UNICO I+D Project
6G-SORUS-RAN

SORUS-RAN-A1.1-E1 (E5)

Final version of the architecture

Abstract

This Deliverable examines the evolution of Radio Access Network architectures and the pivotal role of Artificial Intelligence in shaping modern and future mobile networks. It traces the transition from 3G to 5G, highlighting how rising performance demands have driven the integration of AI-enabled functionalities such as Self-Organizing Networks, traffic prediction, proactive resource management, anomaly detection, and network self-healing. Building on this foundation, the deliverable introduces an architectural framework that separates real-time and non-real-time control layers, aligning with vRAN and O-RAN principles to enable scalable and flexible integration with Technologies such as Unnamed Aerial Vehicles (UAVs) and Reconfigurable Intelligent Surfaces (RIS).

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES

Plan de Recuperación, Transformación y Resiliencia

Document properties

Document number	SORUS-RAN-A1.1-E1 (E4)
Document title	Final version of the architecture
Document responsible	Marco Gramaglia
Document editor	María Molina
Editorial team	Marco Gramaglia, María Molina, Juan Manuel Montes, Pablo Serrano
Target dissemination level	Public
Status of the document	Final
Version	1.0
Delivery date	28-02-2024
Actual delivery date	28-02-2024

Production properties

Reviewers	Andrés Ferré, Jonathan Almodóvar

Disclaimer

This document has been produced in the context of the 6G-SORUS Project. The research leading to these results has received funding from the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO 5G I+D programme.

Contents

List of Figures

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

List of Acronyms

Beyond 5G: B5G

Virtualized Radio Access Networks: vRAN

Reconfigurable Intelligent Surfaces: RIS

Unnamed Aerial Vehicles: UAVs

Central Units: CUs

Distributed Units: DUs

Radio Units: RUs

Radio Network Controller: RNC

New Radio: NR

Cloud RAN: CRAN

Self-Organizing Networks: SON

Automated Neighbor Relation: ANR

Machine Learning: ML

Artificial Intelligence: AI

Real-Time RAN: rtran

RAN Intelligent Controller: RIC

Network Data Analytics Function: NWDAF

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Resumen Ejecutivo

Este documento presentará una visión integral de la evolución de las arquitecturas de Radio Access Network (RAN) y analizará cómo la inteligencia artificial (AI) puede aprovecharse para su integración con tecnologías emergentes como Virtualized RAN (vRAN), Reconfigurable Intelligent Surfaces (RIS) y Unmanned Aerial Vehicles (UAV) en redes Beyond 5G (B5G). Iniciará contextualizando la transición de las arquitecturas RAN tradicionales hacia diseños basados en software y potenciados por AI, destacando cómo el aumento en las demandas de rendimiento ha impulsado la adopción de funcionalidades inteligentes dentro de la red. Posteriormente, el documento abordará casos de uso habilitados por AI, como Self-Organizing Networks (SON), predicción de tráfico, gestión proactiva de recursos, detección de anomalías y capacidades de autorreparación, los cuales son fundamentales para automatizar las operaciones de la red y mejorar su eficiencia.

A partir de los resultados de entregables previos, este trabajo sintetizará los hallazgos de nuestro análisis de escenarios de cobertura asistida por UAV, el estudio de mecanismos inteligentes de control para vRAN y la exploración de las capacidades de RIS para optimizar los entornos inalámbricos. Estos esfuerzos han mostrado cómo los UAV pueden extender dinámicamente la cobertura, cómo los controladores vRAN mejorados con AI pueden facilitar una asignación más eficiente de recursos y cómo RIS puede configurarse para mejorar las condiciones de propagación. Este documento integrará estos resultados en una propuesta de arquitectura que describe las interfaces de control y comunicación necesarias para coordinar vRAN, RIS y UAV de manera conjunta. Así, se establecerá un marco que permita una coordinación adaptativa y multidominio en las redes futuras, mostrando cómo AI puede evolucionar la RAN hacia una plataforma programable y autónoma capaz de responder a las exigencias de los sistemas B5G.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Executive Summary

This deliverable provides a comprehensive overview of the evolution of Radio Access Network (RAN) architectures and examine how Artificial Intelligence (AI) can be leveraged to enable their seamless integration with emerging technologies such as virtualized RAN (vRAN), Reconfigurable Intelligent Surfaces (RIS), and Unmanned Aerial Vehicles (UAVs) in Beyond 5G (B5G) networks. It begins by contextualizing the shift from traditional RAN architectures to AI-enabled, software-driven designs, highlighting how growing performance demands have accelerated the adoption of intelligent functionalities within the network. The document then focuses on AI-driven use cases, such as Self-Organizing Networks (SON), traffic prediction, proactive resource management, anomaly detection, and self-healing, hat serve as foundational elements for automating network operations and enhancing efficiency.

Building on the outcomes of previous deliverables, this work synthesizes insights from our analysis of UAV-assisted coverage scenarios, the study of intelligent vRAN control mechanisms, and the exploration of RIS capabilities for enhancing wireless environments. These prior efforts have shown how UAVs can dynamically extend coverage, how AI-enhanced vRAN controllers can support intelligent resource allocation, and how RIS can be configured to optimize propagation conditions. This deliverable integrates these findings into an architectural perspective that articulates the control and communication interfaces required for AI-native orchestration of vRAN, RIS, and UAVs. By doing so, it lays out a framework for enabling adaptive, multi-domain coordination in future networks, demonstrating how AI can transform the RAN into a programmable and autonomous platform capable of meeting the demands of B5G systems.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

1. Introduction

In this deliverable, we delve into the architecture of Virtualized Radio Access Networks (vRAN), a key innovation in modern wireless networks that enables the integration with new techonologes like Reconfigurable Intelligent Surfaces (RIS), and Unnamed Aerial Vehicles (UAVs).

Unlike traditional Radio Access Networks (RAN), vRAN employs virtualization to split network functionalities across Central Units (CUs), Distributed Units (DUs), and Radio Units (RUs), each optimized for specific tasks. This modular design not only streamlines resource management but also supports dynamic adaptation to real-time traffic and user demands, laying a strong foundation for the advanced requirements of Beyond 5G (B5G) networks.

1.1. Evolution of RAN Architecture from 3G to 5G

Each generation of mobile technology introduced a distinct RAN architecture. The progression from 3G's hierarchical design to 5G's flexible, cloud-friendly architecture reflects increasing demands on the network and advances in processing technology. Below we outline the key architectural features of 3G, 4G, and 5G RANs:

3G RAN (UMTS) – NodeB and RNC Separation

Third-generation networks (UMTS, as standardized by 3GPP) introduced the concept of a two-tier RAN architecture called UTRAN. The Node B is the 3G base station handling the radio transmission/reception, while a separate Radio Network Controller (RNC) manages multiple NodeBs. The RNC is responsible for radio resource control, mobility management (e.g. handovers), and encryption, essentially acting as the 3G equivalent of the earlier 2G base station controller [1]. The RNC connects the RAN to the core network and orchestrates the NodeBs under it. This split architecture meant that 3G had a centralized control entity (the RNC) overseeing the distributed radio sites.

