This deliverable has not yet been externally reviewed and approved by the European Commission.

Project no.:	101095738		
Project full title:	6G Short range extreme communication IN Entities		
Project Acronym:	6G-SHINE		
Project start date:	01/03/2023	**Duration**	30 months

D2.4 IN-X SUBNETWORK ARCHITECTURES AND INTEGRATION INTO 6G 'NETWORKS OF NETWORKS'

Due date	28/02/2025	**Delivery date**	28/02/2025
Work package	WP2		
Responsible Author(s)	Pedro Maia de Sant Ana (BOSCH)		
Contributor(s)	Baldomero Coll-Perales (UMH), Javier Gozálvez (UMH), Filipe Conceicao (IDE), Basuki Priyanto (SONY), Christian Hofmann (APPLE), Dimitrios Alanis (APPLE), Bernhard Raaf (APPLE), Christos Tsakos (COGNINN).		
Version	V1.0		
Reviewer(s)	Davide Dardari (CNIT), Frank Burkhardt (FHG), Gilberto Berardinelli (AAU)		
Dissemination level	Public		

Co-funded by the European Union

Horizon Europe Grant Agreement No. 101095738. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or SNS JU. Neither the European Union nor the granting authority can be held responsible for them.

VERSION AND AMENDMENT HISTORY

Version	Date (DD/MM/YYYY)	Created/Amended by	Changes
V0.1	11.12.2024?	Pedro Maia de Sant Ana	Toc finalisation
V0.2	07.01.2025	Christian Hofmann, Dimitrios Alanis, Baldomero Coll-Perales, Javier Gozalvez, Filipe Conceicao	Initial contributions from partners
V0.3	27.01.2025	Pedro Maia de Sant Ana	Editor revision and feedback
V0.4	31.01.2025	Pedro Maia de Sant Ana	Editor review done
V.07	04.02.2025	Pedro Maia de Sant Ana	Final version ready for internal review
V0.8	23.02.2025	Pedro Maia de Sant Ana	Feedback from the internal reviewers
V0.9	26.02.2025	Berit H. Christensen	Final proofreading and layout check
V1.0	28.02.2025	Pedro Maia de Sant Ana	Submitted version

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

Abbreviation	Meaning
3GPP	3rd Generation Partnership Project
4G	4th Generation
5G	5th Generation
5G-ACIA	5G Alliance for Connected Industries and Automation
6G	6th Generation
AD	Autonomous Driving
AF	Application function
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AMF/AUSF	Access and Mobility Management Function / Authentication Server Function
AMR	Autonomous Mobile Robot
AP	Access Point
AR	Augmented Reality
BSR	UE Buffer Status Report
CCN	Compute Offload Controlling Node
CCREF	Computing & Communication Resources Exposure Functions
CCRMF	Communication & Computational Resources Management Function
CP	Control Plane
CSI	Channel State Information
CompN	6G Network Compute Node
D2D	Device to Device
DCS	Distributed Coordination System
DGF	Device Group Function
DN	Data Network
DRB	Data Radio Bearer
E/E	Electrical/Electronic
ECU	Electronic Control Unit
EN	Entity
FCAPS	Fault, Configuration, Accounting, Performance, and Security
gNB	gNodeB (i.e., the functional equivalent of a base-station)
GW	Gateway Role
HARQ	Hybrid Automatic Repeat request
HC	Element with High Capabilities
IAB	Integrated Access Backhaul
KPIs	Key Performance Indicators
LC	Element with Low Capabilities
LCh	Logical Channel
LU	Location Updates
LiDAR	Light Detection and Ranging
MAC	Medium Access Control
MgtN	Management Node

N3SMF	Non-3GPP Subnetwork Management Function
NEF	Network Exposure Function
NIF	Network Intelligence Functions
NSA	Non-Standalone
NWDAF	Network Data Analytics Function
Near-RT	Near Real-Time
Non-RT	Non-Real-Time
O-RAN	Open Radio Access Network
OFF	Compute Offloading
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
QoS	Quality of Service
RAN	Radio Access Network
RLC	Radio Link Control
RIS	Reconfigurable Intelligent Surface
ROI	Regions of Interest
RRM	Radio Resource Management
RRM	Radio Resource Management
SAF	Sensing Analytics Function
SCF	Sensing Coordination Function
SDAP	Service Data Adaptation Protocol
SL	Sidelink
SMO	Service Management and Orchestration
SN	In-X Wireless Subnetwork
SN-TP	Subnetwork Tunnelling Protocol
SNE	Subnetwork Element
SNM	Subnetwork Management
TB	Transport Block
TSSA	Target Sensing Service Area
TTSA	Target Sensing Service Area
UE	User Equipment
URLLC	Ultra-Reliable Low Latency Communication
VR	Virtual Reality
WLAN	Wireless Local Area Network
XR	Extended Reality

EXECUTIVE SUMMARY

This document presents Deliverable D2.4 of the 6G-SHINE project, a Horizon Europe initiative focused on advancing the development of in-X subnetworks for 6G communication systems. In-X subnetworks, spanning consumer, industrial, and in-vehicle domains, are critical for enabling ultra-reliable, low-latency, and efficient communication tailored to specific use cases. Building upon the foundations laid in prior deliverables, this report emphasizes architectural advancements and strategies for seamless integration into the broader 6G ecosystem.

The report is organized to address several core objectives:

1. Introducing new architectural components, such as Communication and Computing Resource Exposure Functions, to improve scalability and adaptability of in-X subnetworks.
2. Proposing mechanisms for dynamic orchestration of communication and computational resources to meet stringent performance requirements.
3. Outlining methods to integrate subnetworks with edge and cloud infrastructures, forming a cohesive operational continuum.
4. Integrating machine learning into the subnetwork architecture to enhance decision-making for radio resource allocation, interference management, and sensing capabilities.

Key findings of the report include the identification of modular and scalable architectural elements, practical frameworks for network and compute resource allocation, and mechanisms to enable survivability and scalability of subnetworks in diverse operational conditions.

1 INTRODUCTION

6G-SHINE is a European Union-funded project under the Horizon Europe SNS framework, aiming to advance short-range communication technologies as integral components of 6G networks. This deliverable is part of Work Package 2 (WP2) and represents a continuation of prior efforts to refine the architectural components, use cases, and scenarios pertinent to in-X subnetworks. Specifically, this document builds upon Deliverables D2.1 [1] and D2.2 [2], which provided initial definitions and refinements of scenarios, use cases, and requirements for in-X subnetworks. It also directly leverages the architectural implications and technical findings derived from Work Package 4 (WP4) and Work Package 5 (WP5) which contain critical insights obtained during technical studies.

The focus of this deliverable lies in addressing the architectural implications and interfaces required for seamless integration of subnetworks into 6G parent networks, with emphasis on industrial, automotive, and consumer domains. It explores enhancements to subnetwork management, resource allocation, and computing capabilities that facilitate efficient and scalable operations across heterogeneous network environments. Key architectural enablers such as exposure functions, distributed computing roles, and non-3GPP integration mechanisms are elaborated upon, with insights derived from ongoing work in 6G-SHINE.

1.1 Objectives of the Document

The primary objective of this deliverable is to detail the architectural elements, interfaces, and enablers required to operationalize in-X subnetworks within a 6G ecosystem. By identifying and analysing these elements, this document seeks to:

- Define architectural enhancements to current 5G systems to meet the unique requirements of in-X subnetworks.
- Propose solutions for subnetwork integration, including dynamic resource allocation, interoperability with 6G parent networks and sensing capabilities.
- Highlight challenges and opportunities for future developments, with emphasis on scalability, efficiency, and adaptability.

1.2 Structure of the Document

This document is organized as follows:

- **Section 1** provides an introduction, detailing the objectives and scope of the deliverable and its relevance within the 6G-SHINE project.
- **Section 2** provides a review of the key definitions and nomenclature of architectural elements previously established in D2.2 [2]. It outlines the main architectural considerations, which are further expanded in Sections 3 and 4 to illustrate the proposed concepts and requirements. The discussion covers network elements, their roles, and various communication modes.
- **Section 3** discusses the architecture for in-X subnetworks, focusing on core components such as exposure functions, subnetwork management, and distributed computing. Each subsection addresses specific innovations and their implications for seamless integration with 6G parent networks.
- **Section 4** elaborates on architectural implications for interfaces between subnetworks and the 6G ecosystem. This includes discussions on non-transparent subnetwork configurations, tunnelling protocols, and enhanced scheduling mechanisms.

- **Section 5** presents an architectural blueprint that summarizes the key components, elements, and roles discussed throughout the document, comprehensively providing an overview of the architectural framework developed in 6G SHINE.
- **Section 6** presents conclusions and outlines directions for future work, summarizing the key contributions of this deliverable to the overall objectives of the 6G-SHINE project.

2 KEY CONCEPTS FOR IN-X SUBNETWORKS ARCHITECTURE

This deliverable incorporates the key definitions and nomenclature for all architectural elements previously established in D2.2 [2]. This section revisits the main architectural considerations from D2.2, which will be further expanded upon in Sections 3 and 4 to illustrate the proposed concepts and requirements. The discussion includes network elements, their roles, and various communication modes. Additionally, a reference architecture is presented to emphasize the foundational aspects of the initial considerations for an in-X subnetwork architecture.

2.1 Introduction

The diverse range of use cases in 6G-SHINE necessitates a top-down approach to define the key aspects of an in-X subnetwork architecture, emphasizing use case-specific requirements. This approach begins by categorizing network elements based on capabilities - such as computational power, communication features, and energy consumption - and later linking them to functionalities grouped into roles, which combine communication, management, and compute features.

We propose network elements with varying capability levels, defined relative to each use case. These relative comparisons allow for tailored role assignments, ensuring optimal functionality within the specific requirements of each 6G-SHINE use case.

2.2 Network Elements of the Parent Network

For the regular case, we assume that subnetworks are integrated in a 6G parent network, whose elements are described next and summarized in Table 1. We can also assume that the subnetworks can work autonomously, such as when the subnetworks are out of 6G parent network coverage.

2.2.1 6G Base Station (6G BS)

A 6G base station comprises the radio access network components of the parent network, to which subnetworks can be connected. Such a 6G base station can support the existing 5G base station functionalities and interfaces. Apart from the interface to the subnetworks, it should have the interface to the other base station and 6G core network elements. It might realize a subset of the signal processing and coordination functionalities which may influence the subnetworks operation. The details are under investigation in WP3 and WP4.

2.2.2 6G Core Network (6G CN)

The 6G core network comprises all necessary network functions to run, operate and manage the parent network. The core network operates the necessary functions of a mobile network, such as connectivity, mobility management, authentication, authorization, and subscriber data/policy management. It may also include functionality to manage and operate subnetworks associated with the parent network. There might exist two different network types: public networks and non-public networks (enterprise or campus). A public network is designed so that any users can access the network, while a non-public network is designed to serve specific users or organizations. A non-public network can provide a dedicated 6G core network to a certain scenario, such as an industrial factory. Alternatively, the non-public network can be consisted of a private radio access network while utilizing a public 6G core network.

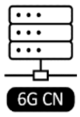
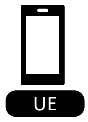
2.2.3 6G Network Compute Node (CompN)

A 6G compute node is a component that offers computing resources, and which is not part of a subnetwork, but instead can be used by multiple subnetworks. A 6G compute node can reside in the core network (e.g., expansion of existing 6G core network node, or a new node) or outside core network (e.g., application server). The 6G core network might be involved in provisioning of the compute resources.

2.2.4 User Equipment (UE)

User equipment (UE) is any device used directly by an end-user to communicate. For example, it can be handheld devices, VR headsets, or IoT devices. We consider standard UEs as defined by 3GPP. The UE does not necessarily have to be part of a subnetwork (e.g., the UE has direct communication from/to a 6G base station) but there is the possibility to have a communication relationship established between such a UE and other elements within subnetworks.

Table 1: Symbols of Parent Network Elements

Element	Abbreviation	Symbol
6G Base Station	6G BS	
6G Core Network	6G CN	
6G Network Compute Node	CompN	
User Equipment	UE	

2.3 Elements in a Subnetwork

We propose four categories of elements being present in a subnetwork, which – as mentioned above – are defined according to their capabilities. They are summarized in Table 2.

2.3.1 Element with High Capabilities (HC)

An element with high capabilities is a device/node with significant capabilities in terms of networking and computation. Such a node might act as the central communication node in a subnetwork and also might offer compute resources to other devices in the subnetwork. Multiple such HCs can be installed in a single subnetwork. An HC element can be a UE as defined by 3GPP or a non-3GPP device, which includes also access points (APs).

2.3.2 Element with Low Capabilities (LC)

An element with low capabilities is like an HC but has limited capabilities in terms of networking and computation. This can limit the functionalities this device provides to the subnetwork. We also assume that there might be no connection between the LC and the 6G base station. In a hierarchical or nested subnetwork, the LC might act as an aggregator. An LC element can be a UE as defined by 3GPP or a non-3GPP device, which also includes access points.

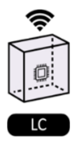
2.3.3 Subnetwork Element (SNE)

Subnetwork elements are computationally constrained devices that have limited form factor, cost footprint, power/energy, and that include devices such as sensors/actuators. An SNE acts as an end device, such that it is either the source or destination of data transmitted over the network. The SNE can also be operated as a relay node. Like the HC and LC, an SNE device can be a UE as defined by 3GPP or a non-3GPP device.

2.3.4 Reconfigurable Intelligent Surfaces (RIS)

A reconfigurable intelligent surface (RIS) is an electromagnetic element that can beneficially influence the wireless channel. RIS properties can be controlled dynamically to 'tune' the incident wireless signals through reflection, refraction, focusing, collimation, modulation or absorption. RIS can be deployed in both indoor and outdoor. RIS can also be operated at any part of the radio spectrum, including frequency below 6 GHz to THz. The descriptions and the fundamental operations of RIS can be found in [3].

Table 2: Symbols of the Elements Involved in Local Communication

Element	Abbreviation	Symbol
Element with High Capabilities	HC	
Element with Low Capabilities	LC	
Subnetwork Element	SNE	
Reconfigurable Intelligent Surface	RIS	

2.4 Further Definitions

In the following, we also introduce two commonly used terms within 6G-SHINE project, entity, and in-X wireless subnetwork, which are important for the definition of the deployment architectures. Their abbreviations and symbols are given in Table *3*.

Table 3: Symbols for Entities and Subnetworks

Element	Abbreviation	Symbol
Entity	EN	
In-X Wireless Subnetwork	SN	

2.4.1 Entity (EN)

The concept of entity stands here for a specific use case unit with limited operational radius, where one or more subnetworks can be installed for a set of coherent and correlated operations. Entities can be robots, production modules, vehicles, but even classrooms for consumer type of applications.

2.4.2 In-X Wireless Subnetwork (SN)

In-X subnetworks are special purpose networks that can support localized and high-performance connectivity within an entity. They are short range (below 10 meters), typically low-power, cells, meant at offering very localized connectivity by offloading larger networks of demanding services. They consist of one or more HCs and/or LCs with integrated edge processing capabilities (in case of HCs) and a potentially large number of cost-effective constrained devices with limited form factor, such as sensors/actuators, called SNEs. They can operate as stand-alone subnetworks in case the connection with a parent 6G network (6G base station) drops or is intermittent. The parent 6G network can take care of efficiently orchestrating the subnetwork operations when connection is available. Non-standalone subnetworks always depend on the umbrella 6G network. Additionally, subnetworks can be nested, meaning that an In-X subnetwork can itself host one or more smaller-scale subnetworks, enabling hierarchical and multi-tiered network architectures that enhance scalability and localized optimization.

2.5 Element Roles

Roles are high-level descriptions of sets of functionalities or features, and they can be assigned to different elements present in the use case. The following roles are defined in the context of 6G-SHINE and summarized in Table *4*. In this section, we emphasize the basic roles, providing an initial classification, while further roles will be defined later as the framework evolves.

Table 4: Symbols for the Different Roles

Role	Abbreviation	Symbol
Gateway Role	GW	
Compute Offloading Role	OFF	
Radio Resource Management Role	RRM	
Subnetwork Management Role	SNM	

2.5.1 The Gateway Role (GW)

An element with the gateway role can manage the data traffic routing within and/or across subnetworks. In special cases, it can act as intra-SN or cross-SN relay as well as gateway towards the 6G parent network.

2.5.2 The Compute Offloading Role (OFF)

An element with the compute offloading role uses its capabilities to orchestrate application and/or network function offloading from source elements to target elements. Furthermore, it can be as a provider or donor of compute resources to another element within the same or another subnetwork. An element with the compute offloading role may also interact with the 6G Network Compute Node.

2.5.3 The Radio Resource Management Role (RRM)

An element with the radio resource management role uses its capabilities to manage the radio resources of one or multiple other elements within a subnetwork. There can be various RRM modes, such as central, hybrid, and distributed mode. Such management functions can also consider jamming/interference handling, RIS operation (if present) and usage of licensed/unlicensed spectrum, including spectrum sharing aspect, in general.

2.5.4 The Subnetwork Management Role (SNM)

An element with SNM functionality manages the operational activities of elements within a subnetwork. This might include authentication, mobility procedures (e.g., handover), master clock roles, and monitoring of network performance.

2.6 Reference Architecture

Figure 1 shows an example architecture as an illustrative reference with three entities employing one or more than one subnetwork each. 6G BS, 6G CN, and CompN are the network nodes in the 6G parent network, in which it can be connected to the subnetwork. The leftmost entity has three subnetworks in a nested setup with HC and LC as "head nodes". These subnetworks are in an entity (EN), such as in a classroom. RISs are either integrated in subnetwork or standalone as part of the parent network. SNEs are usually end nodes, but also exceptions exist where they can relay information to another SNE. There can also be subnetworks without SNEs but only with HCs and LCs instead such as in the rightmost subnetwork. A UE is explicitly considered in case it is outside of any subnetwork but can communicate with elements within a subnetwork, i.e., with an HC, LC, or SNE. In many cases, there exist communication links between HCs and LCs, and with this also between subnetworks of the same entity or different entities. It can also be seen that different communication structures are possible, such as tree-like, meshed or setups with multi-connectivity of SNEs towards more than one HC/LC.

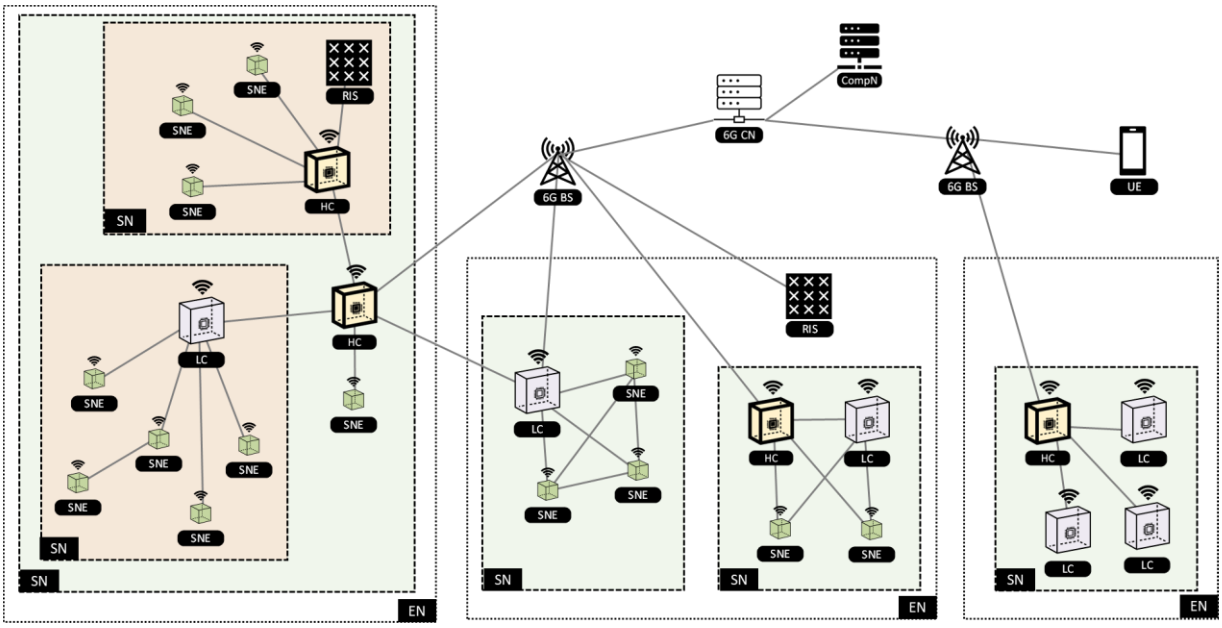


Figure 1: Reference Architecture

In Figure 2, an example role assignment can be seen for the leftmost entity of Figure 1.

Gateway Roles:

In this example, all three "head nodes" (two HCs and one LC) have the GW role as they can relay and distribute information in their subnetworks and among the SNEs. In addition, the HC of the green SN also acts as gateway towards the parent network while there is no other direct communication of any other element in the SNs towards the parent network. In another example, an SNE has the GW role as it relays towards another SNE.

RRM Roles:

The upper HC (being part of the yellow SN) has the RRM role, which means that is can manage radio resources within and beyond its subnetwork, including steering the RIS. At the same time, the 6G BS also employs RRM, so that a hybrid form of radio resource management is installed in this case.

Compute Roles:

Furthermore, two nodes, that are the HC in the green SN and the CompN, have the compute offloading role, which enables them to manage and/or offer computational offloading to other entities within their domain, such as within the green SN (including the yellow SN, as well).

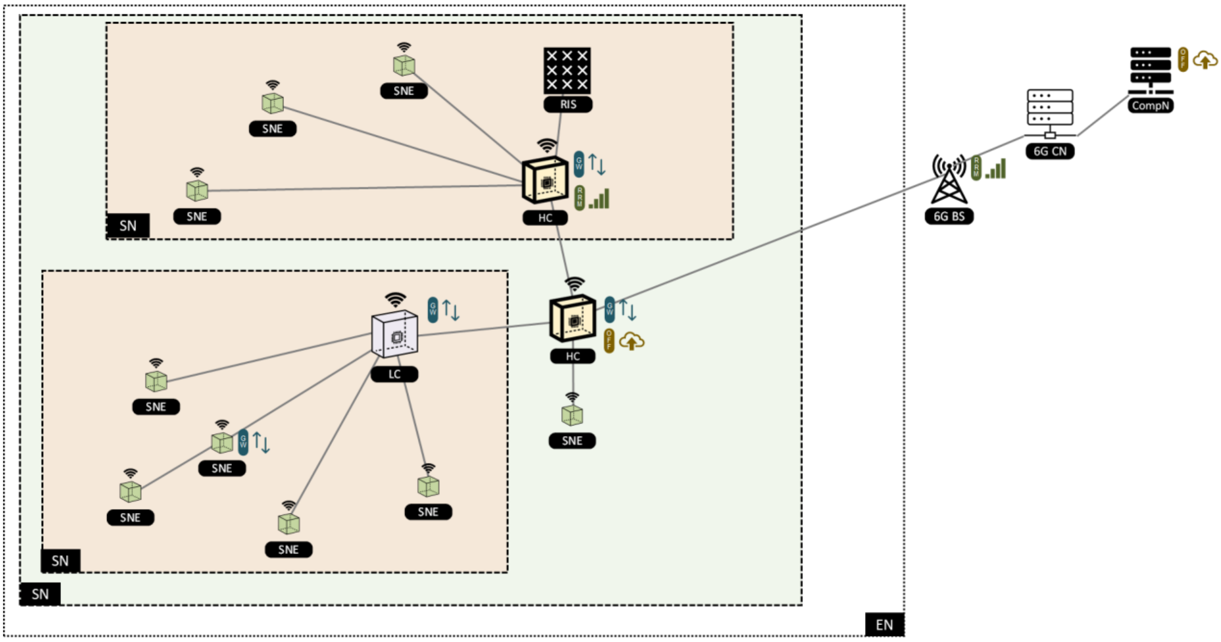


Figure 2: Example Role Assignment

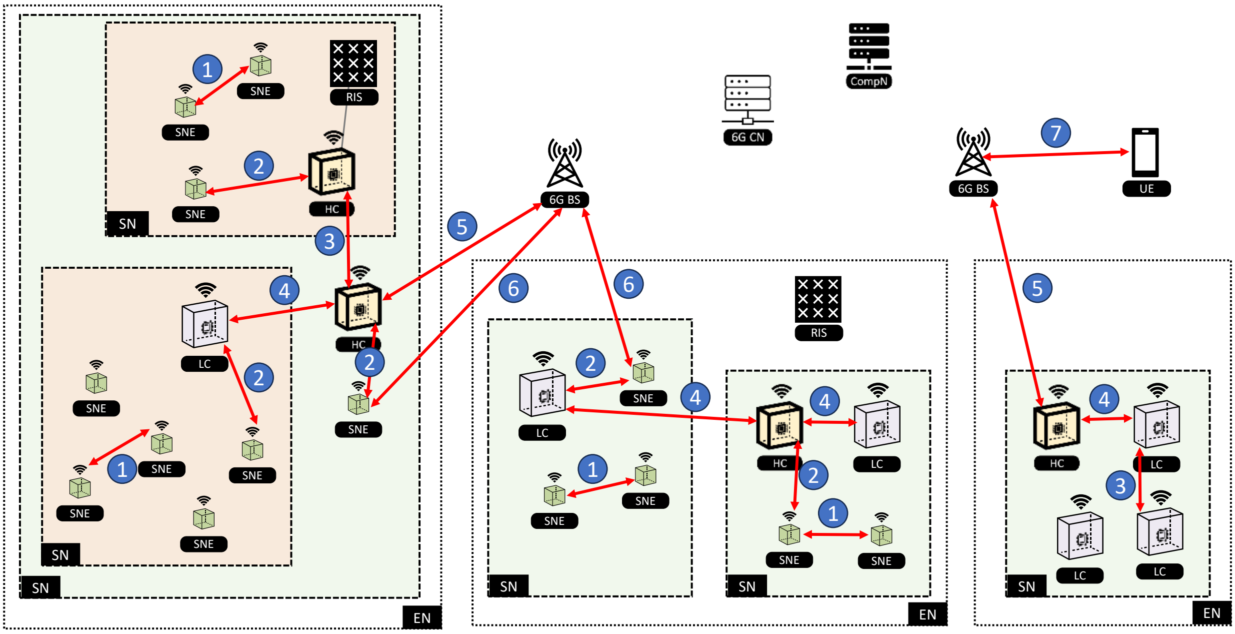
2.7 Communication modes

In general, there exist many different possibilities for communication relationships between the various types of elements. One can distinguish between direct and indirect (such as relayed) communication modes. Table *5* provides an overview of these communication relationships and highlights, to which extent these relationships are relevant for the 6G-SHINE use cases. The different possibilities of the communication modes listed in the table are also illustrated in Figure 3.

