6GSNS

Co-funded by the European Union

SEamless integratioN of efficient 6G WirelesS tEchnologies for Communication and Sensing

D2.3 Final report on 6G-SENSES network architecture evaluation

January 2026

6G-SENSES project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement 101139282

Project Start Date: 2024-01-01 **Duration**: 30 months

Call: HORIZON-JU-SNS-2023 **Date of delivery**: 2026-01-02

Topic: HORIZON-JU-SNS-2023-STREAM-B-01-02 **Version**: 1.0

Type: Report (R)

Grant Agreement Number:	101139282
Project Name:	SEamless integratioN of efficient 6G WirelesS tEchnologies for Communication and Sensing
Project Acronym:	6G-SENSES
Document Number:	D2.3
Document Title:	Final report on 6G-SENSES network architecture evaluation
Version:	1.0
Delivery Date:	2025-10-31 (**2026-01-02**)
Responsible:	Institute of Accellerating Systems and Applications (IASA)
Editor(s):	Pavlos Doanis (IASA)
Authors:	Markos Anastasopoulos, Ilias Floudas, Anna Tzanakaki, Pavlos Doanis (IASA), Jessica Sanson (INT), Novella Bartolini, Federico Trombetti, Salvatore Pontarelli (UNIROMA1), Neco Villegas, Luis Diez, Ramón Agüero (UC), Khai Nguyen, Navid Nikaein (BR), Muhammad Saadi, Shahid Mumtaz (NTU), Ioanna Mesogiti, Elina Theodoropoulou, George Limperopoulos (OTE), Vladica Sark, Jesús Gutiérrez (IHP).
Keywords:	Radio Access Network (RAN), Wireless Access Technologies (WATs), Integrated Sensing and Communication (ISAC), Software Defined Radio (SDR), Cell-Free MIMO (CF-MIMO), Open RAN (O-RAN), Reconfigurable Intelligent Surfaces (RIS), RAN Intelligent Controller (RIC), Near Real Time RIC, Non-Real Time RIC, Control Plane, Service and Management Orchestration (SMO), Multi-access Edge Compute (MEC), Artificial Intelligence (AI), Machine Learning (ML), Spectrum Management (SM).
Status:	Final
Dissemination Level	Public (PU)
Project URL:	https://www.6g-senses.eu

Revision History

Rev. N	Description	Author	Date
0.0	Draft Table of Contents (ToC)	Anna Tzanakaki (IASA)	2025-10-20
0.1	First round of partner contributions	All partners	2025-11-13
0.2	Second round of partner contributions	All partners	2025-12-01
0.3	Harmonization of contributions and first round of editing	Pavlos Doanis (IASA)	2025-12-18
0.4	First revision of the document	Anna Tzanakaki (IASA)	2025-12-20
0.5	Second revision of the document	Ioanna Mesogiti (OTE), Valerio Frascolla (INT), Jesús Gutiérrez (IHP),	2025-12-22
1.0	Final revision and submission of the document	Jesús Gutiérrez (IHP)	2026-01-02

Table of Contents

List of Figures

List of Tables

Executive Summary

6G-SENSES aims to develop and assess an integrated communication and sensing next-generation network architecture capable of supporting advanced services, such as positioning and tracking, environmental awareness, and digital twin (DT) applications. The system design also focuses on enhancing network intelligence and improving sensing accuracy, thereby enabling data-driven and adaptive network operation.

Building on the multi-layer, disaggregated 6G-SENSES architecture aligned with 3GPP and Open Radio Access Network – Open RAN or O-RAN – principles, which was defined in deliverable **D2.2**, this deliverable further refines the architecture to enhance support for Integrated Sensing and Communication (ISAC) and intelligent control. These refinements include the acquisition of sensing data from heterogeneous sensor types, the extension of the Open RAN (O-RAN) E2 interface for sensing data, and support for intent-based control at the Service Management and Orchestration (SMO) layer.

It also provides a systematic evaluation of key user plane, control plane, and end-to-end (E2E) components across multiple Wireless Access Technologies (WATs) and control timescales.

On the user plane, the report evaluates ISAC-enabling technologies, such as multi-Access Point (AP) localization and tracking in the millimeter wave (mmWave) band, Sub-8 GHz Wi-Fi-based human presence detection (HPD), and intelligent sensing-assisted Medium Access Control (MAC) scheduling. It further assesses coordination in Sub-6 GHz Cell-Free Multiple-Input Multiple-Output (CF-MIMO), Reconfigurable Intelligent Surfaces (RISs) capabilities across multiple RIS-enabled scenarios and frequency bands, and Multi-access Edge Computing (MEC)-assisted wireless edge caching (WEC). These evaluations quantify the impact of the proposed solutions on representative performance indicators including positioning, localization, and sensing accuracy, delay, and resource utilization.

On the control plane, the evaluation serves two complementary purposes. First, it defines and applies a structured assessment procedure to verify the compliance of the Non-Real-Time (Non-RT) and Near-Real-Time (Near-RT) RAN Intelligent Controllers (RICs) with O-RAN standards, including functional validation of standardized interfaces and service models (SMs), as well as performance profiling of control-loop latency. Second, it independently evaluates advanced control mechanisms developed within 6G-SENSES, including feedback-based control at the SMO level for enforcing sensing delay and freshness requirements and near-real-time eXtended application (xApp)-based solutions that combine non-3GPP sensing context with 3GPP-native Quality of Service (QoS) information. In particular, Wi-Fi-assisted beamforming and Channel Quality Indicator (CQI) + 5G QoS Identifier (5QI)-aware scheduling are evaluated in a unified control loop, where compact sensing context is conveyed via a Wi-Fi Sensing Indicator (WSI) through the Physical Uplink Control Channel (PUCCH), and are shown to jointly improve channel quality, QoS satisfaction, and fairness.

At the E2E level, the deliverable evaluates how Artificial Intelligence (AI)/Machine Learning (ML)-driven automation enhances the orchestration and operation of ISAC services across the network. A Deep Reinforcement Learning (DRL)-based slice orchestration framework is assessed in terms of its ability to balance energy efficiency and service-level performance for concurrent communication and sensing slices, while a techno-economic analysis provides insights into the broader benefits and trade-offs in ISAC-enabled 6G network deployments.

Overall, this deliverable demonstrates the feasibility and effectiveness of the 6G-SENSES architecture through an extensive, multi-layer evaluation. The results confirm that integrating sensing into both the user plane and control plane, combined with AI/ML-driven automation and O-RAN-compliant control, can significantly enhance network adaptability, efficiency, and service support.

1 Introduction

The emergence of ISAC as a key enabler of future 6G systems introduces new challenges and opportunities across the radio access, control, and service orchestration layers. Supporting advanced services such as positioning, tracking, environmental awareness, and digital twinning requires architectures that natively integrate sensing capabilities with communication, while remaining scalable, programmable, and compliant with the 3GPP and O-RAN frameworks.

Within this context, the 6G-SENSES project has developed a disaggregated, multi-layer architecture that brings together multi-RAT sensing, intelligent control, and AI/ML-driven automation. While the baseline architecture and initial validation were presented in previous deliverables, the focus of this document is on refining that architecture and providing a comprehensive evaluation of its key components and mechanisms.

This deliverable reports the work carried out in the context of Work Package 2 (**WP2**) and concludes the WP, focusing on the refinement and comprehensive evaluation of the 6G-SENSES architecture. It builds directly on the baseline multi-layer, disaggregated architecture defined in deliverable **D2.2** "*System architecture and preliminary evaluations*" [1], which is aligned with 3GPP and O-RAN principles, as well as on the service requirements, scenarios, and KPIs identified in deliverable **D2.1** "*Report on 6G-SENSES use cases, network architecture, KPIs and supported RAN functions*" [2].

In particular, this deliverable extends the architectural work of **D2.2** by introducing targeted refinements that enhance support for ISAC, including mechanisms for ingesting sensing data from heterogeneous sources, sensing-oriented extensions of the O-RAN *E2* interface, and intent-based control capabilities at the SMO layer.

Even more importantly, the deliverable presents a systematic and multi-layer evaluation of the 6G-SENSES architecture, to complement the preliminary evaluation results reported in **D2.2**. The evaluation spans user-plane, control-plane, and E2E perspectives, and leverages a combination of experimental platforms, simulations, and analytical studies. It assesses the performance, feasibility, and impact of key ISAC-enabling technologies across multiple RATs and control timescales, including Near-RT and Non-RT operation. In this way, the deliverable provides an integrated view of how the refined 6G-SENSES architecture supports sensing-aware networking and AI/ML-driven automation.

Overall, the evaluation results reported in this deliverable validate the architectural design choices and provide evidence on the effectiveness of the proposed technologies, user-plane and control-plane mechanisms, as well as on the E2E benefits of AI/ML-driven orchestration for ISAC services, while also introducing a techno-economic evaluation framework that supports the assessment of deployment cost, scalability, and sustainability aspects of large-scale 6G deployments.

1.1 Organisation of the document

This document comprises six chapters. Following the Executive Summary and Introduction sections:

Chapter 2 describes the architectural refinements introduced in this deliverable. It recaps the main architectural components of the 6G-SENSES architecture defined in **D2.2** and presents extensions for integrating sensing data from heterogeneous sources (sensors) and exposing sensing information to the RIC through sensing-oriented E2 SMs. It is also described how the multi-layered 6G-SENSES architecture is general enough to support intent-based control.

Chapter 3 presents the evaluation of key user plane components of the 6G-SENSES architecture. It covers ISAC-related mechanisms, including multi-AP localization and tracking in mmWave, Wi-Fi-based sensing in

Sub-8 GHz, and sensing-assisted DRL-based MAC scheduling, as well as distributed coordination in Sub-6 GHz CF-MIMO, RIS across multiple scenarios and frequency bands, and MEC-based WEC.

Chapter 4 evaluates the control plane components of the 6G-SENSES. It first defines and applies a structured assessment procedure to verify the compliance of the Non-RT and Near-RT RICs with O-RAN standards, including functional validation of standardized interfaces and performance profiling of control-loop latency. The chapter then evaluates advanced control mechanisms developed within 6G-SENSES, including feedback-based sensing-slice control at the SMO layer and near-real-time xApp-based solutions that inject non-3GPP sensing context into the RAN control loop to enable Wi-Fi-assisted beamforming and CQI + 5QI-aware scheduling.

Chapter 5 presents the E2E evaluation of ISAC services supported by the 6G-SENSES architecture. It first evaluates a DRL-based E2E slice orchestration framework, demonstrating how AI/ML-driven automation can optimize the orchestration of communication and sensing slices sharing a common physical network infrastructure, with the objective of minimizing energy consumption while satisfying slice-level Service Level Agreements (SLAs). It then provides a complementary system-wide techno-economic evaluation, assessing large-scale deployment scenarios capable of meeting 6G service requirements and supporting economically viable and sustainable network deployments.

Finally, Chapter 6 summarizes the document.

2 6G-SENSES refined architecture

6G networks will support a broad range of vertical applications with highly diverse requirements. Beyond offering high-performance connectivity, 6G is expected to enable advanced capabilities such as sensing, monitoring, positioning, and real-time DT reconstruction of the environment. To meet these demands, 6G will incorporate ISAC features, allowing the mobile network itself to act as a sensing system.

Two main sensing methods are considered:

Channel State Information (CSI)–based sensing, which uses the communication link between the base station (BS) and the user device to estimate channel properties and derive parameters such as Angle of Arrival (AoA) and Time Difference of Arrival (TDOA). This enables applications including localization, tracking, presence detection, activity recognition, and healthcare support.

Passive radar sensing, where the network analyses reflections of its own communication signals to understand objects and the environment. This information can enhance network operation, improve services like eXtended Reality (XR) and DTs, and introduce new use cases such as object detection, tracking, imaging, and scene reconstruction.

3GPP is already examining ISAC use cases and requirements, highlighting its potential to jointly optimize communication and sensing. In this context, different integration levels are possible, including the following:

- Separate infrastructures for sensing and communication, exchanging information between them.
- Shared hardware, with common spectrum but time-separated sensing and communication transmissions.
- Fully integrated systems, sharing spectrum and time.

Sensing can also involve various device configurations:

- Monostatic (same transmitter and receiver),
- Bistatic/multistatic (one transmitter, multiple distributed receivers),
- Passive sensing (communication signals reused for sensing).

Although early prototypes exist—mainly in Wi-Fi—3GPP-compliant passive-radar ISAC implementations are still immature. These systems face challenges such as high signal-processing complexity and the need to aggregate large volumes of synchronized I/Q echo data, which must be processed at edge servers over high-capacity transport networks.

In view of these, 6G-SENSES introduces a 6G architecture that integrates multi-technology RAN systems (both 3GPP and non-3GPP) to jointly deliver communication and sensing services, as shown in Figure 2-1 and described in detail in **D2.2** [1]. The architecture combines Sub-6, Wi-Fi, mmWave (non-3GPP), and 5G NR (3GPP) technologies within an ISAC framework to build an accurate, real-time view of the surrounding environment.

Non-3GPP sensing relies on enhanced Wi-Fi networks operating as monostatic and multistatic radars. Their sensing data is securely exposed to the RAN via extensions to the E2 interface of the RIC. To support this, 6G-SENSES extends the Non-3GPP InterWorking Function (N3IWF) function so Wi-Fi systems can authenticate and securely provide sensing information to the RIC.

Through this approach, CSI-based information can be ingested into the Near-RT RIC with approx. 1ms delay.

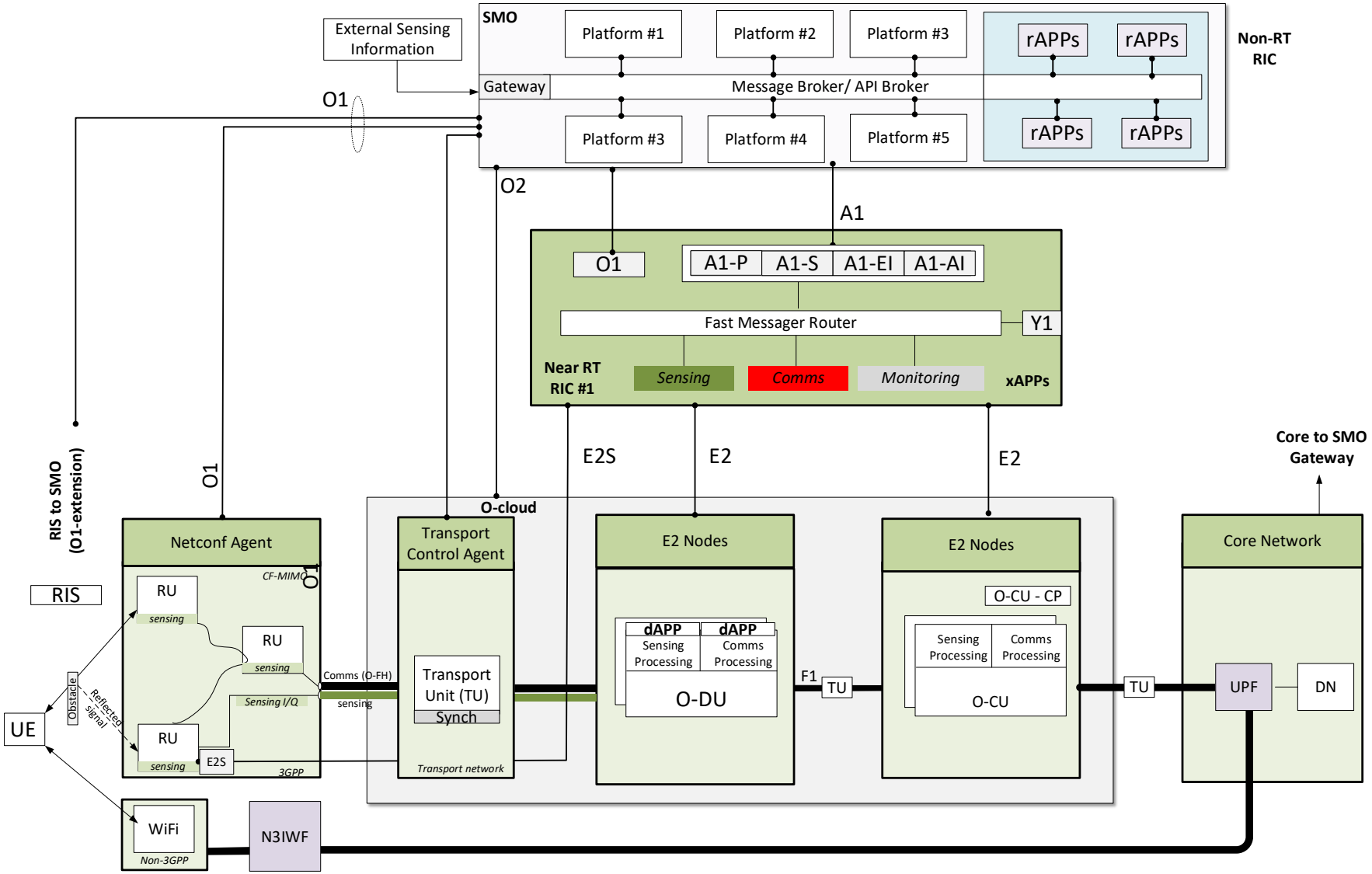


Figure 2-1 Generic 6G-SENSES architecture

In addition, information from 3rd party sensors can be ingested in a secure way directly at the SMO level and consumed by any entity in the 6G-SENSES architecture. Specifically, external sensing data coming from analytics platforms, monitoring systems, or domain-specific sensing functions are integrated into the SMO / Non-RT RIC environment through the Non-RT RIC gateway. The produced data through a message-brokering mechanism can be then consumed by other entities of the 6G-SENSES platform, including rApps and xApps. Within the Non-RT RIC, this sensing data processed by rApps and platform analytics functions operate at time scales of seconds to minutes. The Non-RT RIC correlates sensing inputs with other non-real-time information available through *O1*, such as configuration, performance, and historical measurements, in order to derive higher-level sensing outcomes. These outcomes are delivered as structured information such as sensing indicators, environment states, confidence levels, or predictions that are relevant for downstream RAN optimization or control

Once the sensing information has been abstracted, the Non-RT RIC exposes it to the Near-RT RIC using the *A1* interface. Sensing outputs are conveyed through *A1* where the Non-RT RIC acts as an information producer and defines the corresponding information types and schemas. This allows sensing data to be delivered as enrichment or service information that can be consumed by near-RT functions without embedding control logic in the non-RT domain.

At the Near-RT RIC, the *A1* termination and mediation functions receive the sensing-related *A1* messages and validate them against the registered schemas. The Near-RT RIC then adapts these messages into its internal fast-messaging format and injects them into the fast message router providing connectivity between xApp and terminating protocols (*E2*, *A1*). This step ensures that sensing information coming from the non-RT domain can be distributed with low latency and high reliability to the appropriate near-RT applications.

xApps running in the Near-RT RIC subscribe to the relevant message types and receive the sensing-derived information in near-real time. The xApps treat this information as external context that complements real-time measurements obtained via *E2* from the RAN. By combining non-RT sensing context with near-RT telemetry, xApps can make informed control decisions and issue E2 control actions toward the RAN when required. In this way, sensing data generated and processed entirely in the non-RT domain is systematically ingested, abstracted, transported over *A1*, routed through the fast message router, and consumed by xApps in the near-RT control loop, exactly as illustrated in the architecture.

3GPP-based sensing uses distributed passive radar principles. 6G BSs generate communication signals whose reflections on nearby objects form I/Q echo streams. These streams are forwarded to O-RAN Distributed Units (DUs) for compression and then delivered to the RIC through the E2 interface. Specialized sensing xApps merge Wi-Fi and 3GPP sensing data, assess data quality, and store it in fast memory. The processed sensing information can then optimize RAN operations—such as beamforming, beam steering, and power control—or be provided to external applications via the *Y1* interface.

The sensing outputs are directed to the SMO framework, which configures network resources to support combined sensing–communication services. The SMO automates the lifecycle management of ISAC services and dynamically adjusts E2E slices to meet both communication (e.g., fronthaul, backhaul) and sensing requirements. Relevant demonstrators have been presented in [3], [4].

Building on the baseline 6G-SENSES architecture introduced in deliverable **D2.2** [1], this chapter focuses on the architectural refinements introduced during the current reporting period. Section 2.1 first provides an overview of the 6G-SENSES platform, recalling the main architectural components, capabilities, and technologies already defined, in order to ensure continuity with earlier work. Section 2.2 then introduces architectural extensions that enable the acquisition of sensing data from heterogeneous sensor types and provides a more detailed description of how such data is integrated within the refined 6G-SENSES architecture. Section 2.3 presents extensions to the O-RAN E2 interface through sensing-oriented service models, allowing environmental sensing information to be exposed to the RIC in a technology-agnostic

manner. Finally, Section 2.4 describes intent control and management as a further architectural enhancement at the SMO level, showcasing how high-level service objectives can drive the operation of the 6G-SENSES architecture. Together, these refinements extend the architectural capabilities of 6G-SENSES providing the foundation for the sensing-aware control mechanisms evaluated in subsequent sections.

2.1 Platform overview, capabilities and technologies used

6G-SENSES introduces a multi-layer, multi-technology architecture aligned with O-RAN and 3GPP standards to deliver ISAC capabilities and to expose sensing information both within the system and to third-party applications. The overall architecture, highlighting the main components and interfaces that enable connectivity across the RAN and Core Network (CN) is illustrated in Figure 2-2.

The design follows the O-RAN model, which disaggregates traditional RAN functions and incorporates software-defined intelligence and advanced control mechanisms. A disaggregated gNB comprises the Radio Unit (O-RU), O-DU, and Central Unit (O-CU), interconnected through standardized fronthaul (FH) and F1 interfaces.

At the O-RU level, 6G-SENSES enhances O-RAN specifications to support shared O-RUs capable of serving both communication and sensing functions. As illustrated in Figure 2-3, a shared O-RU can allocate its resources—such as spectrum—across O-DUs that process sensing and communication signals. This shared O-RU model is particularly suitable for deployments where a new O-RU is introduced to provide passive sensing (e.g., monitoring spectrum for sensing-related tasks).

RAN intelligence in 6G-SENSES is enabled through the O-RAN RIC framework, which comprises a Near-RT RIC, responsible for dynamic RAN control and optimization, and a Non-RT RIC, responsible for long-term policies, analytics, resource optimization and RAN configuration (as part of the SMO framework).

Together, these RICs enable AI/ML integration in the RAN: models are trained in the Non-RT RIC and executed in the Near-RT RIC to dynamically fine-tune network operations. Data storage is supported by an in-memory Database as a Service (DBaaS) platform, which ensures ultra-fast access for sensing data, AI/ML features, and geospatial indices used in edge applications.

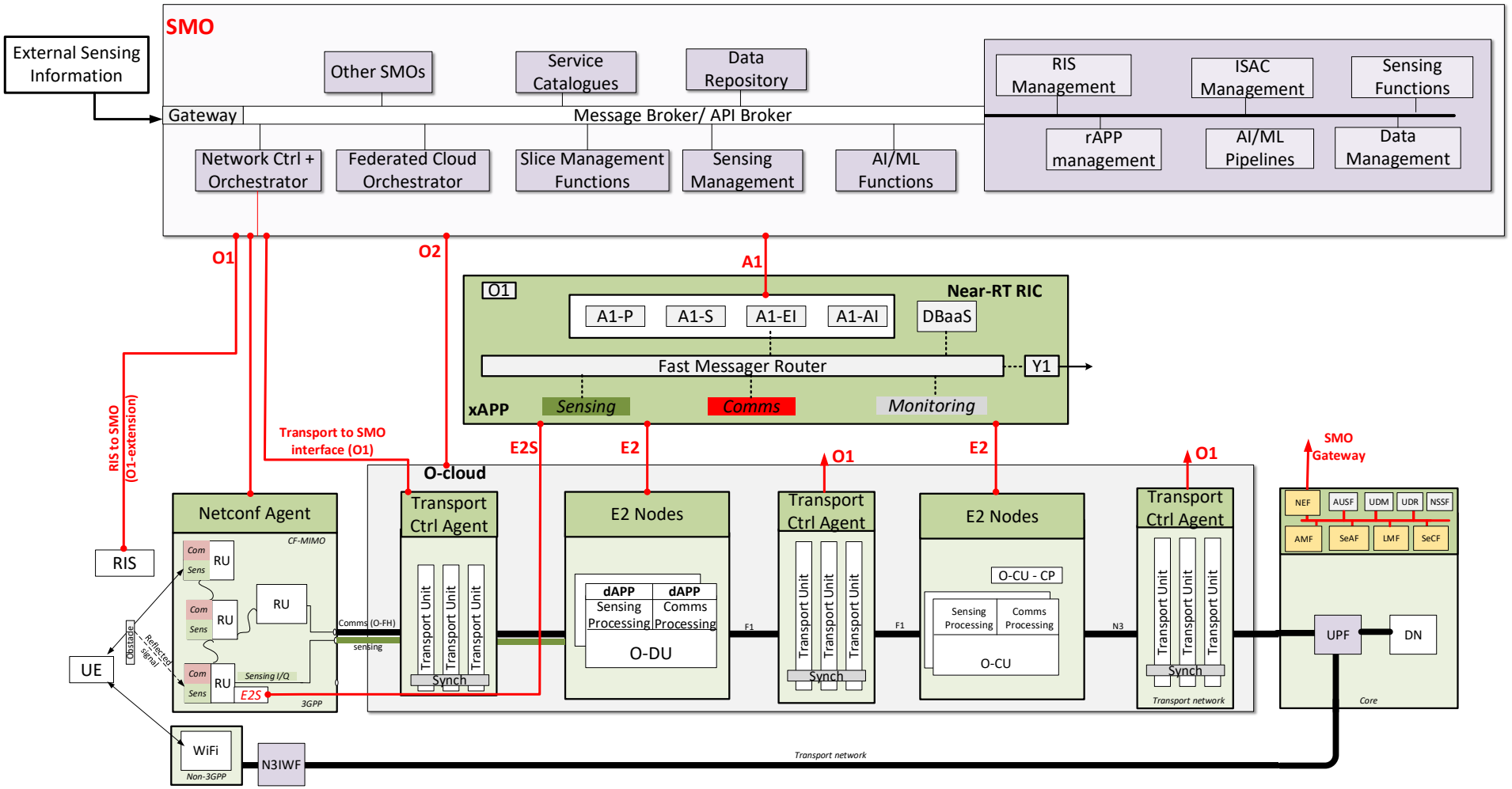


Figure 2-2 Multi-layer 6G-SENSES Architecture

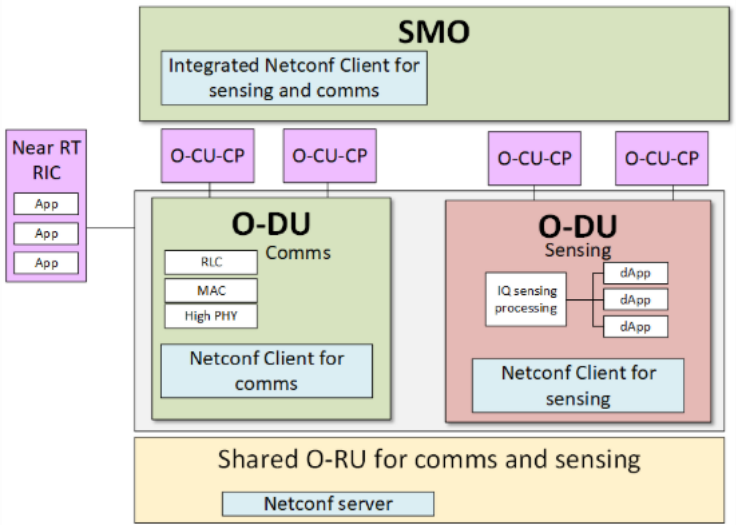


Figure 2-3 O-RU sharing for comms and sensing

Beyond the RAN, the 5G CN provides the service-based control and user-plane functions required for mobility management, session control, policy enforcement, and secure exposure of network capabilities.

To support sensing within the 6G-SENSES framework, new core functions extend 3GPP TR 33.713 [5]:

- Sensing Control Function (SeCF), which provides connectivity for external sensors complementing ISAC, device context and inventory management, authentication/authorisation of sensing devices, and exposure of sensing data to applications via the NEF.
- Sensing Analytics Function (SAF), offering high-level sensing analytics exposed through the NEF.

The user-plane traffic is handled by the User Plane Function (UPF), which ensures low-latency data delivery to external networks and services over *N6*.

To achieve seamless multi-access connectivity, the architecture incorporates non-3GPP technologies (e.g., Wi-Fi, mmWave) through the N3IWF, which provides authentication, encryption, and tunnelling between non-3GPP devices and the 5GC. The *Nwu* interface links devices to the network, while *N2* and *N3* handle control-plane and user-plane communications, respectively.

6G-SENSES further extends the N3IWF with a new E2 interface to securely forward sensing data from non-3GPP Access Points to O-RAN components. This extension uses IPsec-protected tunnels and certificate-based authentication already supported by the N3IWF (as defined in ETSI TS 133 210 [6] and ETSI TS 133 310 [7]. After RAN processing, user traffic reaches the UPF and is routed via *N6* to external applications, enabling low-latency and location-aware services.

2.2 Architecture extension including acquisition of sensing data from various sensor types

During this reporting period, the overall 6G-SENSES architecture has been extended to include the capability of acquiring sensing data from various sensor types and to deliver fused results securely to third parties. In this context, the architecture depicted in Figure 2-4 follows the generic functional decomposition adopted in the first release of the 6G-SENSES architecture inspired by the 3GPP 5G System specifications (primarily TS 23.501 [8], TS 23.502 [9], and TS 38.300 [10] series) and integrates the O-RAN Alliance control, management, and intelligence framework as specified in the O-RAN Architecture Description (O-RAN.WG1.OAD [11]), O-RAN Near-RT RIC Architecture (O-RAN.WG3.E2AP/Service Model documents [12]), and the SMO framework (O-RAN.WG7).

To include the capability of acquiring sensing data from various sensor types, in addition to UE, we also consider a variety of heterogeneous sensing devices that generate application data that enters the system either through 3GPP access or through non-3GPP access technologies. Sensors may connect indirectly via a UE that aggregates or relays data, or via Wi-Fi access points forming part of a non-3GPP access network. The non-3GPP access is integrated with the 5G System via the N3IWF, as defined in 3GPP TS 23.501 §4.2.8 [8]. The N3IWF provides secure connectivity over untrusted access networks, enabling UEs and IoT devices on Wi-Fi to establish an IPsec-protected tunnel toward the 5G CN. Authentication, authorization, and session establishment will be performed following the procedures of TS 23.502 [9], ensuring that non-3GPP devices can seamlessly participate in 5G data sessions and mobility contexts.

On the 3GPP access side, UEs connect to a disaggregated 5G New Radio (NR) BS following the O-RAN split model. The O-RU implements the lower PHY and RF functions as defined in the O-RAN fronthaul specifications (O-RAN.WG4.FrontHaul). The O-DU executes higher PHY, MAC, and portions of Radio Link Control (RLC), consistent with the gNB functional split Option 7-2x defined in TS 38.801 and TS 38.300 [10] The O-CU hosts the RLC, Packet Data Convergence Protocol (PDCP), and Radio Resource Control (RRC) layers (TS 38.331 [13], TS 38.323 [14], TS 38.322 [15]), and splits into the CU-CP and CU-UP functions aligned with the control-plane/user-plane separation introduced in TS 38.470 [16] and TS 38.473. This disaggregation supports the O-RAN objectives of openness, vendor interoperability, and fine-grained deployment flexibility.

The O-DU and O-CU are supervised and optimized by the O-RAN Near-RT RIC operating in the 10 ms to 1 s timescale. The Near-RT RIC communicates with the O-CU and O-DU through the standardized E2 interface (O-RAN.WG3.E2AP specification). Over this interface, the RIC deploys and executes xApps, i.e. modular control applications responsible for functions such as radio resource management, mobility optimization, interference mitigation, and scheduling refinement. The behaviour and capabilities of these applications are defined by O-RAN service models (E2SM-KPM for performance metrics, E2SM-RC for RAN control, etc.). The E2 interface ensures that the RIC receives near real-time measurements while being able to issue policy-based or control-loop actions back to the RAN nodes. To ensure confidentiality and integrity of E2 traffic, IPsec-based transport security is applied as mandated by O-RAN security specifications (O-RAN.WG11).

Above the Near-RT RIC, the refined architecture incorporates also the O-RAN SMO framework, which provides non-real-time network management, analytics, configuration, and orchestration functions. The SMO capabilities follow the O-RAN Operations and Maintenance specifications and are aligned with 3GPP management architecture (TS 28.530 [17], 28.531 [18], 28.532 [19]). The SMO communicates with the Near-RT RIC and the Management Plane (Mn-Plane) of O-RU/O-DU/O-CU using the O1 interface defined in O-RAN.WG1 and WG2. All management traffic between SMO and controlled nodes is secured using mutual Transport Layer Security (mTLS), as required by O-RAN security guidelines. This guarantees authenticated and encrypted management plane communication across multi-vendor RAN components.

The SMO incorporates multiple services essential for data ingestion, storage, model hosting, and orchestration. This includes a Message Broker/Application Programming Interface (API) broker implemented based on Apache Kafka that provides event-driven messaging, telemetry streaming, and asynchronous data exchange between SMO subsystems, analytics engines, and RIC applications. This aligns with the O-RAN concept of data accessibility and the need for consistent mediation of RAN metrics. The DBaaS component acts as the persistent storage layer for operational data, configuration states, performance indicators, model artifacts, and metadata required for management and analytics. It ensures that RIC xApps, SMO processes, and external network functions can retrieve historical or contextual information required for decision-making, in accordance with O-RAN data collection and exposure guidelines.

Intelligence functions inside the SMO may include analytics engines, model execution environments, or non-RT RIC functionality as specified in O-RAN.WG2. Non-RT RIC functions (operating at >1 s timescales) influence the Near-RT RIC via policy-based guidance and enrichment information. These policies are delivered through the *A1* interface (O-RAN.WG2.A1AP [20]), enabling coordination between long-term optimization processes (e.g., ML model management, slice-level performance objectives) and real-time control loops executed by xApps in the near-RT domain.

To 5G CN as in the first release of the 6G-SENSES architecture comprises a set of Service-Based Architecture (SBA) functions in accordance with TS 23.501 [8] and TS 23.502 [9]. The Network Exposure Function (NEF, TS 23.502 §5.12) provides secure northbound interfaces to expose core-level capabilities or network events to authorized Application Functions (AFs) or higher-level orchestrators. The Network Data Analytics Function (NWDAF, TS 23.288 [21]) provides network analytics at the CN level, generating insights about mobility, QoS, slice performance, and traffic patterns. The integration of NWDAF outputs with SMO analytics enables cross-domain closed loops involving both RAN and Core insights. The Unified Data Repository (UDR) and other data-management Network Functions provide structured storage for subscriber data, policy information, and application state, thus supporting consistent data persistence across SBA functions.

User-plane traffic as before is handled by the UPF according to TS 23.501 §5.6. The UPF performs packet routing, forwarding, QoS enforcement, and traffic breakout toward external Data Networks or Application Servers and forms the boundary at which data generated by sensors and UEs is delivered to application ecosystems. Depending on the deployment details, the UPF may be located centrally, regionally, or at the edge to support low-latency requirements.

Overall, the presented architecture is fully aligned with the 3GPP 5G system model and the O-RAN disaggregated and intelligent RAN framework. It supports multi-access integration, near-real-time RAN optimization, cross-layer orchestration, and modular intelligence through standardized interfaces including N3IWF (non-3GPP interworking), *F1*/*E1*/*O1* (intra-RAN management and control), *E2* (RIC–RAN real-time control), and *A1* (non-RT to near-RT guidance). By adhering strictly to the referenced specifications, the architecture maintains full interoperability and enables a multi-vendor, standards-compliant deployment model suitable for research, validation, and future evolution.

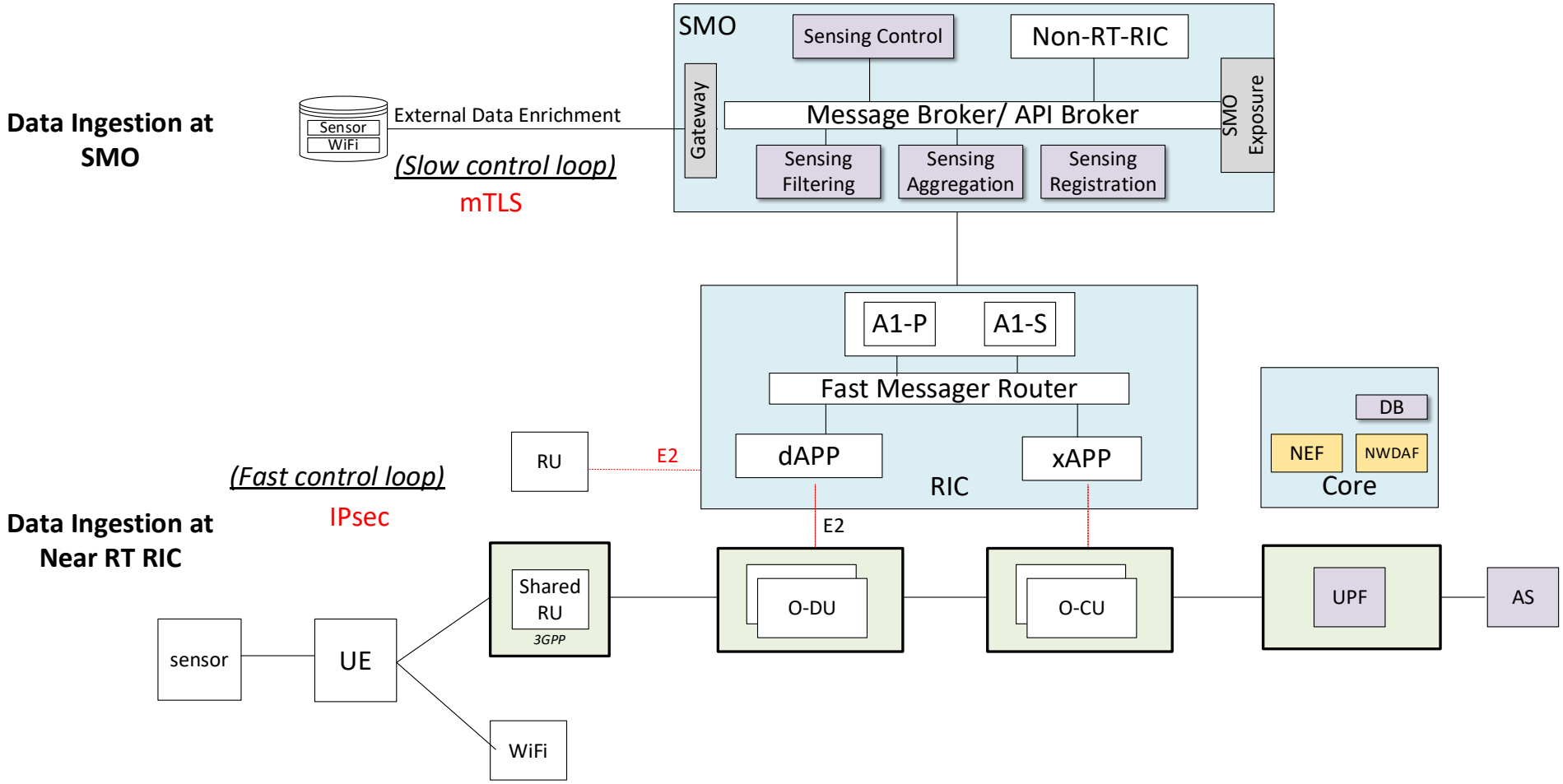


Figure 2-4 6G-SENSES Architecture extension including acquisition of sensing data from various sensor types

2.3 Extended E2 Interface for Sensing-Oriented Service Models

To enable the RAN Intelligent Controller (RIC) to access environmental sensing information independently of the underlying Radio Access Technology (RAT), we propose extending the O-RAN E2 interface through the development of a sensing-oriented service model. The proposed framework supports both 3GPP and non-3GPP nodes, which function as sensing RUs (sRUs) interconnected via the E2 interface, as can be seen in Figure 2-5.

For this purpose, a sensing E2 Service Model (E2SM) has been designed to provide the capability of transmitting physical environment captures to the RIC. This is achieved by leveraging the FlexRIC platform from OpenAirInterface (OAI) software, enabling the definition and implementation of a Service Model (SM) that supports the exchange of E2 Application Protocol (E2AP) messages. These messages facilitate both the collection of sensing data and the control of the sRU, thus integrating sensing capabilities into the O-RAN control architecture.

Sensing information can be collected in different formats, ranging from the rawest to the most processed. The rawest format consists of I/Q symbols, while an intermediate format corresponds to the Channel State Information (CSI), which represents a trade-off between I/Q symbols and heatmaps. Finally, the heatmap format imposes the lowest load on the network but limits the possibilities for exploiting the data.

Specifically, the heatmap is transmitted following a structured format consisting of three components: *radio (initial point, number of samples, and final point), angle (initial point, number of samples and final point),* and *strength (binary buffer),* all included in a `RIC Indication` message. Finally, the service model is extended to include control commands that enable interaction with the sRU, allowing the modification of parameters such as the reporting *periodicity*, *heatmap range (radio, angle)*, and *data resolution.* In addition, the sensing information contains a *timestamp* that specifies the temporal origin of the collected data. These parameters can be included in same `RIC Control` message following a Type-Length-Value (TLV) format.