4G RAN (LTE) – Flat Architecture with eNodeB

Fourth-generation LTE (E-UTRAN) flattened the RAN architecture by eliminating the separate RNC. The base station, now called the eNodeB (Evolved Node B), incorporates both the radio transmission functions and the control/management intelligence that was in the RNC [2]. In other words, an eNodeB is a self-contained cell site that handles everything from signal processing to handover decisions and radio resource management. This simplification to a single-node RAN (often termed a "flat" architecture) reduces latency and complexity in the access network. An LTE eNodeB not only provides the radio link to user devices but also directly performs tasks like scheduling, load balancing between cells, and mobility control, effectively doing the job that NodeB+RNC together did in 3G [2]. This architectural shift was a major enabler for LTE's high data rates and lower latency, but it also meant that as networks grew denser, coordinating many autonomous eNodeBs became a new challenge.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

5G RAN (NR) – Disaggregated and Flexible Architecture

Fifth-generation New Radio (NR) RAN, standardized by 3GPP for 5G, inherits some elements from LTE but introduces greater functional split and flexibility. The main RAN node is the gNodeB (gNB), analogous to an eNodeB in LTE [3]. However, unlike the monolithic 4G eNodeB, a 5G gNodeB can be functionally split into a CU and DU – a major architectural change [4]. The gNB-CU is a centralized logical node that can handle higher-layer RAN protocols (RRC signaling, packet routing, and part of the PDCP/RLC layers), while the gNB-DU is a distributed node closer to the radios, handling real-time lower-layer tasks (the physical layer, MAC scheduling, etc.). A standard interface called F1 links the CU and DU, allowing them to operate as a cohesive gNodeB [5].

This split architecture in 5G enables deployment flexibility. For example, DUs can be located at cell sites or edge data centers for low-latency processing, while CUs might be centralized to aggregate traffic from many DUs and perform coordinated scheduling or mobility management. In practice, one CU can control multiple DUs (supporting multiple cells) [WKM2022], and the CU itself can be further split into control-plane and user-plane components (CU-CP and CU-UP) interconnected by an E1 interface [3GPP2025]. This modular design supports the Cloud RAN (C-RAN) concept: baseband processing (CU functionality) can be pooled in the cloud or data center, while minimal hardware (DU and radio units) reside at remote sites. The result is a more software-driven RAN that can scale and adapt — key for 5G features like network slicing and ultra-low latency use cases. Importantly, 5G still uses the concept of an X2/Xn interface (now Xn in 5G) between gNodeBs for inter-cell coordination, like LTE's X2, ensuring mobility and load management across distributed units.

3GPP Standardization Efforts Driving RAN Evolution

The generational changes in RAN were orchestrated by the 3GPP, the global standards body that develops mobile system specifications. Each major generation corresponds to a set of 3GPP Releases with new architecture definitions and capabilities [9]:

- **3G/UMTS:** Introduced in 3GPP Release 99/4/5 (early 2000s), defining the UTRAN with NodeB/RNC. Subsequent releases added features like HSPA but kept the same basic RAN split.
- **4G/LTE:** Standardized starting in Release 8 (2009), with an all-IP flat architecture (E-UTRAN). Release 10 added LTE-Advanced features (carrier aggregation, etc.), and later releases improved LTE further. [10]
- **5G/NR:** Initial specifications in Release 15 (2018) with the new NG-RAN architecture and service-based 5G core. Release 16 and 17 added enhancements (e.g., URLLC improvements, multi-hop IAB for backhaul). By Release 17, 5G supports almost all LTE services plus new 5G-specific ones, ensuring smooth interworking between 4G and 5G systems.

3GPP's RAN working groups have been pivotal in these transitions, not only defining radio protocols but also standardizing how RAN nodes interconnect. For example, the definition of the gNB CU/DU split and the F1 interface was part of Release 15's NG-RAN architecture work (documented in TS 38.401 and TS 38.300) 3gpp.org. This standardization gives operators confidence that multi-vendor CUs and DUs can interoperate, which is important for the global transition to 5G.

Standardized RAN Management and SON

3GPP also recognized early on that as networks became more complex, automation would be key. A concept called Self-Organizing Networks (SON) was introduced in Release 8 (the same release that brought LTE) [6]. SON refers to a collection of features for automated network configuration, optimization, and healing. However, 3GPP mostly specified SON at a conceptual level – for instance, providing standard measurement reports (performance metrics, KPIs) that an operator's SON algorithms could use [7]. The actual intelligence (the algorithms that adjust parameters or reconfigure the network) was left to implementations, not fixed in the standards. This approach allowed vendors to innovate with proprietary SON solutions on top of a common data framework. Over time, SON features have matured (automatic neighbor relation setup, mobility load balancing, outage detection, etc.), and more operators have adopted them as networks grew denser and more dynamic. For example, LTE introduced features for Automated Neighbor Relation (ANR) and Mobility Robustness Optimization, which are classic SON functions defined in 3GPP specs (like TS 36.300 series) as optional capabilities.

Ongoing 3GPP Work (5G-Advanced)

The story doesn't end with initial 5G rollout. 3GPP continues to enhance the RAN in Release 18 and beyond (5G-Advanced). Two important threads stand out in current standardization: further network automation and AI/ML integration. In Release 16 and 17, features like enhanced Minimization of Drive Tests (MDT) were introduced, allowing the network to collect UE measurements and performance data to drive optimization. Release 18 is continuing this with more SON enhancements – for instance, improving mobility management through features like conditional handover and enhanced reporting [6]. According to the 3GPP RAN3 group, the Rel-18 SON/MDT work focuses on better data collection and signaling to enable advanced automation, helping operators improve network performance and maintenance efficiency [7].

Moreover, Release 18 marked the first explicit study and work item on AI and Machine Learning for NG-RAN. A 3GPP study item (TR 37.817) investigated a framework for using AI/ML in the RAN, which led to a new Release 18 work item specifying how RAN nodes can collect and exchange data to support AI-driven algorithms [8]. In Rel-18, 3GPP standardized signaling and data collection mechanisms for three use cases: network energy saving, load balancing, and mobility optimization. For example, RAN nodes can share information over the Xn interface about traffic loads or mobility events to feed an ML model's inputs. The idea is that by enabling these hooks, the actual AI algorithms (which might run in a network management system or RAN node) have the data they need standardized and can even be multi-vendor. After completing Rel-18, 3GPP immediately launched further study in Rel-19 to extend AI/ML support to new use cases like network slicing

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

optimization and Coverage/Capacity Optimization (CCO), and to address AI in the context of the split RAN architecture (CU/DU). This ongoing work shows that 3GPP is actively embracing AI as part of future RAN design, ensuring that standards keep pace with innovation in network intelligence.

It's also worth noting that global adoption of 3GPP standards has made these transitions truly worldwide. As of recent counts, over 800 operators deployed 4G LTE and hundreds are investing in 5G [9]. Many operators are now sunsetting 3G networks to refarm spectrum for 4G/5G. The standardized path by 3GPP (and backward compatibility features like dual-mode 4G/5G radios or interworking functions) enables this generational coexistence and transition with minimal user disruption [9].

1.2. AI in Radio Access Networks: Use Cases and Developments

As RAN architectures have become more software-driven and data-rich, the door has opened for Artificial Intelligence (AI) and Machine Learning (ML) techniques to play a significant role in RAN operation and optimization. 5G networks, in particular, are exceedingly complex – with a variety of services (eMBB, URLLC, mMTC), massive numbers of parameters, and dynamic conditions. AI/ML promises to handle this complexity by learning patterns, making intelligent decisions in real-time or near-real-time, and automating tasks that traditionally required manual tuning or simple heuristics.

Industry Initiatives and Architectural Support: The integration of AI in RAN is supported by industry initiatives like the O-RAN Alliance, which defines an architecture to inject intelligence into the RAN. A key O-RAN concept is the RAN Intelligent Controller (RIC) – a software-defined component (not part of 3GPP specs but built to work with them) that hosts third-party applications for RAN control [10]. The RIC comes in two flavors: Non-Real-Time RIC (in the operator's management plane, making decisions slower than 1 second) and Near-Real-Time RIC (at the edge of the RAN, making decisions in ~10 ms to 1 s timescales). These RICs allow operators to deploy rApps and xApps – software plugins that use AI/ML to optimize specific aspects of RAN behavior (for example, a handover optimization xApp or an interference mitigation xApp). The RIC framework brings multivendor interoperability, agility, and programmability to RAN control, essentially providing a platform for closed-loop control and analytics in the RAN [10]. While 3GPP provides the data and interface standards, O-RAN provides a practical means to experiment with and deploy AI algorithms in live networks. Indeed, many current prototypes and trials of AI in RAN are built on RIC platforms in Open RAN environments.