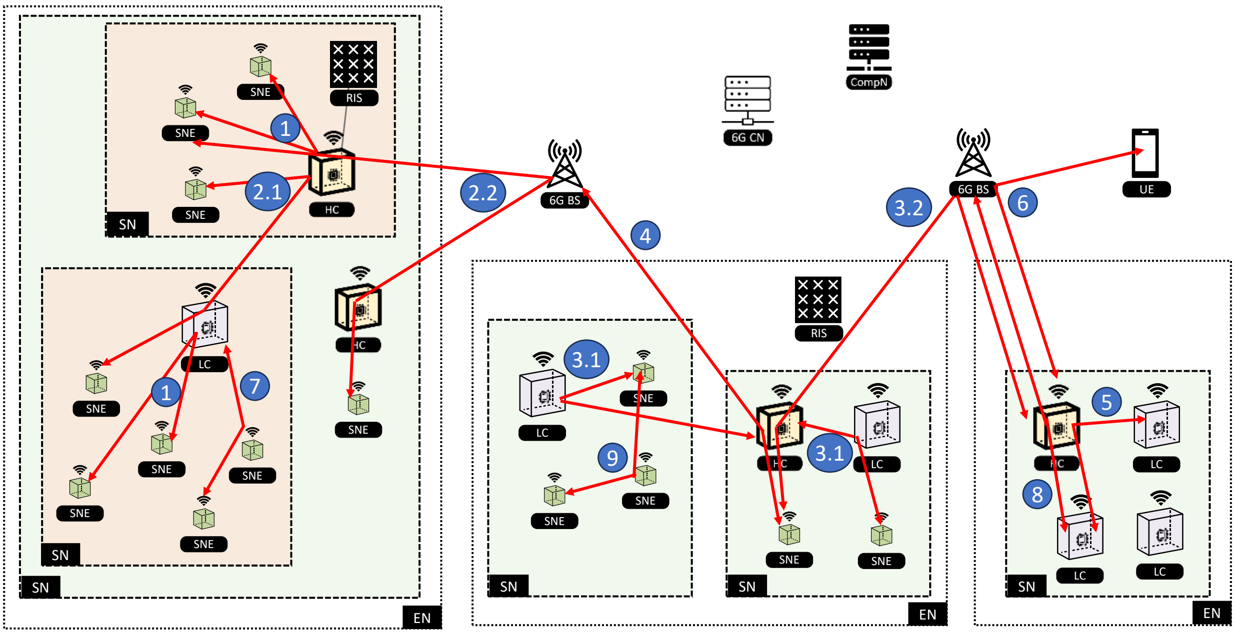
Table 5: Possible Communication Relationships Between Elements.

Communication Relationship		Relevance
	Direct Communication	
1	Between SNEs	Relevant in rare cases.
2	SNE to LC/HC	Standard communication mode present in the use cases.

3	Between LC/HC (same level)	Considered in some special cases, e.g., for load balancing or equally relevant elements.
4	Between LC/HC (different levels)	Standard communication mode present in the use cases.
5	LC/HC to 6G network	Relevant for some of the use cases.
6	SNE to 6G network	Typically, not considered in the use cases.
7	Between UE and 6G network	Relevant for some of the use cases.
	Indirect Communication	
1	Between SNEs via LC/HC	Relevant in rare cases.
2.1 2.2	Between SNEs in different SNs via LC/HC (2.1) or via 6G parent network (2.2)	Relevant in rare cases.
3.1 3.2	SNE to LC/HC via LC/HC (3.1) or via 6G parent network (3.2)	Standard communication mode present in the use cases, especially with LCs acting as data aggregators.
4	SNE to 6G net via LC/HC	Typically, not considered in the use cases.
5	Between LC via HC	Standard communication mode present in the use cases.
6	UE to SNE/LC/HC via 6G parent network	Only in use cases with a UE being present.
7	SNE to LC/HC via SNE	Relevant in exceptional cases.
8	LC/HC to 6G parent network via LC/HC	Relevant for some of the use cases.
9	SNE to SNE via SNE	Typically, not considered in the use cases.



a) Direct communication mode



b) Indirect communication mode

Figure 3: Example of communication modes for the reference architecture

3 ARCHITECTURE FOR IN-X SUBNETWORKS

This section explores the architectural elements designed specifically for in-X subnetworks, which operate with localized autonomy while maintaining essential interconnections with the broader 6G parent network. The focus is on enabling localized communication, computation, and sensing within subnetworks to address diverse use-case requirements, including industrial automation, automotive applications, and immersive education.

In Subsections 3.1 and 3.2, we first examine higher-level architectural aspects, including routing mechanisms, distributed computing capabilities, and dynamic subnetwork grouping. We introduce the Subnetwork Routing Protocol, which facilitates efficient routing, and the deployment of network layers tailored to specific subnetwork needs. Additionally, we present concepts for distributed computing, enabling local task offloading within subnetworks through roles such as Computing Nodes and Compute Offload Controlling Nodes. These mechanisms extend to collaborative computing across subnetworks, optimizing resource utilization and enhancing overall performance in a decentralized manner.

Finally, in Subsection 3.3, we shift focus to RAN-level architectural elements. Inspired by the O-RAN framework [10], we integrate machine learning into the subnetwork architecture to improve decision-making for resource allocation and interference management. This encompasses both real-time and non-real-time control mechanisms, allowing subnetwork RAN operations to dynamically adapt to changing conditions. Additionally, we also introduce the Sensing Coordination Function and Sensing Analytics Function, ensuring that communication and sensing activities are aligned to meet application-specific requirements. We highlight immersive education as a key use case, demonstrating the potential of these functions in optimizing network performance for latency-sensitive and data-intensive applications.

3.1.1 Subnetwork Management

Subnetworks are foreseen as an addition to the existing 3GPP network topology, as mentioned in Section “Flexible Roles for the Subnetwork Nodes” of D4.2 [4]. Subnetworks are distinct entities formed among devices without or with limited Network (NW) configuration or NW awareness. This degree of independence from the NW decreases the NW complexity especially in scenarios with dense deployments. Furthermore, subnetworks and the concept of local communication are essential for achieving the requirements coming from various use cases described in D2.1 [1] and D2.2 [2]. D2.2 defines several classes of devices involved, namely the *High Capability* (HC), the *Low Capability* (LC) and the *Subnetwork Element* (SNE) classes D2.2 [2], which are defined in section 2.3 in more detail. On top of these device classes, there are multiple roles identified and defined in section 2.5, which are required to enable subnetwork operations, such as Subnetwork Management (SNM), Gateway (GW) and Radio Resource Management (RRM) [2].

In D4.2 [4], it is outlined that those roles should be aggregated at the *Management Node* (MgtN), which is inherently an HC device. The MgtN is responsible for the local subnetwork control and routing, while it also acts as a gateway, connecting subnetwork nodes and the overlay 6G NW. Deploying these functions within the subnetwork increases the subnetwork’s survivability in the absence of an overlay 6G network. Figure 4, cited from D4.2 [4] exemplifies this architecture with the MgtN and its various roles.

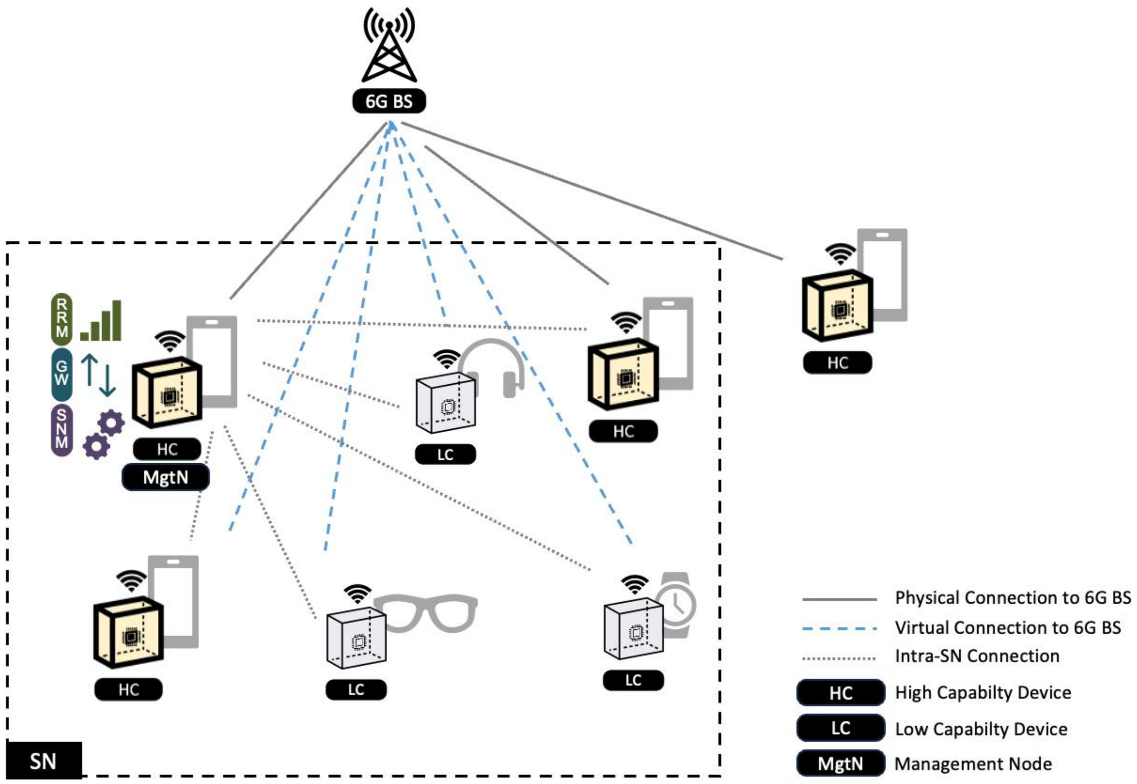


Figure 4: Exemplified single subnetwork architecture for the case of immersive education [4]

3.1.2 Subnetwork Routing Protocol

In D4.2 [4] section 2.2.2.2, it was investigated what kind of user plane functionality is required within the subnetwork and how this *Subnetwork User Plane* (snUP) could be realised. A key point was that it shall be possible that user plane functionality - e.g., provided by sublayers SDAP, PDCP, RLC or MAC - gets deployed flexibly and dynamically within the subnetwork to cover various subnetwork technologies and device capabilities. It also becomes a necessity to enable the various use cases with their specific requirements on, e.g., low latency or high reliability [1]. The detailed deployment shall be transparent to the 6G overlay NW to ensure privacy and to limit the complexity within the NW. The outcome was a novel protocol to accommodate the flexibility of the snUP, namely the *Subnetwork Routing Protocol* (SN-RP). It shall be deployed on top of MAC and accommodate different L2 sublayer deployments (SDAP, PDCP, RLC) across the subnetwork. The SN-RP delivers routing through the subnetwork by carrying mapping and routing information for the different layers per packet between the different devices. In Figure 5, a deployment scheme is showcased, highlighting that the different sublayers can be deployed flexibly, whereas in Figure 6 a detailed example of the SN-RP along with its sublayer deployment is shown.

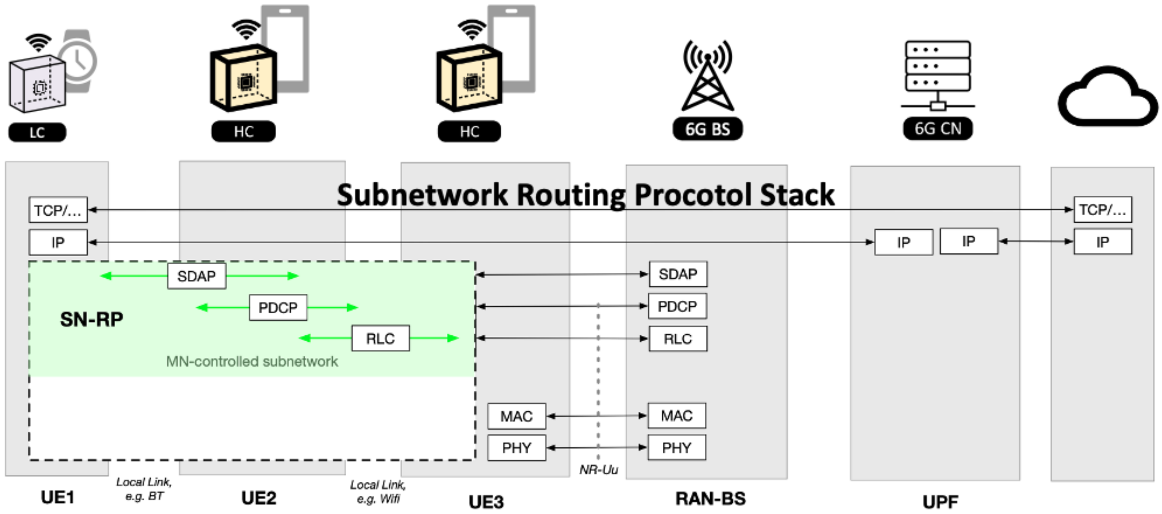


Figure 5: Subnetwork Routing Protocol (SN-RP) stack.

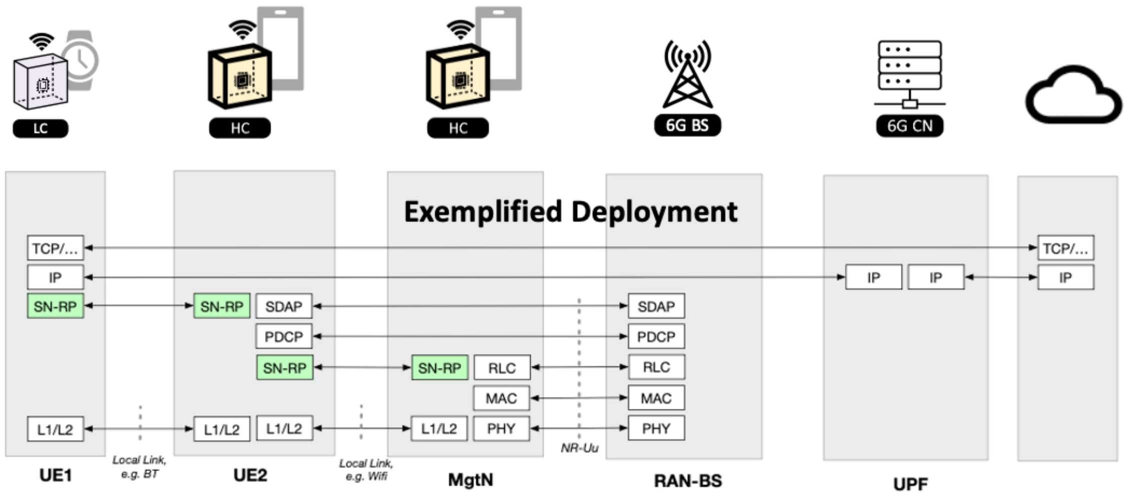


Figure 6: Exemplified deployment with a nested subnetwork and various layer deployments of the SN-RP

To summarize the architectural implications of this approach, each UE that wants to participate in subnetworks shall also support the SN-RP protocol. Additionally, a UE supporting the subnetworks feature shall also be able to indicate in its UE capabilities which layers it supports to be deployed in-device, and which layers are required to terminate on another device. For instance, some devices may only support SN-RP deployments with only SDAP and PDCP.

3.1.3 Device Group Functions

As described in D2.2 [2], many use cases require multiple devices working together in a coordinated fashion as a Device Group (DG) dealing Service Data Flows (SDFs), such as *extended reality* (XR) use cases or interactive gaming. 3GPP also has described such use cases [5] and studied multi-modality requirements in TS 22.847 [6]. In Figure 7, exemplified deployments with video imaging, audio, control signalling for actuators and sensor data are depicted in different colours targeted to different devices within a subnetwork.

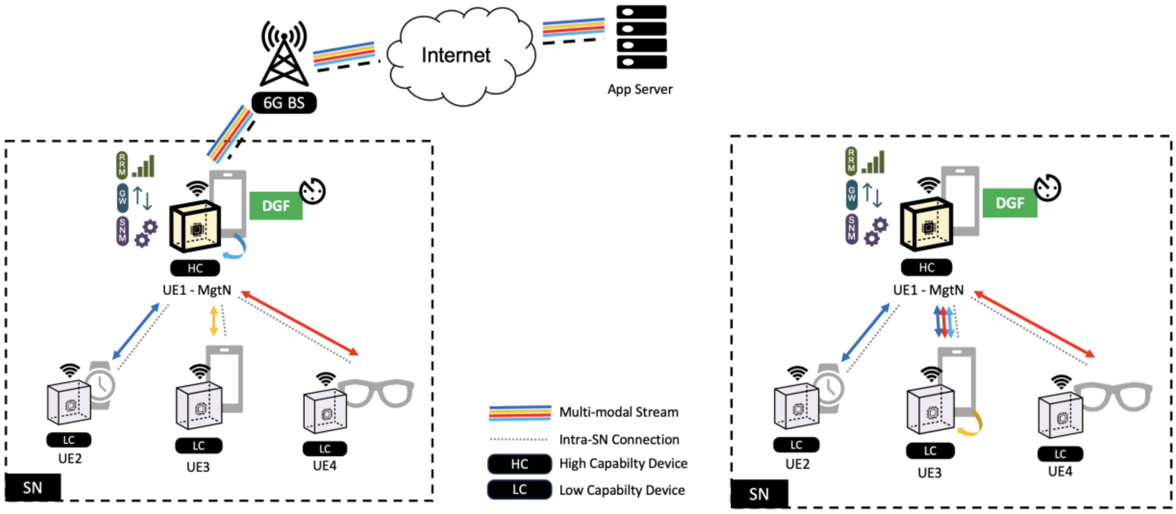


Figure 7: Multi-Modality among Devices in subnetworks.

In the context of 6G SHINE, locality and survivability play an important role and require the introduction of a Device Group Function (DGF) as drafted in Figure 7. This new coordination function, being deployed close to or within the subnetwork, shall ensure the alignment in time of multimodal data streams among the different devices. The details of the DGF as well as the deployment options within the 6G parent NW or the subnetwork will be investigated further in detail in the subsequent deliverable D4.4.

3.2 Local Subnetwork Computing

This section introduces architectural innovations that are necessary to enhance the computational capabilities within subnetworks. In particular, Section 0 introduces architectural innovations to improve the computation capabilities of low-complexity UEs and SNEs, including enablers to allow computational offloading and functional offloading within subnetworks alongside local distributed computing. For example, in immersive education use cases, XR devices worn by teachers and pupils can offload tasks to nearby devices like phones or laptops, reducing delays and meeting stringent requirements for communication and computation. This approach also addresses privacy and security constraints by enabling local processing instead of relying on the edge or cloud infrastructures, which might also incur significant overhead. Additionally, the increasing softwarisation and automation of entities like vehicles and robots require enhanced flexibility and reliability in their subnetworks. Meeting time-sensitive demands necessitates joint orchestration of tasks, communication, and computing resources. 6G-SHINE envisions in Section 3.2 advanced interfaces and exposure functions creating a seamless continuum of resources to enable dynamic allocation, addressing the challenges of automated and softwarised subnetworks efficiently.

3.2.1 Distributed Computing

Apart from architectural enablers introduced to improve the communication capabilities of the subnetworks, the computational capabilities of the UEs should also be improved. This is justified by the fact that LC UEs and SNEs often have reduced computational resources. Additionally, the stringent communication and computational requirements presented in [2] in terms of delay and throughput, demand for enabling computational offloading on top of functional offloading, as will be described in Section 4.3.1. In fact, this is required especially in the consumer use case category [2], such as in immersive education. In this specific use case, both the pupils and the teacher wear XR devices and participate in collaborative tasks. In close vicinity, there are several other nodes, such as phones or even laptops with superior processing capabilities, which can assist the XR devices in the offloading some of their tasks to the neighbouring devices. This is crucial for satisfying the extreme requirements of [2], since the alternative, i.e. offloading tasks to the cloud through the CN, incurs significant overheads in terms of delay and throughput. Furthermore, there may exist privacy and security constraints mandating that computations are executed locally at subnetwork level, instead of using the CN and the cloud infrastructures.

There are no architectural provisions in 3GPP that would enable local distributed computing. Therefore, architectural enablers should be introduced such that local computational offloading becomes possible. In the context of subnetworks, it is foreseen that this would require three stages, shown in Figure 8. Stage 1 covers the subnetwork creation process, i.e., the process where the UEs and the MgtN establish secure connections. Stage 2 covers the subnetwork registration process with the parent 6G network. This stage involves messaging exchange between the MgtN and the base station (BS) to register the MgtN and the devices within its subnetwork with the 6G NW. This stage may be optional, since the subnetworks could be formed independently, in the absence of a parent NW, but here the case of collaboration across SNs via the BS is described, hence it is required. Finally, Stage 3 in Figure 8 covers the process of setting up the compute topology and performing computational offloading. The latter is related to the SN architecture and therefore a more detailed outlook will be given in the following subsections. The complete picture with all details and possible variations of Stage 3 will be covered in the upcoming deliverable D4.4.

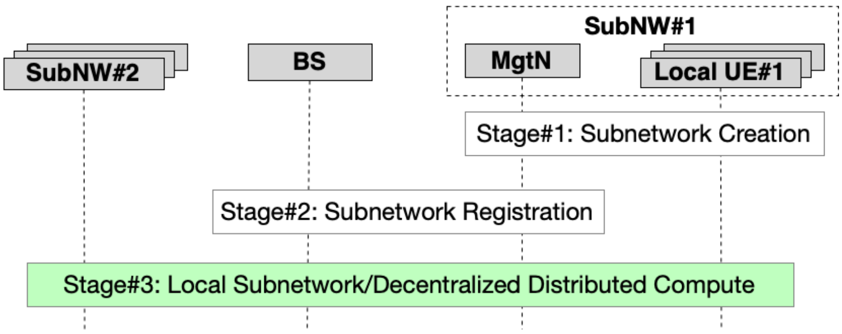


Figure 8: Three stages for enabling local SubNW and decentralized distributed compute.

