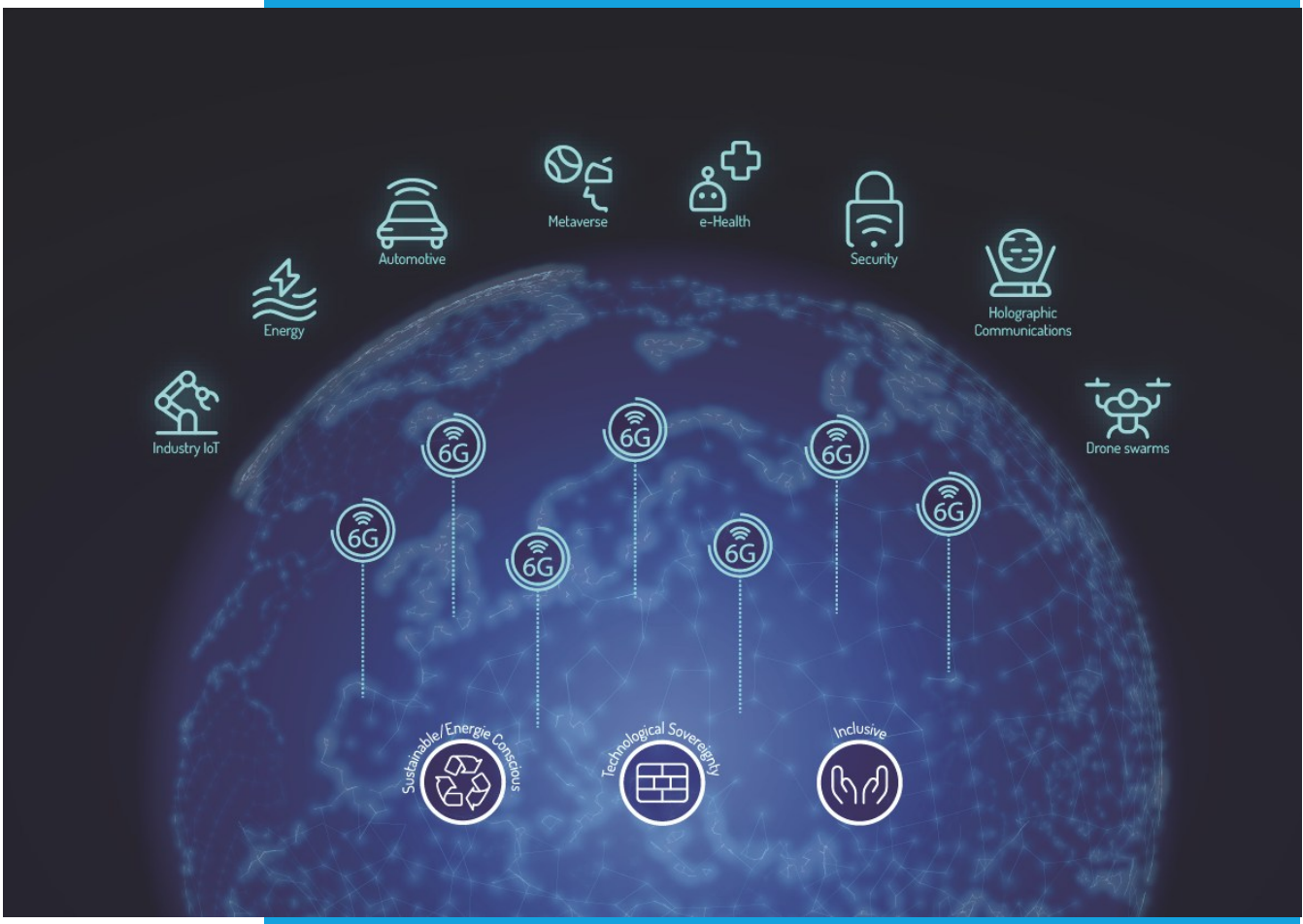
Smart Networks and Services Joint Undertaking (SNS JU)
6G Architecture Working Group



Consolidated Version

30 May 2025

WHITE PAPER

TOWARDS 6G ARCHITECTURE: KEY CONCEPTS, CHALLENGES, AND BUILDING BLOCKS

DOI: 10.5281/zenodo.15001377
URL: https://doi.org/10.5281/zenodo.15001377

Co-funded by the European Union

6GSNS

ACKNOWLEDGEMENT

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the SNS-JU. Neither the European Union nor the SNS-JU can be held responsible for them.

EXECUTIVE SUMMARY

We are entering the standardization phase for the 6th generation (6G) of wireless technologies. While valuable lessons have been learned from the design, deployment, and operation of 5G and Beyond in Europe, new requirements, emerging technologies, and evolving business models must be natively integrated into the next-generation mobile network architecture.

This white paper presents a comprehensive snapshot of the current architectural considerations explored by the Smart Networks and Services Joint Undertaking (SNS-JU) projects. It aims to discuss the rationale for novel architectural components, the ongoing design efforts, and the future outlook for 6G.

We analyse the blueprint for next-generation mobile networks, building on past experiences while integrating cutting-edge advancements. The structure follows the IMT-2030 framework, categorizing insights into the key usage scenarios and overarching architectural aspects.

- Usage Scenarios: We analyse the evolution of IMT-2030 paradigms, focusing on Immersive, Massive, and Hyper-Reliable Low-Latency Communications. We also explore novel 6G enablers, such as Ubiquitous Connectivity, AI-driven Communication, and Integrated Sensing & Communication.
- Overarching Aspects: The next-generation architecture must natively embed Security, Privacy & Trustworthiness, Sustainability, and Network Exposure Capabilities, ensuring these critical aspects are foundationally considered rather than retroactively incorporated.

Finally, we conclude by outlining the essential building blocks shaping the next-generation mobile network architecture, highlighting the most promising research paths identified in this white paper.

TABLE OF CONTENTS

1 INTRODUCTION

As the telecommunications industry advances beyond 5G, the transition to 6G is set to revolutionize the way networks are designed, deployed, and utilized. The 6G Architecture Working Group has prepared this white paper to define the fundamental architectural principles that will guide the development of next-generation mobile networks. The white paper provides an extensive analysis of the key technological enablers, system design choices, and research challenges that will shape the 6G ecosystem.

Unlike previous generations, 6G networks will move beyond connectivity to become intelligent, context-aware, and adaptive systems, leveraging artificial intelligence (AI), deep-edge computing, non-terrestrial networks (NTN), and integrated sensing and communication (ISAC). This transition will require a fundamental rethinking of network architecture, ensuring that 6G can support ubiquitous, sustainable, and resilient communication across heterogeneous environments, from urban landscapes to remote and un(der)served areas.

This white paper begins by analysing the state of current 5G and Beyond 5G (B5G) deployments, highlighting experimental testbeds and federated research platforms across Europe that integrate cutting-edge technologies like network slicing, AI-enabled orchestration, multi-access edge computing (MEC), real-time extended reality (XR) applications, and NTN connectivity. Despite these advancements, critical technology gaps persist in latency, energy efficiency, network scalability, and interoperability with respect to the demanding requirements of existing and future vertical use cases. Key challenges include latency constraints for time-sensitive applications, throughput limitations for future applications like holographic communication, energy efficiency concerns for sustainability of networks, and scalability and interoperability issues. This sets the stage for the 6G system blueprint, detailing how 6G will address these challenges.

The 6G System Blueprint outlines the key architectural principles that define 6G's end-to-end design. These principles include cloud-native, AI-driven networks for dynamic, intelligent network management; modular and scalable architectures to support diverse use cases; intent-based networking (IBN) for dynamic adaptation to user intent; and ISAC for real-time environment perception. The 6G architecture will be built upon a multi-layered framework consisting of an infrastructure layer integrating

terrestrial, aerial, and space-based network components and resources; a network layer with a unified, software-defined design incorporating 6G radio access networks (RANs), core network functions (NFs), and beyond-communication services enablers; an application layer with an AI-driven service framework for customized network functionality; and a security and trust layer with a decentralized zero-trust security model. This blueprint redefines network flexibility, enabling adaptive, programmable, and AI-powered connectivity.

Key enhancements will shape the evolution of 6G, including resource management in the deep-edge-edge-cloud continuum, leveraging distributed computing and AI-driven workload orchestration; zero-trust security and interoperability through zero-trust networking (ZTN) and standardized APIs; and sub-network integration in 6G networks, featuring dynamic, software-defined sub-networks. 6G will extend connectivity beyond terrestrial networks by NTNs for global coverage, enabling direct handheld access to satellites, and incorporating optical wireless communications (OWC) for seamless indoor and outdoor coverage. AI will be native to 6G, embedding intelligence into network automation, intent-based communications, and digital twins for predictive analytics. 6G will merge wireless communication with environmental sensing, enabling smart cities, industrial IoT, vehicular communication, and healthcare innovations.

Security, resilience, and trust in 6G will be ensured through a multi-layered security framework covering end-to-end encryption, post-quantum cryptography, AI-driven threat detection, and resilient network architectures. Sustainability is a core pillar of 6G, with energy-efficient architectures addressing green AI-powered network management, sustainable hardware design, and intelligent resource allocation. Network exposure capabilities, on the other hand, will foster innovation and create an API ecosystem for exposure services, leading to more flexible networks that can accommodate a wider range of use cases. Finally, the white summarizes the main findings on the architectural components, highlighting trends, solutions, and depicting a roadmap towards the 6G architectural design. This white paper serves as a comprehensive guide to 6G architectural advancements, design principles, and emerging technologies, aiming to redefine global connectivity by addressing flexibility, intelligence, security, and sustainability.

1.1 CURRENT DEPLOYMENTS

1.1.1 CURRENT DEPLOYMENTS AND EFFORTS

In the context of the SNS projects ecosystem, there is a continued effort in deploying federated, sustainable experimentation platforms that span across Europe, enabling advanced research and validation of 5G and Beyond 5G (B5G) technologies. They provide decentralized and scalable architecture supporting also Testing as a Service (TaaS), resource sharing, and cross-domain application deployment. Some of them also support portals that serve as the primary entry point, offering readily available functionality and easy-to-use_tools for experimenters, infrastructure owners, and vertical developers.

The testbed deployments integrate cutting-edge technologies such as network slicing, AI-enabled components, and real-time and intelligent orchestration, supporting diverse use cases, including real time edge computing, immersive applications, and data-intensive tasks. These testbeds also facilitate rapid prototyping, KPI tracking, and interoperability testing, bridging the gap between experimental platforms and commercial-grade networks.

This federated experimentation platform [SNS-1, SNS-2, 6G-IA-1] exemplifies a significant leap in telecommunications research, offering a cohesive ecosystem that bridges diverse testbeds across Europe. With its decentralized architecture, advanced tools, and shared resources, the initiative fosters innovation in 5G and B5G technologies, empowering researchers and industry stakeholders to explore emerging use cases and validate cutting-edge solutions. The unique capabilities of each testbed, ranging from AI-enabled orchestration to real-time XR deployments and edge computing, demonstrate the potential for scalable, adaptive, and collaborative research. By aligning with global standards and supporting interoperability through unified APIs, the platform ensures broad accessibility and long-term sustainability.

As the industry transitions towards 6G, this initiative lays a strong foundation for collaborative innovation, addressing the challenges of future networks while driving economic growth and technological advancement in Europe and beyond.

1.1.2 MAIN USE CASES AND TECHNOLOGY GAPS

The experimental facilities discussed above are used to deploy a range of compelling vertical use cases involving high-demand requirements, which are facing gaps in key areas:

- Latency Requirements: Applications like "Smart Crowd Monitoring" and "Remote Proctoring" demand ultra-low latency (<10ms) for real-time operation. Current 5G NSA setups struggle to achieve these thresholds, with upgrades to 5G SA and new network releases (e.g., Rel.-17) anticipated to bridge this gap.
- Throughput Limitations: Use cases involving XR experiences or data-heavy applications in healthcare face challenges in maintaining consistent uplink and downlink throughput, particularly during peak traffic periods. Despite incorporating mid-band and mmWave frequencies, further capacity enhancements are necessary [Fer+22].
- Coverage and Reliability: Mixed urban and rural environments present challenges for achieving consistent service quality. Local breakout and edge computing solutions mitigate some issues, but gaps persist, especially in regions reliant on 5G NSA infrastructure.
- Energy Efficiency: Use cases requiring extensive connectivity (e.g., IoT networks) often struggle with power consumption optimization. Intelligent traffic management and energy-efficient network strategies are crucial for addressing this issue.

High-demand scenarios like live media streaming and uplink-intensive use cases face unique complexities. These include dynamic resource allocation, seamless integration across diverse networks, and programmatic end-to-end (E2E) orchestration. These challenges include:

- Dynamic Resource Allocation: Meeting fluctuating bandwidth needs for uplink-heavy applications while balancing QoS and QoE targets [Mon+22].
- Latency and QoS Guarantees: Managing stringent latency and reliability requirements in shared environments.
- Interoperability: Ensuring seamless integration across multi-vendor equipment and proprietary systems for E2E orchestration.

- Advances in orchestration frameworks and adherence to standards like 3GPP and GSMA/Camara APIs will enable more effective traffic isolation and management for uplink-heavy use cases [TrialsNetD2.3].

Federated experimental platforms require further development to support advanced interoperation and vertical integration. Key gaps include:

- Horizontal and Vertical Interoperation: Platforms must enhance cross-domain coordination to support scalable and dynamic experimentation.
- Service Reliability and Energy Efficiency: Ensuring consistent service quality and power-efficient operations across diverse use case scenarios.
- Network Upgrades: Broad adoption of 5G SA, improved edge computing, and enhanced spectrum usage are critical to addressing latency, throughput, and coverage challenges.

While current platforms demonstrate promising results, gaps in performance, scalability, and resource efficiency must be addressed to fully realize their potential. Integrating emerging solutions such as the ones that we propose in this will play an important role for efficiently integrating this view.

Advancing experimental platforms and leveraging collaborative research across Europe are essential for enabling next-generation communication services. These efforts will drive innovation while ensuring sustainability, adaptability, and interoperability in the evolving 5G and Beyond 5G ecosystem.

1.2 REFERENCES

[6G-IA-1] 6G-IA Trials WG https://6g-ia.eu/6g-ia-working-groups/#trials

[FUDGED3.2] P. Chakraborty and M.-I. Corici, “FUDGE-5G D3.2 On-boarding and Deploying of the Vertical Use Cases”, 2024.

[Fer+22] S. Fernández, M. et al., “Multiparty Holomeetings: Toward a New Era of Low-Cost Volumetric Holographic Meetings in Virtual Reality”, IEEE Access 2022.

[SANDBOXD2.1] G. Makropoulos, “Deliverable D2.1 - Ecosystem analysis and 6G-SANDBOX facility design”, Zenodo, Jul. 2023. doi: 10.5281/zenodo.8366395.

[SNS-1] SNS-JU Project Portfolio https://smart-networks.europa.eu/project-portfolio/

[SNS-2] SNS-JU Trials & Pilots Brochure

[Mon+22] M. Montagud, et al., “Towards SocialVR: Evaluating a Novel Technology for Watching Videos Together”, Virtual Reality (Springer), 2022.

[TrialsNetD2.3] G. Scivoletto, “Deliverable D2.3 - Final design of Platforms and Networks solutions”, Zenodo, Dec. 2024. doi: 10.5281/zenodo.14512906.

2 SYSTEM BLUEPRINT

There is a need for a more flexible and adaptable E2E system architecture for 6G compared to previous generations. 6G will need to support an efficient mobile communication service as well as the new 6G offerings such as beyond communication services, AI and compute offloading. Furthermore, 6G will also be able to process data, generate insights, and deliver value-added services such as spatial/temporal data services, computation services and intelligence services (e.g., such as AI functionality, analysis, and optimizations).

Additionally, as compared to the previous generations, a wider ecosystem collaboration between mobile network operators, cloud providers, enterprises, vertical industries, integrators, application developers, application service providers, end users, etc., will be a prerequisite to creating value for all players. This demands the 6G network to be a versatile platform interfacing with a broad range of applications, which can be tailored to the specific requirements of the ecosystem players.

The architecture should also support the ability to better scale networks than in today's 5G deployments. This should be done dynamically, based on current needs, to improve efficiency.

2.1 ARCHITECTURAL DESIGN PRINCIPLES

The services provided by the 6G platform will be implemented through multiple interacting subsystems, which encompass device evolution, network infrastructure enhancements, advanced network capabilities, and pervasive functionalities including security and privacy, data handling, artificial intelligence frameworks, and end-to-end management and orchestration. In order to ensure flexibility and programmability, the 6G platform will expose a diverse set of Application Programming Interfaces (APIs) to applications, users, and industry verticals using these services, with the capability to extend the APIs over time. The 6G E2E system will adhere to the ten architectural principles detailed in Table 2.1, thus offering an efficient framework to support emerging 6G technologies.

Table 2.1: Mapping of architecture design principles and their impact on 6G E2E system design

Design principles	Impact on 6G E2E system blueprint design	Key values
Support and exposure of 6G services and capabilities	This feature encompasses generic and dynamic exposure functionalities, e.g. simplified APIs to expose capabilities to E2E applications facilitating e.g. seamless integration of beyond communication NFs and hardware capabilities. It emphasizes the inclusion of pervasive AI and a robust compute infrastructure to enhance the overall flexibility for compute offloading of services and improve the performance of the system.	Economic sustainability, environmental sustainability, Trustworthiness, Inclusiveness
Full automation and optimization	The 6G E2E system necessitates a comprehensive and widespread data and analysis framework, complemented by a pervasive AI framework and service management and orchestration. The emphasis is on building an infrastructure that enables efficient data handling, robust AI integration, and seamless service orchestration across diverse scenarios. The pervasive AI framework provides means for predictive orchestration, resorting to distributed AI/ML agents to optimize the system without human interaction. This framework provides continuous orchestration over multiple administrative domains, supporting the multi-stakeholder 6G ecosystem.	Economic sustainability, environmental sustainability, Trustworthiness
Flexibility in integrating different networks	This design principle integrates many different modalities of connectivity, including local and wide area networks, non-terrestrial networks, subnetworks, public and private networks by supporting seamless mobility between them. New spectrum is to be used in an efficient way, implementing programmable transport configurations and having application awareness and adaptive quality of service (QoS) and quality of experience (QoE).	Economic sustainability, environmental sustainability, Inclusiveness
Scalability	This focuses on creating a pervasive service management and orchestration system (e.g., scaling up and down based on mobility and time-varying traffic needs), incorporating a network-centric exposure layer and optimized transport NFs (e.g., over heterogeneous multi-domain/multi-clouds). In addition, network modularity will enable dynamic and efficient introduction and removal of network resources as needed.	Economic sustainability, environmental sustainability, Trustworthiness
Resilience and availability	The 6G E2E system requires pervasive service management and orchestration (e.g., high resilience and availability), encompassing a comprehensive	Trustworthiness

	framework for data analysis, AI integration, and the coordination of RAN functions, transport NFs, 5G/6G cloud native functions (CNFs). E.g., separation of control plane (CP) and user plane (UP), resilient mobility solutions, enhanced redundancy and recovery mechanisms.	
Persistent security and privacy	The objective is to establish a comprehensive framework in the 6G E2E system that ensures security and privacy are integrated across all components with the goal of assuring a trustworthy environment. E.g., address current as well as future threats in a resilient manner and incorporate security fundamentals in its design, inherently support the preservation of privacy, allow different levels of anonymity for future services.	Trustworthiness
Cloud –optimized internal interfaces	This effort centres on the deployment of cloud-native virtual NFs, emphasizing the development of exposure interfaces that facilitate seamless internal communication between different layers of the 6G E2E system.	Economic sustainability, environmental sustainability, Trustworthiness
Separation of concerns of NFs	This refers to the optimized functionality in CN and RAN with bounded context and no duplication, avoiding complex interdependencies and cross-functional signalling- Self-sustained NFs with minimal dependency on other NFs.	Trustworthiness
Network simplification in comparison to previous generations	This initiative aims to avoid many standardized deployment options and protocol splits. It also involves the evolution of the 5GC to accommodate the requirements of 6G RAN. The focus is on simplifying protocols and minimizing User Equipment Network (UENW) signalling.	Economic sustainability, environmental sustainability, Inclusiveness
Minimization of environmental first order effect and enabling sustainable use cases	This principle aims for E2E orchestration, emphasizing energy-efficient and cost-conscious operations. It involves the implementation of a pervasive data and analysis framework alongside the modularization of NFs. The infrastructure layer in the 6G E2E system should optimize both energy consumption and costs for enhanced sustainability and operational efficiency of the use cases.	Economic sustainability, environmental sustainability,

2.2 SYSTEM BLUEPRINT

Starting from the use cases, the underlying architecture design principles and the ecosystem environment, the needed capabilities and requirements can be extracted to develop and assess the system performance as well as showing that the majority of the 6G use cases have a significant positive impact on social, economic and environmental sustainability in light of the offered key values. This is an iterative design process

wherein enablers tackling various aspects of the 6G system are analysed for their seamless integration into the 6G E2E system blueprint. This approach aims to create a system that aligns closely with the real-world needs and expectations of users and stakeholders, fostering a more adaptive and responsive design.

A mature form of 6G E2E system blueprint thus obtained, which is discussed and further detailed in [HEX225-D25], is depicted in Figure 2.1, wherein several novel aspects are encompassed.

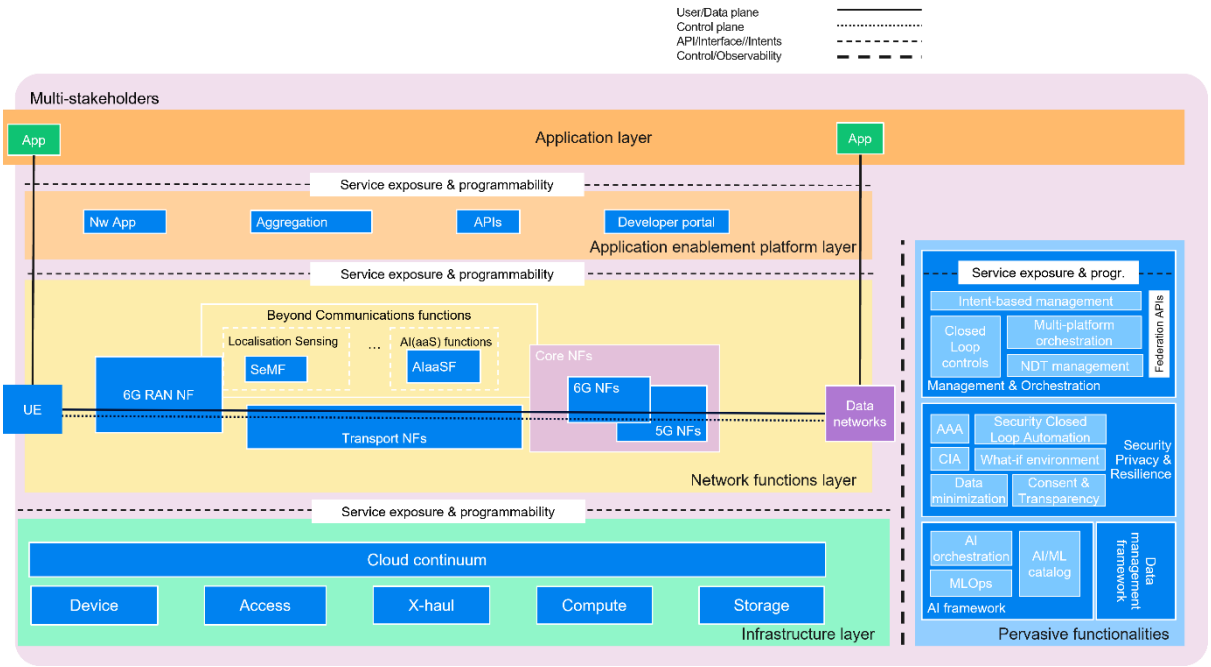


Figure 2.1: E2E System Architecture

The infrastructure layer encompasses all the E2E infrastructure and resources, physical or virtualized, spanning across different technological domains and administrative domains. It introduces the device-edge-cloud continuum, shortened as cloud continuum principle, which will be essential for the 6G E2E system.

The network layer comprises beyond communications functions as well as the 6G RAN and Core NFs blocks. Considering the simplicity of evolution from earlier generations to 6G, the 6G radio access should support a single-RAT architecture only, i.e., a 6G UE that connects via the 6G radio interface establishes a connection to the CN for 6G without any complex inter-RAT multi-connectivity, as compared to the 5G paradigm. The 6G Core NF will be an extension of 5G Core network instead of a subset. The 6G Beyond Communications Functions depicts the functionalities to realize new services expanding beyond the communication capabilities.

The Network-centric application enablement platform layer and the 3rd party application layer enable the full ecosystem applications to interact with the underlying network. The Application Enablement Platform Layer, also known as the Network-Centric Application Layer, serves as the main entry point to the services of the 6G system platform, managed through service exposure. Its primary role is to provide applications with simplified access to 6G capabilities—including management services—via abstracted APIs from the network. Additionally, management and orchestration (M&O) services offer a further layer of abstraction to make these capabilities easier to use for non-experts, and are enhanced with additional services from, for example, network applications, which can be leveraged by developers in the Application Layer. It is providing developer portal, operation, administration, and management (OAM) services (e.g., API discovery, ordering, monitoring). for API consumers, and common functionalities for monetization, authorization, and legislation compliance. Network applications can augment the 6G capabilities with additional services, provided also through APIs.

Aggregation of services, i.e., the action to compose services from a set of capabilities of different 6G service providers, is also provisioned in this layer.

The pervasive functionalities facilitate the four layers of the 6G system blueprint to realize the full potential of the 6G platform, either jointly or independently. Both data management and AI frameworks highlight the AI-centric approach to the 6G system as compared to the 5G system. The data management framework integrating DataOps plays a crucial role in ensuring the effective collection, processing, governance and data quality, which is essential for pervasive AI/ML operations. The 6G system architecture integrates AI/ML into every layer. The AI/ML framework provide a comprehensive suite of tools and components designed to enable advanced AI/ML workflows across the entire ecosystem. It consists of an AI orchestrator, AI/ML catalogue, and MLOps orchestrator, working together to manage the lifecycle of AI/ML services, ensuring efficient deployment, optimization, and continuous improvement, and integrating DataOps and MLOps into the process. The management and orchestration framework moves beyond the traditional relevant functions in 5G networks towards a more intent-based management approach. Hence, the proposed blueprint introduces the Intent-based Management framework as one of the pillars of the 6G management and orchestration framework. The multi-platform orchestration functionality provides a unified management and orchestration of network services and

network applications over a cloud continuum across multiple domains, owned and administered by different stakeholders, and characterized by underlying heterogeneous technologies platforms. The possibly AI/ML based closed loop controls are essential for an increasing level of M&O automation toward autonomy in the 6G network operations. Network Digital Twins (NDTs) have emerged as a tool for applying orchestration actions and observing their effects before carrying them out on the real network. To provide accurate feedback to the M&O, the NDT management creates the Network Digital Twin based on data collected from the real network to accurately reflect the state and behaviour of the network [HEX225-D65]. A broader security and privacy framework for the 6G E2E system is required compared with a more localized and domain specific approach on different layers of the 5G system [HEX225-D25].

Lastly, there will be multiple interactions between the various sets of stakeholders of the 6G ecosystem. As the market is becoming more disaggregated, multiple stakeholders will be engaged in the value creation of the 6G platform, moving away from a linear value chain toward a multi-sided value chain. This leads to the definition of new roles as already described in [HEX223-D22] and multiple multi-stakeholder interfaces between the various layers and domains of the 6G platform, in particular, with federation APIs representing the functionalities required for establishing and managing collaboration between different service or resource stakeholders.

2.3 MIGRATION ASPECTS

In the migration from 4G to 5G, there was a big push for a gradual transition from 4G RAN (LTE) towards a new 5G RAN, which included the move from 4G CN to a new 5G CN (5GC). The main reason was to allow the operators to flexibly upgrade RAN and CN independently. One solution for this was to enable E-UTRA (i.e., 4G) and NR (i.e., 5G) Dual Connectivity, also known as EN-DC. This created a plethora of EN-DC combinations, which had to be standardized, configured, and tested. In the end, only a few combinations were eventually deployed, and the main options were (1) NR connected directly to 5GC (i.e., the 5G stand-alone (SA) option 2) and (2) LTE acting as the Master Node (MN) and NR as the Secondary Node (SN) connected to EPC (i.e., 5G non-standalone (NSA) option 3x). The use of EN-DC as a migration and interworking solution between 4G and 5G also delayed the introduction of 5GC, since NR could be directly connected to the EPC, but this hindered several new 5G services related to the 5GC from being offered. Therefore, to avoid this complex deployment model, and also

to enable a smoother and faster introduction of 6G services, the main option for migration between 5G and 6G is to use a so-called “evolved 5GC, E-5GC” as in Figure 2.1. E-5GC allows the re-use of existing 5GC NFs while introducing new dedicated 6G NFs to support new 6G functionality. In addition, Multi-RAT Spectrum Sharing (MRSS) as a spectrum migration solution can support interworking between 5G and 6G [HEX223-D43], see Figure 2.2.

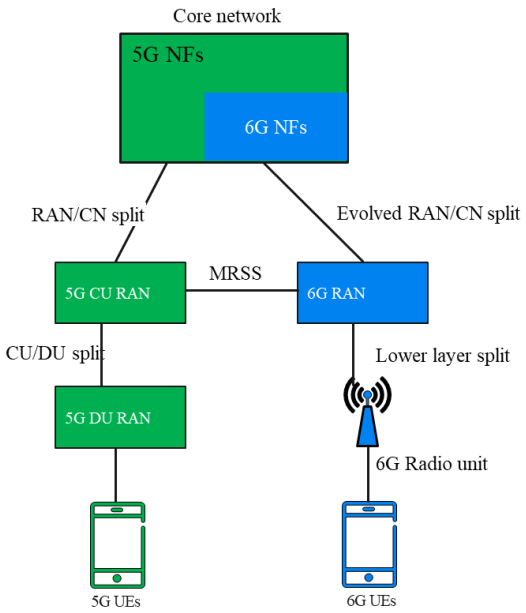


Figure 2.2: Overview of the migration aspects

For MRSS, the legacy 5G UEs should be able to avoid all 6G signals, i.e., the 6G signals are “hidden” by reusing 5G locations in the time frequency grid. This means that the 5G UEs will not be affected by these 6G signals. For example, the 6G CSI-RS locations should be a superset of 5G CSI-RS time/frequency locations. Furthermore, the impact on the network performance and updates to the 5G standard configuration should be minimized. From the 6G perspective, MRSS should preferably not restrict design and operation of a 6G-only carrier and new 6G UEs should be able to avoid 5G signals. It is expected that the overhead using spectrum sharing will be very low and clearly lower than the corresponding spectrum sharing between 4G and 5G in the form of dynamic spectrum sharing (DSS) [HEX225-D35].

Taking into consideration the migration issues and using evolved 5G CN, the expected main interfaces between the UE, RAN and CN in 6G are depicted in Figure 2.2. As can be seen, a Lower Layer Split (LLS) in the RAN will likely be used for 6G, instead of the High-Level Split (HLS) used in 5G. The main reason for this is to simplify the RAN architecture, as well as to support distributed-MIMO. Furthermore, it is expected that

the N2 interface from the User Equipment (UE) to the Access and Mobility Management Function (AMF) may be kept. The main reason for this is a smoother migration from 5G using the same anchor point [HEX225-D35]. However, the N2 interface may need to be evolved compared to 5G in order to handle the demand for a more cloud-friendly environment.

Another aspect to consider is that 6G aims to support new services and use cases. Although AI and ISAC are the most visible examples of new 6G services that will put new requirements on the E2E network, there may be others not yet anticipated. Building upon the Service-Based Architecture (SBA) of 5G, 6G can target streamlined NF design by collocating or refactoring 5G NFs as well as developing new 6G NFs. This streamlined design can target different aspects of the network, e.g. reduced signalling, increased flexibility or communication overhead, etc. Modular networks also bring out challenges in the interactions between different modules and orchestration of the modules and entities.

2.4 MODULAR ARCHITECTURE DESIGN

From the network performance perspective, 6G is envisioned to exceed 5G performance which puts additional requirements to the CN design. In addition to the well-known KPIs, such as lower latency or high throughput, the 6G network needs to ensure streamlined operations with decreased number of standardized APIs, NFs and signalling. Therefore, 6G should further enhance the deployment and operational flexibility by revisiting the 5G CN NF design where needed. It is crucial to find a balance between the granularity of NFs and the number of interactions between system elements. This balance enables flexible and modular addition, update, and replacement of NFs. While a high level of granularity in the control plane has its advantages and disadvantages, understanding the various granularity options and their impact on end-to-end performance is essential to fully grasp the potential of network modularization. The evaluation and analysis of these designs that compass different levels of modular granularity is in line with the pros and cons analysis presented in Table 2.2.

Table 2.2: Advantages and disadvantages of high granularity modules

Advantages	Disadvantages
Efficient resource scaling through scalability management	Need for defined interfaces for cross-vendor deployments

Easier identification and replacement of faulty modules	Increased complexity in integration and testing
Faster development of independent modules	Higher management overhead
Flexible module placement in distributed deployments	Increased signalling and data exchange between modules
Efficient reuse of modules in different implementations	More memory transactions for context data management

A module can encompass different NFs, like microservices, and there are multiple approaches to constructing a network module. Different modularity studies are conducted to investigate the impact of redesigning 5G System (5GS) NFs while varying the granularity of these NFs [HEX225-D35]. One study, called procedure-based decomposition, suggests redesigning the 5G Core NFs so that each NF incorporates the services and processing logic to execute a specific 5G procedure such as UE Registration or PDU Session Establishment, etc. Another study, called Modular UPF, suggests modularizing the 5G UPF to smaller components that includes uplink packet processing logic, downlink packet processing logic, and optional on-demand UPF features such as lawful intercept, etc. On the other hand, it is also possible to preserve highly granular network modules and dynamically allocate them based on the network traffic and required services. Finally, it is also possible to customize both the granularity and the composition of different modules according to the steady state characteristics of the traffic demand and various KPIs. Each one of these methods brings different advantages and challenges, and therefore the method needs to be carefully selected.

A high degree of flexibility is achieved via the interactions between different NFs in 5GC. In 6G, taking advantage of the findings of 5G, the inter-NF or inter-module integration needs to be streamlined, i.e., by optimizing where possible, removing unnecessary interactions, and preserving the already optimum ones. This requires not only optimizing the inter-module interactions but also streamlining the interactions between different entities of the network as the modules can be deployed there (e.g., RAN, CN, edge etc.).

2.5 REFERENCES

[HEX223-D22] HEXA-X-II Deliverable D2.2 – Foundation of overall 6G system design and preliminary evaluation results, https://hexa-x-ii.eu/wp-content/uploads/2024/01/Hexa-X-II_D2.2_FINAL.pdf, December 2023.

[HEX225-D25] HEXA-X-II Deliverable D2.5 – Final overall 6G system design, to be published in April 2025

[HEX223-D43] HEXA-X-II Deliverable D4.3 – Early results of 6G Radio Key Enablers, https://hexa-x-ii.eu/wp-content/uploads/2024/04/Hexa-X-II_D4_3_v1.0_final.pdf April 2024

[HEX225-D35] HEXA-X-II Deliverable D4.3 – Final architectural framework and analysis, https://hexa-x-ii.eu/wp-content/uploads/2025/03/Hexa-X-II_D3.5_v1.0.pdf, February 2025

[HEX225-D65] HEXA-X-II Deliverable D4.3 – Final Design on 6G Smart Network Management Framework, https://hexa-x-ii.eu/wp-content/uploads/2025/02/Hexa-X-II_D6-5_final.pdf February 2025

3 EXTENSIONS

This section focuses on additional innovations and key enhancements needed to enhance the foundational concepts of 6G architecture [NGMN]. These extensions address the integration of emerging technologies, the scalability of existing frameworks, and the adaptation of network functionalities to evolving user demands and environmental constraints. By building on core elements such as resource optimization, security, and interoperability, these extensions aim to future-proof the 6G ecosystem. Key focus areas include expanding the role of AI-driven network intelligence, enhancing the modularity of NFs, and enabling seamless integration with non-terrestrial networks and quantum computing paradigms. Together, these advancements position the 6G architecture to not only meet but exceed the complex demands of next-generation applications and services

3.1 INTEGRATION OF OVERARCHING CONCEPTS INTO THE 6G ARCHITECTURE

This section explores the integration of foundational concepts that guide the 6G architecture, emphasizing the seamless connectivity and interoperability across heterogeneous networks. Key aspects include the unification of distributed sub-networks into a cohesive "network of networks," the adoption of advanced virtualization techniques, and the enhancement of resource management through AI-driven frameworks. These concepts aim to address the growing demands of ultra-reliable, low-latency communication, and energy efficiency while supporting dynamic and adaptive network functionalities. By incorporating overarching goals, the 6G architecture aspires to provide a robust, scalable, and future-ready framework capable of meeting the diverse requirements of next-generation applications.

3.1.1 RESOURCE MANAGEMENT IN THE DEEP-EDGE-EDGE-CLOUD CONTINUUM

The softwarization of entities at the deep edge (such as vehicles or robots), combined with their increasing levels of automation and the virtualization of components, is placing growing flexibility and reliability demands on networks that must support a rising number of compute-intensive applications, functions and control processes (e.g. for autonomous driving), which cannot be efficiently accomplished by

only scaling and re-dimensioning the networks at the deep-edge. The (cost-)efficient execution of these compute-intensive processes can be obtained by leveraging the envisioned 6G deep-edge – edge – cloud continuum, which enables the opportunistic offloading and distribution of processing tasks within the deep-edge, the edge or the cloud [6GSHINE24-D42]. Achieving this vision requires seamless connectivity and integration of the networks at the deep-edge with the 6G parent network. This integration can be facilitated through the (gradual) adoption of wireless sub-networks in entities such as vehicles or robots to the wider 6G network (see Section 3.1.4). For instance, in the scenario of in-vehicle wireless sub-networks, this vision would enable computing entities at the edge or cloud to function as a virtual Electronic Control Unit (ECU), elastically extending the computing and processing capabilities of the in-vehicle network and Electrical/Electronic (E/E) architecture using edge and cloud resources [LCG+24].

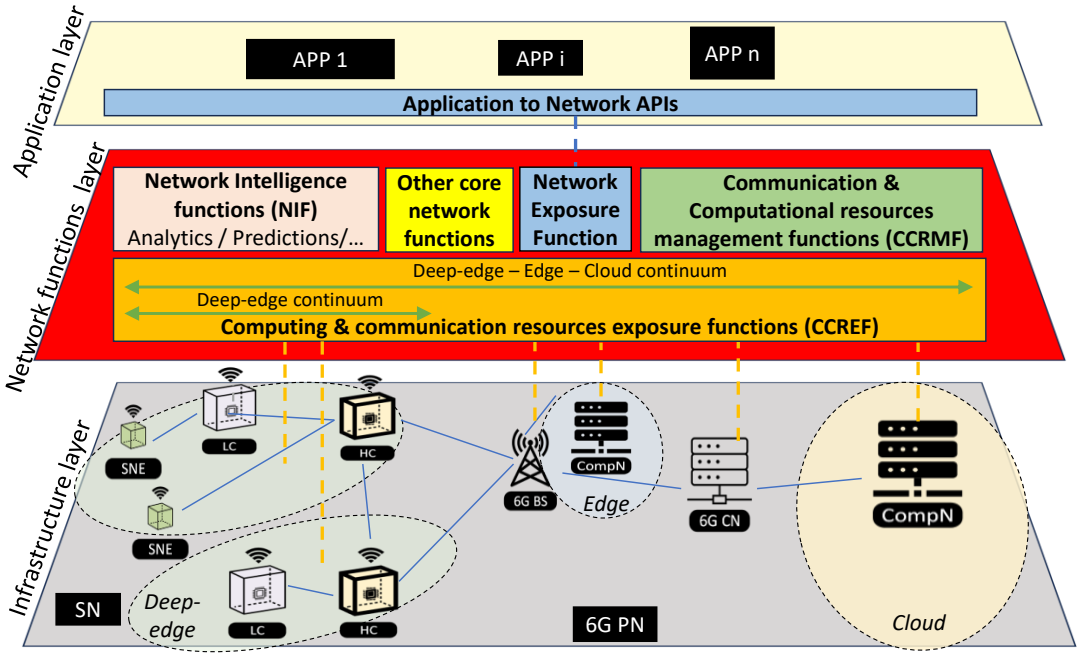


Figure 3.1: System architecture for offloading with dynamic and coordinated resource management in the 6G continuum.

Beyond the architectural components and their associated sub-network control functions required for integrating sub-networks with the 6G parent network [LCG+24], current 5G networks lack the components and interfaces necessary for a joint orchestration/management and context-aware operation of infrastructure (i.e., computing & communication links), NFs and application layers (see Figure 3.2) for the

effective and dynamic completion of compute-intensive (control) process in the deep-edge – edge – cloud continuum.

6G sub-networks at the deep edge will feature distributed communications and processing capabilities for autonomous local data management. They are also being designed to seamlessly integrate with the 6G parent network, forming a deep-edge – edge – cloud continuum. This integration facilitates an elastic continuum orchestration/management approach where computing tasks and communication links are dynamically and jointly scheduled across the nodes forming the continuum. The effective and efficient joint management of connectivity and computing resources across the continuum necessitates the following architectural innovations (see Figure 3.2):

- Communication & Computing Resources Exposure Function (CCREF). 5G SA Network Exposure Function (NEF) mainly focuses on the interface between NFs and Application Function (AF) for traffic management and QoS assignments, through interaction with the policy elements (i.e., Policy Control Function – PCF). In other words, NEF exposes interaction with the NF layer to the application layer. CCREF necessitates open and harmonized interfaces for the exposure of resources and capabilities of any type from the infrastructure layer across the 6G 'network of networks' (i.e., sub-network, 6G parent network, cloud), including computing and communication resources, but potentially others like AI, as well. CCREF is a clear architectural enabler for forming the continuum which could be locally exploited at the deep-edge (deep-edge continuum) and end-to-end (deep-edge – edge – cloud continuum). Network functionalities implemented in the end-to-end system are also necessary to collect those resources and capabilities and make them available to other NFs.
- Communication & Computational Resources Management Function (CCRMF). Extending the capabilities of the 5G PCF, and through the interaction with CCREF, CCRMF enables the implementation of advanced resource management policies that coordinate the scheduling and allocation of the communication and computing resources through the continuum (either at the deep-edge or end-to-end).
- Network Intelligence functions (NIF). Network Data Analytics Function (NWDAF) in 5G networks leverages network data analytics to generate real-time operational intelligence driving network automation and service orchestration.

NIF complements NWDAF by exploiting the CCREF (also NEF from application to NF layers) and deriving AI-driven proactive network management solutions in CCRMF, e.g., through the prediction of the availability of computing and connectivity resources in the deep-edge – edge – cloud continuum.

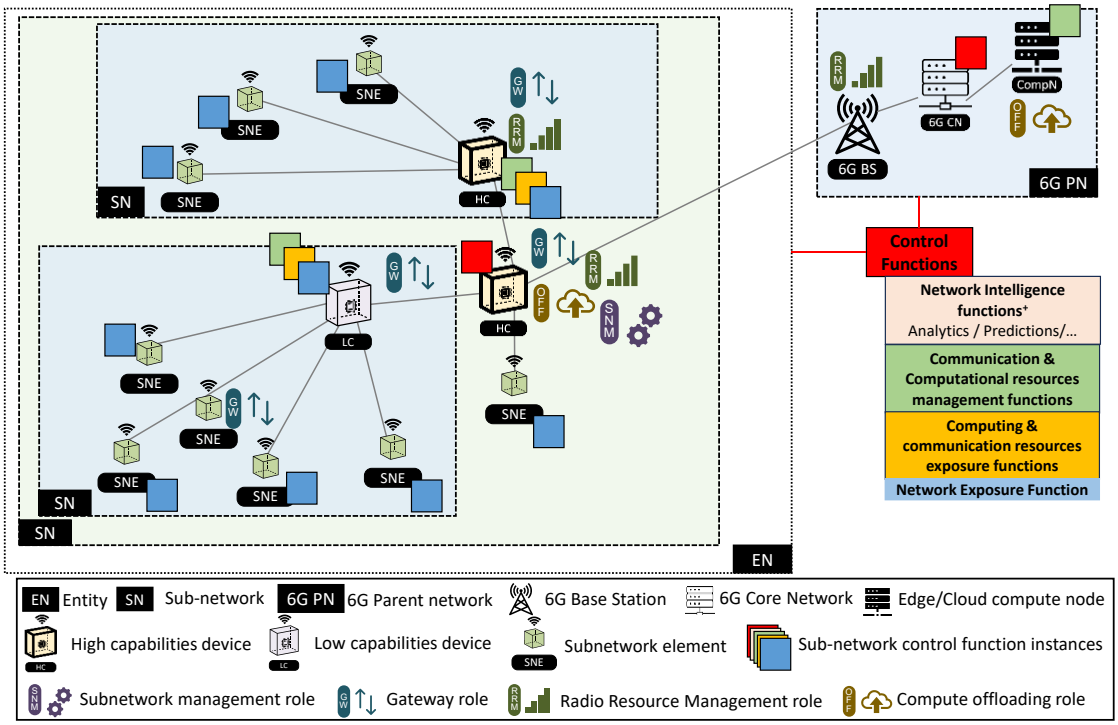


Figure 3.2: Example of realization of the 6G architecture enablers identified in Figure 3.1 for offloading with dynamic and coordinated resource management in the 6G continuum.

Figure 3.2 shows an example of the realization of the system architecture depicted in Figure 3.1 using the sub-network integration with 6G parent network Figure 3.1. For guaranteeing the survivability of sub-networks when the integration with the 6G parent network cannot be established, NFs can flexibly be deployed/instantiated in the sub-network nodes depending on their capabilities. For the isolated operation of sub-networks, Higher Capabilities (HC) devices with SNM/OFF/GW roles are entitled to jointly manage the communication and computing resources exposed by other elements of the sub-network (e.g., LC) to the CCREF, following the policies and procedures of the CCRMF and satisfying the service requirements of the SNEs exposed through the NEF. The flexible deployment of the network layer functions also allows for the centralized and partially decentralized (via functional splitting between the sub-network and the 6G-parent network) management of computing and connectivity resources of the deep-edge – edge – cloud continuum.

3.1.2 SUSTAINABILITY AND COMPUTE CONTINUUM

The interaction of NFs with the underlying infrastructure may pose significant inefficiencies if not properly performed [ORIG24-D21].

The virtualization of Radio Access Networks (vRANs) offers numerous benefits, such as reduced vendor lock-in and resource efficiency, yet it faces significant challenges, particularly with latency-sensitive tasks like LDPC decoding. Current solutions often rely on expensive and energy-intensive hardware accelerators, such as ASICs and GPUs, to meet the strict performance requirements of 5G, which raises concerns about sustainability.

Additionally, the integration of 6G virtual NFs (VNFs) with O-RAN's Radio Intelligent Controller (RIC) is complicated by poor interoperability among network components, leading to potential conflicts and inefficiencies. The disaggregated nature of O-RAN, coupled with the need for real-time AI-driven decision-making, further exacerbates these challenges.

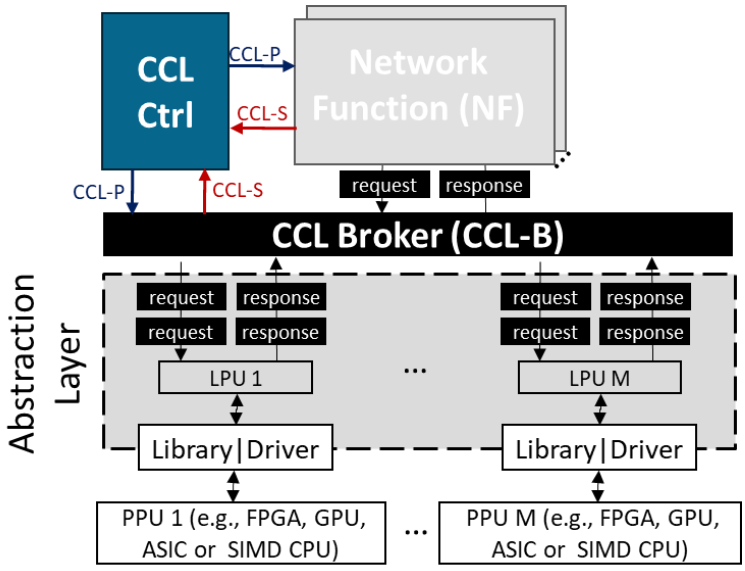


Figure 3.3: The compute continuum layer

Network Intelligence (NI) solutions in 6G also struggle with latency and inefficiency, often due to a misalignment between AI/ML algorithms and the underlying network infrastructure. This misalignment can result in excessive data transmission and suboptimal performance. A more tailored approach that leverages edge computing and optimizes the interaction between NI algorithms and network infrastructure is necessary for improving efficiency and responsiveness.

Finally, the potential of modern programmable transport technologies, like smartNICs and Network Processing Units (NPUs), is underutilized due to the complexity of programming in-band computing models. These limitations prevent the full exploitation of these technologies, hindering the development of innovative 6G services that require real-time processing at line rates. Simplifying the deployment of sophisticated user-plane VNFs is crucial for unlocking new applications and markets.

The Compute Continuum Layer (CCL) [ORIG24-D21] depicted in Figure 3.3 is an architectural innovation designed for 6G systems, facilitating the execution of network processing workloads across diverse computing resources. This architecture aims to streamline resource sharing and fully utilize the capabilities of a heterogeneous computing environment, encompassing GPUs, TPUs, FPGAs, ASICs, NPUs, smartNICs, and quantum computers, alongside traditional CPUs.

The CCL introduces compute-aware network operations while preserving the abstraction of a pure virtualization layer. This allows NFs to be optimized by matching them with appropriate computing resources, leading to enhanced efficiency and reduced resource usage. Real-time resource management is a key feature, enabling dynamic allocation and policy enforcement to maximize resource utilization across the edge-to-cloud continuum.

A centralized abstraction layer within the CCL simplifies the interaction between NFs and hardware components, improving resource management and scalability. Additionally, the CCL accommodates new network sensing functionalities, benefiting tenants interested in metadata over data transport. By regulating virtualized NFs and supporting operational decisions, the CCL ensures their correct performance, particularly in scenarios with stringent time constraints, such as virtualized radio access networks (vRANs).

The CCL provides a robust API that abstracts the heterogeneous and disaggregated computing infrastructure, making it accessible to the 6G network's software components. This API also offers tools for exploiting the underlying infrastructure and implementing policies that govern its usage. This approach prevents vendor lock-in while ensuring optimal performance in the demanding 6G environment.

The CCL architecture addresses critical challenges in 6G networks, such as unsustainable RAN virtualization, poor interoperability, high latency in processing complex network tasks, underutilized programmable transport, and inadequate data

representation. It achieves this by accessing various physical processing units through an abstraction layer, which homogenizes these resources as Logical Processing Units (LPUs). During operation, the CCL manages processing requests from NFs, ensuring that tasks are efficiently routed and executed within set policies and deadlines, thereby optimizing cost and energy consumption while maintaining required performance levels.

3.1.3 ZERO-TRUST SECURITY AND INTEROPERABILITY

The current telecommunications landscape faces challenges due to the lack of standardized global service APIs and the siloed development of network management, orchestration, and control functions. This fragmentation hinders cross-domain interaction, limiting the effectiveness of network slicing, especially in business-to-business contexts. As network slicing becomes more crucial, the need for modular, interoperable NFs and flexible data exchange across domains grows. The adoption of a publish-subscribe methodology and enhanced configurability of NFs are essential steps toward achieving real-time, automated optimization using AI and big data solutions [ORIG24-D21].

Global mobility for IoT devices also presents challenges, as current international roaming models rely on outdated, trust-based agreements between mobile network operators (MNOs), leading to performance penalties and increased costs. A decentralized identity model that decouples user authentication from home operators could enable more efficient local breakout, reducing the need for costly international data routing.

Moreover, the current approach to privacy, security, and data representation in cellular networks is insufficient for the demands of global connectivity and IoT deployments. Improved data governance and high-quality data provisioning are critical for supporting advanced network intelligence (NI) functionalities and enabling precise management of infrastructure.

Finally, while the service-based architecture (SBA) introduced in 5G offers flexibility, it has led to a surge in signalling traffic, challenging scalability and cost efficiency. Optimizing signalling management is vital as networks evolve to support an increasing number of connected devices and complex IT integrations.

To meet the demands of future global connectivity, there is a need for a comprehensive overhaul of network architectures, focusing on standardized APIs,

modular NFs, decentralized identity models, and efficient signalling management. These improvements are essential for unlocking the full potential of 6G networks. Simplifying the deployment of sophisticated user-plane VNFs is crucial for unlocking new applications and markets.