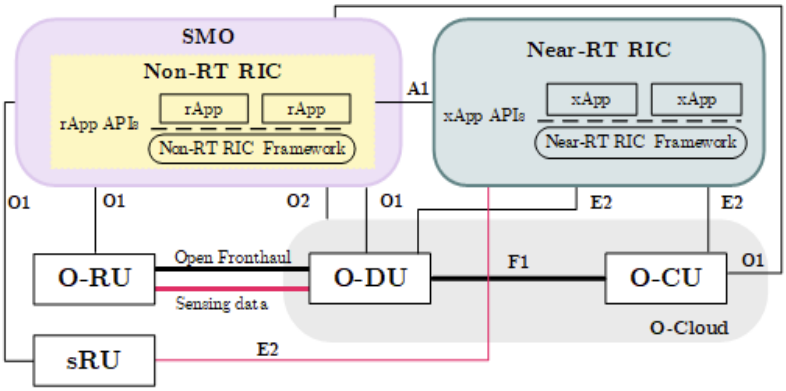


Figure 2-5 Extension of the O-RAN Architecture to Integrate Sensing Capabilities

The communication diagram follows the structure depicted in Figure 2-6. The E2 node, acting as an sRU, runs an E2 agent that operates as a proxy, capable of parsing the information into the corresponding structures defined in the SM. In addition, it can establish communication with the processing unit, enabling both the updating of statistics and the control of the sRU. On the other hand, the xApp can monitor the environment and obtain QoS statistics derived from the sensing data, as well as send RIC Control messages to manage parameters such as periodicity, range, and resolution.

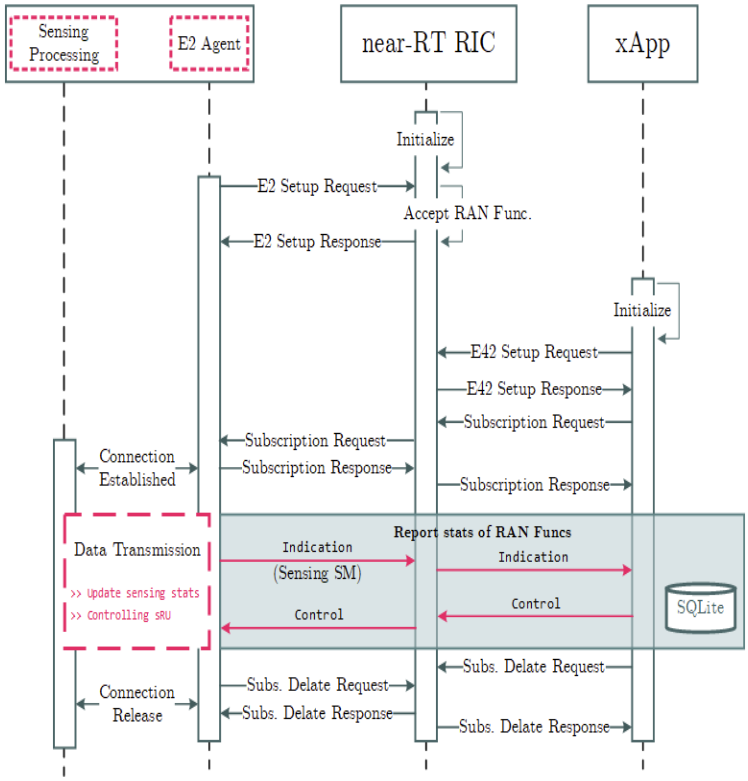


Figure 2-6 Sequence diagram of operations in the Near-RT RIC

As a result of the integration of an sRU, in particular a millimeter wave (mmWave) radar operating in the 60 GHz band with a maximum sensing rate of 2.1 kHz, Figure 2-7 illustrates not only the process of data gathering but also the capability to control what is transmitted over the network in terms of sensing information. It is worth highlighting that the larger the amount of data sent in a transmission opportunity, the higher the delay caused by the micro-burst effect induced by fragmentation. The tests were conducted over an interface configured with a Maximum Transmission Unit (MTU) of 9000 B (jumbo frames) and a capacity of 100 Mbps.

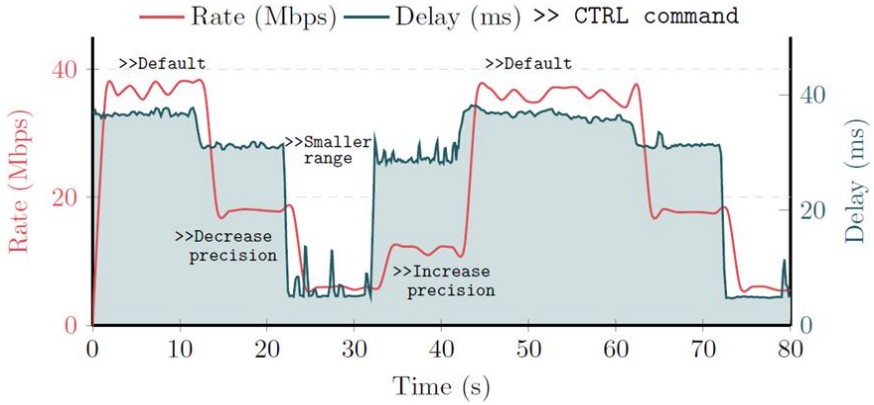


Figure 2-7 Evolution of sensing throughput and delay under configuration changes triggered by RIC Control messages

2.4 Intent Control and Management

An **intent** is a high-level, human-understandable statement that must be interpreted by a machine without ambiguity. It focuses solely on describing desired outcomes (the *what*), in contrast to rules—which specify step-by-step logic (the *how*)—and **policies**, which often blend both the *what* and the *how*.

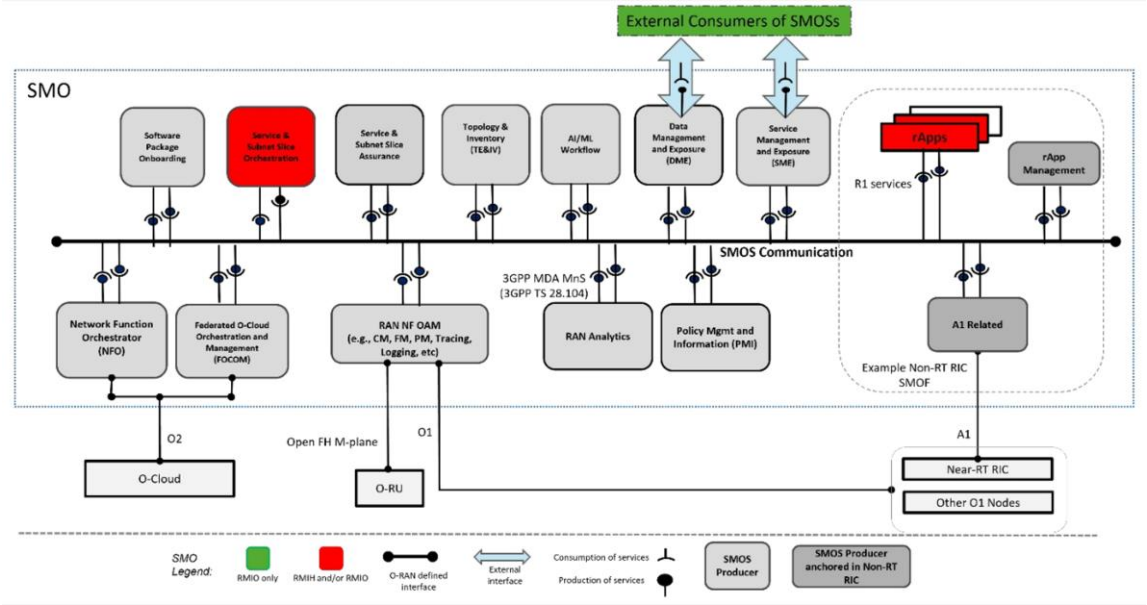


Figure 2-8 Example of SMO SBA including intent-based services

This terminology can be controlled and managed by two SMO services that are defined in the O-RAN standard. **RAN Management Intent Owner (RMIO)**—an SMO service consumer, such as an E2E management system or Business Support System, that provides detailed intent information. Correspondingly, the **RAN Management Intent Handler (RMIH)** is responsible for receiving, interpreting, and processing these requests to ensure the successful fulfilment of RAN management intents.

As illustrated in Figure 2-8, in [22] entities such as **Slicing Operator service** or **rApps** can assume the role of RMIH or RMIO. While external services typically act as the Intent Owner (RMIO), defining the requirement for intents, whereas various SMOs (SMO Service) producers can be realized as RMIHs to process and execute these specific intents

From a design perspective, the **rApp** is the ideal candidate for assuming the role of RMIH (and occasionally RMIO) due to the modular flexibility provided by the O-RAN architecture. rApps reside within the **Non-RT RIC**, which allows them to perform long-term optimization loops (greater than 1 second).

The synergy between **rApp management services** and **A1 interface services** is critical here.

- **rApp Management:** Controls the lifecycle (onboarding, instantiation, termination) of the rApp, ensuring the RMIH logic is healthy and active.
- **A1 Services:** Allow the rApp (acting as RMIH) to translate high-level intents into concrete **A1 Policies**. These policies are then pushed down to the Near-RT RIC to control the RAN nodes.

Therefore, by managing the rApp effectively, the SMO indirectly manages the translation of Intent into Action. These services constitute the core mechanism for the 'Intent Control and Management' of the RAN that this section seeks to define.

Overall, Figure 2-9 below validates rApp that serves as the pivot point (RMIH), translating high-level business intent into actionable network policies.

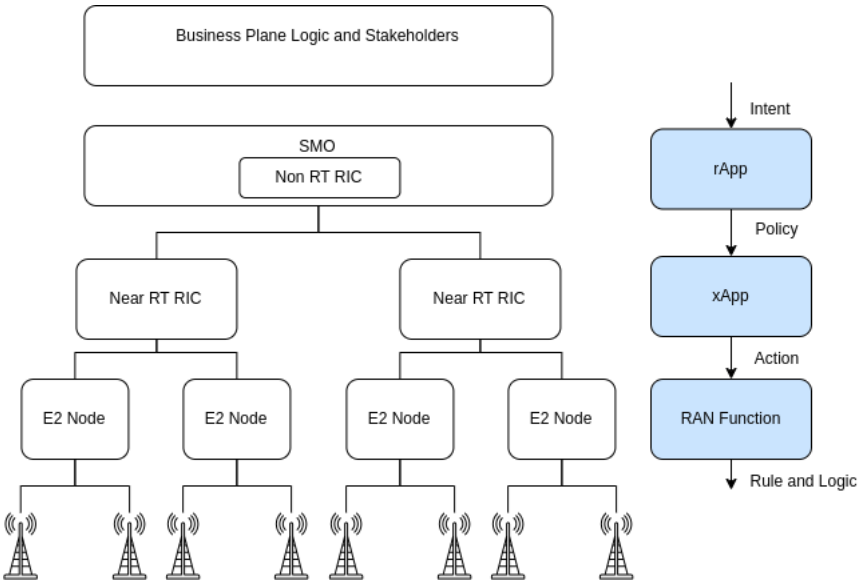


Figure 2-9 Structural Hierarchy (left) and intent-to-action flow (right) for intent-based control

Structural Hierarchy: The left section of Figure 2-9 illustrates the **SMO framework** as the central orchestrator, housing the **Non-RT RIC**.

- **Business Plane Logic:** Represents the **SMO Consumer (RMIO)** (e.g., BSS or E2E management) that initiates the intent.
- **SMO / Non-RT RIC:** Represents the hosting platform for the **RMIH** (rApp), acting as the bridge between business logic and the underlying RAN infrastructure (Near-RT RICs and E2 nodes).

Intent-to-Action Flow: The right section of Figure 2-9 creates a direct visual correlation in promoting rApp as an RMIH.

- **Intent:** The arrow descends from the Business Plane, symbolizing the **RMIO** passing a requirement to the system.
- **rApp (as RMIH):** The rApp receives the "Intent," validating the text's claim that rApps focus on "producing handling services." It processes the intent within the Non-RT RIC.
- **Policy:** The rApp translates the high-level intent into concrete policies (via the *A1* interface) and pushes them to the **xApp** (in the Near-RT RIC).
- **Action:** The xApp executes the necessary changes on the **RAN Function**, converting the abstract intent into tangible network "Rule and Logic."

3 User Plane Components Evaluation

This chapter presents the evaluation of user-plane components within the 6G-SENSES framework, covering a broad set of technologies and use cases. In particular, the section evaluates user-plane mechanisms supporting ISAC activities (Section 3.1), CF-MIMO (Section 3.2), RIS-enabled programmable propagation environments (Section 3.3), and MEC-based wireless edge caching solutions (Section 3.4). Depending on the maturity of each component, the evaluation reports quantitative performance results, summarizes findings, or describes the proposed solutions, adopted evaluation methodologies, and current progress.

The ISAC-related Section evaluates i) different techniques to combine information from multiple Access Points (APs) in order to improve user localization and tracking in the mmWave band (Section 3.1.1); ii) the feasibility and performance of monostatic sensing in the sub-8 GHz band using commercial off-the-shelf Wi-Fi network interface cards (NICs) (Section 3.1.2); iii) sensing-assisted MAC scheduling using a DRL agent (Section 3.1.3); and iv) a solution for ISAC transport network optimization (Section 3.1.4).

The CF-MIMO Section addresses user-plane coordination in Sub-6 GHz CF-MIMO networks through a fully distributed, learning-based architecture. The proposed approach mitigates the scalability and fronthaul limitations of centralized schemes by enabling local AP–user assignment decisions with limited message exchange. By modeling the network as a bipartite AP–user graph and leveraging graph neural networks (GNNs) for distributed inference, the solution supports scalable and near-real-time user-plane operation while respecting AP capacity and user reliability constraints. The section presents the system model, architectural design, and evaluation methodology for the proposed approach.

The RIS Section summarizes results across multiple RIS-enabled scenarios and frequency bands, highlighting the progressive evolution of RIS capabilities. In this progression, RIS gradually acquires new roles: first as an efficient modulation aid, then as a stabilizing element for Unmanned Aerial Vehicle (UAV)-based networks, subsequently as a dual-function component supporting ISAC and, finally, as a fully active and learning-enabled intelligent surface for Terahertz (THz) environments.

Finally, the last section describes the evaluation methodology for WEC in a MEC context, outlining the adopted simulation framework, current validation progress reported in other deliverables, and the planned next steps toward a complete performance assessment.

Together, these contributions provide a comprehensive view of how advanced user-plane technologies can enhance the performance, adaptability, and scalability of the 6G-SENSES architecture.

3.1 ISAC evaluation

In the context of the evaluation of the ISAC activities (**D2.2** [1], **D3.1** [23], **D5.1** [24]), we consider a multi-WAT system comprising:

- radar-based sensing-capable technologies, i.e. Sub-6 GHz, mmWave and SDR-based gNB, and/or
- communication and sensing-capable technologies, i.e. O-RUs, Wi-Fi and experimental O-RUs,

ready to provide sensing data to the near RT-RIC via different paths, either via the E2 interface using a defined sensing SM, or via the O-FH from 3GPP/O-RAN technologies. The sensing information allows us to detect changes in the surroundings, as compared with a reference situation, enabling use cases such as obstacle detection. An xApp subscribed to both data sources is able to fuse the sensing information to predict coming communication conditions and exploit the MAC scheduler SM to adapt the allocation policy.

3.1.1 Localization and tracking of users in the mmWave band

We consider a network-based localization and tracking of a target user with position $p=(x,y)$ moving with a constant velocity. The localization is performed by combining the positioning measurements, TDoA and AoA from $i=1, 2, \ldots, N$ synchronized fixed mmWave APs at positions $p_i=(x_i, y_i)$. The APs are equipped with a uniform linear array (ULA) with M antenna elements, and the target UE is equipped with an omnidirectional antenna ensuring the signal radiates uniformly in all directions. The schematic of the proposed localization system framework in shown in Figure 3-1. The uplink positioning related measurements are computed across each of the APs and fed into the localization server. We then apply a Bayesian state estimation method to localize and track the target user.

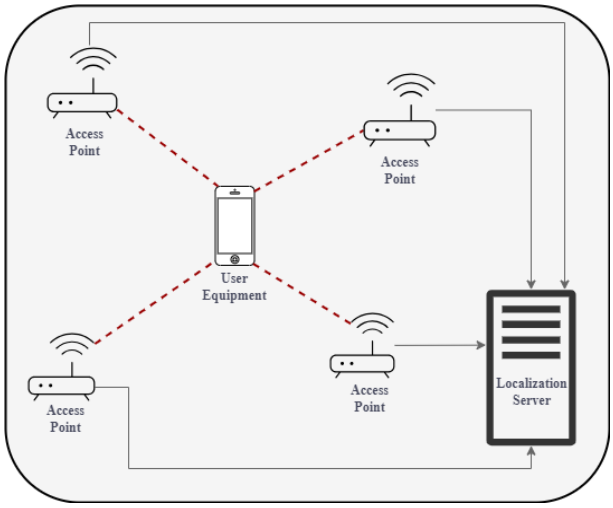


Figure 3-1 Schematic of the proposed localization system framework

We consider 4 mmWave APs within a 100 m by 100 m indoor area. The performance of the algorithm is evaluated over a fixed trajectory of the target UE at a constant velocity of 2 m/sec. Each of the APs provide positioning measurements to the localization server at an update rate of 10 Hz, i.e. with a sampling period of ΔT = 0.1s. The range differences and the AoA estimates are fed into the localization server. In our 2D simulation scenario, the signal attenuation is modeled using the free space path loss (FSPL) model, assuming Line-of-Sight (LoS) conditions between the APs and the user. The path loss exponent is set to 2 for ideal free-space conditions. The target user transmits a waveform based on the pseudo-random sequence with a transmit power of 10 dBm. Each of the APs are equipped with a ULA of 16 antenna elements. The UE transmit and the AP receive antenna gain Gtx and Grx is 3 dBi and 10 dBi respectively. In line with the 5G frequency band, we adopt the mmWave APs to be operating at a channel bandwidth of 160 MHz. Additional state estimation and communication link simulation parameters are provided in Table 3-1.

Table 3-1 Simulation parameters

Parameters	Value
Number of mmWave APs	4
Operating frequency	28 GHz
Bandwidth	160 MHz
Target Velocity	2 m/s
UE Transmit Power	10 dBm
Noise floor at APs	-89 dBm
Number of particles	100

Sampling period ΔT	0.1 s
Total tracking duration T	125 (s)
Process noise σ_v	0.1 m/s
TDoA measurement noise σ_d	50 cm
AoA measurement noise σ_θ	0.5◦

The performance of the proposed algorithm is analysed through Monte Carlo simulations in MATLAB. We analyze the localization error statistics when using hybrid TDoA/AoA measurements using the Weighted Least Squares (WLS) approach. Additionally, the performance of the positioning and tracking using hybrid TDoA/AoA measurements using the Extended Kalman Filter (EKF) and the Particle Filter (PF)-based method are compared and analysed. To do the performance analysis, we compute the Cumulative Distribution Function (CDF) and additional positioning error statistics of the proposed algorithm.

The CDF of the positioning error for the hybrid TDoA/AoA method using WLS, EKF and PF are shown in Figure 3-2. The conventional WLS method with the hybrid TDoA/AoA measurements is simulated by solving the system of equations as in [25]. It can be observed that the mean positioning error for the WLS method is 96.12 cm. However, the combination of TDoA and AoA measurements using EKF and PF results in a mean positioning error of 10.97 cm and 29.27 cm respectively.

The results show a significant improvement in the positioning accuracy by 88.59 % and 69.55 % relative to the conventional WLS technique. Moreover, it can be observed that the EKF-based hybrid TDoA/AoA method with good initialization conditions has a localization accuracy improvement of about 62.52 % as compared to the PF-based method. Table 3-2 summarizes the estimation errors for the EKF, PF and WLS-based method showing the mean, standard deviation and the 95th percentile errors. Furthermore, it can be observed that 90% of the localization error is less than 19 cm for the EKF-based method.

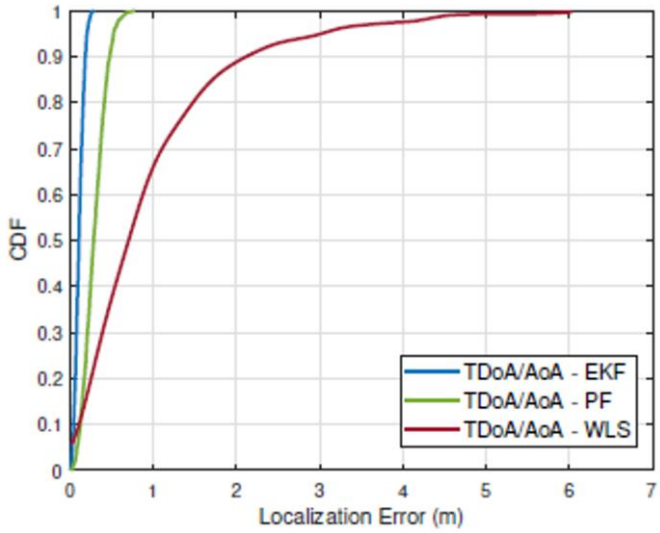


Figure 3-2 CDF of localization error with hybrid TDoA and AoA using WLS, EKF and PF-based algorithm

Figure 3-3 presents the boxplot of the localization error distribution for each of the hybrid methods analyzed in this subsection. The interquartile range (IQR) represents the absolute localization error indicated by the boxes. The width of the box indicates the spread of the location estimation error. The ”+” signs indicate the outliers, that represent the location estimations that have significant deviation from the true position of the

target UE. The percentage of outliers for the WLS method is much higher of about 6.16% while that of EKF and PF is much lower of about 0.24% and 0.65% respectively.

Table 3-2 Positioning Error for different methods

Hybrid Localization Method	Mean Error (cm)	95th Percentile Error (cm)	Standard deviation (cm)
TDoA/AoA-EKF	10.97	20.15	5.37
TDoA/AoA-PF	29.27	52.13	13.58
TDoA/AoA-WLS	96.12	304.6	96.14

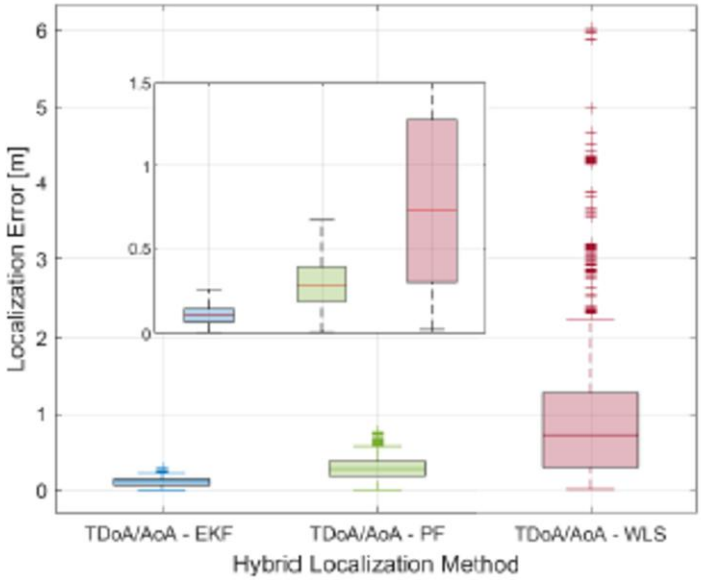


Figure 3-3 Boxplot of the localization estimation errors for each hybrid TDoA and AoA method

Figure 3-4 shows the Root Mean Square Error (RMSE) demonstrating the tracking error of the EKF and PF-based method. The Position Error Bound (PEB) curve, indicated by the red dotted line in the figure, shows the theoretical lower bound of the localization error under the simulated measurement and noise parameters. It can be observed that the EKF-based localization method has a lower tracking error and approaches the theoretical bound as compared to the PF-based method. This is because we consider favorable initial conditions for EKF with linear system dynamics and Gaussian measurement noise model. The PEB reduces as the state estimator is updated over time with new measurement data and the a priori information, therefore resulting in a more accurate positioning as the time progresses.

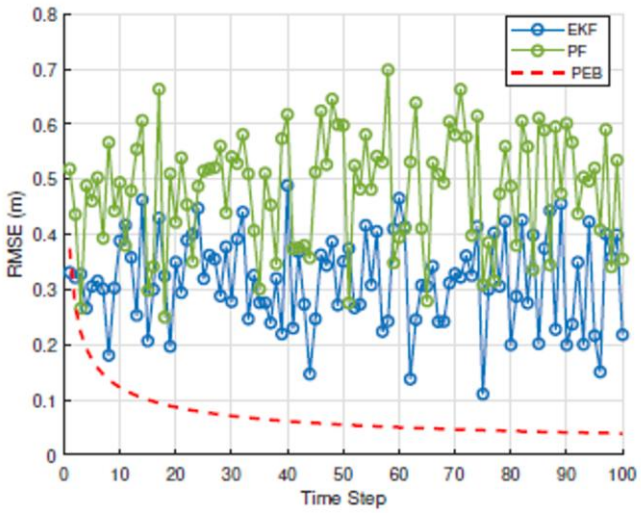


Figure 3-4 The RMSE of the EKF, PF, and PEB along the user trajectory

3.1.2 Wi-Fi Sub-8 GHz Sensing Evaluation

This subsection evaluates the feasibility and performance of monostatic sensing using commercial off-the-shelf (COTS) Wi-Fi NICs in the Sub-8 GHz band. The evaluation focuses on two key aspects: the fundamental accuracy of Range-Doppler estimation compared to reference radar systems, and the application of this estimation to efficient HPD.

3.1.2.1 Range-Doppler Estimation Accuracy and Validation

To validate the capability of commercial NICs to perform radar-like sensing, a signal processing pipeline was developed to extract range and Doppler information from the CSI of Long Training Symbols (LTF). The system addresses hardware asynchronization through time-offset cancellation and phase alignment (see Figure 3-5, which illustrates the successful delay calibration, showing the phase and range bin estimation before and after the calibration, where the maximum correlation peaks are now consistently aligned at range bin zero, confirming accurate time synchronization), while mitigating self-interference (SI) via Direct Current (DC) component removal.

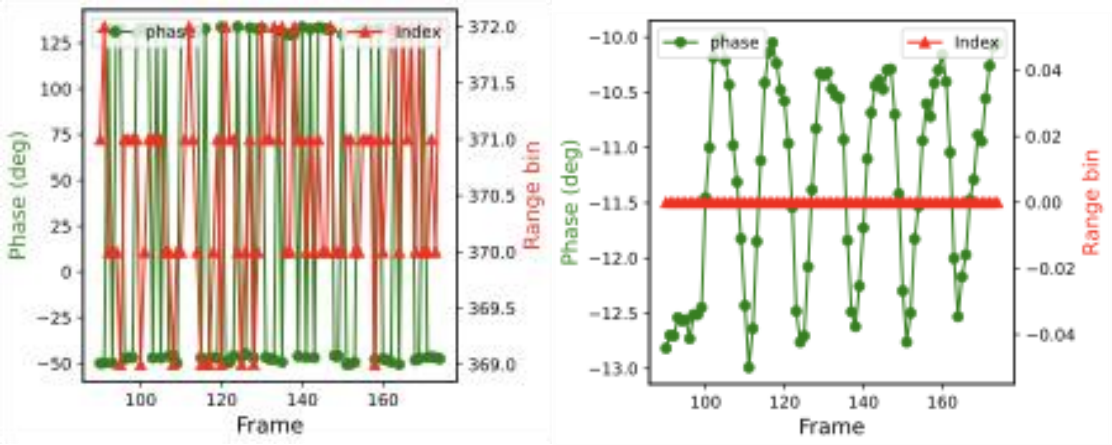


Figure 3-5 Before (left) and After (right) Time-Phase Synchronization

The evaluation of this estimation engine was conducted using an Intel Wi-Fi AX211 NIC and validated against a commercial high-resolution mmWave radar (Infineon BGT60TR13C) serving as ground truth.

- **Evaluation Scenarios:** The system was tested with human hand performed repetitive gestures (0–40 cm range) to evaluate the resolution of complex biological motion.
- **Self-Interference Cancellation Performance:** Analysis of the Range-Doppler Maps (RDM) confirmed that without cancellation, the strong Tx/Rx coupling obscures targets; however, applying the proposed SI mitigation renders moving targets clearly distinguishable, allowing for precise estimation (see Figure 3-6, the figure shows the range-Doppler map estimation for a single frame based on CSI Wi-Fi data without/with SI cancellation. Without the SI cancellation the strong Tx/Rx coupling signal completely obscures the real target, making target detection challenging. With the SI mitigated, the moving target becomes clearly distinguishable, allowing for accurate estimation of its range and Doppler values.).
- **Quantitative Accuracy:** Despite the bandwidth limitations of Wi-Fi (160 MHz) compared to the reference radar (4 GHz), the monostatic Wi-Fi system achieved centimeter-level accuracy. The evaluation yielded median deviation errors of only **0.05 m in range** and **0.03 m/s in velocity** when compared to the mmWave radar ground truth (see Figure 3-7).

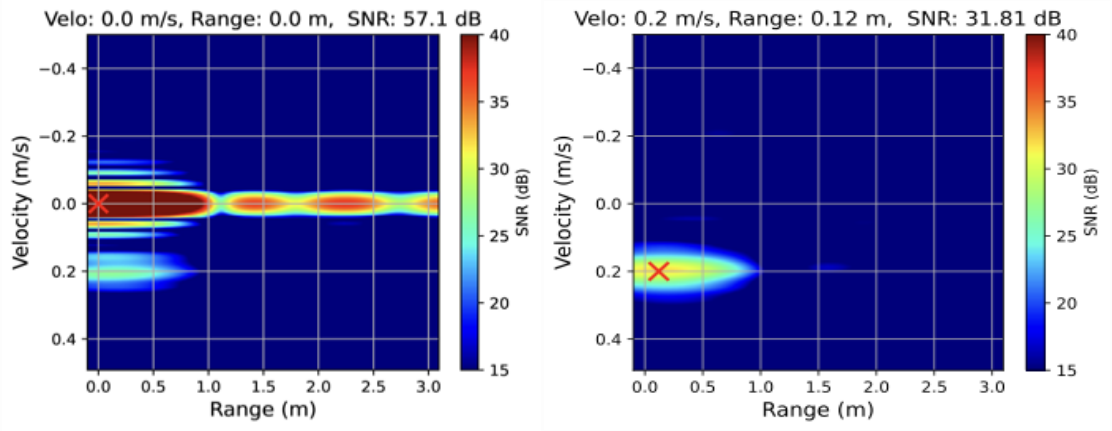


Figure 3-6 Range-Doppler Map Estimation with/without SI Cancellation

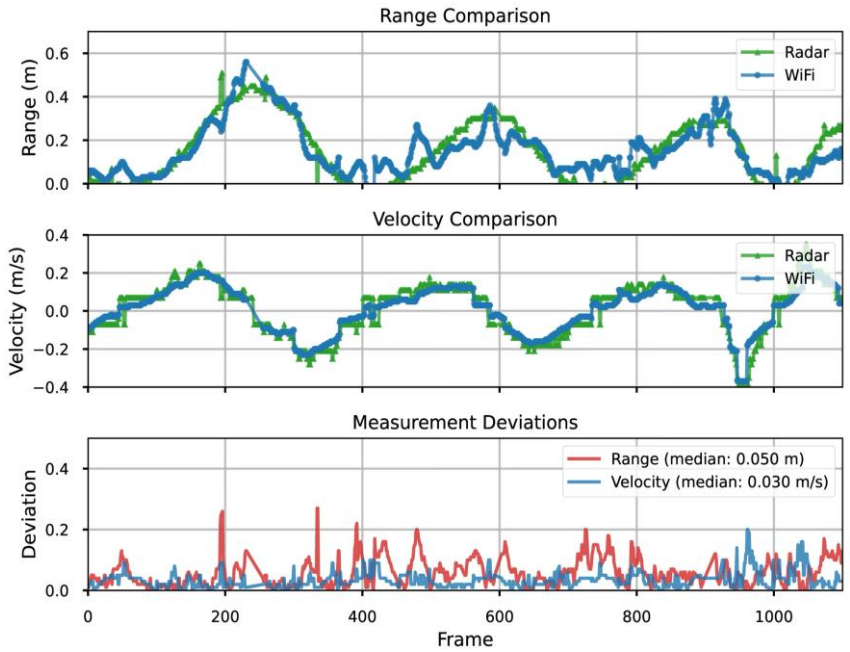


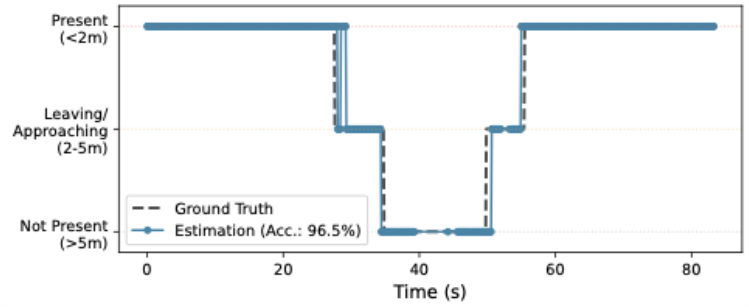
Figure 3-7 Range and Doppler Estimation in Human Gesture

3.1.2.2 *Efficient Human Presence Detection (HPD)*

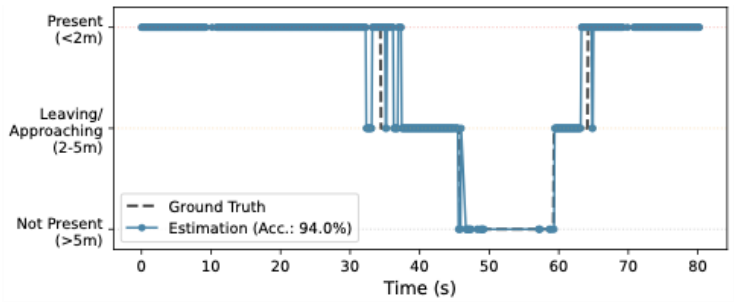
Building upon the validated range-Doppler extraction capabilities, the system was further evaluated for HPD to enable intelligent power management features like "Wake-on-Approach". To achieve this on battery-powered devices, the processing was optimized into a **Range-Filtered Doppler Spectrum (RF-DS)** technique, which replaces computationally expensive full 2D maps with targeted spatial filtering.

The evaluation of this optimized HPD system highlighted several performance enhancements over traditional methods:

1) **Micro-Motion Sensitivity (Breathing Detection):** The evaluation compared standard DC cancellation against a Finite Impulse Response (FIR) based Moving Target Indication (MTI) filter. Results demonstrated that the MTI approach significantly enhanced the visibility of micro-Doppler signatures, successfully detecting subtle user respiration at distances of up to 3 meters.
2) **Macro-Motion Tracking (Approach/Departure):** The system was tested with users moving from a seated position to a distance of 8 meters. The RF-DS method provided stable range and velocity estimates throughout the cycle, maintaining robust detection at extended ranges suitable for automated lock/wake triggers.
3) **Cross-Platform Consistency:** To ensure hardware independence, the solution was validated on laptops equipped with different NIC generations (Wi-Fi 6E and Wi-Fi 7). The system achieved consistent detection accuracy exceeding **94%** on both platforms for identifying presence, approach, and departure events (see Figure 3-8).
4) **Adaptive Power Efficiency:** The framework was validated with an adaptive multi-rate sampling scheme, operating at a low frame rate (10 Hz) during idle states and switching to high fidelity (100 Hz) only upon motion detection, thereby minimizing power consumption while ensuring low-latency response.



(a)



(b)

Figure 3-8 Presence, absence, approach, and leave state estimation for (a) Lenovo Wi-Fi 6E and (b) HP Wi-Fi 7 laptops. Both platforms demonstrate consistent performance, confirming the hardware independence of the proposed approach.

3.1.3 Improved MAC scheduling

This section illustrates the integration of AI/ML in the 6G-SENSES architecture by describing a DRL-based xApp that optimizes MAC scheduling decisions through the dynamic reconfiguration of key parameters based on network dynamics and sensing information.

The considered scheduling architecture, coined as Reinforcement-Based Intelligent Scheduling (RBIS), follows the Drift-plus-penalty (DPP) algorithm [26] which, in turn, relies on Lyapunov Control Theory to balance multiple objectives. In particular, it aims to jointly meet guaranteed bit rates, ensure traffic queues stabilization, delay, and resource efficiency. The DPP behavior can be adjusted through a set of parameters: V, w_Q, and w_G, which influence the scheduler priorities towards resource efficiency, latency control, and throughput, respectively.

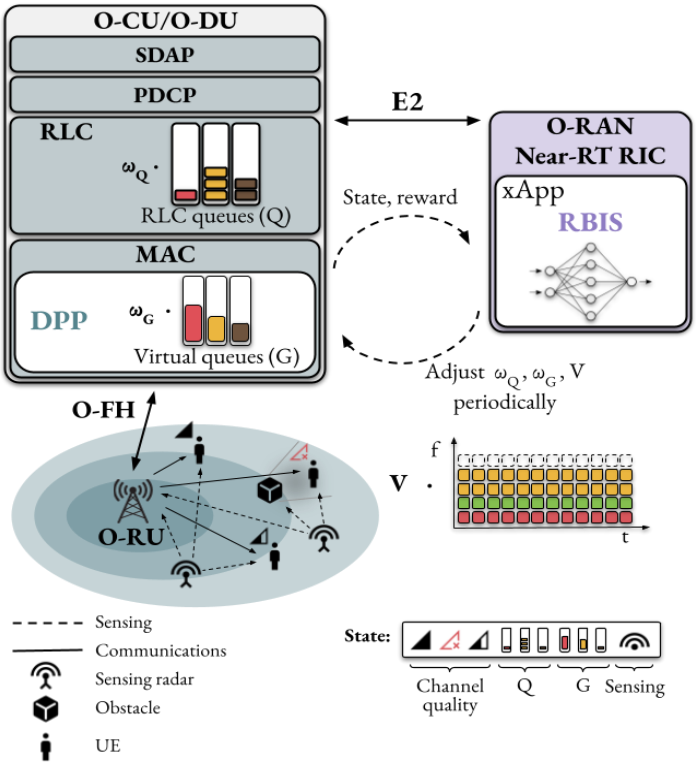


Figure 3-9 Overview of the proposed 5G O-RAN MAC scheduling architecture.

Figure 3-9 shows the overall RBIS architecture, in which an xApp is responsible for the configuration of the MAC scheduler in real-time. For this purpose, several metrics are continuously collected from the RAN, such as the throughput achieved by each UE and the occupancy of their RLC buffers resulting from previous MAC scheduling decisions, among the current channel status for each UE. Furthermore, the proposed DRL agent can leverages ISAC functionalities to gather comprehensive information about the network environment. This sensing data includes user positions, mobility patterns, and obstacle detection. The ISAC-provided sensing data is critical for enhancing the scheduler's situational awareness, allowing it to anticipate channel degradation, identify blockages, and proactively adapt resource allocation.

The operational interaction between the MAC scheduler and the xApp is a cornerstone of this dynamic reconfiguration mechanism. The scheduler, residing within the O-DU, is subscribed to the xApp deployed in the Near-RT RIC. This subscription enables a continuous exchange of information via the *E2* interface, which is specifically designed for near-real-time communication between the RIC and the RAN components, typically operating within control loops of 10ms to 1s. This real-time information exchange is subject to

stringent timing requirements to ensure the effectiveness of the DRL-based scheduling. The scheduler regularly transmits various network state metrics to the xApp. These metrics form the comprehensive "State" input for the DRL agent. Additionally, a reward, calculated based on the impact of the scheduler's decision on the RAN's performance, is sent to the xApp. Upon receiving this aggregated state information, the DRL agent within the xApp processes it to determine the optimal configuration parameters (V, w_Q, w_G) for the MAC scheduler. This decision-making process, while complex, must be executed within tight deadlines to ensure that the scheduler can adapt promptly to rapid changes in network conditions and user demands. The updated configuration parameters are then transmitted back to the MAC scheduler via the E2 interface. Thus, the strict time requirements of the E2 interface and the near-real-time RIC (10ms to 1s control loops) are vital for the proper functioning of the entire system. Any significant delay in the state reporting, decision-making by the DRL agent, or parameter reconfiguration could lead to suboptimal scheduling decisions, increased latency, and a failure to meet QoS requirements. For instance, if the network experiences a sudden surge in traffic or a critical channel degradation due to a moving obstacle, the xApp must quickly process this information and adjust the scheduler's priorities. A delayed response could result in packet drops, congestion, and a degradation of user experience, particularly for delay-sensitive applications.