3.2.1.1 Distributed Computing Roles

Before delving into the specifics of Stage 3 in Figure 8, the associated functional roles of the subnetwork nodes should be defined. A preliminary set of roles has been introduced in D4.2 [4] as follows:

- **Offloading node (ON):** connected to a subnetwork, having a compute task to be offloaded to one or more Computing Nodes
- **SN Computing node (SN CompN):** subnetwork internal node with certain processing capabilities to perform an offloaded compute task and produce compute result (compare to 6G NW CompN in 2.2.3)

- **Compute offload controlling node (CCN):** collects information about all compute capabilities from all available Computing Nodes and makes compute offload decision based on their current load
- **Routing node (RN):** an optional network node via which the compute task/compute result from Offload node/Compute node gets routed to one or more Computing node(s)/Offload Node. Note that this role is only relevant when routing data from compute offloading task.

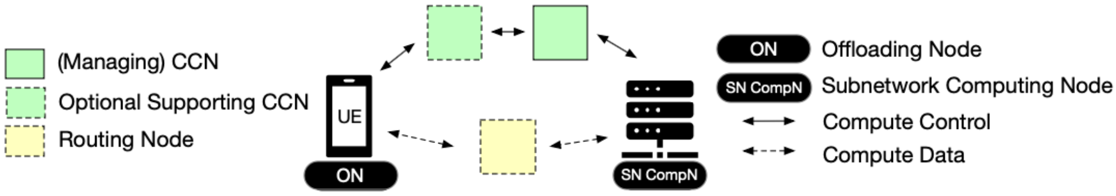


Figure 9: General functional architecture of distributed compute.

This set of functional roles enables the establishment of local offloading within the subnetwork. An ON has a task to offload to a SN CompN. The control of this process is undertaken by the subnetwork CCN, which selects the appropriate SN CompN. Note that the role of the CCN can be undertaken by any of the subnetwork nodes, as long as it has adequate power and computational resources for this role. Then, the computational payload is optionally routed from the ON to the SN CompN with the aid of RNs. However, additional control functionalities are required to enable distributed offloading across neighbouring subnetworks as well. The CCNs of each subnetwork need a way of coordinating their operations. For this reason, we introduce a managing – supporting relationship indicating which CCN will take the overall control, thus coordinating its supporting CCNs. This is achieved with the aid of the following additional roles:

- **Managing CCN:** a CCN that takes control of the overall distribution logic in addition to the handling of the resource and process management functions.
- **Supporting CCN:** a CCN that delegates some or all the compute distribution logic as well as the management of resource and process management functions.

In a nutshell, the managing CCN, which can reside within the same or a neighbouring subnetwork as the ON, controls the overall offloading process, i.e., the selection of an appropriate SNCompN and establishing a route from the ON to the SN CompN. The supporting CCNs assist the managing CCN in creating the compute topology across neighbouring subnetworks, by aggregating and forwarding the computational capabilities of their subnetwork to the managing CCN. These roles are presented in Figure 9, where the CCNs perform the control of the computational offloading task, while the optional RNs are forwarding the associated payloads between the ON and the SN CompN.

3.2.1.2 Local and Decentralized Distributed Computing

Based on the new roles described in the previous section, the processes required to enable local, and decentralized computing can be readily defined. Note that only an overview of these processes related to the overall subnetwork architecture is presented in this deliverable. The finer details of these processes will be presented in D4.4.

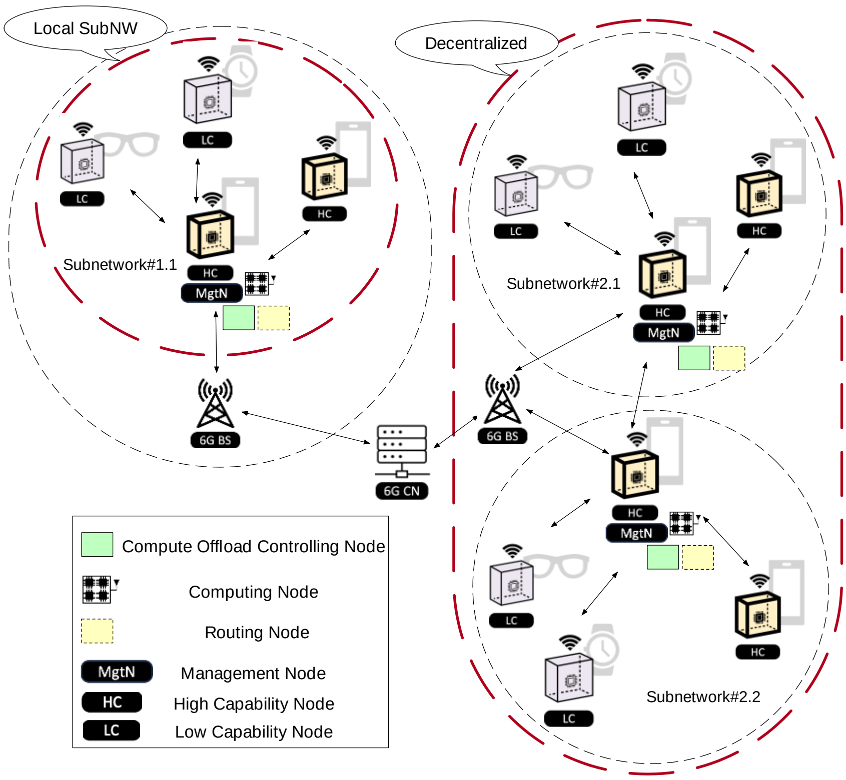


Figure 10: Local subnetwork compute offload and decentralized compute offload

As far as local distributed computing is concerned, the topology of Subnetwork#1.1 in Figure 10 is assumed, where the computational offloading will be carried out exclusively by the subnetwork nodes. Note that in this case, the presence of a 6G BS is optional as all actions for establishing task offloading are restricted to be located within the subnetwork. All devices can become ON nodes, when they have a specific task to offload. Additionally, SN CompNs and CCNs are devices with higher computational capabilities, such as the MgtN and HC devices. The process for establishing local computational offloading from an ON to a SN CompN is portrayed in Figure 11. Naturally, the first step is the subnetwork creation, where the nodes connect to the MgtN and by extension to the subnetwork. During this phase, they establish the communication links required for all the necessary messages exchange among the subnetwork nodes. The MgtN assists in the propagation of the messages from ONs to SN CompNs. Subsequently, a negotiation phase regarding which node will take the role of the subnetwork CCN takes place within the subnetwork. There is a need for this negotiation to take place because other HC nodes besides the MgtN may also act as CCNs, thus offering functional offloading service to the MgtN. Once the CCN is selected across the subnetwork, the CCN gathers the capabilities of all the SN CompNs. When an ON is required by one of its applications to offload a computational task, it forwards this request to the CCN. The latter selects an appropriate SN CompN to offload the task. This could be done based on the SN CompNs' computation capabilities as well as by considering ON's and SN CompNs' security and privacy constraints. Finally, the CCN enables the link between the ON and the SN CompN by appropriately routing the messages between the two endpoints.

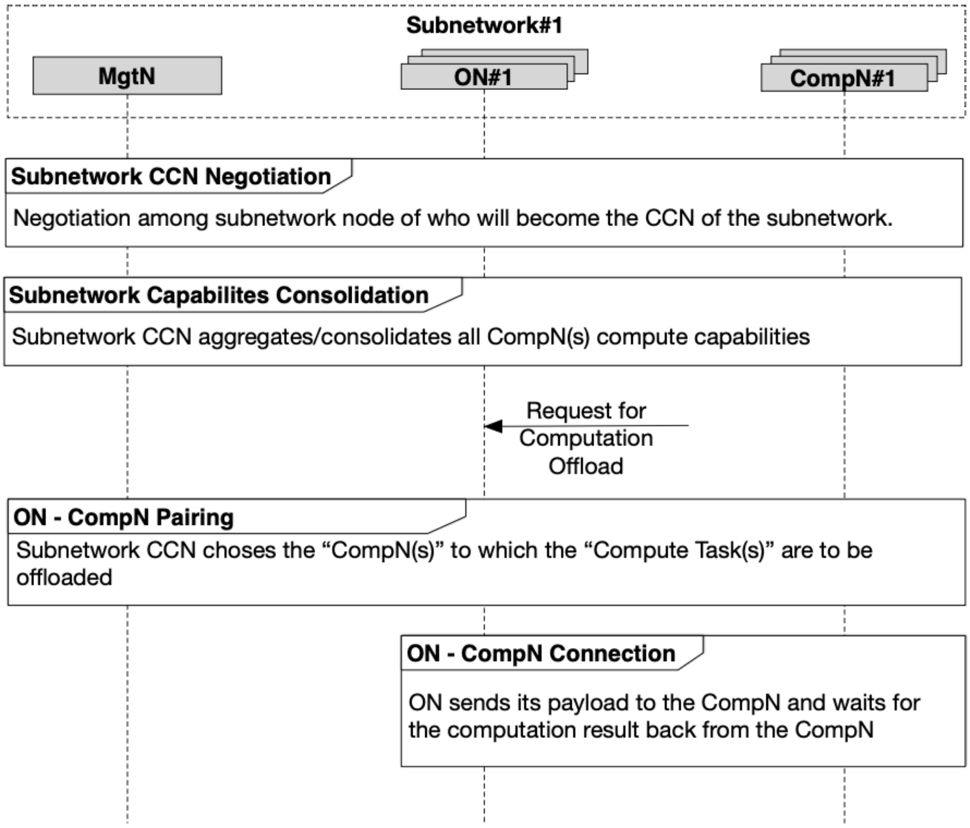


Figure 11: Process for establishing local distributed computing within a subnetwork.

Moving on to enabling the decentralized distributed compute offloading, the design goal is to enable compute offloading not only within a specific subnetwork but also across subnetworks. For instance, in Figure 12 a computational task from Subnetwork#2.1 could be offloaded to the neighbouring Subnetwork#2.1 in case this is not possible to be undertaken within Subnetwork#2.1. This could be done either in a D2D manner - using the mechanism of D4.2, where neighbouring MgtNs authenticate each other and establish direct secure links [4], or with the aid of the 6G BS. The 6G NW has no participation in the decision logic for computational offloading and it is used instead as a communication backhaul. For instance, a task from Subnetwork#2.1 could also be offloaded to Subnetwork#1.1 using the 6G NW's communication infrastructure as a backhaul. It should be noted that Stage#2 "Subnetwork Registration", as shown in Figure 8, is a pre-requisite, when the 6G NW shall be involved as communication backhaul. As multiple subnetworks are now involved, their CCNs should be coordinated. This coordination is enabled with the aid of the newly introduced roles of Section 3.2.1.1. Explicitly, the overall decision logic could be undertaken by a single CCN, i.e., the managing CCN, which could be assisted by the other CCNs, namely the supporting CCNs. For example, the supporting CCNs aggregate the computational capability reports of their subnetworks and forward their reports to the managing CCN.

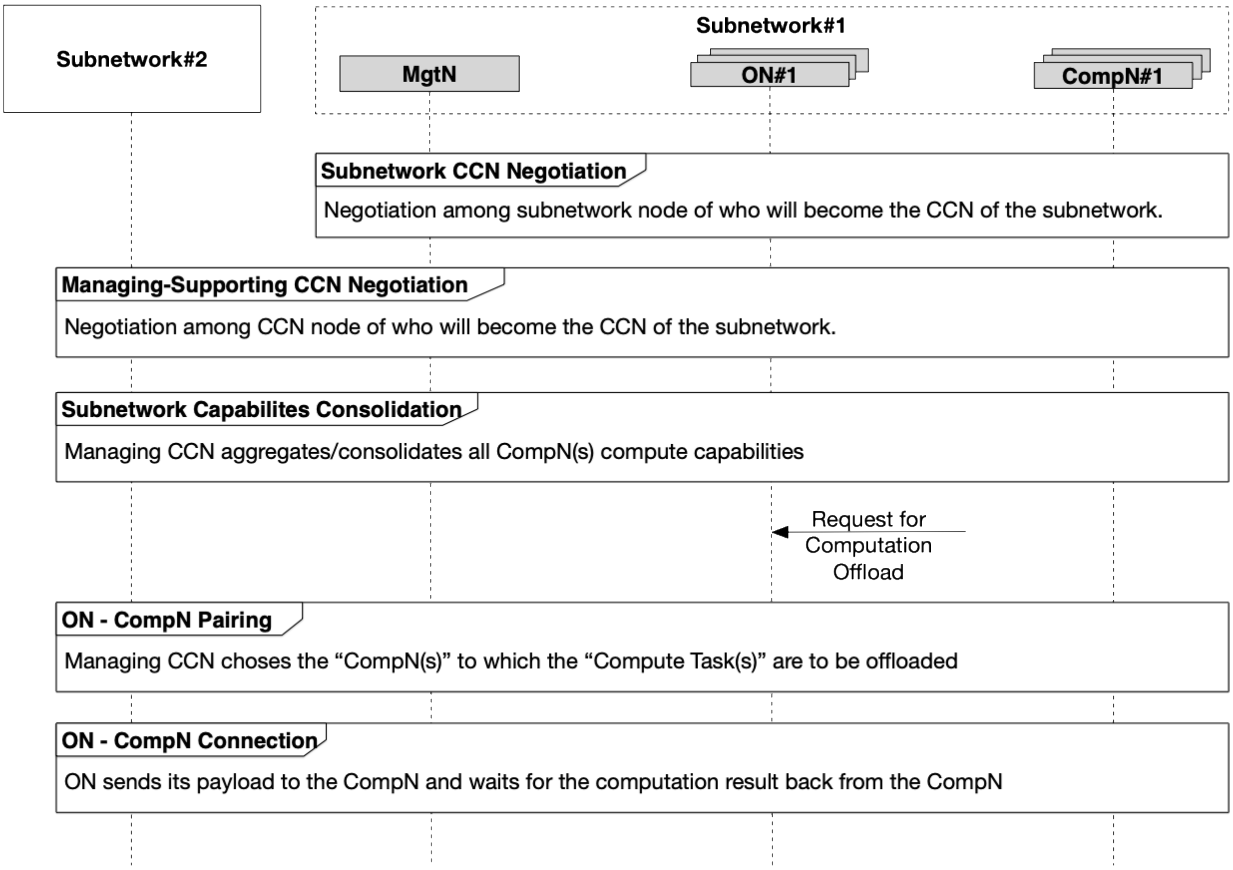


Figure 12: Process for establishing decentralized distributed computing across subnetworks.

The process of establishing decentralized computing offloading across subnetworks is portrayed in Figure 12. Up until the local subnetwork CCN selection, the process is identical to that of Figure 11. After the CCNs are selected per subnetwork, a negotiation phase takes place for the selection of the managing CCN. Once the managing CCN is selected, the rest of the CCNs act as supporting CCNs, thus gathering their subnetworks computational capabilities and sending them to the managing CCN. The managing CCN in turn consolidates the capabilities. When an ON requires computational offloading of a task, this request is forwarded to the managing CCN, which selects an appropriate SN CompN to offload the task. Finally, the managing CCN enables the link between the ON and the SN CompN by appropriately routing the messages between the two endpoints with the aid of the supporting CCNs. Note that details on CCN negotiation process as well as mobility aspects are going to be presented in D4.4.

3.2.2 Control Functions for the Local Subnetwork Continuum

The softwarisation of entities at the deep edge - such as vehicles, robots, and other automated systems - combined with their increasing levels of automation and the virtualization of their components, is imposing growing demands on the flexibility and reliability of their networks or subnetworks. Ensuring the effective execution of these entities' tasks and processing functions within the required time constraints necessitates joint scheduling of tasks, as well as communication and computing resources [4] [8]. To meet the time-sensitive requirements of such systems, advanced mechanisms must be implemented to jointly orchestrate and schedule task execution alongside the optimal allocation of communication and computing resources.

Achieving this requires extending the current exposure functions available in 5G systems to enable the creation of a flexible resource continuum—essentially a virtual pool of resources in the domain that can be dynamically allocated to meet task demands. For example, this subnetwork continuum allows for the elastic extension of the computing and processing capabilities of an in-vehicle Electronic Control Unit (ECU) and enables the seamless execution of tasks on any computing element within the in-vehicle network. This concept could be referred to as an in-vehicle virtual ECU, as described in the defined 6G-SHINE use cases [2].

This section recalls the capabilities of exposure functions available in 5G systems and presents the extended vision of 6G-SHINE for architectural innovations. These innovations aim to enhance subnetworks by creating a seamless continuum of communication and computing resources, allowing for the flexible and efficient allocation required to satisfy the time-sensitive demands of softwarised and automated entities.

3.2.2.1 Exposure functions in 5G systems

In 5G systems, Network Exposure Functions (NEFs) enable the exposure of network capabilities, features, and services to third-party applications, service providers, or other network functions. This is a critical aspect of the Service-Based Architecture (SBA) defined by 3GPP, enabling dynamic service innovation and seamless integration between telecom networks and external ecosystems. As it is represented in the NEF architecture in reference point representation of Figure 13, the NEF exposes capabilities and events of the 5G core network securely to external (e.g. application functions – AF – of 3rd party service providers out of the trust domain) or internal consumers (e.g., network functions – NF), and it provides a means to interact with the network without requiring applications to understand the complexities of network internals.

3GPP defines several key functions and capabilities that NEFs can expose [9]. For example, a NEF allows ex/in-ternal entities (e.g., third-party applications or AFs) to subscribe to, query, or receive notifications about specific events happening in the network, such as: User Equipment (UE) mobility events (e.g., location updates, handovers), Quality of Service (QoS) changes, or network slicing lifecycle events. A NEF can also provide policy and QoS control mechanisms for external entities to influence network policies, such as applying QoS requirements for specific flows, influence traffic routing for specific applications, and request resource reservation or enforce bandwidth or latency guarantees. NEFs can also expose network capabilities such as the available network slices, or supported service types (e.g., Ultra-Reliable Low Latency Communication - URLLC, enhanced Mobile Broadband - eMBB).

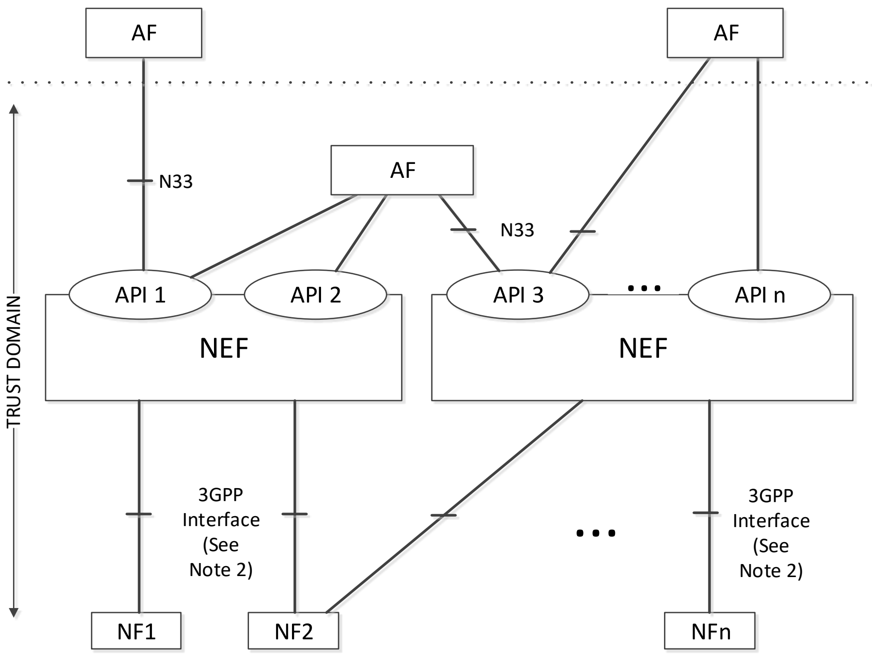


Figure 13: Architecture for Network Exposure Function in reference point representation [9]

3.2.2.2 Advanced exposure functions in subnetworks

The 5G Network Exposure Functions (NEFs) primarily focus on the interface between Network Functions (NFs) and Application Functions (AFs) to facilitate traffic management and Quality of Service (QoS) assignments through interaction with the policy elements (i.e., the Policy Control Function, PCF). As illustrated in Figure 14 using the layered architecture model (or the 5G functional architecture framework), the NEF exposes the application layer for interaction with the network function layer.

In our subnetwork vision, we propose extending NEF capabilities to enable the dynamic scheduling of communication and computing resources within the subnetwork domain for any subnetwork task. The Communication and Computing Resources Exposure Functions (CCREF) in the subnetwork aim to create a localized computing and communication resource continuum that subnetwork management functions can leverage when defining advanced policy control functions.

It is beyond the scope of 6G-SHINE, and hence of this Deliverable, to define the exact mechanisms and Application Programming Interfaces (APIs) that subnetwork elements should use to expose their communication and computing capabilities. However, Figure 14 illustrates an example in which a set of subnetwork elements - namely SNE, LC, and HC - are organized into four distinct subnetworks within the same entity. These elements expose their communication and computing capabilities to the CCREF function. The orange vertical dashed lines in Figure 14 logically represent the exposure of resources from the infrastructure layer (hosting physical resources) to the network function layer (comprising logical network functions). The CCREF enables the definition of a communication and computing subnetwork continuum in which resources are logically pooled as part of the subnetwork, rather than being tied to the individual elements that own them.

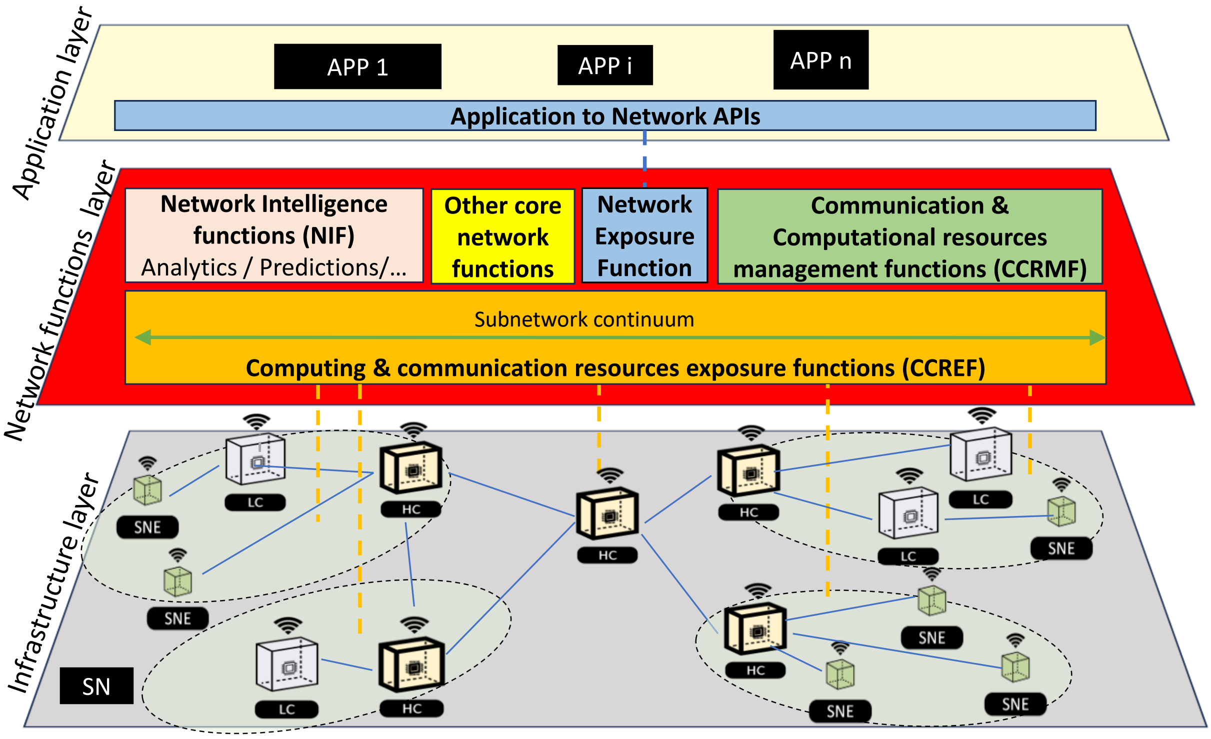


Figure 14:Subnetwork layered architecture model with CCREF continuum (dotted lines represent the logical links of the function matching the colour, while solid lines represent physical wireless links).