Below, we highlight several key use cases of AI in RAN and give examples of development results (from scientific studies, prototypes, or trials) that demonstrate these capabilities:

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Asuntos Económicos y Transformación Digital · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D

Self-Organizing Networks (SON) and Automated Optimization

One of the earliest and broadest areas for RAN intelligence is Self-Organizing Networks (SON). SON encompasses a range of automation techniques for self-configuration, self-optimization, and self-healing in mobile networks [11]. In practice, SON functions include tasks like automatically configuring new cells, optimizing neighbor relations and handover parameters, balancing load between cells, and detecting/recovering from cell outages. AI and advanced algorithms enhance SON by enabling the network to *learn* and adapt policies based on experience, rather than relying only on pre-defined rules.

In LTE and 5G, SON functions are increasingly aided by machine learning:

- **Mobility Load Balancing and Handover Optimization:** ML models can learn from statistics (drops, throughput, signal quality) to adjust handover thresholds or load balancing parameters dynamically. For example, reinforcement learning has been applied to optimize handover decisions and reduce ping-pong events, as noted in research on deep Q-learning for SON fault management and performance improvement [11]. A network that learns the optimal settings for each cell based on time of day or traffic patterns can maintain user experience with minimal human tuning.

- **Coverage and Capacity Optimization (CCO):** AI can analyze coverage maps (including user equipment measurements) to identify coverage holes or areas of overlap and then recommend adjustments like antenna tilt changes or power tweaks. These were traditionally done by drive-testing and human planning, but now networks can use data (including 3GPP MDT reports) to self-optimize coverage. Clustering algorithms or neural networks might detect regions of poor signal and trigger a SON action to fix it.

- **Self-Healing (Fault Management):** When a cell goes down (hardware failure or backhaul issue), SON algorithms attempt to compensate by adjusting neighbor cells. AI-based anomaly detection (discussed more below) can quickly flag a cell outage or performance degradation, and then SON can automatically mitigate it (for example, neighboring cells boost their power to cover the gap). This closed-loop fits naturally with AI: the detection is done via ML, and the response can even be learned (e.g., learning the best compensation strategy for various outage scenarios).

It's important to note that SON as standardized by 3GPP provides the framework (what measurements to use, etc.) but not the brains. Vendors and operators have developed AI-driven SON solutions to fill that gap. In essence, SONs can be made "intelligent" by incorporating machine learning models that adapt to network conditions autonomously [12]. A self-optimizing network might use a neural network to predict future cell traffic and pre-emptively reconfigure parameters, or use an expert system to decide when to switch on small cells in hot-spots. The need for such intelligence is growing – as a 3GPP expert noted, networks are becoming too complex and dynamic to rely on static configurations [7].

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Real-world development examples: Many operators have trialed SON algorithms. For instance, Nokia's SON (e.g., MantaRay SON) and other vendor solutions now embed AI for tasks like mobility management and interference optimization [13]. Another example is from NTT DoCoMo and other carriers experimenting with deep learning to optimize base station parameters in dense urban 5G scenarios (reports have shown prototypes where an AI engine in the RAN adjusts beamforming or scheduling policies in response to live network data). While specific performance metrics are often proprietary, the trend is clear: AI is increasingly the engine underneath SON functionalities in modern networks.

Traffic Prediction and Proactive Resource Management

Mobile traffic loads vary widely by location and time, and being able to predict traffic patterns is highly valuable for network planning and real-time resource allocation. AI, particularly machine learning, has proven adept at forecasting in such complex systems. Traffic prediction in RAN involves using historical data (and possibly external data like events, weather, etc.) to anticipate how many users or how much throughput a cell will need in the near future.

Use cases and prototypes in this area include:

- **Capacity Planning and Network Planning:** Planning teams are using ML models to forecast growth and decide where to add new cell sites or capacity. For example, models can predict that every day at 8 PM a cluster of cells experiences high video traffic, suggesting adding a small cell or carrier. Traditional planning was reactive and coarse; AI allows a more precise and *proactive* approach, considering multidimensional data. The 5G PPP has reported on AI-assisted network planning solutions where algorithms optimize base station placement or configuration by evaluating coverage, demand patterns, and even cost factor [14].
- **Dynamic Resource Allocation:** In operational networks, short-term traffic prediction (on the order of minutes or hours) allows the network to dynamically allocate resources. For example, an AI model might predict a surge of users in a cell (perhaps a train arriving at a station) and can prompt the network to prepare by increasing that cell's capacity (e.g., by adjusting scheduling, pre-empting some bandwidth, or spinning up an extra carrier or small cell). Similarly, predictive models can inform elastic cloud RAN scaling – if a spike in load is predicted, the operator can allocate more cloud resources to the DU/CU pool in advance. [15]
- **Intelligent Scheduling:** At a finer level, real-time traffic prediction feeds could be used by the base station scheduler. For instance, forecasting the data rates or QoE needs of users could allow more efficient scheduling algorithms that improve throughput or fairness. One research direction uses deep learning to forecast user throughput based on past patterns, enabling the scheduler to reduce latency or meet QoS by scheduling users at the optimal times.

Several studies demonstrate these capabilities. A deep learning approach for 5G traffic prediction showed that neural networks can capture complex temporal patterns in base station loads [14]. By

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

predicting traffic with high accuracy, the system can perform *what-if* analyses: e.g., "If current trends continue, cell A will be overloaded in 10 minutes", then proactively offload users or activate a dormant cell. In testbed environments, this has been prototyped to optimize video streaming QoE, where the network pre-positions resources for anticipated demand surges [16]. Another prototype integrated an AI prediction engine with a RAN controller such that it would trigger a scale-out of the CU/DU cloud resources 5 minutes before a busy hour, thus avoiding congestion without permanent over-provisioning.

In summary, traffic prediction with AI is moving from theory to practice. Operators like China Mobile and Orange have publicly discussed using AI to forecast traffic for their 5G networks, enabling dynamic capacity management. This is especially relevant as 5G supports features like on-demand network slicing – accurate predictions can ensure each slice gets the right resources ahead of time. The end goal is a RAN that not only reacts to current traffic, but also prepares for future demand (and does so in an automated fashion).

Anomaly Detection and Network Self-Healing

Operating a RAN involves monitoring vast numbers of metrics and logs to ensure everything is working correctly. Anomaly detection refers to identifying unusual patterns that could indicate faults, performance degradation, or security threats. AI, especially in the form of advanced analytics and machine learning, has become an invaluable tool for anomaly detection in RAN because it can sift through high-dimensional data and find subtle issues far more effectively than manual thresholds.

Key use cases and developments:

- **Fault and Outage Detection:** RAN equipment and links can fail or underperform. Instead of waiting for users to complain or for crude alarms (e.g., "site down"), ML models can learn the normal patterns of network KPIs and detect deviations. For instance, an algorithm monitoring a cell's traffic, throughput, and signal quality might learn what "normal" looks like at various times. If the cell's throughput suddenly drops well below predicted values or the drop-call rate spikes, the system flags an anomaly which could mean a partial outage (like a sector antenna failure) or interference issue. Researchers have applied techniques like autoencoders or clustering to network performance data to detect such faults in near-real-time, often referred to as network anomaly detection using unsupervised learning [17].
- **Intrusion and Attack Detection:** Security is a growing concern in 5G. AI can help detect unusual signaling patterns that might indicate a malicious attack (e.g., jamming, spoofing, or signaling storms). For example, a base station might see an unusual surge of connection requests or repetitive attach/detach attempts – patterns which ML can catch as potential Denial-of-Service attempts. Projects like Simba have used graph neural networks to do root-cause analysis of anomalies in 5G RAN, distinguishing anomalies caused by configuration issues from those caused by security attacks [18]. AI can also monitor for rogue base stations or unauthorized transmitters by learning the spectrum environment.