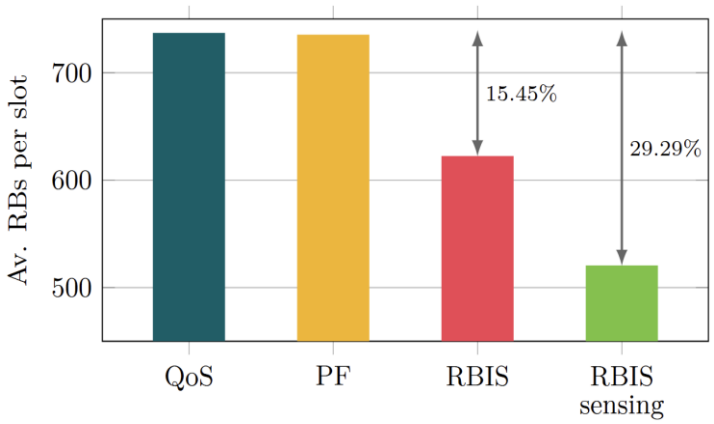


Figure 3-10 Average number of RBs allocated per slot under 3GPP channel and random mobility models.

The evaluation results demonstrate the effectiveness of the scheduling solution. For instance, in a synthetic scenario with controlled channel degradation due to obstacles, it consistently meets the minimum throughput requirements – Guaranteed Flow Bit Rate (GFBR) – while significantly reducing resource usage. As shown in Figure 3-10, the approach enhanced by sensing information (i.e. RBIS sensing) achieves up to 29% fewer average RBs per slot compared to other solutions like Proportional Fairness (PF) or other QoS-aware schedulers, highlighting its superior resource efficiency. Even without sensing (i.e. RBIS), the reduction in RBs usage still achieves a notable 15%.

3.1.4 ISAC transport network optimization

We consider the system architecture shown in Figure 3-11 that exploits an optoelectronic transport network interconnecting the RAN with the core functions located at edge and central cloud compute resources. The hierarchical structure of the proposed architecture offers RAN connectivity and collects and aggregates communication and sensing traffic streams from the RUs, while it transports these to edge and central cloud servers for processing. For the RAN segment, we consider a typical 5G compatible MIMO-Orthogonal Frequency-Division Multiplexing (OFDM) waveform. The primary objective of the system is to successfully establish connections for communication purposes between the 5G RUs and the User Equipment (UEs). The OFDM waveform is organized into frames with 10ms duration each comprising ten subframes of 1ms. The bandwidth (in MHz) allocated per RU i is denoted as W_i. OFDM waveforms transmitted for communication purposes are reflected by obstacles in the environment and received by the RUs. These echo signals can then

be processed to sense the surrounding environment and detect moving targets acting as a doppler OFDM radar [27]. The main parameters characterizing the performance of the OFDM doppler radar are distance and velocity resolutions, Δd_i and Δv_i, respectively, being the lowest distance and velocity such that two targets positioned at d_i and $d + \Delta d_i$ moving with velocities v_i and $v_i + \Delta v_i$ can be distinguished.

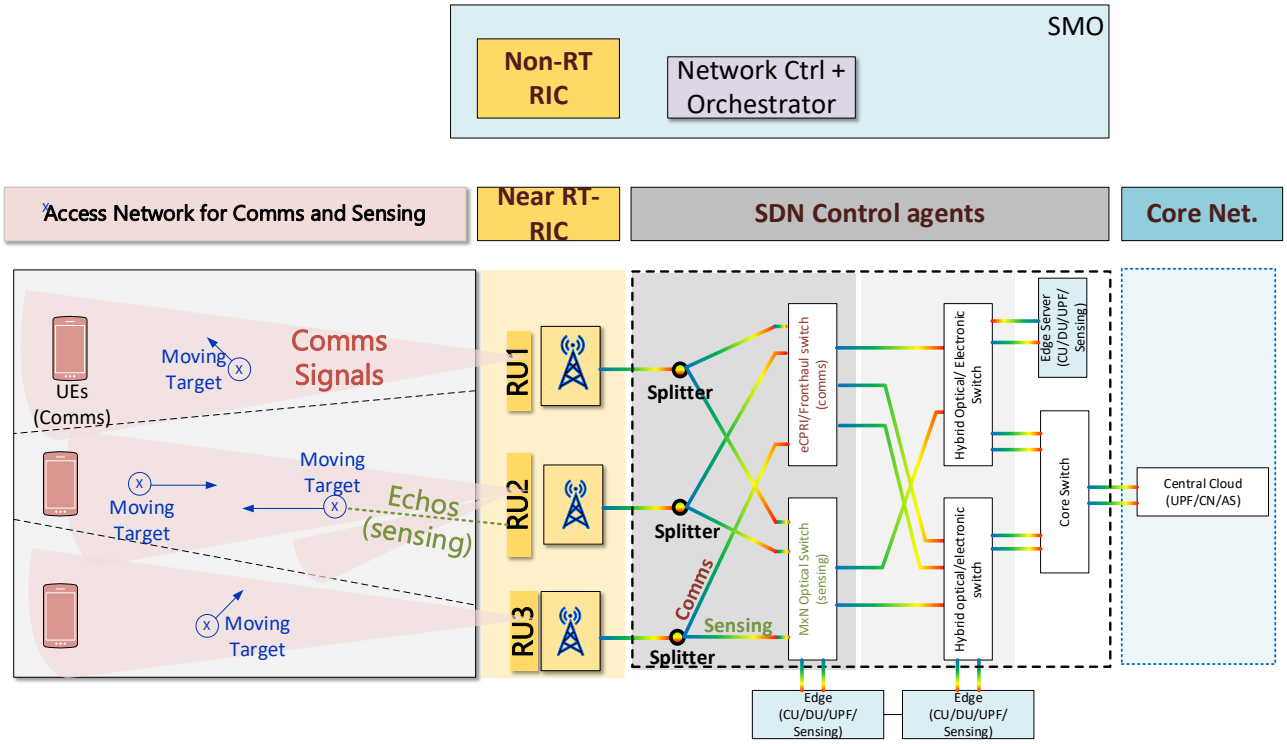


Figure 3-11 Integrated Transport Network for Comms and Sensing

For this type of radar, the range resolution Δd_i for echoes corresponding to communication signals transmitted by RU i is:

$$\Delta d_i = c_0/2W_i, \tag{3-1}$$

where c_0 is the speed of light. Similarly, the speed resolution for the moving targets is given by:

$$\Delta v_i = c_0/2f_c N_{frames} T_s \tag{3-2}$$

where f_c is the central frequency of the wireless system, T_s the duration of the OFDM symbol and N_{frames} the number of frames employed in the velocity estimation process. Therefore, by increasing the number of frames collected and processed by the OFDM radar, higher sensing accuracy can be achieved. However, increasing bandwidth and N_{frames} increases also the volume of sensing IQ streams which are transported and processed by the sensing app implementing the Doppler OFDM radar hosted in the edge.

To address these requirements, optical splitters are employed at the RUs duplicating I/Q streams creating two separate paths for comms and sensing. This allows comms streams to be forwarded through a combination of optoelectronic switches supporting the Common Public Radio Interface (CPRI)/extended CPRI (eCPRI) protocols for further processing at the DUs/CUs whereas sensing streams can be transferred and terminated at the edge server hosting the sensing app. The eCPRI compliant optoelectronic switches can

aggregate I/Q streams from multiple RUs maximizing the utilization of network resources. Aggregation of sensing flows from the RUs is performed adopting all optical switching technologies with switching times in the order of 25 ms-75 ms. This enables the collection of the necessary number of OFDM frames per tracked area allowing the sensing app to achieve the required sensing accuracy level.

This concept is demonstrated in Figure 3-12 where applying appropriate scheduling policies at the optical switches (i.e., policy that connects ingress port 1 to output port 1 for the time interval $(0, t_0)$, ingress port 2 to output port 1 for the interval (t_0, t_1) and ingress port 2 to output port 1 for (t_1, t_2)) we can transfer to the sensing apps the optimal number of N_{frames} from the RUs that are sufficient to detect the moving targets. This concept is shown in Figure 3-11 where for the RAN topology shown in Figure 3-13b, received sensing signals captured from RU4 can be used to detect only 2 targets (Figure 3-13 b) whereas sensing information from RU5 can detect 3 targets. Therefore, combing sensing from multiple RUs all moving targets can be detected.

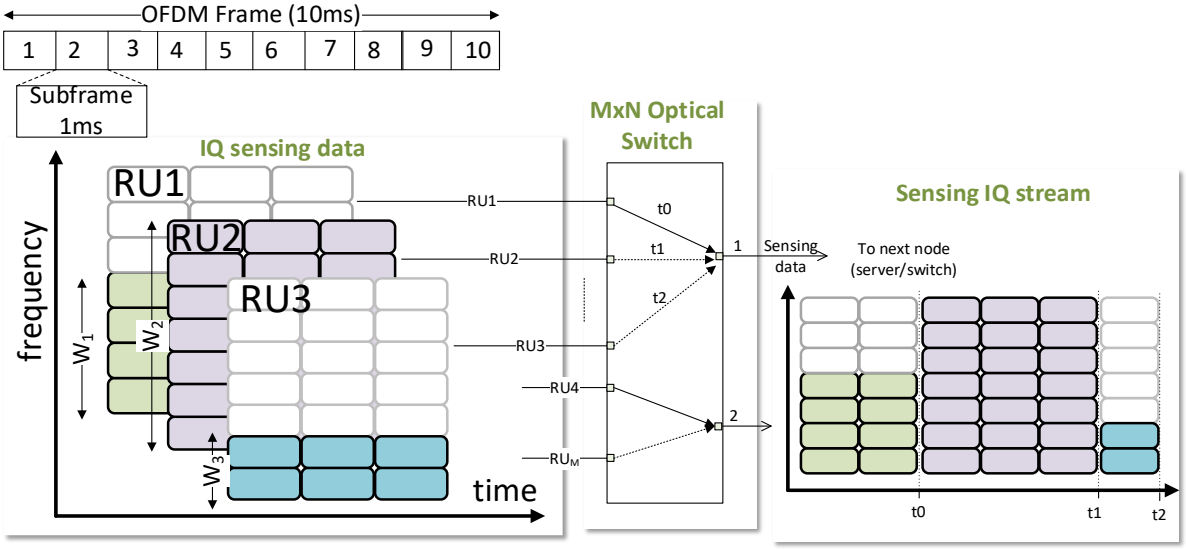
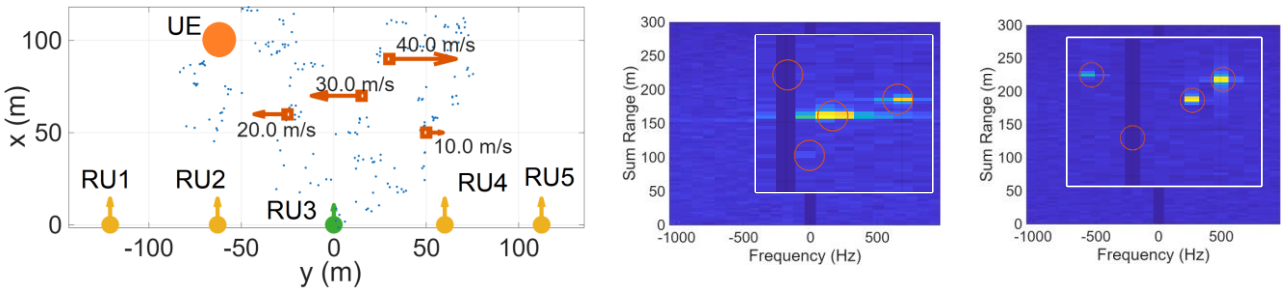


Figure 3-12 Sensing stream scheduling through Optical Transport Switching



a) b) c)

Figure 3-13 a) RAN configuration and moving targets, blue dots are static reflectors, Moving targets as detected by (b) RU4 and (c) RU5, Nframe = 12, W= 100MHz, RAN frequency =3.5GHz

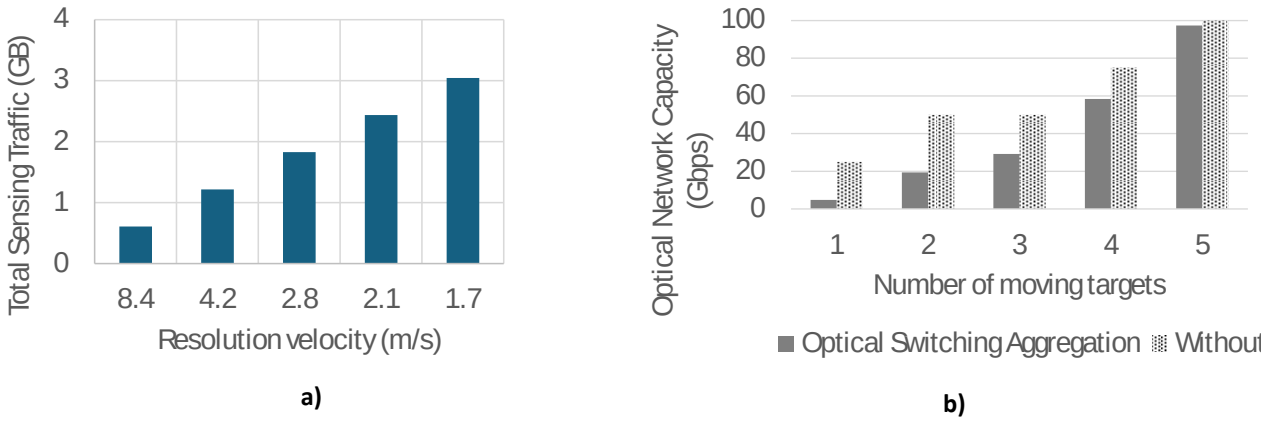


Figure 3-14 a) Volume of sensing traffic vs resolution velocity b) Optical transport network capacity allocated for sensing with and without optical network switching

The total volume of transmitted sensing information per RU for different resolution velocities is shown in Figure 3-14a). Based on (2), OFDM radar resolution can be increased by increasing the number of transmitted N_{frames}. As expected, this increases the volume of transmitted information that should be terminated at the sensing app. Finally, Figure 3-14 b). compares the total volume of transmitted sensing traffic as a function of the number of moving targets for the area shown in Figure 3-14b) with and without the aggregation optical switching functionality. For small number of moving targets, the required number of N_{frames} by the OFDM radar is small allowing the time switched optical network to aggregate sensing flows from a limited number of RUs. Increased number of targets introduces enhanced resolution requirements, in order to maintain the capability of the radar to clearly identify all moving targets. This results in increased number of N_{frames} leading to the need of continuous connection, thus eliminating the benefit of the optical switching aggregation functionality.

3.2 CF-mMIMO evaluation studies

Distributed coordination is a fundamental challenge in Sub-6 GHz CF-mMIMO, particularly in dense deployments with many distributed APs. Centralized schemes for user association and resource allocation become impractical due to fronthaul overhead, scalability constraints, and near-real-time requirements. We address this challenge by developing a fully distributed learning-based coordination architecture, enabling scalable AP–user assignment within the user plane.

3.2.1 System Model

We consider a Sub-6 GHz CF-mMIMO network consisting of N distributed APs serving K users, as sketched in Figure 3-15. APs are lightweight, performing local baseband processing, and exchange only limited features with peers. Users transmit orthogonal pilots, enabling the estimation of channel gain g_{kn}. Each AP can serve at most U users, while each user must be served by at least L APs, ensuring macro-diversity. The network aims at maximizing the sum-rate

$$max_{S} \sum_{k=1}^{K} log_2\left(1+\frac{\sum_{n=1}^{N} g_{kn}\, s_{kn}}{\sigma^2}\right) \quad (3\text{-}3)$$

Subject to:

- $s_{kn} \in \{0,1\}$ (binary AP-user assignment)
- $\sum_{k=1}^{K} s_{kn} \leq U$ (AP capacity constraint)
- $\sum_{n=1}^{N} s_{kn} \geq L$ (user reliability constraint)

Unlike mmWave, phase-coherent combining is feasible at Sub-6 GHz, thus the received power adds coherently.

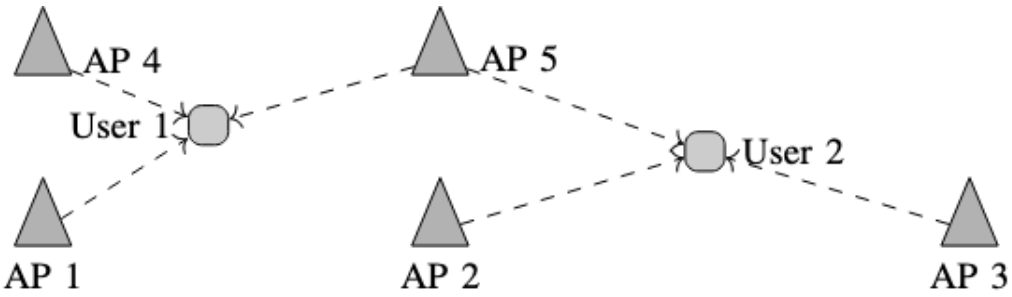


Figure 3-15 Illustration of a CF mmWave deployment in which users can be simultaneously served by several distributed APs, enabling cooperative beamforming.

3.2.2 Distributed Learning-Based Optimization Architecture

To avoid centralized computation and exhaustive search, TUBS proposes a GNN-based distributed optimizer:

- The CF-mMIMO network is represented as a bipartite graph of APs and users.
- Each AP performs local inference, using only neighborhood messages.
- No global controller is required.

Core architectural features:

- Hierarchical Permutation Equivariant (HPE) GNN – supports scalability and arbitrary numbers of APs/users.
- Augmented Lagrangian Method (ALM) – enforces connectivity constraints during inference without explicit discrete optimization.
- Entropy-based relaxation – approximates binary assignments while keeping the solution differentiable.
- Teacher–student graph pruning – learns sparse message-passing topologies, reducing fronthaul overhead.

This enables near-real-time distributed coordination, where each AP autonomously updates its served user set with minimal communication. The iterative inference mechanism of the proposed architecture is illustrated in Figure 3-16.

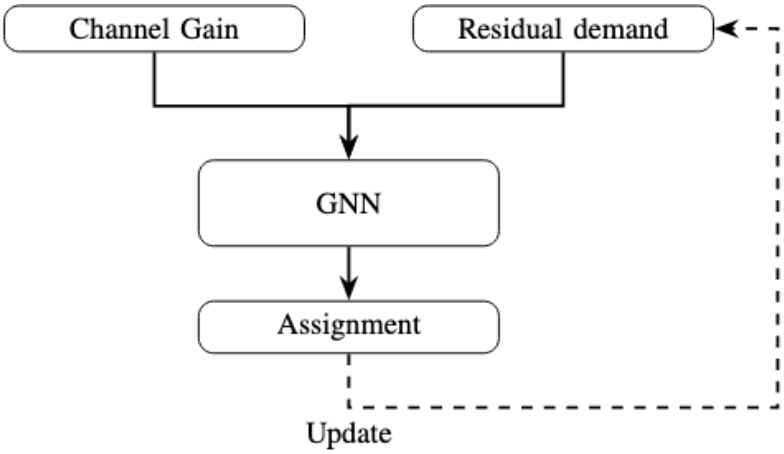


Figure 3-16 Iterative GNN-based AP–user assignment for CF-mMIMO.

Evaluation Plan and Expected Results

The proposed learning-based CF-mMIMO scheme will be evaluated under realistic sub-6 GHz propagation with user mobility and large-scale fading. The assessment will focus on sum-rate performance, fairness, fronthaul efficiency, and scalability with increasing AP and user densities. The expected outcome is a near-optimal yet fully distributed user-association mechanism that significantly reduces coordination overhead and enables real-time user-plane operation for practical 6G deployments. Figure 3-17 and Figure 3-18 illustrate a typical deployment scenario of distributed APs and users considered in our evaluation.

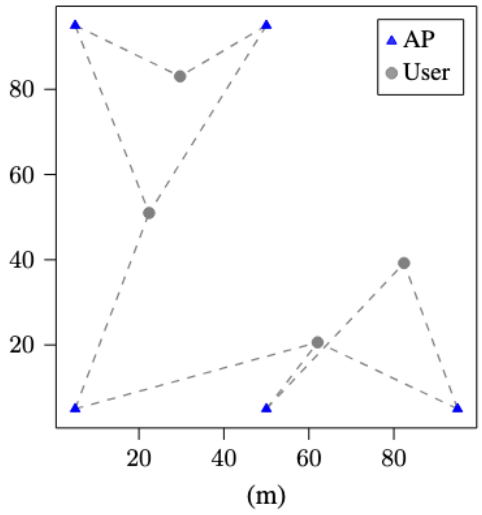


Figure 3-17 Small network scenario with 5 APs.

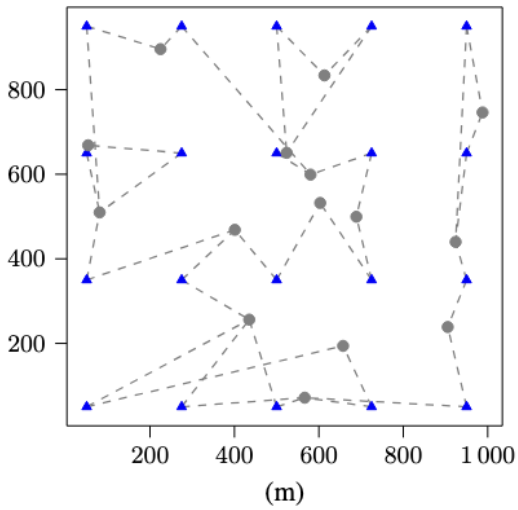


Figure 3-18 Large network with 20 APs.

3.2.3 Summary and Next Steps

The proposed distributed GNN-based optimization provides a promising approach for scalable and autonomous user-plane control in Sub-6 GHz CF-mMIMO networks. This section focused on establishing the core architectural design and algorithmic foundations required for a subsequent comprehensive performance evaluation. The next steps involve implementing the prototype in Python, evaluating its performance under dynamic conditions, and integrating the framework into the 6G-SENSES system-level simulation environment during project months M24-M30.

3.3 Reconfigurable Intelligent Surfaces (RISs)

RISs have become one of the most intriguing developments in advanced wireless research, largely because they point toward a future where the physical environment is no longer a passive bystander but an active participant in communication. The work presented shows the development of RIS steadily acquire more responsibilities: first enhancing and participating in modulation, then stabilizing communication in high-mobility three-dimensional settings, later enabling simultaneous communication and sensing, and eventually operating as active, amplified intelligent surfaces at THz bands.

First, in the field of index modulation, the question comes how to increase spectral efficiency and, at the same time, reduce hardware complexity in MIMO systems. Classical spatial modulation techniques face an inherent limitation—only one transmit antenna is active in a given time slot, and the achievable rate depends heavily on how quickly the system can switch the RF chain among antennas. RIS offer a promising way around this constraint. By dividing the RIS into controllable sub-arrays and treating their indices as part of the modulation alphabet, it becomes possible to embed additional information in the selection of the RIS block

itself. This idea leads to the development of a structure in which the RIS is no longer "just a mirror" but a carrier of information. Introducing an offset between the activated transmit antenna and the selected RIS block further relaxes the RF switching burden while improving reliability and spectral efficiency. Several strategies for choosing the RIS block—ranging from simple fixed selections to more adaptive and gain-aware schemes—illustrate how the surface can be used to strike different balances between complexity, energy consumption, and error performance. What is important here is the shift in perspective: RIS start being treated as a part of the signaling mechanism, integrated into the transceiver architecture rather than appended to it. The core concept is shown in Figure 3-19.

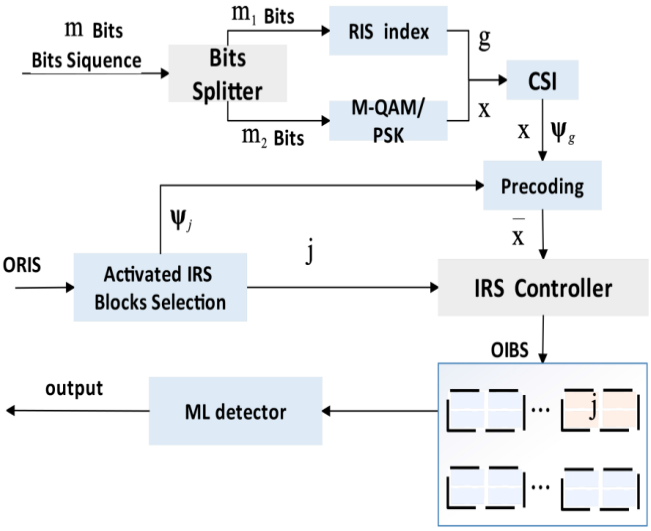


Figure 3-19 System model of the IRS-Offset Index Modulation scheme.

Once RIS are brought into the modulation process, the next obvious question is whether they can be pushed further into real-world, mobility-heavy environments—particularly those involving non-terrestrial networks, UAVs, and complex 3D flight paths. These settings introduce challenges that do not appear in static terrestrial scenarios. A UAV changes altitude, angle, and velocity in ways that cause rapid variations in channel conditions. Doppler shifts become substantial, LoS links can appear and disappear within seconds, and antenna patterns face continuous spatial reorientation. Traditional beam-forming struggles under such volatility, and airborne communication links often suffer from interruptions that directly affect reliability and latency.

RIS can help, but only if they respond fast enough and with sufficient intelligence. The research extends the earlier RIS concepts into this new terrain by pairing them with high-mobility channel models and designing real-time optimization strategies that adapt to the movements of UAVs and other airborne platforms. Instead of assuming static propagation, the RIS is configured to work with path loss variations, aerial shadowing, and interference patterns that vary with trajectory. The benefit is significant: the RIS can reconstruct favourable propagation paths, compensate for Doppler effects, and restore coherent energy focusing even when the UAV follows complex routes.

Rather than restricting the analysis to an abstract environment, the work deliberately examines several representative mobility scenarios. These include dense urban air corridors where buildings introduce intermittent blockages, rural deployments where UAVs must maintain communication across long horizontal stretches, emergency-response situations that require abrupt altitude changes, and multi-UAV swarms where inter-drone interference is a concern. Across these settings, the role of RIS becomes clearer. It is not only a beamforming tool; it becomes a stabilizing influence in a network that is otherwise highly unpredictable. Through joint optimization of RIS phases, UAV positions, and mobility constraints, the

communication links become more resilient, and coverage improves even when the environment is hostile to traditional RF propagation [33]. This step marks a shift from RIS as a clever modulation aid to RIS as a mobility-aware network element, one that supports the integration of terrestrial, aerial, and potentially satellite nodes in future 6G architectures. Figure 3-20 shows the aforementioned scenario.

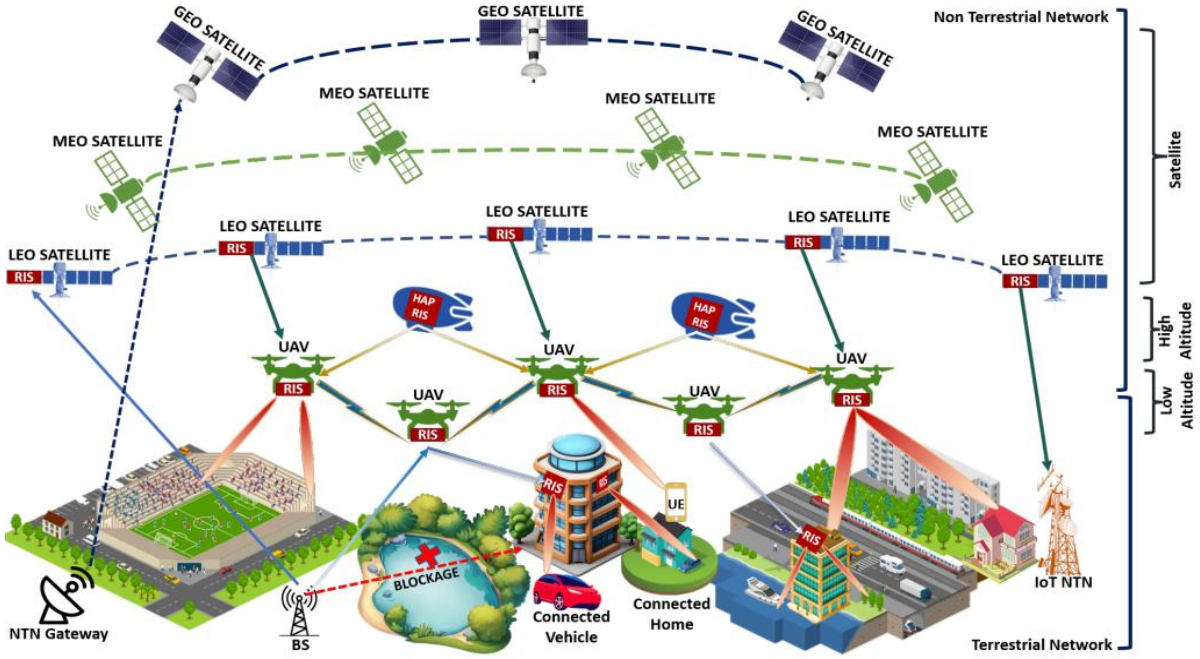


Figure 3-20 RIS-enhanced 6G NTN architecture integrating satellites, UAV-based aerial relays, and terrestrial BSs for seamless connectivity in dynamic environments.

The next stage builds on this mobility foundation and moves toward a central theme in emerging 6G research: the unification of communication and sensing. Instead of deploying separate hardware for radar and data transmission, ISAC attempts to use the same physical channel and often the same waveform for both tasks. For such systems to operate efficiently, especially in full duplex, they require a surface that can manage signals on both sides and allow flexible partitioning of energy between transmission and reflection. This need leads naturally to simultaneous transmitting and reflecting (STAR-RIS).

Compared with the earlier, purely reflective RIS models, STAR-RIS represents a conceptual leap. Each element can direct part of the incoming energy forward (transmission) and part backward (reflection), allowing the surface to cover an entire 360-degree region rather than a single hemisphere. In an ISAC context, this flexibility becomes extremely valuable. Downlink communication, uplink communication, and sensing of a target can occur at the same time, in the same frequencies, with the STAR-RIS shaping how much energy goes toward each task. Managing interference between these functions is not trivial; full-duplex operation introduces strong self-interference, and sensing requires that sufficient signal energy reaches the target and returns to the receiver [34].

To address these intertwined constraints, the optimization problem is approached with learning-based methods rather than static or iterative mathematical techniques. Meta-reinforcement learning is used to adjust the STAR-RIS configuration so that communication quality and sensing accuracy are jointly improved. One major advantage of using learning-based techniques is that they remain effective even when channel state information is imperfect—something particularly important in sensing scenarios, where echoes from targets may not provide clean CSI. Through this approach, the RIS becomes more than a passive optimization variable; it behaves as if it were an adaptive component that continuously improves its response to

environmental changes. In effect, RIS begins to resemble a "soft" programmable layer that links communication and radar functions.

This development naturally sets the stage for the final step in the progression: the use of active intelligent surfaces operating at THz frequencies, coupled with advanced multi-agent reinforcement learning (MARL) and delay-aligned modulation. THz bands offer massive bandwidths for ultra-high-capacity communications, but they also come with severe propagation challenges. Signals attenuate quickly, diffraction is minimal, and even small obstacles can cause complete link failure. For THz ISAC systems, simply reflecting signals is not enough; the surface must also compensate for the harsh propagation losses intrinsic to this band.

Active STAR-RIS (ASRIS) addresses this limitation by equipping each surface element with a low-power amplifier, enabling it to boost signal strength while still offering independent phase control. This changes the nature of the RIS entirely. Instead of a lossless reflective sheet, the surface becomes a semi-active device capable of shaping both energy and phase across transmitted and reflected paths. The amplification allows THz signals to reach wider areas and improves the detectability of sensing echoes, which are otherwise too weak to be useful.

But THz propagation also introduces timing-related challenges. Different multi-path components may arrive with significant delay differences, causing inter-symbol interference and degrading radar resolution. To counter this, the research incorporates a dynamic delay alignment (DDA) modulation technique. Rather than relying on complex channel equalizers, the transmitter intentionally introduces structured delays so that signals from different propagation paths—including those routed through the ASRIS—arrive at the receiver in synchrony. The result is a simplified receiver design and a more robust communication link in fast-changing vehicular or high-mobility scenarios.

Coordinating all these aspects—beamforming, ASRIS coefficients, radar filtering, and mobility effects—creates an optimization problem too large and too coupled for centralized control. This motivates the use of multi-agent deep deterministic policy gradient (MADDPG) algorithms. Each agent learns and manages a part of the system, and through cooperation, they achieve global optimization across sensing and communication tasks. This decentralized learning structure is particularly advantageous in THz ISAC systems, where instantaneous CSI may not be available or where environmental conditions change too rapidly for deterministic solvers [35]. Figure 3-21 captures the whole scenario.

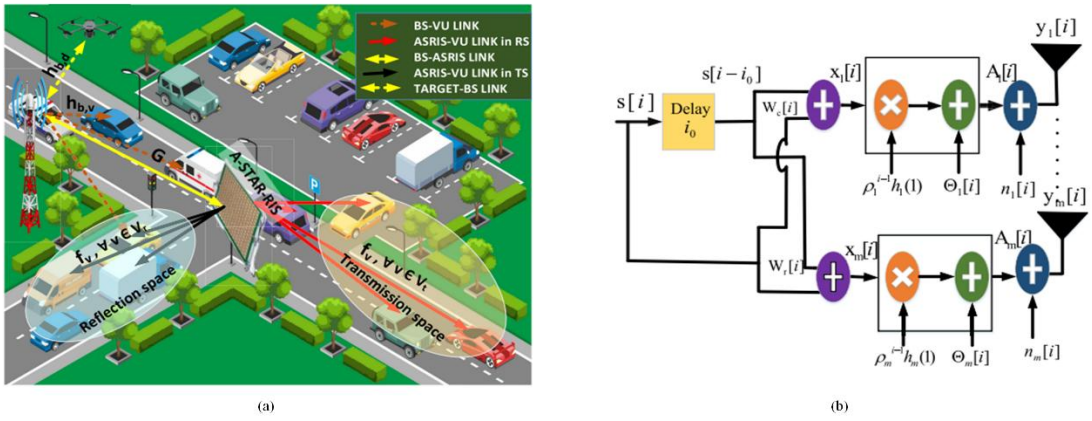


Figure 3-21 (a) ASRIS-aided ISAC system model (b). Block diagram of the proposed DDA modulation.

At this final stage, the RIS concept has transformed substantially from its initial form. What began as a passive surface used to embed additional index bits becomes an intelligent, active, cooperative element that supports THz communication, radar sensing, multipath delay alignment, and mobility awareness—all

optimized through reinforcement learning. The RIS no longer merely reflects signals; it amplifies them, reshapes them, synchronizes them, and adapts them with a high degree of autonomy.

Viewed as a whole, this progression illustrates a clear conceptual evolution. The RIS gradually acquires new roles: first as an efficient modulator, then as a stabilizer for UAV networks, then as a dual-function component in ISAC, and finally as a fully active and learning-enabled intelligent surface for THz environments. Each step introduces new capabilities and addresses new challenges, but the underlying theme remains consistent: the wireless environment is becoming programmable. Instead of treating the channel as something to be estimated and endured, the channel becomes something that can be engineered, shaped, and optimized in real time.

3.4 MEC evaluation

MEC is a network architecture concept that enables computational capabilities at the edge of the network. MEC technology is designed and implemented at the edge, allowing for on-site computation, reducing the workload assigned to central units, and improving the bandwidth usage of the network infrastructure.

We perform evaluation of WEC techniques in the context of MEC. Our solutions, proposed in deliverable **D4.1** [28] and in the Small-scale PoC#1 demo of deliverable **D5.1** [24] are suitable for intelligently using caching strategies at the edge, improving computation workload and bandwidth usage.

In this section, we describe the architecture proposed in order to evaluate our WEC solutions, and the different iterations of the simulation framework used to carry out performance evaluation.

We perform tests in a multi-layered network composed of:

- **Central BS:** the main unit orchestrating the execution of the application
- **Edge Devices:** devices deployed on the edge, suitable for caching-related tasks
- **Sensor Devices:** simple devices which are able to perform sensing-related tasks directly with the end-users. These are generally low-energy devices which can be queried/interrogated for simple tasks.
- **Agents / Mobile Edge Users:** the final end-user leveraging the application.

Our simulation setting is based on multiple iterations of increasing-complexity over ROS and Gazebo, thus allowing for a standardized and easy-to-extend approach:

- **Custom Simulation**: this is the first set of iteration, which focuses on delivering a solid caching mechanism in terms of numerical evaluation. The main focus of this first iteration it is to provide a lightweight simulator which is able to compare caching strategies in a more abstract environment, without taking into consideration the physical characteristics of the RAN sensing use case.
- **ROS-Gazebo Simulation**: the second set of testing iterations, which brings a more realistic environment to the simulation. With the addition of ROS and Gazebo, the simulation is now executed in a more standardized environment, with physically accurate objects taking part in the simulation. This allows us to simulate real scenarios, with multiple actors taking part into the simulation, creating a more realistic environment to test the caching solutions.
- **ROS-Gazebo + RAN Signal Simulation**: the final set of testing iterations, which extends the work to also simulate the RAN wireless sensing signal, and the cache data produced by the sensing use case; aligning with the objectives of the 6G-SENSES project by providing solutions directly tied to the PoC use case.

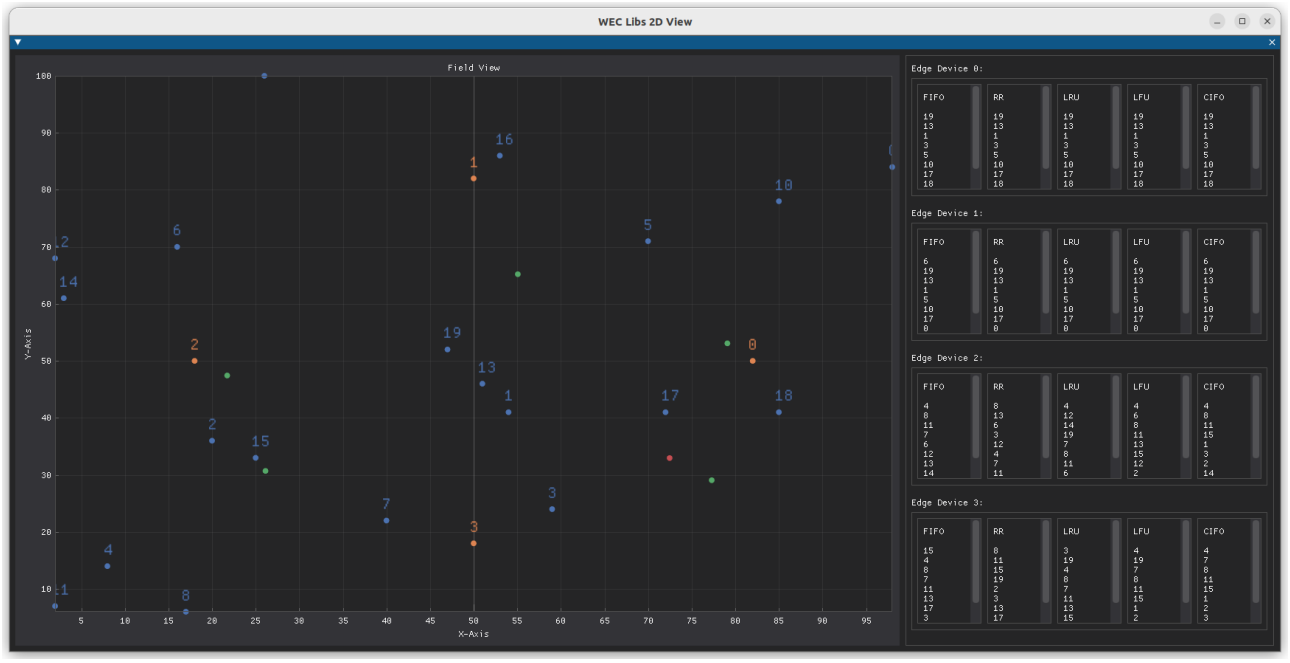


Figure 3-22 A 2D View of the WEC Libs, running the Custom Simulation instance.

In Figure 3-22 we show our libraries running in the Custom Simulation Environment. In this context, a top-down view of the field is displayed, and a live view of the cache content for Edge Devices is reported on the right. This view allows to study in detail how certain caching strategies behave with respect to other traditional ones.