3.2.2.3 Advanced network functions in subnetworks

The continuum established by the CCREF in the subnetwork serves as the foundation for the definition of advanced network functions in subnetworks. Figure 14 considers subnetwork independent operation thanks to the functional splitting and offloading from 6G parent network. The following advanced network functions for subnetworks are represented in Figure 14:

- 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 subnetwork continuum.
- 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 network functions 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 subnetwork continuum.

3.2.2.4 Realization of the subnetwork advanced functions

Figure 15 show an example of the realization of the subnetwork system architecture depicted in Figure 14. The independent or stand-alone operation of the subnetwork is guaranteed in this case with the allocation of all control functions to one of the HC nodes. This node integrates not only the GW, OFF, RRM and SNM functionalities described in Section 2, but it is also responsible for implementing the control functions that allow the creation of the subnetwork continuum (i.e., CCREF). In line with the capabilities of the HC described in Section 2, this HC node could represent a Management Node. The subnetwork elements logically expose their resources directly to this HC node (see Figure 14), but as it can be seen in the example represented Figure 15, the realization of the process to expose resources is done from SNEs/LCs to the HC via intermediate subnetwork elements. To this aim, intermediate LC or HC nodes also embed some control functions which allow them acting as GW (e.g., to interconnect subnetworks) or RRM (e.g., for the management of radio resources) or exposing resources to the HC Management Node (e.g., for the establishment of the communication and compute continuum). It should be noted that the distribution of control functions represented in Figure 15 allows the HC Management Node to jointly manage the communication and computing resources exposed by other elements of the subnetwork (e.g., HC, LC or SNE) to the CCREF following the policies and procedures of the CCRMF and satisfying the service requirements of the SNEs exposed through the NEF.

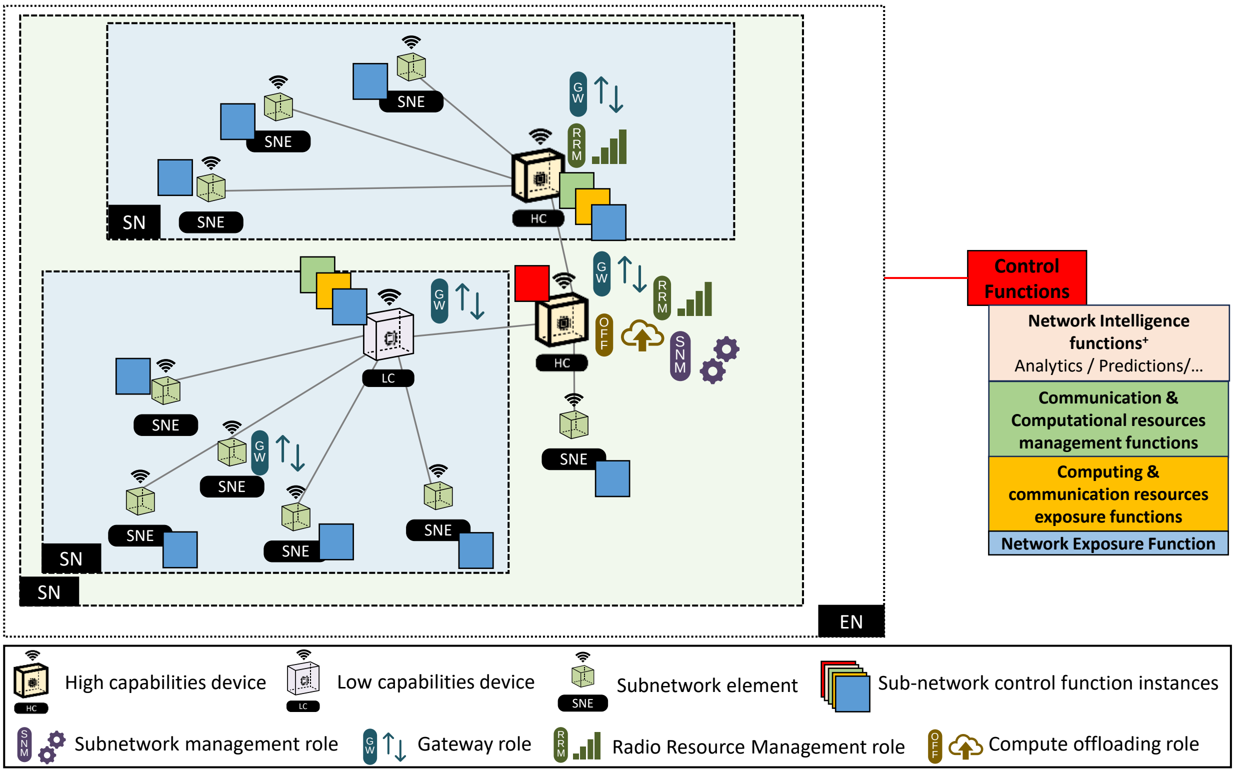


Figure 15: Example of realization of subnetwork layered architecture model with CCREF continuum.

3.3 Advanced Decision-Making in Subnetworks

In the previous subsections, we focused on architectural enhancements that enable subnetworks to internally and dynamically group and distribute compute resources, emphasizing higher-layer optimizations. In this subsection, we shift our focus to RAN-level optimization, specifically leveraging AI techniques to enhance decision-making at the subnetwork RAN level. This approach allows subnetworks to adapt more effectively to dynamic environments while addressing their specific use-case requirements.

Our decision-making framework is fundamentally inspired by the O-RAN architecture [10], utilizing Near-Real-Time and Non-Real-Time Radio Intelligent Controllers (RICs) [10] to manage resource allocation and minimize overall system interference through the implementation of xApps and rApps. By integrating these AI-driven control mechanisms, the subnetwork RAN can make more intelligent, context-aware adjustments.

Additionally, this subsection introduces sensing operations as a crucial component of subnetwork RAN, enabling new applications, for example, in healthcare domain. We present key functions such as the Sensing Coordination Function and the Sensing Analytics Function, which align sensing and communication activities to meet stringent performance requirements. These functions facilitate more efficient spectrum utilization, improve situational awareness, and enhance overall network reliability, reinforcing the role of sensing in future subnetwork architectures.

3.3.1 Radio Intelligent Controller for managing subnetwork operations

3.3.1.1 Introduction

The introduction of a RIC architectural element is motivated by two primary objectives. First, it aims to provide a feasible architectural solution, along with interfaces inspired and validated through the O-RAN framework [10], to effectively manage both subnetworks RAN and their respective elements within a parent umbrella network. This approach aligns with the industrial use cases discussed in Deliverable D2.2 [1]. Second, it seeks to establish a practical method for integrating control applications into network configuration management, thereby enabling goal-oriented network applications. This aligns with the concept of 6G subnetworks as highly specialized network entities, as explored in Deliverable D4.1 [11].

3.3.1.2 Literature Review

The Open Radio Access Network (O-RAN) architecture offers a modular and programmable framework that might serve as a blueprint for future 6G subnetworks. Key components of O-RAN include the Near-Real-Time RAN Intelligent Controller (Near-RT RIC) and the Non-Real-Time RAN Intelligent Controller (Non-RT RIC). The Near-RT RIC enables real-time control and optimization of network nodes via control loops operating on timescales of 10 ms to 1 second. It hosts xApps, which are applications designed to provide specific control and analytics functions and interfaces with the Non-RT RIC through the A1 interface for receiving policy guidance and enrichment data. On the other hand, the Non-RT RIC resides within the Service Management and Orchestration (SMO) framework and supports longer control loop intervals, leveraging rApps for data-driven network optimization. The SMO itself is a central system responsible for managing the RAN domain, handling tasks like Fault, Configuration, Accounting, Performance, and Security (FCAPS) management through interfaces such as O1 for network function management and O2 for cloud and workload management.

The O-RAN architecture emphasizes key interfaces, such as E1 for coordinating control and user plane elements, E2 for connecting Near-RT RIC with network nodes, and O1 for unified management. These interfaces enable modularity and flexibility, ensuring the interoperability of components. The O2 interface connects the SMO to the O-Cloud, a computing platform hosting virtualized O-RAN functions. Together, these elements and interfaces provide a highly programmable and adaptable system that aligns with modern network requirements.

In the context of 6G subnetworks, O-RAN-inspired elements can drive the design of architectures that manage specialized network segments tailored to diverse application needs. RIC-based control mechanisms could facilitate real-time optimization of multiple subnetwork entities, addressing ultra-low latency, high reliability, and dynamic resource allocation requirements. However, decision-making and control actions are expected to occur within a 10-

millisecond timescale. While this latency may be high for certain industrial applications, it is particularly relevant for dynamic spectrum sharing between subnetworks and the parent network, as well as for interference coordination and mobility management.

Furthermore, application-oriented solutions, such as xApps and rApps, could be tailored for high-specialization use cases, including industrial IoT and immersive AR/VR experiences, as defined in 6G SHINE use-cases [2]. The SMO-like orchestration layer could coordinate policies and workloads across subnetworks, ensuring interference mitigation and harmonized operations, as studied by the vehicle use-case in [2].

The principles of openness, modularity, and programmability inherent to O-RAN can be also directly applicable to 6G subnetwork design. By adopting open interfaces and standards, 6G subnetworks can also be able to promote multi-vendor interoperability. Modular controllers inspired by the RIC framework might also enable real-time and non-real-time control, aligning with the diverse and dynamic needs of the control application, as it will be discussed in the next section.

3.3.1.3 Proposed Architectural Elements

Our proposed network architecture is illustrated in Figure 16. At its core, HC devices are equipped with Near-RT RIC functionality, enabling dynamic reconfiguration of subnetwork controller resources in real time. The subnetwork elements correspond to E2 Nodes, where wireless data statistics collected from these devices are processed by xApps hosted on the Near-RT RIC. Through the E2 interface, which can also function as a message broker, the xApp can collect wireless data statistics and reconfigure RAN parameters across different subnetwork RAN function splits, including the O-RAN Central Unit (O-CU), O-RAN Distributed Unit (O-DU), and O-RAN Radio Unit (O-RU) [10]. Additionally, applications can subscribe to the RIC, providing context information to xApps to support decision-making processes.

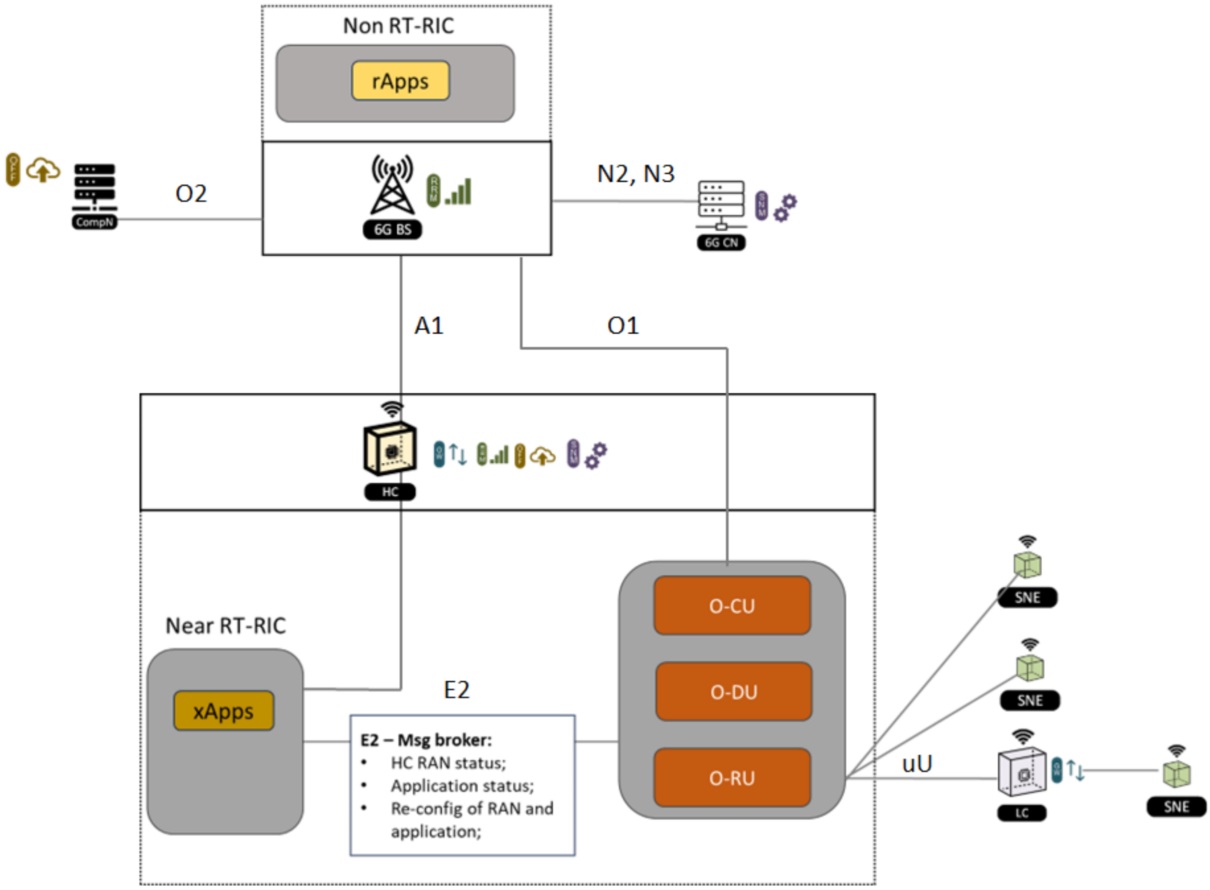


Figure 16: Detailed architectural concept of ORAN-based Subnetwork.

The parent network incorporates an SMO functionality. Through well-defined interfaces, the RIC can report subnetwork statistics to the SMO and receive policy updates and radio reconfiguration commands. This integration also leverages rApps for optimized non-real-time decisions. These architectural concepts and roles facilitate the deployment of solutions to research problems like those analysed in 6G SHINE Deliverable D4.2 [11], which addresses the coexistence of multiple subnetworks in a factory environment. Specifically, in the proposed problem, multiple AGVs (each representing a subnetwork entity) must coexist during their missions within the factory, where a central Radio Resource Management entity controls their velocity to avoid interference [23].

To transition the proposed solution to an actual deployment, the architectural elements can be utilized to design a unified network control approach. In this setup, a central parent network assumes the SMO role for managing multiple AGV subnetworks, each equipped with a Near-RT RIC. A key challenge lies in the implementation of the E2 interface, which can facilitate contextual data sharing between the control application and the RIC while also reporting information to the SMO. For example, this could involve using a message broker protocol for the E2 interface. With contextual data and wireless statistics obtained via the E2 interface, it becomes possible to design xApps and/or rApps with two main objectives:

1. **Adjust wireless communication parameters** directly through the RIC, such as scheduling policies or MIMO configurations, aligning with standard O-RAN RIC functionality.

2. **Send control commands** to the AGV application for physical control, such as speed adjustments.

The proposed deployment's information flow is described in Figure 17.

Actors and Roles

- **Non-RT RIC:** Provides RAN policy control functions.
- **Near-RT RIC:** Implements and enforces RAN policies in near-real-time, while also establishing modules for direct application data exchange.
- **E2 Nodes:** Perform control plane and user plane functions.
- **SMO/Collection & Control:** Acts as the termination point for the O1 interface.

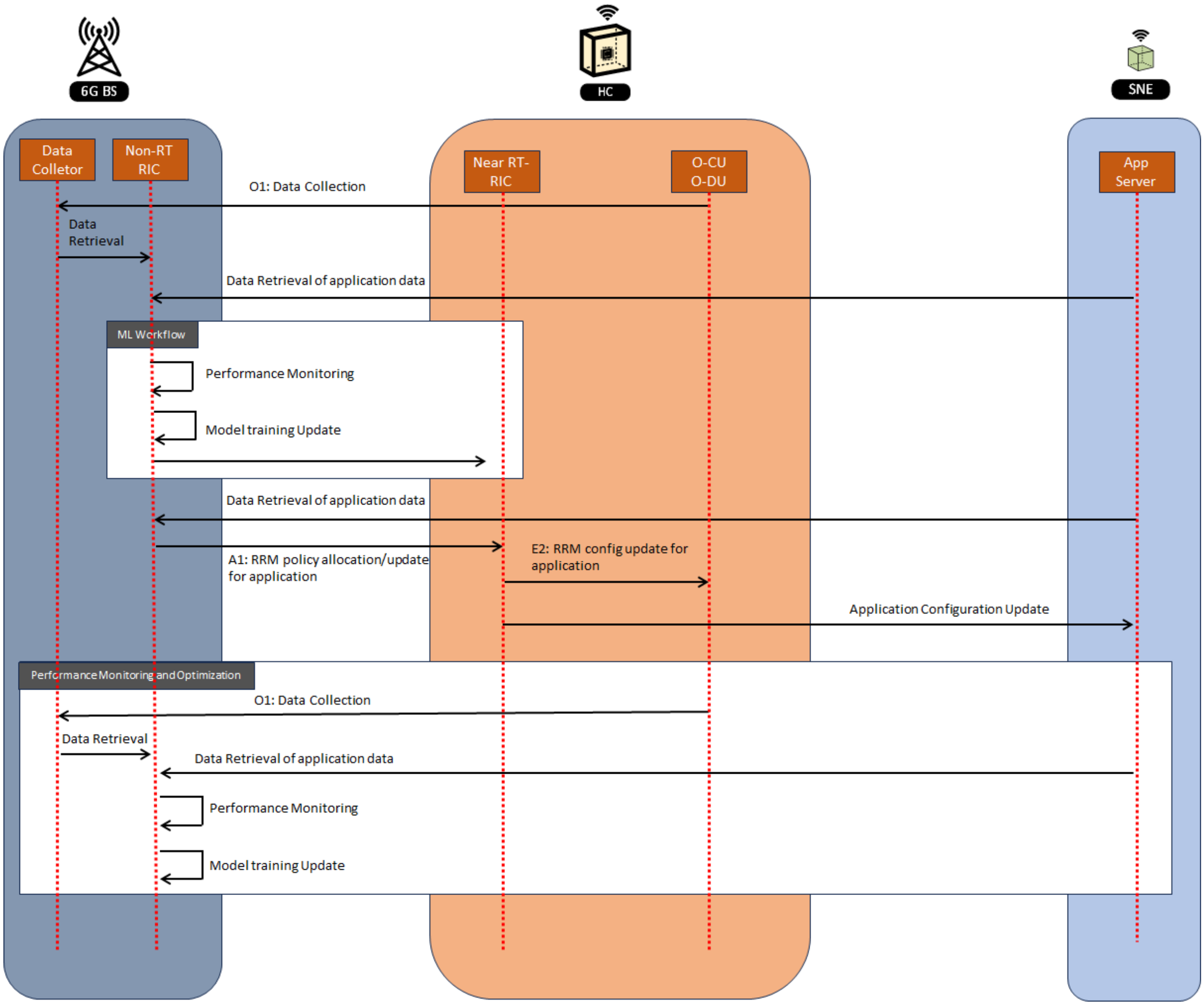


Figure 17: Illustration of the interactions between the SMO, Near-RT RIC, E2 Nodes, and AGV control applications, highlighting the data exchange and control mechanisms via interfaces such as E2, O1, and message broker protocols.

3.3.1.4 Main Benefits

The current gNB/eNB architecture lacks an effective functional module capable of retrieving application-specific information, performing machine learning tasks, and training models. In contrast, the O-RAN architecture introduces components that can collect both application-specific and radio environment information. This is crucial for enabling subnetwork HC devices to tailor network configurations to specific applications. More specifically, by incorporating O-RAN architectural principles into the subnetwork framework, HC devices can utilize application-gathered data to execute ML models, establish a foundation for intelligent, goal-oriented decision-making, and synchronize efficiently with the parent network via the SMO.

Furthermore, the long-term optimization function within O-RAN can process radio and control application data at the subnetwork element level in a unified manner. This function supports both analysis and prediction tasks by leveraging a wide range of AI/ML methods. The Non-RT RIC facilitates long-term analytics and policy-driven optimization, while the Near-RT RIC ensures real-time decision-making based on the processed data. Together, these components enable the transition of the proposed solution into a practical deployment, bridging the gap between theoretical concepts and operational network implementations.

3.3.2 Sensing operation enablement within subnetworks

3.3.2.1 Introduction

Wireless sensing will be a pivotal technology enabler in 6G networks, that will serve a wide range of futuristic applications. Applications span across multiple domains, covering industrial, automotive, health care, safety applications, well aligned with the 6G-SHINE use cases.

The 6G-SHINE use cases have the commonality of extreme requirements in terms of, amongst others, latency and throughput. In sensing, KPIs are different and tailored for the Sensing activity, but in order to be met, they may rely on the fulfilment of underlaying communication tailored KPIs. This is because sensing activities may generate challengingly high amounts of data, that require either processing, and therefore have high compute capability requirements, or transmission over the air for subsequent processing at the destination compute node.

Therefore, communication and sensing KPIs cannot be decoupled or dissociated, as the links that serve the sensing purpose may be the same that serve communications from a logical standpoint and are the same that serve communications from physical standpoint. Hence, sensing KPIs are dependent on the underlaying communication KPIs.

3.3.2.2 Sensing activities and its requirements in the context of subnetworks

Adapted from [22] , Table 6 captures the list of currently defined KPIs for Sensing activities. This includes scenarios (or use case clustering), where particularly *Object Detection and Tracking* and *Motion Monitoring* are sensing activity categories that are included in the set of 6G-SHINE use cases [2].

To name a few examples, *Object Detection and Tracking* is a pertinent sensing activity in the context of automotive and industrial use cases. This can serve the purpose of detecting obstacles on or off a road, e.g., pedestrians, or other vehicles that need to be tracked after initial detection. Those vehicles are themselves subnetworks. In the industrial realm, all use cases in [2] consider sensors, sensor data aggregation units, cameras, and image processing units, which themselves are sensors and perform sensing data collection for the execution of the automated factory use cases. There, *Object Detection and Tracking* is a pertinent sensing activity for tracking AGVs, for example.

Motion Monitoring is particularly relevant in the context of consumer electronics, where the detection of motion can apply to the subnetworks described in [2]. Specific helpful outcomes from sensing that can be used in such use case types include Detection, Localization, and Tracking, both of objects and people, Environment Reconstruction and Imaging, Posture and Gesture recognition, Health monitoring, e.g., breathing and heart rate patterns, weather monitoring, e.g. rain fall, building structural damage, flood detection, amongst others.

Table 6: Performance requirements for 5G Wireless Sensing (adapted from [22])

Scenario	Confidence level [%]	Accuracy of positioning estimate by sensing (for a target confidence level)		Accuracy of velocity estimate by sensing (for a target confidence level)		Sensing resolution		Max sensing service latency [ms]	Refreshing rate [s]	Missed detection [%]	False alarm [%]	Sensing service description in a target sensing service area
		Horizontal [m]	Vertical [m]	Horizontal [m/s]	Vertical [m/s]	Range resolution [m]	Velocity resolution (horizontal/ vertical) [m/s x m/s]					
Object detection and tracking	95	10	10	N/A	N/A	10	5	1000	1	5	2	Indoor/outdoor (e.g., detection of human, UAV)
	95	2	5	1	N/A	1	1	1000	0.2	0.1 to 5	5	Outdoor (e.g., detection of human, UAV)
	95	1	1	1	1	1	1 x 1	100 or 1000; 5000 for detection in highway	0.05 to 1	2	2	Indoor/outdoor (e.g., detection and tracking of human, animal, UAV)
	99 for public safety, otherwise, 95	0.5	0.5	1.5 for pedestrian, 15 for vehicle, otherwise, 0.1	1.5 for pedestrian	0.5	0.5 x 0.5 for factories	100 to 5000	0.1	1	3	Indoor/outdoor (e.g., detection and tracking of human, animal, UAV, AGV, vehicle)
Environment monitoring	95	10	0.2	N/A	N/A	N/A	N/A	60000	60 to 600	0.1 to 5	3	Nature of environments monitored by sensing (e.g. rainfall, flooding monitoring)
Motion monitoring	95	N/A	N/A	N/A	N/A	N/A	N/A	60000	60	5	5	Human motions and activities obtained by sensing
	95	0.2	0.2	0.1	0.1	0.375	0.3	5 to 50	0.1	5	5	Human hand gestures obtained by sensing

Below are considered definitions pertaining to the sensing activity in general and the Sensing KPIs in [22].