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Asuntos Económicos y Transformación Digital · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D

- **Closed-Loop Self-Healing:** Once an anomaly is detected, AI can also assist in taking corrective action. This overlaps with SON self-healing – for example, automatically rebooting a misbehaving cell, or adjusting neighbor parameters if one cell is off. In advanced scenarios, an AI agent might even localize the fault (e.g., pinpoint that a specific antenna module is faulty) by correlating data across the network. [19]

Prototypes and trials

The academic and open-source community has built some prototypes to showcase AI-based anomaly detection. A notable example is MobiWatch, an O-RAN compliant xApp (application on the RAN Intelligent Controller) that uses unsupervised deep learning to detect anomalies and attacks in a 5G network [20]. MobiWatch listens to RAN signaling (RRC/NAS messages) and builds an ML model of normal behavior; it can then flag anomalies like suspicious connection flows that could indicate a security issue. This was demonstrated on a 5G testbed with an open-source RIC, illustrating how operators could deploy AI-driven security monitors in their RAN. Another demonstration (from a 6G security project) showed a digital twin of the 5G RAN detecting anomalies in connectivity in real-time [21], hinting at how future networks might integrate AI for continuous monitoring.

On the commercial side, operators have started using AI in Network Operations Centers (NOCs) to monitor RAN health. For example, Vodafone's trials with an AI system to analyze cell performance data found it could detect issues hours before traditional alarms would fire, allowing proactive maintenance (this trial was part of an AI-based RAN optimization partnership with Nokia, as reported in industry news). The AI-RAN Alliance, a new industry group, is also focusing on use cases like anomaly detection as a key to autonomous networks [22]. And 3GPP's management architecture in 5G includes the concept of a Network Data Analytics Function (NWDAF) which, while core-network oriented, can aggregate data to detect anomalies across the network.

In summary, AI-powered anomaly detection is becoming the "eyes" of modern RANs, enabling quicker and more precise identification of problems, which in turn feeds into automated recovery actions. This significantly reduces downtime and improves security in large, complex 5G networks.

Energy Efficiency and Dynamic Energy Savings

Energy consumption in RAN is a significant operational cost, and it has implications for environmental sustainability. 5G, with its dense small cell deployments and massive MIMO radios, can consume even more power if not managed intelligently. Thus, a compelling use case for AI in the RAN is energy optimization – finding ways to save power during low traffic periods or optimize transmit power in real-time without degrading user experience.

SON mechanisms for energy saving have been discussed for years (sometimes called self-organizing energy saving features). A typical scenario: during late night hours, traffic may be very low in some cells, so an operator could turn off certain carriers or even entire base stations, while neighboring cells expand their coverage to fill the gap. In the morning, the sleeping cells are turned on again as demand rises. Doing this efficiently is a complex decision problem – you must predict when and where to deactivate cells and ensure a smooth transition.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

AI comes into play by using predictive and adaptive algorithms:

- **Traffic Forecast for Energy Saving:** AI can forecast when a cell will have consistently low load, enabling a smarter on/off schedule than fixed time-based rules. Rather than a static nightly schedule, an ML model might learn that some weeknights before a holiday have even lower traffic, or that certain cells see sporadic late usage (preventing an early shutdown).
- **Coverage Optimization during Sleep Mode:** When some cells are turned off, the neighbors need parameter adjustments (tilt, power) to cover the area. AI can assist by learning the optimal configurations that minimize power while maintaining coverage. This might involve reinforcement learning agents controlling antenna parameters in simulation and then applying to the network. [23, 24].
- **Equipment Efficiency Adaptation:** Modern base stations have features like dynamic voltage scaling, sleep modes for components, etc. AI policies can decide, for example, to put a massive MIMO array into a lower-power mode when only a few users are active (since full beamforming gains aren't needed). These decisions could be based on learned policies that trade off energy vs. performance.

Notably, 3GPP's Release 18 AI study explicitly included network energy saving as a use case [23]. The standardization effort is to ensure RAN nodes can exchange information (like which cells are candidates for sleep, or current load levels) to facilitate AI-driven energy management across the network. This shows industry consensus that AI can help reduce RAN energy usage without impacting service – a win-win for operators.

In terms of results, some field trials have been reported. Telefónica, for instance, ran trials of an AI system that dynamically shuts down a portion of its 4G/5G cells during low traffic periods and reported double-digit percentage energy savings while keeping user impact minimal (by carefully timing the shutdowns). Another example is a prototype where an AI agent in the non-RT RIC (O-RAN context) computes an optimal nightly plan for cell activation/deactivation for a cluster of sites, and sends those policies to the network – this prototype was cited in an O-RAN Alliance plugfest demonstrating energy saving use case with AI. On the academic side, research papers have modeled energy-saving as an optimization problem and used algorithms like deep reinforcement learning to solve it, verifying in network simulators that they can turn off ~20% of base stations in a city at night for energy reduction with negligible coverage loss.

The benefits are not just cost savings; they also reduce wear on equipment and cooling systems. As 5G evolves, AI-driven energy management will likely become a standard operations practice, especially as networks densify (and we head towards beyond-5G or 6G where network densification continues).

Emerging AI Use Cases and Trials in RAN

Beyond the major categories above, there are other nascent use cases for AI in the RAN:

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

- **Network Slicing Management:** In 5G, multiple logical networks (slices) share the same RAN. AI could predict slice demand and enforce slice-wise resource allocation to meet SLAs. 3GPP's Rel-19 study is looking at AI for slicing in RAN [25]. Early prototypes use AI to predict if, say, a video slice will need more PRBs (Physical Resource Blocks) and reallocate from a low-use IoT slice temporarily.

- **Advanced Antenna Control:** AI can optimize beamforming in massive MIMO systems. For example, machine learning can help choose the best beam or even shape adaptive beams by learning the environment (this crosses into physical layer research – e.g., deep learning for channel state information feedback or beam prediction). Some demos (like by universities using AI radios) have shown that RL agents can control analog beamforming networks to maintain link quality in mmWave systems with less overhead than exhaustive search.

- **Multi-RAT Coordination:** With 4G and 5G co-existing, AI can assist in deciding which technology serves a user (offloading between LTE and NR, or managing dual connectivity). An AI policy might learn to keep a user on LTE in certain scenarios to offload NR, or vice versa, based on performance data.

- **Field Trials – AI in the Wild:** A relevant initiative in this space is Colosseum, the world's largest wireless network emulator, hosted at Northeastern University. Designed to support AI-native wireless research, Colosseum enables real-time, over-the-air experimentation with AI-enabled Radio Access Networks (AI-RANs) [26]. It allows researchers to deploy AI-driven algorithms directly into the RAN loop, exploring adaptive control, spectrum sharing, and autonomous reconfiguration in complex wireless environments. The testbed integrates software-defined radios, large-scale channel emulation, and edge computing capabilities—creating a flexible platform where communication and computation converge. Colosseum exemplifies how future mobile networks can evolve into intelligent, learning-enabled infrastructures, where the network not only delivers connectivity but also becomes a distributed AI host. This aligns with the vision of next-generation AI-native RANs and supports rapid prototyping of integrated 5G/6G and AI applications.