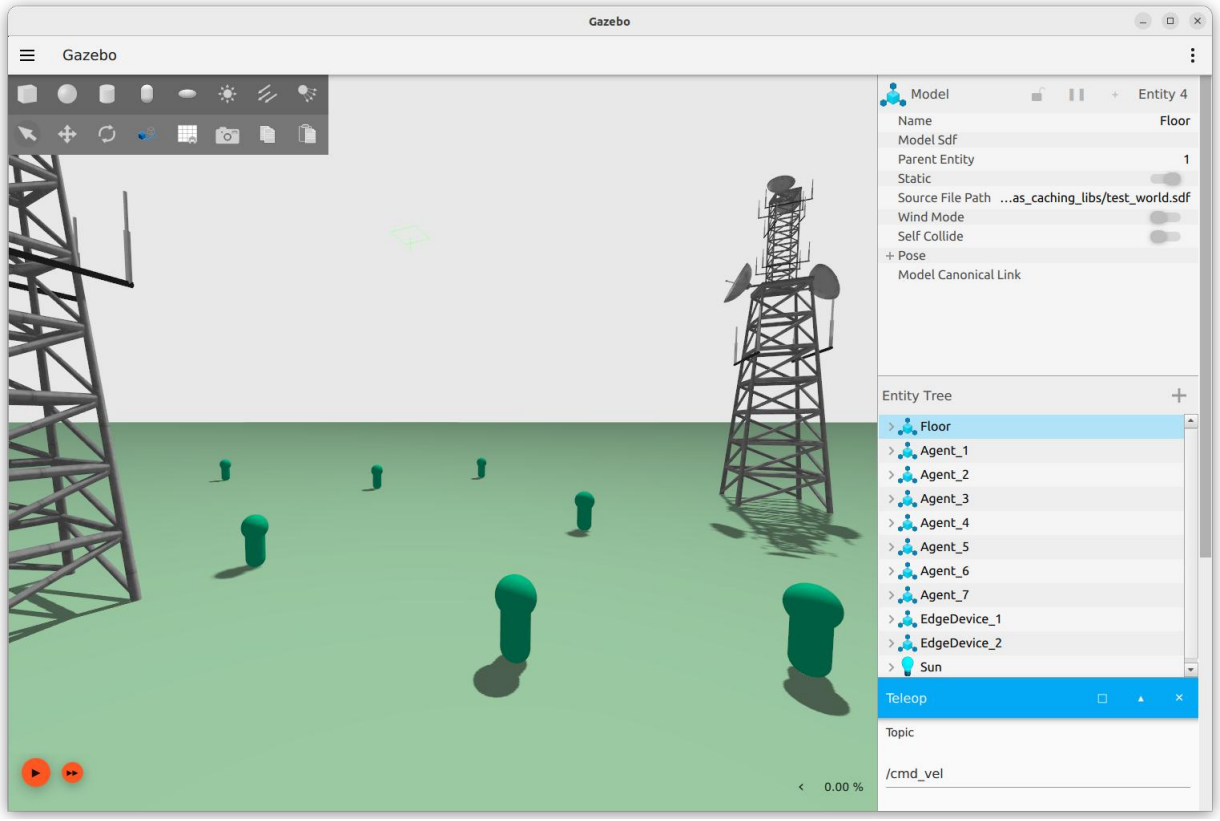


Figure 3-23 WEC Libs running the Gazebo Simulation instance.

Figure 3-23 shows a live view of the Gazebo Simulation in its current state. This simulation instance brings the whole application to a standardized environment. Communication interfaces are implemented through a Gazebo plugin, meaning that their behavior can be modified to reflect the presence of physical objects in the simulation, thus providing a high-fidelity testing instance in the simulation process.

All together, these phases allow us to deliver a comprehensive simulation environment, validating and strengthening our proposed solution through multiple progressive steps. The work of **D4.1** has covered, so far, the first two phases of the work, while the remaining work for the last phase will be addressed in **D4.2**.

4 Control Plane Components Evaluation

This chapter presents the evaluation of the control components of 6G-SENSES, focusing on both standards compliance and the assessment of advanced control-plane mechanisms enabled by the O-RAN architecture.

A first goal of this chapter is to define and apply a procedure for assessing the compliance of the RICs with O-RAN standards, while also verifying that the control components operate within the timing constraints required for hierarchical closed-loop automation. To this end, Section 5.1 introduces an assessment procedure covering both the Non-RT and Near-RT RIC components through dedicated subsections. The evaluation methodology is twofold: a functional audit verifies the availability of standardized interfaces (*R1AP* and *E2AP*) and service models necessary to support advanced use cases such as external sensing data fusion and SDN-based streaming, while a performance profiling regime utilizes Tracy Profiler to quantify the cumulative software latency of the control loops. By isolating the computational overhead of each component, the assessment determines whether the system effectively operates within the designated time domains—greater than one second for the Non-RT layer and sub-second for the near-RT layer.

Moreover, this Chapter evaluates a set of control-plane solutions developed within 6G-SENSES (independently of the standards-compliance procedure). In Section 4.2 we introduce a feedback-based RIC service for enforcing sensing delay requirements over the *E2* interface. Building on the sensing-oriented E2 extensions of Section 2.3 and the observed sensitivity of E2E delay to sensing load, a PID-controlled mechanism is introduced at the SMO layer to dynamically adjust the capacity of a near-real-time sensing slice based on measured delay (age of information). The evaluation assesses its ability to maintain target delay levels under varying sensing rates, background traffic, and latency constraints.

In Section 4.5 we present a Near-RT RIC control framework for multi-RAT indoor 5G deployments, in which non-3GPP sensing context is injected into the RAN control loop through a compact WSI embedded in uplink control signalling. Two cooperating xApps exploit this information: a WSI-driven beamforming xApp that improves link quality using zone-level context, and a CQI + 5QI-aware scheduler xApp that enables QoS-aware resource allocation fully aligned with the 3GPP QoS model. The evaluation assesses the impact of these control functions, individually and jointly, on channel quality, throughput, QoS satisfaction, and fairness.

Overall, this section provides a comprehensive evaluation of the 6G-SENSES control plane components spanning both non-real-time and near-real-time control mechanisms.

4.1 Assessment procedure of RIC compliance with O-RAN standards

4.1.1 Non-RT RIC standards compliance

The evaluation of the Non-RT RIC targets two primary objectives: the functional compliance of the *R1* interface and the computational efficiency of the control loop. This assessment validates the platform's adherence to O-RAN architectural standards for service exposure while ensuring that performance characteristics support effective rApp orchestration and long-term network optimization.

Functionally, we verify the Supported R1 Services across five key domains—SME, DME, RAN Operations And Management (OAM), A1 Policy Management, and AI/ML Workflow—as defined in the O-RAN R1AP specification. It is noted that this Baseline Release targets representative support to validate architectural connectivity rather than full feature coverage. Performance is evaluated via Control Loop Latency, using Tracy Profiler to measure cumulative software processing time against the Non-Real-Time domain requirement (>1 second).

To evaluate the functional compliance of the Non-RT RIC, we utilize an "Audit rApp," a lightweight logical entity designed to validate API behavior against the O-RAN R1AP specification. The Audit rApp performs a systematic verification through three key steps:

1. Bootstrap Enumeration: The rApp queries the Bootstrap API to dynamically retrieve the catalogue of active API entry points, validating that the system supports dynamic service location without hardcoded addressing.
2. Service Discovery Verification: Using the Service Discovery API, the rApp enumerates all registered services to confirm the availability of critical domains, including SME, DME, and A1 Policy Management.
3. Use Case Validation: To verify support for External Sensing and SDN Streaming, the Audit rApp checks that the Data Registration and Data Access endpoints accept specific sensing schemas.

To evaluate the computational efficiency of the control loop, we utilize Tracy Profiler to measure the execution time of distinct software components. By instrumenting critical processing stages—specifically data ingestion, rApp logic execution, and policy formulation—we capture precise latency metrics for each discrete segment. These individual processing times are added together to calculate the total software latency.

4.1.2 Near-RT RIC standards compliance

The evaluation of the Near-Real-Time RIC is driven by three core requirements: functional adherence to the O-RAN E2AP, correct support for E2 SMs, and the high-speed execution of xApp logic. This assessment validates that the controller can correctly terminate E2 interface connections, interpret service-specific RAN definitions, and process events without introducing latency that exceeds the near-real-time budget (10 ms – 1 second).

Functionally, we verify the Supported E2AP Procedures and the Supported E2 SMs. While E2AP manages the connection lifecycle, the E2SMs define the payload structure for specific applications like KPM, RC, CCC, and LLC service model. Performance is evaluated via xApp Loop Latency, using Tracy Profiler to measure the cumulative software processing time from E2 messages.

To evaluate the functional compliance of the Near-RT RIC, we utilize an xApp a reference application designed to exercise the fundamental E2AP message flows and E2SM interpretations defined in O-RAN.WG3.TS.E2AP-R004-v08.00. The Audit xApp performs a systematic verification through three key steps:

- E2 Setup & Service Model Verification: The xApp verifies the E2 Setup Procedure by confirming that the RIC accepts E2 SETUP REQUEST messages. Crucially, it verifies that the RIC correctly parses the RAN Function OID and RAN Function Definition to identify the supported E2SMs (e.g., KPM service model).
- Subscription Management: The xApp validates the RIC Subscription Procedure by successfully sending a RIC SUBSCRIPTION REQUEST containing an E2SM-specific Event Trigger Definition and receiving a valid RIC SUBSCRIPTION RESPONSE.
- Use Case Validation (PUCCH/PUSCH Extensions and Ingestion sensing data at O-DU): xApp tests the data plane integration by verifying the reception of RIC Indication messages (containing E2SM payloads) and the successful transmission of RIC messages.

To evaluate the computational efficiency of the xApp control loop, we utilize Tracy Profiler to measure the execution time of distinct software components. By instrumenting critical processing stages—specifically

ASN.1 decoding (E2AP/E2SM), xApp logic execution, and encoding—we capture precise latency metrics for each discrete segment. These individual processing times are added together to calculate the total software latency. This aggregate value is then validated against the O-RAN Near-Real-Time standard (<1 second) to ensure the system reacts fast enough to RAN events.

Table 4-1 summarizes the results of the xApp verification, including the specific E2 Service Models supported in the BubbleRAN (BR). including the specific E2 SMs supported, enable the PUCCH/PUSCH Extensions and Ingestion sensing data at O-DU (IASA).

Table 4-1 Results of the xApp verification

Component	Standard Reference	O-RAN Spec Ref and custom service model	Supported?
Procedures (E2AP)	E2 Setup	Clause 8.3.1	☑
	RIC Subscription	Clause 8.2.1	☑
	RIC Indication	Clause 8.2.3	☑
	RIC Control	Clause 8.2.4	☑
	E2 Removal	Clause 8.3.7	☑
Service Models (E2SM)	E2SM-KPM (Standard)	TS-E2SM-KPM	☑
	E2SM-RC (Standard)	TS-E2SM-RC	☑
	E2SM-CCC (Standard)	TS-E2SM-CCC	☑
	E2SM-LLC (Standard)	TS-E2SM-LLC	☑

4.2 RIC service to ensure sensing delay requirements over the E2 interface

This subsection builds on the sensing-oriented E2 interface extensions proposed in 2.3 and the observed E2 delay behavior under different sensing loads. With the aim of ensuring the reliability, freshness, and low latency of sensing data over the *E2* interface, a new service has been developed to control the capacity of a near-real-time slice dedicated to sensing.

By gathering statistics from the sensing data, such as the E2E delay over the *E2* interface, which is directly related to the age of information, we propose to control the slice capacity in order to maintain a target delay over time.

For this purpose, we propose to use a Proportional-Integral-Derivative (PID) controller to control the capacity assigned to a slice by obtaining the error *e(t)* between the target delay and the average measured delay *d(t)* at a given instant. The controller then reacts to this error by modifying the resource allocation in real time, combining proportional, integral, and derivative actions to minimise delay variations and stabilise the system performance over time.

$$e(t)=d_{target}\text{-}d(t) \quad (4\text{-}1)$$

In particular, the observation time is divided into discrete slots (Δt). During each slot, the delays of all newly received sensing data are collected, and at the end of each slot, the **average** experienced delay is computed. This average delay is then compared against the **target delay**, allowing the controller to obtain the corresponding **error term** that drives the capacity adjustment process. The controller follows a discrete-time PID formulation, defined as:

$$I(t)=\Delta t \sum_{i=1}^{k} e(t_i)\,,\ D(t)=\frac{e(t)\text{-}e(t\text{-}1)}{\Delta t} \tag{4-2}$$

$$u(t)=K_p e(t)+K_i I(t)+K_d D(t) \tag{4-3}$$

$$r(t+1)=r(t)+u(t) \tag{4-4}$$

Each component of the PID has a specific purpose, expressed in above equations: First, proportional component ($K_p e(t)$) reacts directly to the current delay error. When the measured delay deviates from the target, this term produces an immediate corrective action whose magnitude is proportional to the error. The gain K_p scales the instantaneous deviation between the measured and target delay, determining how aggressively the controller responds to short-term fluctuations. As a result, it provides fast responsiveness and is primarily responsible for reducing short-term deviations.

Second, the integral component ($K_i I(t)$) accounts for the accumulation of past errors, ensuring long-term stability and minimizing steady-state offset, where $I(t)$ represents the accumulated error over time and the gain K_i controls the influence of long-term delay deviations on the control action. Third, derivative component ($K_d D(t)$) predicts future trends of the error by considering its rate of change, thus improving the system's responsiveness and preventing overshooting, where $D(t)$ measures the rate of change of the error between consecutive observation slots and K_d weights this predictive term, allowing the controller to anticipate rapid variations in delay. Finally, the control signal ($u(t)$) is applied incrementally to update the allocated capacity ($r(t)$). This additive update rule allows the controller to perform capacity adjustments, avoiding abrupt changes while continuously steering the system toward the desired delay target.

The controller is deployed in the SMO layer, where it receives delay metrics from the Near-RT RIC through the *A1* interface and allocates the required capacity for the sensing slice via the *O1* interface, as illustrated in Figure 4-1.

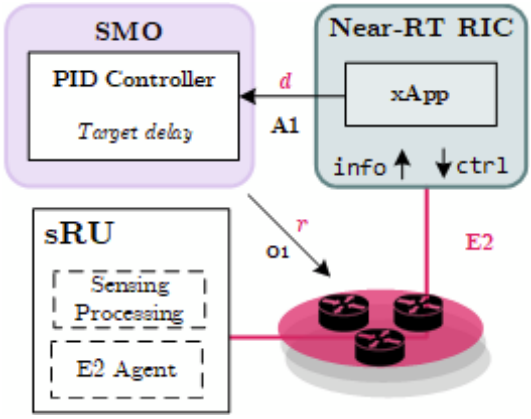


Figure 4-1 O-RAN architecture integrating sensing devices and a RIC service to guarantee sensing delay requirements.

Figure 4-2 shows the evolution of the transmission rate configured by the PID controller, represented in red, together with the delay experienced by the new sensing information for the target flow. In this experiment, the sensing data are generated with a refresh rate of 50 Hz and a heatmap size of 32 KB.

During the first 20 seconds, only the target flow is active. Starting from an initial rate of 10 Mbps, the PID controller progressively adjusts the rate until the delay requirement, indicated by the yellow line, is satisfied.

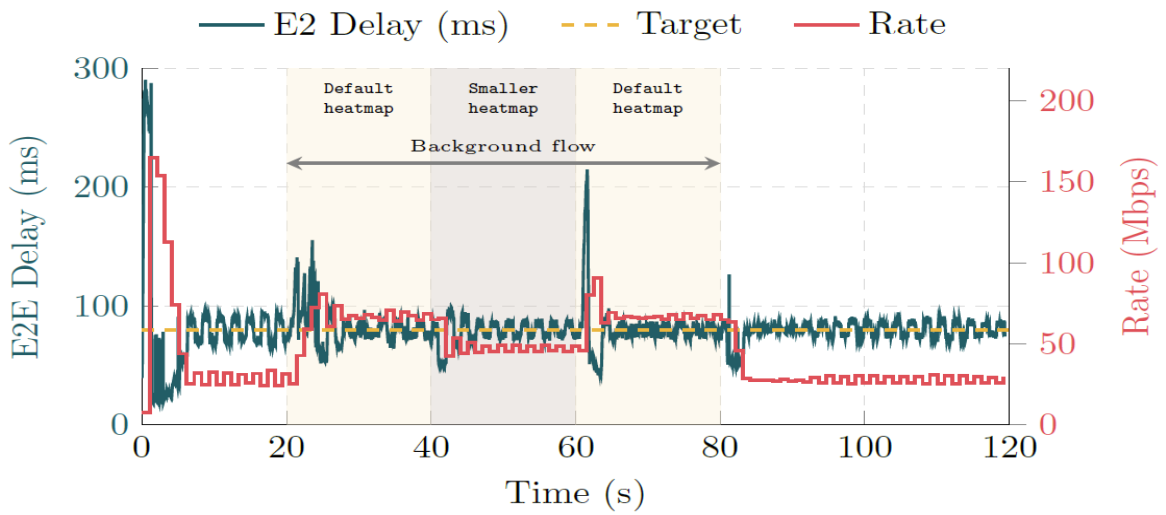


Figure 4-2 Evolution of the rate selected by the PID controller and the delay experienced by the sensing information for the target flow.

In the next 20 seconds, a background flow that is not monitored by the controller is introduced in the same slice. This period is represented by the shaded region in Figure 4-2. The background flow transmits with a refresh rate of 100 Hz and the default heatmap size, temporarily affecting the system delay. The PID controller initially exhibits transient oscillations before converging to a new stable rate that restores the delay to its target value.

During the following 20 seconds, the background flow reduces its heatmap resolution, shown in brown, which decreases its transmission rate. The PID controller reacts accordingly, adapting the allocated rate for the target flow while keeping the delay close to the desired value. Finally, once the background flow stops, the controller restores the rate to its original value, demonstrating the stability and responsiveness of the proposed control mechanism.

Figure 4-3 presents the evaluation of the PID controller when the target delay, represented by the yellow line, varies during the observation period. This scenario emulates the introduction of new services with stricter latency requirements within a specific slice.

As shown in the figure, the controller dynamically adjusts the capacity assigned to the slice according to the target delay. When the target delay decreases, the instantaneous error becomes negative, which causes the proportional term of the controller to increase the control action. This results in a higher allocated capacity, allowing the measured delay to remain close to the target. Conversely, when the target delay increases, the error magnitude becomes smaller, and the control action reduces the allocated capacity accordingly.

It can also be observed that oscillations around the target value become more noticeable when the delay requirement is relaxed. This behavior is mainly due to the looser constraint and the relatively high proportional gain of the controller. Overall, the results highlight the intrinsic trade-off between responsiveness and stability in PID-based delay control mechanisms.

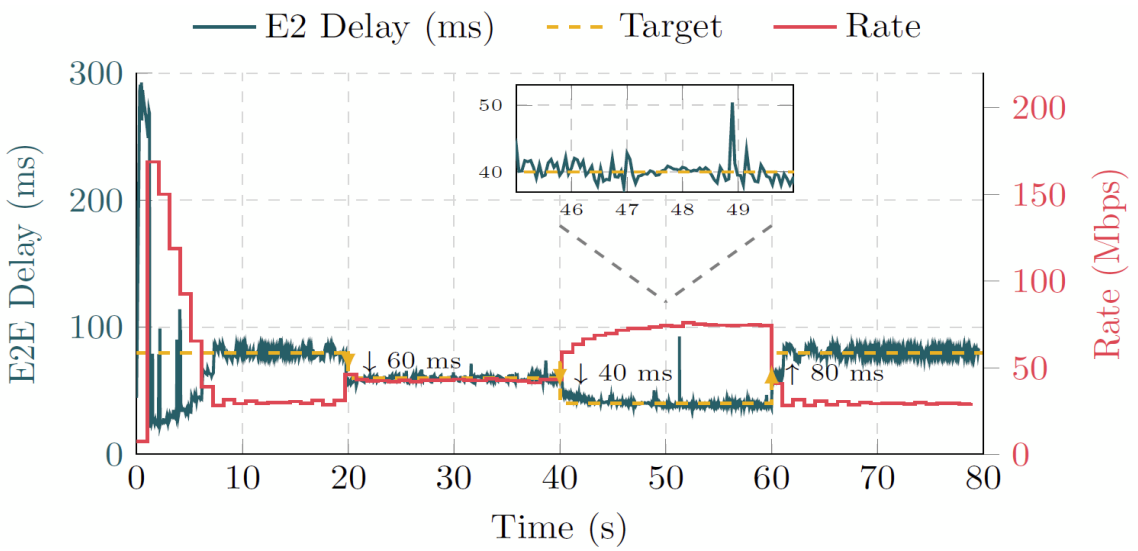


Figure 4-3 Capacity assigned by the PID controller (red) and resulting delay (green) as the target delay (yellow line) varies over time.

4.3 External sensing data fusion

We evaluate the process of ingesting sensing data through the SMO in the 6G-SENSES system, focusing on how sensing information is handled across the non-RT and near-RT domains. The analysis is intentionally split into two parts in order to reflect the architectural separation of control loops. The first part examines the path from the external sensing source towards the *A1* northbound interface, covering sensing ingestion, registration, filtering, aggregation, and exposure within the SMO and Non-RT RIC. The second part analyzes the downstream propagation of sensing information from the *A1* southbound interface to the xApp, including A1 mediation, fast message routing, and delivery to near-RT applications. This decomposition allows a clear assessment of how sensing data is transformed and transferred from the slow control loop in the non-RT domain to actionable context in the Near-RT RIC.

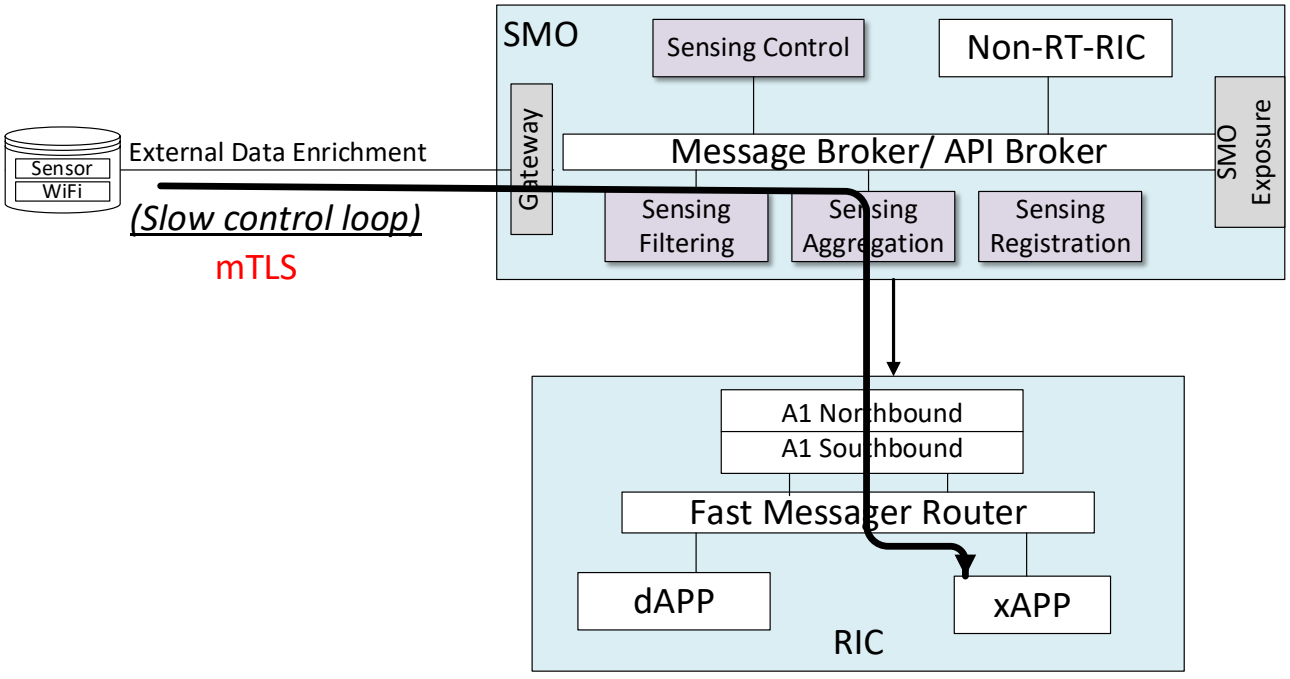


Figure 4-4 Evaluation of external sensing data ingestion through the Non-RT RIC

Figure 4-4 shows the datapath of external sensing data ingestion terminated to an xApp. External sensing data is first terminated at the SMO gateway, where secure connectivity and protocol normalization are enforced before the data enters the Non-RT RIC platform. Within this domain, sensing control functions coordinate sensing registration, filtering, and aggregation components that operate over long-time windows and interact through a message broker / API broker. This enables scalable, decoupled processing and allows rApps in the non-RT RIC to transform raw sensing inputs into structured enrichment information or intent. The *A1* interface acts as the architectural demarcation between the slow and near-real-time domains: *A1* northbound exposes sensing-derived information within the non-RT RIC, while *A1* southbound delivers validated and schema-aligned outputs to the Near-RT RIC. Inside the Near-RT RIC, an A1 mediation layer adapts these inputs to internal message formats and injects them into the fast message router, which ensures low-latency distribution to subscribing xApps. The xApps consume this non-RT sensing context alongside near-RT measurements and, when required, translate it into E2-based control actions toward the RAN. Overall, the architecture ensures modularity, scalability, and strict separation of concerns while enabling sensing-driven intelligence to flow efficiently from the SMO to near-real-time RAN control.

4.3.1 Sensor to Northbound A1

To benchmark the Non-RT RIC in terms of sensing ingestion the following process is executed:

Step 0 – Register sensing Information type (sensing payload schema).

Step 1 – Register sensing Producer (sensor device id, IP address and API endpoints).

Step 2 – Create sensing task (sensing pipeline include source and destination pair).

Step 3 – Instantiate sensing data delivery path.

Once the job is active, the producer begins delivering data. It sends a data sample to SeCF. This sample is specific to the implementation and follows the structure defined by the previously registered schema. Upon receiving the data, SeCF activates again, validates the sample against the schema, identifies which jobs are currently active and relevant for the data, and forwards the payload to the RIC.

Figure 4-5 depicts the total time required by SeCF to create *a set N of* producer–job pairs as a function of *N*, with error bars representing the standard deviation across multiple runs.

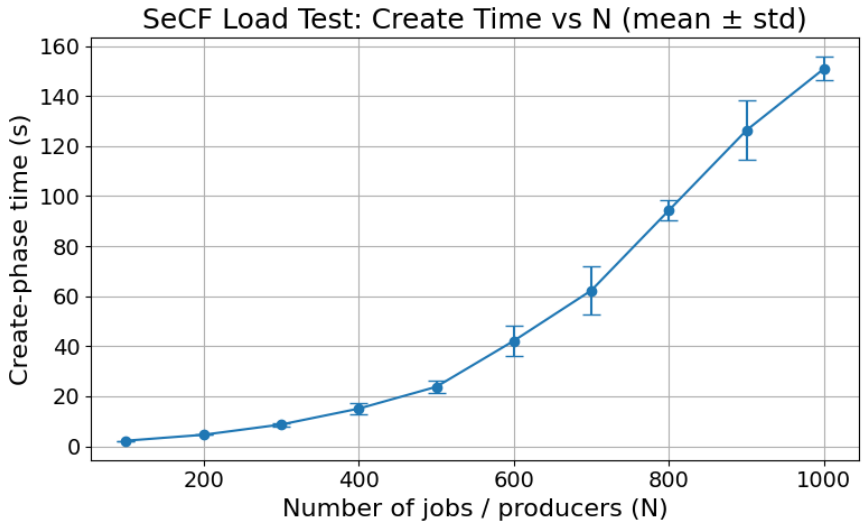


Figure 4-5 SeCF Create-phase Time vs Number of sensing Jobs registered (mean ± std)

For relatively small size of sensing tasks, up to approximately N=300, the create-phase execution time grows in an almost linear manner with increasing N, while the observed variance remains low, indicating stable and predictable behavior. A clear knee point emerges around N≈600 jobs, beyond which the create-phase time begins to increase in a super-linear fashion. After this point, the standard deviation rises markedly, revealing increased run-to-run variability and reduced stability. At N=1000, both the mean create-phase time and its deviation are significantly higher, clearly reflecting unstable system behavior when operating under heavy load.

These results indicate that SeCF operates in a stable regime up to a moderate number of concurrent jobs. Beyond the knee point, internal contention and queueing effects dominate, leading to both increased latency and variability. The growing deviation suggests sensitivity to scheduling, internal locking, and persistence overheads under high control-plane load.

Figure 4-6 presents the effective create-phase throughput (requests per second) as a function of *N*, again with mean and standard deviation across runs. As before, we observe that

- Throughput is high and stable for low values of *N*.
- As *N* increases beyond ~300, throughput begins to degrade.
- The largest variance is observed in the intermediate region (*N ≈ 300–500*), corresponding to the transition from stable to saturated operation.
- At large *N* (≥ 800), throughput collapses to a low but stable value, and variance decreases again.

The throughput behavior reflects classical saturation dynamics. In the transition region, SeCF alternates between productive processing and back-pressure, resulting in higher variability. Once fully saturated, the system converges to a consistently low throughput, indicating that the control plane has reached its capacity limit.

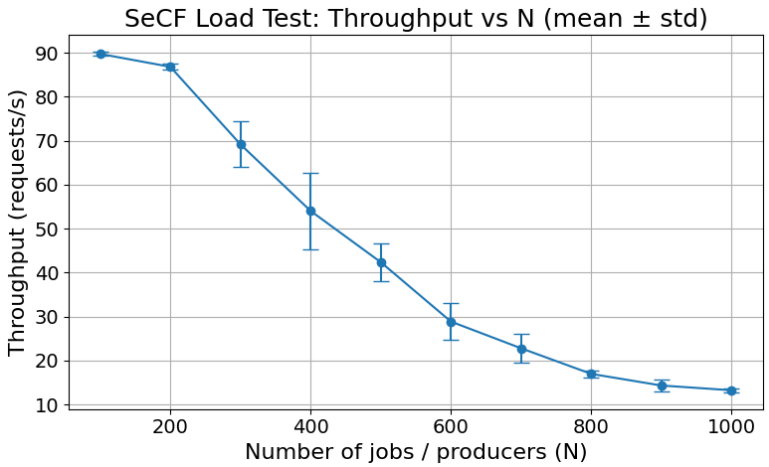


Figure 4-6 SeCF Create-phase Throughput vs Number of Jobs (mean ± std)

Figure 4-7 compares the average latency of **producer registration** and **job creation** as a function of *N*. We measure the average time needed for a sensor to send its measurements as well as the total time required for a sensing task to complete and delivered to its destination. We observe that:

- Producer registration latency remains consistently low (on the order of a few milliseconds) across all values of *N* indicating that its performance is fully deterministic and not affected by the increased number of jobs pushed into the system.
- Job creation latency is significantly higher—by roughly two orders of magnitude—even at small *N*.
- Job creation latency increases steadily with *N*, reaching several seconds at large system sizes.

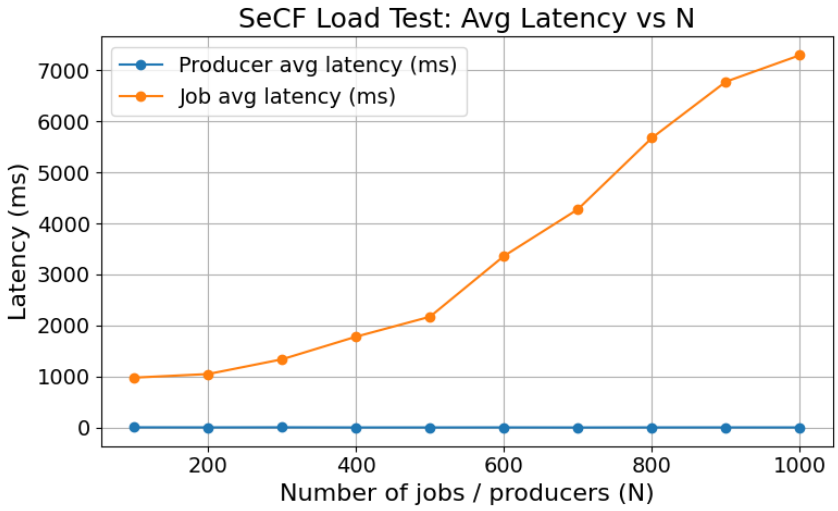


Figure 4-7 SeCF Average Latency vs Number of Jobs (Producer vs Job)

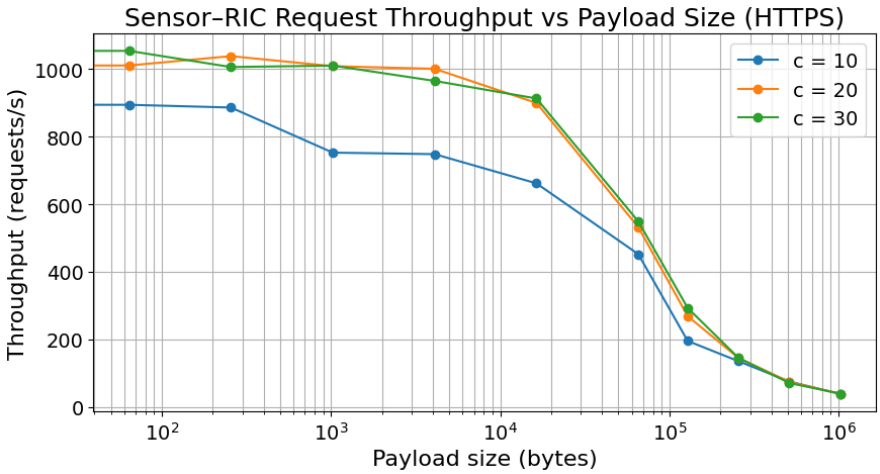


Figure 4-8 Sensor–RIC Throughput vs Payload Size (HTTPS, c = 10 / 20 / 30)

The large disparity between producer and job latency highlights the fundamentally different nature of these operations. **Producer** registration is a lightweight metadata operation, whereas job creation triggers orchestration logic, including producer matching, lifecycle initialization, scheduling, and persistence. The increasing job latency with *N* confirms that job creation is the dominant scalability bottleneck in SeCF.

Figure 4-8 shows the request-level throughput achieved between sensors and the RIC data-delivery endpoint over HTTPS for different payload sizes and concurrent levels (10, 20, and 30 concurrent connections). For small payloads, higher concurrency significantly increases request throughput, reaching approximately 1k requests/s. As payload size increases, throughput decreases smoothly and converges across all concurrency levels, indicating payload-dominated processing costs. Therefore, concurrency improves throughput for small payloads, while payload size dominates performance for large messages.

Figure 4-9 illustrates the aggregate data throughput successfully delivered from sensors to the RIC as a function of payload size. Independent of the number of concurrent connections, aggregate throughput increases with payload size and saturates at approximately 35–38 MB/s. This plateau indicates a processing or protocol bottleneck in the HTTPS ingress path rather than network bandwidth limitations.

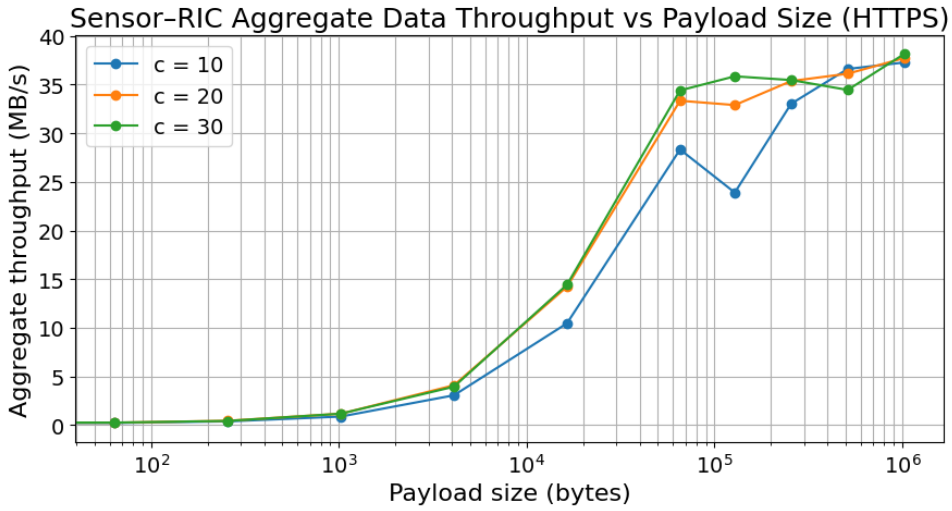


Figure 4-9 Sensor–RIC Aggregate Throughput vs Payload Size (HTTPS)

Finally, Figure 4-10 presents the average E2E latency experienced by sensor data messages delivered to the RIC for varying payload sizes and concurrency levels. Latency increases monotonically with payload size and is higher for larger concurrency values due to queueing effects. No abrupt latency spikes or instability are observed, indicating predictable and stable ingress behavior.

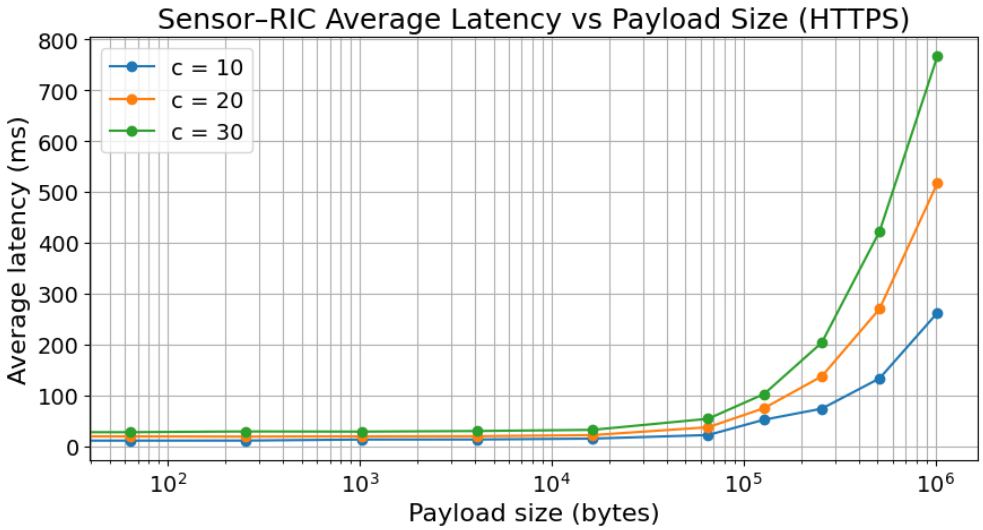


Figure 4-10: Sensor–RIC Average Latency vs Payload Size (HTTPS, c = 10 / 20 / 30)

4.3.2 Southbound A1-S to xApp

The measured process corresponds to the southbound realization of the *A1* interface between the non-RT RIC A1 Mediator and an xApp, implemented using a Fast message router. The following steps have been included in the evaluation process:

1. Policy update Initiation at the non-RT RIC: The process begins when a policy update request is received by the non-RT RIC A1 Mediator (e.g., via the northbound A1-P REST API). At this stage, the mediator validates the request, parses the policy payload, and performs internal bookkeeping, including updating the policy instance state and associated metadata.

2. Southbound policy dispatch via the fast message router: Once the policy update is validated and stored, the A1 Mediator dispatches the policy to the target xApp using the southbound A1 communication path. This

communication is realized over the fast message router which provides asynchronous, message-based transport between RIC components.

3. Policy reception and processing at the xApp: The xApp receives the policy message through the message router. Upon reception, the xApp decodes the message payload, extracts the policy parameters, and applies the policy according to its internal logic. In the evaluated setup, this processing is lightweight and does not involve additional external dependencies.

4. Southbound response reception at the A1 Mediator: The A1 Mediator asynchronously receives the response message from the xApp. Upon reception, it correlates the response with the corresponding policy instance using the policy identifiers and updates the internal policy state accordingly. The reception of this response marks the Round-Trip Time (RTT) measurement end point, as it signifies the completion of the southbound A1 interaction for the given policy update.

5. Status propagation and completion: Following successful response handling, the mediator triggers internal status update mechanisms and, if applicable, policy status notifications. These operations are performed after the RTT measurement window and therefore do not affect the measured latency.

The reported RTT captures only the southbound A1 Mediator–xApp interaction**, including:**

- A1 Mediator internal processing related to policy dispatch
- Fast message router serialization and transport
- xApp policy handling and response generation

As a result, the measurement isolates the **intrinsic latency of the southbound A1 control loop** within the Non-RT RIC.

Under parallel policy update load, multiple policy dispatch and response flows are handled concurrently. The asynchronous nature of the fast message router and the concurrent execution model of the A1 Mediator and xApp result in interleaved log entries. RTT measurements are therefore correlated per policy instance rather than by sequential log order.

Despite concurrent execution, the measured RTT distribution demonstrates consistently low latency with a narrow spread, indicating that the southbound A1 path remains efficient and stable even under increased load.

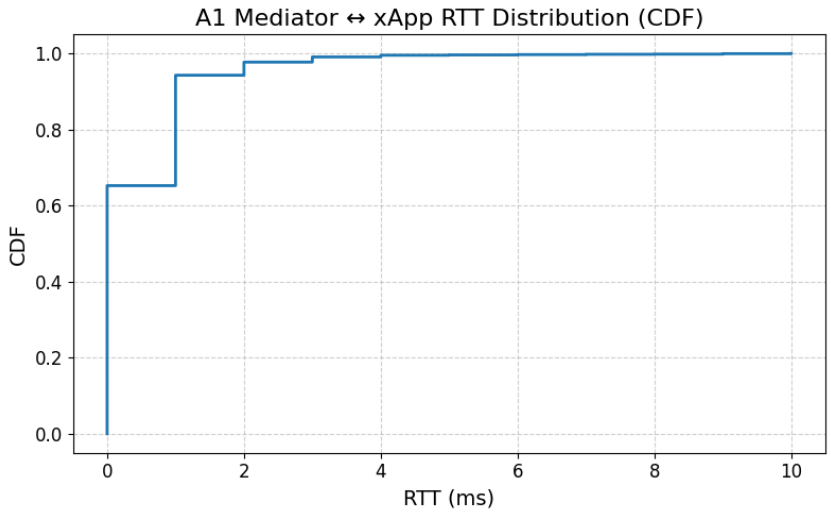


Figure 4-11 CDF of RTT latency in Near-RT RIC

Under increased parallel load, the A1 Mediator–xApp round-trip time remains below 1 ms for 94.3% of requests and below 3 ms for 99.1% of requests, with an average RTT of 0.451 ms. Only 0.37% of requests exceed 6 ms, with a single worst-case outlier at 10 ms.

