, represented by the dashed red line and the dash-dot magenta line, respectively, display better behaviour, approaching the error threshold but never breaching it, making them acceptable choices for scenarios where reliability is crucial but less stringent than those requiring the highest performance. In contrast, FS-8, represented by the dotted orange line, underperforms significantly, with a substantial portion of its cumulative probability curve exceeding the threshold. This is because RU reliance on DU for essential functions like scheduling, channel coding, and modulation can result in contention issues, particularly under high load conditions, leading to resource blocking and decreased reliability. This places FS-8 in the violated region, signalling that it is unsuitable for critical URLLC applications where maintaining a low error rate is essential.

The deviation seen in FS-8 highlights the risks of choosing inappropriate configurations for high-reliability communication scenarios, as its transmission error rates exceed acceptable limits, compromising system performance. Overall, the results demonstrate that while FSs-6 and 2 provide adequate reliability, FS-7.2x offers the best balance between error rates and performance, making it the optimal choice for URLLC systems. Therefore, the analysis supports the recommendation of prioritising FS-7.2x in URLLC systems to ensure that reliability constraints are consistently satisfied, especially in environments demanding ultra-reliable, low-latency communi-

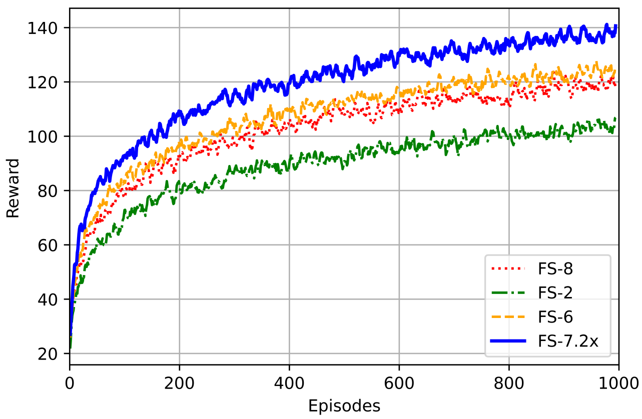


Fig. 4: Covergence behaviour of distributed DRL framework by FS for RRM in NTN

cation. The introduction of TS in TD3 helps the agent explore the action space better, leading to improved decision-making and reduced OP for URLLC. This approach is especially beneficial in environments with high uncertainty or non-constant reward distribution, which is common in URLLC scenarios.

In Fig. 4, the convergence performance of the DRL-based resource allocation framework is evaluated for different O-RAN FS in the NTN scenario. The results indicate that FS-7.2x achieves the highest reward, demonstrating superior learning efficiency and performance stability. This is attributed to its balanced trade-off between onboard processing and front-haul dependency, enabling efficient resource utilisation. FS-6 and FS-8 show moderate convergence, with FS-8 exhibiting slightly lower rewards due to increased front-haul bandwidth requirements. FS-2, which relies more on onboard processing, converges at the lowest reward level, reflecting its limited flexibility in adapting to dynamic network conditions. These findings highlight the importance of functional split selection in optimising URLLC within UAV-based NTNs.

VI. Conclusion

In this paper, we employ the DRL technique to investigate the performance of various FSs within the O-RAN framework for URLLC. We specifically compared the performance of different FS configurations, revealing significant disparities in their ability to meet the 10^{-4} OP threshold essential for URLLC. The analysis highlighted that FS-7.2x consistently outperformed the other configurations, effectively supporting higher packet arrival rates while ensuring compliance with latency requirements. Overall, our findings emphasise the importance of optimising FS configurations in O-RAN for URLLC applications, particularly in contexts requiring ultra-low latency and high reliability. As the demand for URLLC applications continues to rise, our work serves as a foundational study for future research aimed at refining O-RAN architectures and ensuring robust performance across various applications. This research paves the way for further exploration into advanced communication systems capable of meeting the challenges posed by next-generation wireless technologies.