The Zero-Trust Layer (ZTL) [ORIG24-D21] aligns service providers' internal operations with network operators' continuous optimization efforts, thereby unlocking advanced functionalities like remote sensing and digital twinning. By fostering a cooperative control loop, the ZTL enables service providers to have a more direct influence on network operations, while ensuring privacy and security—like how hyperscale cloud services operate today.

The ZTL offers both vertical and horizontal exposure. Vertically, it enhances network analytics by integrating feedback from service providers into the Network Data Analytics Function (NWDAF), enabling more precise customization without compromising confidential information. This approach allows service providers to optimize their own metrics in line with network quality of experience (QoE), which may differ from standard network metrics due to business-specific factors.

Horizontally, the ZTL supports global operations, particularly for IoT devices, by facilitating efficient international roaming and enabling new business models. The architecture envisions a decentralized identity system, decoupling user authentication from the connectivity services provided by home operators. This allows visited operators to directly charge global end-users while giving home operators full visibility into these transactions. The ZTL also incorporates distributed ledger technology for secure, immutable record-keeping and network intelligence modules for real-time anomaly detection, enhancing the security, privacy, and operational efficiency of global networks. This forward-thinking architecture aims to transform traditional network interactions and meet the high demands of next generation 6G networks, ensuring seamless, global connectivity for a diverse range of devices and services.

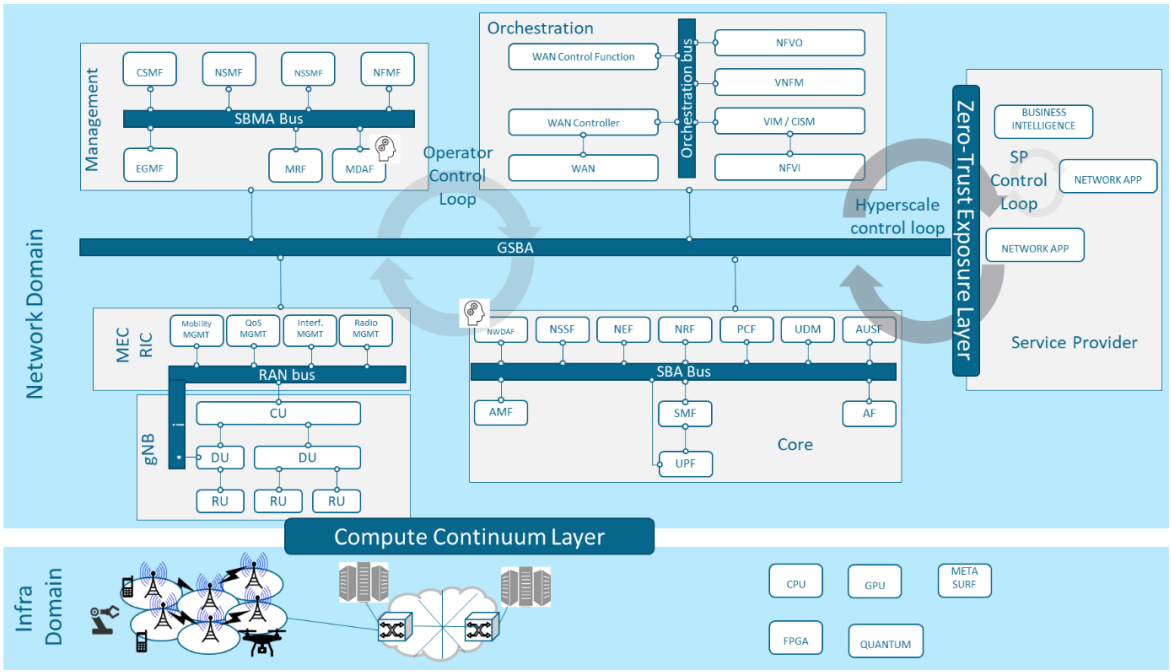


Figure 3.4: The Integration of the ZTL in the Overall 6G Network Architecture.

3.1.4 SUB-NETWORK INTEGRATION IN 6G NETWORKS

6G networks aim to become a "Network of Networks" (NoN) that integrates sub-networks with diverse capabilities, requirements, and characteristics in terms of KPIs/KVIs, coverage, operational purpose, and spectra. These 6G sub-networks could include Non-Terrestrial Networks (NTN) (see Section 4) as well as short-range, low-power radio cells for localized and sustainable deployments in entities at the deep edge, such as within vehicles, robots, or classrooms, replacing their (existing) wired connections [BAC23]. Sub-networks at the deep-edge should be able to operate to certain extent independently while being able to communicate with a parent “umbrella” 6G network. Current 5G and beyond networks lack the necessary mechanisms and architectural enablers for the seamless integration of sub-networks at the deep edge with the 6G parent network, which challenges providing the dependable service levels demanded in sub-networks, comparable to those achieved with wired connections. The service level requirements of these sub-networks could include extreme demands in terms of latency, reliability, throughput, communication cycle, and determinism, which current sub-network-like 5G solutions (such as Personal IoT Networks or Ambient IoT Networks) do not target.

Sub-networks at the deep edge of the 6G NoN will require novel mechanisms and interfaces that enable their efficient, flexible, and resilient integration and (co-) operation/interworking with the 6G parent network. In this respect, suitable control

functions for the interplay between sub-networks and the respective 6G parent network must be defined, for example, in areas such as Radio Resource Management (RRM), security, network management, and interference management, or to support the dynamic and opportunistic offloading of certain functionalities from local processors/computing nodes in the sub-network (e.g., in a vehicle or robot) to remote computing units accessed through the 6G parent network (e.g., edge server/MEC or the cloud).

The integration of sub-networks in the 6G NoN requires specific architectural components, sub-network control functions and interfaces that allow dynamic discovery of the sub-network nodes and their capabilities such as communication, computation, AI and power [6GSHINE24-D22]. Sub-network nodes could be categorized as High Capability (HC), Low Capability (LC) and Sub-Network Elements (SNE) depending on their capabilities and reflecting the degree to which they can (autonomously or in coordination with the 6G parent network) take on certain (sub-) network control and management roles. These roles could be represented as clusters of high-level functionalities such as communication, management and computation. Without loss of generality, these roles could include:

- Sub-Network Management (SNM): A sub-network node with SNM functionality manages the operational activities of nodes within a sub-network. This might include authentication, handover procedures, master clock roles, and monitoring of network performance. SNM nodes can also share the required configuration and management functions with the 6G parent network to control the sub-network.
- Gateway (GW): A sub-network node with the GW role can manage the data traffic routing within and/or across sub-networks. It can act as intra-sub-network or cross-sub-network relay as well as a gateway towards the 6G parent network.
- Radio Resource Management (RRM): A sub-network node with the RRM role uses its capabilities to manage the radio resources of one or multiple other nodes within a sub-network. Distributed and centralized (with the assistance of the 6G parent network) RRM functions aim at maintaining the QoS requirements within the sub-network.
- Compute OFFloading (OFF): A sub-network node with the OFF role uses its capabilities to orchestrate application and/or NF offloading from source

elements to target elements. Target elements can be within the sub-network or can be accessed through the 6G parent network. OFF node is also a provider or donor of computation resources to another element within the same or another sub-network.

6G architectural components and their corresponding sub-network control functions are also necessary for assigning these roles to sub-network nodes (i.e., HC, LC and SNE) and ensuring their discoverability within the sub-network. To enhance survivability, 6G sub-network control functions must also ensure a degree of autonomy from the 6G parent network. Additionally, they should support mechanisms that enable dynamic role changes among sub-network nodes, allowing the sub-network topology and the distribution of application and NFs to be continuously adapted within the sub-network and/or in coordination with the 6G parent network. 6G sub-network control functions should also account for challenging scenarios that arise from the potential mobility of sub-network nodes, both when they join or leave the subnetwork and when the entire sub-network moves across the coverage areas of different 6G parent networks. 6G sub-network control functions should also account for the potential temporal nature of sub-networks such as when they are established for time-bound tasks.

3.2 INTEGRATION OF 6G NETWORK PARADIGMS

This section discusses the seamless blending of network slicing, multi-access edge computing (MEC), and cloud-edge continuum strategies, which will ensure enhanced performance, resource efficiency, and service customization to meet the heterogeneous demands of advanced applications. By aligning these paradigms under a unified framework, 6G networks can provide dynamic, scalable, and intelligent services, supporting challenging use cases such as real-time extended reality, autonomous systems, and smart infrastructure. This integration is critical to unlocking the full potential of 6G networks and achieving global connectivity goals.

3.2.1 NETWORK SLICING AND MULTI-ACCESS EDGE COMPUTING (MEC)

Network slicing is a key feature of 5G and beyond systems, designed to simultaneously support multiple services with heterogeneous requirements (e.g., data rates, latency, reliability, availability). It allows creation of multiple end-to-end logical

networks, denoted as network slices, over a common physical infrastructure. Each network slice is optimized in accordance with the requirements of a set of services to be given within the slice.

While the network slice concept is already quite consolidated, a new dimension of slicing that encompasses the edge computing domain has been recently introduced by the ETSI MEC standardization group with the so-called *MEC Application slices* [ETSI22-MEC038], [ZLL+22]. This new dimension arises from a customer-driven perspective, where the customer uses a virtualized application (e.g., XR application, AI application), referred to as a MEC application (MEC App), that needs to run on the MEC system to perform computations. In this case, the MEC App cannot be considered as a part of the network slice only, since its requirements go beyond those of the network (i.e., data rate, latency, reliability) to include others such as computing resources, isolation at application level, virtualization approach (e.g., deployment as virtual machines, containers), etc. This motivates the introduction of a new MEC application slice as an independent entity from the network slices, which relies on similar concepts in terms of isolation and QoS guarantees but adapted to the MEC system [ZLL+22].

The 3GPP has defined a general architectural framework for management and orchestration [3GPP24-28.533] [3GPP23-28.530], while the concept of network slice management is introduced in [3GPP23-28.530]. Although the 3GPP architecture for network slice management in [3GPP24-28.533] is general and provides room for different implementations, a commonly considered approach is given in [BTB22] and is taken as a reference for the proposed slice manager in Figure 3.5. Specifically, the management of network slices is based on three functions:

- Communication Service Management Function (CSMF): This function is the user interface for slice management and converts the Service Level Agreements (SLAs) (i.e. the business contract between the service provider and the client that specifies the service levels to be ensured) into the Service Level Specification (SLS), which includes the set of technical attributes that have to be satisfied by the network slice.
- Network Slice Management Function (NSMF): This function is responsible for the management and orchestration of the network slice instance (NSI) to fulfil the SLS specified by the CSMF. This includes the different stages of the lifecycle management of an NSI, namely the commissioning (i.e. the slice creation, in which the necessary resources are allocated and configured based on the SLS

requirements), the operation (i.e. the activation, supervision, performance reporting, modification, and de-activation of the NSI) and the decommissioning (i.e. the termination of an NSI when it is no longer needed).

- Network Slice Subnet Management Function (NSSMF): The NSMF splits an NSI into its subnet slice instances, i.e. RAN slice, TN slice and CN slice, indicating for each one the SLS to be fulfilled. Then, there is an NSSMF taking care of the lifecycle management of each subnet slice, namely the RAN NSSMF, the TN NSSMF and the CN NSSMF as depicted in Figure 3.5.

Regarding the MEC App slice management, the following two functionalities defined by ETSI MEC in [ETSI24-MEC044] are to be considered:

- MEC Application Slice Communication Service Management Function (MAS-CSMF): This function is responsible for translating high level service-related QoS requirements into MEC App slice requirements. Moreover, it also facilitates the purchase and monitoring of MEC App slices for the customer, e.g. through the exposure of service performance and alarm information.
- MEC Application Slice Management Function (MAS-MF): This function takes care of the design of the MEC Application Slice template (MAST) and of the lifecycle management of the MEC App slice instances according to the requirements specified by the MAS-CSMF. The MAST is a collection of parameters that define an information model including the MEC App slice identifier, the name, the designer, the version, the release time and the description. In turn, the lifecycle management consists of the creation, activation, operation and release of the MEC App slice instances across the MEC system. This can involve aspects such as the selection of the most appropriate MEC hosts at the edge sites of the edge-to-cloud continuum to fulfil the QoS requirements established in the MAST.

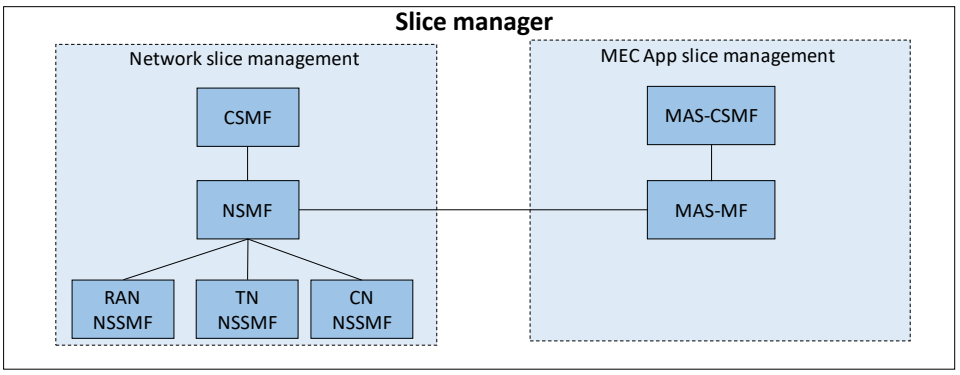


Figure 3.5: Network Slice Management at the Edge

This vision for joint management of MEC App slices and network slices is illustrated in Figure 3.6, which shows an example with two services, namely an XR service and a holographic communications service. Each one is supported by a different network slice, composed of RAN, TN and CN subnet slices, and a MEC App slice available at the MEC host of an edge-site, which in this specific case is co-located with the gNB. Each MEC App slice contains the required MEC Apps to support the computation task associated with the service. The gNB can forward tasks to the MEC host through the local breakout mechanism, which selects the IP-based traffic to be forwarded to the local User Plane Function (UPF) and from there to the MEC host. For instance, the rendering tasks of a user of the XR service are sent to the gNB through RAN slice #i, and then forwarded to the MEC host, which processes them in the rendering MEC App of MEC App slice #X.

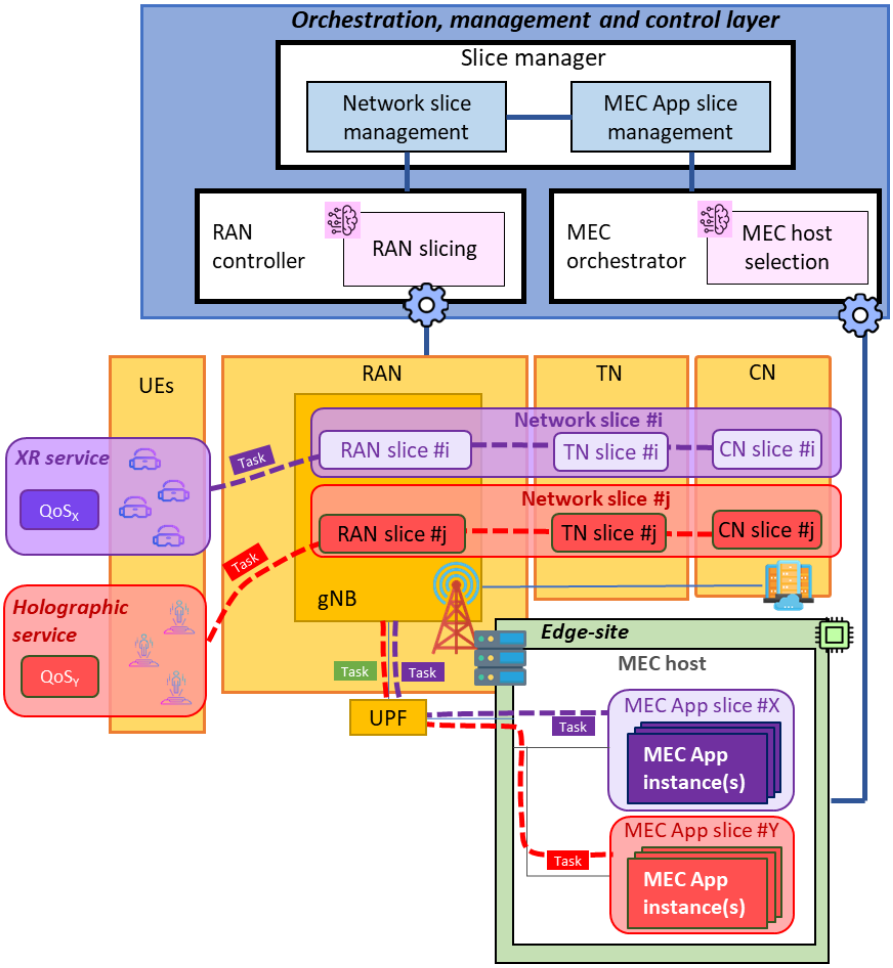


Figure 3.6: Vision for joint management of MEC App slices and network slices

As seen in Figure 3.6, the slice manager needs to interact with other architectural components of the orchestration, management and control layer in order to support the

operation of the different management functionalities. Specifically, the operation of the RAN slices conducted by the RAN NSSMF involves the allocation of the radio resources at the different cells to the RAN slices. For this purpose, the RAN NSSMF can rely on the RAN intelligent controllers to conduct this allocation based on specific algorithmic solutions that can involve the use of AI models. Depending on the solution, this can be done with RAN controllers having a global scope involving multiple gNBs, or having a local scope for solutions involving the allocation of resources to slices separately for each gNB.

For the MEC App slice management, the MAS-MF of the slice manager needs to interwork with the MEC orchestrator (MEO) functionality considered by ETSI MEC in [ETSI24-MEC044]. This orchestrator has an overall view of the MEC system across the different edge sites of the edge-cloud continuum (i.e., deployed MEC hosts, available resources, MEC services and topology) and will take care of supporting the MEC App slice lifecycle management operations of the MAS-MF, e.g. through onboarding application packages and selecting appropriate MEC hosts for the instantiation of MEC App instances.

3.2.2 TIME-CRITICAL AND DETERMINISTIC COMMUNICATIONS

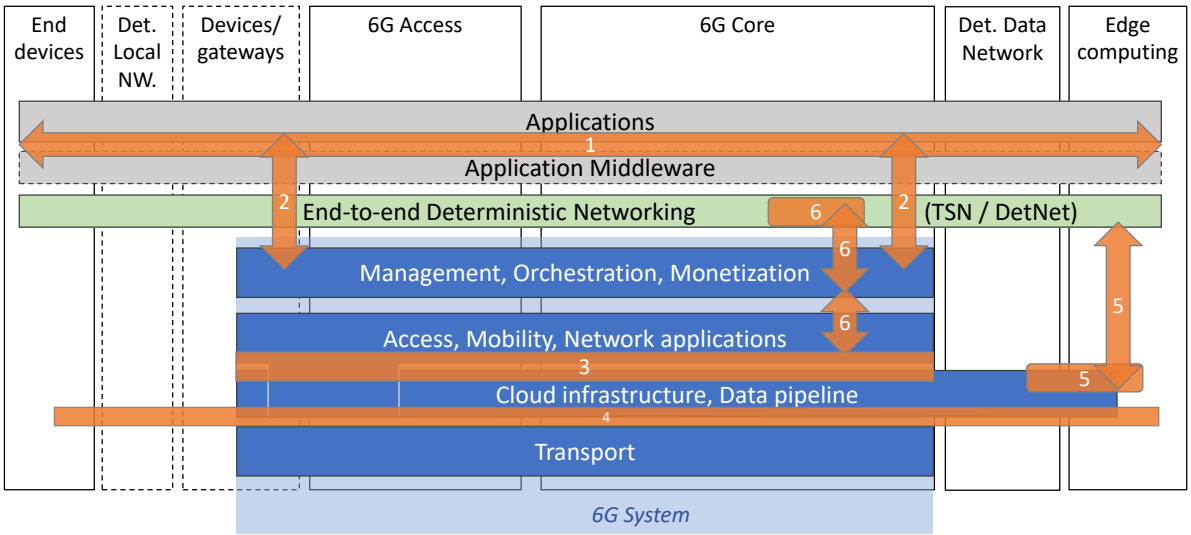


Figure 3.7: Integration of DetNet components in the network architecture

Ensuring dependable end-to-end time-critical communications across diverse communication and compute domains is critical for supporting emerging applications, such as XR, smart farming, and adaptive manufacturing [DET6G23-D11]. These applications demand stringent performance metrics and guarantees, particularly in terms of hyper-reliable and low-latency communications. However, despite advances

in latency and reliability, both computation (e.g., edge computing) and communication (e.g., 5G URLLC) domains exhibit substantial stochastic variations compared to wired deterministic communications technologies like time-sensitive networking (TSN) and deterministic networking (DetNet), particularly in terms of delay characteristics and non-negligible packet delay variations. This leads to the traditional approach for end-to-end resource planning falling short regarding service performance, scalability and efficiency, especially when integrating stochastic elements and deterministic communication technologies to support time-critical services [DET6G23-D22] [DET6G23-D31]. Future networks must address these challenges fundamentally in the design, planning, and operation of time-critical networks, focusing not on achieving purely deterministic performance at the cost of resources but on embracing stochasticity while ensuring application dependability.

In reference to a horizontal 6G architecture, the following areas need attention (as indicated in Figure 3.7 above and described in [DET6G24-D12]):

1. Support for end-to-end time-critical applications [DET6G23-D11].
2. Advanced network configuration via network exposure, to invoke dependable communication in accordance with a well-defined service specification for requesting a dependable communication service from the network [DET6G24-D12].
3. The 6G network needs to provide a dependable communication service, which means that it must be able to comply with and deliver the performance that is requested from the applications by the end users. This includes being able to monitor the KPIs that characterize the delivered service performance and provide a basis for data-driven (latency) performance prediction, so that the 6G network can determine which (latency) performance levels it can promise to what reliability level [DET6G23-D21]. One important characteristic is to also be able to control the packet delay variation as explained in [DET6G23-D31].
4. Dependable time-critical communication builds on time-awareness throughout the system, which is based on robust time synchronization [DET6G23-D22].
5. To enable edge computing-based realizations of time-critical services – including the integration with deterministic TSN/DetNet networking – dependable time-aware compute features need to be provided by the edge [DET6G24-D12].

6. When considering latency variations of sub-components in an end-to-end system, the (TSN/DetNet) end-to-end traffic handling should be aware of the latency characteristics of sub-components in order to enable robust and optimized end-to-end deterministic network configurations [DET6G23-D31].

The proposed approach for end-to-end dependable time-critical communication advocates the following: (I) the acceptance and integration of stochastic elements, such as wireless links and computational elements, with a focus on characterizing their stochastic behaviour, and monitoring and predicting KPIs, such as latency or reliability, which can be leveraged to make individual elements plannable despite the presence of stochastic influences. Nevertheless, system enhancements to mitigate stochastic variances in communication and computational elements are required, by means of mechanisms such as packet delay corrections [DET6G23-D21]. (II) the management of the entire end-to-end interaction loop (e.g., the control loop from the sensor to the controller to the actuator), considering the underlying stochastic characteristics, especially with the integration of compute elements. (III) the adaptation between applications running on converged network infrastructures due to unavoidable stochastic degradations of individual elements. In other words, application requirements can be adjusted based on prevailing system conditions leading to more flexibility in the operations.

The approach builds on the concept of time-awareness by ensuring accurate and reliable time synchronicity while also incorporating security-by-design for dependable time-critical communications. Generally, the notion of deterministic communication, where the behaviour of network, compute nodes, and applications is pre-determined, is extended towards dependable time-critical communication, where the focus is on managing communication (and compute) characteristics to provide the KPIs and reliability levels required by the application. Architectures and algorithms are facilitated for scalable and converged future network infrastructures that enable dependable time-critical communication end-to-end, across domains, including 6G.

3.3 TOWARDS A GLOBAL SBA

This section outlines the vision for a Global Service-Based Architecture (GSBA) that unifies service declaration and management across diverse domains, including network operators, infrastructure providers, and application ecosystems. The GSBA aims to overcome trust and interoperability challenges by leveraging modular, API-driven

frameworks that enhance scalability, flexibility, and automation. By integrating advanced conflict resolution mechanisms, such as those used in RAN Intelligent Controllers (RIC), and ensuring seamless interaction between heterogeneous network components, the GSBA paves the way for innovative business models and efficient resource utilization. This architecture forms the backbone for enabling global connectivity, unlocking new opportunities for collaboration and service delivery in the 6G era.

3.3.1 GLOBAL SERVICE BASED ARCHITECTURE

Both the Cloud Continuum and the Zero Trust Functionality, along with other legacy domain buses such as 3GPP SBA, rely on the Global SBA (GSBA), which is designed to facilitate the management of services across different domains. In this context, a "domain" includes the radio access network, core network, and international carrier network.

A significant challenge within this ecosystem is the inherent lack of trust between these entities, which hampers resource sharing and the adoption of innovative business models. While legacy domain buses like 3GPP SBA provide essential support to GSBA, there are domains where the development of new buses is necessary.

3.3.2 BRINGING THE GSBA TO THE RAN

One possible architectural option is to bring the Global SBA to the RAN, as also recently promoted by the O-RAN Alliance [ORAN-SBA]. The RAN bus operates within the RAN Intelligent Controller (RIC) platform, playing a key role in enabling the collection of key performance measurements (KPM) from RAN nodes and facilitating RAN control (RC) decisions for infrastructure management.

The RIC platform hosts multiple xApps that utilize the RAN bus for a variety of functions. For instance, xApps collect RAN node performance data using E2 service model KPMs (E2SM-KPM), a service defined by the O-RAN Alliance. Additionally, xApps can adjust Information Elements (IEs) within specific signalling messages through the RAN node using E2SM RAN Control (E2SM-RC), avoiding the need to decode entire network messages. Conflicts may arise within the RIC when xApps—especially those developed by third parties—modify IEs in ways that are incompatible with each other. These conflicts are categorized as direct, indirect, and implicit, each requiring specific resolution strategies. The RIC's conflict management mechanisms are critical in addressing these issues. Techniques such as post-action verification and tailored

approaches for managing indirect and implicit conflicts ensure the smooth interoperability of xApps. By effectively resolving these conflicts, third-party xApp developers can enhance RAN node performance, with each xApp focusing on optimizing specific performance metrics.

Introducing SBA into the RAN brings several benefits to network operations and management that are not possible with the current point to point approach.

- Improved Scalability: Service-based architectures (SBA) decouple NFs into modular services, allowing networks to scale dynamically based on demand. This flexibility is essential for handling traffic spikes or expanding capacity without significant hardware investments.
- Enhanced Flexibility and Modularity: The modular design of SBA enables independent development, deployment, and management of NFs. This approach supports agile updates and innovation without disrupting the entire network.
- Improved Automation and Orchestration: SBA supports advanced automation tools and orchestration frameworks. By using programmable interfaces and machine-readable APIs, networks can automate tasks like resource allocation, fault detection, and recovery.
- Better Resource Management: Fine-grained control over individual services enables better monitoring and allocation of resources.
- Resilience and Reliability: SBA supports fault-tolerant designs where failures in one service do not cascade across the network. This architecture improves the overall reliability and uptime of the network.
- Future-Proofing: With its modular and API-driven approach, SBA is well-suited to adapt to evolving standards, protocols, and technologies, ensuring long-term relevance and reduced need for overhaul

3.4 REFERENCES

[3GPP23-28.530] 3GPP TS 28.530 v18.0.0, "Management and orchestration; Concepts, use cases and requirements (Release 18)", December, 2023.

[3GPP24-28.533] 3GPP TS 28.533 v18.2.0, "Management and orchestration; Architecture framework (Release 18)", June, 2024.

[6GSHINE24-D42] 6G-SHINE, Deliverable D4.2, - “Preliminary results on the management of traffic, computational and spectrum resources among subnetworks in the same entity, and between subnetworks and 6G network”, https://6gshine.eu/wp-content/uploads/2024/11/D4.2_Preliminary-results-on-the-management-of-traffic-v1.0.pdf

June 2024.

[6GSHINE24-D22] 6G-SHINE, "D2.2-Refined definition of scenarios, use cases and service requirements for in- X subnetworks", Feb. 2024

[BTB22] S. Bolettieri, D. Thai, R. Bruno, "Towards end-to-end application slicing in Multi-access Edge Computing systems: Architecture discussion and proof-of-concept", in Future Generation Computer Systems, vol. 136, November 2022, pp. 110-127.

[BAC23] G. Berardinelli, R. Adeogun, B. Coll-Perales, J. Gozalvez, D. Dardari, E.M. Vitucci, C. Hofmann, S. Giannoulis, M. Li, F. Burkhardt, B. Priyanto, H. Klessig, O. Ognenoski, Y. Mestrah, T. Jacobsen, R. Abreu, U. Virk, F. Foukalas, "Boosting Short-Range Wireless Communications in Entities: the 6G-SHINE Vision", Proceedings of the IEEE Future Networks World Forum, 13-15 November 2023, Baltimore, MD, USA. DOI: 10.1109/FNWF58287.2023.10520553

[DET6G23-D11] DETERMINISTIC6G, Deliverable 1.1, "DETERMINISTIC6G use cases and architecture principles," Jun. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET6G23-D21] DETERMINISTIC6G, Deliverable 2.1, "First report on 6G centric enablers", Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET6G23-D22] DETERMINISTIC6G, Deliverable 2.2, "First Report on the time synchronization for E2E time awareness," Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET6G23-D31] DETERMINISTIC6G, Deliverable 3.1, "Report on 6G convergence enablers towards deterministic communication standards," Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET6G24-D12] DETERMINISTIC6G, Deliverable 1.2, "First report on DETERMINISTIC6G architecture," April 2024, https://deterministic6g.eu/index.php/library-m/deliverables

[ETSI22-MEC038] ETSI GR MEC 038 v3.1.1, "Multi-access Edge Computing (MEC); MEC in Park Enterprises deployment scenario", November, 2022.

[ETSI24-MEC044] ETSI GR MEC 044 v3.1.1, "Multi-access Edge Computing (MEC); Study on MEC Application Slices", April, 2024.

[LCG+24] L. Lusvarghi, B. Coll-Perales, J. Gozalvez, K. Aghababaiyan, M. Almela, M. Sepulcre, "Characterization of In-Vehicle Network Sensor Data Traffic in Autonomous Vehicles", Proceedings of the 15th IEEE Vehicular Networking Conference (VNC 2024), 29-31 May, 2024, Kobe, Japan.

[NGMN] NGMN "Network Architecture Evolution Towards 6G", Feb 2025.

[ORIG24-D21] ORIGAMI Deliverable D2.1, M. Skarp, "Initial report on requirements and definition of KVIs and KPIs", Zenodo, Jun. 2024. doi: 10.5281/zenodo.12580929

[ORAN-SBA] O-RAN next Generation Research Group (nGRG) "Research Report on Service-based RAN for 6G Network", October, 2024.

[ZLL+22] H. Zhu, J. Liu, Y. Lin, Q. Wang, "MEC Application Slice and Its collaboration with 5G network slice," 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, 2022, pp. 1-6.

4 UBIQUITOUS NETWORKS

6G networks, aiming at the most comprehensive access to digital services in the world, will provide near-universal, seamless wireless connectivity across a wide range of locations, including both urban and rural areas, essentially offering Internet services almost anywhere on Earth, thanks to advanced technologies such as distributed (and cell-free) multiple-input-multiple-output (MIMO) systems for dense urban areas and the integration of non-terrestrial networks (NTN) for areas with poor terrestrial coverage, which will effectively eliminate coverage gaps and deliver consistent high-speed connections regardless of the user's location.

Ubiquitous connectivity is now an explicit usage scenario to be addressed by the IMT-2030/6G system, and supported by 6G technologies. Resiliency in ensuring service availability is a significant challenge for this scenario, where NTN, despite having limited capacity, can contribute as a back-up network infrastructure to the terrestrial network component for conveying network traffic. The NTN component itself can be designed for intrinsic resiliency through a redundant multi-layer infrastructure, combining multi-orbit satellite access (e.g., using both Geostationary Orbit (GSO) and Non-Geostationary Orbit (NGSO) satellite constellations).

Fixed wireless access (FWA), and the integration of transport network with distributed MIMO (dMIMO) and Cell-Free MIMO (CF-MIMO) systems form an innovative architecture to support high-capacity coverage in highly dense urban environments. The use of sensing data in communications, e.g., in Integrated Sensing and Communications (ISAC) systems, can improve network efficiency as well as extending network coverage.

4.1 UBIQUITOUS COVERAGE VIA 6G NTN ARCHITECTURE

For the same use cases, in comparison to 5G NTNs, 6G NTNs can offer higher service performance and Quality of Experience (QoE), including the terminal design that should be adapted to the operational constraints.

The underpinning concept of a 6G NTN is a 3D multi-layered architecture [6GNTN24-D35]. The “3D” characteristic stems from the native unification of the non-terrestrial component with the terrestrial one, while the “multi-layered” feature is related to the

communication nodes flying at different altitudes, i.e., satellites or aerial nodes such as high-altitude platform stations (HAPSs).

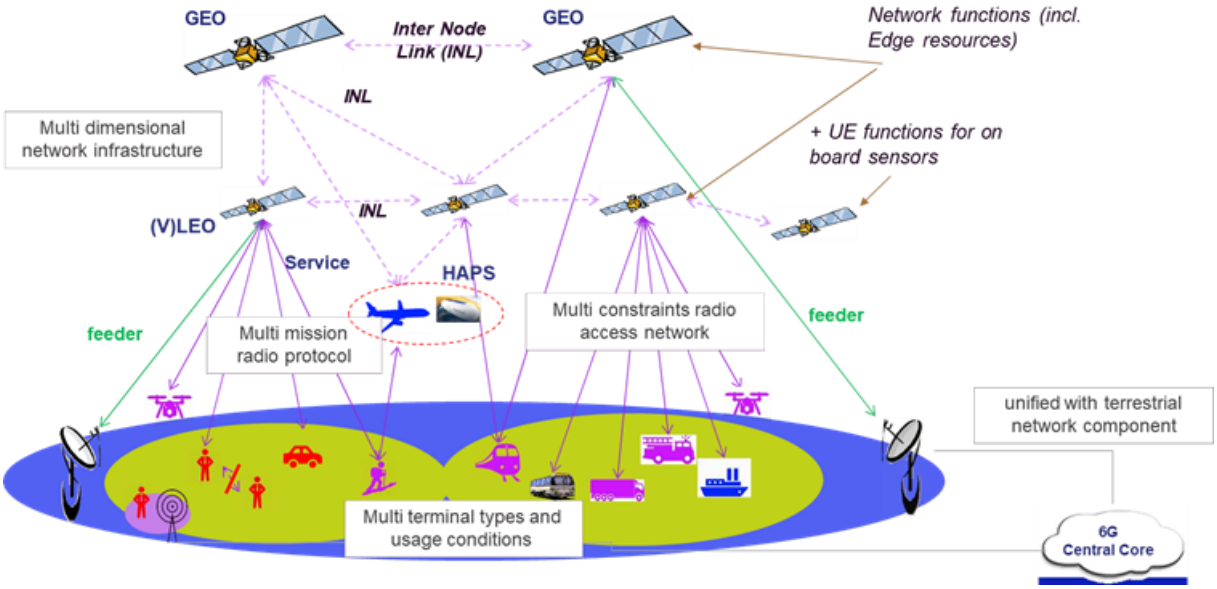


Figure 4.1: The 6G-NTN architecture concept

Two types of non-terrestrial nodes could be considered, namely deterministic nodes with fixed and predictable orbits (both Geostationary and non-GSO satellites in low Earth orbits) and aerial nodes, namely HAPS or special heavy drones, which might be present at different points in time at different locations to extend coverage or enhance the network capacity. The latter are supposed to be deployed "opportunistically" depending on specific needs, and they are not meant to be a permanent infrastructure with global coverage. Target frequency bands are C-band for low data rate services to UEs with hemispherical antennas, and Q/V-band for high data rate services to UEs with directive flat panel antennas. Moreover, optical Inter-Node Links (INL) are foreseen, with the only exception of Low Earth Orbit (LEO)- Geostationary Earth Orbit (GEO) links, where a dedicated radio frequency allocation in Ka-Band can be exploited. All RAN functionalities and eventually some core network functionalities are envisaged to be implemented in space in order to reduce latency (especially on the control plane) and allow connectivity between UEs via only satellites. While the current regenerative satellite architecture generally assigns one gNB per satellite, future deployments, especially in the LEO segment, may adopt a distributed approach, partitioning RAN functionalities across several satellites to address mass and power limitations. Last but not least, the 6G NTN space segment will both provide and rely on reliable UE positioning services, as accurate location information is essential for mobility and

resource allocation decisions, prompting initial 3GPP releases to require UEs to support location-determination capabilities.

The following applications can represent use cases for a unified 3D network integrating terrestrial, airborne and satellite access layers:

- Flexible payload-enabled service provisioning to semantics-aware [Nto+24] and delay-tolerant IoT applications supported with low-density LEO satellite constellations with periods of invisibility (for both the feeder and service links),
- Direct connectivity to smartphones: A certain level of service at outdoor locations and in light indoor environments, and gradually increasing the bandwidth of the service towards broadband levels, using possibly the Ka band for satellite access to realize a unified RAN for handheld devices ,
- Broadband connectivity to air vehicle/drone mounted devices: A service for ultra-small aperture terminals, which can be installed on air vehicles and drones, supporting air-space safety critical operations [6GNTN24-D35].

The design of the NTN component shall consider the following constraints:

- Spectrum scarcity requiring improved coexistence between NTN and TN
- Sustainability considerations (see Section 8) aiming at minimizing the NTN intrinsic first order effect, enabling NTN to minimize the 6G overall first order effect and identifying how NTN can maximize the 6G overall second order effect (addressing some of the SDGs)

In addition to ensuring support for the highly dynamic topology and configuration variability of the multi-layer access network and an effective management of the intra- (horizontal) and inter-layer (vertical) handovers exacerberated by this dynamicity and variability, several enablers must be provided:

- Unified management and orchestration (MANO) for E2E management of both terrestrial/non-terrestrial infrastructure (for the latter, also using the flexible payload feature), the topology of mobile network implemented in software, and services provided by this network,
- Mechanisms of AI to support E2E network performance optimisation, predictive analytics as well as horizontal and vertical handovers,
- Distribution of the core network between terrestrial and non-terrestrial layers,

- Store and Forward mechanisms to support delay-tolerant IoT applications served by periodically intermittent satellite connectivity,
- E2E control of UPF based on multi-domain SDN,
- Semantics-aware analytics,
- Direct handheld access for UEs served by both terrestrial and non-terrestrial RAN, without the need for external antennas.

4.1.1 DISTRIBUTION OF CORE NETWORK BETWEEN TERRESTRIAL AND NON-TERRESTRIAL LAYERS

Recent IoT advancements have created new markets and use cases, but these services are mainly limited to urban areas with existing cellular coverage, leaving 85% of the Earth's rural and offshore regions unserved. This lack of coverage restricts technological and social progress, limits business opportunities, and hinders the potential of massive IoT applications. Achieving global connectivity is crucial for fully leveraging IoT technologies, and satellite-based non-terrestrial networks (NTNs) are key in addressing this. While current proprietary satellite solutions are not cost-effective, 3GPP has standardized cellular NTNs using terrestrial technologies such as NR, eMTC, and NB-IoT. In Release 17, 3GPP focused on transparent payload architectures with no base stations onboard the satellites, which require constant connectivity to the ground on the feeder link to achieve network access, and thus complex and costly mega-constellations for especially the lower Earth orbits that have very short visibility durations [STK+24]. Low-density LEO satellite constellations, on the other hand, offer a simpler approach, reducing costs and enhancing interoperability. Despite the large time gaps between revisits of low-density LEO satellites, it is possible to have several messages per day, which is sufficient for many delay-tolerant applications such as agriculture, livestock monitoring, assets tracking and maritime. However, this also presents challenges, as using the low-density satellite constellation introduces service link discontinuities. Moreover, the feeder link connecting the low-density satellite constellation to the ground station is accessible only at a limited number of locations. To address this, regenerative payloads hosting partial or full base stations functionalities, part of the core network and using a Store and Forward (S&F) mechanism on one or more satellites is proposed, necessitating adaptations in 3GPP standards [KCC+22]. S&F mechanism is indispensable for both User Plane and Control Plane and will enable the flexible payload to provide NB-IoT coverage to delay-tolerant

applications. Implementing these modifications requires adapting the 3GPP standard procedures. Enabling NB-IoT services from low-density satellite constellations requires sustaining service links in the absence of a constant satellite to ground station connectivity, which indicates discontinuous backhauling. Cost-efficient deployment also relies on standard 3GPP interfaces, allowing multiple service providers to share LEO constellations and extend coverage using roaming agreements. A distributed 3GPP architecture, focusing on regenerative payloads and the S&F principle, is being developed and standardized in 3GPP Release 19 to address these challenges.

The main challenge of low-density LEO constellations lies in their discontinuous service and feeder link, which disrupts the assumption of constant connectivity in mobile networks. Key procedures like Attach/Detach, Tracking Area Update, data transmission, and Paging need modifications due to signalling timers that control mobility and sessions, particularly for NAS procedures. These procedures must be completed within the limited visibility period of satellites. To overcome these issues, a distributed 5G core network architecture compatible with 3GPP standards is introduced. It organizes core components into layers, integrating NTNs as a key element. This architecture involves distributing core NFs as AMF, UPF, and SMF between satellites (5G CN-SAT) and the ground (5G CN-GND), ensuring seamless connectivity and efficient operations despite service interruptions as shown in Figure 4.2.

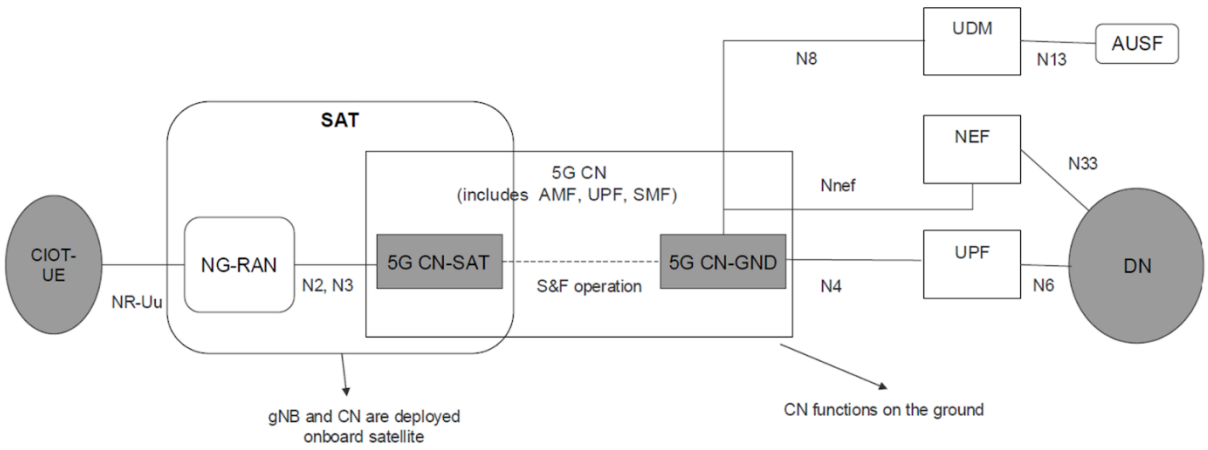


Figure 4.2: Distributed 5G CN architecture between the satellite and the ground satellite operation

4.1.2 UNIFIED MANO

The proposed MANO framework employs a hierarchical model for Management and Orchestration (M&O) of services across ground, aerial, and space domains to tackle the

scalability issues. As shown in Figure 4.3, the M&O in the proposed MANO framework is structured into three hierarchical levels. The E2E level, managed by the E2E Management & Orchestration Component (EMOC), ensures global visibility and control over all domains and their interconnections. EMOC consists of two functional blocks: i) E2E Application Orchestrator (E2EAO), which oversees the Cloud and MEC domains in order to orchestrate MEC Applications. ii) E2E Network Orchestrator (E2ENO), which governs all network orchestration domains and facilitates creation of E2E network slices/services. At the domain level, each of the self-contained domains (including cloud, MEC, RAN, and transport network) is managed by its own Domain Management & Orchestration Component (DMOC) with other domain-specific functions. The infrastructure level comprises Domain Infrastructure (DI) entities that provide an abstracted view of the underlying physical and virtual resources.

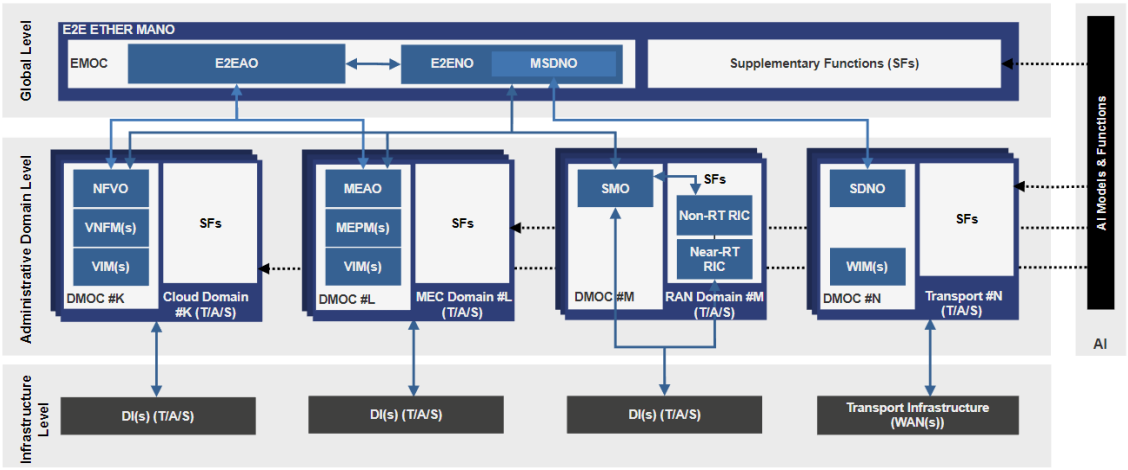


Figure 4.3: High-level view of the proposed MANO with cloud, MEC, RAN and transport domain separation

Furthermore, the Infrastructure Mobility Management (IMM) framework, a part of the Supplementary Functions area, addresses the dynamic management needs of both static and non-static DI resources within 3D networks through a hierarchical structure.

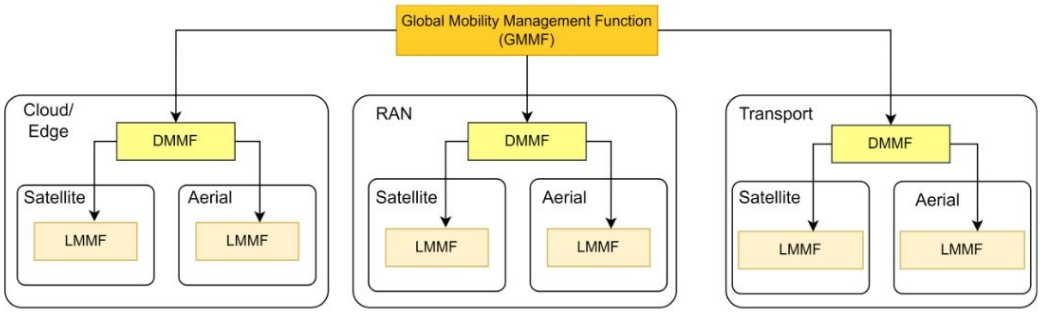


Figure 4.4: Infrastructure mobility management model in a representative case

As illustrated in Figure 4.4, at the top, the Global Mobility Management Function (GMMF) ensures efficient resource allocation and service continuity by registering and discovering available domains as physical infrastructures move between them. The Domain Mobility Management Function (DMMF) then takes over at a more localized level, managing domain-specific mobility and facilitating seamless communication between the local and global mobility management functions. At the bottom, the Local Mobility Management Function (LMMF) tracks the locations and movement patterns of infrastructure components, providing real-time and predictive updates for optimal resource management. This model allows the network to dynamically adapt to changing conditions, ensuring robust service availability and performance across various domains.

The intermittent connectivity challenge and consequently the need for dynamic resource allocation in non-static network infrastructures such as satellites and high-altitude platform systems (HAPS) may be tackled by the Flexible Payload concept. This solution is designed to allow the reconfiguration of onboard hardware, utilizing the Field-Programmable Gate Array (FPGA) technology to adapt to changing service requirements and network conditions dynamically. The Flexible Payload enables non-terrestrial systems (e.g., satellites) to function as NFV Infrastructure (NFVI) nodes, managed by a Virtualized Infrastructure Manager (VIM), enhancing the deployment of virtualized services. This framework supports the virtualization of hardware boards' logical resources and software virtualization on the base operating system, thereby transforming a satellite into a versatile NFVI that can dynamically host and manage various services.

The proposed MANO also addresses the complex challenges of managing and optimizing the trajectories and geographical distribution of non-static network elements, such as satellites, through the integration of a Geographic Information System (GIS). This GIS-based mobility plays a critical role in planning and optimizing the movement paths and spatial distribution of these elements. It also enables NTN operators to effectively simulate communication scenarios and plan orbits, which are integral to managing the mobility of these dynamic infrastructures. The integration of GIS capabilities into the unified MANO framework (Figure 4.5) significantly advances its resource allocation, enhances coverage pattern accuracy, and optimizes network performance. This development effectively streamlines operations and ensures consistent service continuity amid dynamic geographical and environmental changes.

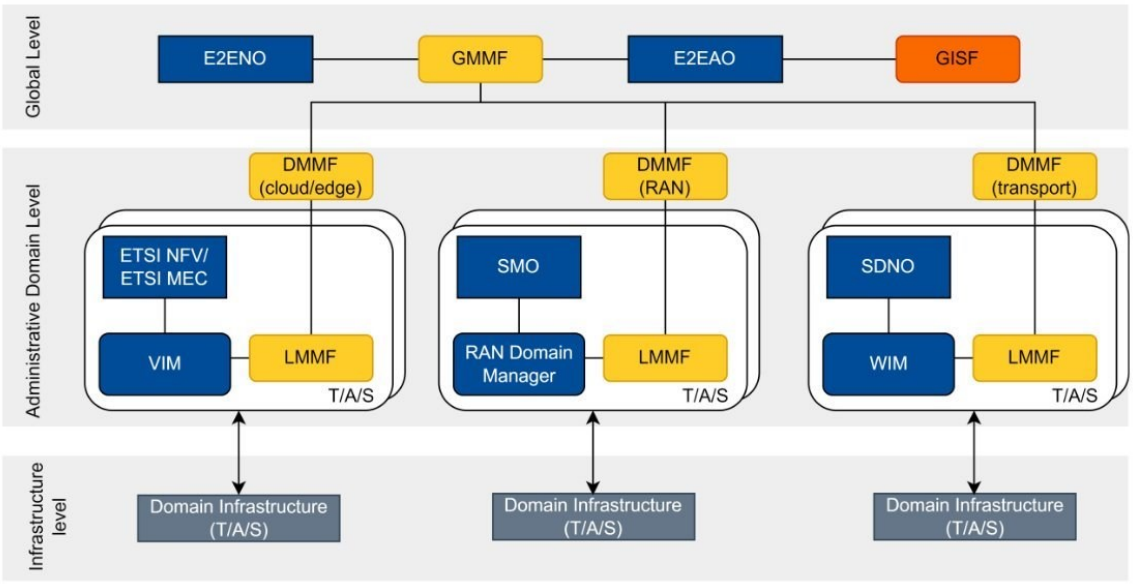


Figure 4.5: Representative case of the proposed MANO architecture that includes mobility and geo-localization management functions

The infrastructure management framework of the proposed MANO follows an approach similar to the 3GPP Management Plane stack. Specifically, thanks to its hierarchical structure, it can directly and natively be interfaced with the stack as depicted in Figure 4.6. In particular, the 3GPP Management Plane functions have interfaces to the corresponding Global, Domain and Local Mobility Management Functions instances to receive the supplementary information used to associate functional instances of the 3GPP framework with their location in the 3D space. Consequently, it is possible to feed the 5G network control and management algorithms with data potentially important for handling dynamic network topology mechanisms, UE handovers, etc.