Definitions pertaining to the sensing activity:

- **Sensing Signal** is a transmitted radio signal from a Sensing Transmitter for the purpose of sensing. The signal can be 6G or non-6G.
- A **Sensing Transmitter** is a 6G or non-6G entity that transmits a Sensing Signal.
- A **Sensing Receiver** is a 6G or non-6G entity that receives a Sensing Signal and produces Sensing Data. A Sensing Receiver can be co-located with a Sensing Transmitter.
- **Sensing Data** is the 6G or non-6G data produced for sensing purposes.
- The **Sensing Results** may include estimations of point cloud, object identification (size, shape, material), properties characteristics of objects (e.g. type, distance, velocity, trajectory, size, shape, material), or other contextual information (e.g. time of generation, environmental information) about objects in the Target Sensing Service Area (TSSA).
- **Fusion** refers to a process to join two or more streams of Sensing Data or Sensing Results together to form one or more Sensing Data or Sensing Result stream(s). Fusion can take place at the origin of the sensing data, along the system entities of a 6GS. The fusion of Sensing Results can also take place along all 6GS system entities. Fusion can also take place in non-6GS entities.

More details on the process that starts with the transmission of a sensing signal and ends with the production of a sensing result are given in Section 4.2.3.

- A **Sensing Service** is a feature of the 6GS that is offered to service consumers. A Sensing Service provides Sensing Results based on communicated requirements and KPIs.
- A **Sensing Task** consists of activities to perform sensing, including the configuration of the required Sensing Transmitter(s) and Sensing Receiver(s) (if applicable), the collection of Sensing Data, the processing of the Sensing Data and the exposure of the Sensing Results. Each Sensing Task fulfils a Sensing Service request.
- **TTSA**: a cartesian location area that needs to be sensed by deriving characteristics of the environment and/or objects within the environment with certain sensing service quality from the impacted (e.g., reflected, refracted, diffracted) 3GPP radio signals. This includes both indoor and outdoor environments.

A TSSA can be a subnetwork, or an area containing the subnetwork that cannot be dissociated from the subnetwork. It represents the area where there is a need to perform sensing tasks, and that comprises the elements of a subnetwork, either the entirety or partially. Examples of this are the TSSA being a vehicle, or an area surrounding a vehicle, pinned to it, i.e., that follows the vehicle as it moves, and remains stationary if the vehicle stays stationary. The TSSA can represent an indoor industrial area, where the sensing activity is carried for the automation of the factory.

The sensing tasks to be performed in the subnetwork, happen following a sensing service request that arrives to a parent 6G network. It then stands out that the management of these activities starts at a System level, at the parent 6G network.

3.3.2.3 Architectural enablers for sensing in subnetworks

To address the execution of sensing tasks within subnetworks, two new components are required to enhance current architectures as described in Figure 18.

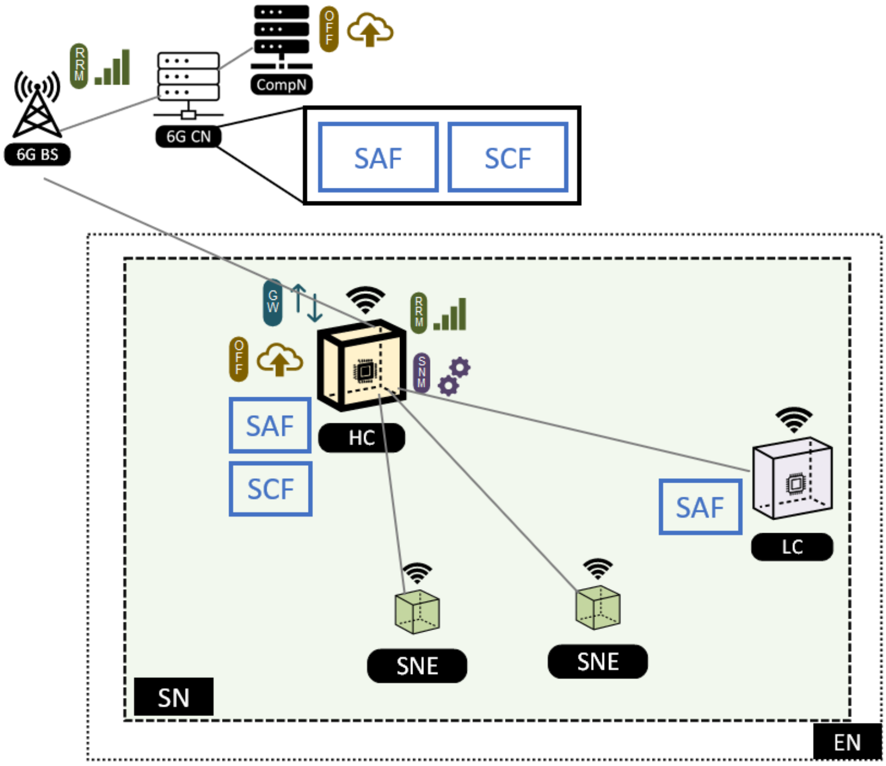


Figure 18: Architectural enablers for Sensing in subnetworks.

A description of these components is given below.

Sensing Coordination Function (SCF)

The SCF coordinates the sensing operation in the parent network and within the subnetworks. This can include full management of Sensing sources of Sensing Data, non-3GPP Sensing Data, Sensing Results, and Sensing contextual information, including source selection, activation, de-activation, configuration and activation/de-activation of reporting from sources, etc., the sources being an individual Sensing transmitter, receiver, or a Sensing group. It manages the activation/de-activation as well and/or switching of Sensing modes. Sensing modes refer to mono-static or multi-static. In mono-static, the sensing receiver node is the same as the sensing transmitter node. In multi-static, there is one node that is the sensing transmitter, and multiple nodes can be the sensing receivers. Bi-static sensing is a subcase of multi-static sensing, where there is only one transmitter and one receiver.

Sensing Analytics Function (**SAF**)

It is described as an end point for sensing data, that requires further processing in order to generate meaningful sensing results (more details in Section 4.2.3). Based on the collected Sensing Data, performs analytics over Sensing Data, processes Sensing Results, or both, being capable of generating additional Sensing Data, Sensing Results, and Sensing contextual information. It can further generate insights over Sensing Data, results and contextual information, e.g., by means of the application of statistical, probabilistic, or AI/ML methods in general. Furthermore, the SAF can perform fusion of Sensing Data from multiple sources, i.e., can combine different Sensing Data, results and contextual information from any Sensing source and generate further data/results from that fusion process.

This functionality can expose the gathered or generated information to application servers in a DN, e.g., via the NEF and/or AF. The SAF can be, just like the SCF, a logical entity rather than a dedicated NF. Amongst the options to realize its processes are enhancements to the current NWDAF, where the described SAF functionalities may be embedded.

Additional considerations on the SAF and SCF

These enhancements require system level architectural changes. However, subnetworks require local management of the described functionalities. Hence, it is considered that the SCF can have a management role in the subnetwork, similar to its CN functionality, via parent network authorized SCF functionality transfer to the subnetwork, e.g., to an HC node.

The process of achieving a sensing result is more complex, as it requires different compute capabilities depending on the target sensing result metric. For this reason, the SAF functionalities can be deployed in the subnetwork, typically in HC and/or LC nodes, that can participate in the sensing task operation.

The management of the compute resources within the subnetwork impacts on the sensing data transfers there. For example, given the topology in Figure 18, the LC node may or not be able to compute sensing data to convert into a sensing result. In case not, it will require compute offloading to the HC node. In this case, there is a trade-off to be addressed in terms of amount of sensing data to offload for compute purposes and the associated total computational latency. This is a non-trivial challenge and for that reason addressed in detail in WP4, D4.4.

3.3.3 AI/ML enabled RRM architecture

In scenarios where multiple subnetworks operate in the same spectrum, efficient **Radio Resource Management (RRM)** becomes critical to ensure reliable communication. Section 2 explores the potential for implementing RRM at the subnetwork level, with the HC node able to take on these functionalities. This responsibility may require architectural modifications to the HC stack- specifically, integrating advanced decision-making processes, such as AI-driven methods, to manage resources dynamically and establishing mechanisms for the exchange of information between different subnetworks. Such an architectural evolution is particularly relevant in environments with high spectrum congestion like industrial ones. A relevant use case is the “Subnetworks Swarms: Subnetwork Coexistence in Factory Hall” scenario, as defined in Deliverable D2.2 [2]. In this setting, multiple robotic control machines communicate wirelessly with one or more robotic arms, with each robotic system forming its own subnetwork. Should these subnetworks coexist within the same frequency bands, the need for efficient RRM becomes essential. A promising approach to address this challenge is distributed power control, as described in Deliverable D4.1 [11]. Further interference management solutions could include, but are not limited to, subband allocation, which dynamically assigns specific frequency segments to subnetworks and adaptive beamforming, which focuses signal transmission to specific directions to maximize spectrum utilization. When combined with AI/ML-driven methods, these strategies enable dynamic, data-informed decision-making to optimize resource allocation. Although interference management through power control serves as just one example of application of the architecture proposed here, the challenge of integrating AI/ML models within subnetworks can be extended to a broader range of resource management problems. Figure 19 illustrates the proposed scheme from D4.1, where each subnetwork is managed by a HC node equipped with AI/ML model.

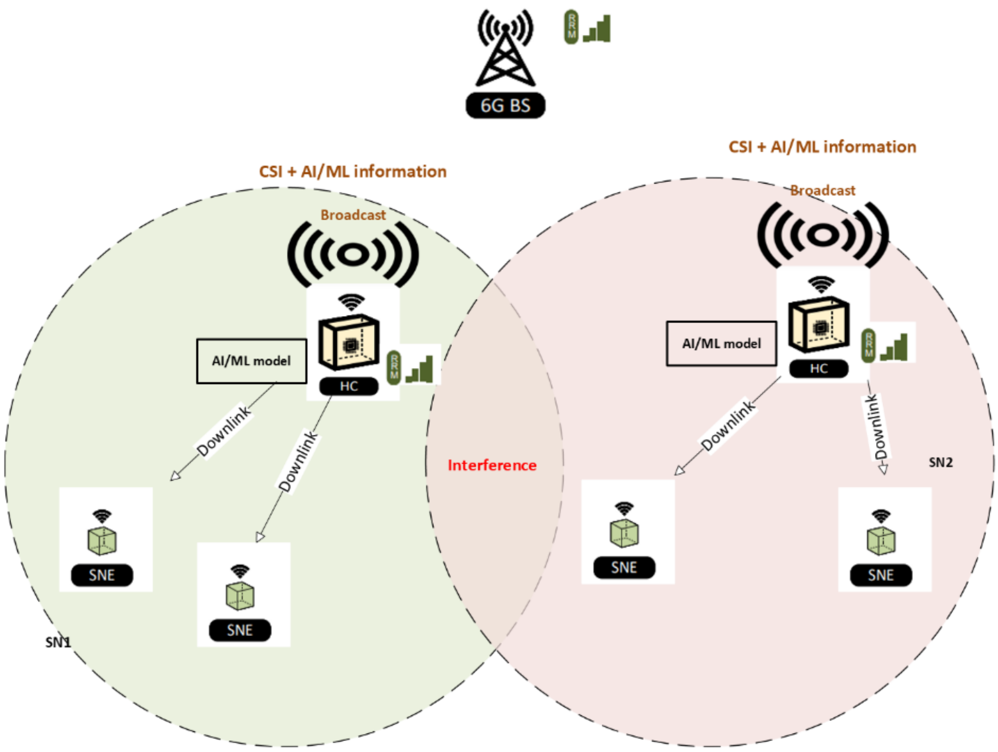


Figure 19: Overall framework for AI/ML enabled RRM

The solution described in D4.1 introduces functional requirements that call for specific architectural enablers within the subnetwork architecture to support its implementation. These enablers are crucial for the integration of AI/ML models, such as Graph Neural Networks (GNNs) and Message Passing Neural Networks (MPNNs), which compute optimal power settings. To achieve this, the HC nodes must be capable of:

- **Processing real-time feedback,** such as Channel State Information (CSI) and node states.
- **Exchanging critical information** to understand channel conditions and interference levels.
- **Dynamically adapting transmission parameters** based on AI/ML model predictions to optimize communication and reduce interference.

To meet these requirements, the AI/ML model is integrated into the MAC layer of the HC node's protocol stack. As shown in Figure 20, the model processes **CSI + AI/ML information feedback** received from the SNEs and dynamically takes RRM decisions for data transmission. In subnetwork deployments legacy CSI gives an immediate snapshot of channel quality for quick decision-making, but it misses broader network dynamics and temporal variations. To address this, AI/ML information is integrated within CSI report. In the solution described in **D4.1** this AI/ML data includes nodes state information to support the Message Passing exchange required. However, in a more general sense this AI/ML information could incorporate a wide range of additional metrics-such as statistical interference patterns, traffic load variations, user mobility trends etc. This combination of CSI and AI/ML information can lead the model to more adaptive and precise resource management strategies. Placing the AI/ML model in the lower levels allow it to decide at every Transmission Time Interval (TTI), enabling the system to respond quickly to changes in channel conditions and interference patterns. Moreover, since the AI/ML model is embedded within the HC node's protocol stack, only the trained model weights need to be transferred to the HC which minimized the overhead for potential updates. In the D4.1 solution, the training phase of the AI/ML model is performed centrally by the network using aggregated subnetwork data, mirroring the NWDAF functionality in 3GPP which can perform such a task. With this approach, the HC node is spared from any extra computational work. The implications of such a choice along with the model sharing to the HC nodes will be discussed in the next section.

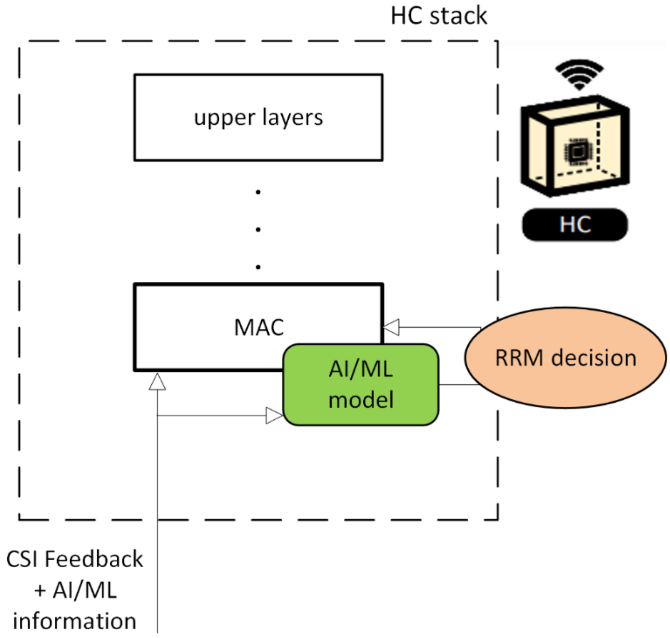


Figure 20: Placement of AI/ML in HC stack.

4 ARCHITECTURAL IMPLICATIONS ON INTERFACES BETWEEN SUBNETWORKS AND 6G PARENT NETWORK

This chapter focuses on the architectural requirements for integrating subnetworks into the 6G parent network. It examines the necessary interface enhancements and network functionalities that support the operation of subnetworks within the 6G framework. While subnetworks can function independently to some extent, their operation often depends on coordination with the parent network for tasks such as resource management, authentication, and data handling. This chapter outlines the key architectural elements needed for this integration.

The discussion includes mechanisms to manage non-transparent subnetworks, allowing the 6G network to maintain virtual connections and enable efficient scheduling and uplink reporting. New approaches for UE-to-UE authentication are introduced to enhance security and reduce network dependence. The chapter also proposes a Non-3GPP Subnetwork Management Function as a central component to coordinate resource allocation and minimize interference between subnetworks and the parent network. The Subnetwork Tunnelling Protocol is described as a method to simplify data aggregation and improve network efficiency.

The following subsections describe the specific architectural components and their integration with the 6G parent network. These include the management of subnetworks through enhanced interfaces, the role of centralized resource coordination frameworks like the N3SMF, and AI/ML-based resource management to support dynamic operations. Additionally, the report covers exposure functions for a deep-edge-to-cloud continuum, ensuring efficient communication and computation across the entire network. These elements aim to establish a practical framework for subnetwork integration within the 6G architecture.

4.1 Integration of Subnetworks into 6G

This section focuses on modifications required in interfaces, authentication mechanisms, and network functionalities, emphasizing the balance between subnetwork autonomy and support from the 6G parent network. The topics include non-transparent subnetwork configurations with virtual connections, subnetwork scheduling through base station-based control, and advanced UE-to-UE authentication for secure subnetwork operations. Additionally, the role of a centralized Non-3GPP Subnetwork Management Function is explored as a core enabler for resource allocation and coordination across heterogeneous networks.

4.1.1 Non-Transparent Subnetworks and Virtual Connections

As highlighted in D4.2 Subsection 2.2.1.1 [4], there are multiple architectural solutions possible to integrate subnetworks into the 6G parent NW [4]. The first option is to make subnetworks completely transparent to the NW. In this way, privacy and flexibility of UEs are improved, since the overlay NW is completely unaware of the subnetwork. This would impose rather demanding capabilities on the MgtN, as this device would be required to impersonate multiple UEs on their behalf and therefore instantiate multiple UP and Control Plane (CP) entities in parallel. Explicitly, this significantly increases the cost of devices that can take the role of the MgtN, while at the same time it reduces the potential of optimization in the RAN, since the 6G BS is incapable of benefiting from the presence of the subnetwork. Hence, the integration of subnetworks into the 6G parent NW in a not fully transparent manner is suggested in D4.2, Subsection 2.2.1.3 [4]. The NW shall be aware of the formation of the subnetwork, the role of the MgtN and the members of the subnetwork without having all the management and control functionality at the 6G BS side. The BS can reduce the size of UE contexts and the related procedural efforts for every device within the subnetwork by deriving the related information from the MgtN context, maintaining only virtual connections for the subnetwork UEs as portrayed in Figure 21.

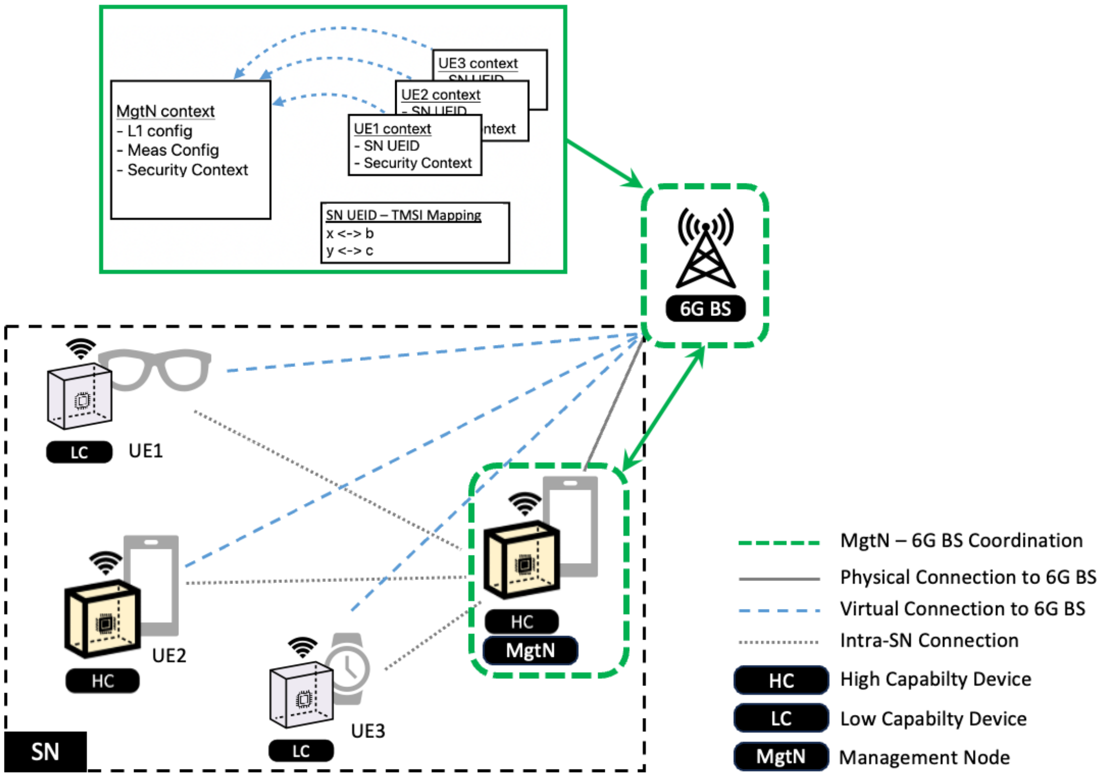


Figure 21: MgtN context stored at the 6G BS side. The green annotations refer to the entities involved in the subnetwork setup

The architectural impact of integrating subnetworks into the 6G NW and, in particular, into the RAN can be summarized like this:

- New UE ID is required within the RAN (SN UE ID), mapped to e.g. Temporary Mobile Subscriber Identity (TMSI)
- Maintenance of single physical link towards the MgtN used for all UEs within the subnetwork
- Lean UE contexts, linked towards the MgtN context.

UEs within the subnetwork can be addressed by their SN UE ID whenever required. For example, when a UE leaves a subnetwork the 6G BS can generate a full UE context and allow for seamless mobility into and out of the subnetwork as described in further detail in D4.2 [4].

4.1.2 UE Dedicated UL Grant and Subnetwork Scheduling

The non-fully transparent subnetwork and virtual connection concept, presented in Section 4.1.1, tries to limit the complexity in the MgtN. Along this line, the mechanisms for scheduling UEs within a subnetwork also need to be evolved. The *Integrated Access Backhaul* (IAB) concept of 3GPP [12] provides a hierarchical architecture of IAB-nodes, where the IAB-nodes themselves perform scheduling individually on every hierarchy level, resulting in several issues. To begin with, IAB nodes need to be more complex devices, since they require their own schedulers, similarly to mobile BSs. Additionally, IAB nodes require a tight integration with their parent node, the IAB donor, via F1 interface [12] to be controlled with respect to scheduling policies.

In comparison to IAB, a MgtN within the subnetwork shall be less complex, hence it is proposed to keep all the scheduling decisions at BS side while the subnetwork is connected to the parent NW. The BS shall schedule all the UEs inside and outside of a subnetwork in the same way. The only difference shall be that UEs inside the subnetwork needs to be scheduled via the MgtN, whereas the others shall be scheduled directly, as shown in Figure 22 (UE3 vs UE5). This allows for End-to-End QoS through the subnetwork, controlled by the BS, even though the subnetwork might not operate on BS-owned resources. To enable this scheme, the interfaces between the 6G overlay NW and the subnetwork should be enhanced. For the Uu-interface, a new per-UE Buffer Status Report (BSR) has to be introduced, as shown in Figure 22, Step (1). This allows the MgtN to report UE-individual buffer status reports, whilst an aggregated BSR is sent in the case of IAB. The 6G BS can use this per-UE information to perform individual scheduling of devices through the subnetwork, as shown in Step (2) of Figure 22.

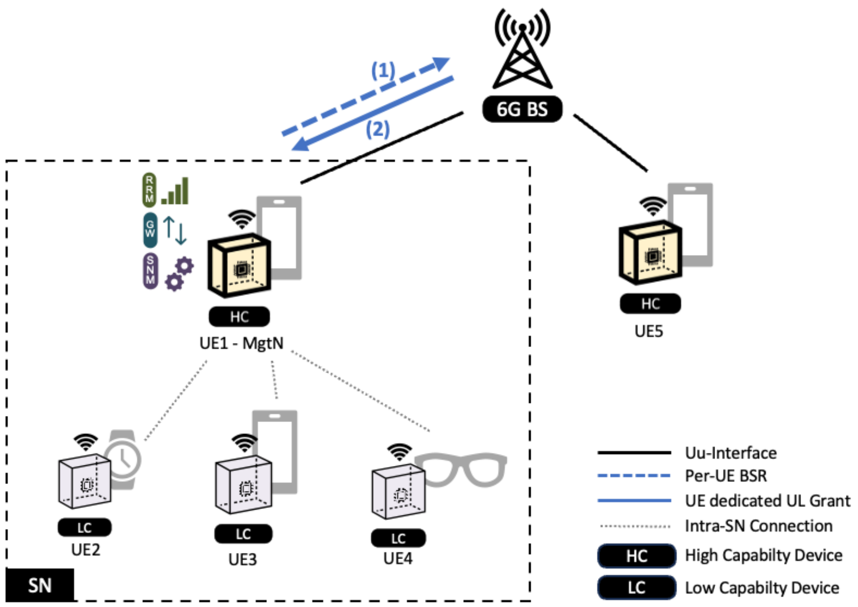


Figure 22: UE dedicated UL Grant with buffered UE Data

More advanced schemes are possible to enhance this solution even further by making it more flexible, where no additional data buffering at the MgtN is required. Instead, the MgtN may pull the data from a device within the subnetwork on demand at the exact moment when the corresponding UE dedicated UL grant is received from the BS in Step (2) of Figure 22. The architectural implications will be further investigated in D4.4.