Acknowledgement

This work was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project number (RSP2025R184).

References

[1] 3rd Generation Partnership Project (3GPP), "NG-RAN; Architecture description (Release 17)," 3GPP, Tech. Rep. TS 38.401 V17.10.0, Sep. 2024, technical Specification. [Online]. Available: https://www.3gpp.org

[2] ——, "Release 17 Description; Summary of Rel-17 Work Items (Release 17)," 3GPP, Tech. Rep. TR 21.917 V17.0.1, Jan. 2023, technical Report. [Online]. Available: https://www.3gpp.org

[3] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[4] S. S. Hassan, Y. M. Park, Y. K. Tun, W. Saad, Z. Han, and C. S. Hong, "Spaceris: Leo satellite coverage maximization in 6g sub-thz networks by mappo drl and whale optimization," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 5, pp. 1262–1278, 2024.

[5] R. M. Sohaib, S. T. Shah, O. Onireti, Y. Sambo, Q. H. Abbasi, and M. A. Imran, "Drl-based joint resource scheduling of embb and urllc in o-ran," in *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2024, pp. 1523–1528.

[6] A. K. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, S. S. Alshamrani, M. Masud, Z. Han, and C. S. Hong, "Coexistence mechanism between embb and urllc in 5g wireless networks," *IEEE transactions on communications*, vol. 69, no. 3, pp. 1736–1749, 2020.

[7] R. Khouli, L. Frank, and A. Hofmann, "Functional split evaluation in ntn for leo satellites," *IET Conference Proceedings*, vol. 2023, pp. 1–9, 2024.

[8] A. Vanelli-Coralli, N. Chuberre, G. Masini, A. Guidotti, and M. E. Jaafari, *5G Non-Terrestrial Networks: Technologies, Standards, and System Design*. Wiley-IEEE Press, January 2024.

[9] R. M. Sohaib, S. Tariq Shah, and P. Yadav, "Towards resilient 6g o-ran: An energy-efficient urllc resource allocation framework," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 7701–7714, 2024.

[10] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*. PMLR, 2018, pp. 1587–1596.

Syed Tariq Shah M.S. & Ph.D. in Electrical & Electronic Engineering, Sungkyunkwan University, Korea (2015, 2018). Lecturer at the University of Essex, U.K. Former Associate Professor at BUITEMS, Pakistan, and Postdoctoral Fellow at the University of Glasgow, U.K. Research focuses on 5G and beyond, Open RAN, AI-enabled networks, RF energy harvesting, and intelligent reflective surfaces. Editor for Electronics and reviewer for IEEE journals.

Rana Muhammad Sohaib Ph.D. in Electronics & Electrical Engineering, University of Glasgow (2024). Research Fellow at the University of York, U.K. Active in Open RAN projects (£4M+). Research focuses on radio resource management, AI-enabled wireless, energy-efficient, and vehicular networks.

Mahmoud A. Shawky B.Sc. in Electronics & Electrical Engineering, Air Defence College, Alexandria University (2012). M.Sc. in Authentication Mechanisms, Alexandria University. Ph.D., James Watt School of Engineering, University of Glasgow. Research focuses on cryptography, number theory, digital signatures, wireless authentication, and cybersecurity.

Poonam Yadav is a Senior Lecturer with the Computer Science Department, University of York, U.K., and a Visiting Research Fellow with the Computer Lab, Cambridge University. Her research interests include IoT and wireless networks, machine leaning (TinyML/UltraML), citizen science, and complex and dynamical systems.

Ahmad Almogren Professor of Computer Science, King Saud University (KSU), Riyadh. Director of Cyber Security Chair at KSU. Ph.D. in Computer Science, Southern Methodist University, USA (2002). Former Dean of CCIS and Vice Dean for Development & Quality. Served as General Chair for IEEE Smart World Symposium and TPC member for IEEE and ACM conferences. Research focuses on mobile-pervasive computing and cybersecurity.