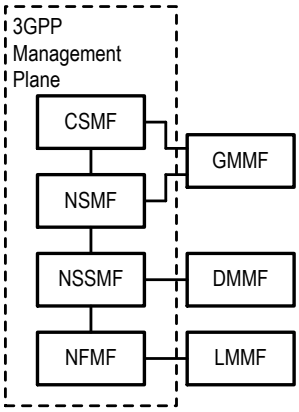


Figure 4.6: 3GPP 5G System Management Plane in the proposed framework and its interconnection with the proposed Infrastructure Mobility Management stack

4.1.3 E2E CONTROL OF UPF BASED ON MULTI-DOMAIN SDN

In the SDN-based UPF, the Control Plane SMF partner behind the N4 interface is the SDN Control Plane in which the Packet Forwarding Control Protocol (PFCP) will be translated to the SDN control protocols, e.g., Open Flow. The SDN Control Plane will be seen by SMF as the control execution enabler of the fundamental packet routing and forwarding functionality of UPF [3GPP24-23501], logically embedded within UPF. In the inherently multi-domain proposed System infrastructure, the SDN-based UPF can be stretched over multiple SDN domains, thus, needing the E2E inter-domain coordination at the Global Level (E2E Network Orchestrator with Master SDN Orchestrator). Then, at the Administrative Domain Level, where the SDN Orchestrator is located, it provides the intra-domain focus and manages the SDN Controller/WIM. It is assumed that UPF can be implemented as a set of dedicated virtual/containerised/physical NFs, i.e., virtualised or physical SDN-enabled Network Elements, SDN Controllers and necessary support functions (in the form of SDN applications) needed to terminate N4 interface and support 3GPP functionalities. It is assumed that the proposed MANO can deploy the UPFs via dynamic orchestration of virtualised components (SDN Controllers, switches) or reuse the already existing ones (e.g., SDN Controllers belonging to the proposed MANO Transport domains and physical SDN switches), ensuring appropriate resource allocation and control privileges over the Control Plane/Data Plane devices. In both cases, the E2E path setup process and session configuration are conducted via the N4 interface following standard 3GPP procedures.

4.1.4 DIRECT HANDHELD ACCESS FOR UE

The 5G NR supports two waveforms, namely Cyclic Prefix OFDM (CP-OFDM) and Discrete Fourier Transform-spread OFDM (DFT-s-OFDM). CP-OFDM is used only for the downlink, whereas either waveform can be selected for the uplinkdepending on the channel conditions in the serving cell, with the gNB instructing the UE to move to the chosen physical uplink shared channel. In the cases of NTNs, both waveforms have been adopted. However, the performance might significantly deteriorate in scenarios with high Doppler shift variations that may arise as the access point of a UE switches from a terrestrial gNB to an NTN node. That is why Orthogonal Time Frequency and Space (OTFS) is proposed as an alternative to OFDM-based waveforms in NTN-related scenarios of high mobility of the NTN platforms, such as LEO satellites. In such a case, through measurements of the Doppler shift by the UE that are reported to the gNB, a threshold-based decision can be taken for whether to switch to OTFS or not (based on

the Doppler shift). Such measurements can be performed under the assumption that the UEs are equipped with a GNSS receiver, so that they know their position, and the satellites transmit their ephemeris data, which contain their position and velocity. In cases, though, where the GNSS signals are weak and the position of the UE cannot be estimated with accuracy, there might be significant residual Doppler effects in the compensation process for OFDM-based waveforms. That is why operating in the delay-Doppler domain, used in OTFS, is advantageous because the channel becomes sparser and varies on a much larger time scale than in the time-frequency domain, but it should be noted that there are several open areas for investigation regarding the introduction of a new waveform, such as OTFS. These include the design of synchronisation algorithms, random access protocols, and reference symbols for the Doppler shift estimation.

Another important feature studied for achieving direct handheld access, namely distributed simultaneous transmission from multiple satellites, is a concept like the standardised coordinated multipoint (CoMP) aspect of LTE-Advanced, which allows joint transmission to a UE from several distributed antennas. Here, the different satellites take the role of distributed flying antenna arrays for which synchronisation in time, frequency, and phase needs to be done.

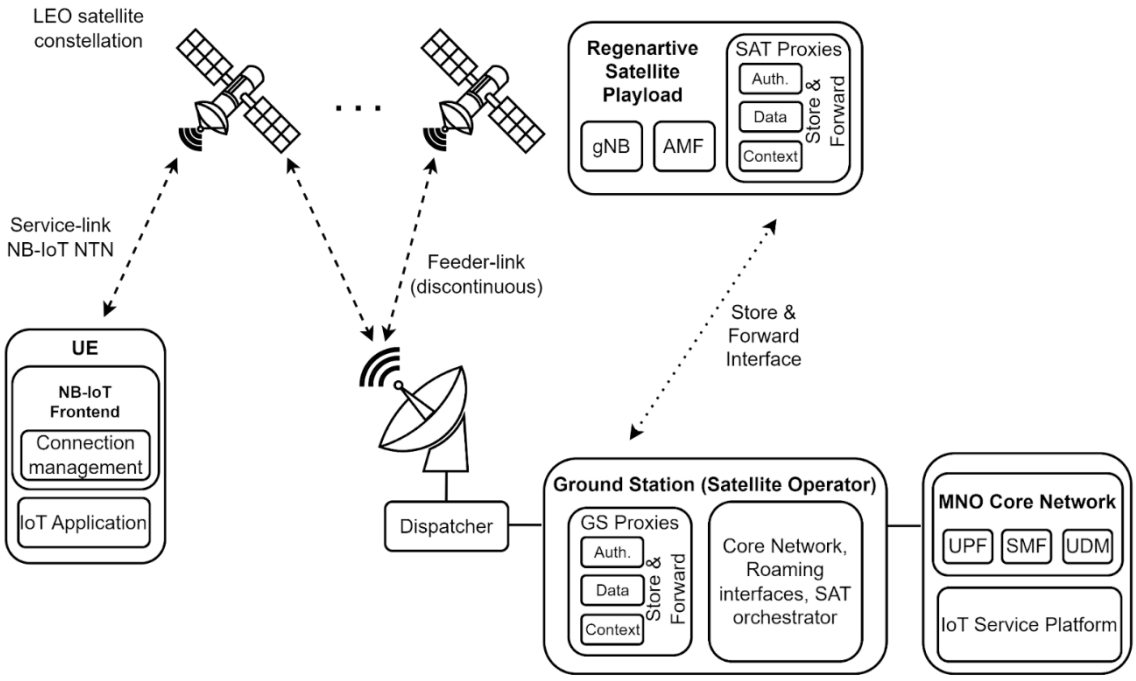


Figure 4.7: Store and Forward mechanism in the distributed 5G CN architecture

The proposed solution for Store and Forward Satellite operations divides the 5G CN functions between satellites (5G CN-SAT) and the ground (5G CN-GND). The satellite (5G CN-SAT) hosts termination endpoints for N2/N3 and NAS protocols, while N8, N4, N6, and Nnef endpoints remain on the ground (5G CN-GND). This architecture also supports multi-satellite scenarios, allowing a single 5G CN-GND instance to interact with multiple satellites. The interface between 5G CN-SAT and 5G CN-GND is flexible, adapting to different constellations and network setups. In one implementation, the core functionality split serves satellites equipped with gNB radio access capabilities and necessary components to complete data and signalling exchanges within the satellite's visibility period. The AMF positioned on the satellites enables completion of NAS procedures within the limited timeframe when the satellite is in contact with the user equipment (UE) as shown in Figure 4.7.

In Store and Forward operation, the core network needs to inform UEs when the satellite operates in the Store and Forward Satellite mode to avoid requests for unsupported services. A satellite cell may switch between Store and Forward and normal modes based on network policies. For example, if the satellite has simultaneous ground network connectivity, it may choose to operate in either mode depending on the situation.

4.1.5 DISTRIBUTED NG-RAN FOR 5G/6G UNIFIED NTN NETWORKS

Network architectures exploiting functional split support implementation of flexible and scalable solutions based on the principles of NF Virtualisation (NFV) and Software Defined Networks (SDN) [ITUR23-M2160]. These allow to tailor the system to the requested use cases and vertical services, and the corresponding Quality of Service (QoS), in addition to an improved end-to-end network management and orchestration. In the context of Non-Terrestrial Networks (NTN), such solutions are particularly promising also to enable the deployment of less complex satellites that carry only the lower layers of the Control Plane (CP) and User Plane (UP) protocol stacks, while leaving the higher layers either on-ground at the gateway side or on-board more complex NTN platforms on the same orbit or higher. The former solution is more oriented to 5G-Advanced NTN, while the latter is a solution that might be feasible for 6G-NTN systems. In the framework of 3GPP New Radio (NR) specifications, only one option is fully enabled, i.e., a split in which the gNB Distributed Unit (gNB-DU) implements up to the Radio Link Control (RLC)/IP layers, where the gNB-DU and the Centralised Unit (CU) are connected via the F1 Air Interface that is persistent, and cannot be closed and re-

established without dropping all of the currently active Packet Data Unit (PDU) sessions serving the User Equipment (UEs) [Nt+24]. This split solution might be particularly challenging in Non-Geosynchronous Orbit (NGSO) NTN implementations due to the payload movement as explained next; in fact, the gNB-DU and the Centralised Unit (CU) are connected via the F1 Air Interface that is persistent, i.e., it cannot be closed and re-established without dropping all of the currently active Packet Data Unit (PDU) sessions serving the User Equipment (UEs), [Nt+24].

With this particular functional split, the gNB-CU is in charge of managing the UE context and requests the gNB-DU to allocate/modify the radio resources for that user. The radio resources are then managed by the gNB-DU based on their availability. As such, the CU and DU are always belonging to the same gNB when the F1 interface connecting them is established. In an NTN scenario, as soon as the NTN node goes beyond the visibility of the serving gateway, the gNB-DU would disconnect from its gNB-CU and, thus, break the gNB, requiring the creation of a new one, which interrupts all of its connections, as shown in Figure 4.8. It is worthwhile highlighting that this is an issue specific to the F1 interface. In fact, handovers at NG interface level (i.e., between the gNB-CU and the 5G Core) are allowed, thanks to NG-flex configurations, i.e., each RAN node is connected to all the Access and Mobility management Functions (AMFs) within an AMF region. In this setting, modifications to the F1 interface procedures or to baseline architectures with one gNB-DU on-board shall be adapted to NGSO scenarios.

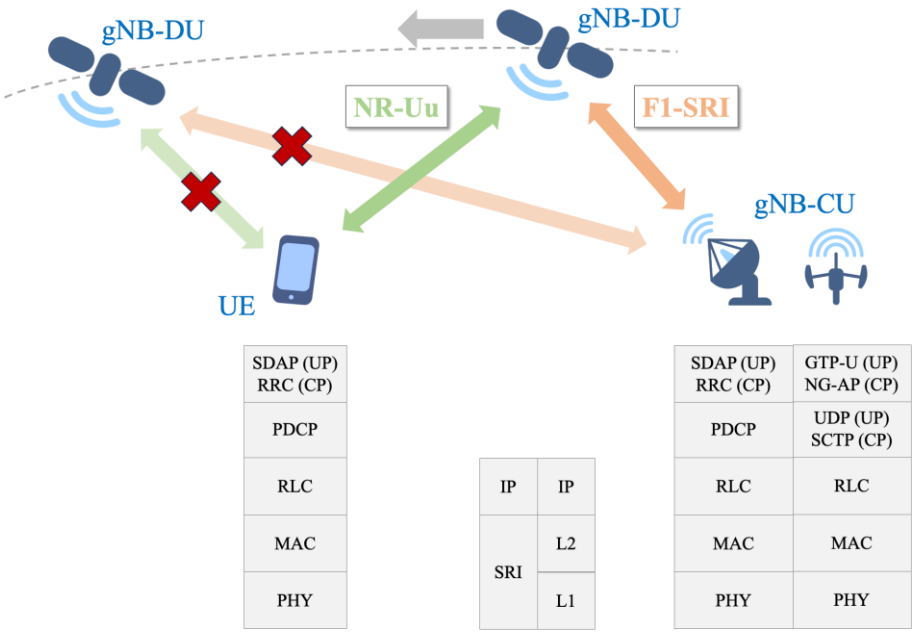


Figure 4.8: F1 persistency in distributed NG-RAN for 5G/6G unified NTN networks

Potential solutions to the F1 persistency problem are based on multiple gNB-CUs or higher layer handover procedures:

- Solution 1a: the single gNB-DU on-board could be connected to multiple on-ground gNB-CUs for resiliency purposes, where as soon as one becomes unavailable due to the feeder link interruption, the second one can carry the F1 interface and its related traffic. However, for this solution to be effective, it is required that the on-board gNB-DU is always in the visibility of at least two gNB-CUs, i.e., two feeder links. This might be challenging due to the very large number of required gateways and traffic overhead on the feeder links. Moreover, the on-ground gNB-CUs shall be connected via Xn. A possible approach to relax the first shortcoming is to also rely on Inter-Satellite Links (ISLs).
- Solution 1b: Again, assuming the availability of multiple feeder links for the NTN node, it might also be possible to implement intra-gNB-CU handover procedures as described in [STK+24]. In this case, a single gNB-CU on-ground shall be connected to multiple gateways and their corresponding feeder links.
- Solution 2: Assuming two separate gNB-CUs on-ground, connected to two different on-board gNB-DUs, it might be possible to implement the inter-gNB handover procedure involving gNB-CU-UP change, described in [3GPP24-38.401]. In this procedure, it is clearly stated that the F1 UE context is modified when sending the handover command to the UE, which also indicates to stop the data transmission for the UE. Once the new gNB-CU-CP is initiated, the radio bearer context modification allows it to retrieve the PDCP status and exchange data forwarding information allowing the completion of the procedure (i.e., to connect the UE to the new DU). Since this procedure involves the interruption of the PDU session and its retrieval, it might be suited for applications that support discontinuous data transmission/reception.

All of the above potential solutions require detailed analyses in terms of architecture feasibility, and the specific 3GPP functions and procedures that will need to be potentially modified.

4.2 MULTI-CONNECTIVITY FOR 6G UBIQUITOUS COVERAGE

4.2.1 6G MULTI-CONNECTIVITY

In 4G and 5G both Dual Connectivity (DC) and Carrier Aggregation (CA) are used as methods to imcrease the amount of spectrum resources in the communication links. The master node can be on a low frequency band, while the secondary node may be a high frequency band cell, with worse coverage. The aim is to achieve efficient usage of the network resources. However, in practice this turns out to be more difficult for especially the DC solution. One drawback with DC (and with EN-DC) is that the master node (i.e., where the connection is terminated) will not have the most recent information about the secondary node performance, since the backhaul (Xn) connection between the nodes may be too slow compared to the time scales of the variations in the radio channel. The communication protocol (i.e., the flow control [3GPP22-38.420], [3GPP24-38.300]) between the master and the secondary node estimates the throughput based on the acknowledgements it receives from the secondary node. In some cases, if for example the secondary node is a cell with high frequency, the coverage may drop quickly and cause long packet delays for the connection over the secondary node. The master node may be unaware of the drastically decreased performance and still send data to the secondary node over the Xn. Another feature of DC is that DL and UL are always coupled and since the secondary connection almost always have worse UL coverage than the master (i.e., the difference may very well be of several dBs, depending on the frequency range), the secondary node feedback may become so bad that this may cause a sharp increase in the round-trip times (or even a timeout), which in turn would result in a decrease in the TCP/IP connection throughput.

6G Multi-connectivity

The multi-connectivity (MC) solution for 6G may be improved and simplified by reducing the number of architecture options to only allowing MC between 6G-enabled base stations and using one type of solution, which should bring the best features from both DC and CA, i.e. a CA/DC evolution. However, it seems natural to base the new 6G MC on the current CA solution in 5G and improve it for 6G. The main reason for using CA as a base is the better UL coverage of CA, where the best UL (i.e., the PCell) can be used for UL response, which means that the UL coverage is often better for CA compared to DC, as the UE does not have to split its limited uplink transmit power

between two concurrent UL connections. The new CA/DC evolution aims to decouple Downlink (DL) and Uplink (UL) (e.g., two DL connections and one UL connection, see Figure 4.9) and employ inherent use of in-active connections. For the in-active connections, the UE only needs to sparsely monitor the control signalling from the network. In addition, the in-active connections should be able to be activated on a short notice. To increase the robustness of the system, there is a need for a more flexible use of the UL so that the SCell may take over the role of control signalling in the UL.

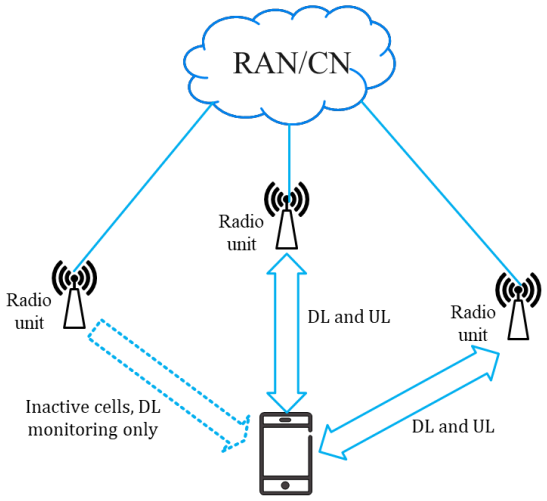


Figure 4.9: Proposed 6G multi-connectivity solutions overview [HEX223-D33]

As part of the CA/DC evolution, faster addition of cells compared to 5G would be beneficial. Furthermore, an enhanced mechanism for PSCell/SCell addition when transitioning from Idle mode to Connected mode could be introduced. With this mechanism, the UE may perform measurements during Idle mode of specific, pre-configured PSCells and not on all frequency layers to avoid higher battery consumption.

Subnetworks

Traditional networks may not be able to efficiently handle the increasing number as well as diversity of devices and applications. Additionally, 6G will introduce requirements for increased coverage, lower power consumption, higher data rates, increased resilience, and increased trustworthiness / user privacy compared to 5G. Subnetworks are formed voluntarily by a set of mutually trusting UEs to aid in achieving these KPIs/KVIs, which are capable of offloading functionalities from one node to another based on the information shared by the nodes to manage the radio resources more efficiently and/or to provide connectivity to devices that are not in network coverage.

Subnetworks architecture and solution

Forming inherently trustworthy subnetworks will also both extend coverage and create a seamless communication system. To achieve the latter, a device may smoothly transition from being served directly by a Base Station (BS) to being served by a Management Node (MgtN) and vice versa. The MgtN is a UE which acts as the subnetwork's primary node, being able to communicate with the BS and other UEs. As an architectural option, the Control Plane (CP) entities of local devices can be flexibly deployed on the MgtN, allowing the subnetwork to use a new lightweight subnetwork CP (snCP) between the MgtN and the UE. In addition to relaying the UEs' UP data to and from the overlay network, the subnetwork and especially the MgtN may assist a UE with multiple CP procedures, such as RRC configuration, mobility, and Idle mode procedures. Such an architecture is illustrated in Figure 4.10, where the snCP is used between the MgtN and the UE. Note that the snCP is transparent to the NW, since it includes configuration and procedures that take place within the subnetwork. The content of the configuration, however, is still managed by the BS.

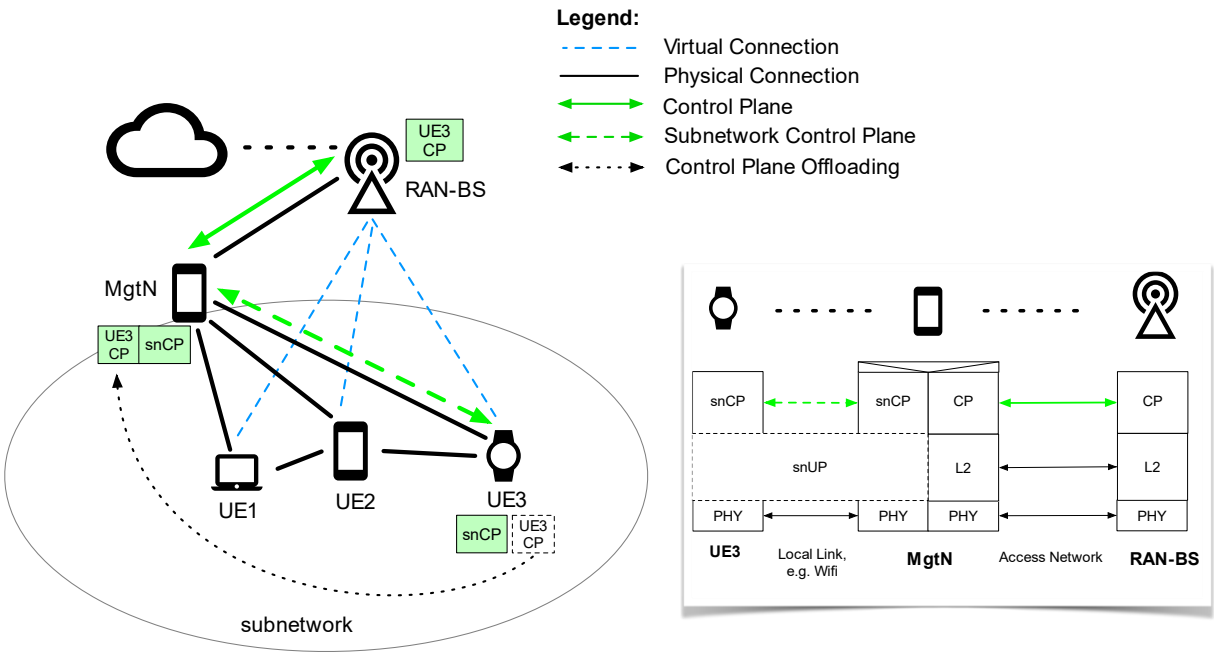


Figure 4.10: UE1 CP deployment at the MgtN and use of snCP within the subnetwork

4.2.2 NTN INTEGRATION FOR 6G MULTI-CONNECTIVITY

The unification of mobile terrestrial network and NTN segments in 6G opens the door to a more effective exploitation of the complete range of network resources available

from both domains, but it also poses formidable challenges in the optimal exploitation of multiple links because of the volatile nature of satellite links and the time-varying network topology, whereby important traffic fluctuations and handovers are dominant. These important network oscillations may severely affect the overall system performance in terms of overall quality of service and experience, and hence a well-defined network architecture able to dynamically react and adapt to such changes is necessary for achieving an effective network unification concept. In this respect, on the one hand the consolidated framework of Access Traffic Steering-Switching-Splitting (ATSSS) represents a viable starting point, but its actual exploitation for scenarios also including 3GPP-native NTN segments is not completely understood or developed. Moreover, the optimization of link splitting, steering, and switching is a complex problem, which demands availability of AI-based solutions to ensure an effective usage of the available network resources. However, the placement of such AI functions in the considered unified NTN-6G network architecture is still an open problem and their interactions with the overall ecosystem needs additional studies to come up with a well-defined and solid architecture.

It is understood that ATSSS was initially conceived to allow native and non-native 3GPP network segments to serve as access networks towards the 5G core network. In this respect, the exploitation of multi-path protocol solutions is considered as particularly appealing because of the intrinsic functionalities of traffic fetching and dispatching to different links by even simultaneously exploiting the resources available from more than one link, and hence naturally achieving load-balancing. In this framework, natural solutions are then represented by Multi-Path TCP (MPTCP) [RFC8684] and Multi-Path QUIC (MPQUIC) [IETF25-QUIC], with the former already is an IETF standard, whereas the latter is still in the approval process that may be completed not earlier than late 2024. Despite this late standardization effort, MPQUIC offers interesting capabilities mostly inherited from the characteristics of QUIC, especially for what concerns the establishment of secure end-to-end data transactions, while MPTPC (similar to TCP alone) must rely on additional protocols or related extensions to provide security features.

Adaptive allocation of traffic flows to the different available network segments necessitates advanced scheduling and fetching mechanisms, which can be suitably coordinated by an end-to-end controller. This opens the door to exploiting AI/ML-based algorithms for optimally selecting the best paths through a dedicated path manager and

making use of the network status (i.e. registered KPI, traffic fluctuations, etc.) recorded through the NWDAF module as input to the actual optimization modules. It is also worth noting that AI functionalities are also attractive for implementing effective traffic prediction algorithms, necessary to forecast possible variations in demands of network resources, and hence accordingly triggering the decision-making process at the path manager for what concerns the adaptation or reinforcement of given traffic splitting and switching policies.

4.3 CONFLUENT TRANSPORT NETWORK

The current vision for 6G includes leveraging new spectral bands, such as high-frequency millimetre-wave and terahertz ranges, to achieve higher peak data rates, and utilizing advanced technologies like ultra-large antenna arrays and cell-free (CF) architectures—enabled by CF-mMIMO—to enhance spectral and energy efficiency while supporting a higher number of simultaneous connections. Sensing at various network segments is also considered an inherent capability to be developed in future 6G networks. At the same time, focus is put on enhancing network orchestration capabilities by adding intelligence at network control and management layers, aiming at optimizing performance and sustainability.

A proposed approach [ECO-eNET] is to integrate advanced technologies at the transport network layer with high-capacity CF-mMIMO structures, advanced Open-Radio Access Network (O-RAN) access nodes, and an intelligent control and network management plane. Sensing data across the network infrastructure will be fused to the management plane to enable confluent-mesh network structures delivering high availability and performance 6G network deployments. These solutions can pave the way towards new service opportunities and features, transforming the 6G ecosystem and advancing its alignment with the IMT-2030 goals and future user demands [ITUR23-M2160].

Developing high-capacity wireless optical and radio fixed wireless access (FWA) technologies and their seamless integration with wired packet optical networks, referred to as "confluence", using optical-spectrum-as-a-service (OSaaS) is the approach to introduce confluent fixed and wireless optical and radio fixed wireless access technologies for 6G front-/mid-/back-haul (xhaul) networks to form cell-free mesh physical layer edge networks as illustrated in Figure 4.11 [Raj+24].

Confluent transmission makes use of a combination of radio frequency wave (RFW, at THz and sub-THz frequencies), free space optical (FSO), switched flex grid wavelength division multiplexed (Flex-WDM) and OSaaS fibre transmission capabilities to form mesh networks offering a flexible management of high-capacity traffic with low latency and high energy-efficiency. In particular, RFW and FSO links enable the formation of mesh networks at the edge, where deployment costs and complexity prohibit wired mesh networks. These wireless links offer the efficiency and latency benefits of mesh data transport, while they can also be used to transmit control plane signals to manage both the wireless and wireline networks. This provides a new degree of freedom in the network control that is exploited to facilitate wireline switching and low latency. Analog radio-over-fibre (aRoF) transmitted over the wired network will efficiently be converted to RFW signals using novel plasmonic devices. Analog signals will be multiplexed with digital signals throughout the confluent xhaul network using a combination of optical and electronic switching to enable highly dynamic power and spectrum management. The efficiency of using such confluent xhaul network with high-density cell-free radio access networks using coordinated multipoint and distributed multiple input multiple output (d-MIMO) techniques needs to be investigated for delivering high data capacity over a wide range of spectral bands, including line-of-sight (LoS) communications subject to severe blocking and fading challenges.

The potential of technologies such as RFW (at THz and sub-THz frequencies), FSO, and Flex-WDM fibre optics to deliver required performance by 6G need to be examined. Electronic RFW transceivers will be replaced by plasmonic-based ones to extend the reach of the RFW link and significantly reduce the energy consumption of the network. Moreover, creating a transparent fibre-FSO interface supporting multiple modulation formats would be desirable. This could be possible by novel, low-cost, adaptive photonic components and lantern technology. Optically switched and modulation format adaptive OSaaS can be used to efficiently multiplex and transport the signals from these diverse transmission links in mesh configurations and connect them with edge computing resources. OSaaS will also enable fibre and radio sensing signals to be carried alongside the communication signals.

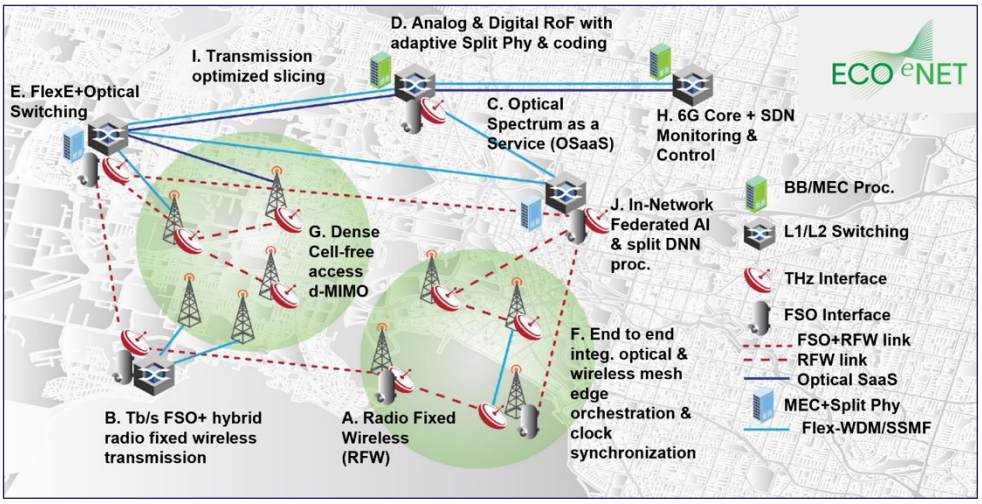


Figure 4.11: Confluent mesh networking

The performance of the physical layer technologies and their potential to deliver improved capacity, energy efficiency, and latency depends strongly on their control and management. This, however, imposes several challenges as radio and transport networks are currently separately engineered and controlled, introducing the need for 5G core, transport and radio access network control to evolve towards increased coordination and compatibility. AI-assisted network control can facilitate optimised control and management capabilities across wireless and wireline systems. Therefore, building tools that can exploit the confluent networking capabilities within this evolving and increasingly integrated control environment is crucial.

A flexible and scalable control framework with extendable Application Programming Interfaces (API) on top of the B5G/6G RAN, the Core, and the transport network, can facilitate monitoring and programmability of the underlying network infrastructure. This will enable an end-to-end platform that can support customised service delivery in response to the service requirements in the most resource and energy-efficient manner. The monitoring data collection ensures appropriate system initialization and allows continuous optimisation of the entire system operation. Such a monitoring system can subscribe and collect both high-level (E2E service related such as throughput, packet latency, jitter, etc.) and low-level statistics (network/compute resource utilisation, bit error rate, packet error rate, power consumption, physical layer characteristics of the RFW links and optical switching nodes). These statistics can be exposed to other NFs, such as the Network Data Analytics Function (NWDAF), to provide recommendation services. Such system capabilities enable lower layer decisions such as optimal mapping of the wireless domain service characteristics to the optical transport

parameter configuration. An integrated control plane development, supporting the seamless cross-domain service delivery through the wireless and the transport network domains, will be beneficial.

4.4 REFERENCES

[3GPP22-38.420] 3GPP TS 38.420: "NG-RAN; Xn general aspects and principles", V17.2.0, September 2022

[3GPP24-23.501] 3GPP, "System architecture for the 5G System (5GS)", 3rd Generation Partnership Project, Technical Specification TS 23.501, ver. 19.1.0, September 2024. [Online]. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144.

[3GPP24-38.300] 3GPP TS 38.300 V18.4.0: "NR and NG-RAN Overall Description", V18.4.0, December 2024

[3GPP24-38.401] 3GPP TS 38.300 V18.4.0: "NG-RAN; Architecture description," July 2024.

[5GSTAR24-D52] 5G-STARDUST Deliverable D5.2, "Preliminary Report on Multi-Connectivity and Software Defined Network Control". Online. https://www.5g-stardust.eu/wp-content/uploads/sites/97/2024/08/5G-STARDUST_D5.2_1.0.F.pdf June 2024.

[6GNTN24-D35] 6G-NTN Deliverable D3.5, "Report on 3D multi layered NTN architecture (2nd version)," March 2024.

[BGG+23] Bahare, M. K., Gavras, A., Gramaglia, M., Cosmas, J., Li, X., Bulakci, Ö., Rahman, A., Kostopoulos, A., Mesodiakaki, A., Tsolkas, D., Ericson, M., Boldi, M., Uusitalo, M., Ghoraishi, M., & Rugeland, P. (2023). The 6G Architecture Landscape - European perspective. Zenodo. https://doi.org/10.5281/zenodo.7313232

[BLG+23] Ömer Bulakçı (ed.), Xi Li (ed.), Marco Gramaglia (ed.), Anastasius Gavras (ed.), Mikko Uusitalo (ed.), Patrik Rugeland (ed.), Mauro Boldi (ed.) (2023), "Towards Sustainable and Trustworthy 6G: Challenges, Enablers, and Architectural Design", Boston-Delft: now publishers, http://dx.doi.org/10.1561/9781638282396

[ECO-eNET] Project ECO-eNET: Developing an intelligent network of technology enablers to seamlessly connect the human, physical, and digital worlds. Online. https://www.eco-enet.eu/

[HEX223-D33] HEXA-X-II Deliverable D3.3, "Initial analysis of architectural enablers and framework" https://hexa-x-ii.eu/wp-content/uploads/2024/04/Hexa-X-II_D3.3_v1.0.pdf, April 30, 2024

[IETF25-QUIC] "Multipath Extension for QUIC", draft-ietf-quic-multipath-10, IETF, Expiration date: 9 January 2025

[ITUR23-M2160] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond", M.2160, November 2023. [Online]. Available: https://www.itu.int/rec/R-REC-M.2160/en.

[KCC+22] T. Kellermann, R. P. Centelles, D. Camps-Mur, R. Ferrús, M. Guadalupi and A. C. Augé, "Novel Architecture for Cellular IoT in Future Non-Terrestrial Networks: Store and Forward Adaptations for Enabling Discontinuous Feeder Link Operation," in IEEE Access, vol. 10, pp. 68922-68936, 2022, doi: 10.1109/ACCESS.2022.3184720.

[Nto+24] K. Ntontin et al., "ETHER: A 6G Architectural Framework for 3D Multi-Layered Networks," 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, 2024, pp. 1-6, doi: 10.1109/WCNC57260.2024.10570731

[Raj+24] R. Raj et al., "Towards Efficient Confluent Edge Networks," 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024, pp. 1163-1168, doi: 10.1109/EuCNC/6GSummit60053.2024.10597093.

[RFC8684] RFC 8684, "TCP Extensions for Multipath Operation with Multiple Addresses", IETF, March 2020

[STK+24] M. M. Saad, M. A. Tariq, M. T. R. Khan and D. Kim, "Non-Terrestrial Networks: An Overview of 3GPP Release 17 & 18," in IEEE Internet of Things Magazine, vol. 7, no. 1, pp. 20-26, January 2024, doi: 10.1109/IOTM.001.2300154.

5 ARTIFICIAL INTELLIGENCE AND COMMUNICATIONS

The 6G architecture must inherently support AI as a Service (AIaaS) for both internal network operations and external applications, thereby enhancing automation and distributed intelligence. A critical application of AI in this context is network automation. The primary challenges to AI adoption for network automation involve the sustainability of the machine learning (ML) training process, particularly in terms of energy consumption and environmental impact, as well as the perceived trustworthiness of AI decisions. To address these challenges, ML training and inference should employ techniques and architectures that optimize computing resource utilization and power consumption. This includes selecting the optimal placement of AI/ML functions based on data location, utilizing computing nodes powered by renewable energy, adopting collaborative learning techniques to distribute the training process, and properly scheduling training and re-training activities. Regarding trustworthiness, explainable AI techniques should be employed to elucidate the rationale behind predictions and automation actions, mitigating potential biases and possibly retaining human oversight. Furthermore, ensuring robustness against security attacks and maintaining the privacy and confidentiality of training data in distributed AI systems are imperative. Federated learning could address these privacy concerns by sharing only the trained models rather than the raw data. Consequently, 6G architecture should emphasize ubiquitous intelligence, sustainability, and security/resilience in AI services.

5.1 AI/ML FRAMEWORK / INTELLIGENCE PLANE

This section discusses several options on how to introduce the AI/ML framework or the intelligence/data plane.

Section 5.1.1 presents several options on how to introduce the AI/ML framework to the 6G architecture. It also outlines the necessary functions or enablers for this and a possible framework. In section 5.1.2 the intelligence plane is introduced, using O-RAN architecture as a basis. Cloud nativeness is an important aspect of 6G, and how to improve the user plane to be more cloud-friendly is handled in section 5.1.3. Finally, section 5.1.4 deals with the RAN aspects for using AI to improve waveforms, transceivers and protocols.

5.1.1 AI/ML ENABLERS AND FRAMEWORK

The AI enablers within the 6G data-driven architecture encompass architectural elements and protocols, Machine Learning Operations (MLOps), Data Operations (DataOps), AI as a Service (AIaaS), and intent-based management [HEX223-D33]. These AI enablers constitute a robust framework that seamlessly integrates AI into the compute continuum of 6G networks, facilitating advanced automation and distributed intelligence.

MLOps is focused on operationalizing machine learning models by ensuring their smooth deployment, version control, and continuous monitoring within the overarching architecture. The architecture requirements for MLOps include access to high-quality data, scalable data storage solutions, computational resources for data processing and model training, and stringent security and trust measures.

DataOps enables efficient data collection, integration, and management, providing MLOps with timely and high-quality data. Essential architectural requirements for DataOps include robust data quality management functionalities, end-to-end data pipelines that effectively serve MLOps, and version control mechanisms for the collected data (refer to Figure 5.1).

Building upon elements such as MLOps, DataOps, the AIaaS framework delivers AI services across various network segments and to end-users. This framework necessitates the development of new APIs for both internal network exposure and external end-user access. Furthermore, AIaaS demands rigorous security measures and regulatory compliance, along with feedback loops for continuous improvement and resource optimization.

In the context of 6G networks, the integration of AIaaS is pivotal for enhancing network performance, reliability, and intelligence. It supports a wide array of applications, from real-time analytics to predictive maintenance, thereby driving the evolution of next-generation network services. By leveraging advanced AI/ML technologies and methodologies, the 6G architecture aims to achieve unprecedented levels of automation, efficiency, and user experience, establishing a foundation for future innovations in telecommunications.

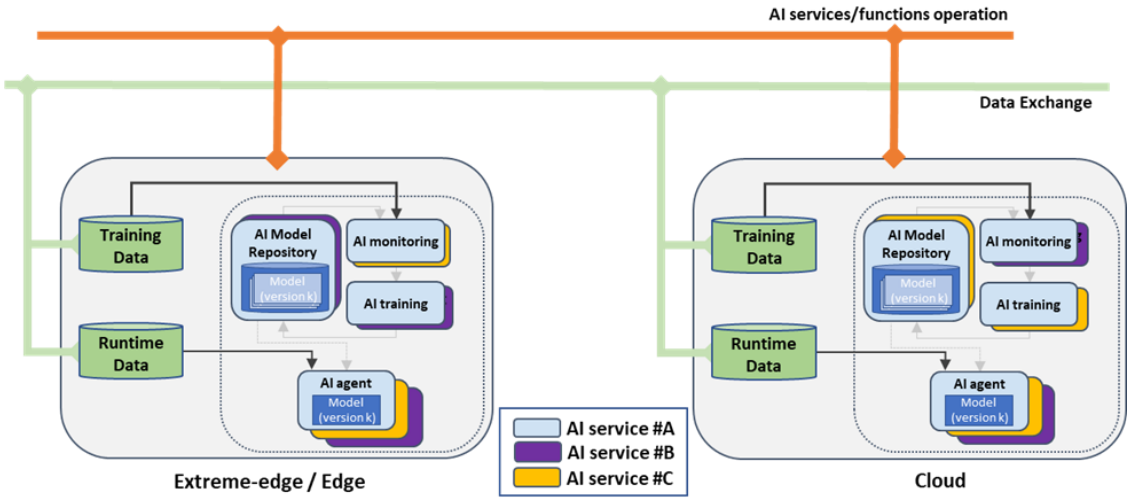


Figure 5.1: AI/ML framework [HEX224-D33]

5.1.2 INTELLIGENCE PLANE IN O-RAN

Intelligence Plane, depicted in Figure 5.2, which works as a cross-domain management entity, integrating control and monitoring functions across RAN, Core and Edge domains and fostering the creation of advanced ML models [BEG23-D41]. The proposed Intelligence Plane incorporates an AI Engine, which provides a serverless execution environment hosting the AI/ML models, offering inference and training services to the rApps/xApps implementing the control loops by following a loosely coupled approach.

The AI Engine hosts the ML models to offload inference tasks from the RICs and implement the necessary AI/ML workflows and services. As shown in Figure 5.2, the AI Engine manages the AI/ML pipelines, including model management, monitoring, training, serving, and a data lake with prepared data. The models are served in a serverless way, which enables efficient scaling of workloads in production. In the case of O-RAN, the inference of the models is exposed to the control rApps/xApps though AIA1 and AIA2 interfaces plus associated AI Engine Assist rApps/xApps, which allows to decouple the implementation of control-loops from the management of ML models.

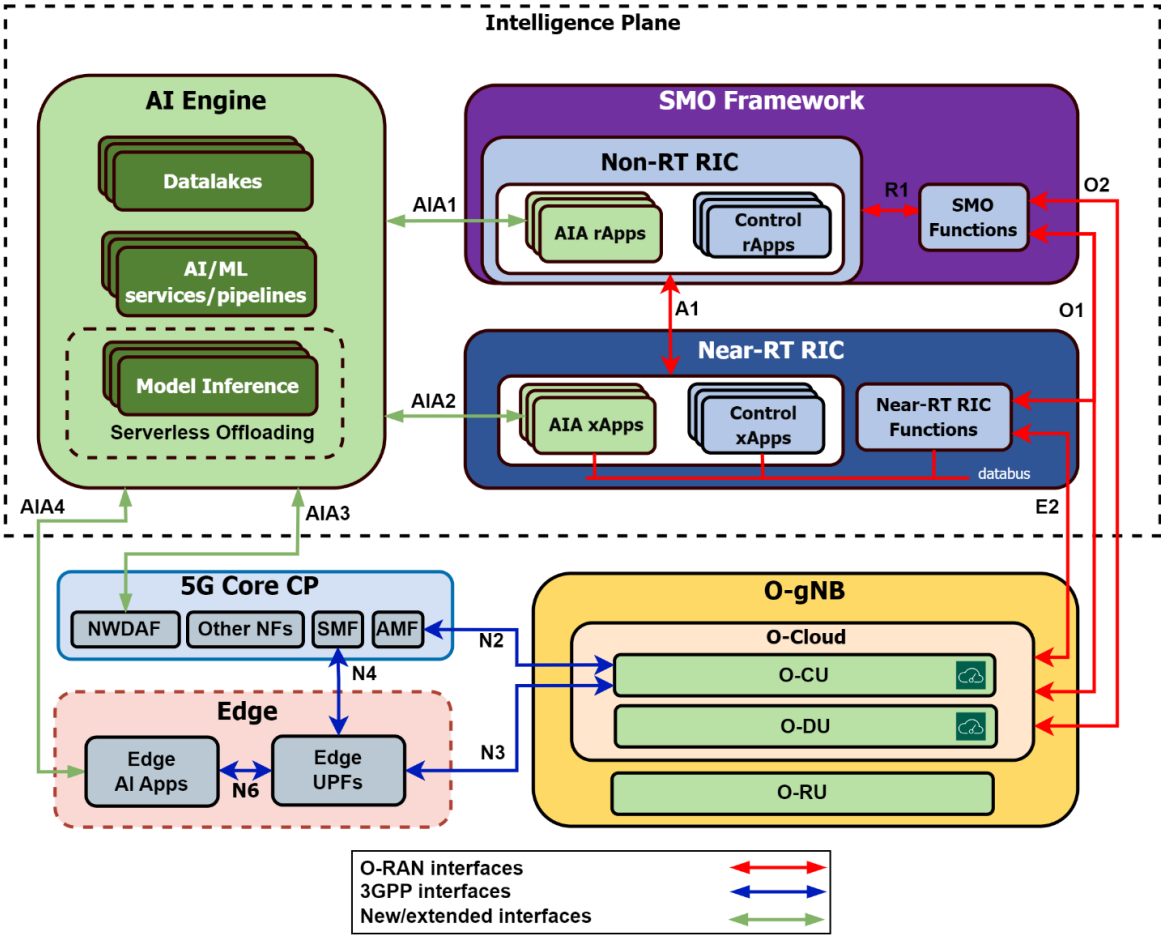


Figure 5.2: Intelligence Plane

This design facilitates model reusability by different control rApps/xApps and allows the integration of AI/ML workflows through the AI Engine independent of the RICs implementation. A similar approach could be adapted for Edge and Core domains, using AIA3 and AIA4 interfaces to expose AI Engine AI/ML services to Edge applications or NWDAF analytics. In [BEG23-D41], this approach is followed to develop energy efficiency optimizations, including the dynamic management of Edge and O-Cloud compute resources, of relay nodes and of Reconfigurable Intelligent Surfaces (RIS) via extended O-RAN interfaces.

5.1.3 AI AIR INTERFACE

AI will be crucial in meeting the technical and societal needs of 6G communication systems, enabling energy-efficient, user-centric communications. By leveraging advanced AI techniques—such as reinforcement learning, transfer learning, and semantic communications— waveforms, transceivers and protocols can be customized

for diverse scenarios, devices, and users. In contrast to 4G and 5G, 6G will benefit from vast datasets and over a decade of machine learning progress [AHM24].

For example, future use cases like mission-critical video streaming in factories with 360° 4K cameras demand extreme bandwidth and low latency, requiring customized physical layer (PHY) and protocol designs. However, today's rigid architecture makes such customization too expensive. AI can optimize waveforms, MIMO processing, and networking protocols for specific devices and environments, offering performance and energy efficiency. This AI-driven approach allows flexible, cost-effective communication systems, addressing challenges like rural coverage and reducing reliance on rigid network architectures.

The Artificial Intelligence native Air Interface (AI-AI) concept places the users' communication needs and application-specific requirements at the centre of the design as depicted in Figure 5.3. Then, tailor-made waveforms, transceivers, signalling, protocols, and hardware implementations are optimized adaptively and on-demand within a modular architecture to support these requirements.

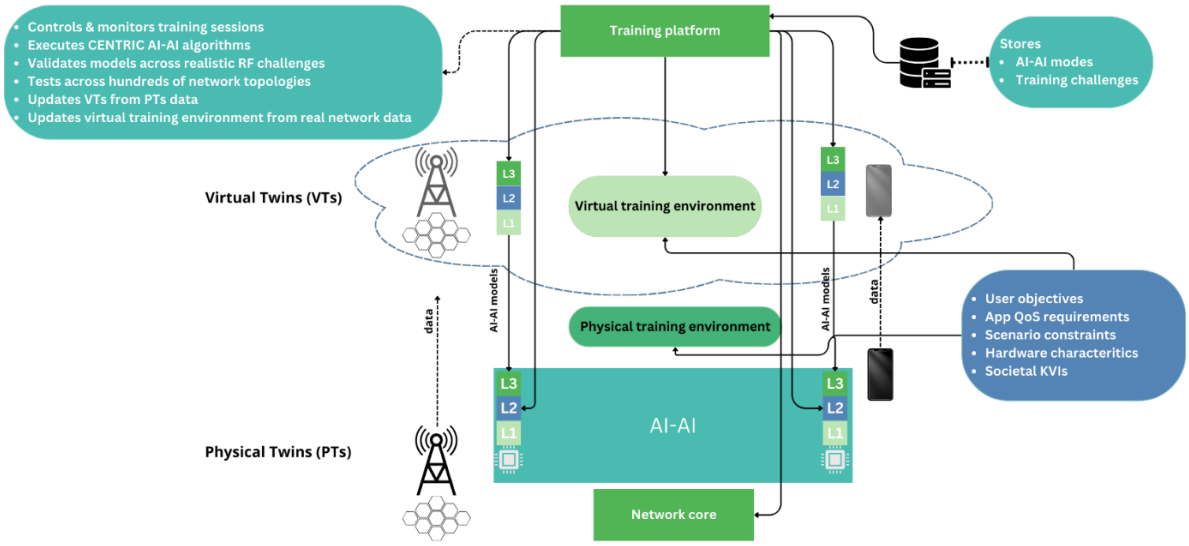


Figure 5.3: Artificial Intelligence native Air-Interface (AI-AI) architecture component

As shown in the AI-AI architecture in Figure 5.3, novel AI-solutions for the physical, MAC and RRM layers must be developed to drive the concept. An important example for such solutions is briefly described below.

Neural network-based receiver (NRX) with end-to-end learning [CENTR24-D21]: An AI-native receiver will constitute a major L1 component of the AI-AI concept. The neural

network-based receiver depicted in Figure 5.4 replaces traditional receiver processing blocks including channel estimation, equalization and demapping in 5G with a single neural network, which is trained in an end-to-end version. The NRX is compatible with 5G NR Physical Uplink Shared CHannel (PUSCH) and can support pilotless communication and custom constellation.

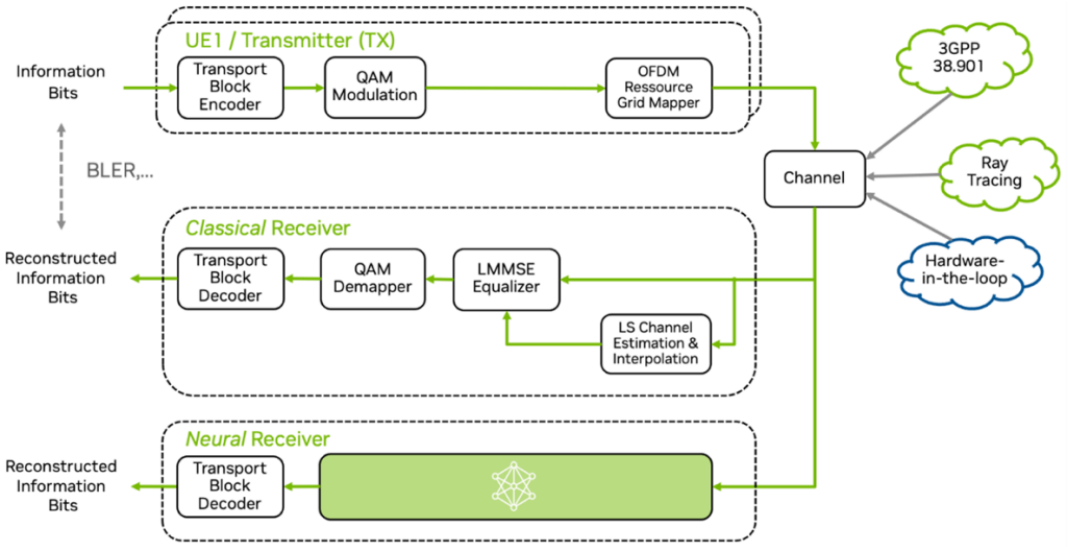


Figure 5.4: Neural network-based receiver

5.1.4 CLOUD NATIVE DATA PLANE

The cloud native principle has widely been accepted by industry and applied to deploy traditional computing applications, enabling better management and utilization of computing resources. Environments like a Kubernetes cluster uses proxy load balancers to ensure seamless load distribution among instances and automatically handles on-demand up and down scaling. This approach supports high-level of flexibility in using computational resources and allows application developers to focus on the business logic's implementation while scalability, performance and deployment questions are handled by the environment. Though cloud native approaches are applied in 5G systems to deploy control plane entities like AMF and SMF, it has not been extended to the user plane yet, where much stricter performance requirements of packet processing logic need to be satisfied. Though the cloud native approach would also be beneficial for user/data plane applications, additional challenges need to be solved to handle heterogeneous programmable targets to be used for packet processing (e.g., programmable switches, CPUs, DPUs, SmartNICs, IPUs, FPGA-based NICs). These challenges include: (i) hiding the implementation and deployment details of the underlying NF data planes, (ii) providing seamless and dynamic offloading and

optimization in the data plane, i.e., hardware/software target selection, disaggregation, (iii) providing tools to better use the resources of hardware data planes with NF isolation, e.g., by enabling the deployment of multiple NF data planes on the same hardware, mimicking data plane hardware virtualization. Current programmable data plane targets do not support multi-tenant usage and virtualization.