4.1.3 Subnetwork Authentication

A solution for authentication of subnetwork devices needs to be provided to enable secure and private communication within the subnetwork. In the current state of the art in 3GPP, the CN is fulfilling the role, using AMF/AUSF in 5G [13] or dedicated servers, such as the Proximity Services (ProSe) application server [14]. To enable more flexible topologies, more UE control and more CN independence from the CN, the authentication needs to be brought closer to the UEs, and in that sense closer or into the subnetwork to achieve:

- Lower NW load: due to less NW interaction
- Less NW complexity: no need to deploy special NF/servers in the CN
- Subnetwork independence (Survivability)
- Faster authentication: due to lower number of hops.

D4.4 investigates a new scheme of mutual authentication of UEs while forming a subnetwork. It is utilizing the fact that those UEs are already registered and trusted by the 6G NW. Figure 23 depicts this new scheme for UE-to-UE authentication on a high level, identifying three alternatives:

1. Authentication via prior paring
2. Authentication with BS Assistance
3. Authentication via application layer.

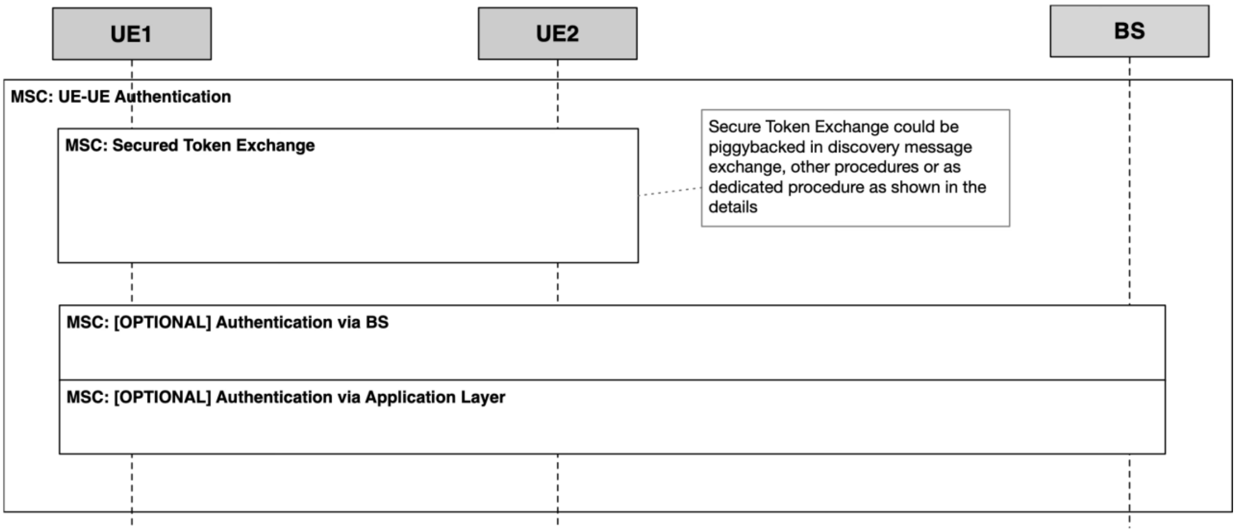


Figure 23: UE-to-UE Authentication Schemes

Architecture-wise, a 6G NW shall provide mutual authentication as a service towards its UEs in order to support them to individually form subnetworks based on their own specific needs. As shown in Figure 23, a RAN node, such as the 6G BS, can provide this functionality, the details will be described in D4.4.

4.1.4 non-3GPP Subnetwork Management Function as a central management component

4.1.4.1 Introduction

We develop architectural solutions for subnetworks primarily operating on non-3GPP technologies (e.g., Wi-Fi) while maintaining connectivity to a 3GPP-based parent network. This approach enables centralized RRM, leveraging the parent network for resource management and coordination. The work is primarily motivated by the Proof-of-Concept (PoC) presented in WP5 [15] of 6G SHINE, where we propose a centralized RRM framework to address the challenges of managing multiple adjacent subnetworks. The goal is to create a 6G-enabled parent network capable of identifying and managing adjacent subnetworks, even those adhering to non-3GPP standards.

More specifically, the proposed solution is built on three key components. First, a centralized RRM entity within the parent network's core oversees resource management across all connected subnetworks, acting as the decision-making hub to ensure efficient allocation and minimize interference. Second, an HC device is responsible for establishing non-3GPP in-X wireless connectivity, serving as a mediator between the parent network and subnetwork elements by translating high-level RRM directives into actionable configurations. Third, Subnetwork Elements represent applications that require wireless connectivity, such as automotive sensors (e.g., cameras and radars), which demand reliable, low-latency communication to support critical functions.

This solution and its components will be implemented and demonstrated in WP5 final deliverable, with a specific focus on showcasing the real-time interplay between the SN building blocks and their ability to deliver an integrated and cohesive solution. Specifically, we expect to provide the following key functionalities:

- Continuous observation of external (non-3GPP) subnetworks to gather real-time insights into their operational state, including traffic load, spectrum usage, and interference levels.
- Establishment of network functions that identify, register, and enable interoperability between non-3GPP subnetworks and the 3GPP parent network. This ensures smooth handover and coordination of resources across heterogeneous domains.
- Once integration is successful, the parent network assumes control over RRM for non-3GPP subnetworks. Using contextual knowledge such as demand patterns, environmental conditions, and interference data, the parent network dynamically allocates resources to optimize performance.

We aim an approach that ensures minimal interference, optimized spectrum usage, and enhanced reliability, particularly in scenarios with high connectivity demands. This solution is particularly beneficial in ultra-dense environments, such as factory floors, where numerous subnetworks operate in proximity, such as Autonomous Guided Vehicles (AGVs), machines, and wireless sensors. Subnetwork interactions in such scenarios can degrade connectivity performance, leading to delays or interruptions in mission-critical tasks.

4.1.4.2 Proposed Architectural Elements

The proposed architecture addresses these challenges by enabling centralized control, where the 6G parent network coordinates resource allocation, manages interference, and ensures reliable connectivity. These features align with use cases outlined in [2] and are further analysed in [4]. Figure 24 illustrates the proposed concept, highlighting three key components:

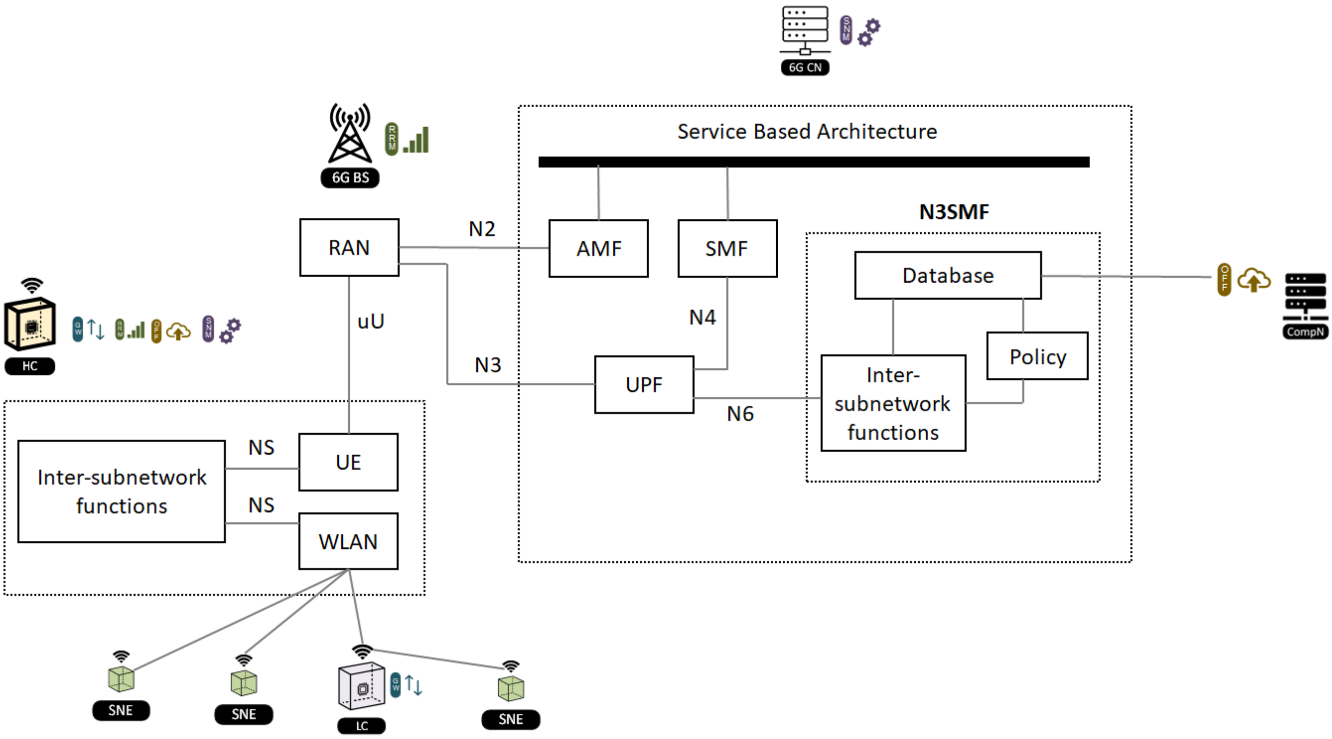


Figure 24: Detailed architectural concept of the proposed N3SMF.

1. **N3SMF Building Block**: Positioned at the 6G Core Network, this is the central enabler of the proposed solution. The N3SMF can be viewed as an advanced evolution of the N3IWF function defined by 3GPP [16]. While the N3IWF primarily facilitates secure WLAN access to the 3GPP Core Network, the N3SMF extends this functionality by managing a broader range of non-3GPP devices beyond Wi-Fi, including wired technologies such as Ethernet, CAN, and LIN, which are critical for automotive applications. Additionally, unlike the N3IWF, the N3SMF is expected to provide not only registration and authentication but also RRM.
2. **Inter-Subnetwork Functionality**: This element is implemented at both the Subnetwork Controller (e.g., the HC device) and the N3SMF. It encompasses a group of network functions essential for full subnetwork integration. Key functionalities include:
    - **Security Mechanisms**: Ensuring secure registration of subnetworks and authentication of their elements.
    - **Subnetwork Discovery and Management**: Identifying available subnetworks and orchestrating their operations.
    - **Monitoring**: Tracking the operational state and performance of each subnetwork.
    - **Data Management**: Maintaining a database to register subnetwork data and store monitored information. This data repository supports advanced policy functions - similar to xApps or rApps in the O-RAN framework—for efficient subnetwork management and optimization.

3. **NS Interface**: Implemented within the HC device, the NS interface acts as a bridge between 3GPP network components and non-3GPP modules. It facilitates seamless communication and resource alignment, enabling interoperability and efficient coordination between the 3GPP core and non-3GPP subnetworks.

4.1.4.3 Flow of Information in the Proposed Solution: An Example Scenario

We can define a use case scenario to exemplify the proposed concepts, as illustrated in Figure 25. Consider a factory floor equipped with multiple Autonomous Guided Vehicles, each carrying an HC device functioning as a Subnetwork Controller. These HCs provide Wi-Fi connectivity to SNEs, such as cameras or sensors. When an HC enters the network, it undergoes an attachment and registration process with the N3SMF, securely integrating into the 6G ecosystem. During this process, the HC shares its capabilities, current configurations, and supported parameters, which are stored in the N3SMF's database for future management. Once registered, the HC begins transmitting periodic updates to the N3SMF, including its operational status, interference levels, traffic demand, and mobility information, enabling the N3SMF to maintain an up-to-date view of the subnetwork environment.

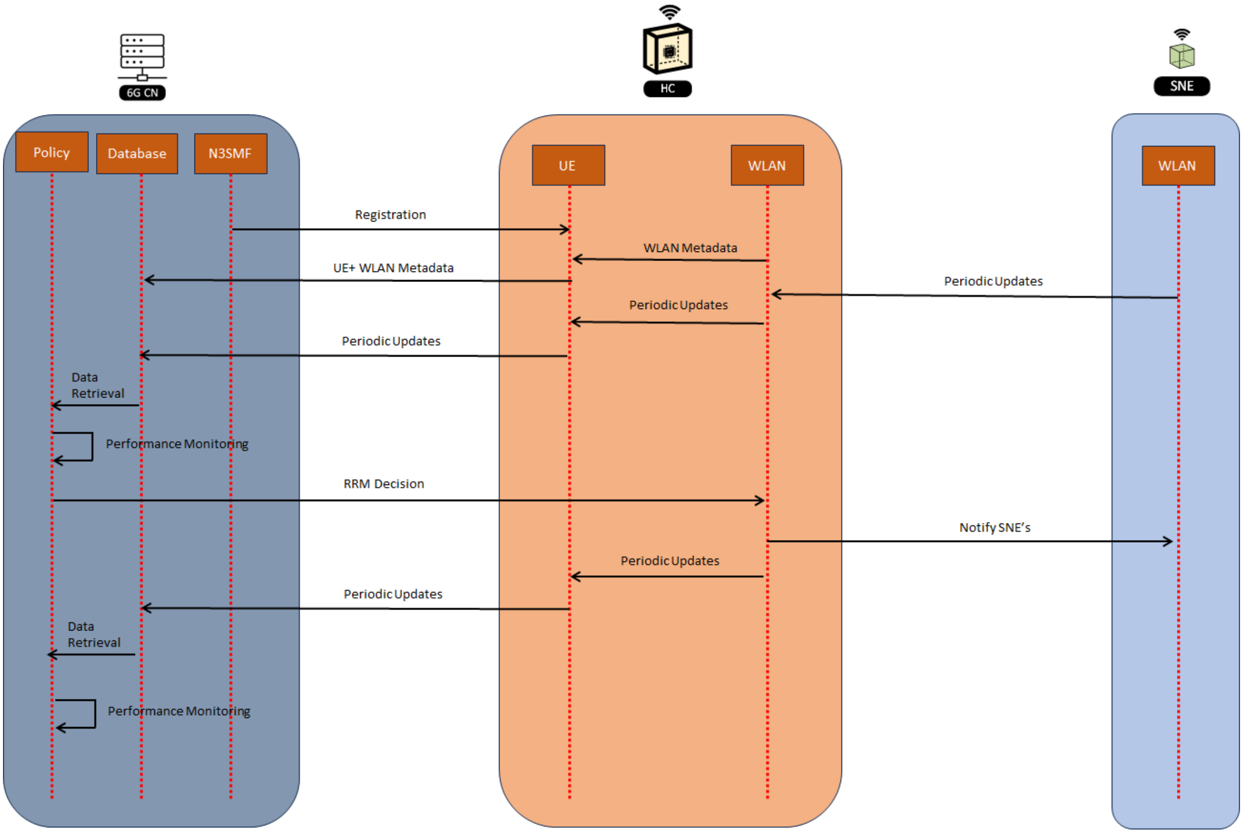


Figure 25: Illustration of the interactions between the N3SMF at core network, HC and SNEs, highlighting the data

As the AGVs move, the N3SMF monitors their proximity using context data, such as location information from localization systems (e.g., UWB or GPS) and interference metrics that can be reported by the HCs or the infrastructure. If the N3SMF detects that two AGVs operating on the same Wi-Fi channel are approaching each other, it identifies the potential for interference and takes proactive measures. Using its policy functions, the N3SMF determines a suitable adjustment to mitigate interference, such as assigning a less congested Wi-Fi channel to one of the AGVs or reconfiguring other parameters like transmission power.

Once a decision is made, the N3SMF sends a reconfiguration command to the target HC via the NS interface. This command specifies the new Wi-Fi channel and any other necessary adjustments. The HC seamlessly implements the change, notifying its connected SNEs to ensure minimal disruption during the transition. After the reconfiguration, the HC resumes normal operation and provides feedback to the N3SMF, including updated metrics on interference levels and subnetwork performance. This feedback allows the N3SMF to confirm the success of the adjustment and refine its decision-making for future scenarios.

4.1.4.4 Main Benefits

By introducing the N3SMF as a key enabler within the 6G Core Network, the architecture ensures that non-3GPP subnetworks, such as Wi-Fi-based systems, can be securely registered, monitored, and managed. This centralized entity can dynamically allocate resources, optimize network performance, and reduce interference by leveraging real-time context awareness (e.g., mobility, spectrum usage, and environmental data). The inter-subnetwork functions further enhance this capability by providing mechanisms for subnetwork discovery, secure registration, and continuous monitoring, enabling efficient decision-making for resource management and policy enforcement.

It is important to highlight that the proposed approach differs from traditional Wi-Fi Distributed Coordination System (DCS) solutions by introducing a centralized, context-aware management framework that extends beyond the reactive mechanisms typically employed in Wi-Fi networks. Unlike DCS, which primarily relies on local channel selection and contention-based access mechanisms to mitigate interference, our solution leverages the N3SMF as a global orchestrator that proactively manages subnetwork configurations based on real-time operational data. By continuously collecting information such as interference levels, traffic demand, and mobility patterns from HCs, the N3SMF enables a more coordinated and adaptive response to network dynamics. Additionally, while Wi-Fi DCS solutions focus on optimizing performance within a single Wi-Fi network, our approach integrates non-3GPP subnetworks into the broader 6G ecosystem. This allows for cross-subnetwork interference mitigation strategies, such as dynamically reconfiguring transmission power, adjusting Wi-Fi channels, or incorporating external context data like localization information (e.g., UWB or GPS).

A major benefit of this architecture is its adaptability to dynamic and high-density environments, such as factory floors, smart transportation systems, and urban IoT deployments. By allowing the N3SMF to dynamically reconfigure Wi-Fi parameters (e.g., channel selection, transmit power), the architecture minimizes spectrum contention and optimizes resource utilization across subnetworks, even in mobility scenarios. This ensures reliable and uninterrupted connectivity for critical applications, such as real-time video streaming or sensor data transmission. Additionally, the centralized database and policy functions enable feedback loops for continuous performance monitoring and optimization, creating a self-adaptive system that can respond to changing conditions in real time, thus improving scalability, efficiency, and overall network performance.

4.1.5 Implications of centralized AI/ML based RRM for interference mitigation

While the enablers described in Section 3.3.3 focus on subnetwork level integration of AI/ML for RRM, the effectiveness of these enablers heavily depends on the interaction and interfaces with the 6G parent network. Figure 26 provides a detailed depiction of how the 6G base station facilitates efficient interference management at the subnetwork level by enabling the integration and operation of AI/ML models within HC nodes. HC nodes utilize the AI/ML models to perform resource management in a distributed manner fulling their RRM role inside subnetworks. In this specific solution, the 6G parent network does not perform RRM itself; it rather supports the RRM role of HC nodes through the centralized offline training and the distribution of updated model parameters to the subnetworks. One potential method for this transfer is through the air interface, which aligns with the broader principles of 3GPP collaboration levels [17]. These levels provide a framework for defining the interaction between the network and UEs to enable efficient AI/ML model management. The levels range from minimal interaction, where AI/ML operations are entirely implementation specific, to extensive collaboration involving the exchange of model structures and parameters. At a high level, the collaboration levels are categorized as follows:

- **Level X**: No collaboration. AI/ML operations are fully implementation- specific with no signalling or information exchange between the network and UEs.
- **Level Y**: Signalling-based collaboration without model transfers. In this level, the network and UEs exchange information to improve AI/ML performance but do not share model data.
- **Level Z:** Collaboration involving the transfer of AI/ML models or parameters. This level supports over-the-air model distribution and updates to UEs.

It is important to note that, in the current 3GPP framework, these levels of collaboration are primarily defined for physical (PHY) layer AI/ML models [18]. These models address tasks such as CSI feedback compression, beam management, and positioning. In these cases, over-the-air model transfer occurs between gNBs and UEs. However, RRM in subnetworks introduces unique requirements: RRM decisions are made at the level of UE- like devices, such as HC nodes, rather than through gNB-to-gNB interactions facilitated traditionally by interfaces like Xn [19]. To this end, we envision collaboration level Z to make possible the efficient distribution of the AI/ML models parameters. Among the subcategories of Level Z [18], Z4 is the most suitable for the RRM framework. This subcategory is designed for cases where:

- The **AI/ML model structure is pre-defined and embedded** within the receiving device, as is the case with HC nodes in this framework.
- **Only the trained parameters of the model need to be updated**, reducing the resources required for updates.
- An open format is used for parameter transmission, **ensuring compatibility across heterogeneous HC nodes and network environments**.

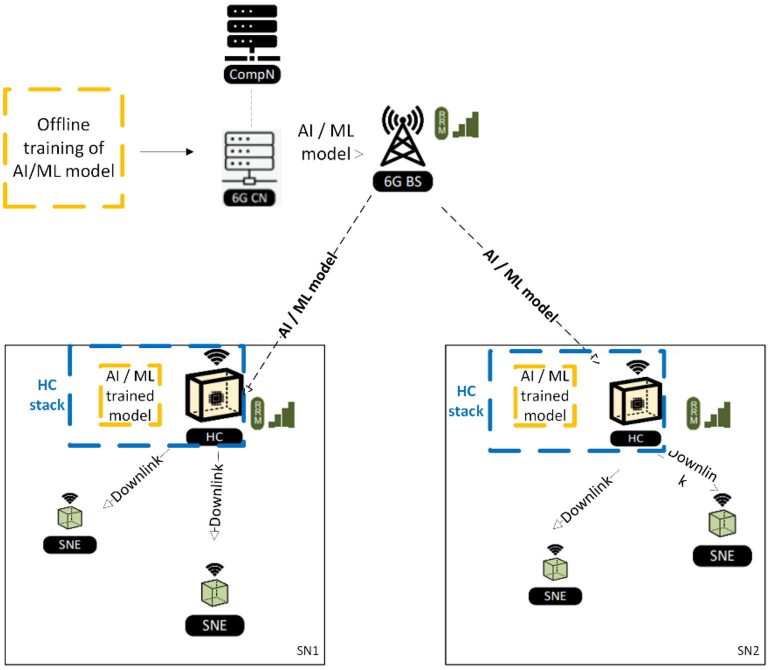


Figure 26: AI/ML Model Training and Distribution

Once deployed, the AI/ML models operate within the HC stack, directly contributing to RRM decisions (power control, resource assignment, etc). When necessary, the 6G parent network initiates model modifications, ensuring alignment with the latest requirements of the dynamic wireless environment. These updates, which may address evolving interference patterns or integrate new learning based on global subnetwork performance data, can be distributed to multiple HC nodes simultaneously using broadcast mechanisms. The sequence of actions involved in this update process - from triggering an update to parameter integration and feedback - is illustrated in Figure 27.