- **RIC and xApp Experiments:** Companies like Vodafone have conducted trials with RIC platforms – for example, testing a vendor's RIC with an admission control xApp and a slice SLA assurance xApp [27]. In one trial, Vodafone, Juniper, and Parallel Wireless showed that a RIC could intelligently control a live network slice's resources to guarantee its throughput. These trials, often showcased in Plugfests or conferences, demonstrate that the AI control loops can be inserted into operational networks without negatively affecting stability – a critical step toward adoption.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

2. General system architecture

Building on this foundation, we explore the integration of vRAN with complementary technologies, including UAVs and RIS. UAVs act as mobile network nodes, extending connectivity to areas with insufficient or temporary infrastructure, while RIS enhances signal propagation by intelligently manipulating electromagnetic waves. By combining these technologies with the core vRAN architecture, we aim to create a robust, adaptive, and energy-efficient system capable of meeting the diverse challenges posed by next-generation networks. This integration leverages the strengths of each component to enhance coverage, improve resource allocation, and optimize network performance in dynamic environments.

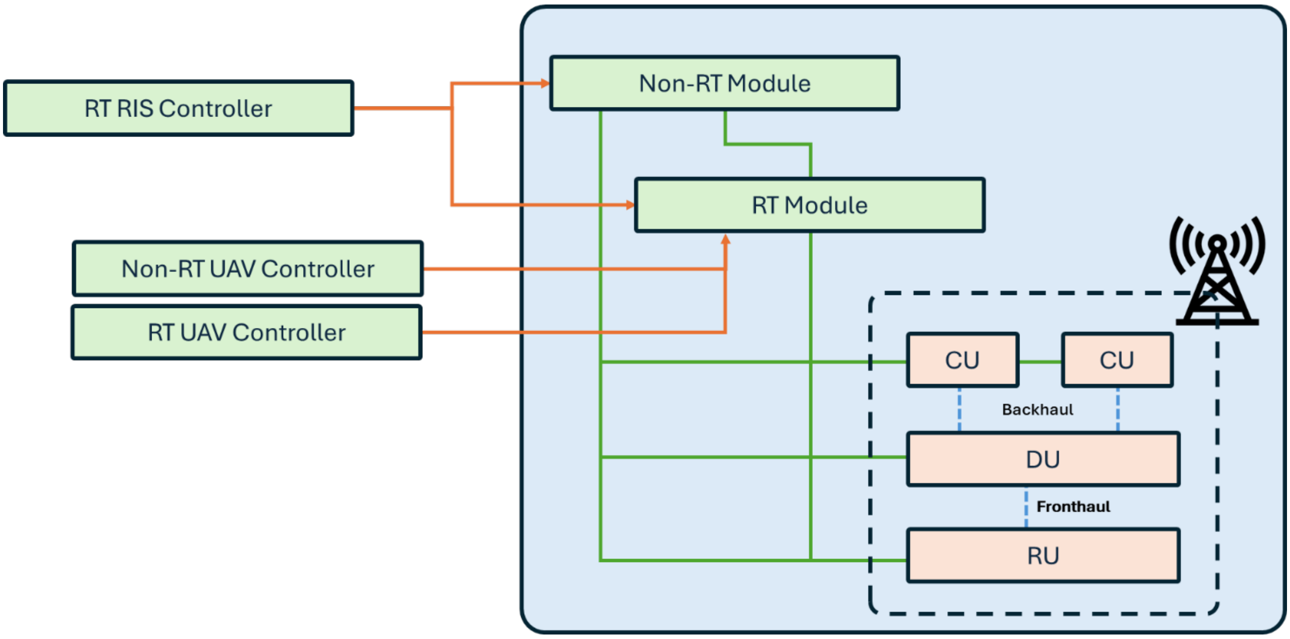

FIGURE 1 INTEGRATED CONTROL ARCHITECTURE FOR UAV-ASSISTED RIS IN VIRTUALIZED RAN (VRAN) ENVIRONMENTS

This figure illustrates the integration of a vRAN architecture with RIS an UAVs, highlighting the interplay between real-time (RT) and non-real-time (Non-RT) control modules to optimize network performance dynamically.

Core vRAN Components (Bottom-Right Box):

- The **CUs**, **DUs**, and **RUs** are shown as part of the vRAN architecture. These components are interconnected via **backhaul** and **fronthaul** links:
  - **CUs**: Perform high-level management tasks like mobility management, resource allocation, and slicing. Multiple CUs can cooperate to manage traffic from the DUs and RUs dynamically.
  - **DUs**: Handle latency-sensitive baseband processing tasks, enabling real-time signal processing and efficient management of user connections.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

- **RUs**: Are deployed at cell sites to perform RF functions like signal transmission and reception. These interface directly with user devices, ensuring optimal connectivity.

Real-Time (RT) and Non-Real-Time (Non-RT) Modules:

- The **RT Module** is responsible for managing time-sensitive operations that require low latency, such as:
  - Adjusting RIS configurations in response to environmental changes (e.g., modifying phase shifts to improve signal propagation).
  - Managing UAV positioning to address dynamic traffic demands and provide on-the-fly coverage enhancements.
- The **Non-RT Module** oversees long-term, higher-level optimization and configuration tasks, such as:
  - Planning RIS deployments and UAV flight paths based on historical traffic data and network conditions.
  - Configuring network slicing policies to allocate resources to different services (eMBB, URLLC, IoT).

Integration with RIS and UAVs:

- **Reconfigurable Intelligent Surfaces (RIS)**:
  - The **RT RIS Controller** is shown interfacing with the Non-RT and RT Modules. It ensures that RIS elements are dynamically reconfigured to optimize signal propagation. For example, RIS can reflect signals toward areas with high user density or mitigate signal blockages in urban environments.
- **Unmanned Aerial Vehicles (UAVs)**:
  - UAVs are controlled through two layers:
    - The **Non-RT UAV Controller** handles strategic decisions, such as determining deployment areas and scheduling UAV operations.
    - The **RT UAV Controller** handles real-time adjustments, such as fine-tuning UAV positions to address temporary demand surges or adapting their role as relay nodes to maintain seamless connectivity.

Interconnections:

- Green lines represent control and communication paths between components, ensuring tight coordination between Non-RT and RT Modules.
- Orange lines show real-time interactions, such as immediate updates from the RT RIS Controller and RT UAV Controller to the RT Module.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

This architecture highlights how RIS and UAVs complement vRAN by adding layers of adaptability and coverage enhancement. RIS improves spectral efficiency by dynamically directing signals, while UAVs provide mobile, on-demand infrastructure to address coverage gaps or congestion. The split between RT and Non-RT control ensures a balance between long-term planning and real-time responsiveness, enabling the network to meet the diverse demands of Beyond 5G (B5G) applications, such as ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and IoT connectivity.

2.1. Real-Time RAN Controller

Figure 2 represents the architecture of the Real-Time RAN (RT RAN) Controller, a critical component in managing and optimizing the operations of a virtualized Radio Access Network (vRAN). It demonstrates the modular and highly adaptable design of the controller, which incorporates multiple functional components that work together to enable efficient real-time network management.