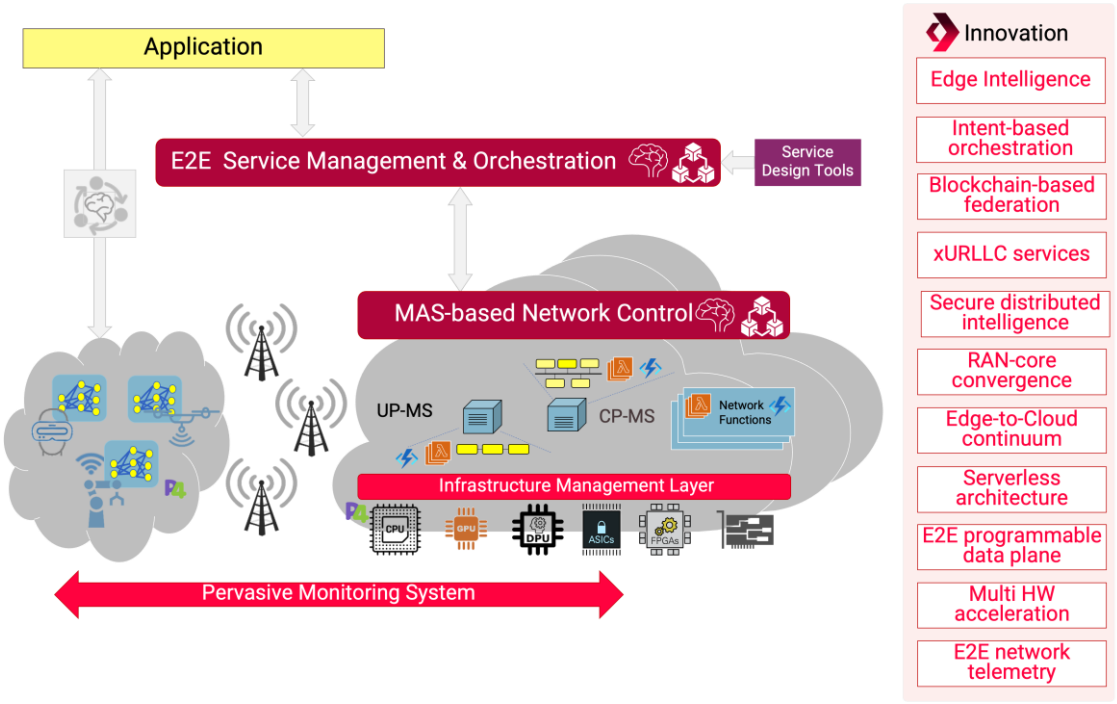


Figure 5.5: High-Level Architecture and Innovations for an Infrastructure Management Layer [DESIRE24-D22]

A new architectural component called infrastructure management layer (IML) was proposed in [DESIRE24-D22] to separate concerns of the packet processing business logic and the infrastructure layer (Figure 5.5). IML basically acts as a combination of a Virtualized Infrastructure Manager (VIM) and a hardware abstraction (HAL) layer. IML is responsible for managing a pool of resources (e.g., located on a 6G site).

In 6G network services, packet traffic needs to be forwarded through different NFs (NFs). Each NF has a control plane and a data plane. IML focuses on the deployment and run-time management of data plane components. An NF data plane component implements the packet processing logic and can be executed on various targets including smartNICs, ASICs, FPGAs, IPUs and DPUs, in addition to traditional CPU resources, IML is responsible for selecting the appropriate target(s) and number of instances to execute the NF data plane and configure the virtual links between them at deployment time. Virtual links are created by infrastructure NFs implementing traffic

forwarding and routing between NFs. To enable run-time optimization and hide the underlying optimization from the NF control plane, IML introduces a control plane proxy using a common northbound API (e.g., P4Runtime [P4R24]) that provides a single-instance view of the data plane component to the NF control plane. The proxy hides the underlying data plane optimization like load balancing between multiple data plane instances of the same NF data plane or offloading heavy hitter users to hardware data planes. To enable the better utilization of data plane hardware resources, IML has a subcomponent called P4-MTAGG [BKL+24a], [BKL+24b] that is a compiler-based virtualization tool for P4 [BDG+14] programmable hardware targets. It enables the deployment and execution of multiple P4 programs on the same P4 hardware in an isolated way. The control plane access to the different data plane programs is also isolated by the IML's control plane proxy component.

5.2 INTENT-/GOALS-DRIVEN COMMUNICATIONS

This section deals with how higher-level languages (such as express intents for network configuration or semantic communication) can be used to control the network. Section 5.2.1 describes intents and the cognitive component, while Section 5.2.2 focuses on information exchange done by using semantic context of AI. Section 5.2.3 discusses an architecture to enable high security and privacy in future 6G networks using intent-based interfaces.

5.2.1 TRUST INTENT-DRIVEN COGNITIVE 6G NETWORK

The work in [GKF+24], [ARF+24] introduce the Cognitive Coordination, a component that is an intent-handling function that comprehends sophisticated and abstract trust intent semantics (divided into the five trustworthiness taxonomies of Safety, Security, Privacy, Resilience, and Reliability), calculates the ideal goal state, and organizes activities to transition the SAFE-6G system into this trustworthy state. The function will be able to research possibilities about the applicability of the five functions in providing the desired degree of trust, learn from precedents, and assess the feasibility of actions based on their expected results.

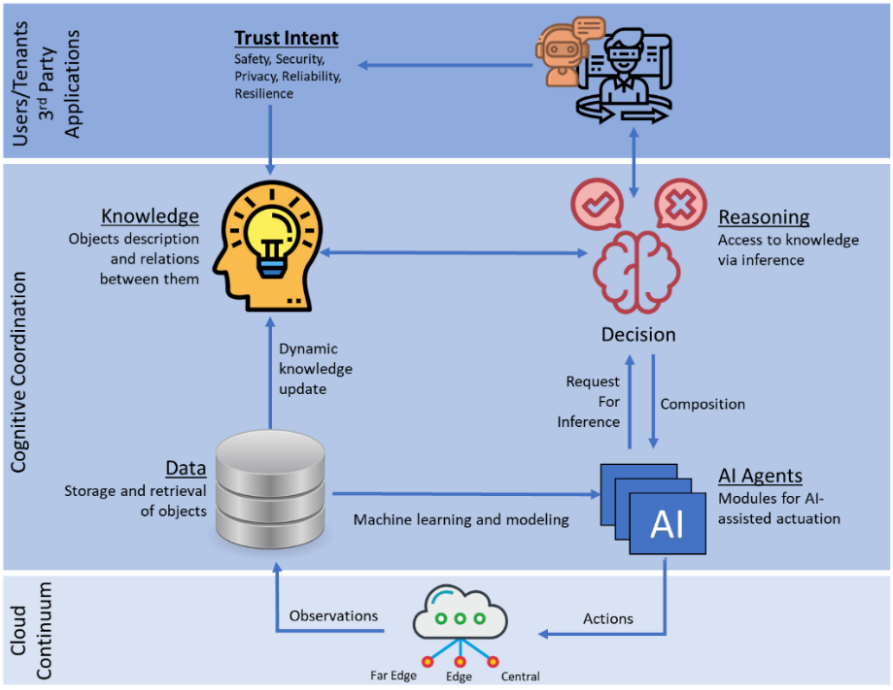


Figure 5.6: SAFE-6G Functional architecture of the cognitive layer

Cognitive Coordination [SAFE6G] is carried out as an autonomous service and network operation by combining well-known AI technologies inside a flexible framework. The cognitive layer, as shown in Figure 5.6, acts as an interface between tenants/users/3rd party apps and the network/environment via the 6G service exposure provider interface. The 6G network architecture in the SAFE-6G project has end-to-end Machine Learning (ML) and model access, encompassing autonomous networking by taking advantage of AI/ML capabilities such as supervised and unsupervised machine learning algorithms as well as reinforcement learning.

AI/ML techniques will help to manage the edge-cloud continuum, dealing with the heterogeneity of data sources and the high number of tenants, correlating data from far-edge, edge, and Core Network towards providing the requested level of trust. Different ML algorithms (e.g., supervised, unsupervised, federated or reinforcement learning) will be considered to efficiently improve the coordination of resource and service orchestration as well as for resource monitoring by implementing orchestration actions (e.g., closed-loop automation, proactive alerting, automated healing, and/or predictive NF scaling/placement, etc). AI will ensure a proficient, optimal, and continuous end-to-end orchestration, which needs to span over different domains, while having a coherent view of its own scope and purpose. In such a context, in SAFE-6G a complex and distributed system will be coordinated by designing cognitive and

effective architectural interfaces to manage and synchronize both operations and communication.

The cognitive coordination layer consists of three major components: a knowledge base, a reasoning engine, and an agent architecture. The knowledge base includes an ontology of trust intents as well as domain-specific knowledge such as the current state of the system. The domain-independent reasoning engine will use the knowledge graph as the primary coordinator function for locating actions, assessing their impact, and ordering their execution in order to provide the requested level of trust to the requested tenant/user and/or third-party application. Finally, the agent design allows for the use of an unlimited number of models and services.

5.2.2 GOAL-ORIENTED AND SEMANTIC COMMUNICATION IN 6G AI-NATIVE NETWORKS

AI's rapid rise impacts communication infrastructure by demanding vast computational resources and data, primarily generated by edge wireless sensors, for training large models. This intensifies network pressure. Furthermore, advanced AI increases device intelligence, shifting communication to information exchange within the semantic context of AI.

With the surge in connected autonomous vehicles, smart wearables, robots, and AR/VR equipment, machine-type communications will dominate networks over the current human-oriented traffic, requiring completely different service KPIs. For instance, video signals for machines focus on task-relevant information, unlike those for human consumption, which must meet latency and quality requirements. This shift redefines communication networks from reliable bit transmission infrastructures to networks of intelligence, blurring the lines between communication, computation, and intelligence, where a goal-oriented design approach should be pursued [Str+24].

[Str+24] outlines contributions to the emergence of this new paradigm with a novel goal-oriented communication architecture, network components and algorithms that make communication converge with computation in a jointly data- and model-driven manner, such as the following:

Semantic Engine will be positioned in the service management and orchestration unit of a network, which is responsible for the efficient and effective delivery of semantic-oriented services, through the orchestration of semantic information resources processing, semantic model’s lifecycle and user experience management.

Semantic Radio Intelligent Controller (S-RIC) will be linked with the control and user planes of O-CU for resource, connection and quality of service management. S-RIC and O-DU connectivity will also be supported for the distributed semantic information processing ability. A rich set of additional interfaces will be defined within the O-RAN architecture for the above connectivity. S-RIC will be designed as a programmable and extensible unit to facilitate the deployment of diverse semantic applications. It will also support different service time requirements, ranging from non-real-time to near-real-time and real-time, for these applications.

Application Plane is orthogonal to the semantic plane in the diagram, which provides the interfaces for the semantic applications across edge devices or user equipment. Meanwhile, this plane will interact with the core network, near-real-time RIC, and newly defined semantic interfaces in the O-RAN for holistic task scheduling and optimised resource allocation in a secured way.

Semantic-powered UE and Edge, which requires devices to be equipped with computational and learning capabilities for the initial semantic information extraction from the raw image, video and sensor data. Meanwhile, an intelligent O-RU will be envisioned for a better real-time semantic processing (including extraction and interpretation) capability.

Knowledge database, whose functionality is implicitly covered in this structure because it is present in almost every semantic processing module. The training and validation of semantic models both require support from a knowledge base, and a unified and consistent knowledge base is crucial for the successful extraction and translation of semantic information by the models.

In a nutshell, integration of semantic and goal-oriented principles into the AI/ML architecture, through the new network components described, optimizes control signalling necessary in the protocols and thus markedly reduces the communication overhead without sacrificing correct and timely operation of the protocols. Examples include innovative L1/L2 protocols for massive access with targeted semantic content selection. To this aim, the novel concept of a semantic RAN intelligent controller (S-RIC) is very critical. The O-RAN architecture introduces two separate RAN Intelligence Controllers (RICs) as ML-based functional blocks operating in a closed-loop scheme at either non-real time (e.g., seconds) or near-real time (e.g., tens of milliseconds) time granularity. However, the interaction among states and messages at different time scales provides an opportunity for bringing in the semantic representations and enable

effective interaction among various control loops. This has a clear potential to progress beyond the state of the art by introducing a novel real-time semantic control layer, applicable to O-RAN. Specifically, this building block can directly interact with L1/L2 DU related tasks, leveraging the semantic communication paradigm to boost the overall network efficiency and performance.

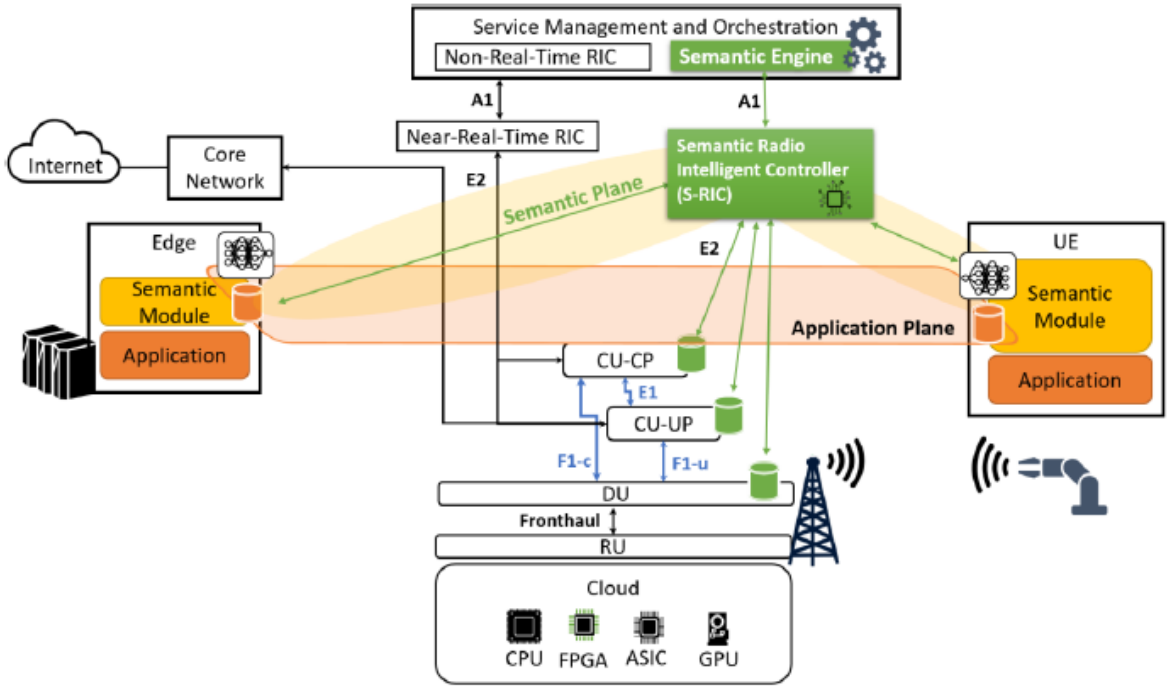


Figure 5.7: The conceptual architecture for Semantic Communications [Str+24]

5.2.3 6G ML TRAINING, INTENT BASED INTERFACE, NETWORK DIGITAL TWIN

Security and privacy enhancement in future 6G networks can be a quite challenging and demanding task due to the vast number of potential threats and attacks and their diverse nature compared to 5G networks (indeed, a larger attack surface is expected in 6G networks). In the same context, the interconnection of a vast number of devices and the support of heterogeneous deployments (exploiting the cloud continuum paradigm), which are both key concepts of the 6G era, escalates security and privacy concerns, since not all devices will have the capability to execute advanced security protocols due to their hardware constrained nature.

To enable high security and privacy in future 6G networks, AI/ML approaches can be deployed. Compared to conventional non-ML detection techniques, ML-based

misbehaviour detection provides both a higher detection accuracy against unknown zero-day attacks as well as a reduced false detection rate.

To be accurate, the AI/ML needs to collect a vast amount of data from the network to train models that can represent input/output pairs with minimum performance loss, mitigating security and privacy concerns via the extraction of abnormal data patterns and the enforcement of appropriate actions.

Deployment of ML approaches for threat detection and mitigation in the 6G landscape is influenced by various key driving factors: i) computational efficiency of the deployed approaches, ii) identification of multiple and even correlated threats and attacks, iii) continuous refinement of the ML approaches and knowledge distillation, and iv) creation of multiple network intents per case for network recovery [GNT+224].

The proposed reference architecture is illustrated in Figure 5.8. Key elements encompass the ML/deep learning (DL) training components dedicated to threat detection, the intent-based networking (IBN) components as well as the digital twin (DT) module.

ML/DL training (Distributed Threat Detection – DTD): This module is responsible for the distributed ML training. To this end, privacy preserving solutions are leveraged such as federated learning [LST+20]. Before the actual training, preprocessing and feature preparation takes place. Trained ML models are stored in a local database, where they can be retrieved on demand. All procedures are orchestrated by the machine learning function orchestrator (MLFO).

Intent-based Threat Mitigation (IBTM): Mitigation and preventive actions will be applied to the network using an intent-based interface (IBI) to facilitate human awareness. An intent should clearly define the desired state of the network while keeping its specification human-readable.

Sandbox – 6G Network DT: The network DT acts as a dynamic representation of the mobile network, constantly learning and evolving alongside the real network environment. ML algorithms, within this DT framework, can leverage historical data, network topologies, and user behaviour patterns to model normal network behaviour and promptly identify deviations that may indicate malicious activities. This integrated approach not only enhances the precision of threat detection but also empowers security systems to both anticipate and proactively mitigate potential risks as well as analyse the impact of any proactive action to be taken.

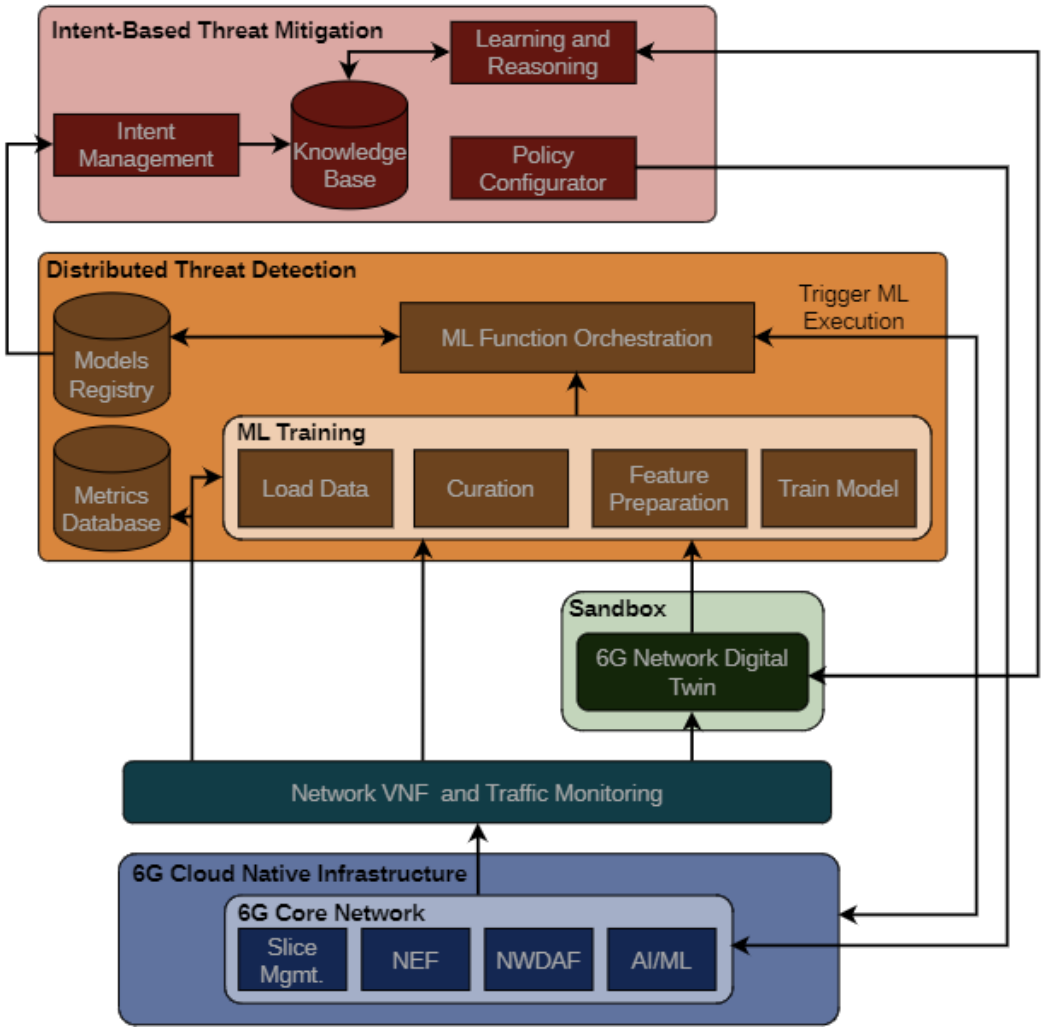


Figure 5.8: Proposed Architectural Approach for the DT integration

Within the DTD module, the data collected either directly from the network or from the emulated context with the help of the digital twin representation is used to train the appropriate ML models for threat mitigation. In this context, the IBTM component interacts with the DTD, where the proposed intents, being outputs of the ML model training, are translated into decisions to be applied to the network as previously mentioned. Therefore, DTD provides a high-level description of the mitigation or preventive actions to be enforced in the different 6G components in the form of an intent. Upon receiving the intent, the intent-based threat mitigation module (IBTM) module matches the received intent with existing information in the knowledge base and selects the proper matching policies. After checking whether the results of policies are as expected and if the new policies are aligned with existing policies and the decision to be taken, the polices can be enforced in the 6G infrastructure. All actions and policies aim either at strategically relocating virtual NFs (VNFs) to alternative cloud-native instances such as containers or other cloud hosts or completely isolating

malicious nodes. Access revocation to the 6G infrastructure is also supported to minimize potential risks and exposure to the threat. All proactive measures are designed to effectively mitigate identified threats and restore the optimal functionality of the 6G Core network.

5.3 MANAGEMENT AND ORCHESTRATION

6G network management needs to integrate several technical enablers, addressing the challenges of the multi-technology and distributed nature of future 6G infrastructures, the diversity of services to be delivered and the variety of stakeholders contributing to the whole ecosystem. Key features are the deep programmability and pervasive monitoring, which jointly enable synergetic, distributed orchestration combined with higher levels of network automation in scalable, multi-domain environments. Techniques like AI/ML algorithms, zero-touch closed loops and network digital twins are applied to bring increasing intelligence in the network, distributed through different layers and domains with functions deployed and configured on-demand following cloud-native, Service-Based Architecture (SBA) and as-a-Service patterns. The usage of service intents mixed with controlled but powerful network exposure capabilities facilitates more effective interactions with verticals and digital service providers, which is a key aspect for the monetization of value-added network services beyond mobile connectivity. Sustainability and trustworthiness follow a pervasive and “by-design” integrated approach. This involves the adoption of unified architectural principles and the embedding of algorithms, protocols, and workflows for user-centric, energy-efficient and secure procedures. Energy efficiency and security are considered not only as primary objectives of provisioning and automation decisions, but also as principles for the design and deployment of Management and Orchestration (M&O) components, introducing elements for sustainable MLOps or Federated Learning for privacy-preserving and explainable AI techniques.

5.3.1 EDGE CONTROL

Supporting edge computing applications (e.g., AI applications) is one of the most exciting features of future mobile networks. These services involve collecting and processing voluminous data streams right at the network edge to offer real-time services to users. However, their widespread deployment is hampered by the energy cost they induce on the network.

Recently, O-RAN has studied collaboration and convergence across domains to enable cross-domain AI optimization [ORAN23-nGRG]. As RAN virtualization enables the use of Commercial-Off-The-Shelf (COTS) hardware for deploying RANs, it opens the door to the joint orchestration and management of RAN, Core, and edge applications. The network's role in these services extends beyond merely transmitting and processing data in transit. Instead, the network must directly enhance AI service performance by optimizing for accuracy (reliable inferences), end-to-end latency (swift inferences), and task throughput (inferences per second) in a resource-efficient manner. This last requirement is critical because these services generate substantial data flows, involve intensive computations, and consume significant energy.

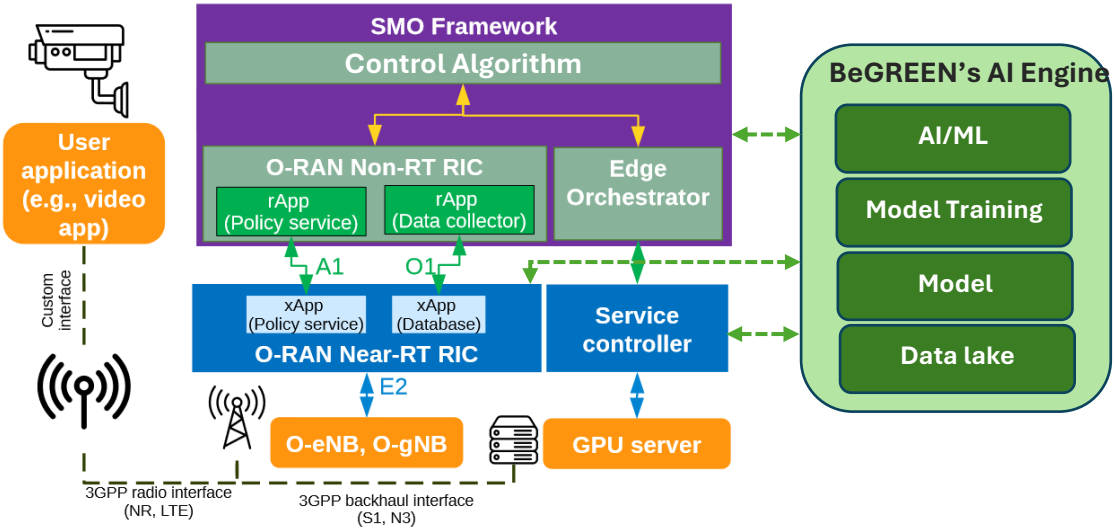


Figure 5.9: Intelligence Plane architecture as proposed in [BEG24-D42]

To this end, the Intelligence Plane [BEG24-D42] includes an Edge Control function in the SMO designed to control the resources of edge servers hosting edge applications dynamically. This function is exposed to the non-RT RIC, whose rApps aim at optimizing edge resources jointly with radio resources to enhance energy efficiency. It also enables the creation of control loops focused on joint optimizations. Since this approach has several similarities with the O-cloud management performed by the SMO, an interface leveraging O2, denoted as O2+, has been specified between the SMO and the edge server. Through this interface, the SMO shall be able to obtain the configuration of the edge, monitor its resources, and apply configurations related to energy-saving modes, CPU frequency control, CPU allocation, or GPU allocation.

Functions for optimising edge applications and RAN configuration policies are deployed as rApps in the O-RAN's non-RT RIC to enforce radio control policies in O-

RAN-compliant eNBs or gNBs. The edge control rApps interact with O-RAN's A1 interface (specifically, the A1's Policy Management Service) to enforce the corresponding radio policies. An xApp handles the A1 service from O-RAN's Near-RT RIC side and uses an E2 interface to forward radio policies to the Base Station. The E2 interface is also used to gather BS KPIs, which are forwarded to the non-RT RIC through the O1 interface. Then, a second xApp manages data KPIs received from the virtual BS and sends them to the Data Lake. Figure 5.9 summarizes the interfaces involved and the overall architecture of this use case for a video application at the edge.

5.3.2 6G NETWORK MANAGEMENT AND AUTOMATION

The overall M&O framework proposed in [HEX224-D63] integrates three main features, supported transversally by several technical enablers: intent-based service management, synergetic orchestration in the computing continuum and cognitive closed loops for network automation at runtime.

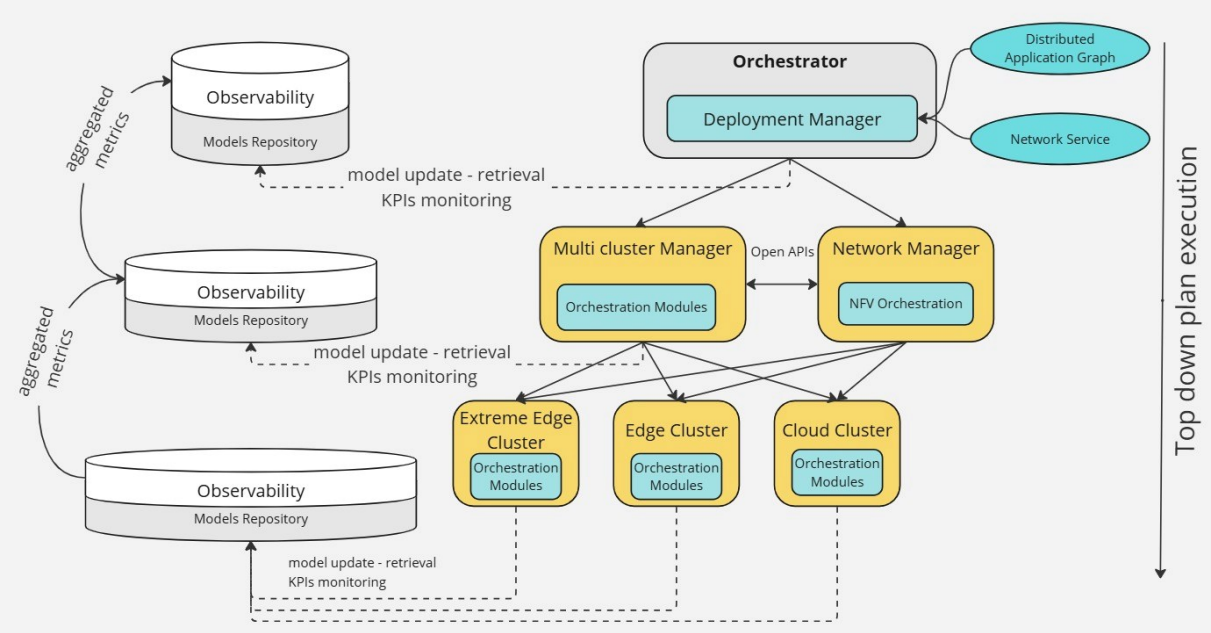


Figure 5.10: Resource orchestration in the computing continuum [HEX224-D63]

The synergetic orchestration enables the management of network services over programmable resources in the computing continuum, using techniques for distributed, de-centralized and/or federated management of the available resources. The scalable management of resources that span across the computing continuum (IoT devices, extreme edge devices, edge/cloud infrastructure shown in Figure 5.10) is crucial, leading to solutions based on hierarchical and multi-agent orchestration mechanisms, distributed mechanisms that build upon decentralized intelligence or federated

orchestration involving interactions between multiple orchestrators under different administrative domains.

In the past years edge and internet of things (IoT) computing have arisen as a paradigm that aims to provide computing, storage and networking capabilities in near proximity to the end-users while providing the same pay-as-you-go model of cloud computing. While edge computing enables application developers and content providers to leverage cloud computing capabilities and an IT service environment at the edge of the network, IoT computing distributes resources and services across the cloud, the edge, and the devices on the field to create the so-called compute continuum. The management and orchestration concepts currently applied to networks must be expanded in 6G to meet the requirements of the compute continuum.

Integration and orchestration of the extreme edge resources in the compute continuum demands M&O capabilities for computing resources beyond the radio access part of the network. The architectural components, interfaces, and mechanisms needed to orchestrate and manage the volatile and resource constrained extreme edge devices to be part of the computing continuum should be developed. For example, in the immersive experience, trusted environments and fully connected world use case families this enabler contributes to the improvement of privacy and security protection, since different applications from these use cases (e.g., eHealth) will be allowed to handle their sensitive information in the device, where it is generated. In addition, the collaborative robots and digital twins use case families require improved M&O and service continuity capabilities, which should provide mechanisms for flexible resource inclusion and allocation in the compute continuum. The compute continuum will have deep implications in the 6G architecture, where new interfaces and mechanisms need to be defined for: 1) exposing the capabilities of the extreme edge devices and 2) for interactions between the extreme edge, edge, and cloud resources.

In the area of network automation, zero-touch closed loops (CL) implement the logic for self-configuration, self-adaptation, and self-optimization towards autonomous, scalable management of dynamic and multi-technology networks. CL functions are deployed on-demand and interact with other M&O functions (e.g., monitoring, data analytics, digital twin, AI/ML functions) to build the four stages of automation workflows: Monitoring, Analysis, Decision, and Execution. CLs can be specialized for several objectives, e.g., SLA or intent assurance, resource usage optimization, etc. They can work in reactive, proactive or predictive mode, operating with different time scales, and

they can be applied to different layers or domains. CL coordination performed by applying techniques for conflict detection and mitigation or arbitration strategies is fundamental to guarantee the consistency and efficiency of decisions coming from concurrent and interdependent CLs.

5.3.3 NATIVE AI - PERVASIVE MONITORING SYSTEM

An AI-native 6G system architecture is defined by intelligence everywhere, distributed data infrastructure, zero touch management and AI as a Service [6GIA24-Vision]. This architecture enables AI/ML capabilities throughout the network, from central nodes to edge devices, supported by a robust data infrastructure for data availability, observability, pre-processing, and model lifecycle management across network layers.

In the context of DESIRE6G, we focus on delivering an AI-native 6G system architecture, revisiting the orchestration management, control and data planes (see Figure 5.11). Following the definition and principles outlined in [ERICSSON23], to make the system perceptive, DESIRE6G introduces a pervasive monitoring system that extends to the user equipment, leveraging in-band network telemetry solutions enabled by data plane programmability for precise, end-to-end information collection. Data access is provided at multiple layers with varying granularity to support decision-making across different levels and timescales, enabling operations at scale. We use an AI-driven Service Management and Orchestration layer (SMO) that supports non-RT decision-making, optimizations and MLOps. We further employ Multi-Agent Systems (MASs) to support intelligent near real-time control loops, pushing network decision making closer to the data plane [BRV24]. This functional split promotes service assurance through enabling faster control loops, ensuring the scalability of the system as its autonomous operation relies mainly on the autonomous coordinated operation of the agents.

At the SMO level, the Optimization Engine is the entity responsible for generic optimizations on medium to long timescales (> 1 sec), while the ML Function Orchestrator (MLFO) is responsible for deploying and, if needed, reconfiguring the MAS of a given service; it is in charge of creating AI/ML pipelines and relating them to the target service. AI/ML pipelines are associated to network entities and need to be deployed and reconfigured properly according to needs, e.g., flow rerouting of a service requires moving agents (with their performance data and models) among different

DESIRE6G sites. Service assurance is achieved mainly by the service-specific MAS, which implements distributed network intelligence closer to the physical infrastructure. MAS is responsible for receiving service-specific monitoring information and fine-tuning the network and compute resources to meet service-level KPIs (e.g. routing [BSM+24], elastic scaling of computing resources [HMP+23]). It configures and uses the pervasive telemetry system to receive service performance indicators, e.g., end-to-end latency for latency-sensitive or latency-critical services.

Additionally, we integrate edge AI capabilities to optimize network services (e.g., RIS configuration [CSB23]) and applications. This includes extending the architectural framework to support in-network machine learning (ML) through the Infrastructure Management Layer that acts as a combination of a Virtualized Infrastructure Manager (VIM) and a hardware abstraction (HAL) layer and the integration of frameworks like SOL and VACCEL, which facilitate rapid cross-framework and cross-hardware execution of AI tasks. Through these innovations, DESIRE6G promotes pervasive AI, fostering collaborative intelligence across the network infrastructure. The description of the DESIRE6G architecture and respective architectural components is provided in [DESIRE24-D22].

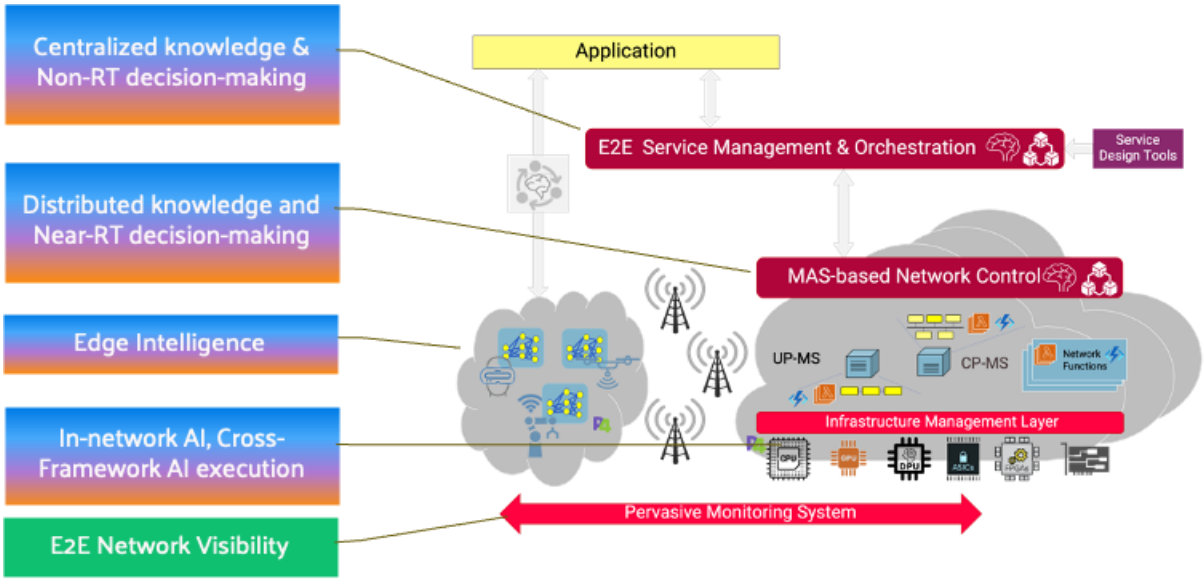


Figure 5.11: AI-native architecture [DESIRE24-D22]

5.4 DIGITAL TWIN

This section discusses several methods on how to introduce a digital twin of the network. Digital twins enable a virtual environment, where AI-driven NFs can be

developed, tested, and optimized without jeopardizing the actual network performance, and they can support automation, real-time decision-making, and dynamic resource allocation through the integration of Network Digital Twins.

5.4.1 INTEGRATING NETWORK DIGITAL TWINS INTO 6G ARCHITECTURES

The increasing complexity and dynamic nature of communication networks as we move towards 6G presents several challenges. Traditional network management methods are becoming insufficient in handling the scale and variety of modern networks, particularly when real-time, high-speed data processing is required.

The challenge also stems from the limitations in current network simulation tools, which are typically designed for specific use cases or domains and often lack the scalability required for larger and more complex scenarios. Current approaches do not fully account for the integration of AI-based management or the seamless coupling between the physical and digital network elements. As a result, networks struggle to adapt to rapidly changing environments, making it essential to develop new architectures that can support automation, real-time decision-making, and dynamic resource allocation through the integration of Network Digital Twins (NDTs) [TCF24].

The concept of NDTs, when integrated into a 6G architecture, addresses these challenges by providing a virtual environment, where AI-driven NFs can be developed, tested, and optimized. This approach allows for real-time adaptation and automation, ensuring networks can evolve and scale to meet the demands of future applications [Fay+24].

To address the outlined challenges, the initial proposed architecture introduces several key components that extend existing standard development organization (SDO) architectures. One of the main architectural innovations is the integration of a Network Digital Twin (NDT), which operates across three distinct layers: the physical network, the digital network, and a federated simulation framework. Figure 5.12 illustrates a high-level architecture of the overall proposed solution, highlighting the interaction between various layers and the integration of NDTs for closed-loop management and control.

The physical layer remains consistent with existing network elements, such as User Equipment (UE), RAN, and core network, while the digital layer introduces a network twin that allows for dynamic simulation and control. The digital layer is built upon the

ITU-T Y.3090 recommendation [ITUT22-Y3090], which outlines two core model types: basic and functional models. A as defined in [ZST24]:

- A basic model of a network element is the collection of data describing its properties, configurations, and operational status, along with any associated algorithms or protocols used to emulate its dynamics and evolution with time. A basic model of a network is the aggregation of basic models of network elements, including their physical and logical relationships and the interactions that occur between them.
- A functional model of a network builds upon basic models, applying advanced processing techniques, often through AI/ML algorithms, under varying operational scenarios. These models are designed for specific objectives such as performance optimization, anomaly detection, or predictive maintenance.

The third layer, the federated simulation framework, enables the coupling of multiple domain-specific simulators, forming a unified system that allows for large-scale scenario testing. This framework supports both online and offline NDTs, enabling networks to perform "what-if" analyses and refine AI-based functions before deploying them in real-world environments. This is critical for the orchestration of AI-driven services, providing a feedback loop for real-time performance optimization.

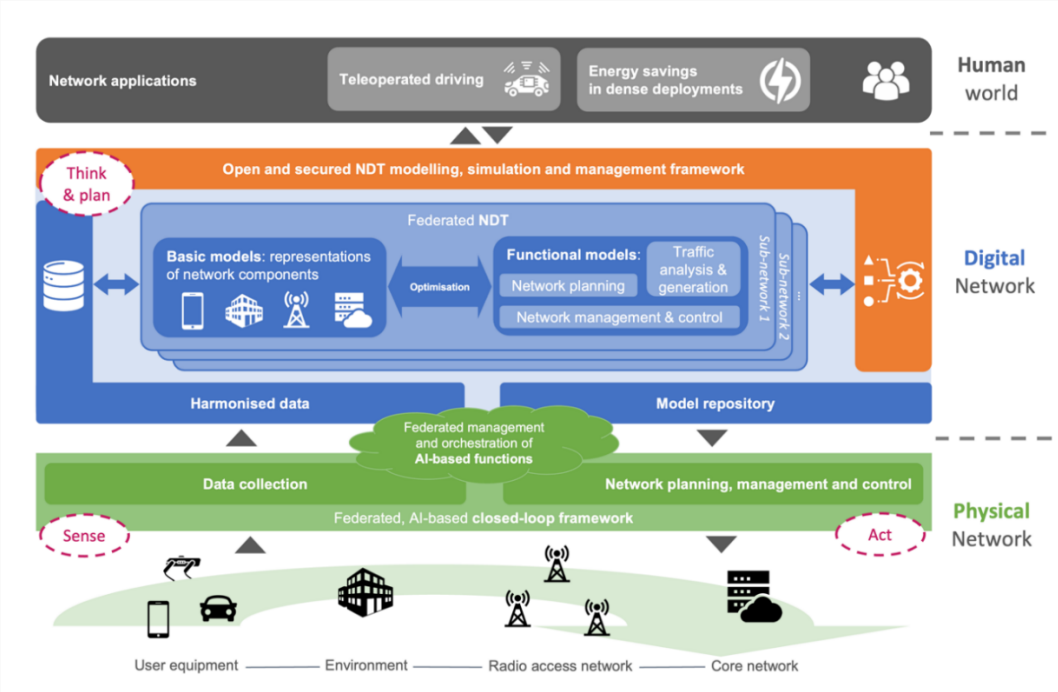


Figure 5.12: A high-level architecture of the proposed network digital twin solution for closed-loop management and control

5.4.2 AI-DRIVEN AND MLOPS-ENABLED NETWORK DIGITAL TWIN FOR MULTI-DOMAIN COMMUNICATIONS

To enhance deterministic communications, the combination of AI/ML capabilities with reliability and time sensitiveness aspects can be introduced in the 6G system to simulate and predict future network status, including KPI prediction for provisioning services, simulation of different types of flows, user demands, network congestion, etc. These capabilities can also be integrated with Network Digital Twins (NDT, [IRTF-Zhu+24]) and monitoring modules, where data can be generated or ingested for training and generating AI/ML models or inferences, or for assessing and validating the outcomes of the models generated across the different network domains. Indeed, the use of an NDT technology can help in estimating the achievable performance of deterministic service traffic flows in near-real-time. The implementation of an NDT in production environments faces several challenges and requirements. Among others, it includes: i) high accuracy in the estimating/prediction of the performance of both already deployed and requested traffic flows; such performance can be defined in terms of KPI, such as end-to-end delay and delay variation; ii) the ability to easily define scenarios for KPI estimation; and, iii) short computation time, (e.g., under 1 min), to provide KPIs when traffic conditions vary, for example, with time.

The introduction and serving of AI/ML capabilities and its combination with NDTs and monitoring components across multiple technology domains poses a new challenge towards the design of 6G system architecture. Such combination of techniques allows to understand the network infrastructure behaviour, anticipate future states and steer the network infrastructure towards the desired goal in terms of service performance.

The concept of MLOps is defined in the literature as the technology paradigm based on the extension of the DevOps methodology for enabling the full lifecycle management of AI/ML models in production environments. As part of this approach, MLOps targets the integration of different processes related to data collection, data transformation or model training towards the development and deployment of AI/ML models [EAD14].

Leveraging this technology concept and considering the significant role to be played by AI/ML techniques in 6G systems [BKJ+23], the design and integration of a specific AI/ML framework solution applying MLOps main principles is seen as a key enabler to support and facilitate the development and delivery of AI/ML services across multiple 6G System domains.

The current contribution proposes the integration of an AI/ML framework in the Management and Orchestration domain control loops of 6G architecture to enable the management of the complete lifecycle of AI/ML models, from design and training to deployment and serving in production environments at different domains. Serving of AI/ML capabilities can be shared either through the serving of AI/ML models in the shape of artifacts or through the delivery of AI/ML inference results, i.e., the resulting predictions or model outputs [BGG+23].

In the context of enhancing determinism, the introduction of AI/ML framework allows the training and distribution of AI/ML models focused on the simulation and prediction of service performance KPIs. The integration of these models with a Network Digital Twin architecture component would allow to trigger the simulation of service KPIs before performing path computation processes as part of the deterministic service provisioning stage. Additionally, monitoring information flows can be used to evaluate and retrain AI/ML models according to the real performance obtained, once specific network and resource configurations for provisioning deterministic services are applied.

From the architectural point of view, the proper deployment of an NDT in support of service KPIs prediction/estimation requires the usage of standardized interfaces to assure: i) interoperability between the NDT and the underlying real network/system to enable real-data monitoring and collection to allow, for example, NDT models adjustment, and ii) integration between the NDT and the network Control Plane (e.g., Service Automation, Path Computation, Monitoring, etc.) to support the internal communications among control plane architectural components for enabling service provisioning workflows.

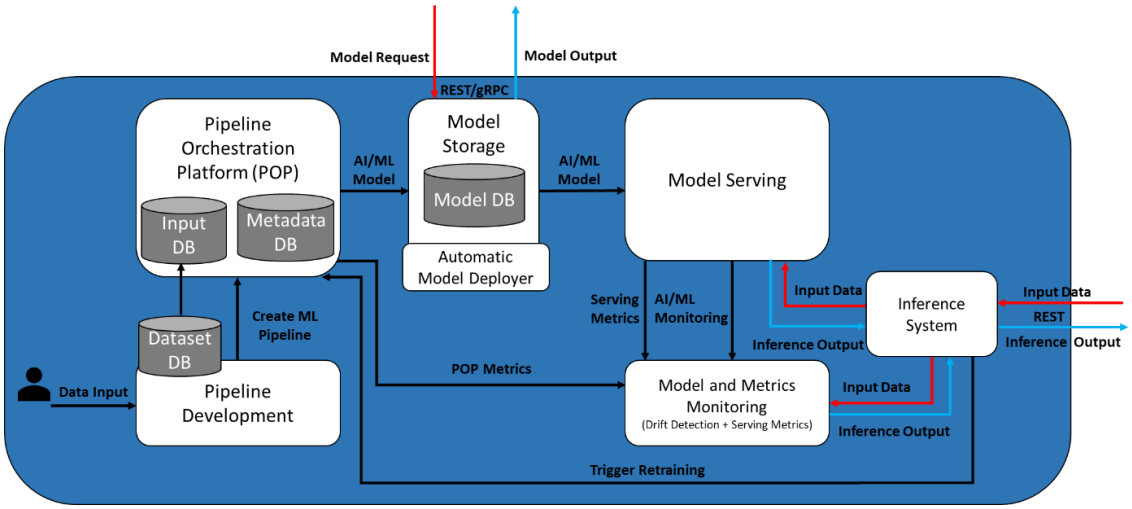


Figure 5.13: MLOps-based AI/ML framework in support of Digital Twinning

The design of the AI/ML framework solution is proposed as a unified AI/ML platform based on a service-oriented design that allows to design, train, serve and monitor different types of AI/ML algorithms (e.g. supervised, unsupervised, hybrid, distributed, etc.) in different network domains as part of the Management & Orchestration domain control loops.

To offer these functionalities, AI/ML framework solution is composed of six main blocks (see Figure 5.13): Pipeline Development, Pipeline Orchestration Framework (POP), Model Storage, Model Serving, Inference System, and Model and Metrics Monitoring. The outputs of the framework can be served to any domain in the shape of models via the REST/gRPC interface available in Model Storage module, or in the shape of inferences via the REST interface available in the Inference System module [PREDICT23-D31], [PREDICT24-D32]. In the context of determinism, this solution is integrated with the NDT and monitoring frameworks, the NDT and the monitoring being in charge of feeding the models for obtaining Service KPI simulations or predictions and for validating the results obtained.

5.5 REFERENCES

[3GPP24-23.501] 3GPP, "System Architecture for the 5G System (5GS)", 3rd Generation Partnership Project, Technical Specification TS 23.501, ver. 19.1.0, September 2024.

[6GIA24-Vision] Uusitalo, M., Bernardos, C. J., Kaloxylos, A., Bourse, D. A., Norp, T., Lønsethagen, H., Hecker, A., Rugeland, P., Papagianni, C., Bulakci, Ö., Li, X., Ericson, M., Anton-Haro, C., Massod Khorsandi, B., Ramos-Lopez, A., Frascolla, V., Marco, G., Gavras, A., & Trichias, K. (2024). European Vision for the 6G Network Ecosystem. Zenodo. https://doi.org/10.5281/zenodo.14230482

[AHM24] R. Adeogun, H. Hrasnica, C. N. Manchon, The CENTRIC Project Vision on Sustainable AI-native Air-Interface for 6G Networks, Feb 2024, aval: https://centric-sns.eu/wp-content/uploads/2024/02/centric_first_vision_paper_v1.pdf

[ARF+24] Ilias Alexandropoulos, Vasiliki Rentoula, Dimitrios Fragkos, Nikolaos Gkatzios, Harilaos Koumaras, "An AI-assisted User-Intent 6G System for Dynamic Throughput Provision", IEEE International Workshop on Computer-Aided Modeling, Analysis, and Design of Communication Links and Networks, CAMAD 2024

[5BDG+14] Bosshart, P., Daly, D., Gibb, G., Izzard, M., McKeown, N., Rexford, J., ... & Walker, D. (2014). P4: Programming protocol-independent packet processors. ACM SIGCOMM Computer Communication Review, 44(3), 87-95.

[BEG23-D41] BeGREEN Deliverable D4.1, "State-of-the-Art Review and Initial Definition of BeGREEN O-RAN Intelligence Plane", December 2023, [Online] Available: https://www.sns-begreen.com/deliverables?id=1008107

[BEG24-D42] BeGREEN Deliverable D4.2, "Initial Evaluation of BeGREEN O-RAN Intelligence Plane, and AI/ML Algorithms for NFV User-Plane And Edge Service Control Energy Efficiency Optimization", [Online] Available: https://www.sns-begreen.com/deliverables?id=1031615

[BGG+23] Bahare, M. K., Gavras, A., Gramaglia, M., Cosmas, J., Li, X., Bulakci, Ö., Rahman, A., Kostopoulos, A., Mesodiakaki, A., Tsolkas, D., Ericson, M., Boldi, M., Uusitalo, M., Ghoraishi, M.,

& Rugeland, P. (2023). The 6G Architecture Landscape - European perspective. Zenodo. https://doi.org/10.5281/zenodo.7313232

[BKJ+23] Ankur Bang, Kapil Kant Kamal, Padmaja Joshi, Kavita Bhatia, 6G: The Next Giant Leap for AI and ML,Procedia Computer Science, Volume 218, 2023, Pages 310-317, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.013.

[3BKL+24a] F. Brisch, A. J. Kassler, S. Laki, P. Hudoba, G. Pongrácz, P4-MTAGG - a Framework for Multi-Tenant P4 Network Devices, Würzburg Workshop on 6G Networks (WueWoWAS'24), 30 September-1 October 2024, Würzburg, 2024.

[4BKL+24b] F. Brisch, A. J. Kassler, S. Laki, P. Hudoba, P4-MTAGG - a Framework for Multi-Tenant P4 Network Devices, 20th International Conference on Network and Service Management (CNSM 2024), DEMO PAPER, 28-31 October, Prague, Czech Republic, 2024.

[BRV24] S. Barzegar, M. Ruiz, and L. Velasco, "Autonomous Flow Routing for Near Real-Time Quality of Service Assurance," IEEE Transactions on Network and Service Management (TNSM), vol. 21, pp. 2504-2514, 2024.

[BSM+24] S. Barzegar, H. Shakespear-Miles, F. Alhamed, F. Paolucci, P. Gonzalez, M. Ruiz, and L. Velasco, "Near Real-Time Autonomous Multi-Flow Routing with Dynamic Optical Bypass Management," in Proc. International Conference on Optical Network Design and Modeling (ONDM), 2024.

[CENTR24-D21] S. Cammerer and J. Hoydis, “D2.1 Evaluation of GPU Implementation of ML-based Receiver”, Zenodo, Jul. 2024. doi: 10.5281/zenodo.12773161.

[CSB23] C. B. Chaaya, S. Samarakoon and M. Bennis, “Federated Learning Games for Reconfigurable Intelligent Surfaces via Causal Representations”. In GLOBECOM 2023-2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 6567-6572.