4.4 SDN controller for sensing streams

The *O1* interface, as defined by the O-RAN Alliance, provides standardized Management Plane (M-plane) connectivity between the SMO framework and O-RAN network elements. NETCONF is used as the underlying management protocol on *O1* to exchange configuration and state information encoded using Yet Another Next Generation (YANG) data models.

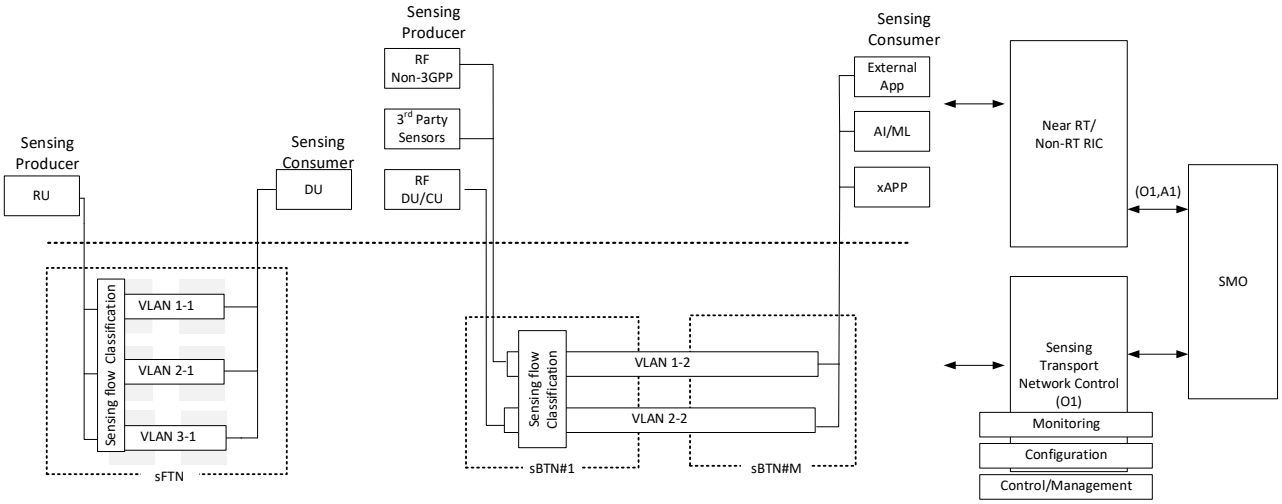


Figure 4-12: SDN Control for Sensing Flows

To establish QoS against remote sensing service consumers and producers VLAN connections are used providing connectivity between RUs generated sensing traffic destinated to DUs over Fronthaul Transport Networks (FTN) and sensing producers (i.e., RF based sensing from Non-3GPP networks, 3rd party sensors, etc.) transferred to sensing consumers over Backhaul Transport Networks (BTN). In our evaluations we consider the scenario of configuring the pipeline between RU-DU to transfer IQ streams using the O1/NETCONF protocol. O1/NETCONF is used exclusively during the configuration phase to instantiate and activate the sensing RU context and its associated VLAN. At runtime, IQ U-plane streams are transported over Ethernet and separated at the DU using VLAN identifiers, enabling per-operator isolation. The analysis covers the part of network configuration to establish a) the required connectivity between RUs and DUs (m-plane) and b) to transfer sensing data to the DU over u-plane.

Figure 4-13 shows the execution latency of representative O1/NETCONF configuration operations on the O-RU. “Create Sensing RU” corresponds to the provisioning of a sensing-specific RU configuration, while “Activate Sensing RU” reflects the runtime activation of the corresponding data-plane functionality. Error bars denote the 95% confidence interval computed across repeated executions. O1 configuration operations that only update existing parameters exhibit low and stable latency, whereas operations that instantiate, activate, or structurally modify RU resources incur higher execution times due to additional internal validation and data-plane reconfiguration.

The **Change Gain** operation exhibits the lowest average latency. This operation corresponds to a partial update of an existing YANG data node within the *tx-array-carriers* container using a NETCONF PATCH operation. Internally, this update does not trigger structural changes in the RU configuration nor require re-initialization of PHY resources. As a result, processing is limited to parameter validation and propagation to the active transmission chain, leading to consistently low and stable execution times with minimal variance.

The **Create Sensing RU** step shows slightly higher latency compared to gain updates but remains relatively low overall. This operation instantiates a new sensing-specific RU context within the O-RU using the *tx-array-carriers* container in the *o-ran-uplane-conf* model. Although this involves object creation and validation, the RU remains in an inactive state, and no data-plane resources are yet allocated. Consequently, the operation primarily affects the management plane and internal configuration datastore, resulting in modest execution latency and low variability.

The **Create VLAN** operation demonstrates moderate latency and higher variance. This step configures or updates a VLAN-enabled Ethernet interface via the *ietf-interfaces* model. In addition to updating YANG configuration data, this operation typically requires interaction with the RU's networking subsystem, including validation of interface state and potential reconfiguration of Ethernet forwarding resources. Depending on the RU implementation, these actions may involve kernel-level or firmware-level operations, which explains both the increased latency and the observed variability across runs.

The **Change Bandwidth** operation is among the most time-consuming and exhibits significant variance. This is expected, as bandwidth changes affect fundamental PHY parameters and may require reallocation of internal buffers, reconfiguration of numerology, or recalculation of scheduling and filtering parameters. In some RU implementations, such changes can temporarily disrupt active transmission paths or require synchronization across multiple internal components, leading to higher and less predictable execution times.

The **Activate Sensing RU** operation consistently exhibits the highest average latency. Activation transitions the RU context from a configured but inactive state into an operational state, enabling the generation of U-plane I/Q samples. This step typically triggers multiple internal actions, including validation of all configured parameters, allocation of PHY and fronthaul resources, and synchronization with timing and transmission subsystems. Because this operation directly impacts the runtime data plane, it is inherently more complex than pure configuration updates, resulting in higher latency compared to other steps.

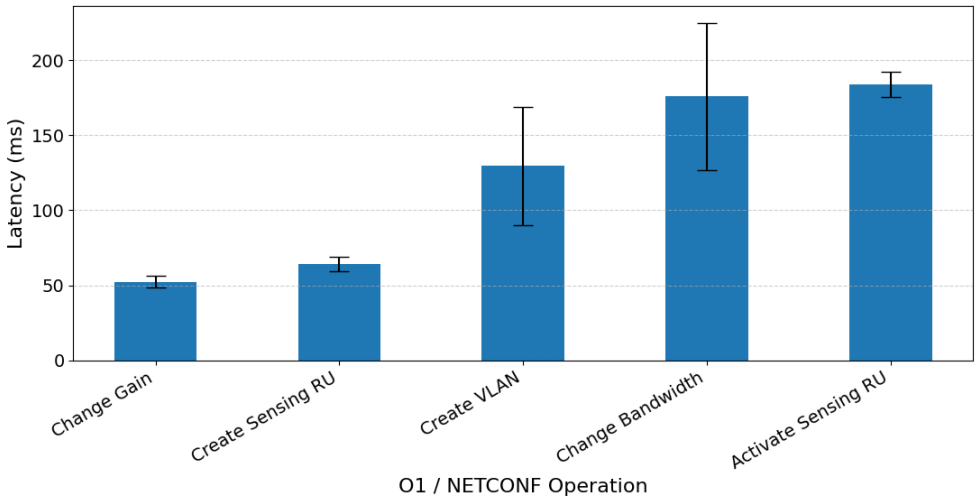


Figure 4-13 O-RU O1 sensing configuration latency (mean ± 95% CI)

Overall, Lower latencies are observed for operations that modify existing configuration parameters without affecting the RU's structural or data-plane state (e.g., gain updates). In contrast, higher latencies and increased variance are associated with operations that require reconfiguration of networking interfaces, PHY resources, or runtime activation of processing chains. These observations are consistent with the separation between management-plane configuration and data-plane activation defined in the O-RAN WG4 architecture.

4.5 Wi-Fi-Assisted, QoS-Aware Control via xApps and PUCCH extension

This section focuses on the design and evaluation of a Wi-Fi-assisted, QoS-aware control framework for beyond-5G communication networks. The framework follows the O-RAN architectural principles and is realized as a set of xApps running on a Near-RT RIC. The goal is to show how non-3GPP sensing and 3GPP-native QoS information can be jointly used in a single control loop to improve channel quality and resource allocation, while preserving conventional performance indicators such as throughput and fairness.

In many realistic deployments, 5G does not live in isolation. Private and campus networks are typically multi-RAT environments, where a 5G gNB and one or more Wi-Fi access points cover the same indoor area. At the same time, emerging applications, such as industrial control, XR and DTs, rely on flows with very different QoS requirements, expressed through the 5QI. Some flows have stringent latency and reliability bounds, while others are best-effort. The RAN must handle this diversity and do so in a way that can evolve with O-RAN and Near-RT RIC control.

Today's 5G schedulers see mainly 3GPP measurements such as CQI and Signal-to-Interference-plus-Noise Ratio (SINR), and they enforce QoS using 5QI inside the gNB. What they typically do not see is the rich information available from Wi-Fi and other non-3GPP sources. At the same time, most hybrid 5G/Wi-Fi positioning or sensing schemes keep the non-3GPP information outside the RAN control loop or rely on additional integration elements such as N3IWF agents. This makes it hard to bring Wi-Fi-assisted intelligence into the O-RAN Near-RT RIC without adding complexity. By moving part of the decision logic into the Near-RT RIC, O-RAN gives us exactly the right place to connect these pieces: xApps can subscribe to detailed telemetry, including uplink control information, and steer beams and scheduling on short time scales.

In this work we leverage that flexibility to combine:

- Wi-Fi-assisted beamforming, based on a compact WSI signaled via PUCCH, and
- CQI + 5QI–aware scheduling, running as an xApp on the Near-RT RIC and fully aligned with the 3GPP QoS model.

WSI is derived from a hybrid fingerprinting method that combines signal strength from both 5G and Wi-Fi access networks, and is embedded into PUCCH using the same CRC and Polar-coding rules we have previously validated for additional control indicators.

Context and motivation

The scenario we consider is deliberately simple but representative of many private and campus 5G deployments. We focus on an indoor environment where:

- a 5G gNB, implemented as an O-RAN-compliant O-CU and O-DU, serves one or more UEs.
- at least one Wi-Fi access point operates in the same area and is visible to the UEs.
- the UEs carry multiple traffic flows mapped to different 5QI classes, ranging from low-latency control traffic to less critical background services.

At first, Wi-Fi measurements are both easy to obtain and highly informative. A UE can periodically scan for Wi-Fi networks and collect RSSI/SNR from nearby APs with minimal overhead. When combined with 5G measurements such as RSRP and SINR, these values form hybrid fingerprints that correlate strongly with the UE's position and with the propagation conditions seen by the gNB. Our previous work [36] has shown that a hybrid fingerprinting method that combines signal strength from both 3GPP and non-3GPP networks can significantly improve indoor positioning accuracy on a private 5G testbed with integrated Wi-Fi access points.

Second, 5QI is already the standard way to express QoS demands in 5G. The 5G QoS Identifier encapsulates delay, reliability and priority characteristics for each QoS Flow. It is natively available at the O-CU-CP and can

be propagated towards the DU and, via O-RAN's E2 interface, towards Near-RT RIC applications. This means that a scheduler xApp can make QoS-aware decisions using 5QI directly, without introducing any new proprietary "priority" indicators.

Taken together, these two aspects suggest a natural control loop:

- use Wi-Fi-assisted sensing to infer where a UE is and what kind of radio environment it sees, and compress that into a small WSI.
- send WSI to the gNB using PUCCH, alongside CSI/CQI, in a way that respects 3GPP payload and coding rules.
- let Near-RT RIC xApps use WSI to improve beamforming (and, therefore, CQI) and use CQI + 5QI together to drive QoS-aware scheduling.

Problem statement and objectives

The problem we address can be stated in three parts. First, the RAN needs a practical way to ingest Wi-Fi-based sensing information into the O-RAN control loop. We do not want to modify Wi-Fi access points or introduce new non-3GPP control anchors. Instead, we want to reuse uplink control signaling mechanisms that already exist in 5G NR, and that are designed for frequent, reliable, small-payload messages.

Second, we want to use this sensing information to improve beamforming. Indoor environments are often dominated by multipath, blockage and fast changes in user position. A beam that performs well in one corridor or zone may be sub-optimal in another. If the Near-RT RIC receives compact WSI values that encode the UE's hybrid 5G/Wi-Fi context, a beamforming xApp can map those to preferred beams or beam patterns and update the configuration in closed loop.

Third, once beamforming has improved CQI for selected UEs, the RAN should be able to schedule radio resources in a QoS-aware way using standard 5QI, without changing the NG-RAN QoS model. In other words, a scheduler xApp should use:

- CQI to understand what is radio-feasible, and
- 5QI to understand which flows are more delay- or loss-sensitive,
- and then compute per-UE or per-flow scheduling weights that reflect both.

To address this, we pursue the following concrete objectives:

- Define and implement a WSI per UE, derived from hybrid 5G/Wi-Fi fingerprints, and encoded as a small number of bits suitable for inclusion in PUCCH.
- Specify how WSI is multiplexed into PUCCH, reusing established 3GPP rules for CRC length, channel coding (small block vs Polar) and PUCCH format selection, as previously validated for similar indicators.
- Design a WSI-driven beamforming xApp that consumes WSI and conventional CSI/CQI at the Near-RT RIC, and issues beam or beamweight hints to the O-DU/O-RU to improve the CQI of selected UEs, especially in challenging indoor zones.
- Design a CQI + 5QI scheduler xApp that uses CQI and 5QI to compute per-UE scheduling weights, enabling QoS-aware resource allocation in an O-RAN-compliant manner, without introducing non-standard QoS parameters.
- Evaluate a set of scenarios that emulate different WSI configurations and 5QI mixes, and quantify the impact on CQI distribution, throughput, QoS satisfaction and fairness.

Architectural approach

The architectural approach is modular and follows the same "xApps on Near-RT RIC" design pattern used in our previous work, but with a different focus. We deploy two xApps that work together on multi-RAT sensing and QoS-aware scheduling:

The WSI-driven beamforming xApp consumes radio measurements such as CQI and CSI, as well as WSI values decoded at the O-DU from Uplink Control Information (UCI) carried on PUCCH. It interprets each WSI as a compact description of the UE's hybrid 5G/Wi-Fi context (typically a zone or cluster ID derived from a hybrid fingerprinting model) and maps it to preferred beam configurations. By updating the beam selection or beamweights in a closed loop, it improves CQI for UEs in difficult indoor areas, without requiring any changes to the PHY layer or the Wi-Fi infrastructure.

The CQI + 5QI scheduler xApp receives CQI reports, basic traffic statistics and QoS context from the O-DU/O-CU. From these inputs it computes per-UE scheduling weights that steer the existing MAC scheduler. In simple terms, CQI tells the xApp "how good the channel is", while 5QI tells it "how strict the QoS is". The xApp combines the two to prioritize flows with tighter delay and reliability budgets when radio conditions allow, while avoiding excessive resource allocation to UEs with persistently poor CQI.

The two xApps sit on top of an O-CU, O-DU and O-RU stack, and they fit naturally into the existing Non-RT and Near-RT RIC framework. The following sections will build on this picture and describe, in turn, the O-RAN architecture, the WSI design and embedding, the xApp logic and the evaluation scenarios.

A key bottleneck is not the learning algorithm itself, but the path by which Wi-Fi sensing becomes available to the near-RT control loop. If Wi-Fi sensing only arrives through heavier user-plane uploads or external pipelines, it becomes difficult to use it as a tight control input. Our work explores a pragmatic alternative: represent Wi-Fi sensing as a compact WSI and carry it in UCI over PUCCH, so that WSI can be exposed to near-RT applications with low latency and minimal integration overhead. A dedicated section later in the document details the WSI-over-PUCCH mapping.

Scope and assumptions

To frame the rest of the section, we state the scope and assumptions of this work.

Our work explains how the WSI is derived from hybrid 5G/Wi-Fi measurements and how it is carried within UCI on PUCCH so that it can reach the O-DU and, from there, the Near-RT RIC. The rest of the control loop—Near-RT RIC operation, xApp lifecycle, CQI feedback, and scheduling influence via E2—follows standard O-RAN design practices and existing implementations.

WSI is explicitly treated as a micro-indicator, encoded in a small number of bits (e.g., 2–8 bits) suitable for inclusion in UCI. Raw Wi-Fi measurement vectors (RSSI per AP, full fingerprints, CSI samples) are not transported over PUCCH and are only used locally (UE side or offline) to derive the compact WSI value.

In the indoor scenario we consider, the UE participates in a hybrid fingerprinting scheme that combines signal measurements from both 5G and Wi-Fi access networks and maps them to discrete WSI values (e.g., zone or position identifiers derived from radio measurements). The underlying 5G/Wi-Fi data acquisition, radio-map construction and classification steps are described later in this document, where we adopt a private 5G testbed with integrated Wi-Fi access points and implement the fingerprinting algorithm as an xApp on the Near-RT RIC.

The WSI-over-PUCCH mechanism is not currently standardized. It is presented as a proof-of-concept, standards-informed design that reuses 5G NR control-channel building blocks (CRC attachment, Polar coding and PUCCH formats) to carry a small, non-3GPP indicator from the UE to the gNB. No changes to the 3GPP

specifications are made, but instead we explore how far we can go by staying within existing degrees of freedom.

Terminology note:

In this document, WSI stands for Wi-Fi Sensing Indicator. It denotes a compact, Wi-Fi-assisted context or zone indicator used for beamforming and RAN control. This meaning is specific to our work and should not be confused with other uses of the abbreviation "WSI" in the broader literature.

4.5.1 Signal Paths and Telemetry Exposure of WSI, CQI, and 5QI

This subsection explains how our proposal fits into a realistic O-RAN deployment. The aim is to make clear where WSI, CQI and 5QI live in the system, and how they reach the Near-RT RIC.

Overall architectural view

We assume a conventional disaggregated O-RAN system (Figure 4-14). For the purposes of this section, we assume a configuration in which the O-DU is the primary E2 node, because this is where CSI/CQI and UCI decoding naturally live, and thus where WSI can be extracted together with standard control information. The O-CU provides QoS context (QFI/5QI, DRB associations) that can be surfaced towards the RIC through the same telemetry path.

The two xApps introduced earlier are embedded in this architecture as follows:

- The WSI-driven beamforming xApp subscribes to UE-level measurements coming from the O-DU, including CQI and the WSI values decoded from uplink control signalling. It processes these inputs and sends back beam-related hints (e.g., preferred beams or beam weights per UE) that the O-DU applies when configuring downlink transmissions.
- The CQI + 5QI scheduler xApp subscribes to CQI reports, traffic and buffer statistics, and QoS-flow context that includes the 5QI. Based on these inputs, it computes per-UE scheduling weights or policies that bias the O-DU scheduler towards QoS-aware decisions without replacing its internal scheduling logic.

Our contribution is confined to the way Wi-Fi sensing information is injected into the uplink control path and how Near-RT RIC applications consume radio and QoS context to refine beamforming and scheduling.

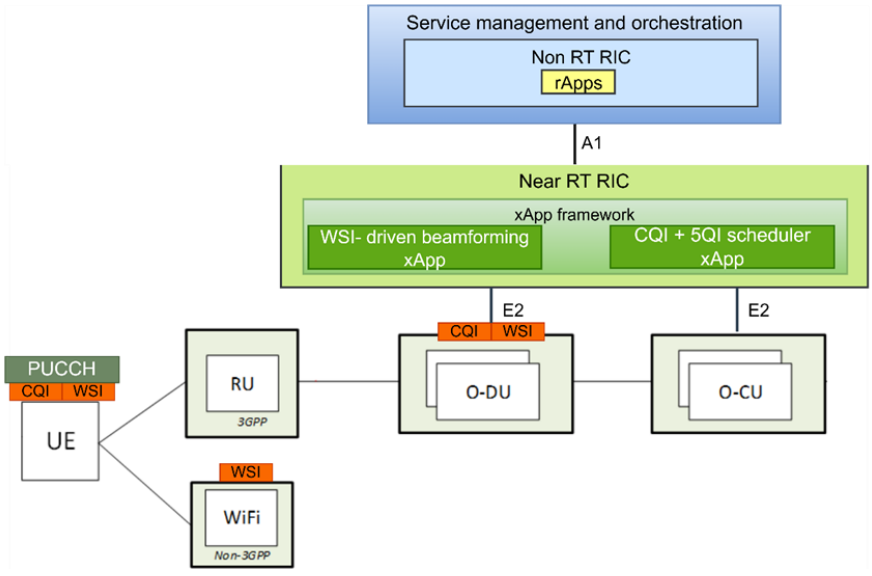


Figure 4-14 O-RAN- architecture with Near-RT RIC xApps (WSI- driven beamforming xApp, CQI + 5QI scheduler xApp).

Where 5QI actually lives

Underneath the O-RAN control framework, the radio access network follows the 3GPP NG-RAN architecture. In this model, 5QI is a property of QoS Flows, not an O-RAN-specific parameter. As discussed in [37], 5G introduced a new QoS architecture based on PDU Sessions and QoS Flows: applications generate Service Data Flows (SDFs) that are mapped to 5G QoS Flows, each identified by a QoS Flow Identifier (QFI) and associated with a standardized 5QI value capturing delay budget, packet error rate and priority characteristics. The QoS Flow is the finest granularity of QoS differentiation within a PDU Session, and flows sharing the same QFI/5QI receive the same forwarding treatment (e.g., scheduling and admission thresholds).

The 5G Core assigns a 5QI to each QoS Flow as part of the QoS profile, and this information is delivered to the gNB via NGAP messages over N2. The CU-CP receives the QoS Flow descriptions, stores the mapping between QFI, 5QI and their QoS characteristics, and configures the CU-UP and DU accordingly. Each DRB is thus associated with one or more QoS Flows and their underlying 5QI values [37].

In our testbed, we exploit the fact that this QoS Flow context is already available inside the RAN: at the CU-CP, as part of the NGAP signaling state, and at the DU, via the DRB and QoS configuration pushed from the CU. We deliberately do not assume that O-RAN "natively" provides 5QI to the Near-RT RIC. Instead, the RAN is instrumented so that the 5QI information contained in the QoS Flow context is retrieved at the CU/DU and added to the telemetry exported to the Near-RT RIC. From the scheduler xApp's perspective, 5QI appears as an attribute attached to each UE or flow in the measurement reports it subscribes to.

E2-Based Telemetry Enrichment and Exposure to Near-RT RIC xApps

The Near-RT RIC terminates the *E2* interface from the E2 node (O-DU/O-CU/gNB) and exposes subscription, reporting, and control hooks to xApps via E2AP. xApps follow the usual pattern: subscribe to specific indications, receive periodic (or filtered) measurement updates over E2SM-KPM, and, when needed, send E2SM-RC control messages to adjust RAN behavior.

We slightly enrich the telemetry exported by the E2 node:

- UE-level CSI/CQI (already used for link adaptation and commonly available via KPM) is made visible to both xApps.
- WSI decoded at the O-DU from UCI on PUCCH, is attached to those UE-level reports so the beamforming xApp sees the latest per-UE WSI value.

The scheduler xApp's QoS context is augmented with the 5QI associated with each DRB / QoS flow aggregate. This does not introduce a new protocol: we surface the existing 5QI from QoS Flow information known at the CU/DU into the RIC-exported data structures.

All of this can be realized with existing service models (e.g., by adding vendor-specific measurement items/labels within E2SM-KPM or through a lightweight custom E2SM. So finally, what we achieve is to make standard 5QI visible to the scheduler xApp and carry WSI as an additional (Wi-Fi–assisted) measurement alongside CQI/CSI.

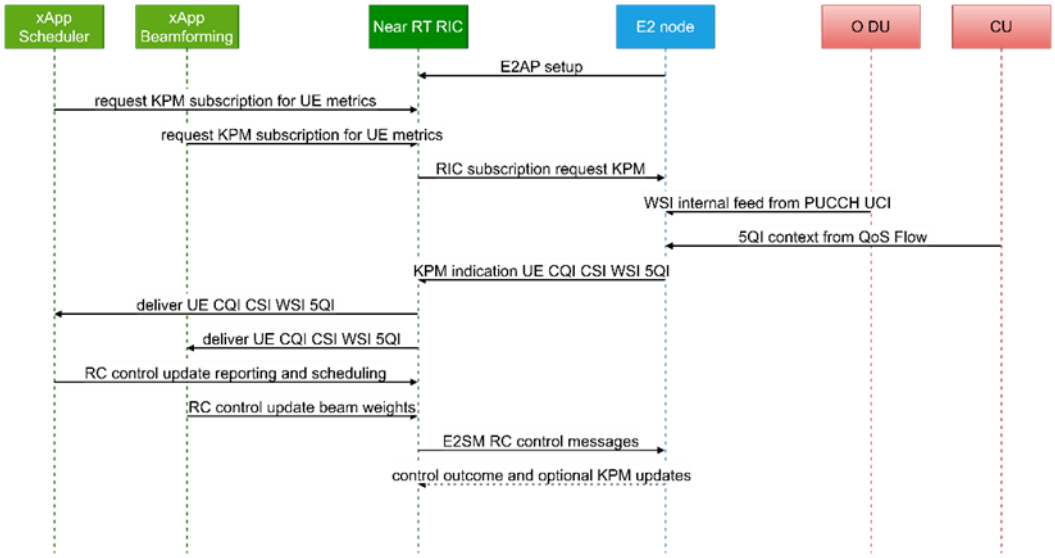


Figure 4-15 Near-RT RIC loop with WSI: O-DU provides WSI, CU adds 5QI, E2 node streams UE CQI CSI WSI 5QI via KPM to the RIC, and xApps return RC controls with outcomes reported back.

Relevant Signal paths

With the architecture of Figure 4-14 in mind, it is useful to summarize the three signal paths that are of high importance for the rest of the section.

The first is the WSI path. At the UE, a hybrid 5G/Wi-Fi sensing component produces a compact WSI value, derived from Wi-Fi and 5G signal measurements and interpreted as a zone or context label. This WSI is embedded into the UCI payload, together with HARQ-ACK, SR and CSI/CQI, and transmitted over PUCCH. At the O-DU, the UCI payload is decoded according to standard NR procedures (including CRC and Polar coding where applicable), and the WSI bits are extracted alongside CQI. The E2 node then includes WSI in the UE-level measurement stream towards the Near-RT RIC, where the beamforming xApp subscribes to it.

The second path is the familiar CQI feedback path. It follows standard NR operation: the UE reports CSI/CQI over UCI (or appropriate CSI reporting channels), the O-DU uses this for its internal scheduler, and the same CQI values are exported towards the Near-RT RIC for use by both xApps. The beamforming xApp uses CQI to evaluate whether WSI-driven beam changes are beneficial or not and the scheduler xApp uses CQI to understand what each UE can realistically achieve at a given time.

The third path is the 5QI exposure path. 5QI is part of the QoS Flow context that the CU-CP learns from the core. In our implementation, we retrieve this QoS context at the CU/DU, maintain a local mapping from DRB/QFI to 5QI, and then surface 5QI to the Near-RT RIC by adding it to the telemetry exported via E2. The scheduler xApp does not “discover” 5QI by magic, but it receives it because the RAN has deliberately included it in the data sent to the RIC.

These three paths together define the information that our xApps can see and act on. WSI gives a compact glimpse of the Wi-Fi-sensed environment, CQI describes the instantaneous radio conditions and 5QI encodes the QoS class of each flow. The next subsection will zoom in on WSI itself: how it is constructed from hybrid 5G/Wi-Fi measurements and how exactly it is packed into UCI and transported over PUCCH.

4.5.2 Wi-Fi Sensing Indicator (WSI)

This subsection describes what the WSI represents, how it is derived from hybrid 5G/Wi-Fi measurements, and how it is transported over the NR uplink control plane. Conceptually, WSI sits at the junction between the hybrid sensing/positioning logic and the O-RAN control loop: at one end it is grounded in the same type

of 5G/Wi-Fi fingerprints we already use for indoor positioning and at the other end it is carried in UCI on PUCCH using a mechanism inspired by our previous work on embedding additional indicators in NR control channels [38].

4.5.2.1 WSI Design

Role of WSI in the overall control loop

WSI is designed as a small, recurrent "hint" about the UE's environment, rather than a full description of its measurements. It captures, in a few bits, the output of a more complex hybrid sensing pipeline and makes it available to the Near-RT RIC at a time scale compatible with beamforming and scheduling decisions.

For the UE, it is the bridge between local hybrid sensing (5G + Wi-Fi measurements) and the 5G control plane. The UE does not send raw Wi-Fi vectors but only sends a compact WSI value instead.

For the O-DU, WSI is simply an additional field in the UCI payload that is decoded together with HARQ-ACK and CSI/CQI. It does not change the way PUCCH is processed. It slightly enlarges the UCI payload only.

For the Near-RT RIC, WSI is a per-UE measurement that the WSI-driven beamforming xApp subscribes to, and that it interprets as a location or context label when deciding which beams to favor.

The key design choice is to keep WSI as a micro-indicator: it uses a small number of bits and is carried over the same control channel as CQI, in order to minimize signalling overhead and leverage existing 3GPP mechanisms.

Hybrid 5G/Wi-Fi sensing as the basis for WSI

We derive the WSI from a hybrid, multi-RAT fingerprinting pipeline that integrates 5G-RAN and non-3GPP connectivity in the indoor setting of Figure 4-16. The setup consists of a single 5G gNB and a Wi-Fi AP, enabling multi-RAT positioning. The UE is equipped with both 5G and Wi-Fi modules, allowing it to collect signal measurements from both technologies. We denote the set of access nodes as $B=\{b_{5G},\ b_{WiFi}\}$ where b_{5G} represents the gNB and b_{WiFi} the Wi-Fi AP.

Fingerprint positioning is based on collecting signal characteristics at different locations and then match them with real-time data in order to estimate the location of the device. This is executed in a two-step process: an offline phase which includes the creation of fingerprint database, and an online phase where the real-time positioning takes place. During the offline phase, a predefined physical location (where the positioning will take place) is depicted as a grid with multiple reference points, where each reference point rp_i acts as a ground truth location in order to train the system. The set of Reference Points is denoted as $RP=\{rp_1, rp_2, \ldots\}$ with each rp_i having known coordinates (x_i, y_i). At every reference point rp_i the xApp collects signal measurements from both the 5G gNB and the Wi-Fi AP at fixed time intervals. Then, the signal measurements, along with the corresponding rp_i, form a record $F_i=\{rp_i, SM_{gNB}, SM_{WiFi}\}$ that is stored in a fingerprint database (radio-map).

The online phase follows once the fingerprint database is ready, where based on actual real-time signal measurements generated from the UE, the estimation of its location takes place. The UE captures its signal parameters (5G/Wi-Fi) in a location that it is unaware of and sends them to the location where the positioning algorithm is executed (fingerprint database). These metrics are then compared with the radio-map measurements and finally the system makes an estimation for the position of the device based on the closest fingerprint match. If multiple locations are similar, weighted averaging or interpolation is used to refine accuracy.

For the execution of the fingerprint positioning algorithm, we begin by dividing a predefined space into multiple grid points (as shown in Figure 4-16). This space covers 4 meters in length and 3 meters in width,

and it is divided into 12 reference points, with equal distances between them. In our setup, the Wi-Fi AP and the gNB are positioned at opposite ends of the grid. The UE used in this experiment was a Quectel device along with a COTS UE that connects to the non-3GPP network.

To build the radio-map (or offline database), we collect multiple signal measurements from each RP, from both the Wi-Fi and 5G networks. Specifically, the 5G signal metrics captured by the Quectel UE are forwarded to the RIC via the E2 interface. For the Wi-Fi part, the COTS UE, collocated with the Quectel device (since the Quectel lacks a Wi-Fi NIC card), is used. The AP records the received signal strength from the connected device, and these metrics are exposed to the RIC.

After receiving the signal measurements from ANs, the RIC stores them in an in-memory database creating a radio-map with multi-technology RAN measurements. The radio-map is then used to train a Support Vector Machine (SVM) classification ML model. The training of the ML model is ported in an xApp that uses as input real time RF measurements which are then mapped into position IDs. Figure 4-17 illustrates the actual trajectory of a UE alongside the trajectory estimated by the xApp.

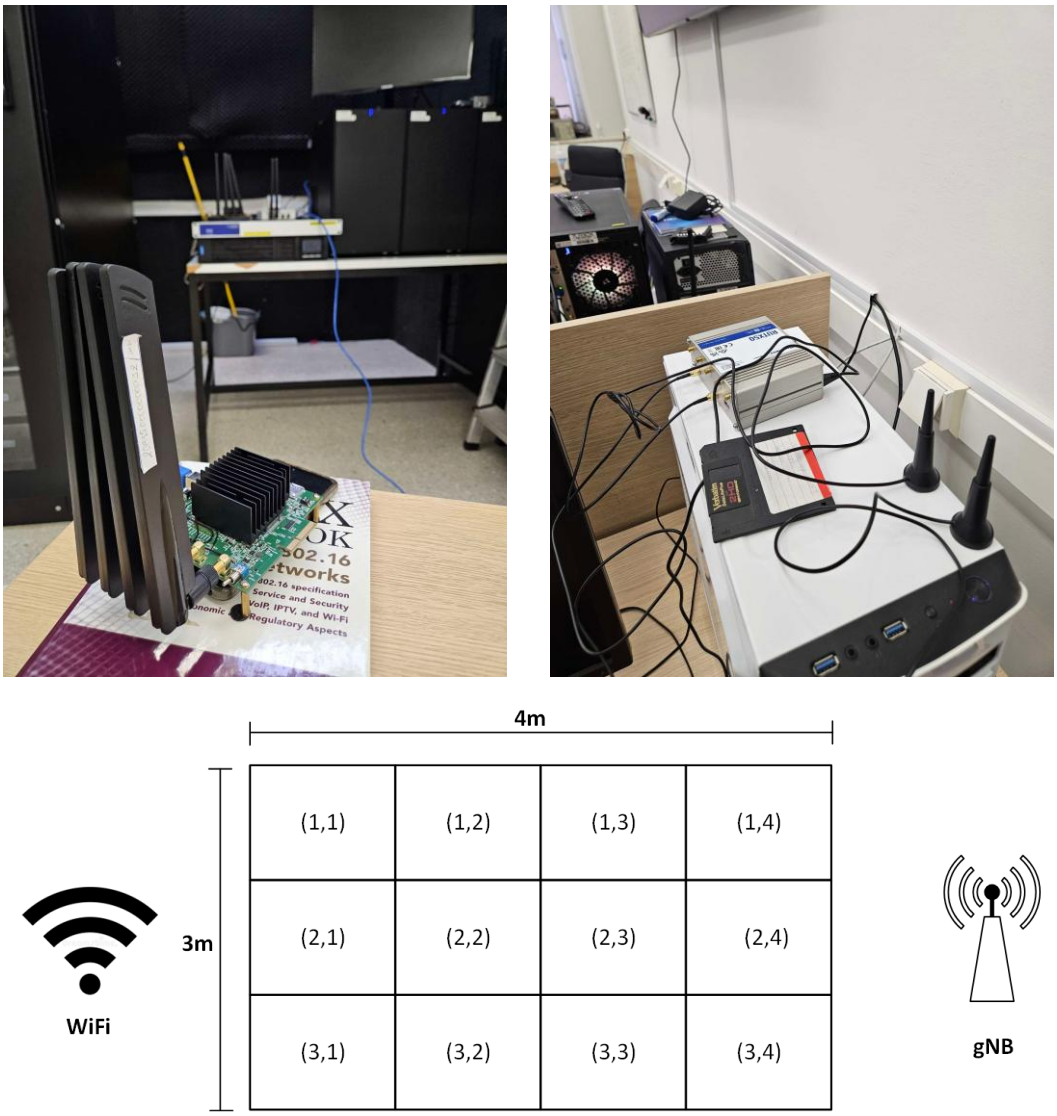


Figure 4-16 Experimental Area for Positioning. The area was divided into a 4x3 grid, with each block measuring 1m x 1m. The 5G gNB and the Wi-Fi AP were placed opposite to each other.

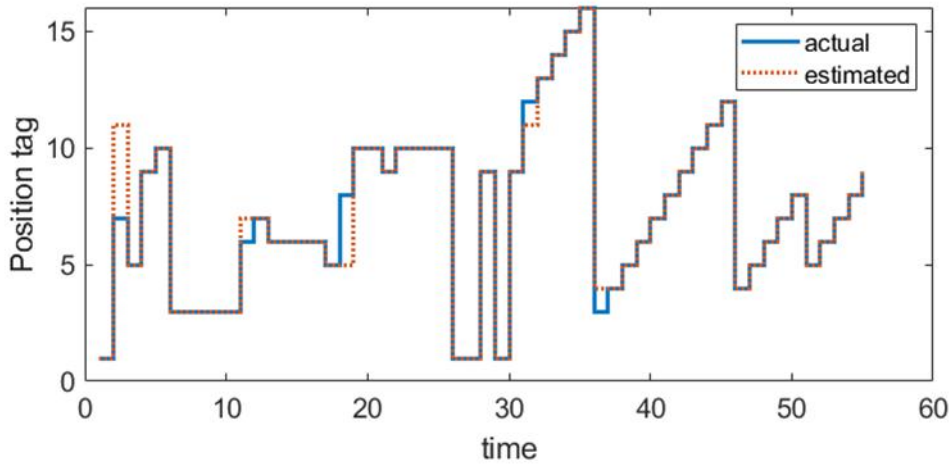


Figure 4-17 Actual and xApp based estimated trajectories for a UE.

WSI as a micro-indicator

Designing WSI as a micro-indicator imposes a few constraints. At first, WSI must be small. We reserve only a handful of bits in the UCI payload, typically in the range of 2 to 8 bits. This is enough to index a meaningful number of zones or clusters (4 to 256, depending on configuration) while keeping the impact on UCI size and PUCCH resources minimal. Second, WSI must be robust enough to be useful, but not necessarily perfect. The hybrid sensing pipeline can be more sophisticated internally, but its output is coarsened into a small set of classes. Third, WSI must be frequent. Because it is a small field embedded on PUCCH, it can be updated at intervals comparable to CQI updates, allowing the Near-RT RIC to track changes in the UE's environment and react by adapting beams.

4.5.2.2 Embedding WSI into UCI on PUCCH

The last part of this subsection focuses on the challenges associated with bringing Wi-Fi–assisted sensing into a 6G-ready O-RAN system. Concretely, we need an effective mechanism for transmitting Wi-Fi-based sensing information from the UE to the network, in a way that is fast, lightweight and does not disrupt existing protocol stacks.

More specifically, our goal is to enable a low-latency and scalable transfer of Wi-Fi sensing context from UEs to the RAN by exploiting standardized control signaling, without requiring changes to Wi-Fi access points or to the 5G physical channels. To this end, we propose to reuse the PUCCH as a real-time signaling vehicle, and to integrate a WSI into the UCI carried on PUCCH.

UCI is already protected by CRC-assisted Polar coding. Building on our previous work [38], we exploit this protection to embed a small WSI payload of 2, 4, 6 or 8 bits inside the UCI. In this way, the UE can periodically convey a compact representation of its Wi-Fi-assisted context (for example, a zone or cluster identifier derived from hybrid 5G/Wi-Fi fingerprints) while the network remains free to use this information to optimize beamforming, link adaptation and scheduling decisions.

Our approach offers a flexible and scalable way to inject non-3GPP sensing into an otherwise standard 5G NR control framework. It leverages existing NR control-channel structures, and extends them just enough to support more intelligent, context-aware transmission behaviors. In contrast to solutions that operate only at the application layer, here the decision-making is explicitly integrated at the PHY/MAC level: WSI is carried within UCI, processed at the O-DU, and then made available to Near-RT RIC xApps that operate in real time on beam selection and QoS-aware scheduling.

Physical Uplink Control Channel and WSI Mapping

Unlike LTE, where PUCCH was limited to fixed spectrum locations, the flexible allocation mechanism in 5G NR allows UCI to be transmitted across different time–frequency resources based on system demands. PUCCH in NR is structured into multiple formats, each optimized for specific transmission needs. Short PUCCH formats, using one or two OFDM symbols, are suited for compact control payloads such as HARQ-ACK and scheduling requests. Long PUCCH formats, spanning four to fourteen OFDM symbols, support larger payloads including HARQ-ACK, scheduling requests and other essential UCI components [39]. These extended formats allow additional information to be transmitted with improved robustness and enhanced range, making them ideal for integrating the proposed WSI without introducing excessive signaling overhead.

The integration of WSI within UCI extends the existing control signaling framework by embedding a variable-length WSI field (e.g., 2, 4, 6 or 8 bits) that informs the O-RU, O-DU and O-CU of the UE's Wi-Fi-sensed zone or environment. While CQI reflects the current channel state, WSI conveys spatial and environmental context derived from hybrid 5G/Wi-Fi measurements, enabling a dynamic adaptation mechanism where the network adjusts beams and scheduling based on both radio conditions and the UE's position within the indoor topology.