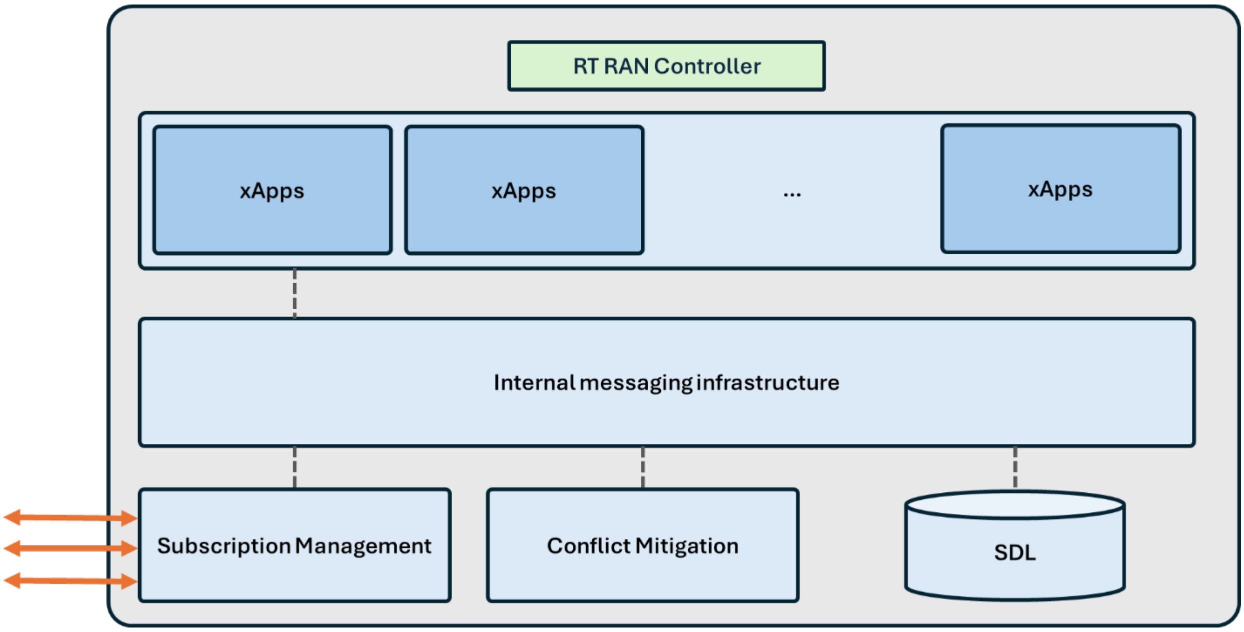

FIGURE 2 REAL-TIME RAN CONTROLLER.

Key Components and Their Roles

1. **RT RAN Controller**:
    - The RT RAN Controller serves as the central decision-making unit, managing and orchestrating real-time operations within the RAN.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

  - It collects input from various components and external sources, processes this information, and generates actionable outputs to optimize network performance, such as resource allocation, handover management, and interference mitigation.

2. **xApps**:
   - **xApps** are modular applications running on the RT RAN Controller that perform specific real-time tasks.
   - These tasks can include load balancing, mobility management, beamforming optimization, or Quality of Service (QoS) adjustments.
   - Multiple xApps can operate concurrently, each addressing a specific network requirement, and they are designed to be independent and extensible, allowing operators to deploy or update applications dynamically.
   - **Interaction**: xApps communicate through the internal messaging infrastructure to coordinate and share information, enabling seamless collaboration and decision-making.
3. **Internal Messaging Infrastructure**:
   - This layer is the backbone of communication within the RT RAN Controller, facilitating the exchange of messages between xApps, subscription management, and conflict mitigation modules.
   - It ensures low-latency and reliable communication, critical for real-time decision-making processes.
   - The infrastructure abstracts the complexity of message routing, allowing xApps and other modules to interact without direct dependencies.
4. **Subscription Management**:
   - The subscription management module handles the registration and notification system for xApps and other components.
   - It allows xApps to subscribe to specific events or data streams, such as user mobility updates, traffic patterns, or radio interference metrics.
   - When relevant events occur, this module pushes the information to subscribed xApps, ensuring they have up-to-date data to make decisions.
5. **Conflict Mitigation**:
   - This module resolves conflicts that may arise from the concurrent operation of multiple xApps, which might generate conflicting commands or compete for the same resources.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

  - For example, if one xApp attempts to allocate additional resources to a high-priority service while another xApp tries to prioritize a low-latency slice, the conflict mitigation module ensures that these actions are harmonized based on predefined policies or real-time network conditions.
  - This ensures a consistent and stable network operation, even under high demand.

6. **SDL (Shared Data Layer)**:
  - The SDL serves as a shared database accessible to all components of the RT RAN Controller.
  - It stores both static information, such as configuration parameters and policies, and dynamic data, such as real-time network metrics, historical performance data, and xApp states.
  - This centralized storage ensures that all components have access to a unified and consistent data source, reducing redundancy and improving coordination.

Interaction Between Components

- The RT RAN Controller acts as the coordinator, enabling each component to contribute to real-time network optimization.
- **xApps** request and receive data via the **Subscription Management** module, ensuring they operate with the latest information about the network state.
- When multiple xApps produce actions that overlap or conflict, the **Conflict Mitigation** module intervenes to enforce policies and resolve contradictions.
- All components rely on the **Internal Messaging Infrastructure** to exchange data and commands efficiently, ensuring low latency and high reliability.
- The **Shared Data Layer (SDL)** provides a central repository for both real-time and historical data, enabling data-driven decisions and improving the performance of algorithms implemented in xApps.

2.2. Non-Real-Time RAN Controller

Figure 2 illustrates the Non-Real-Time RAN Intelligent Controller (Non-RT RIC), a component within the Service Management and Orchestration (SMO) Framework. It operates on a non-real-time scale to provide policy, configuration, and machine learning-based analytics that guide the near-real-time RIC and other network components, including UAVs and RIS. The Non-RT RIC focuses on long-term network optimization and decision-making, complementing real-time and near-real-time operations to ensure the overall efficiency and adaptability of the network.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

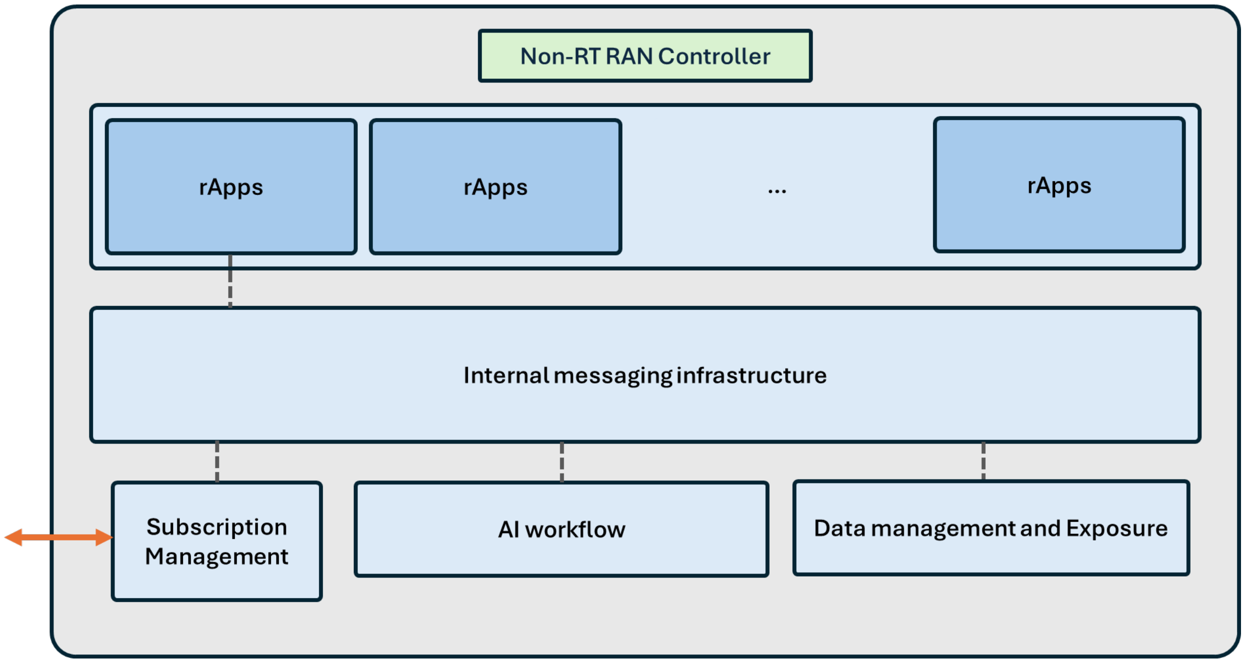

FIGURE 3 NON-RT RAN CONTROLLER AND ITS COMPONENTS

Key Components of the Non-RT RIC and SMO Framework:

1. **Non-Real-Time RIC**:
    - **rApps**: These are applications hosted on the Non-RT RIC, responsible for specific non-real-time tasks such as traffic forecasting, policy generation, and training AI/ML models for network optimization. For example:
        - One rApp could predict long-term user mobility patterns, influencing the strategic positioning of UAVs.
        - Another rApp might optimize the phase configuration of RIS to improve signal propagation in specific areas during peak hours.
    - **R1 Termination**: This interface facilitates communication between the Non-RT RIC and the near-real-time RIC (Near-RT RIC). Policies, configurations, and model updates from the Non-RT RIC are sent to the Near-RT RIC to guide its real-time operations.
2. **SMO/Non-RT RIC Framework Functions**:
    - **Data Management and Exposure**:
        - Collects, processes, and exposes network data from components such as UAVs, RIS, and vRAN elements (CUs, DUs, RUs).
        - This data is critical for training AI/ML models and for providing analytics to rApps.

    - For example, UAV telemetry and RIS performance metrics could be aggregated here to support predictive modeling and long-term optimization.

- **AI/ML Workflow**:

    - The AI/ML framework enables the training and deployment of models used by rApps. Models can predict user behavior, optimize UAV routes, or determine the optimal configuration for RIS to improve network coverage and efficiency.

Two key research contributions provide foundational capabilities for rApps in this architecture:

Infrastructure Optimization rApp: An optimization framework for auto-scaling server farms [28] A recent research contribution introduces an optimization framework for auto-scaling server farms in virtualized mobile networks. The core idea is to determine the best server type and quantity to meet application-specific reliability constraints, while minimizing both capital and operational expenses. This is achieved by combining a queueing-theoretic model, which estimates the resources needed to fulfill reliability guarantees, with a cost model that accounts for both infrastructure and energy consumption. The approach is validated through simulations, showing that it achieves 22% cost reduction while remaining within 3% of exhaustive-search solutions, and with far lower computational complexity.

In the context of the O-RAN architecture, this solution fits naturally as a Non-RT RIC rApp tasked with long-term infrastructure optimization. The rApp would monitor performance metrics (e.g., traffic volume, SLA compliance) exposed via the O1 interface, and periodically run its optimization logic to suggest scaling actions. These recommendations are transmitted either directly to the O-Cloud orchestrator through the O2 interface, or to the Near-RT RIC via A1 policies when short-term adjustments are also needed. This enables more intelligent provisioning of compute resources for functions like UAV control or RIS beam management, ensuring that mission-critical services run with high reliability and at minimal cost.

DiWi: A Privacy-Preserving Mobility Generator: Another important contribution is DiWi, a Transformer-based model that generates synthetic spatiotemporal mobility traces. DiWi is trained on real wireless network logs to learn behavioral patterns of users, but it generates traces that preserve key statistical features without exposing identifiable user information. The model is evaluated through privacy metrics such as membership inference and similarity tests, confirming that it protects privacy while remaining useful for a wide range of applications, such as mobility prediction, resource optimization, or coverage planning.

Within the O-RAN ecosystem, DiWi can be deployed as a data-generation rApp in the SMO's AI/ML framework. It consumes historical logs and produces realistic, privacy-safe synthetic mobility data that other rApps can query via the R1 interface. This

Financiado por la Unión Europea NextGenerationEU · GOBIERNO DE ESPAÑA · MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL · SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D

supports use cases like UAV trajectory optimization or RIS phase tuning, without requiring access to sensitive user traces. DiWi helps maintain regulatory compliance (e.g., GDPR) while enabling continuous model training and simulation. As such, it strengthens the AI-driven capabilities of the Non-RT RIC while embedding privacy-by-design principles into the network's control loop.

Internal Messaging Infrastructure:

- This infrastructure ensures seamless communication between rApps, SMO functions, and other components. It enables rApps to access shared resources and data, ensuring consistency and scalability in decision-making.

3. **SMO's Core Functions**:
   - Includes tasks such as policy generation, inventory management, network design, and configuration. These functions create the high-level guidelines and policies that govern UAV and RIS operations.
4. **External Interfaces**:
   - O2 Termination: Connects the Non-RT RIC to the cloud (O-Cloud), providing access to computational resources for AI/ML training and large-scale data analytics.
   - O1 Termination: Interfaces with network elements like CUs, DUs, and RUs for configuration and management.
   - A1 Termination: Establishes a link with the Near-RT RIC for policy enforcement and near-real-time decision-making.

Integration with UAVs and RIS

The Non-RT RIC plays a pivotal role in managing UAVs and RIS by using its data-driven and AI-enhanced capabilities to optimize their operations:

1. **For UAVs**:
   - **Long-Term Deployment Planning**: The Non-RT RIC, through its rApps, analyzes historical traffic data and user mobility patterns to determine the optimal placement and routes for UAVs. For instance, an rApp might recommend UAV deployments in areas with limited infrastructure during peak traffic hours or emergencies.
   - **Policy and Configuration Updates**: Using the A1 interface, the Non-RT RIC sends policies and model predictions to the Near-RT RIC to enable real-time UAV positioning adjustments based on changing conditions.
2. **For RIS**:

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

- **Dynamic Optimization**: rApps within the Non-RT RIC use AI/ML workflows to train models that determine the optimal phase shifts and configurations for RIS. These models consider long-term environmental changes, user density variations, and traffic patterns.
- **Seamless Integration with Real-Time Systems**: Policies and configurations generated by the Non-RT RIC are communicated to the Near-RT RIC, enabling it to adapt RIS configurations in real time to enhance signal propagation and reduce interference.

3. **Collaborative Operations**:
   - The Non-RT RIC facilitates collaboration between UAVs and RIS. For example:
     - UAVs equipped with RUs can be positioned dynamically based on insights provided by rApps.
     - RIS can be configured to redirect signals toward UAV-mounted RUs, enhancing coverage in underserved

3. Summary and Conclusions

This report provides a comprehensive overview of the evolution of Radio Access Network (RAN) architectures and the transformative role of Artificial Intelligence (AI) in modern and future mobile networks. Beginning with the shift from 3G to 5G, the document outlines how increasing performance demands have led to the integration of intelligent functionalities within the RAN. A significant focus is placed on AI-driven use cases such as Self-Organizing Networks (SON), traffic prediction, proactive resource management, anomaly detection, and network self-healing. These capabilities are instrumental in automating network operation, reducing latency, and enhancing overall reliability and efficiency. The report also references key field trials and testbeds that validate the feasibility of AI-native RANs, marking the transition of these concepts from theory to practice.

The second half of the Deliverable examines the system architecture required to enable AI-native integration within future RANs, with a particular emphasis on combining vRAN, RIS, and UAVs. Building upon the separation of real-time and non-real-time control functions, the analysis explores how RT controllers can manage latency-sensitive tasks such as RIS reconfiguration, UAV-assisted coverage, and mobility management, while non-RT controllers focus on long-term optimization and AI-driven decision-making. This approach, aligned with vRAN and O-RAN principles, offers a scalable framework for embedding intelligence across different layers of the network.