[DESIRE24-D22] Pongrácz, G., & Papagianni, C. (2024). D2.2: DESIRE6G Functional Architecture Definition. Zenodo. https://doi.org/10.5281/zenodo.12784579

[EAD14] Erich F., Amrit C., Daneva M., “A Mapping Study on Cooperation between Information System Development and Operations”, In: Jedlitschka A., Kuvaja P., Kuhrmann M., Männistö T., Münch J., Raatikainen M. (eds) Product-Focused Software Process Improvement. PROFES 2014. Lecture Notes in Computer Science, vol 8892. Springer, Cham, 2014.

[ERICSSON23] Ericsson. (2023). “Defining AI native: A key enabler for advanced intelligent telecom networks [White paper]. https://www.ericsson.com/en/reports-and-papers

[Fay+24] S. Faye et al., "Integrating Network Digital Twinning into Future AI-based 6G Systems: The 6G-TWIN Vision," 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024, pp. 883-888, doi: 10.1109/EuCNC/6GSummit60053.2024.10597058.

[GKF+24] N. Gkatzios, H. Koumaras, D. Fragkos and V. Koumaras, "A Proof of Concept Implementation of an AI-assisted User-Centric 6G Network," 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024, pp. 907-912, doi: 10.1109/EuCNC/6GSummit60053.2024.10597020.

[GNT+224] P. Gkonis, N. Nomikos, P. Trakadas, L. Sarakis, G. Xylouris, X. Masip- Bruin, and J. Martrat, “Leveraging network data analytics function and machine learning for data collection, resource optimization, security and privacy in 6g networks,” IEEE Access, pp. 1–1, 2024.

[HEX223-D33] HEXA-X-II Deliverable D3.3, “Initial analysis of architectural enablers and framework” https://hexa-x-ii.eu/wp-content/uploads/2024/04/Hexa-X-II_D3.3_v1.0.pdf, April 30, 2024

[HEX224-D63] HEXA-X-II Deliverable D6.3 – Initial Design of 6G Smart Network Management Framework, 2024. Online. https://hexa-x-ii.eu/wp-content/uploads/2024/07/Hexa-X-II_D6-3_v1.0.pdf

[HMP+23] Hsu CS, Martín-Pérez J, Papagianni C, Grosso P. V2n service scaling with deep

reinforcement learning. InNOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium 2023 May 8 (pp. 1-5). IEEE.

[IRTF-Zhu+24] C. Zhou et al., "Network Digital Twin: Concepts and Reference Architecture", draft-irtf-nmrg-network-digital-twin-arch-06", July 2024.

[ITUR19-NET2030] White Paper Network 2030: A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond. FG-NET-2030 group - ITU-R, 2019. Available online at https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf

[ITUR23-M2160] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond", M.2160, November 2023. [Online]. Available: https://www.itu.int/rec/R-REC-M.2160/en.

[ITUT22-Y3090] ITUT, "Digital twin network – Requirements and architecture.," ITU, recommendation ITU-T Y.3090, 2022.

[KCC+22] T. Kellermann, R. P. Centelles, D. Camps-Mur, R. Ferrús, M. Guadalupi and A. C. Augé, "Novel Architecture for Cellular IoT in Future Non-Terrestrial Networks: Store and Forward Adaptations for Enabling Discontinuous Feeder Link Operation," in IEEE Access, vol. 10, pp. 68922-68936, 2022, doi: 10.1109/ACCESS.2022.3184720.

[LST+20] [2] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749.

[Mao+22] B. Mao et al., "AI Models for Green Communications Towards 6G," IEEE Communications Surveys Tutorials, vol. 24, no. 1, 2022.

[Nt+24] K. Ntontin et al., "ETHER: A 6G Architectural Framework for 3D Multi-Layered Networks," 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, 2024, pp. 1-6, doi: 10.1109/WCNC57260.2024.10570731

[ORAN23-nGRG] O-RAN Alliance, "Research Report on Native and Cross-domain AI: State of the art and future outlook", nGRG group, September 2023.

[2P4R24] P4Runtime Specification. 2024. https://p4.org/p4-spec/p4runtime/main/P4Runtime-Spec.html

[PREDICT23-D31] PREDICT-6G Consortium. (2023). D3.1 Release 1 of AI-driven inter-domain network control, management, and orchestration innovations. Zenodo. https://doi.org/10.5281/zenodo.12167712

[PREDICT24-D32] PREDICT-6G Consortium. (2024). D3.2 Implementation of selected release 1 AI-driven inter-domain network control, management and orchestration innovations. Zenodo. https://doi.org/10.5281/zenodo.12167665

[SAFE6G] Project SAFE-6G. Online. https://safe-6g.eu/

[STK+24] M. M. Saad, M. A. Tariq, M. T. R. Khan and D. Kim, "Non-Terrestrial Networks: An Overview of 3GPP Release 17 & 18," in IEEE Internet of Things Magazine, vol. 7, no. 1, pp. 20-26, January 2024, doi: 10.1109/IOTM.001.2300154.

[Str+24] E. C. Strinati et al., "Goal-Oriented and Semantic Communication in 6G AI-Native Networks: The 6G-GOALS Approach," 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024, pp. 1-6, doi: 10.1109/EuCNC/6GSummit60053.2024.10597087.

[TCF24] I. Turcanu, G. Castignani and S. Faye, "On the Integration of Digital Twin Networks into City Digital Twins: Benefits and Challenges," 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2024, pp. 752-758, doi: 10.1109/CCNC51664.2024.10454704.

[ZST24] A. Zaki-Hindi, JS. Sottet, I. Turcanu, and S. Faye, " Network Digital Twins for 6G: Defining Data Taxonomy and Data Models," 2024 IEEE Conference on Standards for

Communications and Networking (CSCN), Belgrade, Serbia, 2024.

6 INTEGRATED SENSING AND COMMUNICATION

6G services will be associated with a wide spectrum of vertical applications with greatly varying requirements and will offer advanced features beyond connectivity spanning from sensing to monitoring and positioning. To address these requirements, 6G will feature Integrated Sensing and Communication (ISAC) capabilities, performing sensing through the mobile communication infrastructure. This can be achieved adopting either a Channel State Information (CSI) or a passive radar approach. The CSI sensing approach relies on the connectivity established between the base station (BS) and the user equipment (UE) to estimate channel conditions and extract information for the Angle of Arrival (AoA) and the Time Difference of Arrival (TDoA). This information can be then used to support a set of applications[1] including human localization and tracking, presence detection, activity recognition, healthcare, etc. In the "radar" sensor approach, the network exploits its own radio signals to sense and comprehend the surrounding physical world. The echoes (reflections) and scattering of wireless signals predominately transmitted for communication purposes, provide information related to the characteristics of the environment and/or objects therein [3GPP24-22.837]. The sensing data collected and processed by the network can then be leveraged to enhance the operations of the network, augment existing services such as XR and digital twinning, and enable new services such as object detection and tracking, along with imaging and environment reconstruction. This potential has already attracted a lot of attention from 3GPP, which has initiated a preliminary study on use cases and ISAC requirements, making it a promising candidate to optimize both communications and sensing systems [3GPP24-22.837].

Depending on the level of integration of the sensing functionality into the communication network, different approaches can be adopted as follows:

- Fully separated infrastructures performing sensing and communications functionalities. Based on this approach, information acquired from one infrastructure is used to assist the other.

[1] An Overview on IEEE 802.11bf: WLAN Sensing

- Common hardware supporting sensing and communication capabilities. This approach is implemented by sharing the available spectrum, with the constraint that sensing and communication signals are transmitted over different timeslots.
- Fully integrated systems sharing both spectrum and time domains.

Depending on the number and roles of the devices involved in sensing several options also exist including the following:

- The monostatic case, where a single device is used for transmitting and receiving sensing signals.
- The bi-static/multi-static sensing, where a single transmitter and physically separated single or multiple receivers are used to acquire the sensing signals.
- The passive sensing approach, where signals transmitted primarily for communication purposes can be also used by other devices for sensing.

6.1 INTEGRATION OF NON-3GPP AND 3GPP SENSING

Although some early prototypes are available for validating sensing concepts, these are mostly designed for non-3GPP networks (i.e., Wi-Fi), whereas implementations of 3GPP-compliant passive radar-based ISAC systems are still at a very early stage. The main reason is that these systems demand additional complexity in signal processing and require collection and aggregation of huge volumes of synchronized in-phase and quadrature (IQ) reflected (echo) streams that need to be processed to extract information on the sensed environment. This processing can only be performed at edge servers, introducing the need to transport the IQ streams over flexible high-capacity transport networks.

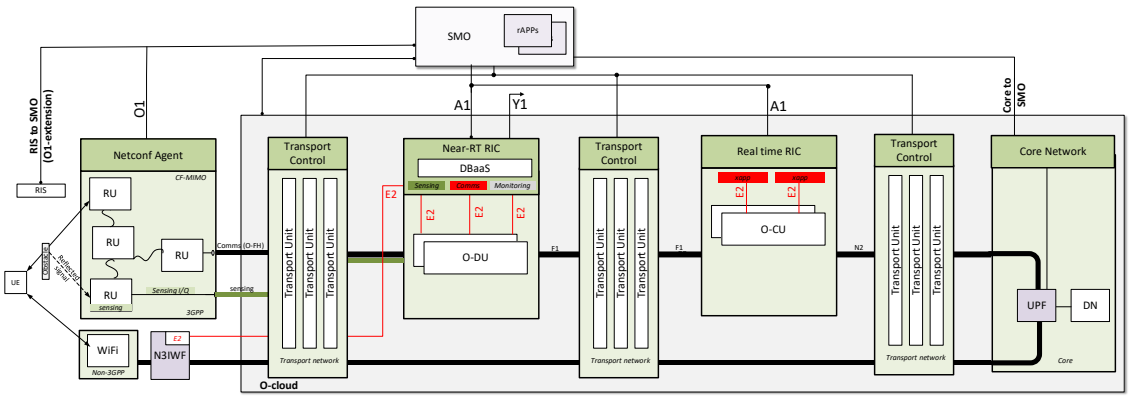


Figure 6.1: Generic architecture of integration of 3GPP and non-3GPP sensing in the core network [6GSENSESD2.2]

Figure 6.1 proposes a 6G architecture that interconnects a multi-technology Radio Access Network (RAN) able to offer sensing functionalities (3GPP and non-3GPP) with core network domains to facilitate joint support of sensing and communication services. The RAN technologies of interest include non-3GPP (WiFi) and 3GPP based (5G NR) networks, which will coexist in an ISAC framework to obtain accurate representation of the surrounding environment.

Non-3GPP based sensing is performed adopting Wi-Fi networks, which are extended to operate as monostatic and multi-static radars. The output of the sensing information from non-3GPP networks is transmitted to the RAN segment through suitable extensions of the E2 interface of the RAN Intelligent Controller (RIC). To achieve this, there is a need to enable Wi-Fi networks to expose their sensing related data in a secure way. Addressing this requirement, in [6GSENSESD2.2] the use of the non-3GPP Inter-Working Function (N3IWF), which is responsible for interworking between untrusted non-3GPP networks and the 5G core, is considered. 6G-SENSES, therefore, proposes to adopt and appropriately extend N3IWF, in order to provide the necessary *access and authentication protocols with new features that will allow Wi-Fi networks to securely expose sensing data to the RIC.* 3GPP-based sensing is performed based on the principle of a distributed passive wireless radar. According to this, 6G BSs generate communication signals reflected on "objects" located in the surrounding area, creating IQ echo streams. These IQ echo streams are transmitted in the form of uplink fronthaul streams to the DUs, where they are compressed (down-sampled) and transmitted through the E2 interface to the RIC. Purposely developed sensing xApps fuse the incoming sensing streams (IQ echo streams and Wi-Fi sending data), analyse their quality and cache data to a fast in-memory database. This data can be then exploited internally by the system to optimize the operational parameters of the various building blocks of the RAN segment (e.g. development of beamforming, beam steering, power control, etc.) or they can be exposed to the vertical applications through the Y1 interface.

The sensing output is also passed to the Service Management and Orchestration (SMO) that decides on the optimal network resource configuration to support both communication and sensing services. To perform this, the SMO provides mechanisms supporting automated lifecycle management (LCM) for ISAC services instantiating and automatically reconfiguring E2E slices considering both communication (i.e. fronthaul,

backhaul) and sensing services requirements. A first concept demonstration of this architectural approach is detailed in [AGT25].

6.2 DISTRIBUTED SENSING ARCHITECTURE

The goal of a distributed sensing system is to perform distributed and cooperative sensing, involving collection and exploitation of data sensed by multiple heterogeneous devices, i.e., sensing receiver nodes (SRNs), as well as tracking many heterogeneous (and mobile) targets, including both passive and active user equipment (UE), over a large area. In this regard, an important focus is to address the challenge of lack of suitable sensing control functions for distributed operation.

Sensing data is initially collected at the Transmission Reception Point (TRP), where some local processing occurs before measurement data is transmitted to the Sensing Management Function (SeMF) in the core network. The two main approaches include i) *Information-Level Fusion*, involving preliminary detection and estimation at each local station (e.g., Angle of Arrival and Time of Arrival), and transmission to a fusion centre (FC). The FC performs global detection and localization through pairing and triangulation of the extracted parameters, and ii) *Signal-Level Fusion*, involving directly and jointly processing the local raw observations at the FC for target detection and localization, thus avoiding the need for local decision fusion or target-measurement association [Y+23].

The trade-off between sensing performance and the volume of data to be stored and exchanged represents a major challenge when choosing the approach to be used within the limits of the application requirements. Moreover, mobility of passive targets, i.e., objects that are not connected with the network, brings the challenge of handover while tracking these objects over space and time to ensure continuity of sensing service.

Another key challenge to be explored is represented by including semantic and goal-oriented functionalities. This impacts the architecture design, requiring the inclusion of a semantic plane, which plays a pivotal role by managing the extraction, interpretation, and transmission of contextual meaning, thereby ensuring that the transmitted information is meaningful and aligned with the system's objectives, rather than just raw data. The main challenge is to effectively insert the semantic plane into a distributed integrated communication and sensing architecture, where additional interfaces and new semantic modules are required to ensure that sensing, communication, and computation tasks are aligned with the system's semantic goals, with the semantic

plane handling the dynamic adaptation of the meaning in changing environments, maintaining coherence across distributed nodes, and optimizing the network's overall performance.

The design of such distributed sensing methods involves accommodating for heterogeneous computing capacities, energy budget constraints as well as sensing and communication capabilities. In this regard, a key example is given by the integration of different types of reconfigurable intelligent surfaces (RIS) [CSA+21] (e.g., passive, active, hybrid, or autonomous [ARD+22]). While existing works generally focus separately either on the communication optimisation, the sensing accuracy or spectrum sharing techniques, the challenge is to develop novel methods to enable joint optimization of ISAC while accounting for such heterogeneous characteristics.

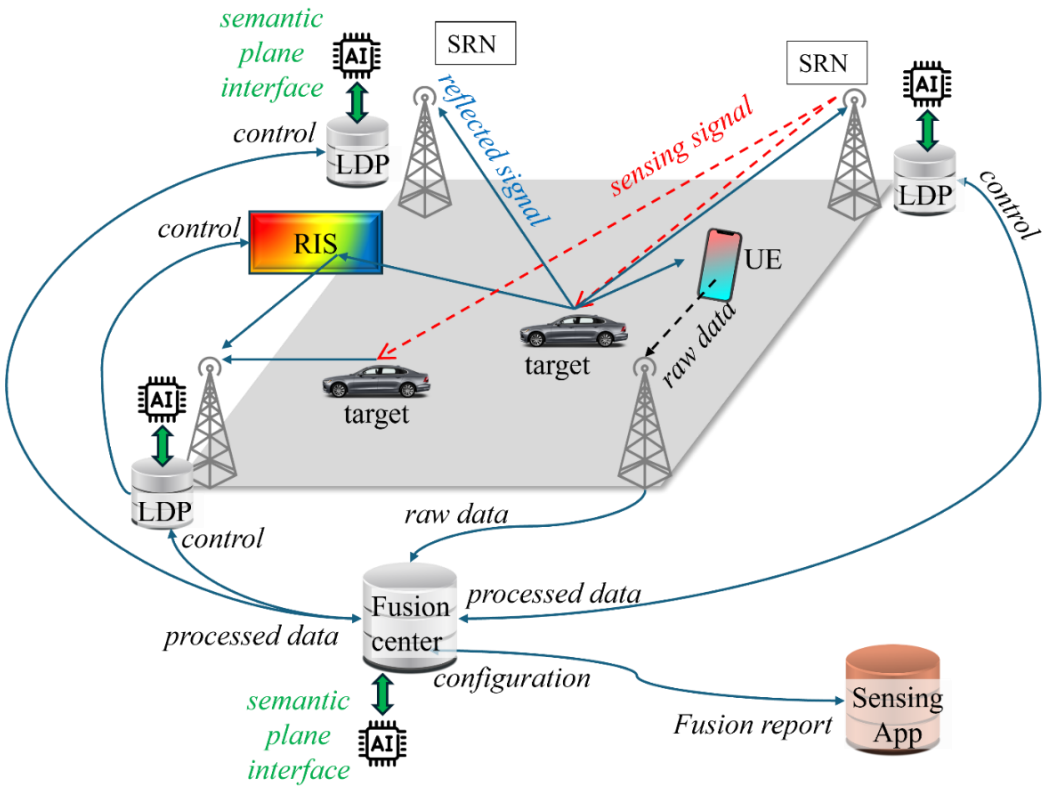


Figure 6.2: Distributed ISAC Architecture [C+25]

Figure 6.2 shows the major features of the distributed architecture such as distributed and intelligent processing, the use of the semantic plane for control and the role of RIS. The anticipated scope of distributed processing involves collaboration among intelligent nodes to provide a multi-perspective and integrated view of the environment of interest, including both active UEs and passive objects, extracting both channel-related propagation parameters such as angles, delays, Dopplers, and

geometric information like location, orientation, and size of the targets. This entails not just fusion of data from multiple SRNs, but also a map of the environment. Given the high dimensionality of the sensing data on MIMO systems, the data transformation is performed at the local (LDP) level to ensure that the communication capacity of the system is not impacted. At the same time, the architecture is flexible enough to support low-capacity nodes, which have limited computing power, thus handling heterogeneous nodes and their associated hardware-related limitations.

Devices that have different communication, sensing, computation, and storage capabilities as well as power consumption are considered in the design of resource allocation and orchestration schemes. This includes both energy-autonomous, energy-supplied, or energy-neutral STNs, devices equipped with advanced or basic sensing capabilities (e.g., almost passive and reflective, simultaneously reflecting and sensing, amplifying reflective RISs, massive MIMO). In the latter case of basic sensing-capable devices, the FC performs the data processing for those nodes. Furthermore, interfaces and protocols enable this intelligence sharing mechanism and cooperation among the different network elements. Such an architecture facilitates sharing of appropriate information from sensing modules to enable substantially enhanced communications via joint optimization of both multi-antenna transmissions and receptions as well as the reflective beamforming of multi-functional RISs for both localisation, sensing, and sensing-aided communications.

The semantic framework provides semantic processing to reduce these data volumes intelligently over time and space. In the context of multi-modal sensing using heterogeneous nodes, the framework provides semantic reasoning to efficiently integrate non-3GPP devices.

Using AI based reasoning, the semantic framework enables extraction of semantic information that uses not only the sensor data but also previously obtained background knowledge (for instance as an inference ML model trained from previous data and subject to rules imposed to the observation environment). This will enable the adaptation of the ISAC sensing parameters (refresh rate, performance criteria) and resource allocation to varying KPI/KVI requirements as well as selective information sharing (semantic information, extracted and processed given accumulated background knowledge) and reasoning about multi-modal sensed information (i.e., generated by different sensor types).

6.3 OPTICAL WIRELESS COMMUNICATION BASED ISAC

Localisation is one of the most essential sensing functions required for enabling the communications-computing continuum in Industry 4.0 use cases such as automated guided vehicles (AGV) navigation for enhancing industrial automation process. In such scenarios, Optical Wireless Communication (OWC) technology can be used as a supplementing radio technology, along with a sub-cm location measurement and sensing solution to enable sensing and localisation in the industry environment.

OWC ITU G.9991 networks are inherently cell-free with access to a cluster of 6 OWC photonic antennas controlled by a single OWC access point to typically support a 4x5 meters coverage area. This paves the way towards a beam-steered OWC system, which will ultimately produce much wider access angles, which would subsequently require fewer OWC APs. Increasing the radius by a factor of 4 with a field of emission angle of 70° and providing a coverage area from one access point of 16x20m = 320m2 would require 340 OWC AP clusters for a factory with a floor space area of 109,100m2, which is much more commercially viable.

The main technique of localization with OWC is mainly the use of the received signal strength (RSS), the TOA or time or phase difference of arrival (TDoA/PDoA) and the AOA to extract position information, either by trilateration or by triangulation. The RSS method is the most popular, and algorithms coupled with Kalman filters have been proposed to improve the positioning resolution and responsiveness to motion. Furthermore, precise round trip (RT) TOA measurements are possible, which are based on several APs and can be implemented using existing fields in the physical layer preamble, together with fine timing measurement (FTM) available in the MAC layer of IEEE 802.11 (Wi-Fi) standard. The aim is to achieve sub-centimetre accuracy in 3D, including the information of orientation angle of the target, based on several improvements [OPTI24-D21]: a) development of a combined method with RSS and RT TOA, which, to our knowledge, has not been addressed yet; b) combination of this first sensing approach with AOA measurements by modulating the angle of emission of the source and, thus, simplifying the receiver, which is an important point in the case of simple embedded receivers.

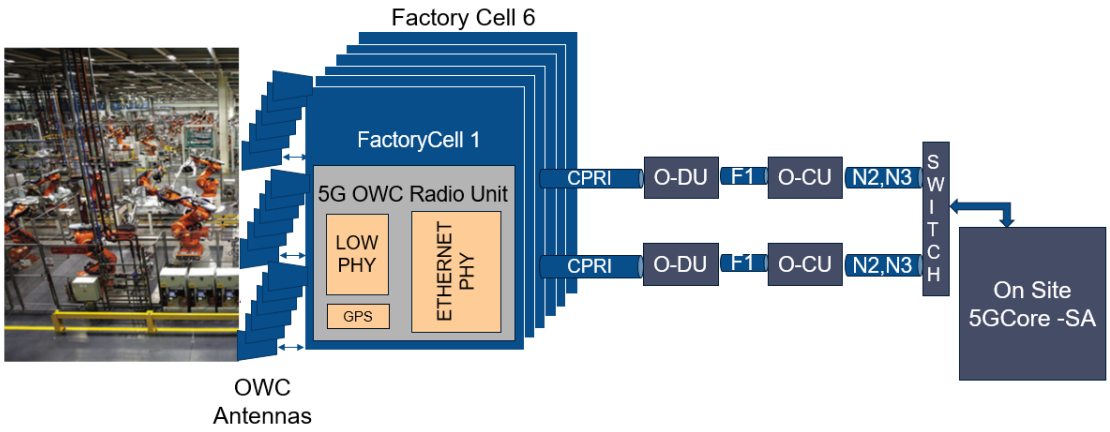


Figure 6.3: OWC-based Localization and Sensing

Localization of UEs using ToA from sub-6 GHz wireless and RSS from Optical Wireless Infrared techniques can be used to obtain an accuracy of less than 1 cm, however this performance is highly dependent on the continuous direct line of sight access between the gNB access points and the UE. In [CAM+24], the authors show in a laboratory experiment that 2 to 3 mm localization accuracy can be achieved using ToA in a sub-6 GHz 5G Orthogonal Frequency Division Multiplexing (OFDM) network. If there is no direct line of sight access to four or more gNB access points, then location ambiguity is introduced, and other techniques can be used to maintain localization such as AoA from received radio signatures. It has been shown that a 2 - 3 degrees orientation accuracy can in principle be obtained in a simulation experiment using Matlab's Siteviewer 3D radio propagation model [CAM+24]. If the received radio signatures estimate AoA from more than two access points or AoA and distance from one access point, then making some form of interim measurements for obtaining location can continue. In [OPTI24-D41], the authors propose to test if more accurate localization of UEs can be obtained using ToA from near IR OWC because of the greater system bandwidth that can be used to lower the levels of multipath reflections and noise.

6.4 SENSING SERVICE PROVISIONING AND EXPOSURE

There is a recent interest for using the mobile network resources for services beyond conventional connectivity, and one example of this is integrated communication and sensing (ICAS). The main challenges identified in this scope are mostly related to the capability of the network to support such challenging services in (i) a sustainable manner, and (ii) a trustworthy, privacy-preserving manner, as large amounts of -beyond communications- data will be required to be transmitted and processed, adhering to

strict latency requirements. In detail, the challenges include (a) the need for novel interfaces that support data collection, (b) data processing, (c) data distribution and scaling of interfaces, (d) trust differentiation when exposing to 3rd party applications, (e) network overload on the exposed Application Programming Interfaces (APIs), (f) privacy risks, and (g) latency/performance.

The Beyond Communications Network (BCN) paradigm will enable new 6G services such as sensing and compute offloading, and how to expose resulting data and relevant service capabilities in a secure, privacy-preserving and efficient manner. The exposure and data management enabler aims to reduce the overhead from data exposure by aggregating and fusing data while ensuring data privacy and trust. This enabler supports the creation of novel services that contribute to societal benefits like safety and sustainability, supported by its capability to efficiently handle and expose data from various producers, including the RAN and sensing nodes.

Any network entity with proper access rights should be able to access data or model(s) from another entity. A form of authorization and/or authentication should be performed when a network entity is trying to access/update/share data, analytics and model from/of/with another entity. Security should be enabled E2E for any operation of the data collection services, including access, exposure, storage, cleaning, processing and encoding. The network should be able to identify energy-aware data collection services and facilitate their operations. Data collection and exposure can be based on a local configuration, or a configuration received from the requester. Different data consumers exist in the network (defined as general network entity) such as UEs, RAN nodes, CN NFs, AFs, 3rd party applications, OAM, etc. Discovery, configuration and in some cases evaluations of such data sources are among the functionalities to cover in 6G.

There is a need to develop novel APIs, enabling both internal NFs as well as third-party applications to request, receive and manage data securely and efficiently, thereby reducing data traffic and overhead significantly. In the ISAC use case, exact information on the position of base stations and UEs can be used to enable some kind of QoS-based sensing. However, when the positioning accuracy of measurements nodes is provided with low confidence score, e.g., a UE position with some uncertainty, sensing services may have to be provided solely by the network on a best-effort basis.

One way to improve sensing quality is obviously to carry out more radio measurements prior to exposure of the measurement report to the requesting

application, preferably measurements that are geographically distributed. Involving more network nodes (UEs or base stations) naturally leads to architectural challenges of centralized vs. distributed inference and processing of the measurements.

The provision of sensing services by next generation communication systems necessitates the introduction of a Sensing Management Function (SeMF) [HEXAD3.3], see Figure 6.4, that will be responsible for facilitating an efficient coordination of sensing procedures, considering various aspects such as sensing requirements, sensing capabilities, sensing constraints, etc. The SeMF can be designed as a dedicated NF, since it is enabling a new functionality for next generation networks. An alternative option would be to integrate the SeMF services as part of the location management function of 5G. Since the amount of data from sensing may not be user specific, one idea can be to transmit sensing data via a "data plane" that is routed via sensing functions in the core network (and not via the UPF as for the user plane).

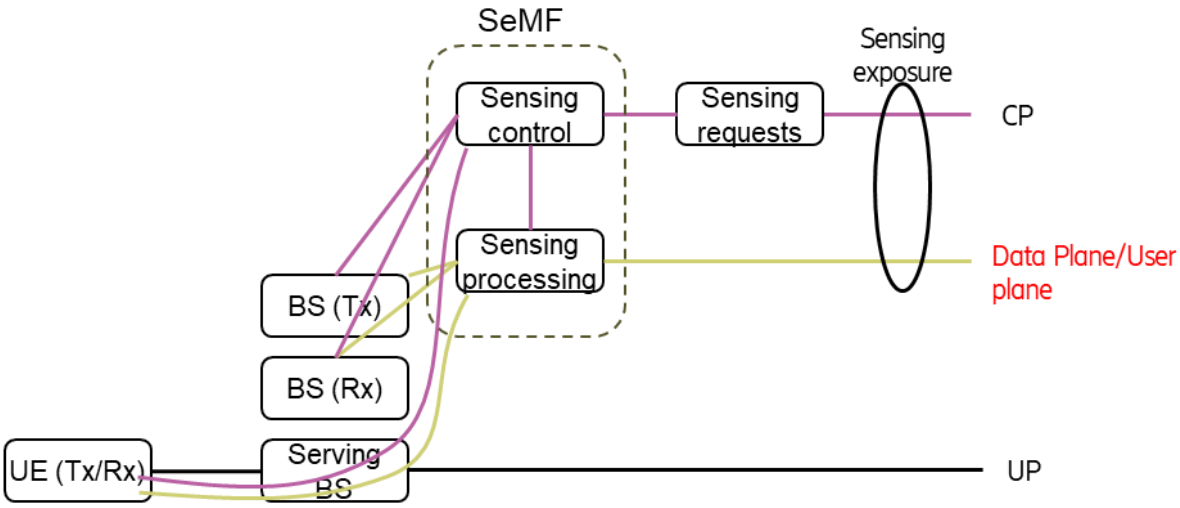


Figure 6.4: Functional Architecture, network-based sensing with UE involvement

The need for identifying and coordinating compute workloads, in the context of sensing service, AI or other computationally heavy tasks, in an efficient manner becomes critical. In particular, different ISAC applications are associated with stringent requirements. On the one hand, heavy computations need to be performed on the sensed data, coming from different sources, to provide with the information on their localization. On the other hand, ISAC applications could also be associated with delay-strict requirements, where the results are expected to be received in real-time (e.g., stopping a robot machine after detecting a human). In addition, although a far edge cloud is located near the end user and comes with promise of reduced communication

delay, it can suffer from scarcity of compute resources, which induces high processing delay. The trade-off between network metrics and compute metrics would call for a new approach that enables the Integration of Network and Compute (INC) domains to perform coordinated optimization.

The introduction of compute offloading in the next generation networks should not increase the complexity of the communication protocol. This can be achieved by tight integration and true convergence of communication and computing and introducing novel architectural components for distributed computing. To satisfy the strict requirements on the computation and communication latency, trustworthiness, power consumption and data accuracy, it is important to introduce a common classification of computing and communication resources of each novel component as well as a common characterization of offloaded compute workload based on predetermined requirements.

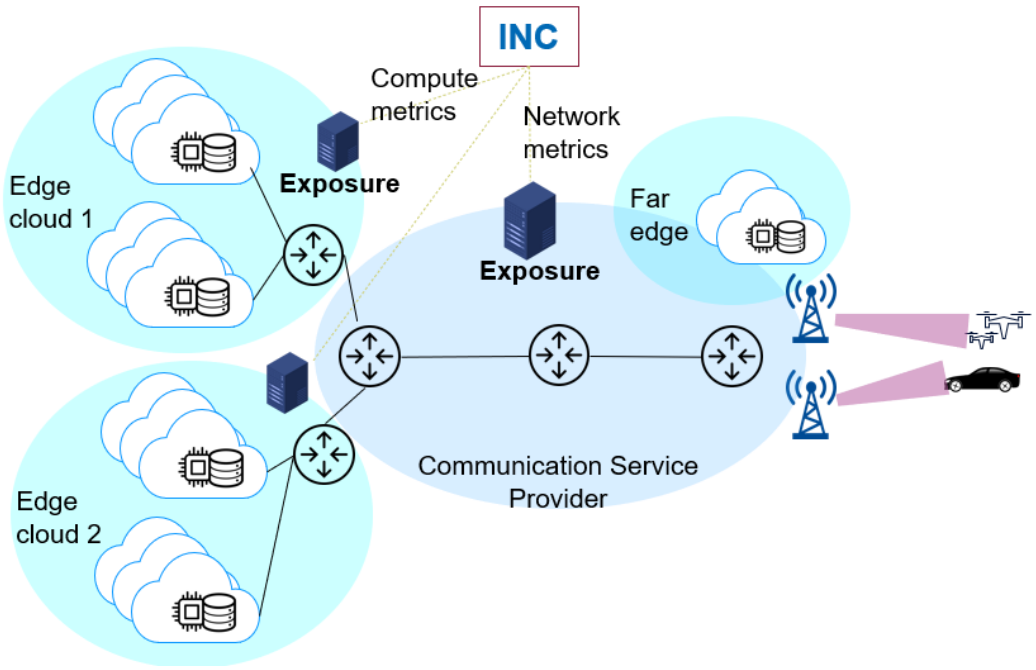


Figure 6.5: Integration of Network and Compute server, collecting network and compute metrics to decide optimised placement of sensing/other BC service consumer placement

When a device, acting as an offloading node, decides to offload a computation, it will have to discover and select the candidate computation nodes, capable of performing the requested computation while satisfying specific KPIs. Each computational node should estimate the task(s) execution complexity and resources demand (i.e., computation and storage) based on a common characterization of the offloaded

compute workloads (i.e., compute tasks) determined by the requirements comprising traffic class (e.g., one time vs multi-iteration, one node vs multi-node, etc.), computation complexity, communication requirements, precision requirement (e.g., quantization level of the compute data and operations), quality of compute service classes (latency-sensitive, precision-sensitive, etc.).

The trade-off between network metrics and compute metrics would call for a new approach that enables the Integration of Network and Compute domains to perform coordinated optimization.

6.5 ISAC APPLICATION FOR V2X USE CASES

The Euro NCAP announced that starting in 2024, new cars in Europe must include Vehicle-to-Everything (V2X) connectivity for a five-star safety rating. In the U.S., the Department of Transportation introduced a National V2X Deployment Plan in August 2024 to enhance road safety, urging collaboration across industries and government. Despite European efforts to deploy V2X, including contributions from major manufacturers, challenges like incompatible technologies and limited consumer awareness have delayed progress. These challenges are now being addressed, paving the way for broader V2X deployment.

The next step is an ambition to improve connectivity and safety for Fully Autonomous Vehicles (FAVs), especially in city intersections, with specific goals of emergency vehicle route prioritization and optimized city-wide traffic flow, real-time data sharing among road users and infrastructure, dynamic traffic management, and the development of a high-capacity, ultra-low latency 6G-V2X network are required. Key components include advanced LiDARs and RaDARs for object detection, Optical Wireless Communication (OWC) for high-speed data exchange, and AI-based 3D mapping for real-time environmental awareness.

Challenges for implementing technologies like LTE-V2X and OWC include achieving ultra-low latency in dense environments, maintaining signal clarity in adverse weather, and integrating diverse systems for seamless communication. Efficient computational resources and data management will be crucial for real-time collective perception among vehicles and infrastructure.

Ultimately, the 6G architecture envisions a robust, AI-driven network supporting V2X in dense urban areas, enabling collision-free navigation, improved traffic flow, and

enhanced safety for vehicles, Vulnerable Road Users (VRUs), and emergency services across interconnected city areas.

Specifically, the focus on the support the next generation of (fully) autonomous driving aims at the significant improvement of the road safety (and connectivity) for all road participants, especially at busy city intersections. To achieve this, real-time, ultra-high speed and low latency with ultra-high reliability communication and data exchange between the FAVs, the VRUs and the road/network infrastructure is mandatory. On top of the above, the system shall allow for seamless coordination in critical scenarios like emergency vehicle prioritization and large-scale vehicular traffic flow optimization across urban areas.

The proposed high-level network architecture, depicted in Figure 6.6, ensures that all road participants (FAVs L0-L5, VRUs, etc.) can communicate between each other and with the road/mobile network infrastructure via either/both LTE/NR-V2X technologies (via Uu and/or PC5) and/or via OWC/VLC while the stringent QoS requirements posed by the safety related applications -based on the collective perception concept- can be fulfilled by:

- The deployment of the OWC/VLC APs - to satisfy throughput and latency-related requirements.
- High-speed fronthaul/backhaul, photonic-based network which will ensure the resilient, high-capacity, low-latency communication across the system - meeting throughput, latency and reliability-related demands.
- The co-deployment of LTE-V2X 5G/NR-V2X base stations/RUs and OWC/VLC APs and the capability of the FAVs' OBUs to transmit data concurrently via both radio technologies - throughput, latency and reliability-related QoS requirements are to be addressed.
- Deployment of AI-based data processing infrastructure / data centres (incl. compute nodes, application servers) capable of processing huge volumes of sensor data in real-time, close to the road participants (see extreme/far-edge DCs)- ultra-low latency required for critical tasks such as those related to JCAS and SLAM, accurate, real-time 3D maps of the surrounding environment, decision-making for FAVs, real-time updates for VRUs will be achieved.

On top of the above, regional DCs and Central Cloud(s) are deployed at higher-levels for data aggregation and data processing and supporting less critical applications such

as traffic flow optimization at wider-city areas, but these are also utilized for traffic management, large-scale data analytics, long-term storage, etc. The central cloud, in addition to the 4G/5G/6G core NFs, hosts the TMS, which, by gathering and processing data from various sources, can adjust the traffic lights and signals, aiming at congestion minimization and vehicular traffic flow improvement, along with emergency vehicle prioritization related functionalities. Finally, the AI-based network management and orchestration functions will ensure the optimal allocation of the optical network resources in real-time (and/or proactively) based on the current and predicted traffic demands to guarantee the applications' QoS end-to-end, but also to improve the overall network efficiency.

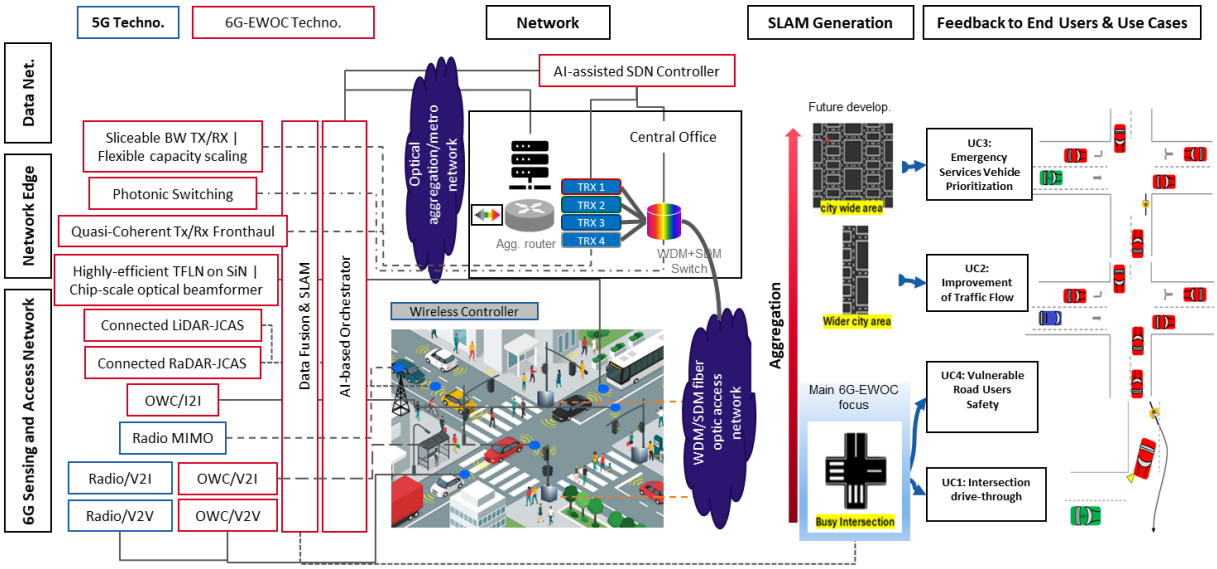


Figure 6.6: ISAC-based V2X network architecture

In such a deployment plan, fibre is essential as a high-bandwidth, low-latency medium connecting Roadside Units (RSUs) to cloud processing platforms, enabling robust V2X communication. The system relies on flexible optical networks, which support long-distance, low-latency data transmission through optical amplifiers across a wide spectral band. However, fibre access is limited near RSUs, so Free-Space Optics (FSO) is used to extend fibre-like connectivity to the last metre. The architecture thus benefits from FSO's ability to provide high-speed, fibre-equivalent wireless communication for challenging network segments and enhances V2X services.

To bridge network gaps, a Fi-Wi-Fi bridge is used, allowing single-mode transmission through transparent FSO channels, maintaining low latency, energy efficiency, and compatibility with wideband optical transmission schemes. This connectivity, digitizing

radio signals over fibre or air, is critical for high-demand applications such as 5G and 6G services at remote radio heads.

In areas requiring longer-range communication, FSO is enhanced with MIMO or hybrid RF links to ensure resilience against adverse weather. Short-range FSO applications connect RSUs in dense urban areas, supporting high data loads without interfering with RF channels. Optical Wireless Communication (OWC) complements RF by handling excess data traffic at intersections, where high user density can cause RF congestion.

Fronthaul connectivity relies on quasi-coherent transceivers for high-speed, low-latency fibre links, enabling real-time collective perception and SLAM processing in autonomous vehicles. Initial deployment includes 50 Gb/s transceivers, with future upgrades to 100 Gb/s to handle increased sensor data. These transceivers enable modular Bandwidth Variable Transceivers (BVTs) at central offices, delivering variable capacities of 2 Tb/s to 380 Tb/s across cell sites, adapting to wireless network demands and future-proofing for data-heavy autonomous driving applications.

6.6 MULTIX -ADVANCING ISAC THROUGH MULTI-TECHNOLOGY, MULTI-SENSOR FUSION, MULTI-BAND AND MULTI-STATIC PERCEPTION

The vision of 6G is to integrate sensing with communication in a single system. Sensing and native Artificial Intelligence (AI) operations are the two key aspects to build the connected intelligence in 6G. For sensing, the use of multiband (sub-6 GHz band and higher frequency bands – from millimetre wave (mmWave) up to THz), wider bandwidth, and massive antenna arrays will enable high accuracy and high-resolution sensing, which can help implement the ISAC in a single system for mutual benefit. To enable such a vision, the ability to utilize a range of technologies to perform sensing and the ability to process the sensing data into meaningful sensing results is a key challenge for 6G systems. Furthermore, how to manage such heterogeneity and fully integrate various sensing sources and ISAC technologies throughout the RAN up to the User Equipment (UE) across different layers of the RAN stack, and thus how to explore the full potential of each sensor, each node, each band, each technology as well as their combination to achieve the maximum spectrum and system efficiency, energy

efficiency and resource usage across the entire RAN system is a big question, but also a huge opportunity.

To address the outlined challenges, a redesign of the 3GPP RAN system is required that can aggregate a set of diverse design concepts in a seamless way (defining the so-called MultiX concept) to create an integrated multi-sensor, multi-band, multi-static, and multi-technology paradigm to enable multi-sensorial perception for future 6G sensing applications. Based on a system architecture built on the upcoming 3GPP Release 20 (R20) and O-RAN specifications, architectural design enhancements are required to support aggregation of perception data from multi-technologies and multi-sensors, their processing and exposure to third parties in a secure, privacy-preserving and trustworthy way. Furthermore, support for vertical and horizontal handovers and network selection procedures considering perception requirements are envisaged while exploring the use of AI/ML for novel connectivity options to enhance the perception capabilities of the network.

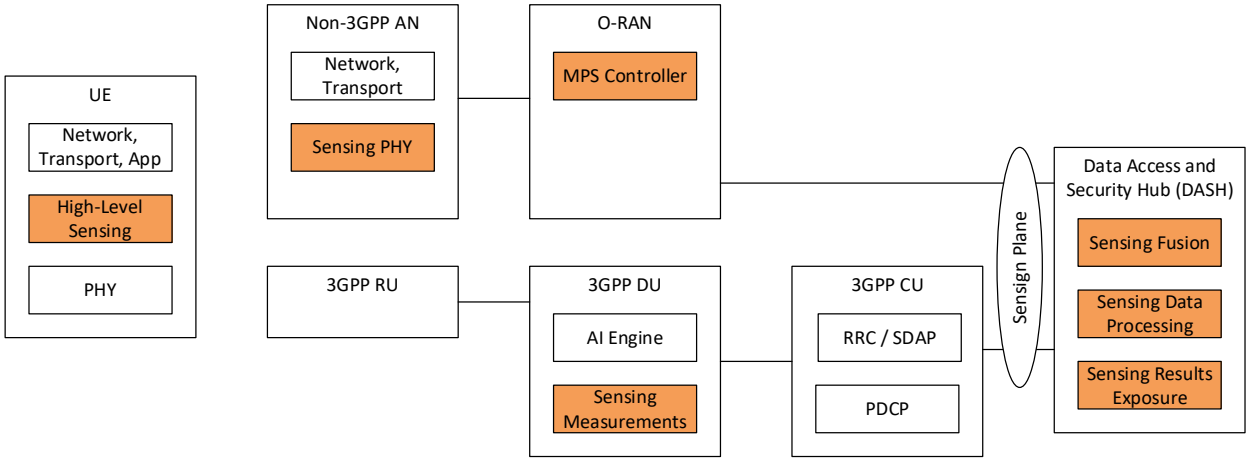


Figure 6.7: High-level architecture of 6G system to enable multi-perception sensing

Supporting the above functionalities, data fabric and pipeline solutions are to be developed to integrate mechanisms and artifacts to enable advanced data collection via novel paradigms (e.g., ProSE and Integrated Access and Backhaul (IAB)), processing, distribution, management (e.g., data curation and enrichment), and exposure for third-party access applications, ensuring privacy, security, and trustworthiness. A Multi-Perception System (MPS) is envisaged for providing perception to the network across different access technologies. The architectural components of such an MPS is illustrated in Figure 6.7 (MPS components are given in orange). Current sensing approaches will be improved in the MPS-enhanced 6G system with multi-band, OTFS-

based, programmable environments and disaggregated architectures, and integration of multi-technology RATs in the MPS, considering multi-static deployments, where synchronization is key. AI is considered for the distributed processing of the raw data into actionable information, including natively energy-efficient AI design of receiver architectures that jointly leverage signal processing domain knowledge and brain-inspired, event-based operation. Finally, all sensing data is fused, processed and exposed in the Data Access and Security Hub (DASH).

6.7 REFERENCES

[3GPP24-22.837] V19.4.0 (2024-06) Feasibility Study on Integrated Sensing and Communication (Release 19).

[6GSENSESD2.2] 6G-SENSES Deliverable D2.2, "System architecture and preliminary evaluations", March 2025.

[AGT25] Markos Anastasopoulos, Jesús Gutiérrez, Anna Tzanakaki, "Optical Transport Network Optimization Supporting Integrated Sensing and Communication Services", OFC 2025.

[Y+23] Yao, Shanliang et al. "Exploring Radar Data Representations in Autonomous Driving: A Comprehensive Review." (2023).

[CAM+24] John Cosmas, Kareem Ali, Hussein Malki, Adam Kapovits, Israel Koffman, Benjamin Azoulay, Clément Lartigue, Emmanuel Plascencia, Bastien Béchadergue, Barthélemy Cagneau, Luc Chassagne, Anastasius Gavras "Optical 6G Cell Free Networks for High Performance Communications and Sensing in Industry 4.0" IEEE Future Networks World Forum 2024, 15-17 Oct 2024, Dubai UAE

[CSA+21] E. Calvanese Strinati, G. C. Alexandropoulos, H. Wymeersch, B. Denis, V. Sciancalepore, R. D'Errico, A. Clemente, D.-T. Phan-Huy, E. De Carvalho, and P. Popovski, "Reconfigurable, intelligent, and sustainable wireless environments for 6G smart connectivity," IEEE Communications Magazine, vol. 59, no. 10, pp. 99–105, Oct. 2021.

[ADR+22] A. Albanese, F. Devoti, V. Sciancalepore, M. Di Renzo and X. Costa-Pérez, "MARISA: A Selfconfiguring Metasurfaces Absorption and Reflection Solution Towards 6G," IEEE INFOCOM 2022 - IEEE Conference on Computer Communications, London, United Kingdom, 2022, pp. 250-259, doi: 10.1109/INFOCOM48880.2022.9796976

[HEXAD3.3] Hexa-X Deliverable D3.3 "Initial analysis of architectural enablers and framework" https://hexa-x-ii.eu/wp-content/uploads/2024/04/Hexa-X-II_D3.3_v1.0.pdf, April 30, 2024

[OPTI24-D21] OPTI-6G Deliverable D2.1. Cosmas, J., Koffman, I., Lartigue, C., Béchardergue, B., Kapovits, A., & Bienvenu, A. (2024). D2.1 OWC Cell Free Network Use Cases and User, Functional and Technical Requirements (1.0). Zenodo. https://doi.org/10.5281/zenodo.14654936

[OPTI24-D41] OPTI-6G Deliverable D4.1. Koffman, I., Globen, B., Cosmas, J., Mahbas, A., Ali, K., Lartigue, C., & Béchadergue, B. (2025). D4.1: OWC Localization Sensing System Design (1.0). Zenodo. https://doi.org/10.5281/zenodo.14851632

[C+25] E. C. Strinati et al., "Toward Distributed and Intelligent Integrated Sensing and Communications for 6G Networks," in IEEE Wireless Communications, vol. 32, no. 1, pp. 60-67, February 2025, doi: 10.1109/MWC.001.2400056.

7 SECURITY, RESILIENCY, PRIVACY AND TRUSTWORTHINESS

7.1 SECURITY CONCERNS INTRODUCED BY 6G

The advent of 6G incorporates new paradigms especially focused on a greater capillarization of services. This new offering encompasses mass sensorisation, ubiquitous connectivity provisioning with new types of connectivity, native AI integration, network exposure and programmability to enable interaction between various functions, and new forms of businesses focused on coordination among different providers. As with all other generations of mobile networks, these new expansions bring multiple ambitions as well as threats and risks. Vulnerabilities, as yet unaddressed in its predecessor, are accentuated by the magnitude of data that will be collected from all parts of the infrastructure. In this respect, security becomes an even more critical issue. Trustworthiness is one of the main properties, on which security efforts are being focused in this early phase of 6G. Establishing a solid foundation on which to build secure and resilient solutions serves to ensure that properties such as privacy or software genuineness are stable and reliable. Besides, security solutions must cover complex aspects more than ever, and vulnerabilities can impact the system in such ways that endanger the whole system safety, for instance by poisoning data, leading to non-intended learning. To this aim, empowering resiliency has become principal to face security challenges by providing autonomous reaction to threats, and even leveraging AI skins to prevent breaches in advance. In security concerns, privacy has been one of the principal topics of 5G, whose importance increases when multistakeholder and ubiquitous connectivity is added into the scene.

Correct addressing of these considerations starts with the identification of the threat vector and weaknesses newly incorporated to the landscape. Within the whole 6G security picture, enablers and technologies are studied to establish synergies and build architectural blocks that supports the main 6G functions by adding security at the design phase.

To this end, we firstly introduce the main threats of the groups identified for the IMT's six usage scenarios (i) extension from IMT-2020 (5G) including immersive communication, massive communication, and hyper reliable & low latency

communication; (ii) ubiquitous connectivity; (ii) AI and Communication; (iv) and Integrated Sensing and Communication:

IMT-2020 extension

The extension of IMT-2020 to meet the needs of 6G has raised a number of security issues. Rigidly structured and poorly bonded networks find it difficult to adjust to the dynamicity of the environment, especially in terms of security. The adoption of post-quantum cryptography (PQC), which is still under study, may lead to signalling overload and new, as yet unknown attack vectors, leaving the network vulnerable to unforeseen threats. In addition, the increasing reliance on cloud-native architectures, virtualisation and multi-site orchestration introduce complexity and vulnerabilities such as insufficient isolation, miss-configured systems and unsecured APIs - the perfect testing scenario for an attacker to attempt to disrupt services or gain unauthorised access.

Ubiquitous Connectivity

A key component of 6G networks is ubiquitous connectivity, where devices will be able to connect from anywhere, inevitably widening the attack surface and increasing the likelihood of exploitation. While facilitating seamless communication, the constant generation and transmission of huge volumes of sensitive data raises privacy concerns. Since any device or node can be an attacker's access point, it is difficult to effectively secure this vast distributed environment.

Integrated Sensing and Communication

ISAC combines communication and detection capabilities, but doing so carries significant risks. Sensitive information such as location and identity could be exposed due to inadequate data fusion procedures, with a lack of obfuscation. Moreover, as hackers can exploit unsecured data flows, the huge volumes of data produced by IoT and ubiquitous networks pose privacy risks. A security breach at this level can compromise further processes in the system, so there is a need to maintain the security and accuracy of the merged data.

Artificial Intelligence

Risks associated with AI integration in 6G networks include adversarial AI, where attackers alter AI models to generate unreliable results and lead decision processes to intentional failures. AI exploitation, where malicious actors use AI to bypass detection measures, allows them to make much more sophisticated attacks. In addition, it is

important to note the problem of data poisoning, which occurs when compromised training datasets compromise network security and inadequately train AI models or trick analysis engines into believing that unrealistic situations are occurring.