The selection of a specific PUCCH format for WSI transmission depends on the required UCI payload size and accompanying control elements. Short PUCCH formats cannot be used for WSI transmission when WSI is combined with CSI/CQI, because the additional bits push the payload beyond the capacity of short formats. When WSI is carried together with CQI and CSI reports, a long PUCCH format is therefore necessary to accommodate the additional control signaling while maintaining transmission reliability.

Our proposed design aligns with ongoing 3GPP standardization directions on NR control channels, introducing a practical method for enhancing uplink control efficiency through context-driven, Wi-Fi–assisted adaptation. It provides a seamless and adaptive approach to resource allocation and beam management in next-generation wireless systems.

Impact of WSI Payload on CRC

Reliable transmission of WSI within PUCCH requires robust error detection mechanisms to maintain the integrity of control signaling. In accordance with [39][40], UCI payloads transmitted via PUCCH undergo CRC attachment, which is essential for detecting transmission errors and ensuring correct reception. The CRC length is determined by the total size of the UCI payload, following a structured approach that minimizes error propagation while maintaining efficient signaling.

When the total UCI payload is 11 bits or fewer, no CRC is applied, as the likelihood of undetected error is minimal. For UCI payloads ranging between 12 and 19 bits, a 6-bit CRC is appended to enhance error detection capability, while payloads of 20 bits or more require an 11-bit CRC to provide stronger protection against transmission errors [40].

The introduction of WSI bits influences CRC selection when multiplexed with CQI and other UCI elements. Given that WSI is designed to be transmitted alongside CQI in long PUCCH formats, it follows the CRC attachment rules applicable to larger UCI payloads. For example, adding a 4-bit WSI to a 10-bit UCI payload increases the total size to 14 bits, triggering the need for a 6-bit CRC. Similarly, a UCI payload of 16 bits, once combined with WSI, reaches 20 bits, requiring an 11-bit CRC.

The CRC-protected UCI payload, including WSI and CQI, enables robust error detection and retransmission mechanisms that prevent Wi-Fi sensing information from being lost or misinterpreted due to transmission impairments. PUCCH's design follows the repetition procedure outlined in [42], ensuring that the entire UCI payload is retransmitted, if necessary, rather than selectively retransmitting erroneous segments. Unlike the

procedure followed for the Physical Uplink Shared Channel (PUSCH), where retransmissions typically handle data payloads via HARQ [40][41], the retransmission strategy for PUCCH is performed by retransmitting the complete UCI payload, including the WSI and associated control information. This comprehensive approach assures control-information integrity and correctness that is critical for WSI-driven beamforming and scheduling. By embedding WSI within the existing CRC-based error-control framework, the proposed solution aligns with 3GPP standardization objectives and maintains a scalable, backward-compatible path for Wi-Fi–assisted control in future 6G networks.

WSI Integration in Channel Coding

The integration of WSI within UCI requires an optimized channel coding strategy to ensure robust and efficient transmission, particularly in the context of the O-RAN framework. In accordance with [40], the coding scheme applied to the UCI depends on the overall payload length, which drives the selection of the appropriate encoding method to align with standardized uplink control procedures.

For UCI payloads of 11 bits or fewer, channel coding of small block lengths is applied, facilitating efficient transmission with minimal computational complexity. When the UCI payload exceeds 11 bits, due to multiplexing with additional elements such as HARQ-ACK, SR, CSI and WSI, Polar coding is required, as defined in [40], ensuring enhanced error resilience and reliable reception under dynamic radio conditions.

Prior to transmission over PUCCH, rate matching is performed to optimize bit allocation and spectral utilization. As outlined in [40], this process ensures that UCI elements are allocated efficiently within the available uplink control resources. PUCCH format selection then depends on the resulting payload size [41]: shorter payloads utilize Format 1 or 2, whereas larger UCI payloads, including those carrying WSI, are mapped to Format 3 or 4 to ensure robust transmission.

Upon reception, the O-DU decodes the UCI according to the encoding scheme used. For small payloads, decoding is straightforward and enables low-latency retrieval of control elements. When Polar coding is employed, successive cancellation decoding reconstructs the UCI, enabling real-time adaptation of beamforming and scheduling within the O-RAN framework. Embedding WSI within the existing UCI encoding structure gives this approach the potential to contribute to ongoing efforts in advancing 3GPP standards towards more context-aware, multi-RAT control.

4.5.3 xApp Design and Implementation

This subsection describes how the two main control components of the proposed framework are realized as xApps on the Near-RT RIC. The first xApp exploits the WSI to improve downlink beamforming decisions, while the second xApp uses CQI and 5QI to steer scheduling in a QoS-aware manner. Both xApps operate on top of an O-RU/O-DU/O-CU stack and interact with the RAN through standard E2-based telemetry and control procedures, without modifying the underlying 3GPP protocol layers.

The design follows a common pattern: the E2 node exports UE-level measurements and QoS context towards the Near-RT RIC; the xApps process this information at near-real-time timescales (tens to hundreds of milliseconds) and the resulting control decisions are pushed back to the O-DU and O-RU. In this way, WSI and 5QI become first-class inputs to the RAN control loop while preserving compatibility with existing O-RAN deployments.

4.5.3.1 WSI- driven beamforming xApp

The WSI-driven beamforming xApp is responsible for exploiting Wi-Fi-assisted sensing to keep UEs in favorable radio conditions. Conceptually, it treats WSI as a compact descriptor of the UE's indoor zone or propagation environment and uses this descriptor, together with CQI and CSI measurements, to select or refine the beams applied at the O-RUs.

At run time, the xApp receives a continuous stream of measurements from the E2 node. For each active UE, these measurements include the most recent WSI value decoded at the O-DU from UCI on PUCCH, the associated CQI/CSI reports and, where available, simple performance indicators such as average SINR or recent block error rates. The WSI identifies the zone in which the UE is currently located according to the hybrid 5G/Wi-Fi fingerprinting process described earlier in the document. The xApp maintains, for each such zone, a small set of candidate beams or beamweight vectors that have proven effective in that zone in the past.

When a new measurement report is received, the xApp interprets the WSI value as a key into this zone-specific beam knowledge base. For the current WSI, it retrieves the corresponding beam candidates and selects one of them as the preferred configuration for the UE, considering the current and recent CQI values. If CQI is already high under the existing configuration, the xApp can choose to keep the beam unchanged, merely confirming the current mapping between WSI and beam. If CQI is poor, it may switch to an alternative candidate beam for that zone, with the explicit goal of improving the channel quality experienced by the UE.

Once a decision is made, the xApp formulates an E2 control message that encodes the recommended beam index or beamweight adjustment and sends it to the E2 node. This message is then translated into a concrete configuration update at the O-RU, for example by updating the beamforming weights used for downlink transmissions towards the target UE. No changes to the physical layer are required: the xApp simply exploits the vendor's existing ability to reconfigure beam tables or digital precoders via O-RAN-compliant control hooks.

The beamforming logic operates in closed loop (see Figure 4-18). After applying a beam recommendation, the system continues to report CQI and related metrics for that UE. The xApp compares the observed CQI evolution with its expectations and gradually refines its internal view of which beams are effective for each WSI zone. If a particular (zone, beam) combination consistently leads to poor CQI, that combination is downgraded in the internal ranking and alternative beams are favored in subsequent iterations. Conversely, if a configuration repeatedly yields good CQI and stable throughput, the association between the WSI zone and that beam is reinforced.

Over time, this process leads to a self-adapting mapping between WSI zones and beam patterns. The mapping is robust to changes in the environment: if furniture is moved, walls are modified or additional Wi-Fi equipment is deployed, the hybrid sensing pipeline will produce different WSI–measurement relationships, and the closed-loop beamforming logic will gradually learn new preferred beams for the updated zones. Importantly, all of this learning happens within the RAN and the Near-RT RIC.

The net effect is that downlink transmissions become explicitly aware of the UE's Wi-Fi-sensed environment. UEs located in zones known to be challenging (for example, behind walls or in corners) can benefit from more suitable beams, increasing their CQI and improving the effectiveness of subsequent QoS-aware scheduling.

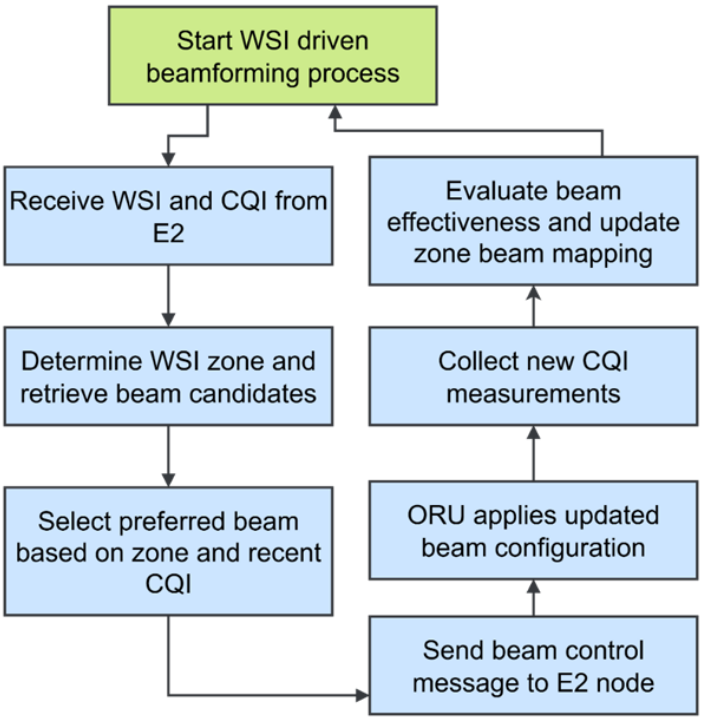


Figure 4-18 WSI-driven beamforming xApp, which uses WSI and CQI measurements to select and refine downlink beam configurations in a closed control loop.

4.5.3.2 CQI + 5QI scheduler xApp

The second xApp, the CQI and 5QI scheduler, focuses on resource allocation. It does not replace the MAC scheduler in the O-DU but acts as a policy layer that biases the scheduler through per-UE weighting, based on both instantaneous radio conditions and the QoS class of each flow.

The scheduler xApp receives, again via E2, a set of per-UE indicators at regular intervals. These indicators include CQI (and possibly related radio metrics such as MCS history or wideband SINR), basic traffic statistics from which average throughput can be inferred, and the 5QI associated with each QoS Flow or DRB. As explained earlier, 5QI is obtained from the QoS Flow context maintained at the CU-CP and DU and exposed to the Near-RT RIC.

Using this information, the xApp periodically computes a scheduling “score” for each UE or flow. The computation is policy-driven but follows a common structure. First, CQI and 5QI-related parameters are normalized so that they can be combined. CQI indicates how efficiently the network can serve the UE from a radio perspective, while 5QI encodes the QoS strictness of the associated flow (for example, lower 5QI indices correspond to tighter delay or reliability budgets). A fairness term based on historical throughput can optionally be included to prevent starvation of UEs that have been underserved in the past.

On top of this normalized information, the xApp can implement different scheduling behaviors. In a baseline, CQI-only mode, the score may be dominated by CQI and aim to mimic a traditional channel-aware scheduler: UEs with good radio conditions are favored, and 5QI is effectively ignored. This provides a reference for comparing the impact of exposing 5QI to the RIC.

In a more balanced CQI + 5QI mode, the scheduler xApp combines the two signals so that flows with stricter QoS requirements are more likely to be prioritized, provided their CQI is not prohibitively low. For example, a UE carrying a low-5QI flow with acceptable CQI may receive a higher score than a UE carrying a best-effort

flow with slightly better CQI. In this way, the scheduler can protect critical services while still taking advantage of good channels when they are available.

Finally, a QoS-first mode can be used when stringent flows must be protected aggressively. In this configuration, the score calculation puts significant weight on 5QI, subject to a minimum CQI threshold. UEs associated with low 5QI classes are strongly favored as long as their radio conditions are above that threshold, whereas best-effort traffic is scheduled more opportunistically. This mode is useful for exploring the trade-off between aggregate throughput and strict QoS satisfaction.

Once scores have been computed for all UEs or flows, they are normalized into a weight vector. This vector is sent back to the O-DU using an E2 control message that carries per-UE scheduling weights. The O-DU's internal scheduler then uses these weights as inputs when selecting which UEs to serve in the next scheduling interval. The exact mapping between weights and resource assignments is implementation-specific and does not need to be standardized. From the xApp's perspective, it suffices that higher weights translate into a higher probability of being scheduled or a larger share of resources.

By repeatedly adjusting these weights based on up-to-date CQI and 5QI, the scheduler xApp keeps the RAN operation aligned with both radio conditions and the QoS configuration provided by the Core. When combined with the WSI-driven beamforming xApp, which seeks to improve CQI in the first place, the result is a two-stage control process: beams are adapted to make the channel as favourable as possible for each UE in its current zone, and then scheduling weights are computed to distribute resources in a way that respects QoS classes while maintaining reasonable efficiency and fairness.

A schematic of the scheduler xApp closed loop is depicted in Figure 4-19.

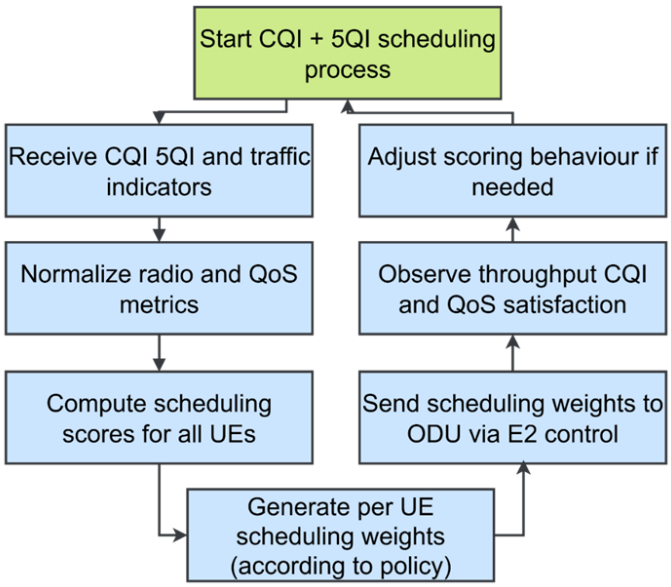


Figure 4-19 CQI + 5QI scheduler xApp, which computes per-UE scheduling weights by combining radio conditions and QoS class information to guide MAC-layer resource allocation.

4.5.4 Evaluation

This section reports the main results obtained from the scenarios described earlier. We first examine the impact of embedding WSI into the UCI payload on the reliability of the uplink control channel, and then move to the system-level effects observed when WSI-driven beamforming and CQI+5QI scheduling are enabled in the Near-RT RIC loop.

Impact of WSI on the uplink control channel

The first set of experiments focuses on the control-plane implications of adding WSI bits to UCI, before analyzing any E2E QoS effects. The objective is to verify that carrying WSI over PUCCH does not compromise the robustness of uplink control signaling beyond acceptable limits.

The integration of WSI within UCI, before its transmission via PUCCH, introduces a minimal control overhead that does not significantly degrade system performance. PUCCH is optimized for control signaling, ensuring that the additional bits reserved for WSI remain within acceptable signaling constraints. As mentioned above, Polar codes are employed and are considered to be particularly effective in maintaining low Bit Error Rate (BER) for short UCI payloads in 5G NR systems.

To evaluate the impact of WSI integration on BER performance, Figure 4-20 illustrates the BER vs. SNR for various input bit sizes. The green curve represents the BER for a baseline UCI payload (without WSI), while the additional curves depict BER degradation when 2, 4, 6, and 8 bits of WSI are added. The results confirm that increasing the number of transmitted bits slightly degrades BER due to the higher probability of bit errors in transmission. The slight increase in BER can also be due to the change in the effective coding rate resulting from the extra bits. However, the observed BER increase remains within tolerable limits, which is explained by the robustness of Polar coding in handling UCI transmission efficiently.

Our results confirm that adding WSI bits introduces only a marginal increase in BER. This rise in BER however comes with significant benefits in network efficiency, Wi-Fi-assisted adaptation, and improved spectral resource management. It should be noted that the controlled BER increase remains within acceptable limits, ensuring that WSI-enhanced control signaling in PUCCH does not notably compromise the reliability of the system.

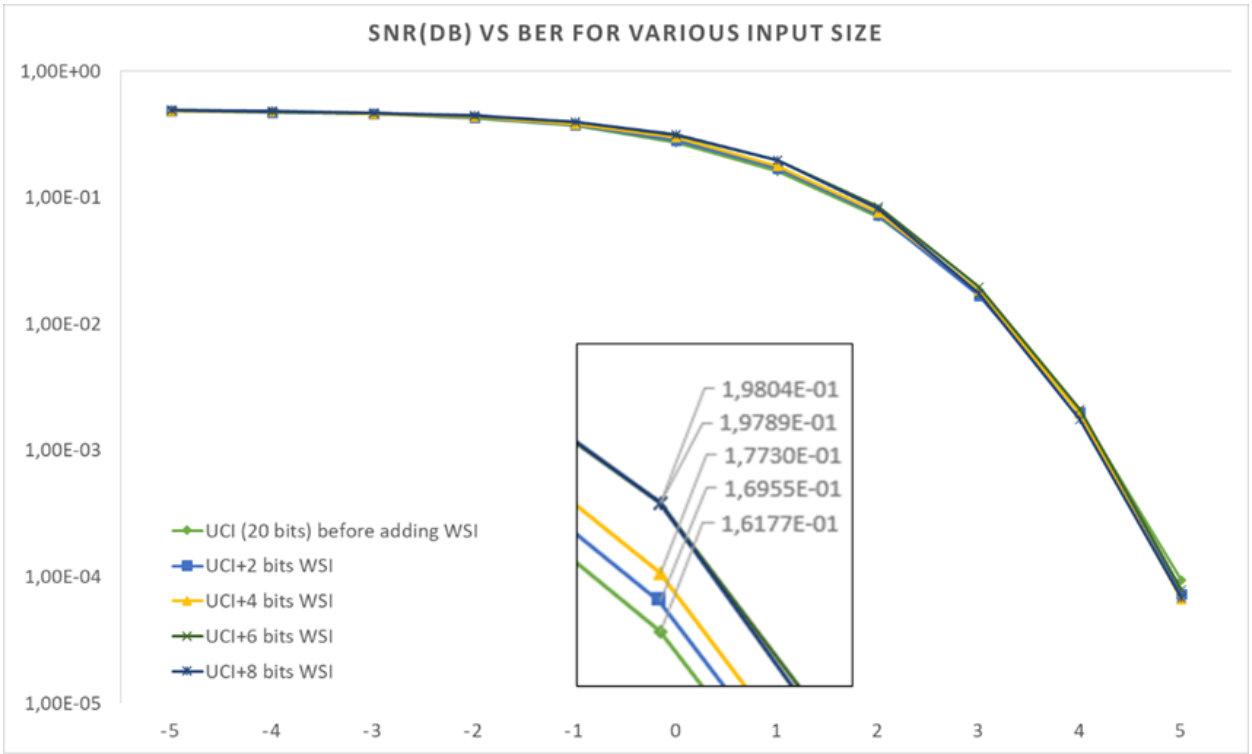


Figure 4-20 BER vs SNR performance of Polar coding for various block lengths. Integration of WSI slightly increases BER due to larger payload sizes, but the impact remains limited thanks to the robustness of Polar coding.

System-level evaluation objective and high-level setup

The remainder of this section evaluates the E2E behavior of the proposed near-RT control framework when (i) the WSI-driven beamforming xApp and (ii) the CQI + 5QI scheduler xApp operate in the same O-RAN control loop. The objective is twofold. First, we quantify the system-level benefit of injecting Wi-Fi–assisted context (WSI) into the Near-RT RIC loop and using it for beam targeting, with the expected outcome being an improvement in the observed CQI evolution for UEs that traverse challenging indoor zones. Second, we quantify how exposing 5QI to the scheduler xApp allows the RAN to translate recovered radio capacity into improved QoS handling, by prioritizing flows with stricter QoS requirements when radio conditions permit.

To this end, we considered two main configurations. In the WSI-OFF baseline, the UE does not embed WSI in UCI and the RAN operates with its default beam management, relying on CSI/CQI as provided by the implementation. In the WSI-ON configuration, the UE periodically reports WSI over PUCCH, the O-DU decodes and forwards WSI to the Near-RT RIC, and the beamforming xApp uses this information, together with CQI, to select beams from the zone-specific codebook described in the previous section. In both cases, UEs follow trajectories that move them between favorable and unfavorable propagation regions, including Non-Line-of-Sight (NLoS) segments and corners of the indoor area.

The role of WSI-ON is to improve the observed CQI evolution, primarily by reducing the duration and frequency of low-CQI episodes after zone transitions through timely WSI-triggered beam targeting updates. Each UE may traverse both favorable and challenging regions at different times, creating runs where degraded CQI segments occur for different users and may overlap. This makes the scenario representative of realistic indoor deployments, where mobility and local geometry determine when and how each UE enters disadvantaged propagation conditions.

Finally, to enable the CQI + 5QI scheduler xApp, 5QI is available to the Near-RT RIC as part of the flow/QoS context (via CU telemetry exposure in the testbed). The scheduler uses 5QI jointly with CQI and traffic indicators to compute scheduling weights, allowing comparisons between channel-driven and QoS-aware policies under the same dynamic radio conditions.

WSI- driven beamforming xApp: Beam OFF vs Beam ON (WSI-driven beamforming disabled/enabled)

To align the evaluation with the proposed WSI mechanism, we map the “beam OFF / beam ON” modes to two operational states of the beamforming control loop:

- In Beam OFF (WSI-OFF), the WSI-driven beamforming xApp is disabled and no WSI is conveyed to the Near-RT RIC. Beam management therefore follows the default/non-WSI behavior of the underlying RAN implementation, relying on conventional measurements (e.g., CSI/CQI) and whatever baseline beam selection/tracking procedures are provided. Under indoor mobility, UEs may enter unfavorable regions (e.g., NLoS segments, corners, blockage pockets) where channel quality drops and may remain low for non-negligible intervals, until baseline procedures restore a better beam alignment.
- In Beam ON (WSI-ON), WSI is periodically embedded by the UE into UCI and reported over PUCCH. The O-DU decodes the UCI payload, extracts WSI, and forwards it to the Near-RT RIC. The beamforming xApp consumes WSI together with CQI and uses the WSI value as a compact zone/context key to select or recommend beams from the zone-specific codebook described previously. As UEs move across zones, changes in WSI trigger beam targeting updates that aim to shorten and reduce low-CQI intervals by converging faster to a beam configuration that is better matched to the local propagation conditions. CQI remains a dynamic outcome of the environment and mobility in both configurations. The role of Beam ON is therefore evaluated through its effect on the CQI trajectory (e.g., fewer deep CQI drops and faster recoveries after zone transitions).

CQI + 5QI scheduler xApp: inputs, operation, and policies (CQI + 5QI)

The downlink scheduler xApp is evaluated as a policy layer that biases scheduling decisions through per-UE (or per-aggregate) weight updates, without replacing the O-DU scheduler. At each control step, the xApp retrieves the following inputs from the RAN/near-RT telemetry stream:

- CQI, representing instantaneous radio feasibility and achievable MCS,
- Traffic/load indicators, such as recent throughput estimates
- 5QI, representing the QoS class of the active traffic (QoS flow / QFI) and its relative strictness.

Since 5QI is defined at the QoS-flow level, while the weight interface is typically applied at a coarser granularity, the xApp maps per-flow 5QI information into a per-UE scheduling preference. In this evaluation, this mapping is implemented by selecting the dominant active 5QI per UE (e.g., the strictest class currently active, or a weighted aggregation across the UE's active flows), and using that value to compute the UE's QoS priority term. Using the above inputs, the xApp computes a normalized score per UE and converts scores into a weight vector that is sent to the RAN over E2.

To keep the evaluation comparable across scenarios, we consider the following scheduler policy modes:

- cqi_only: channel-driven baseline that uses CQI to maximize instantaneous radio efficiency, without using 5QI.
- balanced (CQI + 5QI): combines CQI and 5QI, prioritizing stricter QoS classes while still respecting feasibility and avoiding excessive allocation under very low CQI.
- QoS_first (5QI-first): prioritizes stricter 5QI classes more aggressively, subject to a minimum CQI feasibility condition to prevent persistent resource waste.

These policy modes are evaluated under both Beam OFF and Beam ON. This separation is essential: Beam OFF/ON captures the contribution of WSI-driven beam targeting to the CQI trajectory, while the policy modes capture how effectively the scheduler translates the available (and potentially improved) capacity into QoS-aware resource allocation.

Mapping to policies and runs

The next table summarizes the set of scenarios S1–S6 used in the evaluation, showing for each case whether the WSI-driven beamforming xApp is active (WSI-ON or WSI-OFF) and which scheduling policy is applied.

Table 4-2 Set of scenarios S1-S6 used in the system-level evaluation.

Scenario	Beam state	Scheduler policy
S1	WSI-OFF (Beam OFF)	cqi_only
S2	WSI-OFF (Beam OFF)	balanced (CQI + 5QI)
S3	WSI-OFF (Beam OFF)	QoS_first
S4	WSI-ON (Beam ON)	cqi_only
S5	WSI-ON (Beam ON)	balanced (CQI + 5QI)
S6	WSI-ON (Beam ON)	QoS_first

The first group of scenarios explores how the scheduler behaves under dynamic indoor conditions when WSI-driven beam targeting is disabled.

- **S1 – WSI-OFF, cqi_only**: WSI is not reported and the beamforming xApp is disabled. CQI evolves under baseline beam management and reflects the natural indoor dynamics (zone transitions, blockage, multipath). The scheduler uses a CQI-only policy and therefore acts as a QoS-blind baseline.
- **S2 – WSI-OFF, balanced (CQI + 5QI):** Same WSI-OFF beam state and the same type of dynamic CQI evolution, but the scheduler combines CQI with 5QI. This scenario shows how much QoS handling can be improved purely by changing the scheduler policy, even when beam control does not exploit WSI.
- **S3 – WSI-OFF, QoS_first:** Again WSI-OFF, but the scheduler prioritizes strict 5QI classes more aggressively, subject to a CQI feasibility condition. This is the "aggressive QoS" configuration without WSI-driven beam targeting.

The second group explores the joint effect of WSI-driven beamforming and scheduler policy under the same type of dynamic indoor conditions.

- **S4 – WSI-ON, cqi_only:** WSI is periodically reported over PUCCH and consumed by the beamforming xApp. Beam targeting updates are triggered on zone changes and use the zone-specific codebook. The scheduler remains CQI-only. This scenario isolates the pure beamforming gain: improved CQI dynamics (faster recovery, fewer sustained low-CQI periods) should translate into higher throughput even without QoS-aware scheduling.
- **S5 – WSI-ON, balanced (CQI + 5QI):** Same WSI-ON beam targeting as S4, but the scheduler combines CQI and 5QI. This scenario represents the intended "full-stack" configuration: WSI-driven beamforming stabilizes CQI during challenging segments, and 5QI-aware scheduling exploits the recovered capacity to improve QoS satisfaction for strict flows while respecting radio feasibility.
- **S6 – WSI-ON, QoS_first:** WSI-driven beam targeting is active and the scheduler prioritizes strict 5QI classes. This run explores the extreme QoS-driven configuration in which strict flows are protected, with the expectation that WSI-ON reduces the frequency of infeasible intervals and thereby reduces the penalty of aggressive QoS prioritization.

System-level evaluation Results

Below we report how the proposed framework behaves across the scenarios S1–S6. For each scenario, the RIC Supervisor aggregates the logs from the CQI+5QI scheduler xApp and the RAN, and computes two main metrics over each run: (i) SumDL [Mbit/s], i.e., aggregate downlink throughput across the UEs, and (ii) 5QI satisfaction (5QI-Sat), i.e., a scalar KPI that summarizes how well traffic with stricter QoS requirements is served compared to best-effort classes, based on delivered traffic per 5QI class.

In all scenarios, UEs generate constant downlink traffic (iperf3) so that differences in SumDL and 5QI-Sat reflect control decisions rather than traffic intermittency, while at the same time, flows are configured with distinct 5QI classes so that the scheduler's QoS-aware modes can be meaningfully evaluated.

Global behavior across S1–S6: Across the full set of scenarios S1–S6 the behavior is very consistent and tells a clear story (see Figure 4-21). When the WSI-driven beamforming xApp is disabled (WSI-OFF/Beam OFF, S1–S3), SumDL stays around 13.5 Mbit/s, regardless of the scheduler policy. In this regime, policy choice mainly shows up in the 5QI-Sat values: even with Beam OFF, QoS-aware policies improve 5QI-Sat compared to the QoS-blind baseline, with 5QI-Sat increasing from 7.274 (S1) to 8.22–8.502 (S2–S3) when 5QI is accounted for (balanced and QoS_first). Once the WSI-driven beamforming xApp is enabled (WSI-ON/Beam ON, S4–S6), SumDL jumps to about 17.7 Mbit/s in all cases, roughly a 28–31% gain over the WSI-OFF scenarios, while 5QI-Sat also rises sharply. Here we see a clear QoS benefit: for the same beam state and very similar throughput, the QoS-aware runs reach 5QI-Sat = 11.595 (S5) and 12.613 (S6), whereas the QoS-blind baseline (cqi_only) is 11.024 (S4). The QoS_first policy in S6 gives the highest 5QI satisfaction overall, without any throughput

penalty. Figure 4-22 highlights the combined benefit of WSI-driven beamforming and 5QI-aware scheduling. Moving from the baseline (WSI-OFF, cqi_only, S1) to the showcase configuration (WSI-ON, QoS_first, S6) increases the aggregate downlink throughput from 13.557 to 17.762 Mbit/s and raises 5QI satisfaction from 7.274 to 12.613, i.e., significantly more useful (QoS-compliant) bits are delivered. At the same time, the throughput shares of UE1 and UE2 stay close to 50/50 in both cases, showing that the gain in performance and QoS is achieved without sacrificing fairness between users.

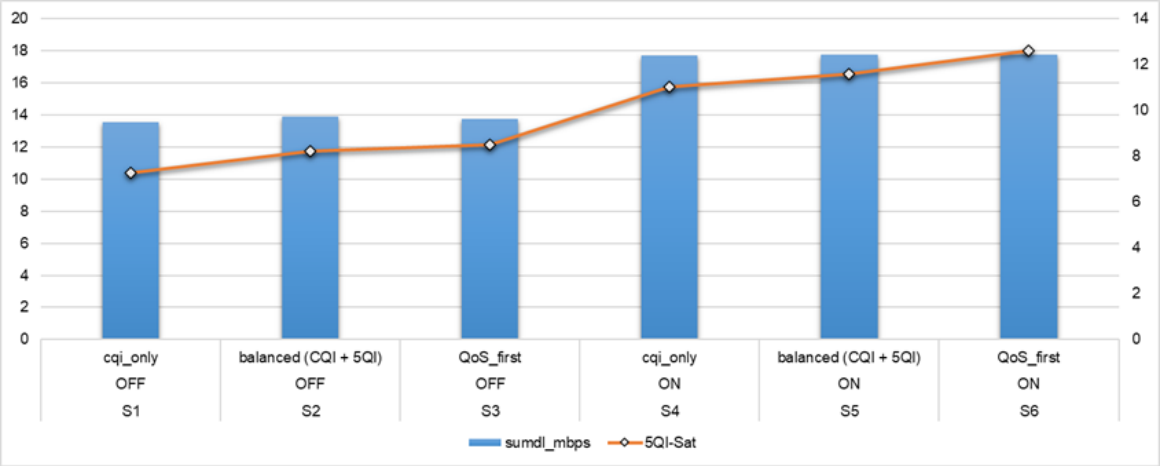


Figure 4-21 Average downlink throughput and 5QI satisfaction across scenarios S1–S6, comparing WSI-OFF/WSI-ON and contrasting scheduler policies (cqi_only vs balanced vs QoS_first).

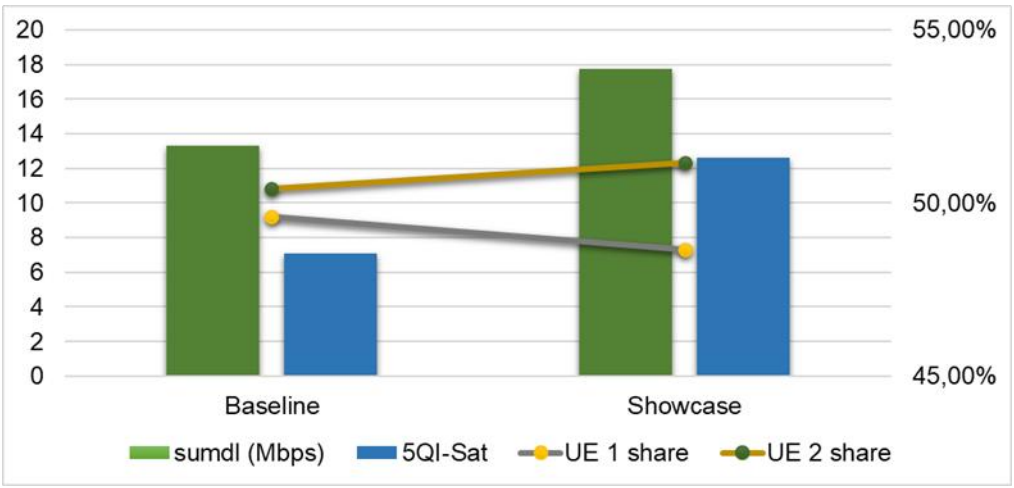


Figure 4-22 Baseline versus showcase: joint gain in throughput and 5QI satisfaction with preserved UE fairness.

Incremental gains from S1 to S6: The impact of enabling the WSI-driven beamforming xApp can be seen clearly when comparing S1–S3 (WSI-OFF) with S4–S6 (WSI-ON) on a per-policy basis (see Figure 4-23). Enabling WSI-driven beam targeting is the dominant lever across S1–S6: turning WSI ON boosts SumDL by roughly 27.8–30.8% for every policy and lifts 5QI-Sat by 41.1–45.8%. Concretely, the measured incremental gains from WSI-OFF to WSI-ON are:

- cqi_only: ΔSumDL = +30.8%, Δ5QI-Sat = +45.8%
- balanced (CQI + 5QI): ΔSumDL = +27.8%, Δ5QI-Sat = +41.1%
- QoS_first: ΔSumDL = +29.2%, Δ5QI-Sat = +43.3%

These results confirm that WSI-driven beamforming provides a largely policy-independent throughput lift by improving the CQI trajectory under dynamic indoor conditions, while the scheduler policy remains a second-order lever for throughput and primarily affects the relative QoS benefit once radio feasibility is improved.

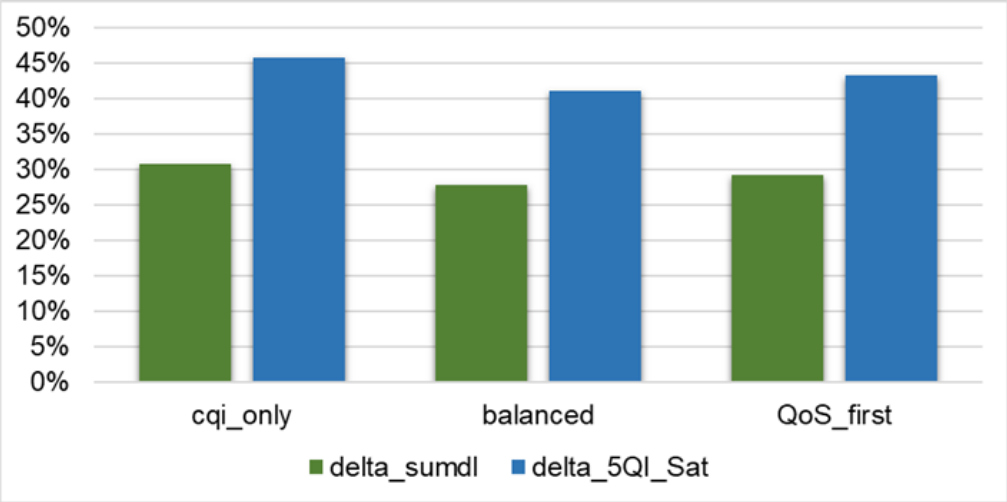


Figure 4-23 Percentage gain from WSI-OFF to WSI-ON per policy, shown for SumDL and 5QI-Sat.

Cross-scenario discussion (WSI + CQI + 5QI): Across scenarios S1–S6, the results show a clear split of roles between the two control functions. Enabling WSI-driven beam targeting is the main robustness lever: switching from WSI-OFF to WSI-ON moves the system to a higher-throughput operating point, with SumDL increasing by ~27.8–30.8% across all policies. This confirms that improving the physical layer through faster, zone-aligned beam selection is a first-order requirement under dynamic indoor CQI variability.

Within a fixed beam state, 5QI-aware scheduling mainly provides a "vertical" gain in 5QI-Sat rather than a throughput gain. While SumDL changes little across policies, 5QI-Sat improves, and the WSI-OFF → WSI-ON transition also lifts 5QI-Sat by 41.1–45.8%, because stricter QoS classes become schedulable more often once CQI is stabilized.

Figure 4-24 below shows a representative run at the per-UE level. The downlink throughput alternates between the two UEs as channel conditions evolve with mobility and zone transitions, revealing the short-time-scale behavior. In periods where one UE becomes disadvantaged (e.g., entering an NLoS segment or corner), its throughput drops sharply, while the other UE temporarily dominates the delivered traffic. Under WSI-ON, these disadvantaged intervals are typically shorter and recover faster due to WSI-triggered beam targeting, which stabilizes CQI and restores throughput more quickly.

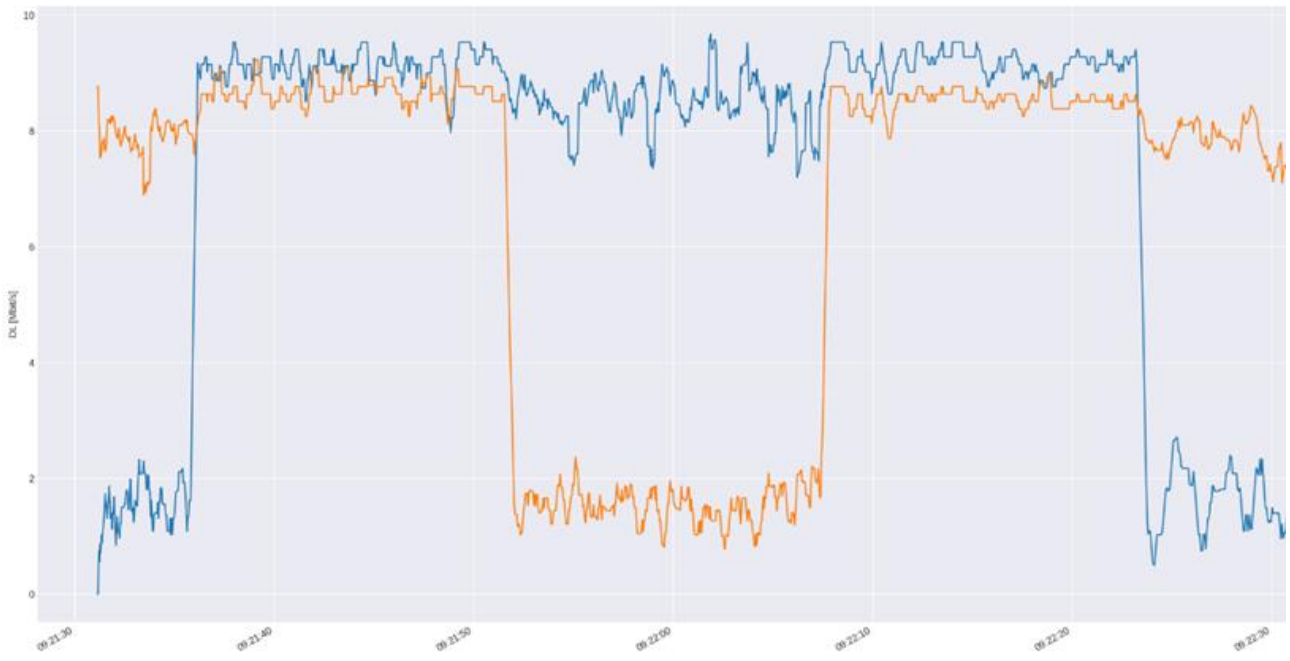


Figure 4-24 DL (Mbits/s) per UE time series.

Conclusions

In this section we evaluated an O-RAN–aligned Near-RT RIC control framework that combines WSI-driven beam targeting with CQI + 5QI aware scheduling. The solution is implemented as 2 xApps (beamforming and scheduler) supervised through KPI aggregation, forming a closed-loop stack that first stabilizes link quality and then uses QoS class information to steer resources toward stricter flows.

Across all scenarios, the results are consistent. Enabling WSI-driven beamforming (WSI-OFF → WSI-ON) is the primary performance lever, increasing SumDL by ~27.8–30.8% across policies, confirming that improving the physical layer (via faster, zone-aligned beam selection) is the first-order requirement under dynamic indoor CQI variability. In parallel, 5QI-Sat increases by ~41.1–45.8%, because stricter QoS classes become schedulable more often once low-CQI episodes are reduced.