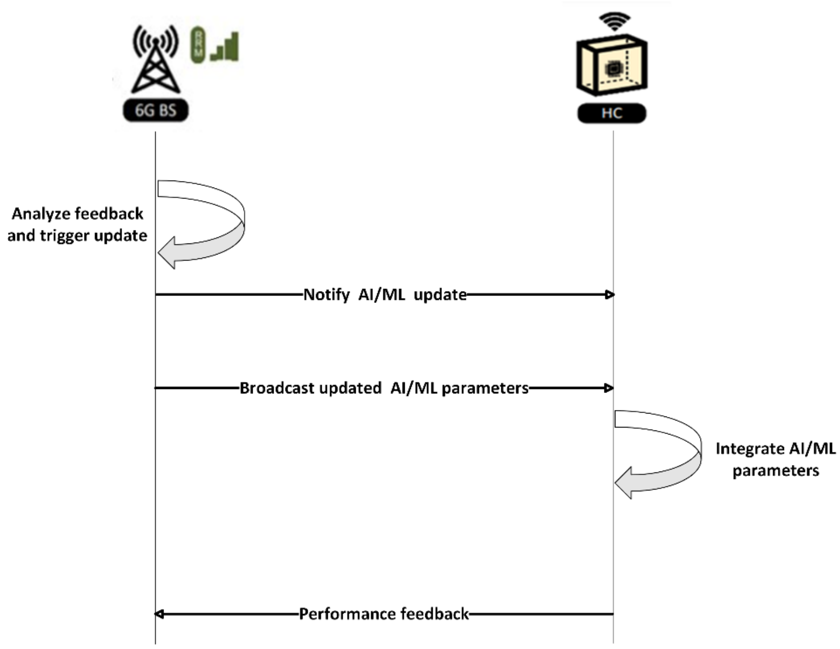


Figure 27: AI/ML model update procedure

4.2 End-to-End Data Handling

The following subsections focus on the data handling and how to enable subnetworks that shall be integrated into 6G. In Section 4.2.1, the interface between the SN and the 6G NW is presented in the form of the Subnetwork tunnelling Protocol (SN-TP), as highlighted in D4.2 [4]. Section 4.2.2 shows how LC devices can be integrated into the 6G NW with the aid of the SNs. Finally, in Section 4.2.3 it is shown how sensing data can be handled in the SN context.

4.2.1 Subnetwork Tunnelling Protocol

The way subnetwork may be integrated in the 6G overlay NW is depicted in more detail in [4] Section 2.2.2.1. There, the limitations of 3GPP *Sidelink* (SL) and IAB are discussed and the new protocol, referred to as *Subnetwork Tunnelling Protocol* (SN-TP), is introduced. The SN-TP aims to have more flexibility on how to aggregate the data of different UEs on different levels in the protocol stack. The proposal suggests aggregating different UEs, or UE QoS Flows into fewer (even one) Data Radio Bearers (DRB), Radio Link Control (RLC) channels or Logical Channel (LCh) entities between MgtN and 6G BS. In this way, the MgtN complexity in terms of number of DRB/LCh entities is limited for each device in its subnetwork. The SN-TP protocol provides UE and QoS flow mapping information per packet at the link between MgtN and 6G BS. This information is portrayed in Figure 28 for both UL and DL. It could be deployed on different layers depending on where to aggregate the UEs and their respective QoS Flows within the subnetwork. Two possible deployments are shown in the right hand-side of Figure 28. Example 1 shows the deployment of SN-TP above RLC, where the whole SN traffic can be multiplexed into a single RLC channel or alternatively a single RLC channel per UE could be established with all DRBs of a UE aggregated. In Example 2, the SN-TP is deployed below RLC, where UEs' Logical Channels are multiplexed into a common Transport Block (TB) per UE, thus maintaining different HARQ (Hybrid Automatic Repeat reQuest) processes for different UEs.

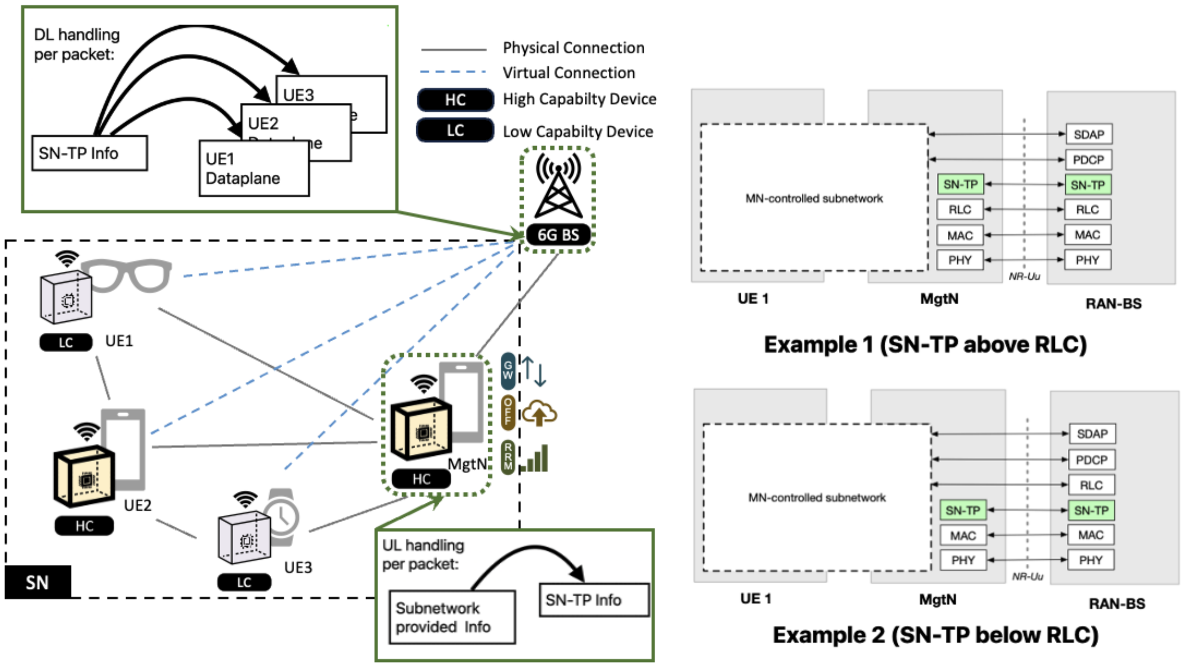


Figure 28: Subnetwork Tunnelling (SN-TP) configuration and subnetwork setup (left) and UP protocol stack (right).

More details on SN-TP are outlined in D4.2 [4]. In summary, both the UEs that wish act as a MgtN connecting a subnetwork to the 6G overlay NW and the 6G BSs need to support of the SN-TP protocol.

4.2.2 Non-Standalone Low Capability Devices Within Subnetworks

With the introduction of subnetworks as a topology option a new category of LC devices, the so-called Non-Standalone (NSA) devices can be enabled as described in D4.2 [4]. In Figure 29, such NSA LC devices that have the ability to access the 6G NW for data communication while not being fully capable standalone cellular devices are shown, namely UE2 and UE3. They may require direct 6G BS connections to remove unnecessary hops including the associated delays to fulfil the low-latency requirements imposed by the various use cases explained in D2.1 [1] and D2.2 [2].

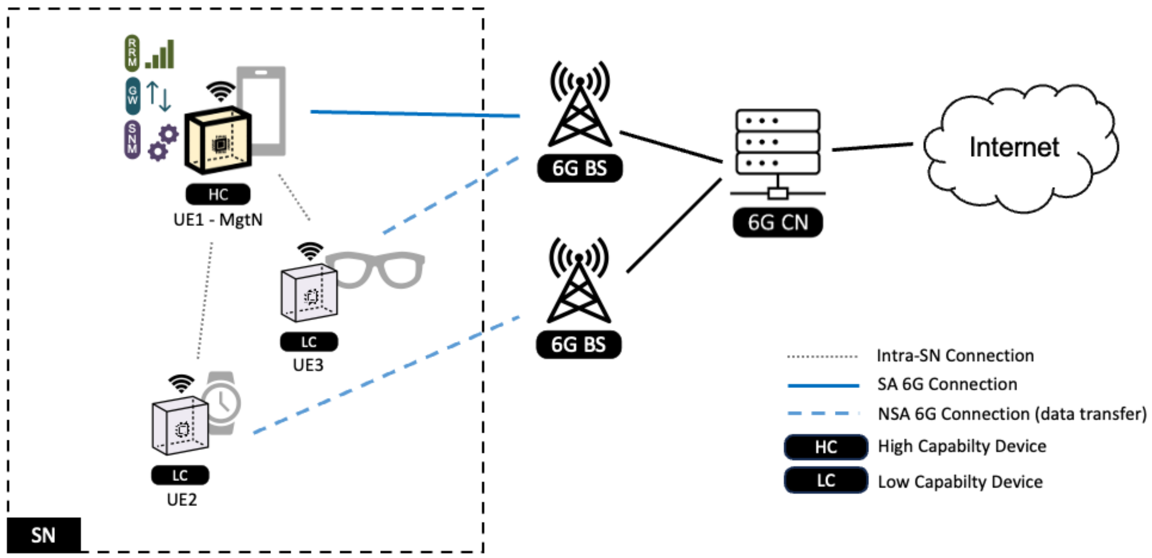


Figure 29: Subnetwork enabling Non-Standalone LC devices (UE2 and UE3) for direct 6G connections [4]

Figure 30 shows the envisioned high-level configuration architecture and the necessary steps for establishment of a direct link between NSA LC UE and 6G NW. The basic flow involves:

- an NSA-UE requesting a direct radio link from its MgtN in Step (1) of Figure 30
- the MgtN translating and forwarding that to the NW in Step (2) of Figure 30
- a configuration from the NW towards the NSA-UE in Steps (3) and (4) of Figure 30

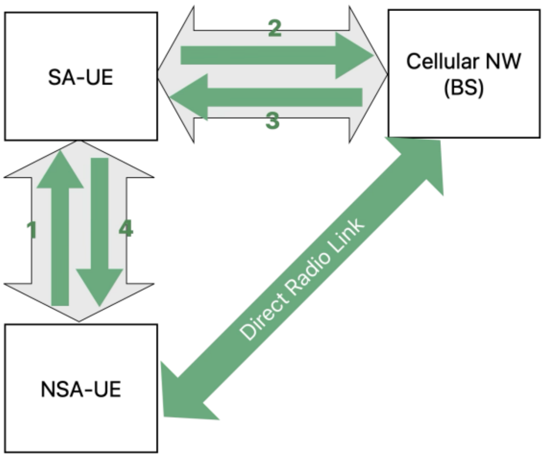


Figure 30: High-level Configuration Architecture for direct links to NSA LC devices

Naturally, certain procedural enhancements in the 6G NW as well as the MgtN are required to enable the deployment such NSA within a subnetwork:

- Authentication/Security enhancements
- Configuration procedures
- Direct Link establishment
- User Plane enhancements.

The details of the above-mentioned procedures and the message flow between the involved devices shown in Figure 30 will be investigated in more detail in the D4.4.

4.2.3 Processing of Sensing data

The sensing use cases presented in [22] include a list of sensing KPIs. These KPIs are sensing results that cannot be estimated without computation of Sensing data. Figure 31 illustrates a workflow for sensing tasks, that starts with the transmission of a sensing signal and ends with the generation of an estimated sensing result.

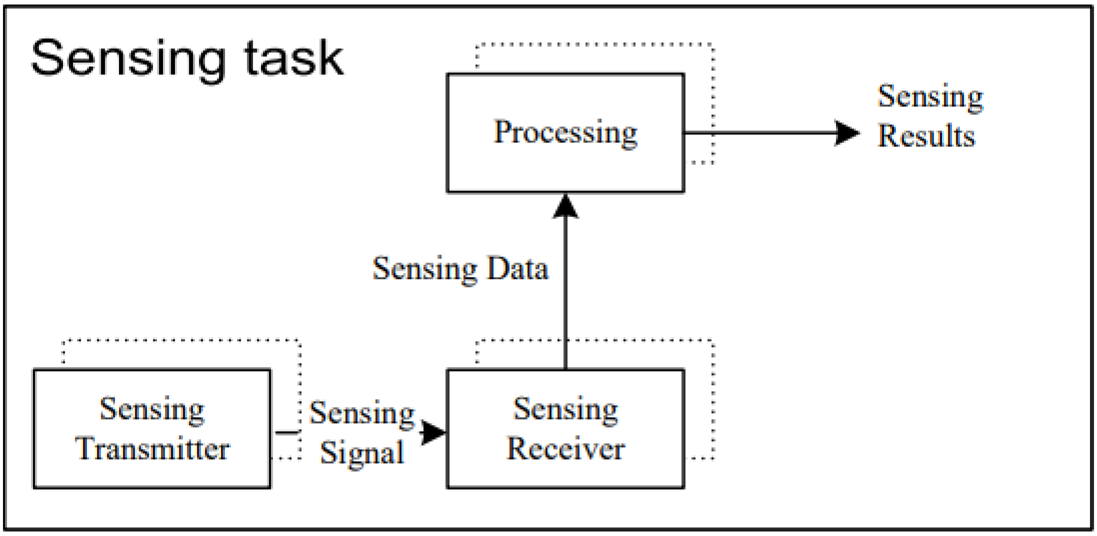


Figure 31: Sensing task workflow for Sensing KPI value generation

The sensing transmitter and receiver pair is suitable for RF based sensing within the subnetwork. It is however not required in cases where the available sensors are non-6G based sensors such as cameras or LiDAR, that generate video frames and 3D point cloud data, respectively. In these cases, it is still the sensing receiver that generates sensing data that in turn requires processing, to generate a sensing result. Available sensing results, regardless of the sensing source type are then used for assessment of sensing KPIs in [22].

Processing of sensing data can happen in any capable node within the subnetwork, or in the CN, where processing capabilities are higher, as illustrated in Figure 18, via the SAF functionality. The challenge of processing sensing data is complex and addressed in more details in WP4 reports. In some cases, however, where the processing of sensing data happens in the CN, it is important to assign compute roles to elements in the subnetwork. An example of such cases is given in section 4.3.3.

4.3 Offloading to/from the 6G Network

This section introduces architecture innovations arising for supporting control plane (CP) function and data plane offloading to/from the 6G network from/to the in-X subnetworks. This section identifies and shows approaches to handle the security challenges of CP offloading when CP functionalities are delegated to other devices within a subnetwork. This section also extends the interfaces and exposure functions introduced in Section 3.2.2 to expand the continuum of communication and computing resources across the integrated subnetwork-edge-cloud (also referred to as deep edge-edge-cloud or IoT-edge-cloud [20][21]) for the advanced and efficient data plane offloading. These architecture innovations are meant to address the increasing flexibility and dynamic demands of softwarised entities like vehicles and robots, supporting compute-intensive applications such as autonomous driving. The envisioned 6G subnetwork–edge–cloud continuum facilitates opportunistic offloading and distribution of processing tasks across the continuum, providing a cost-efficient solution for executing resource-intensive processes that cannot be handled solely within the subnetworks. Offloading can also be necessary from the 6G parent network to the subnetwork. This is the case for example when the subnetwork sensing conditions are poor. Transferring the sensing results to the 6G parent network could be considered a waste of resources, especially for resource-constrained devices like the ones in subnetworks. The proposal is then to transfer SAF functionalities normally deployed in the 6G parent network to the subnetwork when this poor sensing conditions are identified. The offloading of the SAF functionalities to the subnetwork enables local decision-making about the sensing's results KPI and hence avoids sending poor data from the subnetwork to the parent 6G network in this case.

4.3.1 Control Plane Function Offloading

The Control Plane (CP) offloading was described in [4]. When a certain CP functionality involving communication with the overlay NW needs to be offloaded to other devices in the subnetwork, certain security implications arise. The device that takes over a certain function must be able to communicate with the NW on behalf of the offloading devices which requires enhancements to the security in 6G. In the example shown in Figure 32, a set of devices, like a group of pupils in the immersive education use case, is moving together as a group and each device is performing Location Updates (LU) individually. As an improvement, these devices could instead form a subnetwork and offload the corresponding CP functionality for LU to their MgtN.

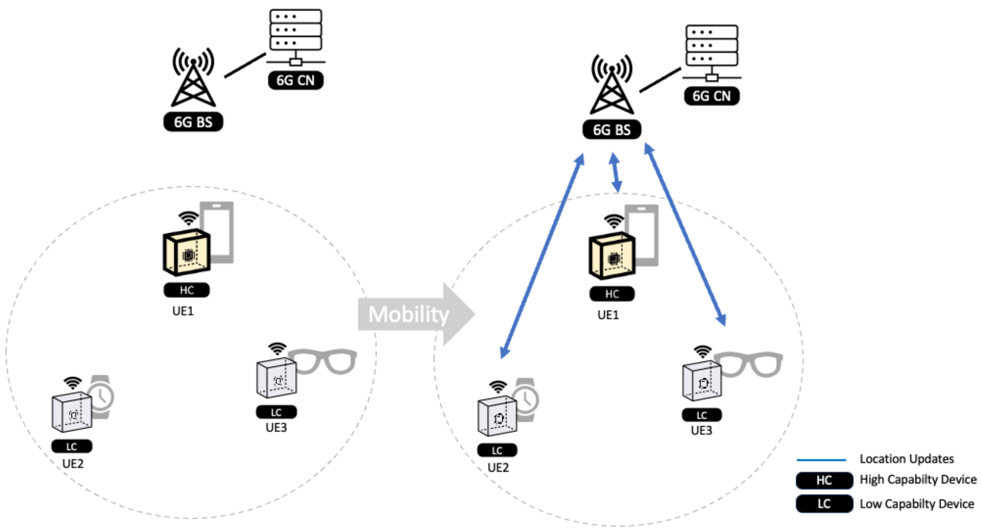


Figure 32: Individual Location Updates by nodes moving together

This would imply that the MgtN performs those procedures on behalf of other devices in a privacy-preserving and secure manner. Consequently, the CN requires enhancements to distinguish that a certain procedure, like an LU, is meant for a group of UEs and not only of the device that performs the procedure, like the MgtN in this case. Those enhancements shall ensure no sharing of security-related details, such as security contexts, among UEs within the subnetwork. In the ongoing work, different approaches have been identified:

- exchange of SecuredIDs between the UE and the CN, e.g. during registration
- establishment of a GroupID while forming a subnetwork.

Both mechanisms can be used for mapping and verification by the NW as shown in Figure 33. A more detailed description and the architectural options will be presented in D4.4.

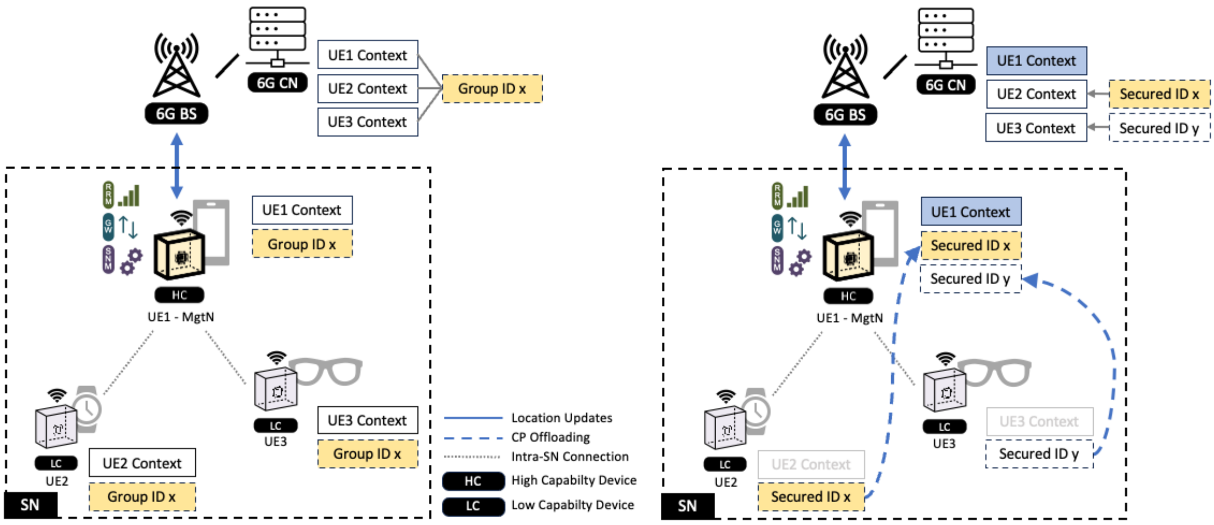


Figure 33: CN Enhancements for CP offloading

4.3.2 Control Functions for the Subnetwork – Edge – Cloud Continuum

Section 3.2 has presented subnetwork architectural innovations for extending network exposure functions currently available in 5G(-Advanced) systems. The exposure functions described in Section 3.2 enable a communication and computing continuum at the deep edge or the subnetwork domain and are suitable for supporting dynamic and increasing flexibility demands of softwarised entities at the deep edge (such as vehicles or robots). The growing number of compute intense applications, functions and control processes at the deep edge (e.g. for autonomous driving) could not be efficiently accomplished by only scaling, re-dimensioning and re-designing the networks at the deep edge. The cost-efficient execution of these compute-intense processes can be enhanced by leveraging the envisioned 6G deep-edge – edge – cloud continuum, which enables the opportunistic offloading and distribution of processing tasks within the deep edge and to the edge or cloud [4].

4.3.2.1 Advanced exposure functions towards the subnetwork – edge – cloud continuum

Achieving the 6G deep edge – edge – cloud continuum vision requires of seamless connectivity and integration of networks at the deep-edge with the 6G parent network. This integration can be facilitated through the (gradual) adoption of wireless subnetworks in entities such as vehicles or robots. For instance, in the scenario of in-vehicle wireless subnetworks, 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 [2]. This vision requires architectural components and their associated subnetwork control functions for the seamless integration of subnetworks with the 6G parent network, as well as the components and interfaces necessary for a joint orchestration/management and context-aware operation of infrastructure (i.e., computing & communication links), network functions and application layers for the effective and dynamic completion of compute-intense (control) process in the deep-edge – edge – cloud continuum.

This vision is represented in Figure 34. In comparison to the stand-alone operation of subnetworks described in Section 3.2, the integrated operation of subnetwork-6G network with edge and cloud resources requires extending the exposure of network functions to the complete end-to-end domain. In the layered architecture model represented in Figure 34, the extended exposure functions from the infrastructure layer to the network functions layer logically expose communication and computing resources of the subnetwork, the edge nodes, core network nodes and cloud nodes to the CCREF. The exposure of resources is represented in Figure 34 with the orange vertical dashed lines from the infrastructure layer to the network functions layer. The definition of the specific mechanisms and APIs to realize the exposure of these resources is beyond our goal though. Such realization could benefit from progress towards open and harmonized APIs made in projects like CAMARA (https://camaraproject.org/).

It should be also noted that advanced control functions at the network functions layer, like the CCRMF and the NIF (see Section 3.2), could exploit the communication and compute continuum made available by the CCREF in the entire subnetwork-6G network-cloud continuum for the flexible design of policies that satisfy the processing, computing and QoS requirements of task and functions generated in any of the domains. In other words, the integration between advanced control functions CCREF, CCRMF and NIF facilitates an elastic continuum orchestration/management approach where computing tasks and communication links are intelligently, dynamically and jointly scheduled across the nodes and links forming the continuum.

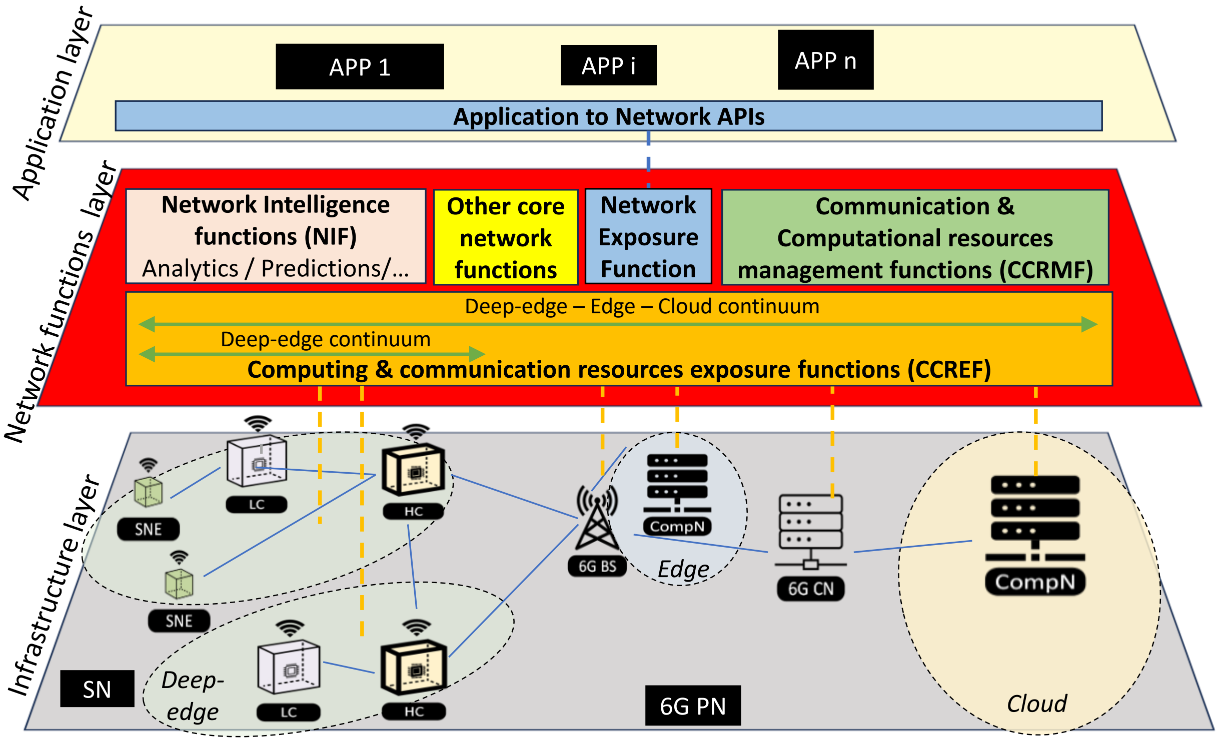


Figure 34: End-to-end layered architecture model with CCREF continuum in the subnetwork-6G edge-cloud domains (dotted lines represent the logical links of the function matching the colour, while solid lines represent physical wireless links).