Network Exposure & Programmability

6G networks are more susceptible to attacks, based on distributed denial of service (DDoS) or man-in-the-middle (MITM) attacks and data breaches, due to their programmability and exposure to external environments, as this increases the distributed nature by adding new attack points accessible to all. As networks become more open and programmable, attackers can use flaws in software-defined NFs, service interfaces and APIs to intercept private information, compromising the integrity of communications or simplifying processes that overload network or compute resources.

New Business models

Finally, it is not included in the groups mentioned but it is important to highlight the complex trust management issues arising between operators and stakeholders as a result of emerging business models in 6G networks, particularly in shared infrastructure scenarios. Possible risks associated with the implementation of cloud-native architectures and SLA management include impersonation, unauthorized access, and in general terms, insufficient authentication, authorization, and accounting (AAA). Smart pricing techniques could also be exploited to manipulate invoicing or service access, underlining the necessity for comprehensive security measures to assure trust and reliability within the ecosystem.

7.2 TRUSTWORTHINESS AS THE PILLAR TO BUILD SAFE, RESILIENT & RELIABLE SECURITY AND PRIVACY SOLUTIONS

The 6G vision of making 5G an open, multi-operator, user-centric network extends one of the main security concerns: trust. The incorporation of collaboration with other operators and third parties to offer advanced services and reuse infrastructures in a transparent way in what is known as cloud-continuum, as well as the softwarization of IT-based infrastructure, makes trust a central element that should be ensured throughout the operations cycle.

The 5G security approach follows a centralized network architecture, and trusted connections between network parts are created at the protocol level, rather than depending on device and network behaviour. In the envisioned 6G ecosystem, trusted connections are critical for all parties involved, extending security and privacy to a more inclusive framework such as trustworthiness, which should be assured as a native feature.

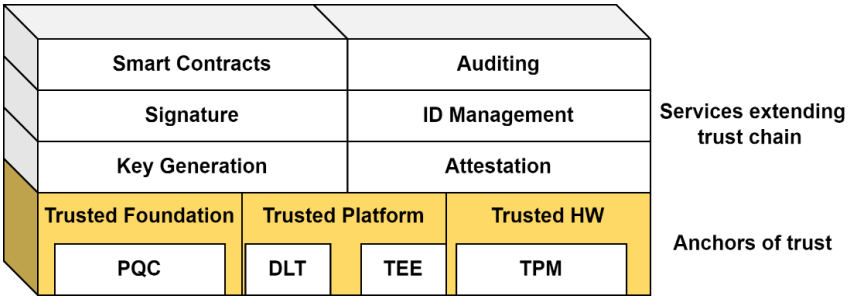


Figure 7.1: Anchors of trusts & services built upon them extend trust chain

The most significant paradigm adjustments in the envisioned 6G system are the shift from a security-only focus to a broader scope of native trustworthiness, clarifying that the term "trustworthiness" refers to a holistic approach, building safety, security, privacy, resilience and reliability upon the anchors of trust. Trustworthiness then refers to the solutions coming from the trusted foundation, platforms and hardware, extending the chain of trust from the basis towards the user-centric perspective.

Trustworthiness challenge must come with a realistic solution, recognizing all security measures (i.e. safety, security, privacy, resilience and reliability) come at a cost in terms of usability, agility, or swiftness. As a result, the envisioned trustworthiness framework should provide a balance between the various security measures by dealing with a security-by-design approach as well as a wide range of themes such as the trust model and the application of new cognitive coordination technologies (e.g., Intent-based trustworthiness based on AI and ML techniques).

7.2.1 TRUSTWORTHINESS AND LEVEL OF TRUST RELATION

IMT 2030 promotes trustworthiness as a new attribute for the 6G vision, and thus numerous standard groups [ISO16], including 3GPP, ETSI, and IEEE, are working on trustworthiness issues. Meanwhile, the world's main communications suppliers explicitly underline the importance of 6G trustworthiness in their 6G projects, proposals, and white papers. In addition, several scholars produced technical papers on the definition, generation, protection, and optimization of trustworthiness. All of these

suggests that trustworthiness will become an essential characteristic in 6G, but still, the usage of trustworthiness and trust terms is confusing for many people. Despite the discrete meaning and scope of trustworthiness, it can be misused as trust. For this reason, it is necessary to start by clearly defining the meanings of and the relation between trustworthiness and trust.

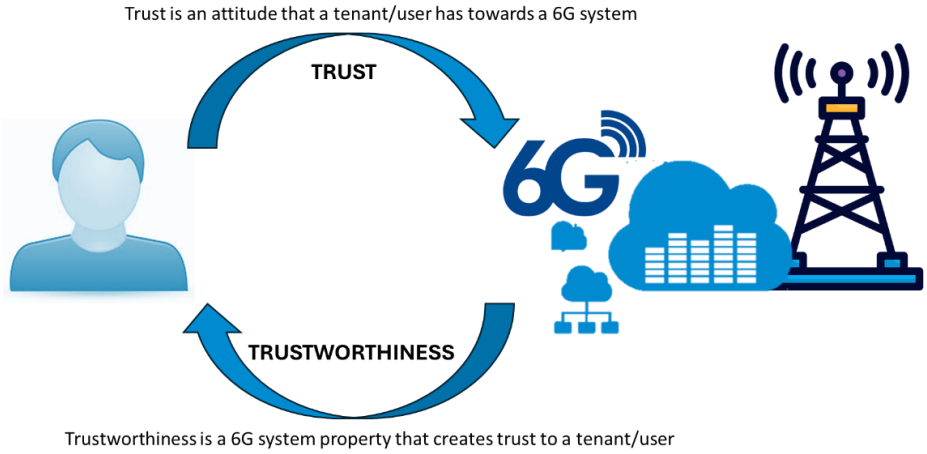


Figure 7.2: Trust and Trustworthiness definitions

Trust is an attitude that a tenant has towards a 6G system. In contrast, trustworthiness is a system property that creates trust within the 6G tenant/user towards the system. Thus, a user/tenant trusts (or requires trusting the system at a specific level) a 6G system because the 6G system is trustworthy. In other words, the trustworthiness of a 6G system contributes to building the trust level of the tenant/user of the specific system. Therefore, the more trustworthy the 6G system is, the higher the trust level of the tenant/user will be [K+21].

7.2.2 6G PROPERTIES IN TRUSTWORTHINESS

User-centric and AI-Assisted Coordination

Each tenant/user having different requirements in terms of trust level from a 6G system creates the need for the 6G system to be capable of being adapted to the specific needs (i.e., trustworthiness level) that reflect the level of trust and requirements of these tenants/users. Therefore, the 6G system should not be only trustworthy in a static way, but it should become user-centric, dynamically adapting the trustworthiness level to the requirements of each tenant. The user-centric approach in the 6G system allows for dedicated network services to be provided at the user-granularity by configuring the 6G system per single user needs (i.e., the core NFs) [YAZ+23].

In order to achieve user-centric adaptability, intent-based driven trustworthiness is required for dynamic configuration of 6G systems. Both paradigms are necessary for the introduction of the trustworthy 6G concept to be driven by AI/ML-assisted coordination component, which will act as an intent-handling function that comprehends sophisticated and abstract trust intent semantics and calculates the ideal goal state to organize activities for transitioning the 6G system into this trustworthy state.

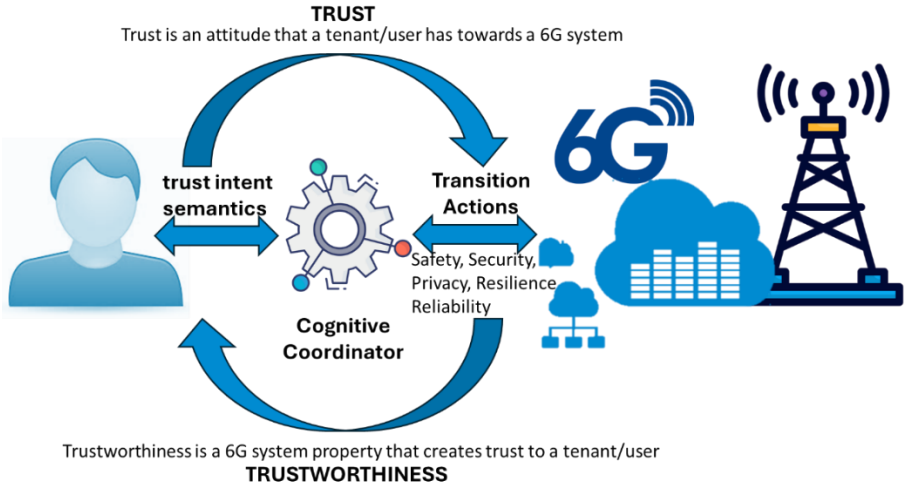


Figure 7.3: User-centric and AI-assisted coordination of 6G trustworthiness

The AI-assisted coordinator should perform the process of mapping the trust-intent semantics received from the tenant/user into transition actions of the 6G system via configurations into the trustworthy dimensions of the 6G system, (i.e., into the safety, security, privacy, resilience, and reliability domains). Given that the autonomous mapping between intents and trustworthiness dimensions is a technical application of cognition (since it is designed to perform the operational tasks of understanding by experiencing and monitoring), the envisioned AI-assisted coordinator is named Cognitive Coordinator. Besides, the generation of trust intents shall be allowed to be autonomous. According to the requested service, a minimum granted trustworthiness should be given with regard to the criticality of data, service, and networking aspects such as multistakeholder cooperation. In granting that trustworthiness, the system is protected against malicious users that could try to intentionally degrade the level of trust of a given service.

Explainability

The use of trustworthiness measures by the system could impact users' perceived level of trust, especially if users cannot understand trustworthy actions taken. In this context, explainability is often viewed as an effective way to build trust among stakeholders. If users have a better understanding of the process by which the system generates its outputs and the explanation provided for a particular result aligns with their preconceptions of what constitutes a proper decision, then the level of trust of the system has been improved. The literature does, in fact, frequently link explainability to trust [CBS21], [Pie11], and many researchers—at least tacitly—assume that explainability and trust are strongly related [GRR+19], [RSS16]. This relationship is known as the Explainability-Trust-Hypothesis, which states that "explainability is a suitable means for facilitating trust in a stakeholder" [LOS+21].

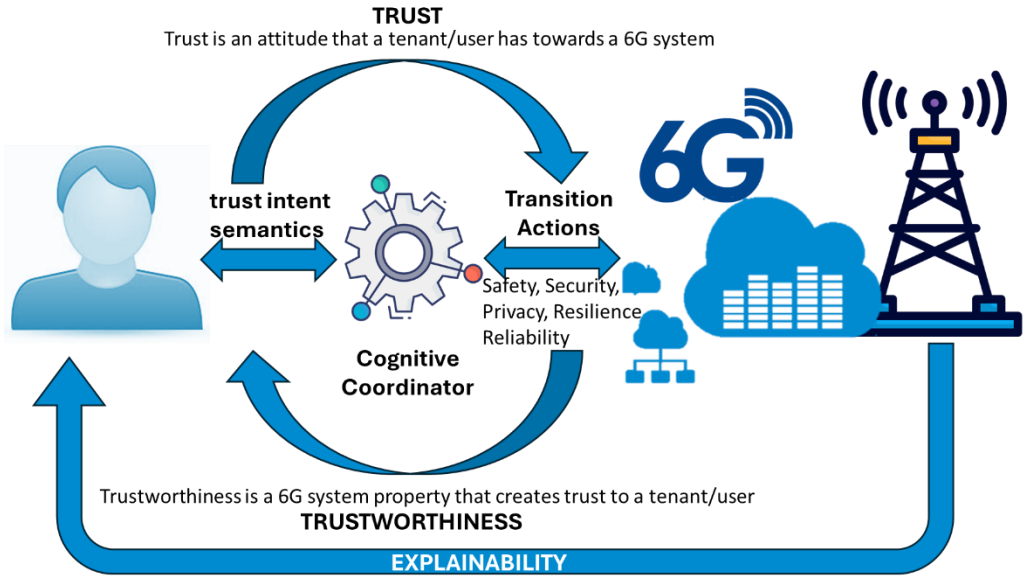


Figure 7.4: Explainability as an additional means for improving the Level of Trust in 6G networks

Among the various explainability tools, there is a tendency for modern systems to select eXplainable AI (XAI) [LOS+21] to complement the operation of the Cognitive Coordinator and contribute to the improvement of the tenant's/user's Level of Trust over the Trustworthy 6G System. As evidenced by the "right to explanation" outlined in the General Data Protection Regulation (GDPR) and the European Commission's (EC) Technical Study on "Ethics recommendations for trustworthy AI" [KP20], trustworthiness has become crucial for both users and governmental organizations. They claim that explainability is a crucial element for trustworthiness. As a result, XAI or an AI "that creates information or reasoning to make its working obvious or easy to

understand" is receiving more and more interest from both industry and academia. In this context, two strategies for achieving explainability can be identified: The adoption of post-hoc explainability techniques (i.e., the "explaining black-box" strategy) and the design of inherently interpretable models (i.e., "transparent box design" strategy). These approaches allow understanding the model behaviour and can be integrated in the model training or applied as post-hoc approaches after ML training.

7.3 CHALLENGES

Adaptive and unattended cybersecurity

Regular 5G orchestration frameworks do not contemplate security as a native element in service orchestration. In addition, very specific solutions that only work under certain circumstances make it difficult to use them in other scenarios, lacking adaptability and automation with heterogeneous infrastructure and conditions. A clear integration of security modules into a ZSM architecture is mandatory to be able to respond to further 6G security challenges.

Following a modular approach, ZSM security framework could be extended to incorporate new security capabilities and assets, propelling the flexibility to adapt security to novel scenarios and technologies with an E2E perspective as well as having several alternatives to responding to and evaluating threats.

ZSM as a service: AI driven intent management

5G lacks coordination between third parties, also third parties' domains and networks that would compose a 6G Network Service could potentially lack self-driving capabilities (e.g. self-healing). ZSM as a service is a challenging task that aims to provide such autonomous AI based intent management on demand. This ambitious service involves some other challenges, such as, the common representation of the system's exposure functions under an expandable and well-defined format. Such a system model needs to contain: services running, resources available, lower-level orchestrators... among others. In this context, to autonomously manage the security of dynamic topologies is a daunting challenge.

Security SLA (SSLA) negotiation

Normally, Service Level Agreements (SLAs) have been widely used to represent contract between customer and providers, but 5G lacks on a common format to model interactions between different stakeholders and customers. Optimized resource sharing

and network service compositions won't be possible until communication is automated through a common format. Defining the standardized model to cover these aspects present some real hurdles due to its complexity, leading into difficulties for automating the negotiation between the participants as well as the definition of a baseline to define the expected system and service behaviour. Furthermore, regarding security, the definition of a common abstraction model used to define the security aspects of a security service deployment and management forming the Security and SLAs (SSLAs) is a challenging task.

Resilience and Service continuity

Even though 5G solution proposes dynamic network service deployment, they are normally realized by statics templates. Dynamic NS composition tailored to SSLA/SLA, possess a significant challenge specially on how to include multiple domains and grant service duplicity or alternatives NS to build resilience service delivery.

Application of Quantum Key

Quantum computing is pushing forwards, traditional forms of encryption will become obsolete, new wave of key generation and encryption resistant to quantum computing styles are needed, how to integrate the novel PQC in constrained devices. Ensuring seamless and secured transition to quantum-resistant system introduce unique challenges. Performance overhead of PQC algorithms, especially lattice-based and hash-based, is one major drawback. Larger key sizes and therefore increased needs on computational power can put resource-constrained devices like smartcards and IoT gadgets under stress. This increased demand complicates the deployment of PQC in environments where high efficiency and low latency are required.

Besides performance-related concerns, migrating to PQC requires planning in ensuring interoperability between current systems and novel quantum-resistance algorithms. Coexistence during this phase will happen, hybrid cryptographic schemes are very often adopted, which embed traditional methods along with the post-quantum ones. This adds complexity for uninterrupted and secure management operations.

The security of PQC algorithms relies on different mathematical foundations than classical cryptography. Building confidence in these novel algorithms is an ongoing challenge. Continuous research and cryptanalysis are necessary to validate these algorithms against evolving quantum threats [KP20].

The successful adoption of PQC will, in the end, be a function of how industries and government agencies can effectively incorporate new standards such as FIPS 203, 204, and 205 into their existing cryptographic infrastructures. This would require updating of legacy systems, ensuring compliance with evolving protocols, and achieving coordination across diverse stakeholders. The interaction of these challenges underlines the complexity of implementing PQC and underlines comprehensive strategies for a secure and efficient transition to quantum-safe technologies.

DLT-based trustworthiness

5G has made significant strides in data speed and service management, but it falls short in addressing security concerns produced by next-generation connectivity. With the envisioned heterogeneity of devices and new shapes of service delivery, perimetral-based 5G security do not cover 6G attack surface. Zero Trust Architecture and its integration with DLT-based solutions are a challenging task to ensure trustworthiness of the system. The strict access policies provided by ZTA as well as the verifications needed possess a challenge in energy consumption and scalability. In this sense, the need for trustworthiness in multi-domain environments becomes a key challenge, specifically where Multi-Agent Systems take place, enabling dedicated management but generating the challenging task of enforcing trust between the agents. Integrating such attestation processes in distributed environments with added difficulty of secured key distribution. The integration of DLT to solve these challenges also come with scalability problems, especially to support geographically distributed agents in these multi-domain environments [DES23].

Continuum of Trust

Since the digitalisation is impacting an ever-increasing part of daily life, connectivity and supporting network have become a daily necessity and support critical activities and have become a vital asset. For instance, energy, transport, finance and public administration are recognised by NIS2 [NIS24] legislative framework as essential entities having to comply with the constraint of resilience, while heavily relying on telecommunications. Given the importance of the service they deliver, network infrastructure must be accepted and trusted by their users.

However, as the 6th generation of mobile networks is anticipated arising around 2030 [Eri24], several fundamental paradigm changes is foreseen hampering the trust in the resources. Specifically,

The openness of the network, due to the (i) API-ification of the network capabilities to third-party services and (ii) the increased involvement of diversified solution providers (including open-source communities). This situation exposes the network to the composition and interaction with software having a different degree of resilience and robustness.

The involvement of multiple providers, providing and supervising their own network segment accounting for access network, transport technologies to be dynamically integrated to ensure constant and optimal connectivity (corresponding to the network of the network paradigm). This understands the involvement of actors enforcing different security management, which requires different trust model to cooperate.

The cloudification of the network services and their continuous desegregation throughout the cloud continuum will increase the dynamism of applications deployment and the providers' resource usage. This dynamism will make it infeasible to establish a clear boundary for network service and therefore outwitting any attempts for a perimeter-based security approach.

Cross domain high-scale monitoring & Analysis

Security and privacy enhancement in future 6G networks can be a quite challenging and demanding task, due to the vast number of potential threats and attacks and their diverse nature compared to 5G networks (indeed, a larger attack surface is expected in 6G networks). In the same context, the interconnection of a vast number of devices and the support of heterogeneous deployments (exploiting the cloud continuum paradigm), which are both key concepts of the 6G era, might leverage security and privacy concerns, since not all devices will have the capability to execute advanced security protocols due to their hardware constrained nature. In this context, artificial intelligence/machine learning (AI/ML) approaches that collect a vast amount of data from the network to train models that can represent input/output pairs with minimum performance loss, can leverage security and privacy mitigation via the extraction of abnormal data patterns and the enforcement of the appropriate actions. Compared to conventional non-ML detection techniques, ML-based misbehaviour detection provides both a higher detection accuracy against unknown zero-day attacks, as well as a reduced false detection rate.

The deployment of ML approaches for threat detection and mitigation in the 6G landscape is a quite challenging procedure, dictated by various key driving factors: i)

computational efficiency of the deployed approaches, ii) identification of multiple and even correlated threats and attacks, iii) continuous refinement of the ML approaches and knowledge distillation, as well as iv) creation of multiple network intents per case for network recovery [GNT+24]

Extreme virtualization and softwarization

The trustworthy 6G moves beyond the current NF-centric core network towards a user-centric evolution of the B5G/6G system over the recently researched edge-cloud-continuum, which is expected to be the primary option as infrastructure for deploying the softwarised components of a distributed 6G network. Therefore, for 6G to become the human-centric system of systems requires significant architectural redesign based on the user-centric (i.e., per-user perspective), given that the network intrinsically handles the state of each UE or user. A user-centric design is specifically capable of providing to each user a complete instance of a personalised 6G system through a user-specific core-network synthesis, supporting for example personal data management, policy control, session control, and mobility management per-user. These customized nodes are the so-called User Service Nodes (USNs), while regular centralized core NFs are defined as Network Service Nodes (NSN).

It is in the transformation of NF-centric to user-centric architectures where the paradigm shift in establishing, maintaining, and scaling network trustworthiness occurs for 6G. Such architectural evolution introduces several challenges concerning trustworthiness assurance in a personalized, user-centric network environment.

The challenges in designing a robust trust model that accounts for the dynamic and heterogeneous nature of USNs lie in how each USN is responsible for the implementation of user-specific services, policies, and mobility management, often tailored in real time to the evolving trustworthiness levels and requirements of the user. How would the system ensure mutual trust between the user and the network, considering when the network becomes distributed, personalized, and softwarized?

For example, challenges like personal data sovereignty, secure multi-tenancy, and real-time computation of trust for individual user sessions have to be solved without sacrificing scalability. Also, embedding trust functions like safety, resilience, and reliability across the NSN-USN continuum introduces potential trade-offs between personalization, performance, and security. It will be exciting and also critical to create a scalable and adaptive trust framework for USNs that is capable of guaranteeing

integrity and authenticity while considering user-controlled data ownership. How could this framework balance user control, adaptability of the network, and intrinsic trust in this highly personalized environment?

Security of software, virtualized environments and hardware-based platforms

Building trustworthiness on 6G relies on the use of anchor of trust and applications that extend them to build a chain of trust towards higher layers. Guaranteeing that software deployed is trustworthy is a challenge in which the integration between trusted execution environments (TEEs) and attestation form part of the addressing approach. Still, interoperability, scalability and lifecycle management are challenging tasks related to the integration of both. As TEEs are growing attention for 6G networks, finding practical solutions is a key element.

Scalability is one of the main concerns. Attesting thousands of applications and TEE instances in real-time may result in a system overloading. As well as, properly ensuring isolation in multi-tenant environments, where different users share the same hardware functions. In this context, preventing malicious actors from exploiting shared resources and vulnerable separation of processes and data is of the utmost importance.

Life-cycle management is equally important. Initializing, updating, and decommissioning TEEs and trusted applications need to be handled prioritizing trustworthiness, protecting sensitive data from leaks. This becomes intricate specially during workloads migration, where other security vulnerabilities could emerge. In addition to this challenge, the diversity of hardware platforms, including Intel SGX, AMD SEV, and Arm TrustZone, requires a unified approach to ensure interoperability and security across different implementations.

Privacy preserving approaches

The research developments towards the vision of 6G networks represent a substantial advancement in communication technology, for significant improvements in connectivity, speed, and innovation. However, this progression also introduces security and privacy challenges. As 6G integrates an expansive network of devices and services, protecting sensitive information becomes paramount. Traditional security frameworks are inadequate to address the complex threats and privacy risks inherent in 6G ecosystems [NLC+21].

At the same time, Privacy is considered a key pillar in EU research and development activities towards 6G, as privacy enablement is considered a top societal aspect in the

EU 6G vision [BCG+24]. 6G is envisaged to comprise a decentralized, zero-trust, globally connected continuum of heterogeneous environments involving several actors across the service chain (core/edge/RAN infrastructure providers, service providers). In such a pluralistic environment, privacy is pivotal, not only for the end users but also for all involved stakeholders; and it needs to be considered as a critical requirement in all technologies of the network stack, including security mechanisms.

In other words, the challenge for security enablers in future networks is, on the one hand, to address the significantly widened 6G threat landscape, while, on the other hand, to preserve the privacy of all actors in the 6G chain. Intrusive security cannot be anymore considered acceptable.

Identity management

The forthcoming 6G networks are expected to be accompanied by extensive collaboration between stakeholders from different domains. From large infrastructure providers to specialised microservices. This cooperation will enable transparent services to be offered to a large number of users of all types, irrespective of the subscriber's home operator. This development brings many advantages, such as offering advanced, ubiquitous, resource- and price-optimised services, but it also introduces major security challenges. The heterogeneity of stakeholders and customers hampers trust management and security arrangements, exposing part of the root of the challenge in having a compatible AAA system for all participants. Authentication, Authorization and Accounting (AAA) between stakeholders and users must be fully reliable and trustworthy. To address the security challenges in 6G, ensuring user privacy at all stages of service provisioning is of the upmost importance. To this aim, given the vast offer of services and participants, fine-grained permissions must ensure the subscriber only displays the minimal necessary permission in each request, possibly through the use of attribute-based access control schemes (ABAC). For further emphasis, 6G network will deal with sensitive and high-data volume user data, therefore protecting privacy requires not just minimizing the data exposure but also ensuring unlikability across the service chain and correlated metadata analysis. In this context, traditional AAA systems do not embrace the decentralized nature of 6G [BCG+24]. New wave of mechanisms, integrating cryptographic methods with DLT and Decentralized Identity (DiD), can offer more robust and tamper-proof domain-less authentication to the decentralized services and infrastructure, also providing non-repudiation of actions

and transactions occurred. But always bearing in mind that the use of these solutions entails a notable increase in energy consumption and scalability challenges.

Automation of Federation

Orchestration has been widely applied in recent years, encompassing several layers, trying to cover every single domain and possible tasks of cellular networks. But still, because of its increasing complexity, security is a major challenge usually omitted in orchestration studies. Massive incorporation of devices of all kinds, as well as the integration of new or evolved technologies, in the 6G picture expand the attack surface to unknown limits. Massive sensor data collection and processing is a clear example. Interactions between orchestrators of different levels and cross domains represent another challenge, as each domain may have its own security protocols, assets, and physical devices. The lack of a common model to format data and coordinate the exchange of information expose a handicap of current networks. The absence of such a model, not only complicates coordination but also limits the information sharing, particularly important in terms of security. Besides, conflict and dependency detection and resolution are a challenging problem that accentuate the difficulties of the coordination. Described security orchestration challenges form a subset of challenges inherit from the Zero-touch Service and network Management (ZSM) paradigm. In this sense, the hierarchical view of ZSM needs to be updated integrating novel approaches that enable the horizontal learning and action enforcement. The use of federated agents in ZSM to share knowledge and information on security management tasks could help decompress this challenge. Orchestrating the interaction between FL agents while ensuring privacy-preserving Cyber Threat Intelligence sharing introduces even more complexity. Each agent may use different models and security protocols, making it difficult to align the goals of privacy, security, and efficiency [Z+24]. Besides, the dynamic nature of 6G deployments makes the monitoring system a challenging task, where the criticality of data flow in real time for mitigation and decision-making urgent for a correlation between network topologies and information gathered. Following the cross-domain federation approach, dedicated agents can minimize delays in coordination and control. However, this distributed approach introduces several security challenges, making multi-agent systems (MAS) vulnerable to attacks targeting (i) agent integrity and (ii) inter-agent communication [Z+24].

7.4 OVERARCHING BLOCKS

The keystone for the security and system management in current and future network is the Security, Orchestration, Administration, and Response (SOAR) closed loops. SOAR avoid the definition of complex E2E protocols, but rather specify how system components, grouped in different categories, should be interconnected to accomplish a specific policy/intent. SOAR specifies how the following layers are interconnected:

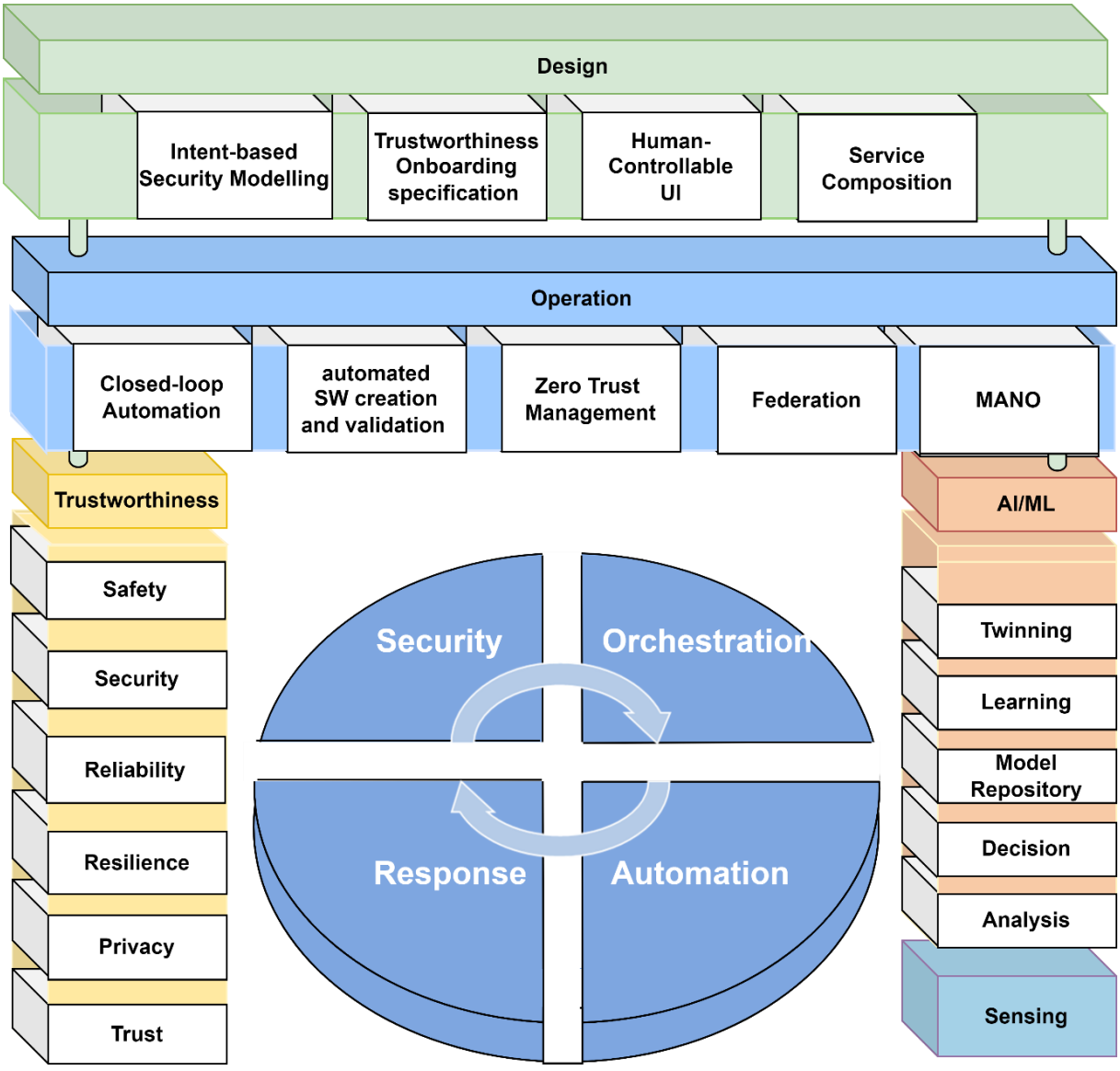


Figure 7.5: Overarching Security Blocks

7.4.1 TRUSTWORTHINESS

At the heart of the 6G security architecture lies the Trustworthiness pillar, which serves as the foundational motivation driving the design, implementation and application of all other pillars. Trustworthiness, is divided in several domains, including

Safety, Security, Reliability, Resilience, Privacy and Trust. Therefore, Trustworthiness solutions lie on one or more of these domains. Building trustworthiness is of the upmost importance, as confidence from customers and stakeholders relies on it. Trustworthiness must be built upon the anchors of trusts, which means effectiveness of applied domain properties can be proof. Thus, trustworthiness can be measured maximizing the trust posture over different targets, for instance between administrative domains, customers or final service consumers. Therefore, applying trustworthiness solutions in the different fields and layers, facilitate the collaboration for cohesive and unified security strategies. As mentioned, trustworthiness is present in the rest of the layers as the driving factor to apply measurements, but in particular, we identify several efforts to apply it through:

- Policy Administration and Enforcement Inspection [iTrust]: Specifies intent-based security and trust policies while providing explanations for their usage. This pursuit transparency and confidence in the system, as allow to stakeholders to understand the implications of the intent in the system.
- Static and Dynamic Trust Assessment [iTrust]: capability to conduct evaluation of posture of assets depending on their design (e.g. conformity to referential), their current posture (e.g. resource integrity and behaviour), and support forensic evidence.
- Collaborative Cyber Threat Intelligence Sharing [iTrust]: Establishes a baseline for trust and security-oriented collaboration among multiple providers. It leverages standards such as OASIS STIX, TAXI or MISP to enrich information on threat to account for trust context.
- DLT-Based Trust Infrastructure [Privateer, Desire6G]: By using one of the trusted anchors, it facilitates secure data exchange with transparency, traceability and accountability.
- Privacy-Aware Orchestration [Privateer]: which manages network services in compliance with privacy regulations like GDPR, guided by Level of Trust (LoT) assessments.
- Proof-of-Transit Mechanism [Privateer]: Ensures data flows through secure, predefined paths to safeguard privacy against potential attacks.

- Trust Exposure Layer [iTrust]: Limits shared information with external entities, exposing only necessary trust metrics without revealing sensitive infrastructure details.
- Privacy-Granting AAA Management: Manages the identity cross-domain and for third party services in a privacy preserving approach through 3 main components:
    - Issuer: Central authority for certificate creation using permissioned blockchain to generate and register subscriber's DiD and Verifiable Credentials (VCs).
    - Subscriber's Wallet: Empowers users with Self-Sovereign Identity (SSI) principles, allowing full control over ID management and minimizing information exposure through pseudo-attributes based on p-ABC schemes.
    - Verifiers: Distributed across the service infrastructure to manage access requests by validating pseudo-attributes without revealing original identities, ensuring trustworthy service access.

7.4.2 DESIGN LAYER

With a clear integration of human in the loop, and the objective of providing security modelling, covering from the onboarding specifications for devices and applications in a trusted way to the policy modelling with different levels of abstraction, enabling the scalability and interaction between domain's orchestrators. Principally this pillar are the mechanisms that define how network components are structured, interconnected, and managed to uphold trustworthiness. SOAR uses intents that belong to one trustworthiness domain to drive the interactions and applications, setting a groundwork for a secure and trustworthy network architecture. Main functional blocks for this block are:

- Intent-based modelling: Defining the models used for intents to define, desired system status through security objectives and also more specific definition for protection mechanisms, such as technologies or assets to be used. This block guarantee that security states are clearly articulated and understandable by domain's security orchestrators, propelling scalability and adaptability across diverse domains.

- Trustworthiness Onboarding Specification: Specifies onboarding requirements that align with trustworthiness domains, ensuring that all assets meet predefined security standards before being up and running, and accessible. This involves the trusted integration of devices and applications into the network.
- Human-Controllable UI: Incorporates a human-readable and manageable interface to manage aspects of the system, such as explainability of security AI processes. As well as allow overseeing enforced policies, or build new ones. This functional block is the main interface for enabling human-in-the loop.
- Service Composition: Defines how network services can be chained, and configurations can be applied based on the security intents. It reduces the complexity of orchestration and allows for different stakeholders to agree on how their domains will be integrated.

7.4.3 OPERATION LAYER

Receive intents from the design layers or alerts from the analysis assets. Interact with the AI/ML Layer to drive the enforcement of the high-level intent in the optimal way, considering several factors such as the network/device conditions and security requirements specified in the intent. This conforms a Cognitive MANO that through the security orchestrator and the intent-based security management, reproduces and communicate actions in two ways, to the twinning deployment to evaluate impact of actions and search for better solutions, and to the real infrastructural domain, composing and reconfiguring service compositions.

- Security Closed-Loop Automation: Implements intent-based automated set of recurrent steps interconnecting different logical entities, deployed assets and infrastructure producing a feedback loop that continuously monitor and adjust security measures. It ensures that the network remains resilient against evolving threats by automating responses based on triggered security intents and service compositions.
- Automated Software (SW) Creation and Validation: Leverages automation to develop and validate security-related software components. Trusted libraries as well as a set of validation tests described for the domain that the SW belongs to. This automation accelerates the deployment of security measures while ensuring their reliability and proper performance.

- Zero Trust Management: Enforces Zero-Trust principles by managing access controls and continuously verifying assets. Thus, maintaining minimal trust assumption, adequation isolation methods.
- Federation: This functional block is a key component with the envisioned horizontal nature for domain composition in the continuum context for 6G networks. The management of federated agents that enables CTI sharing and cross-domain security coordination.
- Management and Orchestration (MANO): Orchestrators mesh is envisioned for 6G. Orchestrators for different levels and tasks are envisioned to work together in protocol-less manner, for instance, ML orchestrators that manages the life-cycle operation of ML model (MLFO), orchestrators focused exclusively on applying security offering Security as a Service or effectively applying federation.

7.4.4 AI/ML LAYER

Represents the reaction plane, where security mitigations and service compositions are first evaluated in a Digital Twin infrastructure, allowing the impact evaluation and recalculation of intents enforcement through the virtual infrastructure. Monitoring and Analysis, in which we highlight the need of a topology inventory agent, tracking the network topology and providing updated information for threat detection and mitigation planning. Also, the network security analysis, mainly classifying data traffic from genuine or anomalous, by extracting and analysing metadata. This component can take into consideration the federation of agents, for data and analysis correlation.

- Twinning: The network DT acts as a dynamic representation of the mobile network, constantly learning and evolving alongside the real network environment. ML algorithms, within this DT framework, can leverage historical data, network topologies, and user behaviour patterns to model normal network behaviour and promptly identify deviations that may indicate malicious activities. This integrated approach not only enhances the precision of threat detection but also empowers security systems to both anticipate and proactively mitigate potential risks as well as analyse the impact of any proactive action to be taken.
- Learning: Processes enabling the federated learning between distributed agents and models. This module is responsible for the distributed ML training. To this end, privacy preserving solutions are leveraged, such as federated learning.

Before the actual training takes place, preprocessing and feature preparation takes place. Trained ML models are stored in a local database, where they can be retrieved on demand. All procedures are orchestrated by the machine learning function orchestrator (MLFO). Adversarial training is used to protect AI/ML models against attacks, while privacy-preserving mechanisms are integrated to protect heterogeneous data types

- Model Repository: Place in which to hold models, accessible by stakeholders and orchestration processes. This enables a place to share trained models, selecting optimal models given the infrastructure and assets used in the service composition. Allows for on-demand retrieval and deployment of the latest threat detection models, ensuring they are readily available when needed.
- Decision: 6G Leverages AI-driven decision-making processes to interpret data from various sources, decision is driven by one or more of trustworthiness domains, such as privacy-sensitive decisions or trust-based decisions. Prioritising giving trustworthiness properties to the system while maintaining consistency and reliability. Decisions retrieve analytics, topology, assets etc. Which complexity lead into having specialized engines that manages concrete tasks instead of a general decision-making engine. Decisions are envisioned to target first DT domain to evaluate impact and look for better solutions.
- Analysis: Analytics based on AI and machine learning models also encompasses detection of anomalies in user equipment (UE) and the network, processing data in a federated manner to enhance privacy, and following the federated NWDAF deployment scenario of the 3GPP.

7.4.5 SENSING

Sensing is another pillar to cover over-architectural security. Sensing in 6G extends from the centralised infrastructure to the far-edge devices. Where security capabilities must be able to extract metrics from the latter nodes, which are usually the most vulnerable as they are generally dynamic and mobile in nature. Sensing is one of the main enablers of the cloud in continuum for these devices, since it is necessary to know their state before forming a chain of services using them. Sensing is also of particular importance to detect threats outside the security perimeter of the operators. Sensing itself do not represent a security mechanism but enable new ways of studying security for early detection and correlation. On the other hand, emerging sensing capabilities

can be used by attackers to incorporate new layer of information from the surroundings, for instance doing a mobile network mapping, inferring what kind of devices are connected to the network or also their location. Beside the new threats incorporated by sensing, it is the main enabler for the following functions:

- Physical Layer Anomaly Detection: Detection capabilities can be used to detect threats at the physical layer, such as jamming, spoofing, eavesdropping, semantic communication inference, denial of service attacks and more. By analysing environmental metrics such as signal patterns, jamming levels and anomalous behaviour at the physical layer, this module facilitates real-time identification of attacks that could compromise the security level of the network by threatening communications integrity or degrading network performance.
- External Perimeter Surveillance: Detection extends security capabilities beyond the traditional network perimeter by enabling detection of external threats. This includes the identification of entities not authorised to broadcast on certain frequencies, malicious devices or environmental disturbances near critical infrastructure. By continuously monitoring the external environment, this module improves proactive threat detection and overall network resilience.

7.5 REFERENCES

[ISO16] ISO Central Secretary, “Systems and software engineering – Systems and software quality requirements and evaluation (SQuaRE) – Measurement of quality in use,” International Organization for Standardization, Standard ISO/IEC 25022, 2016. [Online]. Available: https://www.iso.org/standard/35746.html

[K+21] L. Kastner, et al., "On the Relation of Trust and Explainability: Why to Engineer for Trustworthiness," in 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), Notre Dame, IN, USA, 2021 pp. 169-175. doi: 10.1109/REW53955.2021.00031

[YAZ+23] X. Yan, X. An, W. Ye, M. Zhao, Y. Xi and J. Wu, "User-Centric Network Architecture Design for 6G Mobile Communication Systems," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 305-310, doi: 10.1109/EuCNC/6GSummit58263.2023.10188283.

[CBS21] L. Chazette, W. Brunotte, and T. Speith, “Exploring explainability: A definition, a model, and a knowledge catalogue,” in IEEE 29th International Requirements Engineering Conference (RE). IEEE, 2021.

[Pie11] W. Pieters, “Explanation and trust: What to tell the user in security and AI?” Ethics and Information Technology, vol. 13, no. 1, pp. 53–64, 2011.

[GRR+19] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models,” ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2019.

[RSS16] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the predictions of any classifier,” in Proceedings of the 22nd ACM SIGKDD International Conference

on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery, 2016, pp. 1135–1144.

[LOS+21] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kastner, E. Schmidt, A. Sesing, and K. Baum, "What do we want from explainable artificial intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research," Articifial Intelligence, vol. 296, 2021.

[KP20] M. Kumar and P. Pattnaik, "Post Quantum Cryptography(PQC) - An overview: (Invited Paper)," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 2020, pp. 1-9, doi: 10.1109/HPEC43674.2020.9286147.

[DES23] DESIRE6G, "D3.1: Initial report on the intelligent and secure management, orchestration, and control platform", 2023. Online: https://zenodo.org/records/10356033

[NIS24] "NIS 2 Directive." Accessed: Aug. 06, 2024. [Online]. Available: https://www.nis-2-directive.com/

[Eri24] "6G standardization – an overview of timeline and high-level technology principles", Ericsson, March 22, 2024. [Online] Available (Accessed: Sept. 13, 2024): https://www.ericsson.com/en/blog/2024/3/6g-standardization-timeline-and-technology-principles

[GNT+24] P. Gkonis, N. Nomikos, P. Trakadas, L. Sarakis, G. Xylouris, X. Masip- Bruin, and J. Martrat, "Leveraging network data analytics function and machine learning for data collection, resource optimization, security and privacy in 6g networks," IEEE Access, pp. 1–1, 2024.

[NLC+21] V.-L. Nguyen, P.-C. Lin, B.-C. Cheng, R.-H. Hwang, and Y.-D. Lin, "Security and Privacy for 6G: A Survey on Prospective Technologies and Challenges," IEEE Commun. Surv. Tutor., vol. 23, no. 4, pp. 2384–2428, 2021, doi: 10.1109/COMST.2021.3108618

[BCG+24] Jose Manuel Bernabé, Eduardo Cánovas, Jesus Garcia-Rodriguez, Alejandro M.Zarca, Antonio Skarmeta (2024), "Decentralised Identity Management solution for zero-trust multi-domain Computing Continuum frameworks", Future Generation Computer Systems, https://doi.org/10.1016/j.future.2024.08.003

[Z+24] A. Zahir et al., "Distributed Genuine Intelligence: From Agent Integrity to Secure Inter-Agent Communications," European Conf. on Network and Communication (EuCNC), 2024.

8 SUSTAINABILITY

8.1 INTRODUCTION

As we advance toward the design and development of 6G, sustainability has emerged as an essential pillar of next-generation smart networks and services (SNS) and their connectivity strategies. The digital transformation of our society and economy must align with European norms and values, including the European Green Deal principles and the global Sustainable Development Goals (UN SDGs), while ensuring no one is left behind.

The development of more advanced communication networks faces compound challenges: increased energy demands, complex resource allocation, impacts on biodiversity, e-waste management, and the *rebound effects* produced by ICT (referring to the case where the efficiency improvements in ICT technologies lead to increased overall resource consumption due to higher usage, offsetting potential environmental benefits), all while meeting heightened social expectations surrounding digital inclusivity. The SNS JU initiative seeks to incorporate sustainability at the forefront of its research and innovation agenda, supporting projects that both minimise their environmental impact (reduced first order effect) and maximise their positive contribution to sustainability challenges (increased second order effect).

In this document we refrain from using the terms *footprint* and *handprint*. Instead of footprint, we use *first order effect* defined as the direct economic, societal or environmental outcome associated with the existence of an ICT based solution, and generic processes supporting the deployment and operation of the ICT based solution. These could be positive and/or negative for a stakeholder. Instead of handprint we use *second order effect* defined as the indirect outcome created by the use and application of ICT based solutions, which includes economic, societal or environmental changes. These could be positive and/or negative for a stakeholder.

Environmental sustainability challenges: first order effect

6G networks promise substantial advancements while presenting significant environmental sustainability challenges. The network's complex infrastructure, coupled with exponential increases in data traffic, could dramatically expand its environmental first order effect. Key environmental sustainability drivers include energy consumption, resource use, biodiversity conservation, and electronic waste reduction. These factors

underscore the critical need to minimise the environmental impact of network infrastructure while supporting responsible digital evolution.

The environmental first order effect of 5G and upcoming 6G networks is multifaceted, encompassing energy consumption in production and use, e-waste, rare resource extraction, air and water pollution, and biodiversity impacts. To address these challenges, SNS JU projects are developing innovative solutions focused on reducing energy and more general environmental first order effect through optimised resource allocation, relay management, reusability, and AI-driven network intelligence.

However, significant challenges remain. The production and disposal of electronic equipment for 6G networks may contribute substantially to e-waste, natural resource depletion, and biodiversity loss. Moreover, as a rebound effect, the infrastructure required to support 6G systems and related device production is expected to increase Greenhouse Gas (GHG) emissions. These elements — resource consumption, rebound effect, biodiversity conservation, trust, security, inclusivity, and affordability — are integral to the larger sustainability conversation surrounding 6G.

In the field of data transmission, scalable and sustainable optical transport networks are being designed to handle the massive bandwidth requirements anticipated in 6G networks. Through optical switching and control protocols, these networks enable efficient data transmission via multi-granular optical nodes and flexible wavelength allocation. Additionally, applying photonic solutions to the front-haul, mid-haul and back-haul segments, also known as X-haul, promises to minimise energy consumption by replacing power-intensive electronic processing with highly efficient optical solutions.

Societal and economic dimensions: second order effect

While environmental sustainability through enhanced energy efficiency is crucial, the impact of 6G on societal and economic sustainability is equally important. To ensure sustainability by design, 6G must address the digital divide, manage the rebound effect, and promote equitable access to and benefits from digital resources. Ensuring sustainability thus necessitates a two-fold approach: not only must systems be accessible, but users must also possess the requisite capacity, infrastructure, and contextual support to derive measurable value from them. To that end, SNS Ensuring sustainability thus necessitates a two-fold approach: not only must systems be accessible, but users must also possess the requisite capacity, infrastructure, and

contextual support to derive measurable value from them. Accordingly, SNS JU projects are working to provide affordable, high-quality connectivity, particularly in underserved and remote areas, thus fostering digital inclusivity. The potential to impact societal equity extends beyond mere connectivity, encompassing issues of trust, security, affordability and digital literacy.

Unlike previous generations, 6G networks will deeply integrate non-terrestrial (NTN) components, including satellites, aerial networks, and high-density terrestrial infrastructure. Such full integration proves particularly impactful in bridging the digital divide, extending connectivity to remote and underserved regions where terrestrial infrastructure deployment would be both economically and environmentally costly. Furthermore, the NTN component's adaptable architecture enhances sustainability across the transportation sector, supporting efficient operations in aeronautical, maritime, railway, and land vehicle systems.

Regulatory framework

Finally, navigating the regulatory landscape of sustainability is critical for the deployment of 6G networks. Compliance with the European Green Deal, UN SDGs, and various national policies presents both challenges and opportunities. The alignment of 6G strategies with broader sustainability policies requires careful consideration of trade-offs between accelerated digitalization, societal acceptance, and environmental responsibility. As 6G networks mature, achieving compliance will demand a comprehensive approach that addresses technological, societal, and economic impacts.

This chapter explores the environmental considerations and broader socio-economic frameworks being addressed by SNS JU projects such as BeGREEN [BeG23], ETHER [ETH23], FLEX-SCALE [FLE23], HEXA-X-II [HEX23], ORIGAMI [ORI24], PROTEUS-6G [PRO24], 6G-NTN [NTN23], and 6G4Society [6G4S24], examining how these initiatives contribute to shaping a more sustainable future for digital infrastructure, examining specific technological innovations for reducing environmental first order effect, analysing solutions for enhancing positive second order effect through digital inclusion and societal benefits, and discussing concrete approaches to regulatory compliance. A strong emphasis is placed on contributions promoting energy and power efficiency, indicating that these are seen as the most immediate and tangible areas for sustainability impact across the projects. In this context, each section provides insights

into project achievements, methodologies, and contributions to building a sustainable 6G ecosystem.

Summary of the chapter

The evolution of B5G and 6G networks is driven by the need for sustainability and efficiency. Incorporating relay nodes and innovative optical X-haul technologies can significantly reduce energy consumption and Operational Expenditure (OPEX). These advancements, along with AI/ML techniques, enhance network performance and support sustainable operations. By unifying terrestrial and non-terrestrial domains, 6G networks aim to create a more energy-efficient and sustainable wireless infrastructure, addressing the growing demands of mobile data traffic. This section provides technology options pertaining to sustainability focusing on optimizing different parts of the network, i.e., the Radio Access Network (RAN), the transport or entirely the End-to-End (E2E) network.

RAN sustainability advancements: Incorporating relay nodes in future B5G RAN networks can significantly enhance sustainability. By mitigating signal blocking and increasing coverage in high-density areas, relay nodes reduce the need for additional base stations, leading to lower energy consumption and operational costs. This energy efficiency extends to user equipment, which transmits less power when connected to a relay, thus conserving battery life. Studies have shown energy savings of up to 90% with the use of relays. Additionally, AI/ML techniques can optimize relay functionalities, further improving system performance and reducing energy consumption. Overall, relay nodes contribute to a more sustainable and energy-efficient wireless network infrastructure.

Transport network sustainability advancements: The scalability challenges posed by 6G networks, driven by high end-user rates and massive small-cell deployments, necessitate innovative optical X-haul network technologies. These technologies aim to provide ultra-high-speed, energy-efficient all-optical bypasses, significantly reducing power consumption and enhancing performance. By integrating multiple optical switching granularities and dynamically adapting bandwidth, 6G networks can achieve high energy savings, contributing to sustainable and efficient network operations. In parallel, by enabling dynamic functional splits and passive traffic distribution elements, these technologies can adapt to varying traffic requirements, leading to more efficient energy use. This approach not only reduces operational costs

but also contributes to the development of sustainable and energy-efficient network infrastructures.

E2E sustainability advancements: The rapid growth of mobile data traffic is pushing current networks to their limits, necessitating the development of 6G networks that unify terrestrial and non-terrestrial domains, i.e., aerial and space layers. This transformation will create a 3D architecture, enhancing link capabilities and resource allocation. To ensure sustainability, 6G networks need to maximize energy efficiency, reducing Total Cost of Ownership (TCO) and energy consumption. Strategies for efficient and low-complexity resource allocation should consider various resource types, i.e., network, compute and storage, as well as related constraints, enabling real-time decision-making and guaranteeing end-to-end optimality. Overall, integrated Terrestrial and non-Terrestrial 6G Networks (TN-NTNs) aim to develop sustainable solutions for user association, traffic routing, and NF placement, ensuring sustainable and efficient network operations.