Within a fixed WSI state, scheduler policy has limited impact on throughput but clearly affects QoS outcomes: balanced (CQI+5QI) and especially QoS_first improve 5QI satisfaction compared to cqi_only, with QoS_first achieving the highest 5QI-Sat without a throughput penalty in the tested setup. Overall, the evaluation supports a clean split of roles: WSI-driven beamforming delivers the main throughput lift by stabilizing CQI, while 5QI-aware scheduling converts the recovered capacity into improved QoS satisfaction.

5 End-to-end evaluation results for ISAC services

This chapter presents the evaluation of the E2E performance of ISAC services supported by the 6G-SENSES architecture. The focus is on assessing how AI/ML-driven automation can enhance the joint provisioning of communication and sensing across the network.

Specifically, section 5.1 evaluates a DRL–based approach for E2E slice orchestration, demonstrating how dynamic control of network resources can reduce energy consumption while simultaneously satisfying the QoS requirements of both communication and sensing slices. Section 5.2 complements this evaluation with a system-wide techno-economic analysis, assessing the potential benefits, trade-offs, and implications for supporting ISAC services in large-scale 6G network deployments.

5.1 AI-driven Network Slicing in support of ISAC services

The introduction of ISAC services, as proposed by the 6G-SENSES architecture, will bring new capabilities to enhance network functionality, but at the same time will introduce additional requirements on the underlying mobile network infrastructure. The need to transfer gigabytes of I/Q-echo streams from the RU to the DU or even the CU, to be processed by sensing functions, will significantly increase both throughput and processing requirements. At the same time, sensing and communication services sharing the physical network resources will impose diverse QoS requirements that should be fulfilled. For example, a communication service typically requires specific throughput and E2E delay for its end-users, while a sensing service may on top of that demand high sensing accuracy or even stricter E2E delay constraints.

The Network Slicing paradigm offers the required flexibility to host services tailored to the specific needs of different verticals with diverse QoS requirements [43]. It facilitates the creation of multiple logically independent virtual networks, called slices, on top of a shared physical network infrastructure. Each slice may have different E2E SLAs that should be satisfied, and the network operator pays penalties in case of SLA violations.

Considering that traffic demands are time-varying or even unknown, and that slices compete for common network and compute resources, intelligent AI/ML-based orchestration of E2E communication and sensing slices becomes crucial. Such orchestration must ensure both the per-slice QoS, as determined by its E2E SLAs, and the overall energy and resource-utilization efficiency of the system.

In this section we investigate dynamic network slicing for Integrated ISAC using DRL at the SMO layer. The DRL agent observes both the current (and potentially recent) throughput requirements of the communication slice, sensing-related information that determine the throughput of the sensing slice, and the network configuration. Based on these observations, it anticipates future demands and selects an appropriate new network configuration. This includes both sensing-related parameters and the embedding of slices onto the physical network (i.e., mapping virtual to physical nodes).

As an example, consider a scenario where traffic demand is initially low, enabling the agent to place the Virtual Network Functions (VNFs) of the two slices on a small number of servers to save energy. As demand increases and congestion becomes imminent (leading to potential SLA violations), the agent has two options: i) increase the sensing SNR threshold in selected BSs, which reduces the fraction of reflected I/Q echo streams transported to the sensing app for processing (**D2.2**, Section 7.3), and therefore slightly sacrifice sensing accuracy to reduce the load; or ii) migrate VNFs to previously inactive servers (a costly action that may be suboptimal in the short term but beneficial if high traffic persists). Through these dynamic decisions, the agent continuously balances energy efficiency, migration cost, and service quality across both slices.

In what follows we will describe the system model considered and the RL problem formulation.

5.1.1 System model

Physical Network

The physical network comprises the RAN, MEC, Transport, and CN domains, as illustrated in Figure 5-1. The RAN consists of BSs, each hosting an integrated Radio Unit / Distributed Unit (RU/DU). The MEC is located close to the RAN and offers processing capabilities through servers running containerized RAN VNFs with stringent delay requirements, such as the CU and the sensing function. The CN is farther from the RAN and includes servers hosting the CN Functions (for simplicity and without loss of generality we consider just the UPF in this scenario).

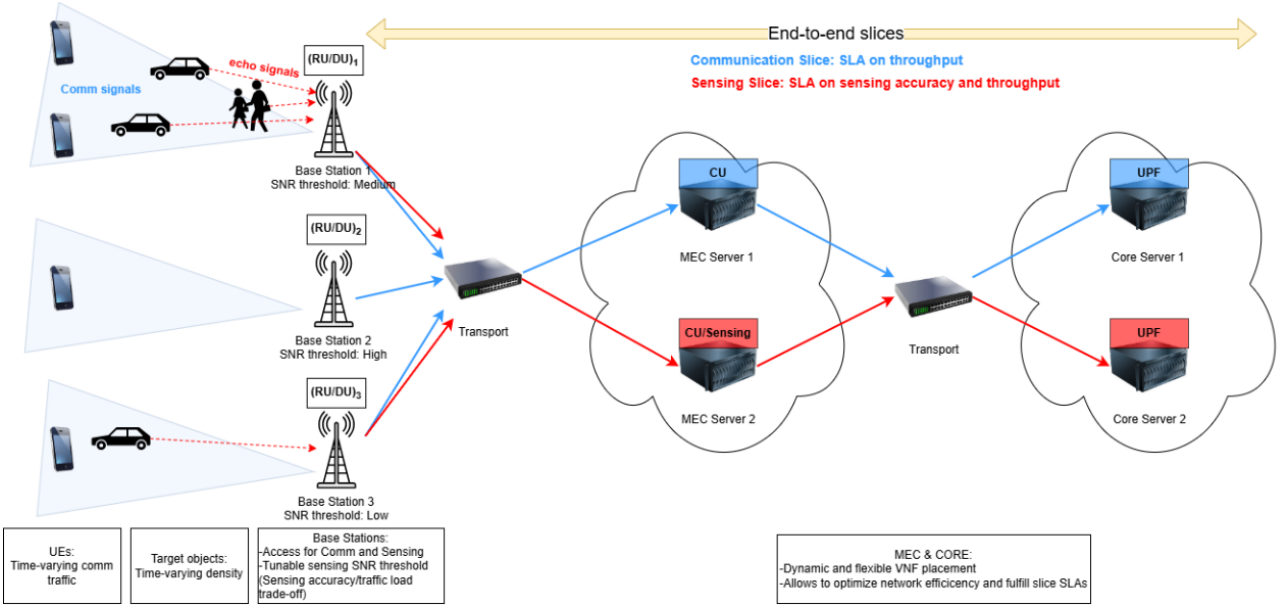


Figure 5-1 System model overview

The transport network interconnects the RAN with the MEC and the 5G CN domains, and it may use two types of links: either an optoelectronic link or an all-optical fast link. We assume that the routing of the flows within the transport network is handled by a lower-level control entity and is outside the scope of our higher-level orchestrator. So, we abstract away all internal switching and interconnection details and consider only the resulting E2E transport characteristics. Each node (BS or server) and each link is associated with some throughput capacity C_e, which indicates the maximum throughput that the element can serve. Additionally, each network element introduces some delay (e.g. queuing, processing, propagation).

Network Slices

We consider two types of slices, a communication slice and a sensing slice. Each slice is a VNF chain, i.e., a directed graph where the nodes represent the processing stages required for each flow (e.g. RU, DU, CU, sensing function, UPF), and the links determine the order that this processing must take place. Time is slotted and at each time slot t throughput demands T_r^{sens} and T_r^{comm} arrive at each BS r for the sensing and communication slices respectively. These translate into respective throughput requirements for each subsequent VNF and Virtual Link along the VNF-chain. A VNF $n \epsilon N$ may optionally apply a throughput-transformation factor $\gamma_n \in [0,1]$ to model processing steps that reduce (or preserve) the data volume they forward to the next stage.

Regarding the topologies of the two slices, as it can be seen in Figure 5-1, the communication slice is a chain of RU/DU→CU→UPF, while the sensing slice is a chain of RU/DU→CU/sensing→UPF.

Moreover, each slice has a number of specified SLAs. Both types of slices may come with SLAs on throughput and delay. Any SLA violations are penalized by linear, or even quadratic costs on the excessive metric. The slice-specific E2E delivered data rate (throughput) equals the minimum rate supported by the most congested physical network element along the VNF chain. If the capacity of the physical node or link is at least equal to the sum of the corresponding individual demands by the collocated VNFs or VLs then the throughput requirements are satisfied, otherwise the data rate is reduced proportionally among the collocated elements. Regarding the per slice E2E delay, this is the total delay accumulated across all nodes and links (sum of queuing, processing, and transport delays; includes a load-dependent term). Finally, the sensing slice has additionally an SLA on the sensing accuracy, which is determined by the sensing throughput that will be forwarded to the sensing app (the higher the throughput per BS and the more BSs involved in sensing the higher the accuracy).

Traffic modeling

The throughput $T_r^{comm}(t)$ of the communication slice in our experiments will be driven by a real-traffic dataset (Milano dataset) [44]. The throughput at each BS is essentially a normalized trace taken from this dataset that demonstrates variations based on user mobility and activity.

The load of the sensing slice $T_r^{sens}(t)$ is a fraction of the RU→DU throughput of the communication slice. We assume that the RU→DU throughput is constant as it depends merely on numerology and not on traffic load. The variations of the sensing throughput depend on two factors:

- External factors, beyond the control of the orchestrator, such as the position and number of detected target objects and the position of the connected UEs.
- A controllable parameter, namely the sensing SNR threshold θ_r at each BS r. Setting a high θ_r results in dropping I/Q-echo streams with a lower SNR and therefore not forwarding them from the DU to the sensing app for processing. This reduces the throughput demand of the sensing slice but can potentially degrade sensing accuracy (the modeling of the sensing accuracy will follow shortly).

We use a logistic (sigmoid) function to model the fraction of reflected I/Q-echo streams whose SNR exceeds the sensing threshold. This choice is justified by the fact that the SNR of multipath reflections in practical environments follows a composite fading process that combines large-scale shadowing (log-normal) with small-scale multipath fading (Rayleigh/Rician/Nakagami), resulting in composite SNR distributions [45]. These distributions typically exhibit a smooth but relatively sharp transition in the proportion of echoes that lie above or below a given threshold. Accordingly, the sensing throughput generated at BS r at time slot t can be modeled as:

$$T_r^{sens}(t) = \beta FH_r \cdot min(1, \alpha \cdot O_r(t)) \cdot (1/(1+e^{k(\theta_r(t)-\theta_{cut})})),$$

where:

- FH_r: is the Fronthaul (FH) traffic of the communication slice at BS r. We assume it is constant (depends only on numerology and not traffic load). Typically, FH_r is order(s) of magnitude larger than the communication slice throughput.
- β: represents the structural sensing fraction, i.e., the maximum portion of the FH I/Q stream that can ever be sensing-relevant. Therefore, βFH_r boils down to: if object activity is maximum and the SNR threshold is at its lowest, what is the maximum I/Q data rate that could be forwarded?
- $O_r(t)$: Normalized number of target objects detected at RU r and time t. It may evolve according to a stochastic process (e.g. Markov) or imported from real data.

- α: Scaling factor to convert object activity into the number of strong radar echoes. Determines how aggressively object count increases throughput.
- $\theta_r(t)$: SNR threshold for BS r in time slot t.
- θ_{cut}: Critical threshold where half the I/Q echo streams pass.
- k: Slope (how fast throughput drops around θ_{cut}).

In Figure 5-2 we plot an example of how the sensing throughput changes with the SNR threshold and different object activities.

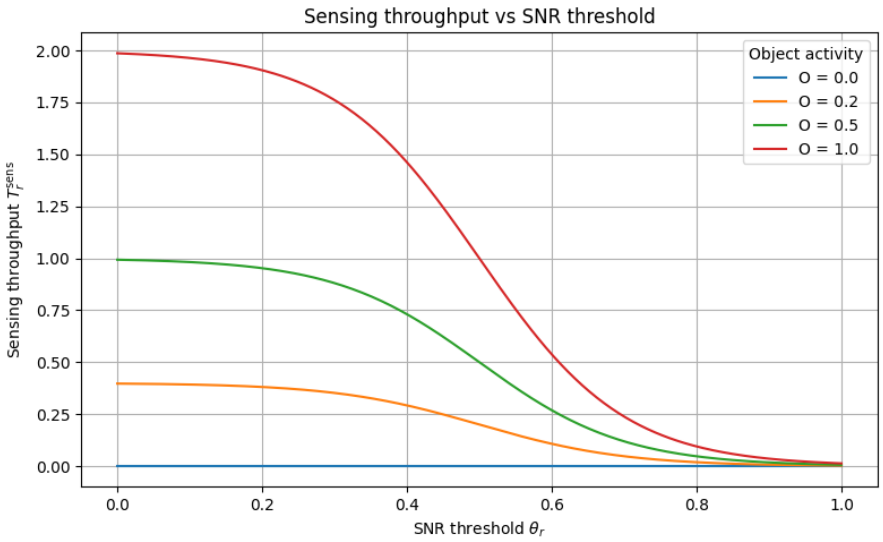


Figure 5-2. Plot of the sensing throughput at a BS as a function of the corresponding SNR threshold for different object densities.

Sensing Accuracy modeling

We model sensing accuracy both per BS and globally as a monotonic, saturating function of the sensing throughput (after filtering out the I/Q echo streams that fall below the selected SNR threshold per BS). For BS r at time slot t, the local sensing accuracy is:

$$Acc_r(t)=1\text{-}e^{-cT_r^{sens}(t)} \quad (5\text{-}1)$$

where $c>0$ controls how fast accuracy saturates with the sensing throughput (additional retained I/Q-echo streams). This choice captures the well-known ISAC behavior that keeping more high-SNR echo samples improves detection quality, but with diminishing returns once sufficient sensing information has been collected (low SNR I/Q-echo streams don't improve much sensing accuracy). In Figure 5-3 we plot the sensing accuracy function for different saturation factors c.

To capture multi-BS sensing gains, we define a global accuracy by applying the same formula to the aggregate sensing throughput over all BSs:

$$Acc_{glob}(t)=1\text{-}e^{-c\sum_{r\epsilon R}T_r^{sens}(t)} \quad (5\text{-}2)$$

reflecting that sensing information from different BSs can add up and improve accuracy but after a point the global accuracy also saturates.

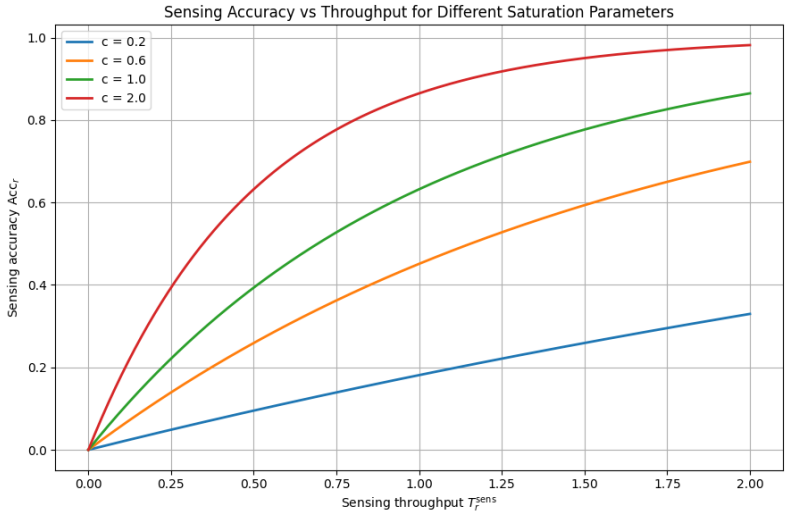


Figure 5-3 Plot of the sensing accuracy as a function of the sensing throughput for different saturation factors c. Sensing accuracy increases with throughput (fraction of the FH traffic) but eventually saturates, as low SNR I/Q-echo streams don't significantly improve it

All the above sensing accuracy functions are needed in the RL reward to enforce both global performance and per BS fairness. As an example, each BS should be able to capture the strong reflections by the objects that are located within its area but the utilization of multiple BSs that capture also reflections from different angles can boost performance. What we want to avoid is to minimize the SNR threshold of a BS and starve the rest.

5.1.2 RL formulation

Time is slotted, and the RL agent decides for a network configuration at the beginning of each slot. Since the problem requires a global, E2E view of the network, and typically VNF migration decisions occur in non-real time intervals, the agent can be hosted the SMO layer.

At each time slot t, the agent observes the system state s_t, which includes:

- the average traffic demand (throughput) $T_r^{comm}(t)$ of the communication slice at each BS r measured over the previous time interval $[t\text{-}1, t]$, and optionally the N most recent historical traffic demands ($T_r^{comm}(t\text{-}1), T_r^{comm}(t\text{-}2), \ldots, T_r^{comm}(t\text{-}N)$) or a respective running average,
- the number of detected objects (object density) $O_r(t)$ per BS r over the same horizon, considering that each detection is associated with the closest BS; this is an elegant way to encode both object count and spatial distribution,
- the current VNF embedding, i.e., placement of all VNFs (DU, CU, UPF, sensing functions) to servers.

Based on this state, the agent selects an action a_t, that defines:

- A new embedding for each VNF (placement of VNFs to servers),
- The SNR threshold for the sensing slice, controlling the trade-off between sensing accuracy and processing load,
- (Optionally) the transport-network type (all-optical or opto-electronic) assigned to each slice.

This configuration is applied to the network during the subsequent time window $[t, t+1]$.

At $t+1$ the new traffic demands are revealed, and the agent receives a reward r_{t+1}, reflecting the performance achieved by the configuration a_t under the actual load observed in $[t, t+1]$. The reward is defined as the negative weighted sum of the following cost components:

- cost proportional to the number of active servers and physical links (energy consumption),
- penalty for VNF migrations between timeslots,
- penalties for sensing slice SLA violations (e.g. throughput, delay, sensing accuracy),
- penalties for communication slice SLA violations (e.g. throughput, delay).

Consequently, the agent must make each decision based solely on past observations, as future traffic conditions are unknown. Through repeated interaction with the environment, the agent gradually learns a policy $\pi(a_t|s_t)$ that maximizes the expected cumulative reward (or, equivalently, minimizes the cumulative cost) over time. In doing so, it implicitly learns the underlying system dynamics and balances energy efficiency, migration overhead, and service quality across both slices.

DRL agent

Due to the combinatorial nature of the problem at hand, tabular RL solutions with convergence guarantees, such as Q-learning, are impractical and suffer from exploding Q-value (state-action) look-up tables [46]. Such agents impose huge memory and computational requirements, while their convergence is extremely slow (both in terms of time and sample efficiency). Therefore, we need to rely on function approximation-based RL methods that may lead to suboptimal solutions but converge in reasonable time. To this end, we employ Deep Q Learning (DQN)-based methods [47] that are well established and known for their sample efficiency. Although the action space is still manageable in our two-slice setting, the state space is effectively continuous due to the real-valued traffic demands, making approximation necessary even in a single-agent design. For scenarios involving a large number of slices a multi-agent approach would be required.

5.1.3 Evaluation Results

We will now evaluate the performance of the proposed DQN-based dynamic orchestration policy in an ISAC scenario comprising one communication slice and one sensing slice. The goal is to assess the ability of the learning-based policy to jointly manage heterogeneous slice requirements under time-varying traffic, and to compare its performance against a set of carefully designed static baseline policies.

System Setup

The RAN consists of three BSs, while the MEC and the CNs comprise two servers each (Figure 5-1). The SNR threshold at each BS is quantized and may take one of three discrete levels: low, medium, or high. Moreover, CU/sensing VNFs can be placed only at the MEC, while UPFs can be placed only at the Core. Under these constraints, and accounting for all possible VNF placements and sensing configurations, the agent has a total of 432 possible actions.

-Traffic details. Both the traffic demands of the communication slice and the object activity of the sensing slice are imported from the real-traffic Milano dataset [44]. Since object activity is often correlated with the traffic demand, we choose highly correlated traces from the Milano dataset. This doesn't limit the applicability of our method, as the object activity could be driven by different random processes (e.g. Markov) in future experiments.

-Slice SLAs. The sensing slice is associated with two SLAs: one on throughput and one on sensing accuracy, considering both local (per BS) and global accuracy. The communication slice, on the other hand, is associated with a single SLA on throughput. Note that the required throughput for each slice is not fixed but varies

dynamically according to the traffic demand. Whenever the required throughput is not met, an SLA violation penalty is incurred. Additional SLAs could be incorporated in future extensions.

Static Baselines

We compare the proposed DQN policy against six static baseline policies, grouped into two families, each highlighting different operational trade-offs.

-Opt. Energy. This family of static policies places the VNFs on the minimum possible number of nodes, minimizing energy consumption but potentially incurring throughput SLA violations. Combined with different sensing configurations (SNR thresholds), this results to the following policies:

- **Opt. Energy - Min. Sensing:** SNR thresholds are set to high values, causing almost all reflected I/Q streams to be dropped by the sensing application. This minimizes sensing accuracy while further reducing network load.
- **Opt. Energy - Balanced Sensing:** SNR thresholds are set to medium values, aiming to strike a balance between sensing accuracy and network load.
- **Opt. Energy - Max. Sensing:** SNR thresholds are set to low values, maximizing sensing accuracy but significantly increasing sensing traffic, which may lead to throughput SLA violations.

-Opt. Throughput. This family of static policies distributes VNFs across all available network nodes, minimizing throughput SLA violations at the expense of higher energy consumption. As above, three sensing configurations are considered:

- **Opt. Throughput - Min. Sensing:** High SNR thresholds.
- **Opt. Throughput - Balanced Sensing:** Medium SNR thresholds.
- **Opt. Throughput - Max. Sensing:** Low SNR thresholds.

Training

We train the DQN agent over 10 individual runs with different random seeds. Each run consists of 200 episodes, and all reported cost-related results are averaged across the 10 runs. Regarding exploration, we apply an exponentially decaying exploration rate, starting from 1 at the beginning of the training to 0.001 near the end (episode 180). The final policy used for evaluation corresponds to the policy obtained at the end of episode 200.

Results

The convergence behavior of DQN during training is depicted in Figure 5-4, which confirms the ability of the agent to improve its policy by interacting with the system and finally learn an effective dynamic orchestration policy that outperforms all static baselines. DQN starts with a relatively high-cost policy (left side of the plot) but steadily reduces the normalized cost over time and converges to a value significantly lower than all static baselines. The reported cost is a weighted sum of multiple components, including SLA violation penalties, the number of active nodes, and VNF migration costs, and therefore captures the overall system performance.

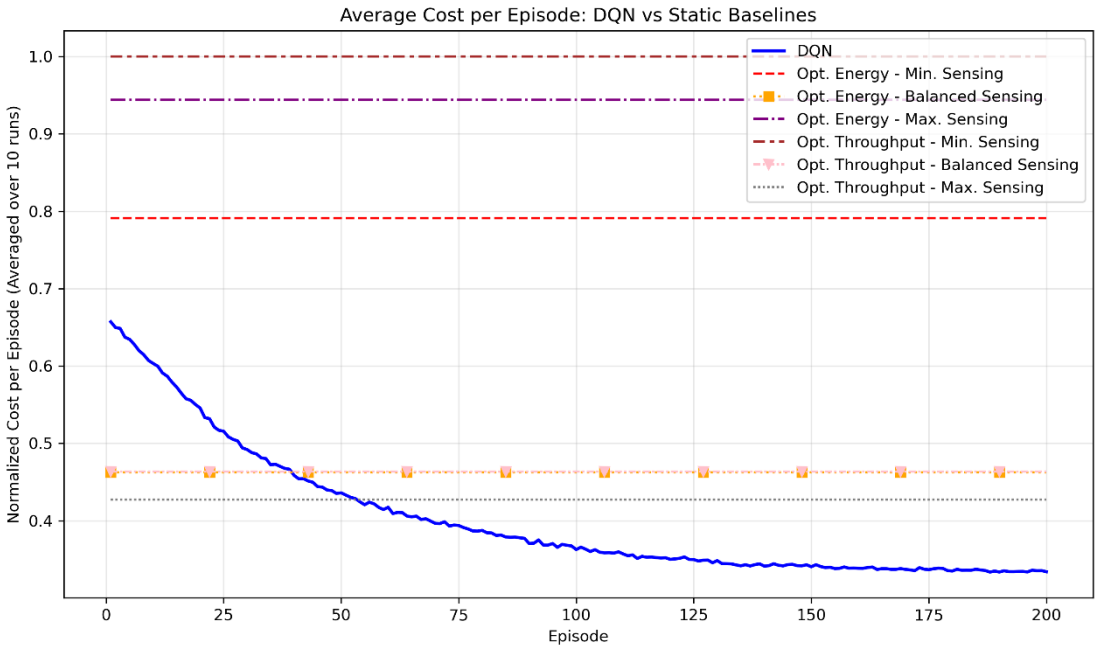


Figure 5-4 DQN convergence plot and comparison with static baseline policies.

To further analyze the obtained DQN policy we provide Table 5-1, which summarizes the normalized performance metrics across all policies, including overall cost, sensing and communication slice performance, and network resource utilization.

Table 5-1 Evaluation results summary. All the performance metrics reported below are normalized, while the respective DQN metrics correspond to the final obtained policy at the end of the training phase.

		Overall	**Sensing Slice**				**Comm Slice**		**Network**
	Policy	Cost	Data rate	Data rate violations	Sensing acc.	Sensing acc. violations	Data rate	Data rate violations	Active nodes
Dynamic Policy	**DQN**	**0,329**	0,514	0,055	0,796	0,637	1,000	0,029	0,555
Static Baseline Policies	**Opt. Energy Min. Sensing**	0,791	0,007	0,000	0,021	1,000	1,000	0,000	0,500
	Opt. Energy Balanced Sensing	0,463	0,501	0,307	0,820	0,750	0,968	0,307	0,500
	Opt. Energy Max. Sensing	0,944	0,802	1,000	1,000	0,000	0,780	1,000	0,500
	Opt. Throughput Min. Sensing	1,000	0,007	0,000	0,021	1,000	1,000	0,000	1,000
	Opt. Throughput Balanced Sensing	0,463	0,521	0,000	0,820	0,750	1,000	0,000	1,000
	Opt. Throughput Max. Sensing	0,428	1,000	0,246	1,000	0,000	1,000	0,000	1,000

The results highlight the advantages of the proposed DQN-based dynamic policy compared to the static baselines. Overall, DQN achieves roughly a 30% improvement compared to the best static baseline (Opt. Throughput-Max Sensing policy incurs 30% higher cost than DQN), demonstrating its ability to jointly balance energy consumption and slice-level QoS. While energy-optimized static policies minimize the number of active nodes, they do so at the expense of slice performance, leading either to severe sensing SLA violations (min. sensing variant) or to high throughput SLA violations (max. and balanced sensing variants). On the other hand, throughput-optimized static policies successfully satisfy communication SLAs but consistently over-provision network resources by activating all available nodes, resulting in higher operational cost.

In contrast, the DQN policy dynamically adapts both the number of active nodes and the sensing configuration. As a result, it manages to maintain high communication throughput without inducing significant SLA violations, while simultaneously achieving high average sensing accuracy. Although some sensing SLA violations are observed, their magnitude is small, indicating that the learned policy operates close to the sensing SLA boundary in order to reduce cost. While the two balanced-sensing static variants may have slightly higher average sensing accuracy than DQN, in practice they still demonstrate higher sensing accuracy SLA violations. DQN often avoids such violations by reducing the SNR thresholds when traffic demands of the communication slice are low.

This dynamic behavior is also reflected in network resource utilization. The DQN agent activates significantly fewer nodes than throughput-optimized policies, and only slightly more than energy-optimized ones, while achieving a QoS performance that is close to the former. This demonstrates that DQN effectively adapts resource usage to current traffic conditions

Overall, these results indicate that the DQN agent successfully learns an efficient dynamic policy that balances energy efficiency, sensing accuracy, and communication QoS, outperforming static policies that are inherently limited to single-objective optimizations.

5.2 Techno-economic analysis

5.2.1 Scope and objective

It is a fact that 6G network services will pose stringent coverage and data rate requirements to access network segments. In parallel, the inclusion of ISAC as main capability of 6G access networks can pose new challenges in the deployment of the networks – related to sensing performance, sensing coverage, capacity requirements for sensing etc. On the other hand, AI/ML techniques can play a key role in operational network deployments. This landscape of 6G requirements on one hand and enablers on the other, can lead to very dissimilar network deployments in terms of technologies used, performance achieved and capabilities that are enabled.

To this end, for transport network segments it will be necessary to support high capacity macro-coverage Radio Units, dense layers of high capacity micro-coverage Radio Units, or/and versatile Distributed and Cloud-based Radio Access Network function (D-RAN, O-RAN) setups, while transporting sensing alongside communication services. For this purpose, the transport network shall be equipped with mechanisms to support flexible and scalable access network deployments, to converge fronthaul (FH) and backhaul (BH) traffic of various functional splits (FS) alongside sensing-related channels over a single infrastructure consisting of various wireless and optical technologies.

In this context, the delivery of radio access and transport network deployments that are capable to serve the 6G service requirements is not a straightforward task. In practice, many factors shall be considered, such as area specifics, deployment feasibility, long-term service roadmaps, traffic demand and growth patterns/forecasts, as well as infrastructure availability, scaling capabilities and the implicated costs. The

scaling capabilities and the identification of the implicated costs shall take into consideration the various deployment phases (over time), in order to pre-estimate the critical, high cost factors and to extract deployment guidelines at network pre-planning stages – prior to the definition of the specific network deployments to be realized.

In this context, the techno-economic evaluation of large-scale alternative network deployments is necessary to identify cost optimizations and to investigate various technological aspects and critical parameters (e.g. Sensing technologies, sensing functions placement, sensing-related performance parameters) at early stages towards delivering economically viable and sustainable deployments.

5.2.2 Specification of dimension

The techno-economic evaluation of 6G-SENSES addresses the infrastructure and network deployment (level) as this constitutes the main, necessary investment for any service provisioning on top at the premises of the verticals under study. The key factors (KPIs) that are evaluated are:

- CAPital EXpenditures (CAPEX) of alternative network deployments
- OPerational EXpenditures (OPEX) of alternative network deployments
- Total Cost of Ownership (TCO) of alternative network deployments.

These factors – KPIs, are broadly used in the conduction of techno-economic studies. The scope of the study in the context of the project is to assess comparable deployments (with specific coverage, capacity, sensing-enabling capability, AI/ML techniques incorporated in RAN operation etc.) and analyze the results to optimize techno-economic factors, given certain service levels.

5.2.3 Methodology

In general, performing techno-economic analysis is seldom a straightforward task, the underlying reasons being many and versatile. Some of these reasons are related to the scope and goal of the analysis, the level of detail of input information in terms of deployment area specifics, the network segment and the technologies in focus, the scale of the system and so on. Therefore, there are various types of analysis applied at different stages of a solution commercialization, with different focuses depending also on the entity performing the analysis (external or internal department, strategy/financial or technical department etc.).

In the context of 6G-SENSES, the methodology that has been implemented with the technoeconomic analysis tool comprises the following steps:

Step 1. Definition of the area / vertical premises under study. This step aims to define the coverage area and model the physical / vertical premises environment specifics. 6G-SENSES analysis focuses on two main environments: a large-scale clutter-ignorant deployment, and a hotspot (mall/ stadium type) deployment.

Step 2. Definition of the service scenarios in terms of coverage area, traffic demand to be served, sensing coverage area, sensing performance parameters etc. and their scaling over years.

Step 3. Definition of the access and transport deployment blueprint, providing a general model of alternative standard RAN / O-RAN/ CF-MIMO RAN, Backhaul/ Fronthaul, Edge/5G CN Options that can co-exist in a deployment. At this point, alternative placement of sensing function options are modelled.

Step 4. Definition of the scaling rules for each technology and the dimensioning rules for each segment. In practice this includes:

- RAN segment dimensioning:

- as access network nodes' elements calculation based on capacity and coverage increase over years; considering access radio units capacity, gNBs disaggregation model, existence or not of AI/ML at RAN operation etc.
- as RAN sensing elements calculation based on sensing coverage increase over years;

- Transport segments dimensioning: as capacity and required links calculation considering the adoption of various Functional Splits, various sensing channel requirements, various transport aggregation levels, various transport technologies link capacity/ range/ hops, etc.
- Edge segment dimensioning: as compute resources calculation considering the application services and Network Functions requirements, loading factors, etc.
- Core segment placement and dimensioning of UPFs to serve the Edges.

Step 5. Definition of the cost information and its scaling rules for all elements that are modelled/analyzed.

Step 6. Setting of analysis timeframe; commonly set to 5-10 years for telecom network deployments.

Step 7. Definition of the deployment scenario in terms of technologies mix, including selection of Functional Splits, wireless / optical technologies, etc.

Step 8. Iteration of all steps and collection of analytical cost results, in order to identify key factors influencing cost and extract deployment recommendations.

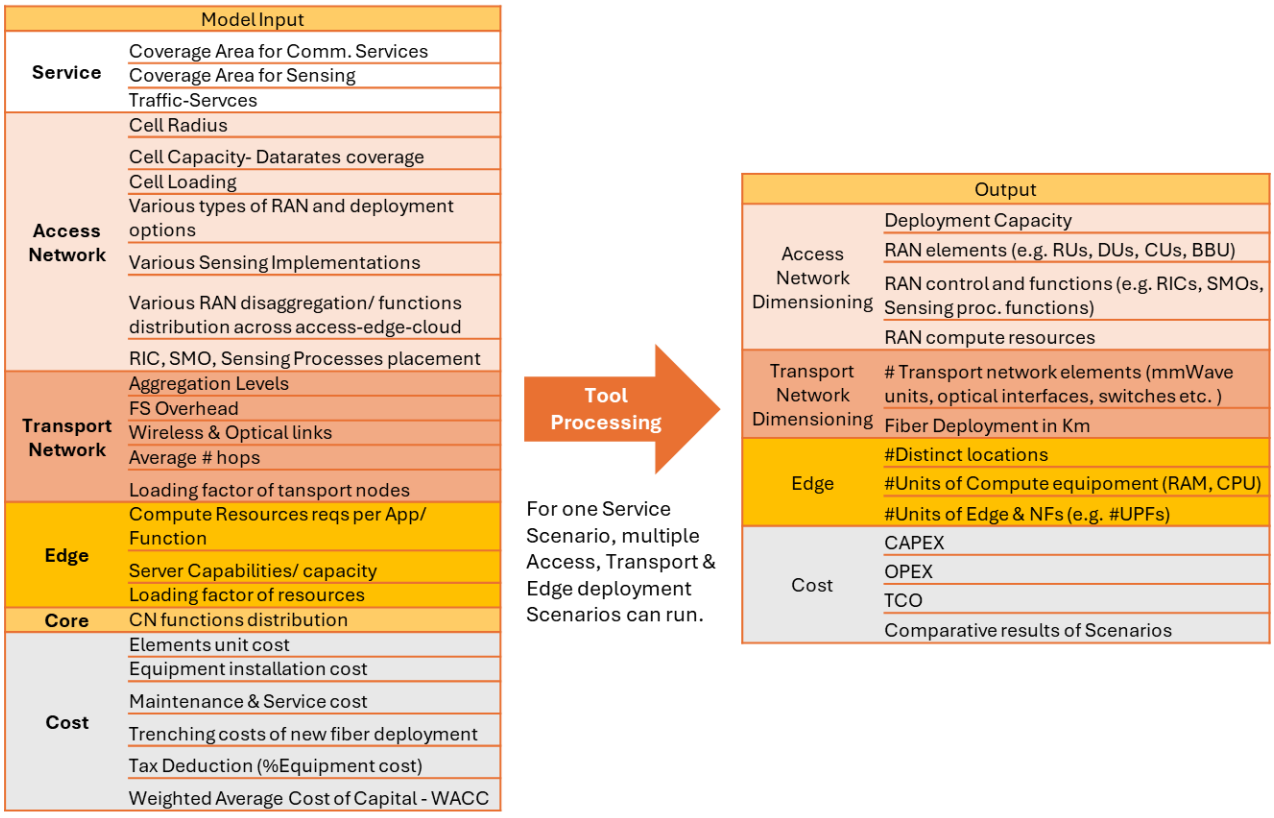


Figure 5-5 6G-SENSES techno-economic Analysis Input and Output.

5.2.4 6G-SENSES Deployment Blueprint and Modelling Aspects

The techno-economic approach outlined in Steps 3 and 4 involves modelling the physical architecture of a 6G network, which includes the essential RAN with or without sensing capabilities, various transport network segments (access, aggregation and core), as well as edge and core components. The foundational principles guiding 6G network architectures assume multiple network and computing infrastructure configurations at a highly granular level, incorporating disaggregated pools of networking, computing, and storage resources. The baseline principles for both physical and logical/functional architectures, also adopted by 6G-SENSES, have been thoroughly detailed in deliverable **D2.2** [1]. The deployment options implementing the high-level architecture can be many.

The principles and these deployment options are reflected in a general deployment blueprint. More specifically, the model encompasses a wide range of RAN configurations to support advanced deployment scenarios. These include non-3GPP access using IEEE 802.11ad as a RU integrated to the 6G network via N3IWF, traditional gNB or disaggregated RAN setups with or without sensing capabilities, Sub-6 GHz RUs with integrated sensing, and O-RAN architectures enhanced with sensing functionality. The blueprint takes also into consideration the widely adopted network roll-out principle for transport network deployments that comprise multiple hierarchical transport network segments. In general, considering the current trend of minimizing the number of segments from the access to the CN, the transport network model comprises three levels/ segments – the access transport segment, the aggregation transport segment and the core transport segment. The overall blueprint that has been used in 6G-SENSES techno-economic analysis is Figure 5-6.

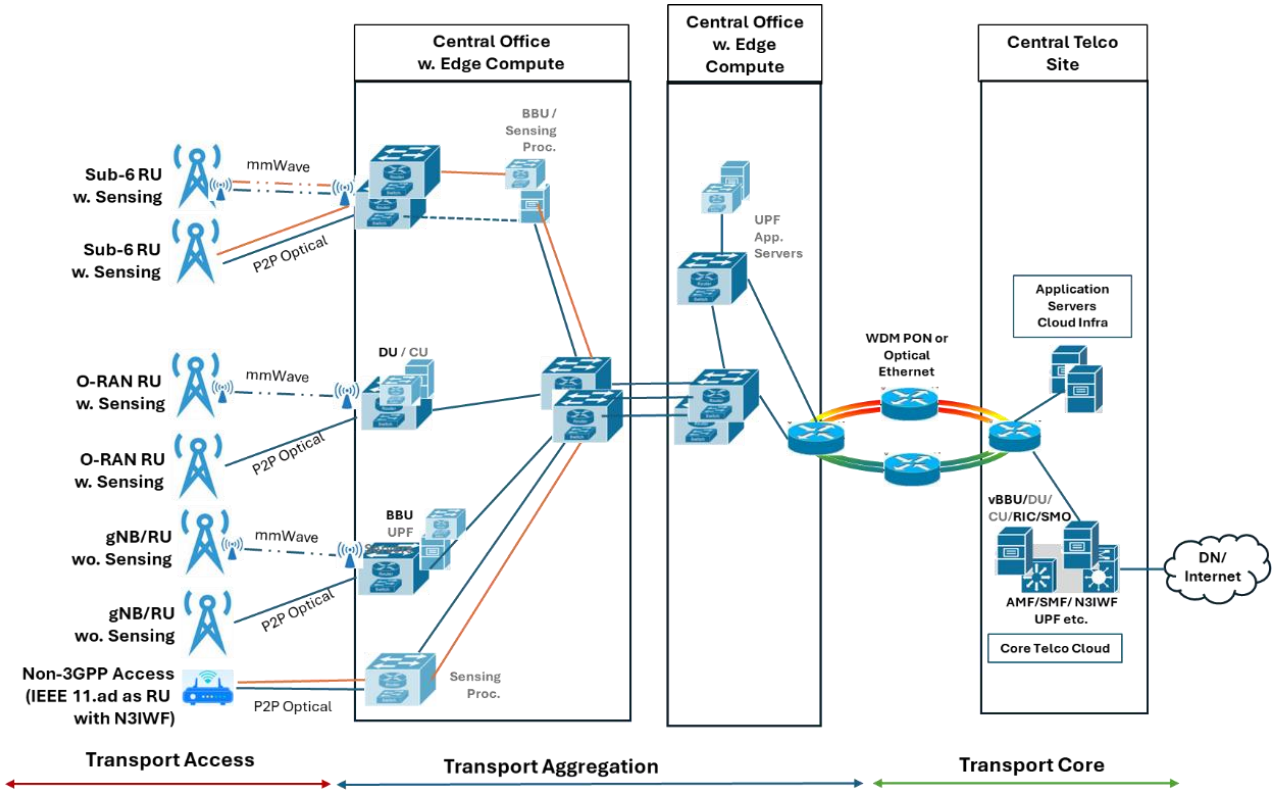


Figure 5-6 6G-SENSES physical architecture blueprint

Access Network Deployment Principles/Assumptions

Especially, considering only the provisioning of communication services the 6G access network comprises a dense layer of short-range access network nodes (RRUs and gNBs) to serve the high traffic demand at specific hotspot areas and vertical premises. This layer is complemented by a macro layer of longer range access

network nodes (RRUs or gNBs) for maximizing coverage over the wider area. It shall be noted that the discrimination between RRUs and gNBs is modelled with the transport network modelling where various functional splits (FSs) are considered; at radio network the access nodes' radio coverage characteristics are modelled. Considering the provisioning of sensing capabilities, multiple technology options can be considered in 6G-SENSES namely: O-RAN gNBs with sensing capabilities, non-3GPP access network nodes with sensing capabilities (based on IEEE 11.ad) integrated to the 6G network via N3IWF, and standard 3GPP RAN nodes (sub-6 based) with sensing capabilities.