This deliverable provides an architectural perspective on how AI can be natively integrated with vRAN, RIS, and UAVs to support highly dynamic and adaptive B5G networks. By articulating the control and communication interfaces required for their coordination, it outlines a path toward intelligent, multi-domain orchestration capable of leveraging UAV mobility and RIS programmability alongside AI-driven optimization in vRAN.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

References

[1] Mushiba, Abel. (2024). ADVANCED TELECOMMUNICATIONS TECHNOLOGIES - A Comprehensive Guide to GSM, UMTS, LTE, and 5G.

[2] Vaezi, M., Zhang, Y. (2017). Radio Access Network Evolution. In: Cloud Mobile Networks. Wireless Networks. Springer, Cham. https://doi.org/10.1007/978-3-319-54496-0_6

[3] Lin, X., Li, J., Baldemair, R., Cheng, J. F. T., Parkvall, S., Larsson, D. C., ... & Werner, K. (2019). 5G new radio: Unveiling the essentials of the next generation wireless access technology. IEEE Communications Standards Magazine, 3(3), 30-37.

[4] Wypiór, D., Klinkowski, M., & Michalski, I. (2022). Open RAN—Radio Access Network Evolution, Benefits and Market Trends. Applied Sciences, 12(1), 408. https://doi.org/10.3390/app12010408

[5] 3GPP. “Open RAN.” [Online]. Available: https://www.3gpp.org/news-events/3gpp-news/open-ran. [Accessed: 2-Jul-2025].

[6] A. Chaoub et al., "Hybrid Self-Organizing Networks: Evolution, Standardization Trends, and a 6G Architecture Vision," in IEEE Communications Standards Magazine, vol. 7, no. 1, pp. 14-22, March 2023, doi: 10.1109/MCOMSTD.0002.2200049.

[7] 3GPP. “Self-Organizing Networks (SON).” [Online]. Available: https://www.3gpp.org/technologies/son. [Accessed: 2-Jul-2025].

[8] 3GPP. “RAN3 Led Features in Release 18.” [Online]. Available: https://www.3gpp.org/technologies/ran3-rel-18. [Accessed: 2-Jul-2025].

[9] 3GPP. “5G System Overview.” [Online]. Available: https://www.3gpp.org/technologies/5g-system-overview. [Accessed: 2-Jul-2025].

[10] 3GPP TR 36.912, "Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced)", V10.0.0, March 2011.

[11] Juniper. “What Is a RAN Intelligent Controller (RIC)?.” [Online]. Available:https://www.juniper.net/us/en/research-topics/what-is-ric.html. [Accessed: 2-Jul-2025].

[12] Asghar, M. Z., Abbas, M., Zeeshan, K., Kotilainen, P., & Hämäläinen, T. (2019). Assessment of Deep Learning Methodology for Self-Organizing 5G Networks. Applied Sciences, 9(15), 2975. https://doi.org/10.3390/app9152975

[13] Fourati, H., Maaloul, R., Chaari, L., & Jmaiel, M. (2021). Comprehensive survey on self-organizing cellular network approaches applied to 5G networks. Computer Networks, 199, 108435.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

[14] Nokia, “MantaRay SON,” in Intelligent RAN operations, Nokia. [Online]. Available: https://www.nokia.com/mobile-networks/ran-operations/network-management-son/. [Accessed: 2-Jul-2025].

[15] 5G PPP Technology Board. (2021, May 11). AI and ML – Enablers for Beyond 5G Networks (Version 1.0) (White paper). https://doi.org/10.5281/zenodo.4299895

[16] A. Calvillo-Fernandez, T. Dimitrovski, M. Groshev, A. Ganesh, C. Ayimba and A. de la Oliva, "Attention to Virtualization: Making Network Digital Twins Aware of Network Slicing," in IEEE Network, vol. 39, no. 3, pp. 134-139, May 2025, doi: 10.1109/MNET.2025.3552137

[17] N. Apostolakis, M. Gramaglia, L. E. Chatzieleftheriou, T. Subramanya, A. Banchs and H. Sanneck, "ATHENA: Machine Learning and Reasoning for Radio Resources Scheduling in vRAN Systems," in IEEE Journal on Selected Areas in Communications, vol. 42, no. 2, pp. 263-279, Feb. 2024, doi: 10.1109/JSAC.2023.3336155.

[18] Asghar, M. Z., Abbas, M., Zeeshan, K., Kotilainen, P., & Hämäläinen, T. (2019). Assessment of Deep Learning Methodology for Self-Organizing 5G Networks. Applied Sciences, 9(15), 2975. https://doi.org/10.3390/app9152975

[19] Hasan, A., Boeira, C., Papry, K., Ju, Y., Zhu, Z., & Haque, I. (2024). Root Cause Analysis of Anomalies in 5G RAN Using Graph Neural Network and Transformer. arXiv preprint arXiv:2406.15638.

[20] ETSI. (2024, October). Zero-touch network and Service Management (ZSM); Intent-driven Closed Loops (ETSI GS ZSM 016 V1.1.1). European Telecommunications Standards Institute. Retrieved from ETSI website

[21] Wen, H., Sharma, P., Yegneswaran, V., Porras, P., Gehani, A., & Lin, Z. (2024, November). 6G-XSec: Explainable Edge Security for Emerging OpenRAN Architectures. In Proceedings of the 23rd ACM Workshop on Hot Topics in Networks (pp. 77-85).

[22] Li, P., Aijaz, A., Farnham, T., Gufran, S., & Chintalapati, S. (2023, October). A Digital Twin of the 5G Radio Access Network for Anomaly Detection Functionality. In 2023 IEEE 31st International Conference on Network Protocols (ICNP) (pp. 1-2). IEEE.

[23] AI-RAN Alliance. “AI-RAN Alliance Vision and Mission White Paper.” [Online]. Available for download: https://ai-ran.org/publications/. [Accessed: 2-Jul-2025].

[24] Martínez-Durive, O. E., Suárez-Varela, J., Omaña Iglesias, J., Lutu, A., & Fiore, M. (2024, December). An Evaluation of RAN Sustainability Strategies in Production Networks. In IEEE International Conference on Computer Communications.

[25] Wu, Q., Chen, X., Zhou, Z., Chen, L., & Zhang, J. (2021). Deep reinforcement learning with spatio-temporal traffic forecasting for data-driven base station sleep control. IEEE/ACM transactions on networking, 29(2), 935-948.

[26] D. Bega, M. Gramaglia, A. Garcia-Saavedra, M. Fiore, A. Banchs and X. Costa-Perez, "Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management," in IEEE

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Communications Magazine, vol. 58, no. 6, pp. 32-38, June 2020, doi: 10.1109/MCOM.001.1900653.

[27] M. Polese et al., "Colosseum: The Open RAN Digital Twin," in IEEE Open Journal of the Communications Society, vol. 5, pp. 5452-5466, 2024, doi: 10.1109/OJCOMS.2024.3447472.

[28] Bakri, S., Dey, I., Siljak, H., Ruffini, M., & Marchetti, N. (2025). Mitigating xApp conflicts for efficient network slicing in 6G O-RAN: a graph convolutional-based attention network approach. arXiv preprint arXiv:2504.17590.

[29] J. Perez-Valero, P. Serrano, J. Garcia-Reinoso, A. Banchs and X. Costa-Perez, "Minimum-Cost Design of Auto-Scaling Server Farms Providing Reliability Guarantees," in IEEE Open Journal of the Communications Society, doi: 10.1109/OJCOMS.2025.3586088.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES

Plan de Recuperación, Transformación y Resiliencia

UNICO I+D