8.2 RAN SUSTAINABILITY ADVANCEMENTS

8.2.1 RELAY NODES FOR ENERGY-EFFICIENT RAN

The inclusion of relay nodes in future B5G RAN can be useful for different purposes such as mitigating signal blocking in millimetre wave deployments or increasing coverage and capacity in high-density areas, leading to a reduction in the number of BSs to deploy. Besides that, the use of relays is a cost-efficient option for energy saving as it allows reducing the base station transmit power consumption and, consequently, it facilitates a reduction of MNOs OPEX. At the same time, User Equipment (UE) that is connected to a relay also transmits less power, thanks to the better propagation conditions, thus reducing the UE battery consumption. The power consumption improvements that can be obtained through the use of relays in a university campus scenario have been studied in [BeG24], considering indoor relays at different positions and buildings. Energy savings ranging between 50% and 90% with respect to not using relays were observed, depending on the power consumption model parameters and the required bit rate, and improvements in the energy efficiency in a factor around 2.6 were obtained.

The deployment of relay nodes in wireless networks can be done with different types of relays. The first approach involves the use of fixed relays, where the MNO chooses the position of the relay as an extension of the currently deployed base stations. The

second approach consists of installing relays within a moving element (Moving Relay), such as a bus or a train, as exemplified in [NCK20]. The last approach involves equipping UEs with a relay functionality, in a way that some of the UEs may become relay-UE (RUE) and act as a relay between the BS and neighbour UEs [PS21]. Standardization for relay support in 5G has been introduced by 3GPP in the so-called Integrated Access and Backhaul (IAB) technology since Rel-16.

The definition of specific functionalities for the control of the relays is crucial to improve the system performance and reduce energy consumption. The incorporation of AI/ML techniques in these relay control functionalities is a key concept to make them more efficient. These relay functionalities cover different aspects. On the one hand, a coverage hole detection functionality makes use of a set of collected measurements with the aim of identifying geographical regions with large traffic demands and poor coverage. On the other hand, a fixed relay placement functionality is in charge of determining adequate geographical locations to place a new fixed relay and establishing their initial configuration parameters. Additionally, a candidate RUE identification functionality aims to identify UEs that can be good candidates to act as relays between the network and other UEs in their proximity. Finally, a relay activation/deactivation functionality is in charge of the dynamic activation/deactivation of these relays/RUEs with the objective of improving the network performance and reducing the energy consumption.

8.2.1.1 ARCHITECTURAL COMPONENTS

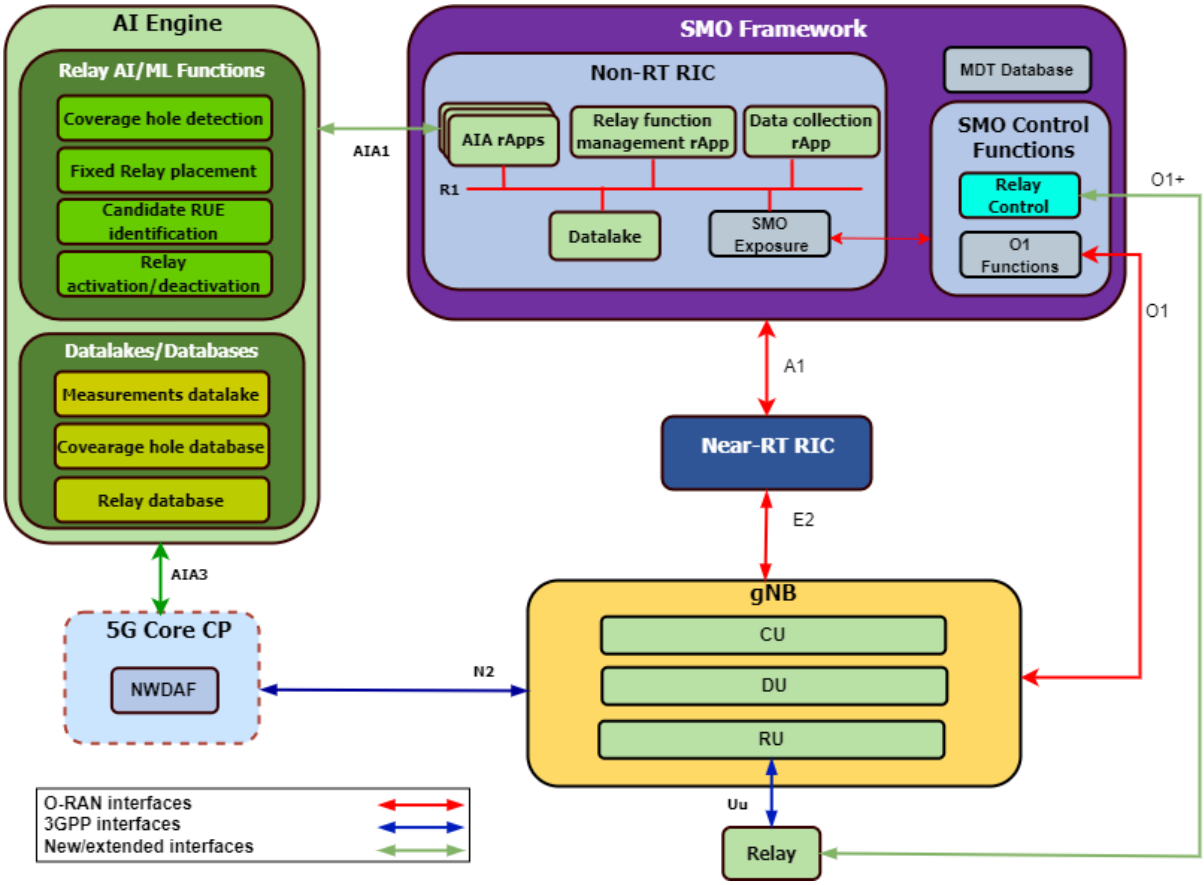


Figure 8.1: Relay control components and functionalities

Figure 8.1 details the components that support the implementation of the above-mentioned relay control functionalities. On the one hand, as shown in Figure 8.1, a Relay Control entity placed at the Service Management and Orchestration (SMO) is in charge of the interaction with the relays for the collection of the network measurements and the relay reconfigurations through an extended O1 interface, denoted as O1+. In turn, gNB measurements are sent to the SMO through the O1 interface. On the other hand, the relay control functionalities are sustained by different rApps in the non-RT RIC. In particular, the Data Collection (DC) rApp is in charge of the management of the different processes related to the collection of measurements. Moreover, the Relay Function Management (RFM) rApp is in charge of the coordination and management of all the functionalities related to the control of the relays. This RFM rApp decides when and where to trigger the execution of each functionality. Each AI/ML-based relay control functionality is sustained by a different AI Engine Assist rApp (AIA rApp), which makes the AI/ML workflow services of the AI Engine accessible to the non-RT RIC. These AIA rApps cover different aspects such as data pre-processing, model triggering,

performance monitoring of the AI/ML models and determining the necessity of model updates or model retraining. For each relay control functionality, the associated AIA rApp triggers the execution of the corresponding AI/ML function hosted in the AI Engine. These AI/ML functions make use of collected information stored in the AI Engine datalakes/databases to obtain an output (e.g. the result of a clustering, a prediction or a recommendation) useful for taking adequate decisions of relay deployment or relay reconfiguration.

8.2.1.2 SOLUTION

This section describes how the components mentioned above can fulfil the different proposed relay control functionalities. With certain periodicity, a network monitoring process is run to identify cells that require the activation of some of the proposed relay control functionalities. Then, for these bad performing cells, a process of data collection is done. It consists of the collection of geo-located measurements e.g. based on Radio Resource Control (RRC) UE measurement reports. Collected measurements are stored in a measurements datalake in the AI Engine (see Figure 8.1). Then, the RFM rApp, activates the coverage hole detection process that is managed by an AIA Coverage Hole Detection (AIA CHD) rApp, which triggers the execution of the process in the AI Engine (see Figure 8.1) making use of the collected measurements stored in the AI datalake. The result of this process is a characterization of the coverage holes identified in each cell and is stored in the Coverage Hole Database. After the identification of a coverage hole, the RFM rApp oversees deciding the most adequate solution to address this problem. One possible solution is the use of RUEs. For this purpose, the RFM rApp may trigger the process of Candidate RUE identification, with the objective of identifying UEs that may be good candidates to become a RUE. In general, UEs with good propagation conditions with its associated BS, a static/semi-static mobility pattern, a periodical and predictable space-time location and large session durations may be good candidates to become RUE. The list and characterization of candidate RUEs for each coverage hole is stored in the Relay database. In case no suitable RUE has been found to address a specific coverage hole, the RFM rApp may trigger the Fixed Relay Placement functionality to determine an adequate geographical location to place a new relay. In case a new fixed relay is deployed, the geographical location and the configuration parameters of this new relay are added in the Relay database. With the aim of improving the system performance and reducing the overall energy consumption, both fixed relays and RUEs are dynamically activated/deactivated depending on the number of users in

their surroundings. In order to do this, the Relay activation/deactivation process makes use of recently collected measurements and information related to the relays status and uses an AI/ML model to take smart relay activation/deactivation decisions.

8.3 TRANSPORT NETWORK SUSTAINABILITY ADVANCEMENTS

8.3.1 OPTICAL TRANSPORT NETWORKS SUPPORTING SUSTAINABLE CAPACITY SCALING

The target end-user rates and massive small-cell deployments envisioned for 6G, as well as the rates generated by future immersive AR/VR and holographic services supported by next generations of FTTH networks, pose scalability challenges to the electronic packet layer in terms of performance and power consumption. This is especially the case of metro aggregation-core segments where traffic flows in a totally hierarchical way to and from either WAN transit nodes or CDN caches. In this case, traffic is not meshed but concentrates into a few core nodes, which simplifies the task but creates hot spots with huge capacity needs. The challenge is even more complex if the evolution toward centralization of radio processing functions becomes mainstream in 6G settings. This situation calls for performing disruptive research on optical X-haul network technologies of Optical Switching Nodes and Transceiver Interfaces to enable flexible capacity scaling. Concrete goals set to cope with the 6G vision are: ≥10 Tb/s rate per optoelectronic interface, ≥1 Pb/s capacity per link (utilizing ultra-wideband (UWB) transmission and space division multiplexed (SDM) fibre solution) and ≥10 Pb/s throughput per multi-granular optical node (MG-ON) utilizing new waveband-selective switch (WBSS)). These rates go beyond the capacities of the conventional C-band employed in DWDM systems. The objective is providing ultra-high-speed energy-efficient all-optical bypasses to 90% of the traffic destined and coming from the core, seamlessly integrated with the IP layer by means of a smart control plane. In this context, the support of multiple optical switching granularities is essential to achieve high energy savings and dynamically adapt the bandwidth used to the traffic flows served.

8.3.1.1 ARCHITECTURAL COMPONENTS

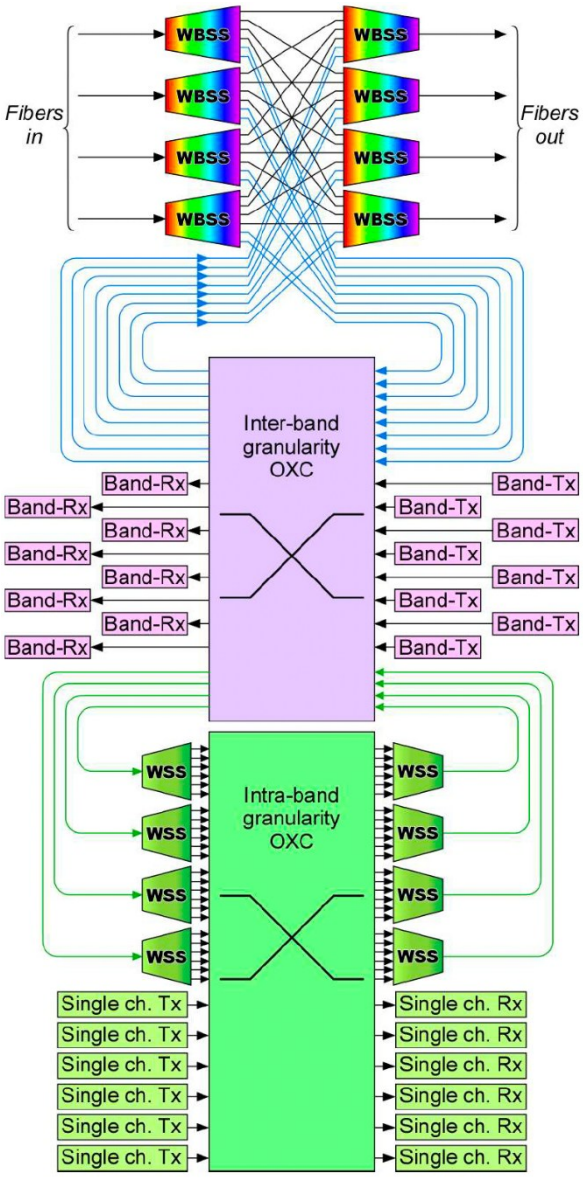


Figure 8.2: An optical node architecture featuring multiple switching granularities [TPU+24].

Most innovation and changes required in existing standards involves the control plane of the transport network. Current 5G schemes are in many senses technology-agnostic with respect to the fixed part of the network. This makes the end-to-end guarantees envisioned for real-time applications hard to achieve. However, transport networks are going through a revolution towards disaggregation, openness and programmability that enables unprecedented seamless integration with IP and e2e 6G services. The control of multi-granular optical nodes combining wavelength, waveband, and spatial switching is complex given the combinatorial possibilities and

implications in the physical layer of multigranular configurations. The Optical SDN controller, other than traditional wavelength Routing and Spectrum Assignment (RSA), needs to support dynamic optical band switching (FLEX-RSA), where an optical-band circuit can be dynamically configured to accommodate an increasing (or decreasing) number of wavelength channels (over the same path and assigned to the same band) depending on the traffic needs. Recovery of the optical band, i.e., an entire band (e.g., C band) or a wide portion of it, is also considered. For example, in case of a C-band amplification malfunction, the optical band could be recovered over the S band along the same route, or in case of a soft-failure Quality of Transmission (QoT) degradation, the spectral window could be changed where QoT is superior.

8.3.1.2 SOLUTION

Achieving the target rates in a cost-effective and energy-efficient way can be achieved through the utilization of ultra-high bandwidth transceivers employing photonic/plasmonic technologies and the efficient exploitation of optical spatial and spectral switching (UltraWide-Band Spectral & Spatial Lanes Multiplexing; UWB/SDM). The target is achieving record energy efficiency (sub-pJ per switched/transmitted bit) and low cost, enabled by photonic integration and optical transparency, replacing/bypassing power-hungry and costly electronic processing systems (e.g., electronic routers/switches). The Optical Nodes and their Transceiver Interfaces should be controlled by a Machine Learning-enabled SDN control plane running smart resource allocation algorithms, which will optimize traffic flow routing across network layers and segments, improving network QoS (high rates, low latency, high reliability/availability) and low cost/power consumption, as required by 6G specifications.

Network nodes supporting MG-ON capacities ≥10 Pb/s are envisioned to be capable of switching at multiple granularities, ranging from individual channels to full fibres through wide spectral super-channels (flexible bands) withing the Ultra-wide band (UWB) window (1460-1625 nm, (S, C, and L bands)) and must support Spatial Division Multiplexing (SDM) to generate flexible reconfigurable add-drop in colour/direction/contention-less (C/D/C) ROADMs.

Control and orchestration should rely on a cloud-native architecture such as ETSI TeraFlowSDN (TFS) controller. TFS is an open-source cloud-native SDN controller able to scale the management of a large number of flows. Its modular architecture is based on microservices making use of containers to isolate the functionalities of the components. Each microservice is independently deployed and the communication

between them occurs through a custom open interface based on Google Remote Procedure Call (gRPC). TFS needs to be extended with new or restructured microservices such as optical for multi-granular optical support, end-to-end orchestrator for packet/optical support, and a ZSM (Zero-touch network and Service Management)-aligned monitoring-analytics-automation loop. This enables to collect information related to network services, operations, and devices in real-time, analyse data in real-time to detect any undesired conditions, and autonomously reconfigure/reoptimize packet and multi-granular optical networks based on the defined policies for the monitored key performance indicators.

8.3.2 OPTICAL TRANSPORT NETWORKS SUPPORTING FLEXIBLE FUNCTIONAL SPLITS

The latest 5G new radio numerology and, extrapolating, 6G radio cell-free massive MIMO schemes, will boost dramatically the capacity requirements of fronthaul networks. Irrespective of whether Distributed or Cloud RAN will be a widely deployed technology in 6G, it is clear that dynamic functional splits are a useful tool to move the radio processing functionality deep into the MAN in search of reduced energy consumptions (due to the disconnection of elements in the Distributed RAN) when the network load allows such centralisation. This would not be possible without a dynamic, flexible, scalable, cost-effective, high-bandwidth, and low-latency optical transmission and switching technology in the fronthaul network that makes use of passive traffic distribution elements as we get close to the edge (interfacing the RUs), alternative to high energy consuming packet switches. Once accomplished, it would provide opportunities to better exploit the multiplexing gain across multiple RUs by adapting the individual dynamic functional splits to the varying traffic requirements of all cells fed by a common set of fronthaul/midhaul links. Achieving such inexpensive technology is extremely complex.

This would not be possible without a suitable agile service management, orchestration and control system to enable dynamic reconfiguration of the functional splits in the RU/DU/CU and dynamic reconfiguration of the packet-optical X-haul network to deliver the required transport capacity demanded by the selected split level.

8.3.2.1 ARCHITECTURAL COMPONENTS

A photonic element that can help to build a fronthaul network to cope with the aforementioned challenges is some sort of spatially-diverse point-to-multi-point

(SDPtMP) optical fronthaul distribution network as depicted in Figure 8.3. This element is intended to distribute efficiently the energy (rather than through a splitter) to the edges and it is connected to an advanced ROADM (Reconfigurable Add-Drop Multiplexer) at a Central Office. Alternative, splitter-based Digital Subcarrier Multiplexing (DSCM) can perform this same function at a shorter reach, lower speed but with more flexible bandwidth allocation options.

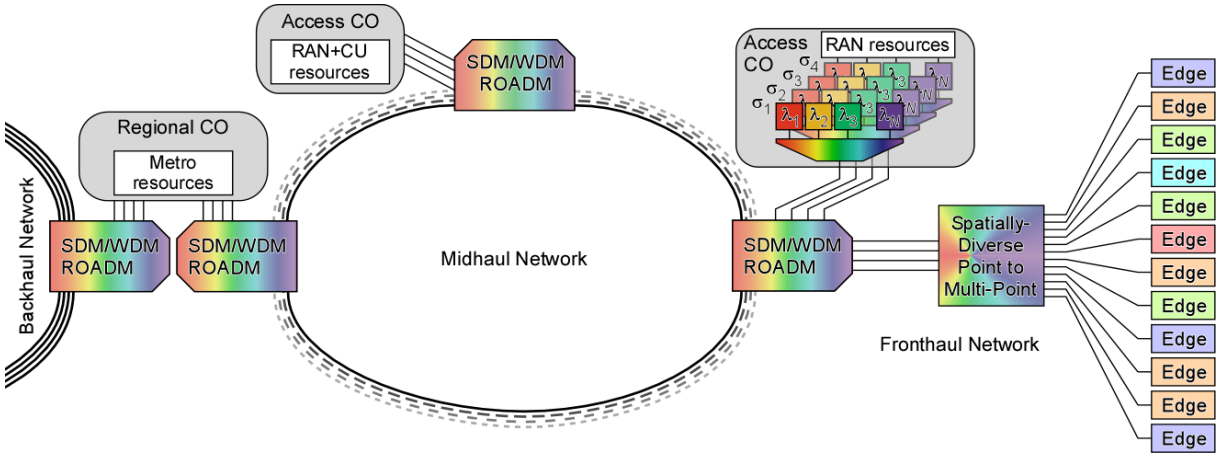


Figure 8.3: Location of spatially-diverse point-to-multipoint devices in the overall X-haul network.

Innovations require proper standard protocols to realize the flexible functional split concept, possibly via O-RAN, both from the functionality migration point of view and from the ability to configure the network to adapt the capacity to the traffic profile according to functional split and load.

8.3.2.2 SOLUTION

A possible implementation of the aforementioned concept is a spatially-diverse point-to-multi-point (SDPtMP) optical fronthaul distribution network, with degree-four SDM (Spatial Division Multiplexing) feed for a first capacity multiplier, and WDM for a second capacity multiplier [CGG+23]. Introducing an SDM feed to the ODN and establishing capacity allocations in digital subcarrier allocations per RU and CU, can allow all RUs to operate on a common optical wavelength without interference. This can be accomplished by optical devices that optically separates transmitted digital subcarriers using an array of precise optical interleaving filters. These filters are designed to jointly operate as a circular subcarrier mux/demux. Once the capacity is exhausted across the SDM group, a second optical wavelength may be introduced and assigned to particular RUs using a reprogrammable WSS or fixed demultiplexer. The

SDPtMP architecture benefits from SDM/WDM capacity scaling and the assignment of spatial/spectral RU-CU connections, without the need for splitters/combiners in the light-path.

Next-generation Digital Subcarrier Multiplexing (DSCM) transceivers can be used for dynamic functional split up to option 7-2. Innovative ultra-high-speed, low-latency, low-cost, and power-efficient Lite-Coherent (LITE-COH) transceivers (TXR) can be developed as key enablers for the realization of cell-free MIMO. This can allow for ultra-high fronthaul capacities, as specified in functional split option 8, enabling 6.4 Tbps per fibre over 8 wavelengths (l). The development of the LITE-COH TXR will feature 0.8Tb/s per-l, while also contributing to reduced latencies and significant cost and power-reductions by virtue of all-optical-signal-processing (AOSP) functionality, which will replace the bandwidth-limited and power-hungry DSP used to process I/Q quadratures and orthogonal polarizations in conventional coherent TXRs. As a result, the new TXRs will save an estimated 50% power.

The control plane should also account for the design of data models and protocols of the key optical network elements to enable fully programmable and monitorable X-haul network deployment. Autonomous networking architectures and optimization algorithms for efficient packet-optical network resource, NF, and service management will also be addressed.

8.4 E2E SUSTAINABILITY ADVANCEMENTS

8.4.1 REAL-TIME SUSTAINABLE RESOURCE ALLOCATION IN INTEGRATED TN-NTNS

The rapid growth of mobile data traffic is expected to push existing network capacities to their limits in the coming years. To address this challenge, 6G networks aim to unify terrestrial and satellite domains, providing a broader range of access points. This shift will drive a transformation of the current Radio Access Network (RAN), leading to a 3D architecture—a “network of networks”—that enhances X-haul (front/mid/back-haul) link capabilities through technologies such as Inter-Satellite Links (ISLs).

In this context, resource allocation becomes even more challenging, due to the additional space and aerial base stations (BSs) and X-haul links, requiring new channel modelling to support them. In parallel, the User Equipment (UE) Service Function Chains (SFCs) and their associated x-Network Functions (xNFs), which can be NFs of any type,

e.g., physical (PNF), virtual (VNF) or cloud-native (CNF), have to be deployed ensuring that the capacity and computational constraints of each node are met and that the xNFs are executed in the same order as in the SFC without violating any link constraints [LGLY21]. xNFs can be placed either to BSs (in both terrestrial, aerial and space domain) colocated with Multi-Access Edge Computing (MEC) capabilities for lower latency, or to farther cloud computing nodes with greater capabilities but with higher latency or to fog computing nodes in between offering a trade-off between computing capacity and latency. For wireless links, the use of the mmWave band is favoured, while for the ISL links higher frequencies, i.e., optical connections from 20 to 375 THz [ITU02], will be preferred exploiting the lack of atmosphere in space.

Due to the network densification dictated by the unprecedented data traffic growth, 6G networks are forced to maximize their energy efficiency, mainly to: a) reduce the associated Operational Expenditure (OPEX) of the network and b) decrease the associated energy and power consumption, a critical determinant in achieving sustainable network design and operation. As a result, due to these imminent additions and modifications to future mobile networks, strategies for online resource allocation should be designed that: i) consider different resource types, such as computational, communication and storage, as well as 6G technologies, e.g., THz bands, and their constraints, ii) enable real-time network decision-making by optimizing the algorithmic computational complexity, iii) guarantee end-to-end (E2E) optimality by taking into account the E2E latency and data rate needs and iv) maximize network energy efficiency, while guaranteeing the user QoS satisfaction. In a nutshell, integrated Terrestrial and non-Terrestrial 6G Networks (TN-NTNs) call for the development of energy-friendly solutions for online user association, traffic routing and xNF placement, while guaranteeing the QoS of the UE and the SFC chaining.

8.4.1.1 ARCHITECTURAL COMPONENTS

For real-time E2E network optimization, 6G networks set forth a vision for end-to-end Network Intelligence (NI) enabling zero-touch management, which requires the coordination of many NI algorithms running across schedulers, controllers, and orchestrators, ensuring their conflict-free and synergic operation. Yet, current frameworks by main Standards Developing Organizations (SDOs) are far from supporting a native NI integration (e.g., 3GPP Rel. 19 will not include centralised AI/ML control, which will thus not be considered by the SDO until 2026 [Chu24]). The proposed approach builds instead on proposals by 5G PPP/6G IA [BGG+23], which are

inspired by the results of [DAE25]. A new NI Stratum is introduced for the control of AI/ML models deployed across the network. The proposed AI Layer (Figure 8.4) is aligned with the internal organization, AI/ML model representation and operation of the NI Stratum. The AI layer (internal or AI as a Service, AIaaS) supports the E2E cross-domain Management and Network Orchestration (MANO) framework to coordinate on multiple levels, which are shown in Figure 8.4, and include:

- Infrastructure Layer – It includes the TN and NTN assets (Core/Central Cloud, Transport, Edge, Extreme Edge infrastructure), External Infrastructure, Virtualised Infrastructure, i.e., NFVI and non-virtualised resources.
- Network Layer –NFs (e.g., 3GPP Core Network (CN), RAN, transport) and third-party functions
- Service Layer – Network Slice Instances (NSIs) composed of NFs residing in the Network Layer, slice management and exposure mechanisms.
- Application Layer – applications using functionalities offered by Network Slices (NSs).
- Business Layer – business actors: Mobile Network Operators (MNOs), verticals, etc.

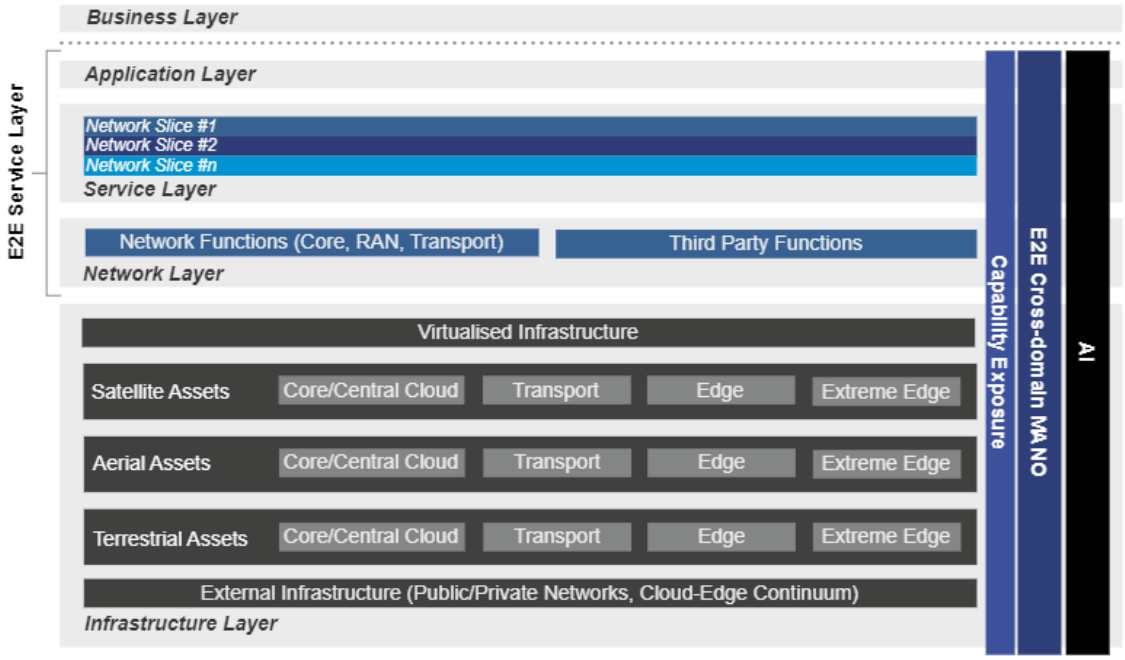


Figure 8.4: Proposed architecture for E2E real-time network optimization.

8.4.1.2 SOLUTION

Both an optimal solution that solves the joint real-time user association, traffic routing and xNF placement as well as a heuristic algorithm for low complexity are proposed. The optimal solution is based on the system being modelled as a directed graph, with the set of non-UE nodes and the set of links among them, solving the joint problem subject to flow conservation, power, capacity and causality constraints. The heuristic, which is named as Online Power-efficient Terrestrial Non-terrestrial Resource Allocation Heuristic Algorithm (PETA), studies the joint problem, aiming to maximize user acceptance ratio as well as energy efficiency. As illustrated in Figure 8.5, PETA is split into two main steps: a) first the algorithm decides upon the user association and traffic routing path and then, b) it places the xNFs of the SFC required by each UE, in the exact order specified by the SFC.

In the first step, every time a new UE SR arrives, PETA constructs a weighted graph and examines all available paths from the source to the destination based on their power consumption. In each path, all feasible wireless and fibre X-Haul transport links are included, as well as the AN link between the serving BS and the UE. The shortest-weighted path, i.e., with the minimum power consumption, is then selected to satisfy the UE demands, as long as the capacity and delay constraints are not violated. In case of a violation, PETA selects the next available shortest path, with no constraint violations and proceeds or, otherwise, if there is no other path to select, PETA blocks the UE and checks for new SR arrivals.

Once a path has been selected, PETA moves to the next step, i.e., the xNF placement. In order to place each xNF of the requested SFC in the available nodes specified by the selected path, the nodes are being sorted by a parameter denoted by Ω, which consists of the node's closeness centrality, the maximum computational capacity and the load of CPU. As for the CPU load, four values are allowed: a) 1 (high priority) when the node has enough computational capacity and can host the xNF without the need for a new xNF instance initiation, b) 0.5 (low priority) when the node has enough computational capacity but needs a new instance initiation to host the xNF, c) 0.1 (very low priority) when the node has enough computational capacity but needs a new instance initiation to host the xNF and the node was previously inactive, and d) 0 (no priority) when the node cannot host the xNF. When sorting is finished, PETA selects the highest ranked node and places the xNF, provided that the computational constraints are satisfied. If such a node cannot be found, PETA returns to the first step and selects the next shortest

path, while repeating the process in the second step until all xNFs are placed or there are no other available paths, in which case it blocks the UE. If all xNFs from the SFC are placed, PETA updates the network state and waits for new SR arrivals, repeating the same steps.

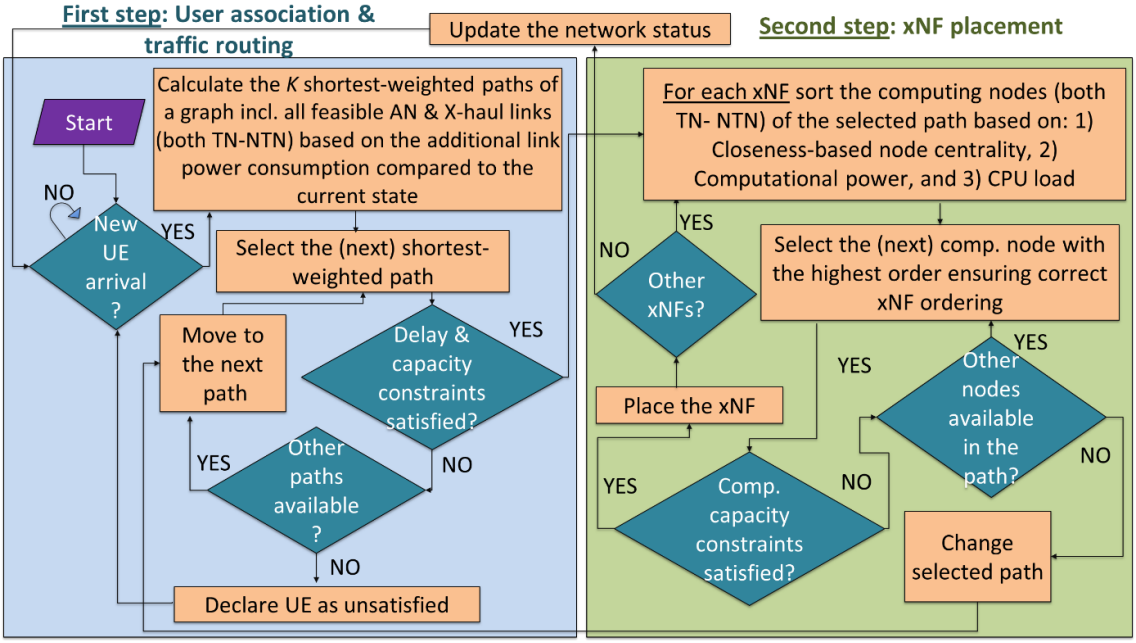


Figure 8.5: PETA's flowchart

8.5 CONCLUSIONS

This chapter has highlighted the pivotal role of sustainability in shaping the development and deployment of 6G networks. A high focus has been given on the European projects' advancements which mainly focus on how to make the 6G networks sustainable, resource-conscious, and environmentally responsible by addressing what is called Sustainable 6G. These advancements may focus on specific parts of the network such as the RAN or the transport or aim to optimize it E2E in terms of energy-efficiency.

Although energy-efficiency by design has been a key sustainability target among European projects, its additional role as an enabler, other than a goal, is expected to gain ground during the next years, towards what is called 6G for Sustainability—leveraging 6G technology to drive sustainable growth across industries. As we look ahead, challenges like optimizing energy use in constrained environments and enhancing network efficiency will demand innovative solutions, with technologies like AI expected to play a significant role. Ultimately, this chapter reinforces the notion that

sustainability will be a foundational pillar of 6G networks, just as 6G will be a vital enabler of sustainability across the digital ecosystem.

8.6 REFERENCES

[6G4S24] 6G4Society project "Towards a sustainable and accepted 6G for Society", 2024. Available: https://6g4society.eu/

[BeG23] BeGreen project "Beyond 5G Artificial Intelligence Assisted Energy Efficient Open Radio Access Network", 2023. Available: https://www.sns-begreen.com/

[BeG24] BeGREEN, D2.2, "Evolved Architecture and Power Enhancement Mechanisms", July 2024, (Online) Available: https://www.sns-BeGREEN.com/deliverables

[BGG+23] M. K. Bahare et al., "The 6G Architecture Landscape - European perspective." Feb 2023. https://doi.org/10.5281/zenodo.7313232

[CGG+23] C. Christofidis, G. Gorgias, H. Georgopoulos, K. Moschopoulos, D. M. Marom and I. Tomkos, "Spatially-Diverse Point-to-MultiPoint Optical Distribution Network for Enhanced 6G Fronthaul," 2023 International Conference on Photonics in Switching and Computing (PSC), Mantova, Italy, 2023, pp. 1-3, doi: 10.1109/PSC57974.2023.10297167.

[Chu24] J. Chuang, "AI In Telcom & RAN - Standard Evolution." Mar 2024. https://www.linkedin.com/pulse/ai-telcom-ran-standard-evolution-jessica-chuang-0pi4c/

[DAE25] H2020 ICT-52 DAEMON, "Network intelligence for aDAptive and sElf-Learning MObile Networks", https://h2020daemon.eu/

[ETH23] ETHER project "ETHER – sElf-evolving terrestrial/non-Terrestrial Hybrid nEtwoRks", 2023. Available: https://ether-project.eu/

[FLE23] FLEX-SCALE project "Flexibly Scalable Energy Efficient Networking", 2023. Available: https://6g-flexscale.eu/en

[HEX23] HEXA-X-II project "A holistic flagship towards the 6G network platform and system, to inspire digital transformation, for the world to act together in meeting needs in society and ecosystems with novel 6G services", 2023. Available: https://hexa-x-ii.eu/

[ITU02] ITU-R S.1590, "Technical and operational characteristics of satellites operating in the range 20-375 THz," ITU-R, Tech. Rep., 2002.

[LGLY21] L. Liu et al., "Joint Dynamical VNF Placement and SFC Routing in NFV-Enabled SDNs," IEEE Transactions on Network and Service Management, vol. 18, no. 4, pp. 4263–4276, Dec. 2021.

[NCK20] G. Noh, H. Chung and I. Kim, "Mobile Relay Technology for 5G," in IEEE Wireless Communications, vol. 27, no. 3, pp. 6-7, June 2020, doi: 10.1109/MWC.2020.9116079

[NTN23] 6G-NTN project "6G Non Terrestrial Networks", 2023. Available: https://6g-ntn.eu/

[ORI24] ORIGAMI project "Optimized Resource Integration and Global Architecture for Mobile Infrastructure for 6G", 2024. Available: https://sns-origami.eu/

[PRO24] PROTEUS-6G project "Programmable Reconfigurable Optical Transport for Efficiently offering Unconstrained Services in 6G", 2024. Available: https://proteus-6g.eu/

[PS21] J. Pérez-Romero and O. Sallent, "Leveraging User Equipment for Radio Access Network Augmentation," 2021 IEEE Conference on Standards for Communications and Networking (CSCN), Thessaloniki, Greece, 2021, pp. 83-87, doi: 10.1109/CSCN53733.2021.9686119

[TPU+24] Tomkos, I., Papapavlou, C., Uzunidis, D., Moschopoulos, K., Muñoz, R., Marom, D.M., & Nazarathy, M. (2024). Towards Multi-Pbps Backbone Optical Networks in Support of Future 6G Networks. CLEO 2024.

9 NETWORK EXPOSURE CAPABILITIES

Network openness has emerged as a significant integral part of 6G networks. A major enabler for this potential has been the network exposure capabilities, though the emerging ecosystem of network Application Programming Interfaces (APIs). Taking advantage of exposure APIs, various advancements enabled, such as deterministic networking, programmability at data and control plane, as well as digital representation of the network infrastructure. Overall, a technological and business osmosis is being conducted, leading to new architectural approaches.

9.1 NETWORK EXPOSURE CAPABILITIES

9.1.1 ENABLING DETERMINISTIC NETWORKING BY EFFICIENTLY BRIDGING MULTIPLE NETWORK DOMAINS

Resource management and control exposure has been a challenging task for multi-domain environments. Each domain often has its own control mechanisms and protocols, which makes the coordination between domains more complex. As network infrastructures evolve toward 6G, there is an increasing need for a framework that allows dynamic, real-time exposure and management of deterministic network capabilities across diverse domains. Such frameworks are indeed the basis to implement pervasive automation mechanisms characterizing the 6G systems. Focusing on deterministic networks, it is crucial for the end-to-end service automation of flow management to have at any moment an accurate picture of the status of the deterministic service parameters e.g., latency/RTT, jitter, data rate etc. in any domain. This challenging objective requires that the collection and exposure of data must happen in real-time and in a synchronized manner. Specific elaborations are furthermore required to calculate the impact of local domain parameters on an E2E Deterministic service provisioned across multiple network technologies.

The proposed architecture builds on IETF DetNet to deliver end-to-end deterministic services across multi-domain environments, integrating diverse technologies such as IEEE 802.1 TSN, 5G TSN, and IP-based networks. By unifying control and data planes, the architecture ensures low latency, minimal jitter, and high reliability across different segments.

- AI-driven Multi-Stakeholder Inter-Domain Control-Plane (AICP). The AICP provides the intelligence needed to manage deterministic paths across different network domains. Using AI/ML algorithms, it dynamically allocates resources, predicts network load, and proactively manages service-level agreements (SLAs). It ensures cross-domain coordination, abstracting technology-specific complexities and providing a unified control interface that seamlessly orchestrates deterministic services across diverse technologies like TSN and 5G. The AICP includes a suite of Management Services (MS) that handle critical tasks such as Time Synchronization, Path Computation, and Service Automation. These services ensure that network devices across domains are synchronized and that deterministic paths are optimally configured. Time-sensitive applications benefit from precise timing, while path computation ensures deterministic flows are prioritized across domains. The monitoring task is addressed by this part of the architecture, where dedicated MSs [1] collect the parameters from the Data Plane (see point 2) exposing them in real-time so that enabling analytics and decision process for service assurance.
- Multi-Domain Data Plane (MDP). The MDP ensures that deterministic traffic can traverse heterogeneous networks while maintaining strict quality-of-service (QoS). It leverages DetNet Layer-3 capabilities to create deterministic paths using Packet Replication, Elimination, and Ordering Functions (PREOF), which prevent packet loss and ensure consistent data delivery across domains. Furthermore, a 3GPP's PDU Set approach has been adopted as Data Unit Groups (DUGs) into an IPv6 header solution, enabling DetNet L3 capabilities across a set of IP packets instead of individual packets within a service flow only [2]. The MDP handles flow-based routing to guarantee minimal latency, coordinating closely with the AICP to maintain deterministic paths across different domains.
- Interoperability Between Domains. A major challenge is ensuring that deterministic services can be maintained across different network domains. The architecture integrates gateways (*DetNet Extended routers*) at domain borders to translate QoS requirements and service parameters, ensuring that traffic retains its deterministic properties when transitioning between technologies like TSN and 5G, as well as doubling as NW-TT and DS-TT Translators between domains. PREOF and DUG mechanisms ensure high reliability by replicating

packets along multiple paths and eliminating duplicates to avoid packet loss and out-of-order delivery.

- Scalability and Extensibility. The architecture is designed to be modular and extensible, with model-driven APIs that allow easy integration of new technologies (e.g., future 6G innovations). This flexibility ensures that deterministic services can scale to meet the demands of complex, large-scale multi-domain environments without sacrificing performance.

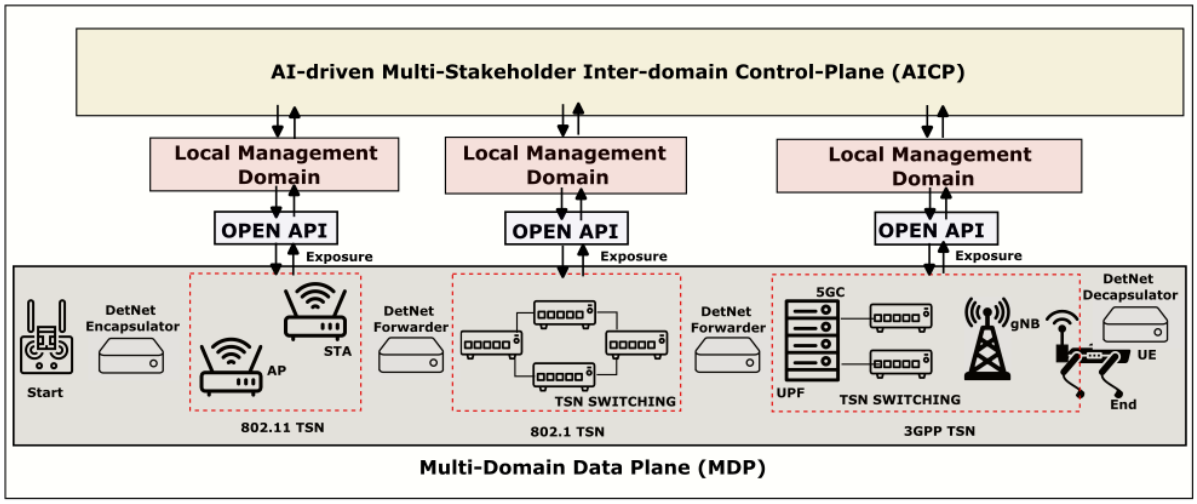


Figure 9.1: System Blueprint of a Multi-Domain Data Plane for Deterministic Networking

9.1.2 API ECOSYSTEM FOR EXPOSURE AND INTERCONNECTION SERVICES

A wide set of RESTful APIs has emerged to support northbound interface, allowing external (third party) systems to easily integrate and automate procedures related to configuration, performance, and fault management. Already a NaaS approach has been described and is being developed by GSMA, CAMARA, TM Forum, and other fora, identifying three main API types.

- The Service APIs provide a purpose-specific capability to third parties, including management APIs, allowing the application developer to run certain management functions from within the application. The CAMARA project has been the major contributor for the definition, development, and validation of the Service APIs.
- The OAM APIs offer programmable access to Operation, Administration and Management (OAM) capabilities to facilitate the integration of the Open Gateway NaaS Platform with portals, marketplaces and other aggregation platforms.

- The Technology-specific APIs refer to operator internal APIs offering programmable access to telco infrastructure and network, service and IT capabilities. These APIs are typically defined in standardization bodies (e.g., 3GPP, IETF, ETSI, TM Forum) and cloud communities (CNCF) and are typically tied to the underlying technology.

SNS JU SoftNet WG, has released a relevant white paper [SoftNet24] where the related API ecosystem and the emerging capabilities are presented. From the architectural perspective, the efforts are led by the Operator Platform group (GSMA) where the target is a framework that can unify the external integration and exposure, allowing operators to offer their services or collaborate with hyperscalers and other service providers.

9.1.3 EXPOSURE SERVICES TO ENABLE ADVANCED EXTENDED REALITY APPLICATIONS

The eXtended Reality (XR) ecosystem is still facing network performance, interoperability, sustainability, and cost barriers when targeting ubiquitous networked services over heterogeneous environments [Mon24].

Novel modular and standards-compliant architectural innovations can be seamlessly integrated to B5G – towards 6G - networks for an enhanced and more flexible support for XR services, by exploiting network exposure, edge computing / federation, and Network as a Service (NaaS) principle, and by additionally abstracting service developers from requiring an in-depth knowledge of underlying technologies and systems, and of associated low-level and domain-specific APIs. Such API services can support Network-assisted Rate Control as well as Edge Selection and Lifecycle Management.

Network-assisted Rate Control. Current Over-the-Top (OTT) rate adaptation mechanisms in XR services can lead to unfairness and stability limitations [Lop24]. A new Network-assisted Rate Control API is envisioned so that an XR Control Plane (CP) Application Function (AF), like an XR Orchestrator [Fer23, Mon24], can subscribe to network and metrics exposure functions informing about service-related Quality of Service (QoS) drops or network-level congestion situations. Upon detection / estimation, two main mitigation actions can be triggered: 1) The XR CP AF can enforce rate adaptations by client-side or in-cloud XR User plane (UP) Afs, based on specific recommendations by a network element with an holistic view of the used resources; 2)

The XR CP AF can request Quality of Service (QOS) to the network, e.g., employing Quality on Demand (QoD) APIs [Mon24], so that the underlying resources (network slices, compute nodes, etc.) are re-configured accordingly.

Edge Selection and Lifecycle Management API. XR services can exploit Edge Computing paradigms to offload processing functions from the clients (thus reducing their computational resources and favouring interoperability and sustainability) and to bring communication modules from far Cloud to close-by Edge servers. Relevant examples include the instantiation of: (i) Multipoint Control Units (MCU) [Fer23] so that smart mixing, transcoding and/or forwarding media functions allow reducing the computational and bandwidth requirements on the client side; (ii) Remote Renderers [Yer24], so that efficient support for untethered and lightweight (e.g., smartphones) XR devices can be provided. In this context, Edge-Cloud APIs [EdgeCloud] allow to discover the available Edge resources, facilitating the selection of the desired ones based on specific criteria (e.g., delay, cost, etc.) and managing the lifecycle of the virtualized AFs (e.g., MCUs, Remote Renderers) to be instantiated in those Edge servers. In addition, the XR CP AF can subscribe to network exposure functions (e.g., NEF), like User Equipment (UE) mobility detection, so that Edge migration and Traffic Influence actions can be triggered to further improve performance, e.g. selection of optimal routing path for lower latency.

9.1.4 EXPOSURE SERVICES TO ENABLE CONNECTED AND AUTOMATED MOBILITY (CAM)

6G aims to expand the set of supported verticals and provide enhanced capabilities beyond connectivity. Already, 5G System (5GS) has been built as a modular architecture to support any vertical running on top in a vertical-agnostic manner. In this context, Connected Automated Mobility (CAM) vertical services, have emerged as a broad range of services in and around vehicles, including both safety-related and other services enabled or supported by the 5GS. However, it is realized that certain verticals (like CAM) have specific and strict requirements. Thus, although significant progress has been made in supporting verticals, the corresponding necessary configuration of the network and end-devices is a time-consuming manual process that requires tight coordination at technical and business levels across the verticals, the vendors, the network operator, and even the end-users. This hinders not only the greater adoption of 5GS but also the uptake of novel CAM Use Cases (UCs) and the modernization of existing ones that require a tighter integration with the underlying network.

Thus, a main objective towards 6G is to open up the reference 5G adv. architecture, and also to transform it into a vertical-oriented with the necessary interfaces tailored to the CAM UCs that i) expose network capabilities to verticals, ii) provide vertical-information to the network; iii) enable verticals to dynamically request and modify certain network aspects in an open, transparent and easy to use, semi-automated way. This requires dedicated APIs that can act as an intermediate abstraction layer that translates the complicated 5GS interfaces and services into easy to consume services accessible by the vertical domain. The experimentation framework and the main innovations developed in the project are: Multi-access Edge Computing (MEC) with service continuity support, zero-touch management, multi-connectivity and predictive Quality of Service (pQoS).

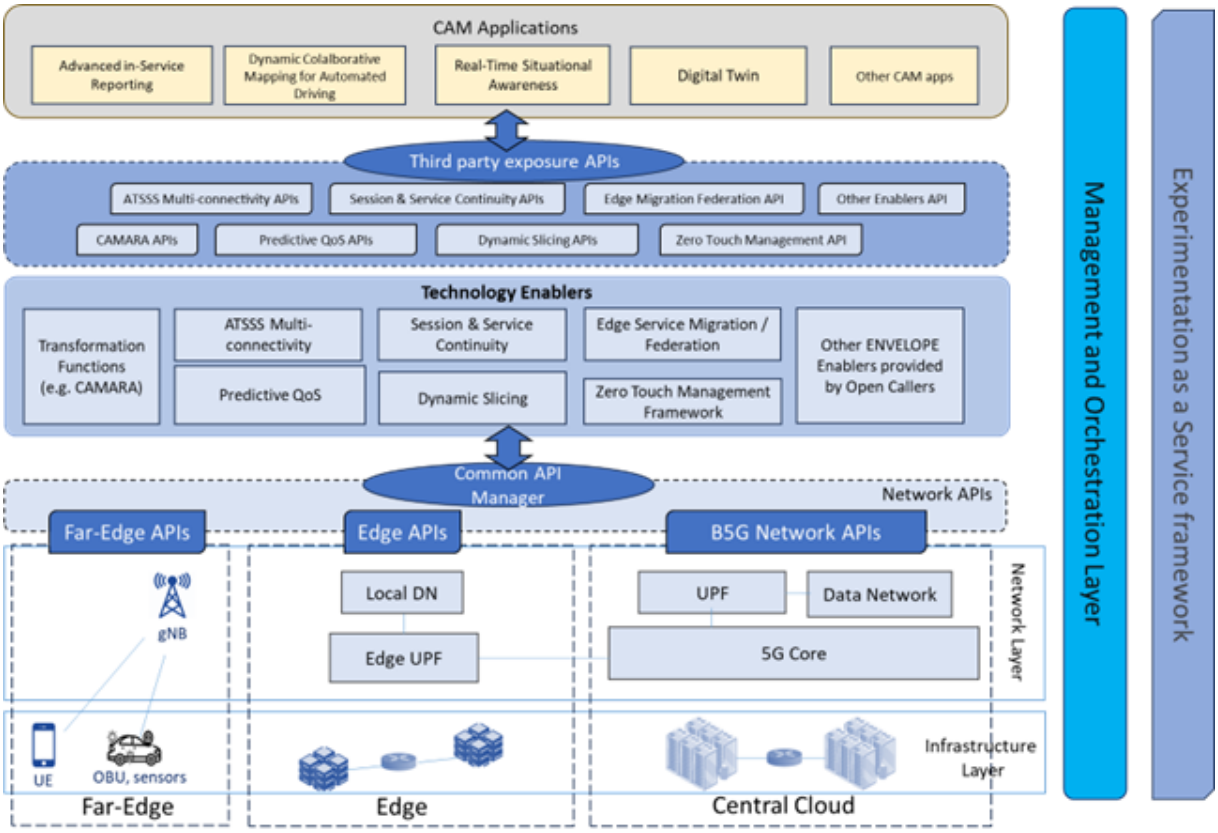


Figure 9.2 Key architectural components for enabling Network-aware CAM applications

9.2 PROGRAMMABILITY ENABLING FEATURES

9.2.1 INFRASTRUCTURE MANAGEMENT LAYER

In 6G network services, packet traffic needs to be forwarded through different NFs. Each NF has a control plane and a data plane. A new architectural component called infrastructure management layer (IML) was proposed in [1] to separate concerns of the packet processing business logic and the infrastructure layer. IML basically acts as a combination of a Virtualized Infrastructure Manager (VIM) and a hardware abstraction (HAL) layer. IML is responsible for managing a pool of resources. IML focuses on the deployment and run-time management of data plane components. An NF data plane component implements the packet processing logic and can be executed on various targets including smartNICs, ASICs, FPGAs, IPUs and DPUs, in addition to traditional CPU resources. IML is responsible for selecting the appropriate target(s) and number of instances to execute the NF data plane and configure the virtual links between them at deployment time. Virtual links are created by infrastructure NFs implementing traffic forwarding and routing between NFs. To enable run-time optimization and hide the underlying optimization from the NF control plane, IML introduces a control plane proxy using a common northbound API (e.g., P4Runtime [2]) that provides a single-instance view of the data plane component to the NF control plane. The proxy hides the underlying data plane optimization like load balancing between multiple data plane instances of the same NF data plane or offloading heavy hitter users to hardware data planes. To enable the better utilization of data plane hardware resources, IML has a subcomponent called P4-MTAGG [3, 4] that is a compiler-based virtualization tool for P4 [5] programmable hardware targets. It enables the deployment and execution of multiple P4 programs on the same P4 hardware in an isolated way. The control plane access to the different data plane programs is also isolated by the IML's control plane proxy component.