The key technical characteristics of the radio access nodes have been modelled are related to:

- Capacity of RUs – vendor specific input.
- Data rates achieved at cell-edge and range of cell-edge; Network Operator specific constraint deriving from communication service-driven radio network planning.
- Maximum expected loading factor of RUs – Network Operator specific constraint.
- Scalability of various Radio elements and functions (i.e. ratios of: RU:BBU, RU:DU:CU, Antennas:RU)
- Scalability of Radio Control functions, i.e. RIC, SMO, etc.
- Different Functional Splitting (FS) options (referring to eCPRI A to E options).

The access network deployment dimensioning is performed considering generic radio-coverage calculations for a certain area especially given a service to be provided at cell edges, as well as forecasts of traffic demand density.

Transport Network Deployment Principles/Assumptions

Adhering to general transport network deployment principles, the transport network comprises three main segments:

- Transport Access, providing connectivity from the access network nodes to the 1st level aggregation sites (Central Offices).
- Transport Aggregation, aggregating transport access connections at next hop Central Offices.
- Transport Core (3rd level transport aggregation) is normally used in large scale, public networks, providing connectivity/routing traffic between COs to the CN).

Transport Network Deployment Options

In general, the data-plane architecture considers an integrated optical (P2P or PON based) and wireless (mmWave, Sub-6) network topology and infrastructure to support jointly backhauling of gNBs, and, fronthauling of various functional Splits of Remote Radio Units (RRUs) to Baseband Units (BBUs) or O-RAN DUs/CUs. The 6G-SENSES transport access deployment options are related to the incorporation of sensing capabilities at RAN (at various sensing service/availability levels), to the usage of different FSs, to the usage of various wireless and optical technologies and to other deployment specificities.

Deployment Options related to Multiple FSs

More specifically, the 6G-SENSES concept, architecture and functionality support the coexistence of different FSs over the same infrastructure, namely eCPRI A, eCPRI B, eCPRI C, eCPRI D, eCPRI Id, eCPRI IID/IU, eCPRI E. In practice, the throughput overhead of each functional split needs to be considered in the dimensioning of the transport links between the access network node and the BBU. A simple formula, calculating the

transport links throughput on the basis of the access network nodes data bandwidth or capacity and of the splitting option overhead, has been used as follows:

T_FH=(aT_cell+bC_cell)(1+〖OH〗_(split_i))

where:

a,b = {0,1}.

Tcell : data throughput of the cell.

Ccell : maximum cell capacity.

OHsplit_i : overhead, depending by the adopted splitting option.

Indicative values of parameters for some functional splitting options are shown in Table 5-2.

Table 5-2 Indicative Functional Splits parameters

Splitting options	a	b	OH
Split E: CPRI / eCPRI	0	1	600%
Split II$_D$ / I$_U$: load-dependent FH	1	0	80%
Split A: BH with centralized scheduling	1	0	25%

Deployment Options related to Various Transport Access Technologies

Considering the general mobile networks' deployment models and the most widely considered technologies, the network deployment options and technologies scaling rules for the transport access are:

- mmWave links for BH and/or FH providing ~10 Gbps average data rates per link; depending highly on the spectrum (frequency, bandwidth), the distance between the two nodes and the radio environment. mmWave links are implemented by pairs of transceivers, while links' capacity scales by adding new pairs of transceivers.
- P2P optical or WDM PON connections for BH/FH links at 10 Gbps or 25 Gbps, given the wide penetration of this technology in the market, and the availability of the relevant information. Scaling of these links is performed by adding connections interfaces and connecting them to optical switches. The latter vary depending on the transport segment that are placed. To this end at transport aggregation level (Leaf Type) switches can be used aggregating 18 P2P interfaces of 10 or 25 Gbps to up to 6 aggregate interfaces of 40 Gbps. At transport core level, Spine type switches are considered switching/aggregating 32 interfaces of 40 or 100 Gbps.

The scenarios can differ in the grade at which each of these technologies is used.

The wireless transport connections may consist of one or multiple hops, defined as deployment options in the techno-economic analysis. Moreover, multiple links per connection/hop may be considered, in case the traffic to be served exceeds the capacity limits of the equipment.

The access transport links are aggregated at 1st level at COs (edge sites), and then at 2nd level towards the Central Telco Site where also the CN resides. At 1st aggregation level we consider 18 mmWave or P2P links aggregated to Leaf type optical switches. Spine type switches are considered to be used at transport core segments to aggregate, at 2nd level, Optical interfaces with flexible degrees of aggregation (reaching 32x 40Gbps or 100Gbps optical interfaces).

For all the transport segments, other deployment specifics are modelled, such as the maximum link utilization (usually set to 80%) for the access transport links, the maximum loading factor for the aggregation nodes connected interfaces, etc.

Deployment Options related to Fiber Deployment

Given the fact that the main technology competitors to wireless backhauling are the optical network technologies, as well as the fact that fiber deployment is continuously expanding to the deployments' last mile optical network technologies are modelled in the various transport segments. To this end, given that fiber deployment is not ubiquitous, instead it incurs extra costs where not present, various fiber deployment scenarios/options are considered - in terms of cost of civil works. Typically, two main scenarios:

- Greenfield scenario: This scenario assumes that there is no previously deployed infrastructure and it is necessary to consider not only the fiber as such, but also different elements that intervene like digging and trenching, law permits, technical personnel. The deployment cost depends highly on the type of area to be covered, as well as the applicable construction costs, which vary highly between countries/areas, etc.
- Brownfield scenario: In this scenario, the deployment assumes an existing fiber infrastructure, thus what is modelled is the aforementioned cost for the required extension of the fiber deployment.

Deployment Options related to Usage and Placement of vBBUs and MEC

Compute resources are present at various physical and logical locations of the network for hosting applications and/or network services such as vBBUs and UPFs. These are indicated as Edge Compute at locations denoted as "Central Offices (COs)". Similarly to large scale deployments' common practice, COs host the transport aggregation equipment.

By modifying the deployment parameters, different access network, access transport and aggregation transport deployment scenarios can be defined, dimensioned and evaluated.

Other Input – Modelling related Information

Space in Central Offices – Rack Footprint Cost

The footprint cost per year (OpEx) of a rack can to be considered, corresponding to expenses for cooling, personnel, security, etc.

Installation

Installation costs can be either modeled as a percentage of the equipment CapEx, usually 2%. In case the value is 0%, the cost is considered embedded into the CapEx.

Maintenance

In order to simplify the cost evaluation, expenditures for maintenance are assumed to be equal to 2% of CapEx for each year.

Cost information and Financial Figures

To proceed with the cost estimation of a specific deployment, input is required regarding the per-unit equipment cost and its escalation model, installation and maintenance/service costs for each technology, along with the estimated annual prices erosion or escalation. These costs are kept as Confidential among the Project members. Other financial figures introducing or representing extra costs such as Weighted Average Cost Of Capital (WACC) and Tax Deduction (% of Equipment Cost) are also considered.

5.2.5 Deployment Scenarios Modelling

The stepwise methodology was used to perform a technoeconomic assessment of the various network deployments at a wide dense urban area as an example case, for a 5-year timeframe.

Step 1 & 2: Definition of the area under study.

The first step for the evaluation is the definition of the area of deployment. The definition refers to the total surface (km^2) that needs to be covered. Based on the area definition and the assumed equipment used in the access network nodes, the radio coverage parameters (i.e. coverage radius of access network nodes) can be evaluated as average pre-planning assumptions.

Although the tool provides this option, given the fact that in the 6G era dense urban areas traffic demand will drive network planning, in the area under study we extract the RAN deployment on traffic demand basis. In particular, in the example case, as area under coverage/study we consider a dense urban area, over which a peak aggregate traffic demand data rate will reach 1Tbps at peak hour. RAN and transport network dimensioning is performed on this basis, while auxiliary information such as the current or tentative deployment of cabinets as "Central Offices" (COs) is also defined at this point. In the case of the area under study, the assumptions are presented in the following table.

Table 5-3 Model Area under study - Assumptions

Radio Network related Assumptions	
Dense Urban Area	100%
Total Traffic Demand (Peak Hour) Gbps	1000
Average Cell Range Type A	500
Average Cell Range Type B	200
Average Cell Range Type C	200
Average Cell Range Type D	70
Average distance to CO (m)	500
#1st Level COs	15
#2nd Level COs	2

Step 2: Definition of the service scenarios in terms of coverage and traffic demand.

At this step, also the targeted radio coverage and traffic serving assumptions are recorded. As aforementioned we base the study on service demand. Despite the traffic demand we model scenarios that consider the capability or not of the network to provide sensing (in the form of ISAC adhering to the 6G-SENSES technologies/options).

Step 3: Definition of the Access and Transport deployment blueprint.

This step is based on the 6G-SENSES deployment blueprint, with the following area specific assumptions:

- The COs host vBBUs and UPFs. We model 15 in total COs, which host vBBUs/ DUs/CUs, while in 5 of them a UPF is also present to allow for traffic offloading at the edge.
- At access network segment. The model also considers various RAN deployment options depending on the level of usage of the various types of RAN, and the level of support of sensing (ISAC):
    - Type A RAN represents a standard 3GPP gNBs model (composed of RUs and BBUs or silo gNBs) with no sensing capabilities, but with high RU aggregation capabilities reflecting the

commercial large vendor trends in B5G RAN. We consider this type of RAN to serve the high traffic demand at specific hotspot areas and vertical premises.

  - Type B RAN represents an O-RAN gNB model composed of RUs/DUs/CUs, with sensing capabilities but with less aggregation capabilities.
  - Type C RAN represents an ISAC (3GPP or ORAN) RAN node with sensing capabilities.
  - Type D represents a non-3GPP (based on IEEE 11.ad), ISAC enabled node that can be connected to the CN via a N3IWF, while communicating sensing information to the O-RAN RIC.
- To remove the complexity related to the transport network and focus on assessment of the techno-economic impact of sensing (and the various technologies options) we considered only specific Functional splitting options for Type A and Type B RAN, while we considered baseband processing taking place at the 1st hop of COs. However, the study holistically modelled in detail the complete transport network deployment in order to study the technoeconomic impact of sensing not only on RAN but also on transport network costs. In the scenarios under study, we considered a technology mix of the prevailing last mile/access transport network technologies i.e. mmWave, P2P optical and WDM PON, while at the 1st and 2nd transport aggregation levels we considered a mix of optical transport technologies.

Step 4: Definition of the technology scaling and deployment dimensioning rules.

This step is based on the 6G-SENSES deployment blueprint.

Access and Transport Network Deployment Definition

The Access Network Deployment (5-years roll-out) definition follows the network architecture and common access network deployment principles. The access network dimensioning is performed in terms of actual number of RUs (of any RAN technology Type) needed for the deployment, calculated on the basis of traffic forecasts. The number of RUs can be also defined using external sophisticated radio network planning tools and inserted as input in the techno-economic-analysis tool. Further configuration information that is considered includes:

- The average number of RRUs per BBU. For Type A we assume where we assume an average of 18 RRUs per BBU; for Type B we assume 3 RUs/DU and 5 DUs/CU; For Type C 4 Antennas/BBU. Type D is an Access Point type of node so no aggregation in terms of BBU.
- The capacity of the nodes also derive from existing estimates, being for Type A: 2Gbps/RU, for Type B: 1Gbps/RU, for Type C: 1Gbps/RU with sensing and 4Gbps without sensing, for Type D:~2Gbps.
- The average maximum cell loading we assume is 80% in all cases following the telecom operator deployment principles.

Also, considering the transport network configuration we assume:

- Transport links utilization (as 80%-100% of link capacity) defined per technology.
- Aggregation and core transport equipment loading factors (interfaces loading) defined per technology per year.

Regarding the fiber deployment, various scenarios are considered ranging from Greenfield (0%) to Brownfield (50%) scenarios. The extend of the deployment is calculated taking into consideration the total number of sites to be (fiber) reached.

Transport network dimensioning rules consider technologies' capabilities/restrictions/default dimensioning rules and telecom operator imposed principles such as the component's threshold loading factors before scaling. The dimensioning is calculated in terms of transport equipment elements, for the various transport network links/ segments/ technologies.

Step 5: Definition of the cost information.

For the economic part of the analysis, input is required regarding the per-unit equipment purchase cost, installation cost and maintenance/service cost for each technology, along with the estimated annual prices erosion or escalation. The per-unit costs can be calculated then for a 5-year timeframe. In the context of the current analysis, cost information is collected from the industry partners. This information is confidential as it may vary depending on the requested equipment volumes, possible bilateral Business-to-Business agreements, etc.

Other financial figures introducing or representing extra costs such as WACC and Tax Deduction (% of Equipment Cost) can be also considered.

Step 6: Setting of analysis timeframe.

For telecom network deployments a 5-year timeframe is usually considered.

Step 7: Definition of the deployment scenario and extraction of results. & Step 8: Iteration of all steps and collection of analytical cost results.

The definition of the deployment scenario to be evaluated constitutes in setting the values in the aforementioned assumption parameters in the aforementioned steps 1-4. Based on the assumptions, the Tool provides:

- Dimensioning of radio and access / aggregation / core transport segments in terms of number of transport links, equipment units per year.
- Cost estimation of the deployment scenario per segment, including CAPEX/OPEX breakdown per technology/year/etc.
- Comparative cost results of a number of deployment scenarios.

The tool output/results are elaborated in the following section.

5.2.6 Techno-economic Analysis Scenarios and Results

In general, the scenarios on which our study has focused assume radio network capacity and traffic increase over the 5-year period, and radio network deployment performed at Year 1. However, over the 5 -year period the radio network capacity increase and the transport network dimensioning increase (in terms of links addition and capacity expansion) is considered.

The evaluation covered multiple RAN deployment scenarios, varying by RAN type composition, fiber transport configuration, and sensing capability. Scenarios included single-type deployments such as Type A (Typical RAN) and Type B (O-RAN), as well as mixed-type configurations like Type CD (50% Type C and 50% Type D), Type BCD (50% Type B, 30% Type C, 20% Type D), and Type ABCD (35% Type A, 15% Type B, 20% Type C, 20% Type D). For each scenario, two fiber transport options were considered—GF and BF (50%)—and both sensing-enabled and non-sensing configurations were evaluated, providing a comprehensive view of performance across diverse network compositions and capabilities. The extracted results provided absolute cost numbers of CAPEX, OPEX, TCO, RAN deployment costs, Transport deployments costs, CN deployment costs.

Table 5-4 Technoeconomic Scenarios for 6G-SENSES study

RAN Scenarios Evaluated	Fiber Transport Deployment	Sensing Capability
Type A: Only Typical RAN in service	GF	With Sensing
		No Sensing
	BF - 50%	With Sensing
		No Sensing
Type B: Only O-RAN in service	GF	With Sensing
		No Sensing
	BF - 50%	With Sensing
		No Sensing
Type CD: In service: Type C RAN - 50%; Type D RAN - 50%	GF	With Sensing
		No Sensing
	BF - 50%	With Sensing
		No Sensing
Type BCD: In service: Type B RAN 50%; Type C RAN - 30%; Type D RAN - 20%	GF	With Sensing
		No Sensing
	BF - 50%	With Sensing
		No Sensing
Type ABCD: In service: Type A RAN 35%; Type B RAN 15%; Type C RAN - 20%; Type D RAN - 20%	GF	With Sensing
		No Sensing
	BF - 50%	With Sensing
		No Sensing

The following paragraphs detail the analyzed scenario sets and the results analysis.

Scenarios Sets 1: Optical network deployment Greenfield vs Brownfield

This set of scenarios focus on the evaluation of a Greenfield optical network deployment compared to various Brownfield options. The Greenfield scenario refers to no existing fiber deployment in the area while the Brownfield scenario refers to the existence of fiber deployment in 50% of the locations of the RAN. We extracted the TCO for the various mix of RAN options, with and without sensing capabilities.

Table 5-5 Scenarios Set 1 – Comparative Results

	Impact of cost of fiber on TCO - BF vs GF	
	With Sensing	**No Sensing**
Type A	7%	7%
Type B	2%	2%
Type CD	6%	4%
Type BCD	5%	3%
Type ABCD	5%	4%

Unlike the results of previous generation mobile networks, as indicated by the comparative analysis results, the cost of fiber deployment is no more a critical cost factor, compared to the TCO. The reason is to be found not in the large capacities provided by fiber links, as well as in the comparatively high cost of the RAN. Apparently, the extra capacity required for sensing impacts the cost of the transport deployment, that's why we observed a slightly higher impact of fiber transport on TCO in the case of sensing-enabled deployment.

Scenarios Sets 2: Impact of Sensing on Cost

This set of scenarios focus on the evaluation the cost of sensing on various cost factors, and for various RAN deployment options. In practice, the results provided in Table 5-6 represent the percentage of cost overhead incurred when sensing is introduced in each of the following RAN deployment options vs. the non-sensing deployment.

Table 5-6 Scenarios Set 2 – Comparative Results

Cost Factor	Impact of sensing on Cost for various RAN deployment Options				
	Type A	Type B	Type CD	Type BCD	Type ABCD
CAPEX	0%	0%	214%	109%	133%
OPEX	0%	0%	210%	106%	130%
TCO	0%	0%	214%	109%	133%
RAN Cost	0%	0%	205%	106%	136%
Transport Cost	0%	0%	347%	116%	81%

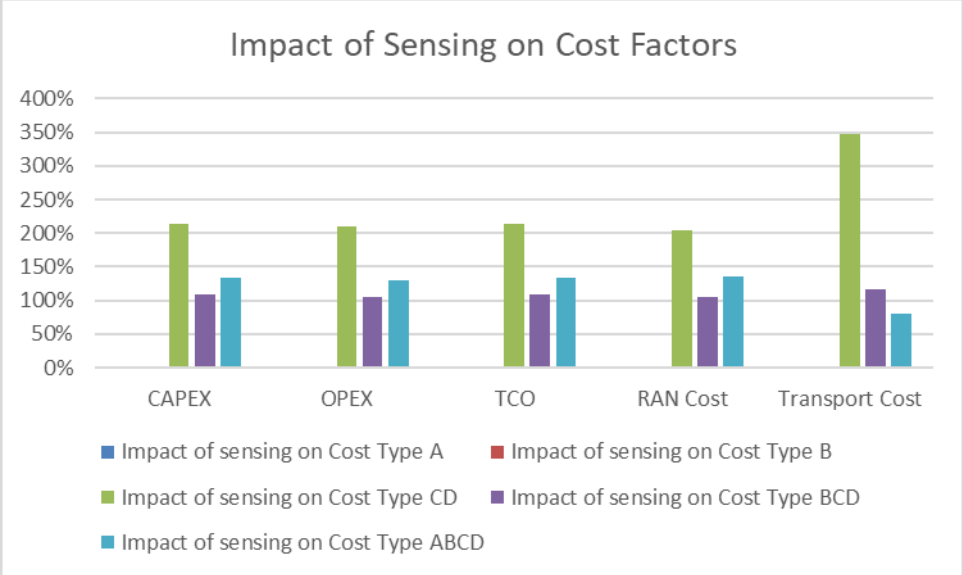


Figure 5-7 Scenarios Set 2 – Comparative Results

Through the analysis we observe that:

- For Type A (Typical RAN) and Type B (O-RAN): Introducing sensing is cost-neutral—there is 0% overhead across CAPEX, OPEX, TCO, RAN, and Transport. These architectures are the most favorable starting points for sensing enablement from a cost perspective.

- For Type CD (50% Type C / 50% Type D): Sensing introduces the largest cost burden of all scenarios, driven primarily by Transport (+347%) and substantial increases in CAPEX/TCO (+214%) and OPEX (+210%). This indicates significant infrastructure augmentation needs to support sensing.
- For Type BCD (50% B / 30% C / 20% D): Sensing causes significant cost growth across all factors—roughly ~+106–116%—with Transport (+116%) still the leading driver, but much less extreme than Type CD.
- For Type ABCD (35% A / 15% B / 20% C / 20% D): While sensing drives high RAN-side overhead (+136%) and CAPEX/TCO (~+133%), Transport overhead is comparatively lower (+81%) among the mixed scenarios—suggesting diversified RAN mixes reduce transport pressure but shift incremental cost to the RAN domain.

Scenarios Sets 3: Highly Aggregating RAN models vs. distributed and sensing enabling RAN models

This set of scenarios focuses on the evaluation of various RAN deployment options compared to the commercially available highly aggregated legacy RAN option. The results are extracted both in the case of sensing and without sensing capability. Table 5-7 and Table 5-8 present the cost overhead for various cost factors of various RAN deployment scenarios over Type A RAN (baseline scenario).

Table 5-7 Scenarios Set 3a – Comparative Results – no Sensing

Comparative study of various RAN options over highly aggregated RAN				
	Type B	Type CD	Type BCD	Type ABCD
CAPEX	694%	722%	863%	424%
TCO	696%	723%	865%	425%
RAN Cost	2064%	1961%	2304%	1142%
Transport Cost	33%	-43%	-4%	-13%

Table 5-8 Scenarios Set 3a – Comparative Results – with Sensing

Comparative study of various RAN options over highly aggregated RAN				
	Type B	Type CD	Type BCD	Type ABCD
CAPEX	694%	2479%	1915%	1122%
TCO	696%	2481%	1917%	1123%
RAN Cost	2064%	6186%	4840%	2830%
Transport Cost	33%	153%	106%	58%

The comparison highlights that moving from a traditional RAN (Type A) to more advanced or highly aggregated architectures significantly increases costs across most factors, particularly in the RAN domain. Type B (O-RAN) and mixed configurations such as Type BCD exhibit the highest overheads, with RAN costs surging by over 2,300% for Type BCD and 2,000% for Type B, indicating that disaggregation and virtualization drive substantial complexity and expense. CAPEX also rises sharply—while transport costs show a contrasting trend, with reductions in scenarios involving Type C and D (e.g., -43% for Type CD), suggesting that

aggregation can reduce transport requirements but at the expense of much higher RAN-side investments. This contrast is due to the significant capacity increase of Type C RAN when no-sensing capability is introduced. Overall, the analysis underscores that while advanced RAN architectures may optimize transport, they impose dramatic cost burdens on RAN infrastructure and operations.

The comparative analysis shows that when including sensing in the RAN, adopting highly disaggregated RAN architectures dramatically increases costs across all major factors compared to the baseline. RAN-specific costs exhibit the most significant escalation, rising by over 6,000% for Type CD and 4,840% for Type BCD, indicating that sensing in disaggregated RAN options impose substantial complexity and infrastructure requirements. CAPEX and TCO also surge, with Type CD reaching ~2,480%, followed by Type BCD (~1,917%) and Type B, reflecting the heavy upfront investment and overall ownership cost of these advanced architectures. Transport costs, while also higher, grow at a much slower rate—between 33% and 153%—suggesting that sensing has a major cost impact on the transport network cost. Overall, the study underscores the need for holistic technoeconomic studies (including transport and core besides RAN costs) when assessing the introduction of new RAN options.

In general, the scenario analysis reveals that sensing introduces no additional cost for Type A and Type B deployments, making them ideal for rapid and low-risk implementation. Especially Type B RAN being sensing-enabling or providing low resolution sensing with zero cost overhead across CAPEX, OPEX, TCO, RAN, and transport costs can be considered an investment safe option. In contrast, Type CD exhibits the highest sensitivity to sensing, with transport costs increasing by approximately 347% and other cost factors—such as CAPEX, OPEX, and TCO—rising by more than 200%.

Mixed deployments show varying cost patterns. Type BCD incurs balanced overheads of around 106–116% across all major cost factors, suggesting that sensing roughly doubles the total cost while distributing the impact evenly between transport and RAN domains. Type ABCD, on the other hand, shifts the burden primarily to the RAN side, with RAN costs increasing by 136% and CAPEX/TCO by approximately 133%, while transport costs remain relatively modest at 81%, although not introducing sensing to the complete service area. These findings underscore that the degree of RAN diversity and aggregation strongly influences cost distribution, with transport becoming the critical factor in scenarios dominated by Type C/D components, whereas mixed architectures mitigate transport upgrades but intensify RAN-related expenses.

6 Conclusions

This deliverable has presented the refinement and comprehensive evaluation of the 6G-SENSES architecture, concluding the work carried out in **WP2**. Building on the baseline architectural design and preliminary evaluation introduced in Deliverable **D2.2**, this document has focused on assessing the feasibility, performance, and impact of the proposed ISAC–enabled architecture across the user plane, control plane, and E2E service levels.

From an architectural perspective, the deliverable has introduced targeted refinements that strengthen the support of ISAC within a multi-layer, disaggregated framework aligned with 3GPP and O-RAN principles. These refinements include mechanisms for integrating sensing data from heterogeneous sensor sources, extensions of the O-RAN E2 interface to expose sensing information to the RIC, and intent-based control capabilities at the SMO layer. Together, these enhancements improve the architectural readiness of 6G-SENSES for sensing-aware networking and intelligent automation.

At the user-plane level, the deliverable has evaluated a broad set of ISAC-enabling technologies across multiple RATs. The results demonstrate the feasibility and benefits of multi-AP localization and tracking in the mmWave band, Wi-Fi–based sensing in Sub-8 GHz environments, DRL-based sensing-assisted MAC scheduling, distributed coordination in Sub-6 GHz CF-mMIMO, RIS across different frequency bands and deployment scenarios, and MEC-assisted wireless edge caching. These contributions highlight how sensing-aware processing, programmable propagation environments, and distributed coordination can improve accuracy, efficiency, and resource utilization in future 6G systems.

At the control-plane level, the deliverable has provided a twofold contribution. First, it has defined and applied a structured assessment procedure to verify the compliance of the Non-RT and Near-RT RICs with O-RAN standards, including functional validation of standardized interfaces and profiling of control-loop latency. Second, it has evaluated advanced control mechanisms developed within 6G-SENSES, independently of the compliance procedure. These include feedback-based sensing-slice control at the SMO layer for enforcing delay and freshness requirements, as well as near-real-time xApp-based solutions that inject non-3GPP sensing context into the RAN control loop. In particular, the joint evaluation of Wi-Fi–assisted beamforming and CQI + 5QI–aware scheduling, enabled through a compact WSI embedded in uplink control signaling, demonstrates tangible gains in channel quality, QoS satisfaction, and fairness while remaining aligned with O-RAN and 3GPP design principles.

At the E2E level, the deliverable has assessed how AI/ML-driven automation enhances the orchestration and operation of ISAC services. A DRL–based E2E slice orchestration framework has been evaluated, showing that dynamic and intelligent control of shared network resources can reduce energy consumption while satisfying service-level requirements for concurrent communication and sensing slices. In addition, a system-wide techno-economic evaluation framework has been introduced to assess deployment cost, scalability, and sustainability aspects of large-scale 6G deployments, complementing the technical performance analysis with economic considerations.

Overall, the results reported in this deliverable validate the refined architectural design choices of 6G-SENSES and provide concrete evidence of the effectiveness of the proposed user-plane and control-plane mechanisms, as well as the E2E benefits of AI/ML-driven orchestration for ISAC services. The comprehensive evaluation and benchmarking performed in this document confirm that the 6G-SENSES architecture constitutes a viable foundation for future 6G networks, capable of natively integrating sensing and communication while supporting intelligent, adaptive, and economically sustainable network operation.

7 References

[1] 6G-SENSES deliverable D2.2, "System architecture and preliminary evaluations", March 2025, https://6g-senses.eu/wp-content/uploads/2025/11/2025-03-05-6G-SENSES_Deliverable_2_2_vf_pending_EU-review.pdf

[2] 6G-SENSES deliverable D2.1, "Report on 6G-SENSES use cases, requirements and KPIs, and key technological advancements", September 2024, https://6g-senses.eu/wp-content/uploads/2024/10/2024-09-30-6G-SENSES_Deliverable_2_1_vf.pdf

[3] M. Anastasopoulos, J. Gutiérrez, A. Tzanakaki, "Optical Transport Network Optimization Supporting Integrated Sensing and Communication Services", in *Optical Fiber Communications Conference and Exhibition (OFC)*, San Francisco, CA, USA, 2025, pp. 1-3.

[4] A. Tzanakaki, M. Anastasopoulos, "Optical Transport Networks Supporting Integrated Communications and Sensing in 6G", in *European Conference on Optical Communications (ECOC)*, Copenhagen, Denmark, 2025, pp. 1-4, doi: 10.1109/ECOC66593.2025.11263441.

[5] 3GPP TR 33.713, "Study on security aspects of Ambient Internet of Things (AIoT) Services in 5G," 3rd Generation Partnership Project.

[6] 3GPP TS 33.210, "3G Security; Network Domain Security (NDS); IP Network Layer Security," 3rd Generation Partnership Project. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/33_series/33.210/

[7] 3GPP TS 33.310, "Network Domain Security (NDS); Authentication Framework (AF)," 3rd Generation Partnership Project. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/33_series/33.310/

[8] 3GPP TS 23.501, "System Architecture for the 5G System (5GS)." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/23_series/23.501/

[9] 3GPP TS 23.502, "Procedures for the 5G System (5GS)." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/23_series/23.502/

[10] 3GPP TS 38.300, "NR and NG-RAN Overall Description." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/

[11] O-RAN Alliance, "O-RAN Architecture Description," O-RAN.WG1.OAD. [Online]. Available: https://www.o-ran.org/specifications

[12] O-RAN Alliance, "O-RAN E2 Application Protocol (E2AP) Specification," O-RAN.WG3.TS.E2AP, v08.00. [Online]. Available: https://www.o-ran.org/specifications

[13] 3GPP TS 38.331, "NR; Radio Resource Control (RRC) Protocol Specification." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.331/

[14] 3GPP TS 38.323, "NR; Packet Data Convergence Protocol (PDCP) Specification." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.323/

[15] 3GPP TS 38.322, "NR; Radio Link Control (RLC) Protocol Specification." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.322/

[16] 3GPP TS 38.470, "NG-RAN; F1 General Aspects and Principles." [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.470/

[17] 3GPP TS 28.530, “Management and Orchestration; Concepts and Requirements.” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/28_series/28.530/

[18] 3GPP TS 28.531, “Management and Orchestration; Provisioning.” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/28_series/28.531/

[19] 3GPP TS 28.532, “Management and Orchestration; Assurance.” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/28_series/28.532/

[20] O-RAN Alliance, “Non-Real-Time RIC Architecture and A1 Interface,” O-RAN.WG2 and O-RAN.WG2.A1AP. [Online]. Available: https://www.o-ran.org/specifications

[21] 3GPP TS 23.288, “Architecture Enhancements for 5G System (5GS) to Support Network Data Analytics Services (NWDAF).” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/23_series/23.288/

[22] O-RAN Alliance, “Service Management and Orchestration (SMO) Integration – Technical Report,” O-RAN.WG1.TR.SMO-INT-R004, v06.00. [Online]. Available: https://www.o-ran.org/specifications

[23] 6G-SENSES deliverable D3.1, “Initial report on the development of 6G-SENSES infrastructure building blocks”, January 2025, https://6g-senses.eu/wp-content/uploads/2025/07/2025-01-15-6G-SENSES_D3.1_v1_0-FINAL.pdf

[24] 6G-SENSES deliverable D5.1, “Testing Methodologies and Testbed Setup”, November 2025, https://6g-senses.eu/wp-content/uploads/2025/11/2025-11-09-6G-SENSES_Deliverable_5_1_vf_pending_EU-review.pdf

[25] Y. Chan and K. Ho, “A simple and efficient estimator for hyperbolic location,” in *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, 1994.

[26] M. J. Neely. Stochastic Network Optimization with Application to Communication and Queueing Systems. 2010. url:http://dx.doi.org/10.2200/S00271ED1V01Y201006CNT007.

[27] C. Knill, B. Schweizer, S. Sparrer, F. Roos, R. F. H. Fischer and C. Waldschmidt, "High Range and Doppler Resolution by Application of Compressed Sensing Using Low Baseband Bandwidth OFDM Radar," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3535-3546, July 2018.

[28] 6G-SENSES deliverable D4.1, “Initial SoTA and design of wireless edge caching solutions”, January 2025, http://6g-senses.eu/wp-content/uploads/2025/02/2025-01-15-6G-SENSES_Deliverable_4.1_MAIN_FINAL.pdf

[29] O-RAN Alliance, “Near-Real-Time RAN Intelligent Controller Architecture,” O-RAN.WG3. [Online]. Available: https://www.o-ran.org/specifications

[30] 3GPP TS 38.801, “Study on New Radio Access Technology; Radio Access Architecture and Interfaces.” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.801/

[31] 3GPP TS 38.473, “NG-RAN; F1 Application Protocol (F1AP).” [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.473/

[32] G. Zhang, X. Q., Jiang, H. Hai, L. Xu, S. Mumtaz, “Intelligent reflecting surfaces based offset index modulation for MIMO systems”, in *IEEE Transactions on Vehicular Technology*, 73(8), pp.11735-11742.

[33] M.S. Ayub, M. Saadi, and I. Koo, “Optimization of RIS-Assisted 6G NTN Architectures for High-Mobility UAV Communication Scenarios”. *Drones*, *9*(7), p.486.

[34] P. Saikia, A. Jee, K. Singh, S. Mumtaz, and W.J. Huang, “STAR-RIS-aided full-duplex ISAC systems: A novel meta reinforcement learning approach”, in *IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 5086-5091.

[35] S. Kurma, K. Singh, S. Mumtaz, T.A. Tsiftsis, and C.P. Li, “Resource optimization in active-STAR-RIS-aided THz ISAC systems with DDA modulation: A machine-learning approach”, in *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 15291-15307, Oct. 2024, doi: 10.1109/TWC.2024.3428493.

[36] V. -M. Alevizaki, A. -I. Manolopoulos, M. Anastasopoulos and A. Tzanakaki, "Integrating 3GPP and Non-3GPP Technologies for Hybrid Positioning in B5G Networks," *in Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Poznan, Poland, 2025, pp. 333-338, doi: 10.1109/EuCNC/6GSummit63408.2025.11037103.

[37] V.-M. Alevizaki *et al.*, “Distributed Service Provisioning for Disaggregated 6G Network Infrastructures,” in *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, 2023.

[38] I. Floudas, M. Anastasopoulos, A. Tzanakaki and J. G. Terán, "Experimental Evaluation of Semantic Communications for 6G Networks in Railway Systems," in *IEEE Communications Standards Magazine*, vol. 9, no. 4, pp. 34-40, Dec. 2025, doi: 10.1109/MCOMSTD.2025.3601193.

[39] 3GPP TS 38.213 V17.12.0 Physical layer procedures for control, Technical Specification Group Radio Access Network, 2025.

[40] 3GPP TS 38.212 V17.10.0 Multiplexing and channel coding, Technical Specification Group Radio Access Network, 2025.

[41] 3GPP TS 38.211 V18.5.0 Physical channels and modulation, Technical Specification Group Radio Access Network, 2025.

[42] O-RAN Alliance, O-RAN Working Group 1, Technical Specification OAD-R004-v13.00, Feb. 2025.

[43] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, ‘‘Network slicing and softwarization: A survey on principles, enabling technologies, and solutions,’’ *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 3rd Quart., 2018.

[44] Telecom Italia, “Milano Grid,” 2015.

[45] S. Atapattu, C. Tellambura and H. Jiang, "Representation of Composite Fading and Shadowing Distributions by Using Mixtures of Gamma Distributions," in *IEEE Wireless Communication and Networking Conference*, Sydney, NSW, Australia, 2010, pp. 1-5, doi: 10.1109/WCNC.2010.5506173.

[46] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[47] V. Mnih et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

8 Acronyms

Acronym	Description
2D	Two-Dimensional
3GPP	3rd Generation Partnership Project
5GC	5G Core Network
5GNR	5G New Radio
5QI	5G QoS Identifier
6G-IA	6G Infrastructure Association
AI	Artificial Intelligence
ALM	Augmented Lagrangian Method
AoA	Angle of Arrival
AP	Access Point
API	Application Programming Interface
ASRIS	Active STAR-RIS
BER	Bit Error Rate
BI	*Barkhausen Institut* (6G-SENSES Beneficiary)
BR	BubbleRAN (6G-SENSES Beneficiary)
BS	Base Station
BTN	Backhaul Transport Network
CAPEX	CAPital EXpenditures
CCC	Cell Configuration and Control
CF	Cell-Free
CF-mMIMO	Cell-Free massive MIMO
CIR	Channel Impulse Response
CN	Core Network
CO	Central Office
COTS	Commercial Off-The-Shelf
CPRI	Common Public Radio Interface
CQI	Channel Quality Indicator
CSI	Channel State Information
DBaaS	Database as a Service
DDA	Dynamic Delay Alignment
DNN	Deep Neural Network
DoA	Direction of Arrival
DoW	Description of Work
DPP	Drift-Plus-Penalty
DQN	Deep Q Learning

DRL	Deep Reinforcement Learning
DT	Digital Twin
E2E	End-to-End
EC	European Commission
EKF	Extended Kalman Filter
FIR	Finite Impulse Response
FoV	Field of View
FTN	Fronthaul Transport Network
HW	Hardware
HPD	Human Presence Detection
HPE	Hierarchical Permutation Equivariant
IASA	Institute of Accelerating Systems and Applications (6G-SENSES Beneficiary)
ICT	Information and Communication Technology
IHP	*IHP – Leibniz Institut für innovative Mikroelektronik* (6G-SENSES Beneficiary)
INT	*Intel Deutschland GmbH* (6G-SENSES Beneficiary)
IoT	Internet of Things
I/Q	In-Phase and Quadrature
IQR	interquartile range
ISAC	Integrated Sensing and Communication
ISM	Industrial, Scientific and Medical
IT	Information Technology
ITU	International Telecommunications Union
KPI	Key Performance Indicator
KV	Key Value
KVI	Key Value Indicator
LO	Local Oscillator
LoS	Line-of-Sight
LTF	Long Training Field
m/eMTC	massive/enhanced Machine Type Communications
MAC	Medium Access Control
MADDPG	multi-agent deep deterministic policy gradient
MARL	Multi-Agent Reinforcement Learning
MEC	Multi-access Edge Computing
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
mMIMO	Massive MIMO
MNO	Mobile Network Operator
MTI	Moving Target Indication

mTLS	mutual Transport Layer Security
MTU	Maximum Transmission Unit
MUI	MultiUser Interference
MX	Multi-x (i.e. multi-vendor, multi-version, multi-node, multi-distribution, multi-runtime, multi-cloud, and multi-instance)
N3IWF	Non-3GPP InterWorking Function
NLoS	Non-Line-of-Sight
NN	Neural Network
NR	New Radio
NTU	Nottingham Trent University
O-RAN	Open Radio Access Network
OAI	OpenAirInterface
OAM	Operations And Management
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	OPerational EXpenditures
OS	Operating System
OSS	Operations Support System
OTE	*ORGANISMOS TILEPIKOINONION TIS ELLADOS OTE AE* (6G-SENSES Beneficiary)
PDSCH	Physical Downlink Shared Channel
PEB	Position Error Bound
PF	Proportional Fairness
PHY	Physical
PID	Proportional-Integral-Derivative
PoC	Proof-of-Concept
PUSCH	Physical Uplink Shared Channel
RAN	Radio Access Network
RB	Resource Block
RBIS	Reinforcement-Based Intelligent Scheduling
RC	RAN Control
RF	Radio Frequency
RF-DS	Range-Filtered Doppler Spectrum
RIS	Reconfigurable Intelligent Surface
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RTT	Round-Trip Time
RU	Radio Unit
SA	Stand Alone
SAF	Sensing Analytics Function

SBA	Service-Based Architecture
SCWD	Sensing-Centric Waveform Design
SDN	Software Defined Networking
SDR	Software Defined Radio
SE	Spectral Efficiency
SeCF	Sensing Control Function
SINR	Signal-to-Interference-plus-Noise Ratio
SLA	Service Level Agreement
SM	Service Model
SMO	Service Management and Orchestration
SNS JU	Smart Networks and Services Joint Undertaking
STAR-RIS	Simultaneously Transmitting and Reflecting RIS
SVM	Support Vector Machine
SW	Software
TCO	Total Cost of Ownership
TDOA	Time Difference of Arrival
THz	Terahertz
TUBS	*Technische Universität Braunschweig* (6G-SENSES Beneficiary)
UAV	Unmanned Aerial Vehicles
UC	University of Cantabria (6G-SENSES Beneficiary)
UCI	Uplink Control Information
UN	United Nations
UNIROMA1	*Sapienza Università di Roma* (6G-SENSES Beneficiary)
UPF	User Plane Function
URA	Unsourced Random Access
URLLC	Ultra-Reliable and Low-Latency Communications
VNF	Virtual Network Function
WACC	Weighted Average Cost Of Capital
WAT	Wireless Access Technology
WG	Work Group
Wi-Fi	Wireless-Fidelity
WLAN	Wireless Local Area Network
WLS	Weighted Least Squares
WP	Work Package
WSI	Wi-Fi Sensing Indicator
XR	eXtended Reality
YANG	Yet Another Next Generation