9.2.2 PROGRAMMABLE TRANSCEIVERS IN OPTICAL TRANSPORT NETWORKS

The sustainable scaling of the capacity, to support 5G+/6G, requires combining Wavelength Division Multiplexing (WDM) with Space Division Multiplexing (SDM) to exploit the spectral and the spatial dimension of the fibre (i.e., frequencies, cores, and modes) using multicore fibres (MCF), multimode fibres (MMF), or combining cores and modes in few-mode multicore fibres (FM-MCFs), or bundles of SSMFs. A key challenge

is to design and produce a transport network infrastructure able to support beyond 5G and new emerging services, relying on the joint usage of Multi-Band and SDM, spanning the access, aggregation, and metro/long-haul segments, supporting the requirements for X-haul, further integrating the packet/optical and computing layers, and targeting efficient networks in terms of capacity and energy efficiency. In this view, a converged packet-optical transport is need based on resources, so that drastically reducing the presence of aggregation routers and O/E/O conversions, capable of removing boundaries between different network domains and between networks and computing resources.

9.3 NETWORK REPRESENTATION AND FUNCTIONAL STRUCTURE

9.3.1 NETWORK DIGITAL TWIN

The digital twinning concept brings real time monitoring and prediction capabilities down to the network infrastructure. Towards 6G, the integration of the so-called Network Digital Twin (NDT), is expected to operate across three distinct layers: the physical network, the digital network, and a federated simulation framework.

The physical layer remains consistent with existing network elements, such as User Equipment (UE), RAN, and core network, while the digital layer introduces a network twin that allows for dynamic simulation and control.

The digital layer is built upon the ITU-T Y.3090 recommendation [ITU-T-Y3090], which outlines two core model types, basic and functional models:

- A basic model of a network element is the collection of data describing its properties, configurations, and operational status, along with any associated algorithms or protocols used to emulate its dynamics and evolution with time. A basic model of a network is the aggregation of basic models of network elements, including their physical and logical relationships and the interactions that occur between them.
- A functional model of a network builds upon basic models, applying advanced processing techniques, often through AI/ML algorithms, under varying operational scenarios. These models are designed for specific objectives such as performance optimization, anomaly detection, or predictive maintenance.

The third layer, the federated simulation framework, enables the coupling of multiple domain-specific simulators, forming a unified system that allows for large-scale scenario testing. This framework supports both online and offline NDTs, enabling networks to perform "what-if" analyses and refine AI-based functions before deploying them in real-world environments. This is critical for the orchestration of AI-driven services, providing a feedback loop for real-time performance optimization.

9.3.2 NETWORK ABSTRACTION TO SUPPORT TRIALS

When it comes to network exposure for experimentation purposes, network and compute infrastructure should facilitate medium- to long-term experiments without frequent manual reconfiguration and giving experimenters access to internal network configuration. In this context the Trial Network concept has been introduced [Tso24].

A *Trial Network* (TN) represents an end-to-end network with physical and virtual components dedicated to experimentation purposes. The TNs are fully configurable, manageable, controllable and automatically deployable networks combining virtual, physical, and emulated resources.

The TN software components are described in a common repository, called in [Tso24] as *6G Library*, which eases an experimenter to perform a modular and automatic deployment of a Trial Network by selecting on demand the required elements from the library.

The 6G library's objects are curated and designed to serve as the foundational building blocks for building the Trial Networks. For the library implementation Github serves as a sophisticated version control system essential for monitoring changes within computer files, primarily employed in managing source code during software development. Each element within the 6G library embodies the *Everything as a code (EaC)* philosophy, designed as self-contained unit equipped with the necessary automations and scripts for deployment within a network and compute infrastructure.

The architectural blueprint of a component (6G library element) has been meticulously crafted with a focus on simplicity, clarity, versatility, scalability, and adaptability. In this view, every element in the library is becoming Trial Network-ready by complying with a specific predefined toolset (namely Terraform, Ansible, Jenkins, and Ansible) and is hosted under a common folder structure.

At functional level, in [Tso24] a common API framework is defined for all the components of the 6G-library that need to interact at application level with third party software / experimenter.

The brain for the realization of the Trial Network Concept is the Trial Network Life-Cycle Manager (TNLCM). The TNLCM is, within the [Tso24] Architecture, the entity that ensures that every Trial Network in each platform is accessible and in working order, as well as orchestrates the necessary actions required for changing the state of a TN when necessary.

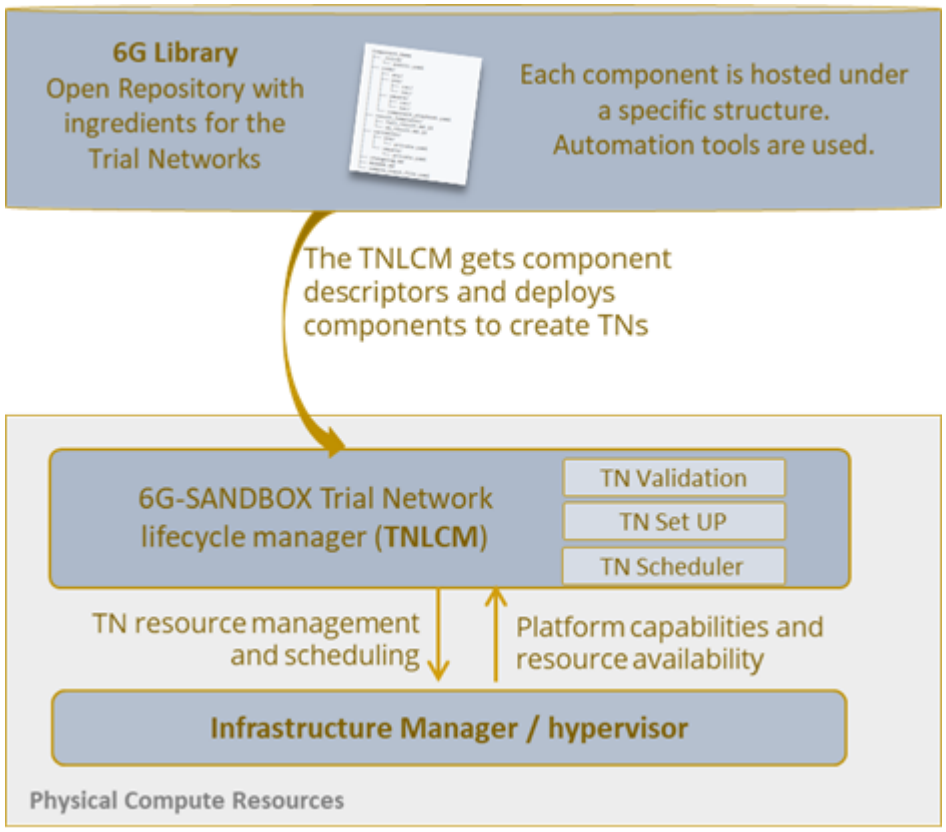


Figure 9.3 Abstract representation of the components that realise the Trial Networks concept

9.3.3 USER CENTRIC NETWORK STRUCTURING

Network services will be recentred on users, following a user-centric approach, distinguishing characteristics of the 6G architecture, enabling user-definition, user-configuration, and user-control. The user-centric architecture in 6G will alter how users, network services, and apps now communicate, which will have an influence on the ownership of personal digital assets, network access, and mobility management. The softwarized telecom service enables deployment of fundamental NFs (including

forwarding, session management, and policy management) without regard to location, in line with the current trends of NF modularization and cloudification. With a shared context and a modular design, the per-user network will do away with message exchange between conventional NFs.

To realize the user-centric 6G network, architectural redesign of the core network is required, following the paradigm shift from "NF-focus" to "user-focus", allowing users to participate in network service creation and operation, while also giving users full control over data ownership. To achieve this design, the network architecture is envisioned to be separated into user service nodes (USN) and network service nodes (NSN), which will be adaptable for activities, such as collaborative sensing and distributed learning in order to spread AI applications on a broad scale across the edge-cloud continuum, in accordance with the EUCloudEdgeIoT [EoCIoT] European initiative.

9.4 REFERENCES

[Fer23] S. Fernandez, M. Montagud, D. Rincón, J. Moragues, G. Cernigliaro, "Addressing scalability for real-time multiuser holo-portation: Introducing and assessing a multipoint control unit (MCU) for volumetric video", ACM Multimedia'23, October 2023

[Lop24] A. López, A. AbdelNabi, D. Camps-Mur, M. Catalan-Cid, M. Montagud, "NetXRate: O-RAN enabled network assisted rate control for XR services", IEEE GLOBECOM'24, December 2024

[Mon24] M. Montagud, et al., "AwareXR: A NaaS architecture to enhance XR services over beyond 5G networks", IEEE Network, To Appear in 2024

[Yer24] I. Yeregui, D. Mejias, G. Pacho, R. Viola, J. Astorga, and M. Montagud, "Edge Rendering Architecture for multiuser XR Experiences and E2E Performance Assessment", IEEE BMSB'24, June 2024

[Tso24] Tsolkas, D., Merino, P., & Dieudonne, M. (2024). THIRD PARTY EXPERIMENTATION: Engagement process and technical information. Zenodo. https://doi.org/10.5281/zenodo.13594165

[SoftNet24] Tsolkas, Dimitris, David Artuñedo Guillen, Anastasius Gavras, Christos Tranoris, Sándor Laki, Antonio Skarmeta Gómez, João Paulo Barraca, George Makropoulos, e Ricard Vilalta. «Network & Service Management Advancements - Key Frameworks and Interfaces Towards Open, Intelligent and Reliable 6G Networks». Zenodo, 28 dicembre 2024. https://doi.org/10.5281/zenodo.14234898.

[EdgeCloud] CAMARA Edge Cloud https://github.com/camaraproject/EdgeCloud/

[ITU-T-Y3090] ITU-T "Digital twin network – Requirements and architecture", 2023

[EoCIoT] The EUCloudEdgeIoT.eu initiative https://eucloudedgeiot.eu/

10 6G ARCHITECTURAL DEFINITION

In the following we analyse and discuss the major architectural trends and opportunities that emerge towards the upcoming standardization work on 6G, starting from the transitions from the 5G technology. To allow for a smooth and faster introduction of 6G services, the main option for migration between 5G and 6G is to use a so called "evolved 5GC", where 6G can reuse existing 5GC NFs if possible and introducing new dedicated 6G NFs to support new 6G functionality, see section 2.3 for more details. In addition, for interworking between 5G and 6G, Multi-Radio Spectrum Sharing (MRSS) is seen as the main option, see Figure 2.2.

10.1 MODULAR ARCHITECTURE DESIGN

10.1.1 MULTI LAYERED ARCHITECTURES FOR NTN INTEGRATION

A core objective of future 6G networks is truly global, resilient coverage. This will be achieved by natively integrating Non-Terrestrial Networks (NTNs) with terrestrial infrastructure, creating a multi-layered 3D architecture. As discussed in Section 4.1, This architecture consists of satellites, High-Altitude Platforms (HAPs), airborne nodes, and traditional ground-based cells. This discussion explores the vision, challenges, and enabling technologies behind these multi-layered NTNs, emphasizing how converging orbital, aerial, and terrestrial segments can deliver ubiquitous connectivity.

The concept of a 3D multi-layered NTN architecture unifies terrestrial networks with various non-terrestrial nodes, such as satellites and HAPs at different altitudes. This multi-layered configuration expands coverage to remote and sparsely populated regions, including maritime areas, and provides backup and resiliency for critical communications. Key design considerations include multi-orbit integration, combining Geostationary (GSO) and Non-Geostationary Orbit (NGSO) satellites (LEO, MEO) for both wide-area broadcast services and low-latency links. Flexible airborne nodes, like HAPs, specialized drones, or other aerial platforms, can be deployed opportunistically to manage capacity surges or event-based connectivity, rather than establishing permanent global coverage layers. Unified radio and access are expected, with native convergence at the radio interface, enabling handsets and IoT devices to seamlessly connect to terrestrial gNBs or non-terrestrial nodes. By combining multiple layers, each optimized for different coverage and performance targets, 6G networks can

dynamically adapt resources to localized needs while maintaining wide-reaching coverage.

Multi-layered architectures inherently enhance resiliency by offering failover paths and diverse connectivity options. In emergencies or disasters where terrestrial networks fail, satellites or HAPs can maintain at least minimal service, providing backup infrastructure. Store and forward mechanisms, using periodic or intermittent satellite links, enable local data storage and forwarding upon re-establishing ground connectivity, which is vital for delay-tolerant IoT applications like agriculture, livestock tracking, and maritime operations. Multi-layer redundancy, using multi-orbit constellations (GEO and LEO), ensures that if one link is compromised due to congestion or poor channel conditions, other layers can carry essential traffic.

To accommodate seamless integration, 3GPP is developing standards to support non-terrestrial access within the 6G system. Distributed 5G/6G core functionalities (AMF, SMF, UPF) can be partially hosted in space to reduce latency and manage intermittent feeder links. A split-core architecture, with both on-ground and on-board NFs, allows flexible deployment, especially for low-density satellite constellations. A hierarchical MANO framework will cover ground, aerial, and space infrastructure. Domain-level orchestrators manage local resources, while a global-level orchestrator coordinates cross-domain and mobility management, ensuring smooth handovers as nodes traverse orbits or airspace. Extending SDN principles to the user plane enables fine-grained forwarding control across multiple domains and network segments, simplifying end-to-end provisioning of data paths, including satellite feeder links and terrestrial backhaul.

This unified 3D architecture positions 6G to serve diverse use cases. It enables direct smartphone connectivity, extending 6G coverage to remote areas with a consistent user experience, potentially including limited indoor connectivity. For delay-tolerant IoT, large revisit times of low-density LEO constellations and store-and-forward mechanisms support agriculture monitoring, livestock management, maritime navigation, and asset tracking. Flexible broadband connectivity through ultra-small aperture terminals, suitable for in-vehicle, drone-mounted, or airborne platforms, is provided for vehicle and drone broadband. Finally, it supports safety-critical operations by connecting aviation and space control systems, ensuring uninterrupted data exchange even when conventional terrestrial links are unavailable.

10.1.2 TIME CRITICAL AND DETERMINISTIC NETWORKING INTEGRATION

Emerging 6G applications, such as extended reality (XR), smart farming, and adaptive manufacturing, require end-to-end time-critical communications with stringent latency, reliability, and deterministic performance. Achieving these goals in inherently stochastic wireless systems, especially when also integrating compute elements (like edge computing) and deterministic networking technologies (like TSN and DetNet) is a significant challenge.

Future 6G networks must allow to incorporate stochastic components, accepting that latency variation, packet loss, and jitter are inevitable in wireless and distributed compute environments. These networks must also characterize and predict these stochastic behaviours, such as latency distributions and reliability levels, to enable proactive resource planning and management, mitigating variances using mechanisms like packet delay correction or buffering to offset jitter and keep time-critical packets within acceptable bounds.

Dependable time-critical services must account for both network and computational latency introduced by edge or cloud processing. Time-aware edge computing, embedding time-awareness into compute infrastructure (like time-synchronized edge nodes), ensures tasks are scheduled and completed within predictable deadlines. This requires that end-to-end feedback loops are established, considering the entire chain from sensors and controllers to actuators, allow the network to dynamically adjust resource allocations and QoS parameters in real time, adapting to latency variations within any sub-component and tune application-level components according to the system load or changes in network characteristic.

To unify deterministic and non-deterministic domains, a horizontal 6G architecture should expose network capabilities to higher-layer applications through a well-defined service interface. Performance monitoring and prediction, using continuous monitoring of KPIs (latency, packet delay variation, reliability), enables data-driven forecasting of whether the network can sustain the requested QoS. Time-aware configuration allows network control entities to orchestrate resources, enforcing tight latency bounds across TSN/DetNet and wireless sections, creating consistent performance "islands" end to end.

Future 6G systems will integrate TSN/DetNet-based deterministic networks with wireless stochastic segments and compute nodes. This requires holistic traffic handling, where TSN or DetNet understands each sub-component's latency characteristics, including variable wireless links, edge compute tasks, and buffering points. Scalable architectures are necessary, as algorithms and protocols must handle large-scale deployments of sensors, machines, and devices. Architectural components should plan resources based on probabilistic performance while meeting application-specific reliability targets. End-to-end orchestration, through a unified management framework, ensures cross-domain configuration, monitoring, and adaptation at scale, bridging traditional deterministic subnetworks with next-generation, AI-driven wireless domains.

As discussed in this white paper "deterministic networking" will transition into a dependable time-critical communications paradigm, where network performance becomes predictable and can be matched to the application needs. This is maintained through time-awareness, network performance observability, adaptive management, and security-by-design. Time-awareness ensures consistent, synchronized operations across network and compute layers. Adaptive management allows fine-tuned responses to real-world variations in traffic, mobility, and compute load. Security-by-design recognizes the critical nature of time synchronization and performance assurances. By orchestrating communications and computations holistically, 6G will deliver next-level support for demanding verticals, reliably connecting sensors, controllers, and actuators across diverse domains and advancing a new era of networked intelligence.

10.1.3 INTEGRATION OF SENSING AND DIGITAL TWINNING

Building on the vision of Integrated Sensing and Communication (ISAC), future 6G networks will incorporate distributed sensing architectures spanning heterogeneous devices, reconfigurable surfaces, and multi-modal data sources. Achieving time-critical and deterministic performance in such complex environments requires an end-to-end design that accounts for varying capabilities, semantic-aware processing, and advanced orchestration of communication, computation, and sensing resources.

A core objective of distributed sensing is to collect, fuse, and exploit data from multiple heterogeneous devices (Sensing Receiver Nodes, SRNs) to track both passive and active targets over large areas. To optimize distributed sensing under the high

dimensionality of multi-modal data, a semantic plane is introduced. Its key functionalities include context extraction, interpreting and managing the "meaning" of data rather than raw bits, reducing overhead and improving relevance. It also includes dynamic adaptation, aligning sensing, communication, and computation tasks with specific system goals (e.g., continuous target tracking, minimal latency). Semantic modules and interfaces ensure that all nodes cooperate under common semantic goals, enabling flexible data sharing and efficient resource utilization.

Future networks will expose sensing-as-a-service capabilities to internal NFs and third-party applications. A dedicated Sensing Management Function (SeMF), or an extension of existing location services, will coordinate sensing procedures, manage data flows, and enforce privacy and security. Secure interfaces for collecting, processing, and distributing sensing data will accommodate dynamic trust levels and avoid network overload.

A way of achieving precise network configuration without overloading the network is to introduce AI-Driven Network Digital Twins (NDTs) which offer a virtualized replica of the physical network and enable predictive optimization. Closed-loop management, using simulations running in real time (online NDT) or offline "what-if" scenarios, will inform dynamic resource reallocation or reconfiguration. Federated simulation, integrating multiple domain-specific simulators (RAN, optical, compute, etc.), will allow large-scale scenario testing. MLOps principles will streamline the design, training, and deployment of AI models across network domains, ensuring consistent performance monitoring and retraining based on real measurements.

10.1.4 SUSTAINABILITY MANAGEMENT

Enforcing sustainability in the network operation has to be achieved with specific architectural components [NGMN-GREEN], such as a Sustainability Monitoring Plane (SMP) designed to manage different sustainability needs coming from telecommunications networks, industry verticals, end users, and associated services in a 6G network. This includes comprehensively addressing six dimensions of sustainability: environmental (resource efficiency, energy consumption), economic (cost-effectiveness, profitability), societal (user accessibility, inclusivity), technological (innovation, reliability), regulatory (compliance, standards adherence), and ethical (privacy, transparency). This generalized monitoring plane is similar to well-known control and data planes, and will continuously gather and analyse data from multiple

sources, such as industrial sectors, consumer activities, network operators, and smart grid operators.

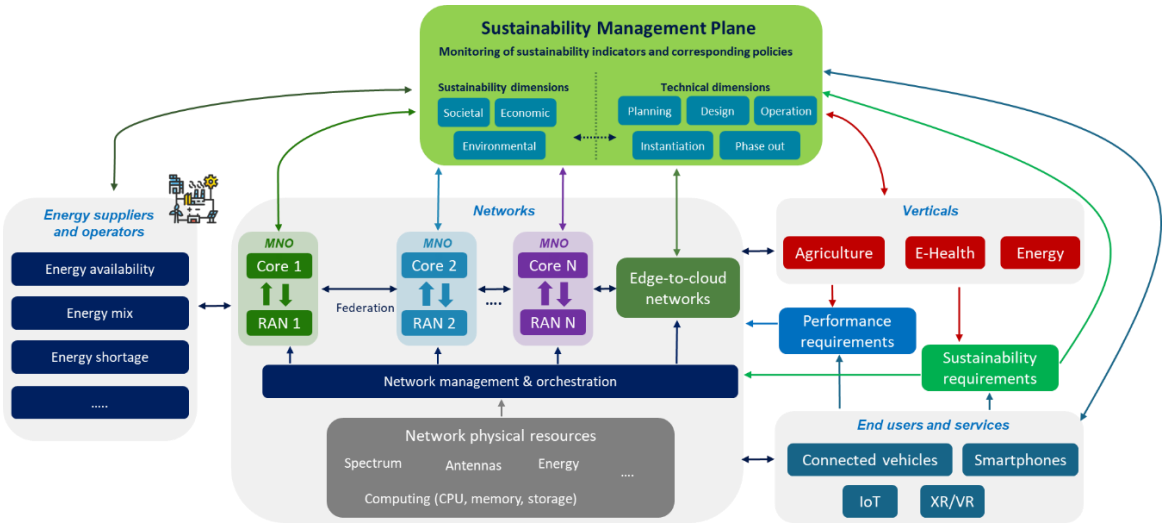


Figure 10.1 SMP overarching networks, verticals, end user, and energy suppliers to holistically monitor sustainability indicators

As illustrated in Figure 10.1, the SMP will establish bidirectional communication with multiple stakeholders—including network operators, vertical industries, end users, and energy suppliers—to systematically collect and evaluate sustainability indicators spanning environmental, societal, and economic dimensions. Hence, the work on the definition of the SMP shall work on:

- Functionalities and Interfaces: SMP capabilities and its integration with diverse systems and domains.
- Data Exchange: Relevant information flows among stakeholders to effectively monitor sustainability indicators at different temporal scales.
- Sustainability Metrics: Criteria for assessing and evaluating the various sustainability dimensions.
- Data Management Policies: Guidelines governing data sharing and handling practices aimed at enhancing overall sustainability efforts.

10.1.5 GLOBAL SERVICE BASED ARCHITECTURE

The application of the SBA architecture opens the way for the possibility of promoting it towards other domains in the network. The capability of efficiently promote the consumer producer paradigm in a data-driven manner makes this technology a good candidate for its inclusion in other domains beyond Core, as in 5G. In particular,

integrating the GSBA concept in the RAN has the potential to improve several aspects of the radio access network.

- Improved Scalability: SBA decouple NFs into modular services, allowing networks to scale dynamically based on demand. This flexibility is essential for handling traffic spikes or expanding capacity without significant hardware investments.
- Enhanced Flexibility and Modularity: The modular design of SBA enables independent development, deployment, and management of NFs. This approach supports agile updates and innovation without disrupting the entire network.
- Improved Automation and Orchestration: SBA supports advanced automation tools and orchestration frameworks. By using programmable interfaces and machine-readable APIs, networks can automate tasks like resource allocation, fault detection, and recovery.
- Better Resource Management: Fine-grained control over individual services enables better monitoring and allocation of resources.
- Resilience and Reliability: SBA supports fault-tolerant designs where failures in one service do not cascade across the network. This architecture improves the overall reliability and uptime of the network.
- Future-Proofing: With its modular and API-driven approach, SBA is well-suited to adapt to evolving standards, protocols, and technologies, ensuring long-term relevance and reduced need for overhauls.

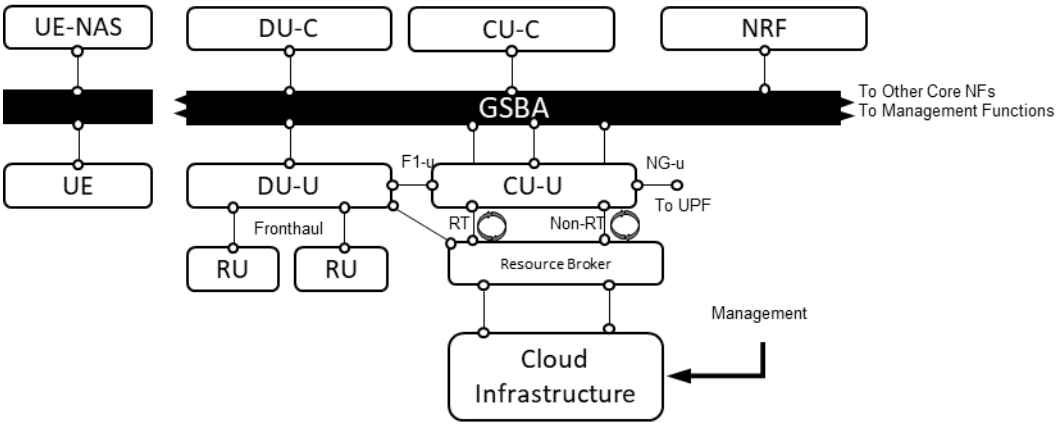


Figure 10.2 A possible architecture for the GSBA integration in the RAN

10.2 CLOUD CONTINUUM MANAGEMENT

The 6G network architecture will be very integrated, glueing together advanced connectivity and computational power from the most deep-edge sub-networks, like those found in vehicles or robots, through edge sites, and all the way to cloud infrastructure. This integration allows for a seamless flow of data and processing, enabling tasks like autonomous driving or robotic control to be executed with optimal efficiency.

To achieve this, the proposals listed in this white paper suggest the utilization of a resource pool across this continuum, where tasks can be dynamically offloaded to the most suitable location and key to this vision are new architectural enablers. The Communication & Computing Resources Exposure Function (CCREF) extends network exposure to provide real-time awareness of diverse resources, including computing, connectivity, and AI, across the entire 6G network. The Communication & Computational Resources Management Function (CCRMF) orchestrates advanced resource-sharing policies, dynamically allocating capacity to meet performance demands. Network Intelligence Functions (NIF) leverage AI and machine learning for proactive network management, enhancing existing analytics. The introduction of a Compute Continuum Layer (CCL) abstracts heterogeneous computing elements, ensuring seamless access to optimal resources for NFs and applications, leading to performance and energy efficiency gains. Additionally, Trusted Execution Environments (TEEs) provides a hardware-based secured platform for confidential computing, reinforcing integrity and confidentiality of data as it moves across the network continuum.

Besides these novel components, multi-access Edge Computing (MEC) is enhanced by defining application slices that incorporate both network and computing requirements, working in tandem with network slicing to fulfill diverse QoS and compute demands. AI/ML-driven orchestration enables real-time data analytics, proactive resource scaling, automated healing, and threat detection, with distributed and federated learning ensuring privacy and efficiency. Zero-touch closed loops support self-configuration and optimization, while hierarchical orchestrators coordinate across administrative boundaries, allowing all devices, from resource-constrained IoT to cloud sites, to participate in collaborative processing. In essence, this 6G management approach unites deep edge, edge, and cloud resources into a unified framework, facilitating the dynamic and efficient allocation of connectivity and computing

resources, thus ensuring high performance, flexibility, and sustainability for a wide array of services.

10.3 INTEROPERABILITY AND GLOBAL OPERATION

6G architecture should enable zero-trust principles to enforce how infrastructure capabilities are shared and consumed across diverse stakeholders and domains. By integrating a Zero-Trust Layer (ZTL) into the architecture, the network achieves granular security, continuous evaluation, advanced analytics, and adaptable business models, moving beyond traditional perimeter-based trust systems.

This approach fosters cooperative control between network operators and service providers, mirroring hyperscale cloud operational models. Instead of requiring complete trust, participants share only essential performance and analytics data through managed interfaces. This ensures secure feedback loops, feeding service provider data into the Network Data Analytics Function (NWDAF) for personalized network optimization without compromising confidential business information. This cooperative loop enables continuous optimization, automating resource adjustments to meet each provider's unique quality of experience (QoE) metrics. The ZTL framework emphasizes both vertical and horizontal exposure, catering to a wide range of use cases.

In this context, vertical exposure integrates feedback from service providers into AI-driven network analytics and management, allowing providers to adjust NFs, slices, or other configurations to align with their application-level metrics. Horizontal exposure facilitates global operation in multi-operator scenarios, such as international roaming, enabling direct, secure interaction among different network operators and service providers without relying on vulnerable trust-based models.

A decentralized identity model replaces legacy roaming agreements, enabling visited operators to directly charge global end-users while providing full visibility to the home operator. This eliminates inefficient data routing, reducing latency and transport costs, and fosters new business relationships among operators, hyperscalers, and vertical providers. By decoupling user identity, integrating AI/ML-driven analytics, and utilizing distributed ledgers, this approach paves the way for real-time, zero-trust operations in future communications networks.

Besides, in line with Zero Trust principles, decentralized identity management can be integrated within an authentication framework that operates seamlessly across multiple

proprietary domains. Proposed framework leverages attribute derivation techniques. Based on the self-sovereign identity paradigm, attributes ensures that only the minimal and essential information required to validate access to network resources and services is disclosed. It safeguards customer identity, mitigating traceability and linkability while supporting privacy-preserving operations. From an operator's perspective, the architecture incorporates dedicated issuers and verifiers to manage part of the credential lifecycle. Issuers generate and cryptographically sign verifiable credentials containing the derived attributes, thereby providing a secure proof of identity and access rights. Verifiers, strategically deployed across the network, authenticate these credentials in real time, ensuring that only users with valid credentials can access network services.

This new wave of authentication schemes is the principal enabler for operator-agnostic access, allowing users to select and access services from any operator based on the current network state and service requirements. Additionally, the framework marks a significant advancement for proximity services, delivering fast and privacy-preserving authentication that supports highly dynamic service compositions driven by users' locations.

10.4 AI DRIVEN NETWORK MANAGEMENT AND ORCHESTRATION

With the advent of 6G, network management and AI-driven orchestration are key components for accomplishing autonomous and dynamic network operations. In this whitepaper we outline the principal role of AI to enable, Intent-based closed-loop management, where ongoing monitoring, analysis, decision-making, and execution are integrated within an orchestration framework

In presented approaches, AI is not only responsible for automation but also allows for comprehensive operational management of NFs and resources distributed across the cloud continuum (Far-edge, RAN, Edge, transport, Core, etc.). A critical function is the integration with advanced AI/ML frameworks, comprising MLOps for continuous model deployment and retraining, and DataOps for robust, high-quality data pipelines, which work together to ensure that real-time telemetry is accurately processed and actionable insights are derived. Key to the concept is the use of intent-based management. High-level service intents, reflecting user requirements and network

policies, are translated through standardized APIs into detailed network and service configurations. This process is supported by a distributed, multi-agent orchestration model that coordinates the control of both intra- and inter-domain resources. The architecture supports flexible closed-loop mechanisms, allowing its features to be expanded as new technologies or tools are developed, as well as proactive fault management, domain optimization, resource reassignment, and service migration, which are capabilities necessary to adapt the network to the continuous changes it faces.

Furthermore, the AI-driven orchestration framework natively incorporates elements of trustworthiness. Starting from intent modelling, involving the five dimensions of trustworthiness to express requirements. Passing through explainable AI, which is embedded in the system to provide transparency in the decision-making process, enabling oversight by management teams. In addition to privacy methods implemented to ensure that sensitive data remains secure throughout the operation's lifecycle, especially useful in federation scenarios. Moreover, integration of digital twin technologies enables real-time simulation and reliable predictions of KPIs; these digital twins allow for the evaluation of measures and countermeasures before they are applied, as well as enabling the training of ML agents at an unprecedented speed.

10.5 REFERENCES

[NGMN-GREEN] NGMN, "Green Future Networks: A Roadmap to Energy Efficient Mobile Networks", June 2024.

[SUST6G] SUSTAIN-6G project "SUSTainability Advanced and Innovative Networking with 6G", 2025. Available: https://sustain-6g.eu/

11 ABBREVIATIONS AND ACRONYMS

Acronym	Description
3GPP	3rd Generation Partnership Project
4G	Fourth Generation Mobile Network
5G	Fifth Generation Mobile Network
5GC	5G Core
5GCN	5G Core Network
5GS	5G System
6G	Sixth Generation Mobile Network
AAA	Authentication, Authorization, and Accounting
ABAC	Attribute-Based Access Control
ADR	Adaptive Data Rate
AF	Application Function
AI	Artificial Intelligence
AICP	AI-driven Multi-Stakeholder Inter-Domain Control-Plane
AMF	Access and Mobility Management Function
AN	Access Network
AOA	Angle of Arrival
AOSP	Android Open Source Project
AP	Access Point
API	Application Programming Interface
AR	Augmented Reality
ASIC	Application-Specific Integrated Circuit
ATSSS	Access Traffic Steering, Switching, and Splitting
B5G	Beyond 5G
BC	Blockchain
BCN	Blockchain Network

BMSB	Broadcast and Multicast Service Broadcast
BS	Base Station
BSM	Basic Safety Message
BVT	Bandwidth Variable Transceivers
CA	Carrier Aggregation
CAM	Cooperative Awareness Message
CAMARA	Cloud Application Management for Open Networks
CCL	Compute Continuum Layer
CCREF	Communication & Computing Resources Exposure Function
CCRMF	Communication & Computational Resources Management Function
CDN	Content Delivery Network
CF	Computing Function
CL	Cloud Layer
CN	Core Network
CNCF	Cloud Native Computing Foundation
CNF	Cloud-Native Function
CNSM	Conference on Network and Service Management
COH	Coherent Optical
COTS	Commercial Off-The-Shelf
CP	Control Plane
CPU	Central Processing Unit
CS	Communication Service
CSA	Cloud Service Architecture
CSMF	Communication Service Management Function
CTI	Cyber Threat Intelligence
CU	Centralized Unit
DAR	Data Analysis and Reporting
DASH	Dynamic Adaptive Streaming over HTTP

DC	Data Center
DD	Data Distribution
DFT	Discrete Fourier Transform
DI	Data Integration
DL	Downlink
DLT	Distributed Ledger Technology
DMMF	Domain Mobility Management Function
DMOC	Domain Management & Orchestration Component
DPU	Data Processing Unit
DSCM	Digital Subcarrier Multiplexing
DSP	Digital Signal Processing
DSS	Dynamic Spectrum Sharing
DT	Digital Twin
DTD	Document Type Definition
DU	Distributed Unit
DUG	Data Unit Groups
DWDM	Dense Wavelength Division Multiplexing
E2E	End-to-End
E2SM	E2 Service Model
EC	Edge Computing
ECU	Electronic Control Unit
EMOC	E2E Management & Orchestration Component
EN	Edge Node
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
EUC	End User Computing
FAV	Fully Autonomous Vehicles
FC	Forwarding Controller

FG	Focus Group
FGCS	Future Generation Computer Systems
FIPS	Federal Information Processing Standards
FL	Federated Learning
FLEX	Flexible Networks
FM	Fault Management
FPGA	Field Programmable Gate Array
FR1	Frequency Range 1
FR2	Frequency Range 2
FSO	Free-Space Optics
FTM	Fine Timing Measurement
FTTH	Fiber to the Home
FWA	Fixed Wireless Access
GDPR	General Data Protection Regulation
GEO	Geostationary Orbit
GH	Green Hydrogen
GIS	Geographic Information System
GMMF	Global Mobility Management Function
GND	Ground
GNSS	Global Navigation Satellite System
GPU	Graphics Processing Unit
GSBA	Global Service-Based Architecture
GSMA	GSM Association
GSO	Geostationary Satellite Orbit
GW	Gateway
HAL	Hardware Abstraction Layer
HAP	High Altitude Platform
HC	Higher Capabilities

HLS	High-Level Synthesis
HMP	Hybrid Multiplexing Processing
HPEC	High-Performance Embedded Computing
IAB	Integrated Access and Backhaul
IBI	Information-Based Interface
IBN	Intent-Based Networking
IBTM	intent-based threat mitigation module
ICAS	Integrated Communication and Sensing
ICT	Information and Communication Technology
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IML	Interactive Machine Learning
IMT	International Mobile Telecommunications
IN	Intelligent Network
IOTM	Internet of Things Management
IP	Internet Protocol
IPU	Intelligence Processing Unit
IR	Infrared
IRTF	Internet Research Task Force
ISAC	Integrated Sensing and Communication
ISL	Inter-Satellite Link
ISO	International Organization for Standardization
IT	Information Technology
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JSON	JavaScript Object Notation
KPI	Key Performance Indicator

LAN	Local Area Network
LBS	Location-Based Services
LIDAR	Light Detection and Ranging
LLS	Lower Layer Split
LPWAN	Low-Power Wide-Area Network
MAC	Medium Access Control
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MLOps	Machine Learning Operations
MME	Mobility Management Entity
MNO	Mobile Network Operator
MRF	Media Resource Function
MRSS	Multi-Radio Spectrum Sharing
NAS	Non-Access Stratum
NF	Network Function
NFV	Network Function Virtualization
NIC	Network Interface Card
NR	New Radio
NSA	Non-Standalone
NSMF	Network Slice Management Function
NTN	Non-Terrestrial Networks
OAM	Operations, Administration, and Maintenance
O-RAN	Open Radio Access Network
OSS	Operations Support System
OTFS	Orthogonal Time Frequency Space
PDU	Protocol Data Unit

PHY	Physical Layer
PKI	Public Key Infrastructure
PON	Passive Optical Network
PPDR	Public Protection and Disaster Relief
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RF	Radio Frequency
RIC	RAN Intelligent Controller
RLC	Radio Link Control
RRM	Radio Resource Management
SBA	Service-Based Architecture
SDN	Software-Defined Networking
SIM	Subscriber Identity Module
SLA	Service Level Agreement
SMF	Session Management Function
SON	Self-Organizing Network
SMP	Sustainability Management Plane
SRv6	Segment Routing over IPv6
TCO	Total Cost of Ownership
TN	Terrestrial Network
TPM	Trusted Platform Module
TSN	Time-Sensitive Networking
UE	User Equipment
ULP	Ultra-Low Power
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communication

V2X	Vehicle-to-Everything
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VN	Virtual Network
VR	Virtual Reality
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
XAI	Explainable Artificial Intelligence
XR	Extended Reality
ZTA	Zero Trust Architecture

12 LIST OF EDITORS & REVIEWERS

Name	Company / Institute / University	Country
Document editors		
Marco Gramaglia	Universidad Carlos III de Madrid	Spain
Ömer Bulakci	Nokia	Germany
Xi Li	NEC Laboratories Europe	Germany
Anastasius Gavras	Eurescom GmbH	Germany
Chapter 2 editors		
Mårten Ericson	Ericsson	Sweden
Sylvaine Kerboeuf	Nokia Bell Labs	France
Chapter 3 editors		
Marco Gramaglia	Universidad Carlos III de Madrid	Spain
David Larrabeiti	Universidad Carlos III de Madrid	Spain
Chapter 4 editors		
Mir Ghoraishi	Gigays	UK
Agapi Mesodiakaki	Aristotle University of Thessaloniki	Greece
Chapter 5 editors		
Mårten Ericson	Ericsson	Sweden
Anastasius Gavras	Eurescom GmbH	Germany
Harilaos G. Koumaras	NCSR Demokritos	Greece
Chapter 6 editors		
Xi Li	NEC Laboratories Europe	Germany
Sebastian Robitzsch	InterDigital	UK
Chapter 7 editors		
Rodrigo Asensio Garriga	University of Murcia	Spain
Antonio Skarmeta	University of Murcia	Spain
Harilaos G. Koumaras	NCSR Demokritos	Greece

Chapter 8 editors		
Anastasius Gavras	Eurescom GmbH	Germany
Agapi Mesodiakaki	Aristotle University of Thessaloniki	Greece
Chapter 9 editors		
Dimitris Tsolkas	Fogus Innovations & services	Greece
Chapter 10 editor		
Marco Gramaglia	Universidad Carlos III de Madrid	Spain

Name	Company / Institute / University	Country
External Reviewer		
Nazli Güney	Turk Telekom	Türkiye

13 LIST OF CONTRIBUTORS

Name	Company / Institute / University	Country
Adam Zahir	Universidad Carlos III de Madrid	Spain
Adriana Fernández-Fernández	i2CAT Foundation	Spain
Agapi Mesodiakaki	Aristotle University of Thessaloniki	Greece
Alessandro Guidotti	CNIT	Italy
Alessandro Vanelli-Coralli	CNIT / University of Bologna	Italy
Alexandros Kostopoulos	Hellenic Telecommunications Organization (OTE) Group	Greece
Almudena Diaz Zayas	University of Malaga	Spain
Amr AbdelNabi	i2CAT Foundation	Spain
Anastasios Zafeiropoulos	Institute of Communication and Computer Systems	Greece
Anastasius Gavras	Eurescom GmbH	Germany
Andra Lutu	Telefonica	Spain
Andres Saavedra	NEC Laboratories Europe	Germany
Angelos Antonopoulos	Nearby Computing	Spain
Anna Tzanakaki	UOA	Greece
Anna Umbert	Universitat Politecnica de Catalunya	Spain
Antonio de la Oliva	Universidad Carlos III de Madrid	Spain
Antonio Skarmeta	University of Murcia	Spain
Aurora Ramos	Capgemini	Spain
Baldomero Coll-Perales	Universidad Miguel Hernandez de Elche	Spain
Benjamin Barth	German Aerospace Center (DLR)	Germany
Carlos J. Bernardos	Universidad Carlos III de Madrid	Spain
Christoph Schmelz	Nokia	Germany
Christoph Sommer	TU Dresden	Germany

Christos Verikoukis	Industrial Systems Institute	Greece
Chrysa Papagianni	University of Amsterdam	The Netherlands
Dan Marom	HUJI	Israel
Daniel Camps-Mur	i2CAT Foundation	Spain
Daniel Kilper	Trinity College Dublin	Ireland
David Larrabeiti	Universidad Carlos III de Madrid	Spain
David Rico Menendez	Universidad Carlos III de Madrid	Spain
Diego San Cristobal	Ericsson	Spain
Dimitris Pliatsios	University of Western Macedonia	Greece
Dimitris Tsolkas	Fogus Innovations & services	Greece
Elli Kartsakli	Barcelona Supercomputing Center	Spain
Emilio Calvanese Strinati	CEA	France
Fabrizio Granelli	CNIT	Italy
Fernando Agraz	Universitat Politecnica de Catalunya	Spain
Filippo Cugini	CNIT	Italy
George Alexandropoulos	NKUA	Greece
George Xilouria	Space Hellas	Greece
Georgios Gardikis	Space Hellas	Greece
Gergely Pongracz	Ericsson Research	Hungary
Giada Landi	Nextworks	Italy
Gilberto Berardinelli	Aalborg University	Denmark
Giyyarpuram Madhusudan	Orange	France
Gourav Prateek Sharma	KTH Royal Institute of Technology	Sweden
Harilaos G. Koumaras	NCSR Demokritos	Greece
Henk Wymeersch	Chalmers University	Sweden
Ignacio Labrador Pavon	ATOS	Spain

Ioannis Chochliouros	Hellenic Telecommunications Organization (OTE) Group	Greece
Ioannis Tomkos	UPAT	Greece
Ion Turcanu	Luxembourg Institute of Science and Technology	Luxembourg
Irene Vila	Universitat Politecnica de Catalunya	Spain
James Gross	KTH Royal Institute of Technology	Sweden
Jarno Pinola	VTT	Finland
Javier Garcia Rodrigo	Telefonica	Spain
Javier Gozalvez	Universidad Miguel Hernandez de Elche	Spain
Joachim Sachs	Ericsson Research	Sweden
Joan A. Ruiz-de-Azua	i2CAT Foundation	Spain
John Cosmas	Brunel University	UK
Jordi Perez-Romero	Universitat Politecnica de Catalunya	Spain
Jose A. Ayala-Romero	NEC Laboratories Europe	Germany
Jose A. Lazaro	Universitat Politecnica de Catalunya	Spain
Jose M. Alcaraz Calero	University of the West of Scotland	UK
Josep Maria Fabrega	CTTC	Spain
Josep Martrat	Eviden	Spain
Josep Xavier Salvat Lozano	NEC Laboratories Europe	Germany
Joao Fernandes	OneSource Consultoria Informática	Portugal
Juan Sanchez-Gonzalez	Universitat Politecnica de Catalunya	Spain
Janos Harmatos	Ericsson Research	Sweden
Kyriakos Vlachos	CNIT	Italy
Konstantinos Ntontin	University of Luxembourg	Luxembourg
Kostas Ramantas	Iquadrat	Spain
Lanfranco Zanzi	NEC Laboratories Europe	Germany
Lechoslaw Tomaszewski	Orange Polska	Poland

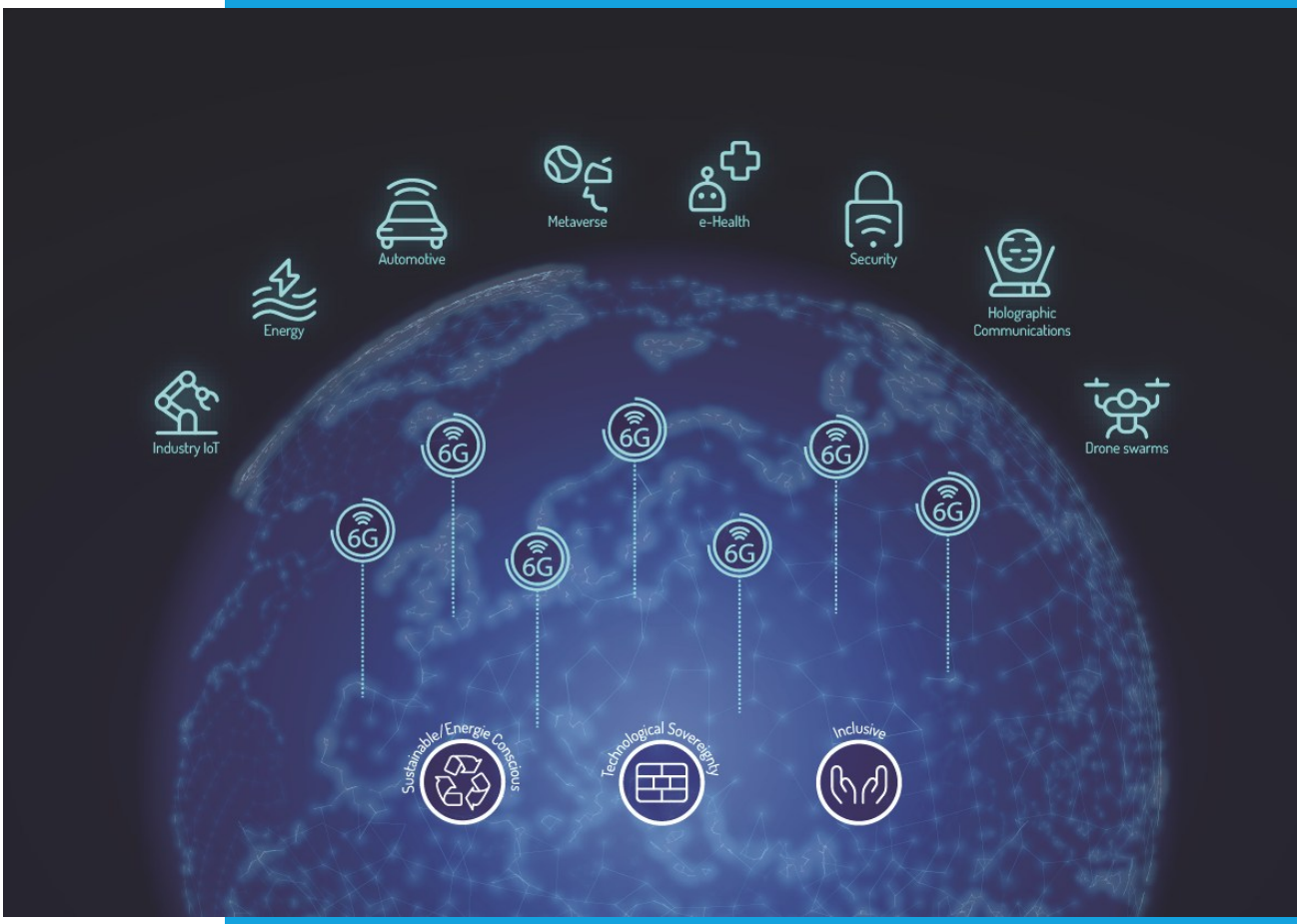
Luis Cordeiro	OneSource Consultoria Informática	Portugal
Luis M. Pessoa	INESCTEC	Portugal
Luis Velasco	Universitat Politecnica de Catalunya	Spain
Marco Fiore	IMDEA Networks Institute	Spain
Marco Gramaglia	Universidad Carlos III de Madrid	Spain
Marco Quagliotti	Telecom Italia	Italy
Maria Christopoulou	Space Hellas	Greece
Marilet De Andrade Jardim	Ericsson Research	Sweden
Mario Montagud	i2CAT Foundation & University of Valencia	Spain
Mark Angoustures	Solidshield	France
Maxime Compastie	i2CAT	Spain
Merve Saimler	Ericsson Research	Türkiye
Miguel Camelo	Imec	Belgium
Miguel Catalan-Cid	i2CAT Foundation	Spain
Mir Ghoraishi	Gigays	UK
Marten Ericson	Ericsson Research	Sweden
Mohammadreza Mosahebfard	i2CAT Foundation	Spain
Nicolas Chuberre	Thales Alenia Space	France
Nikolaos Nomikos	National and Kapodistrian University of Athens	Greece
Oriol Sallent	Universitat Politecnica de Catalunya	Spain
Oscar Gonzalez de Dios	Telefonica	Spain
Ömer Bulakci	Nokia	Germany
Özgür Akgül	Nokia	Finland
Pablo Picazo Martinez	Universidad Carlos III de Madrid	Spain
Panagiotis Botsinis	Apple	Germany
Panagiotis Gkonis	National and Kapodistrian University of Athens	Greece
Panagiotis Trakadas	National and Kapodistrian University of Athens	Greece

Paolo Di Lorenzo	Sapienza University of Rome	Italy
Paolo Monti	Chalmers University of Technology	Sweden
Pietro G. Giardina	Nextworks	Italy
Placido Mursia	NEC Laboratories Europe	Germany
Qi Wang	University of the West of Scotland	UK
Rafał Tępiński	Orange Polska	Poland
Ramon Casellas	CTTC	Spain
Ramoni Adeogun	Aalborg University	Denmark
Robert Kolakowski	Orange Polska	Poland
Roberto Viola	Vicomtech	Spain
Rodrigo Asensio Garriga	University of Murcia	Spain
Ryan Husbands	British Telecom	UK
Salvatore Spadaro	Universitat Politecnica de Catalunya	Spain
Sameh Eldessoki	Apple	Germany
Sandor Laki	ELTE	Hungary
Sandro Scalise	German Aerospace Center (DLR)	Germany
Sean Aherne	DELL	Ireland
Sebastian Robitzsch	Interdigital	UK
Sergio Aguilar	Sateliot	Spain
Shubham Gupta	Sateliot	Spain
Sokratis Barmpounakis	WINGS	Greece
Stefan Wänstedt	Ericsson	Sweden
Sylvaine Kerboeuf	Nokia France	France
Tarik Taleb	ICT-FI	Finland
Tomaso de Cola	German Aerospace Center (DLR)	Germany
Vasilis Tsekenis	WINGS	Greece
Vincent Lefebvre	Solidshield	France
Xavi Masip-Bruin	Universitat Politecnica de Catalunya	Spain

Xi Li	NEC Laboratories Europe	Germany

SUPPORTING PROJECTS

The SNS JU projects have received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme.



Website: https://smart-networks.europa.eu/sns-ju-working-groups/

Co-funded by the European Union