Figure *35* shows an example of the realization of the system architecture depicted in Figure *34* in the integrated subnetwork - 6G parent network ecosystem. The figure represents the flexibility of such integrated ecosystem that allows flexible functional splitting and implementation of network functionalities in the end-to-end system, e.g., for the collection of communication and computing resources and capabilities (i.e. CCREF) and the management and orchestration of those (i.e., CCRMF and NIF). The example illustrated in Figure 35 shows a possible realization for guaranteeing the survivability of subnetworks when the integration with the 6G parent network cannot be established. This is achieved through the deployment/instantiation of network functions in the subnetwork nodes depending on their capabilities. As described in Section 3.2, the isolated operation of sub-networks can be achieved through the coordinated management of communication and communication resources. HC devices with SNM/OFF/GW roles (or Management nodes as defined in Section 3.1.1) play a key role in this process. These devices jointly manage the resources exposed by other sub-network elements (e.g., LC) to the CCREF. The integrated system can then leverage the interaction between CCREF and CCRMF to implement advanced policies and procedures, ensuring that the service requirements of the SNEs exposed through the NEF are met. The flexible deployment of the network function layer functions also allows for the centralized and partially decentralized (via functional splitting between the subnetwork and the 6G-parent network) management of computing and connectivity resources of the deep-edge – edge – cloud continuum. This is represented in Figure *35* with the allocation of network functions also in the core network nodes of the 6G parent network.

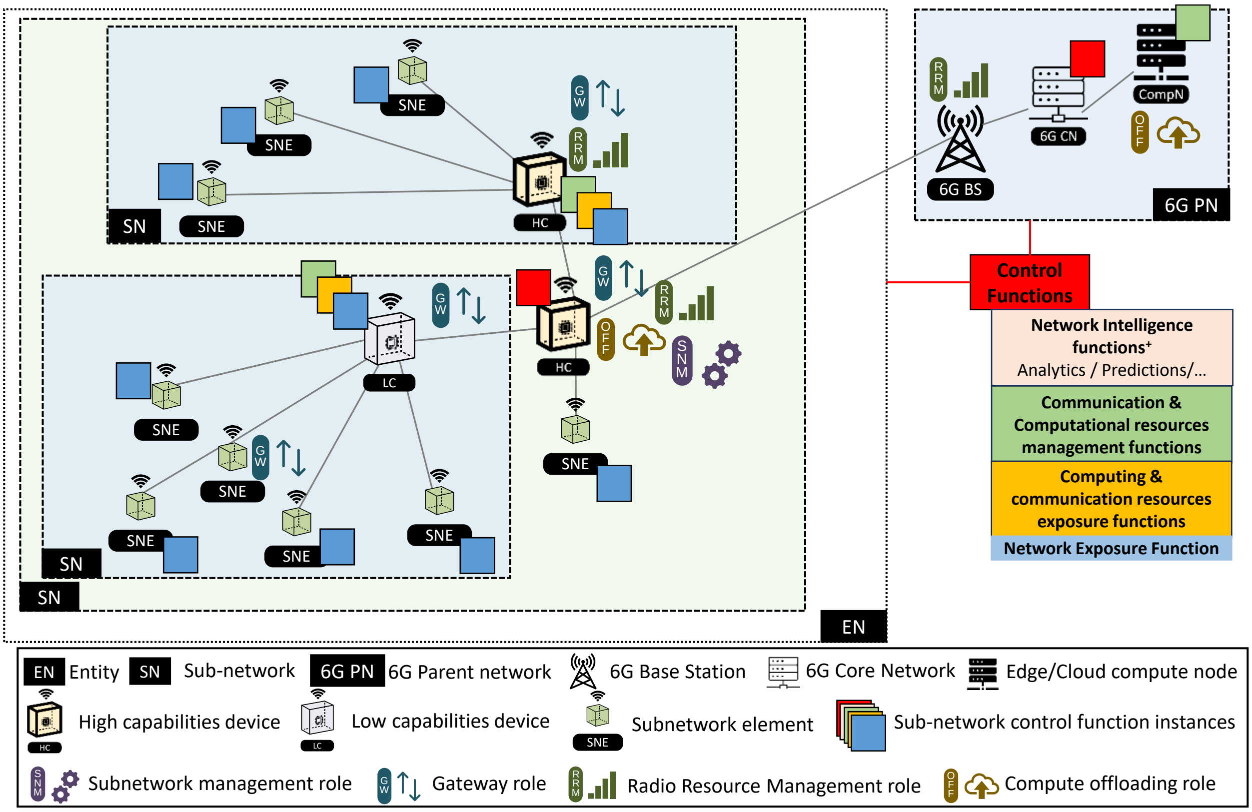


Figure 35: Example of realization of end-to-end layered architecture model with CCREF continuum.

4.3.3 Compute offloading from the parent to the subnetwork

In the context of subnetworks, that require a certain degree of autonomy, the radio resource allocation for sensing can be managed locally, e.g., by an HC node, but the assessment of the sensing operations, after converting sensing data into sensing results, might still make sense to be performed at the parent 6G network.

Examples of usage scenarios where the parent 6G network assesses the sensing operations include sensing operations within a subnetwork that pertains to use cases involving sensing in multiple subnetworks. In case the distance between subnetworks is such that the subnetworks are not able to exchange information, a node in the parent network needs to be responsible for the oversight. With subnetworks closely located, there could be other constraints to sensing information exchanges between subnetworks, such as lack of secure link establishment, the subnetworks pertaining to different domains (e.g., operators), localized exhaustion of radio resources for sensing information sharing, etc.

The oversight and management of the sensing operation is therefore, by default, in the CN of a 6G System. There are however situations where the production of sensing data is not suitable for the generation of sensing results, or not suitable for the generation of trustable sensing results. This can happen due to several factors. Examples include the presence of interference, either legitimate or malicious, insufficiency radio resource allocation that cannot be improved because of, e.g., extreme communication requirements, or poor radio conditions in the link that serves transport of sensing data towards the parent 6G network. But under poor sensing conditions, that cannot be improved, continuous usage of quite valuable radio resources is wasteful, as the subnetwork already has extreme radio resource requirements for communication operations. In these situations, it is more efficient to temporarily handover the oversight of sensing tasks (i.e., the analysis if sensing KPIs are being met, for example) to the subnetwork. Without handing over the oversight of sensing tasks to the subnetwork, the sensing data originated there will continue to be transferred to the parent network, and this could have a significant impact on communication requirements as it represents a transfer over the air of faulty sensing data to the parent 6G network. This assumes that there is no possibility to perform further radio resource management to correct the problem, and this could happen if e.g., the subnetwork is using all allocated radio resources, the subnetwork goes out of coverage, there is a radio resource overload problem in the subnetwork or in the supporting parent network coverage resources, etc.

It is then more efficient in these cases to transfer the sensing KPI assessment to the subnetwork. This can be executed by transferring SAF functionalities to the subnetwork, to either HC, LC or both types of nodes. Many are the potential factors that can influence the node selection. Hence, more details are to be reported in WP4, D4.4. For the purposes of this document, LC/HC are bundled together for illustration purposes.

Figure 36 illustrates an example of sensing compute offload from the parent 6G network to the subnetwork, under poor sensing conditions.

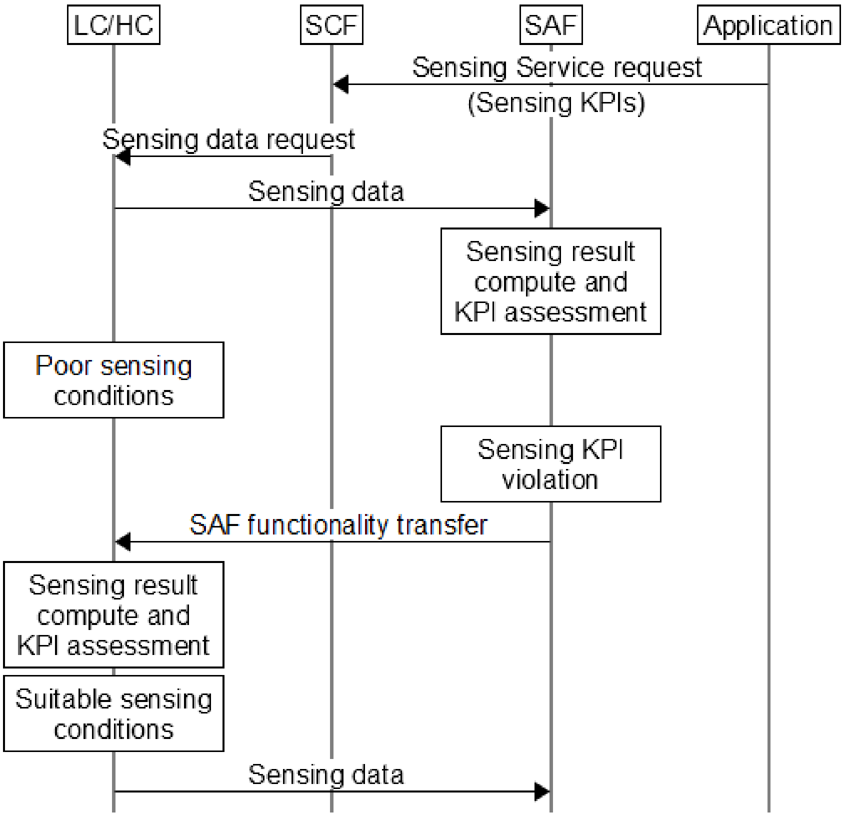


Figure 36: An example of sensing compute offload from the parent 6G network to the subnetwork

The example procedure considers a sensing task that needs to be performed in a subnetwork because of a request from an external application, that uses sensing as a feature. The application is in charge of communicating the sensing KPI targets to the 6G System. The SCF communicates roles to the subnetwork, by activating the reporting of sensing data to SNEs, LCs, and/or HC nodes, providing necessary radio resource configuration. The target nodes in the subnetwork then start this report to the SAF, that will convert sensing data into sensing results and assess if the subnetwork nodes produce meaningful sensing data, i.e., data that is sufficiently good to produce estimated sensing results or data that is sufficiently good to produce estimated sensing results that meet the KPI requirements of the application.

As a result of the sensing conditions degradation, the SAF will detect violation of sensing KPIs. It represents a waste of resources to continue sensing data reporting from the subnetwork to the parent 6G network in this case. The trivial solution would be to switch off this reporting activity but if applied, then the SAF in the CN would not be able to assess exactly when to resume reporting, as soon as the sensing conditions become suitable again - it would need to probe the subnetwork constantly to determine if suitable sensing conditions are again met, which is inefficient in itself due to resulting subnetwork probing signalling overhead.

A more efficient solution is therefore to transfer temporarily the SAF functionalities to the subnetwork. This includes the configuration of compute needs and functions that are in place at a system level, in the subnetwork, e.g., by assigning one or more subnetwork nodes with these compute tasks, that would execute the sensing result processing flow in Figure 31.

With the compute functionalities now active in the subnetwork nodes, the reporting of sensing data is paused. This enables immediate detection as soon as the subnetwork nodes recover, e.g., from the interference problem, and enables sensing data reporting towards the parent 6G network with zero delay.

5 6G-SHINE ARCHITECTURE BLUEPRINT

The work presented in this document pertains to architectural aspects related to the subnetwork concept, including its management and impacted interfaces, both for single subnetwork operation and the interactions between subnetworks. It includes proposals for system level improvements that are required for their integration to a parent 6G network, both at RAN and CN levels, without which subnetworks would not be able to be integrated. It also describes various aspects of computing operations both in local (i.e., subnetwork level) and in an end-to-end manner (i.e., subnetwork-edge-cloud), as compute capable elements are embedded across the system, including cloud, CN, RAN, edge compute fabric, subnetwork, and subnetwork elements.

For a complete appreciation of the works proposed in this document on a high-level, end-to-end manner, Figure 37 depicts in green colour and with letters A-D, the high-level system components where impacts are proposed in this document, as architectural component categories. They represent the enhancements proposed in this document, focusing on:

- A- Intra-subnetwork management – a cluster with all enhancements required for the management of a subnetwork. These include fundamental aspects towards enabling subnetworks such as authentication, node and node group management, routing, tunnelling, virtual connections, compute roles, AI/ML aspects and radio resource allocation;
- B- Inter-subnetwork management – a cluster focusing on control functions for subnetworks connection, operability and service continuity, as well as distributed compute enablers across multiple subnetworks;
- C- Parent 6G network management (RAN level) – a cluster of enablers for subnetworks at RAN level. Many of the aspects covered in A and B have a RAN impact when integrating the subnetworks, as detailed in each section. However, bigger focus is given to AI/ML related aspects, as well as enabling compatibility with O-RAN architectures;
- D- Parent 6G network management (CN level) - a cluster of enablers for subnetworks at CN level. Many of the aspects covered in A and B have a CN impact when integrating the subnetworks, as detailed in each section. However, bigger focus is given in this cluster to CN aspects pertaining to enabling of sensing and ISAC, compute service in the edge-cloud fabric, compute offloading from the subnetwork to the CN and from the CN to the subnetwork, integration and management of non-3GPP subnetworks, and AI/ML services to the subnetworks.

Figure 37 also shows in blue, and with numbers 1-20, the main component(s) and/or interface(s) where the enhancements have impact. This represents the 6G-SHINE architectural blueprint, and it reflects work and implications, including those of interfaces.

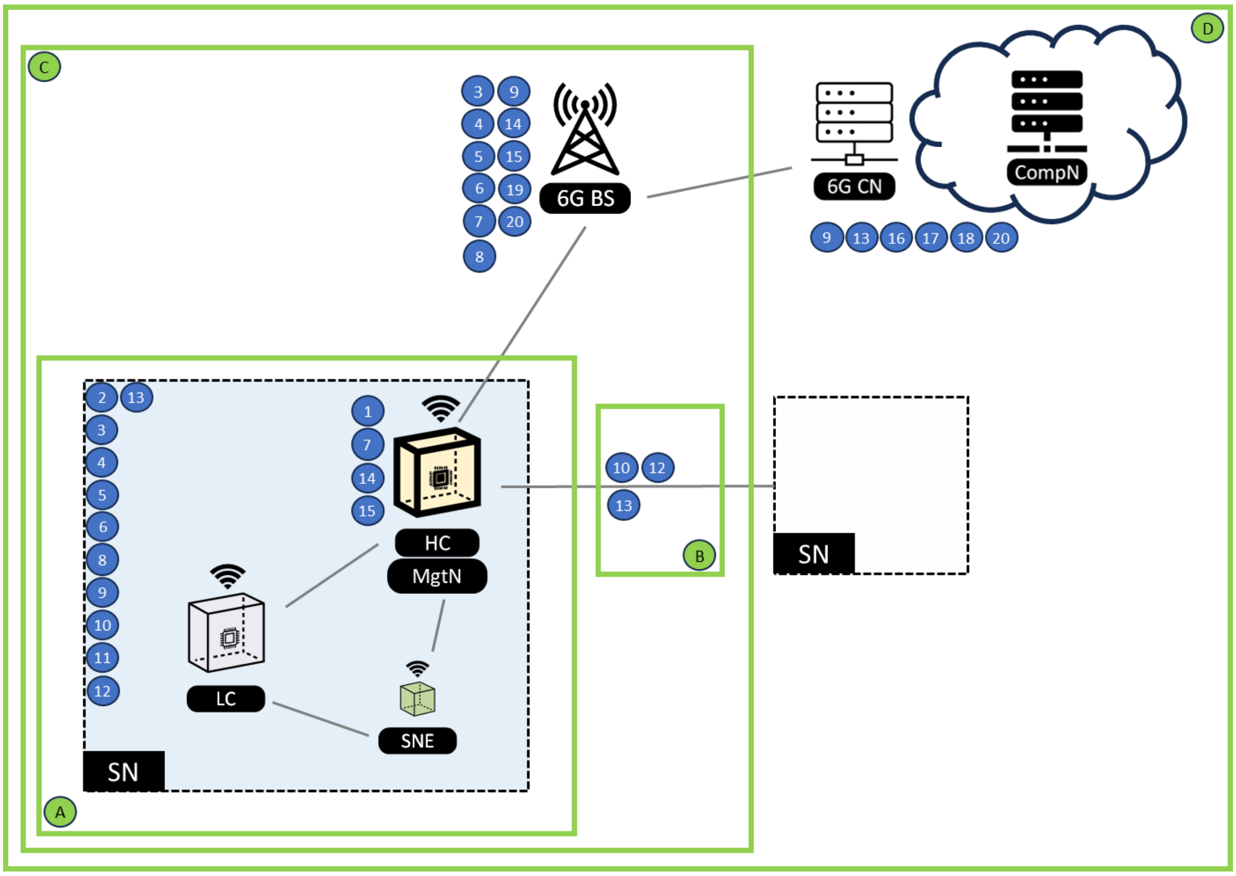


Figure 37: 6G-SHINE architectural blueprint - End-to-end 6G network architecture for subnetwork integration.

Table 7 and Table 8 summarize the 6G-SHINE architecture blueprint and provides a mapping between enhancements proposed in this document, system components, interfaces, impacted nodes and respective sections for details. They highlight the weight of the work done by the consortium, in each of the architectural components. Naturally, focus is put on the management of each subnetwork (see architectural component A in Figure 37), as expected, as this is a new concept. Additionally, good balance is observed between all other components (see architectural component mapping numbers C and D in Figure 37), that are required to integrate subnetworks in 6G, and whose impact is covered is all aspects of the end-to-end 6G system.

Table 7: 6G-SHINE architectural blueprint - mapping between enhancements, system components, interfaces, impacted nodes and document sections.

Enhancement mapping number	Proposed enhancement	Nodes impacted	Section
1	SN Management	HC	3.1.1
2	Subnetwork Routing Protocol	SNE, LC, HC	3.1.2
3	Device Group Functions	SNE, LC, HC, 6G-BS	3.1.3
4	Non-Transparent Subnetworks and Virtual Connections	SNE, LC, HC, 6G-BS	4.1.1
5	UE Dedicated UL Grant and Subnetwork Scheduling	SNE, LC, HC, 6G-BS	4.1.2
6	Subnetwork Authentication	SNE, LC, HC, 6G-BS	4.1.3
7	Subnetwork Tunnelling Protocol	HC, 6G-BS	4.2.1
8	Non-Standalone Low Capability Devices Within Subnetworks	SNE, LC, HC, 6G-BS	4.2.2
9	Control Plane Function Offloading	SNE, LC, HC, 6G-BS	4.3.1
10	Distributed Computing Roles and Local and Decentralized Distributed Computing	SNE, LC, HC	3.2.1
11	AI/ML enabled RRM architecture	SNE, HC	3.3.3
12	Control functions for local subnetwork continuum	SNE, LC, HC	3.2.2
13	Control functions for subnetwork-edge-cloud continuum	SNE, LC, HC, 6G-CN, CompN	4.3.2
14	Utilization of a RIC for managing SN resources and integrating with the parent network	HC, 6G-BS	3.3.1
15	AI/ML model transfer to subnetwork	HC, 6G-BS	4.1.5
16	Sensing Coordination functions Sensing Analytics functions	SNE, LC, HC, 6G-CN, SAF, SCF, CompN	3.3.2
17	Processing of Sensing data	SNE, LC, HC, 6G-CN, SAF, SCF, CompN	4.2.3
18	Compute offloading from the parent to the subnetwork	SNE, LC, HC, 6G-CN, SAF, SCF, CompN	4.3.3
19	Integration and management of non-3GPP subnetworks.	HC, 6G-BS	4.1.4
20	AI/ML model training for subnetworks	HC, 6G-BS, 6G-CN, CompN	4.1.5

Table 8: 6G-SHINE Architectural component focus.

Architectural component focus	Architectural component category	Interfaces impacted
A	Intra-subnetwork management	SNE-LC, SNE-HC, LC-HC, HC-6G-BS
B	Inter-subnetwork management	HC-HC, HC-6G-BS
C	Parent 6G network management (RAN level)	HC-6G-BS, HC-LC,
D	Parent 6G network management (CN level)	LC-HC, HC-6G-BS, 6G-BS-6G-CN

Additional definitions from the table are noteworthy:

Interfaces impacted – represents all possible interfaces where the described enhancements have a visible impact (at transmission level). Depending on the subnetwork topology, some interfaces may not be present (e.g., a subnetwork without an SNE will not have SNE-X interface impact), and therefore, the interface may not apply. Depending on the proposed enhancement, there could be specific target interfaces. Details need to be consulted from the respective section.

Nodes impacted – represents all possible nodes that could be executing the proposed enhancement, i.e., that could be part of the proposed protocols and/or solutions. This can include any node from the 6G-SHINE reference architecture and new additional nodes proposed as part of this document. Depending on the subnetwork topology, some nodes may not be present. Depending on the proposed enhancement, there could be specific nodes targeted. Details need to be consulted from the respective section.

6 CONCLUSIONS

This deliverable systematically addressed the architectural and functional aspects required to integrate in-X subnetworks into the emerging 6G ecosystem. The contributions are structured across key chapters, each focusing on a specific aspect of this integration and its practical implications.

Chapter 1 and Chapter 2 introduced the objectives and scope of the deliverable, establishing the foundation for the rest of the document by defining the necessary architectural components and functional requirements.

Chapter 3 focused on the internal architecture of in-X subnetworks. It described the development of resource exposure functions, such as Communication and Computing Resource Exposure Functions, to facilitate efficient resource management within subnetworks. Advanced subnetwork management strategies, routing protocols, and distributed computing roles were also proposed, offering methods for improving communication reliability, computational efficiency, and autonomy. The chapter also discussed the role of centralized and decentralized controllers, such as High-Capability devices, in managing these tasks.

Chapter 4 examined the implications of interfacing between subnetworks and the 6G parent network. This includes the design of new protocols and functions that enable subnetworks to operate efficiently while maintaining interoperability with the broader network. Specific topics included subnetwork tunnelling protocols, enhanced authentication mechanisms, and scheduling techniques for managing data flows between devices within subnetworks and their parent networks. The chapter also introduced the concept of non-3GPP Subnetwork Management Function, which provides a centralized mechanism for integration and resource allocation of non-3GPP subnetworks, such as Wi-Fi.

To conclude, this deliverable supports the core objectives of the 6G-SHINE project by detailing the architectural elements, resource management strategies, and integration methods necessary for the effective operation of in-X subnetworks. It provides a structured framework for enabling key features, such as communication and computing resource exposure, advanced subnetwork management, and efficient interface designs, which directly address the challenges of integrating subnetworks into the 6G ecosystem.

REFERENCES

[1] B. Priyanto et al., “D2.1. – Initial Definition of Scenarios, Use Cases and Service Requirements for in-X Subnetworks,” 6G-SHINE, August 2023.

[2] B. Priyanto et al., “D2.2 – Refined Definition of Scenarios, Use Cases and Service Requirements for in-X Subnetworks,” 6G-SHINE, February 2024.

[3] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. -S. Alouini and R. Zhang, "Wireless Communications Through Reconfigurable Intelligent Surfaces," in IEEE Access, vol. 7, pp. 116753-116773, 2019, doi: 10.1109/ACCESS.2019.2935192.

[4] D. Alanis et al., "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” 6G SHINE, June 2024

[5] 3GPP TR 38.838 V17.0.0, “Tech. Report Group Radio Access Network; Study on XR (Extended Reality) Evaluations for NR (Rel. 17)”, December 2021.

[6] 3GPP TS 22.847 v18.1.0, “Study on supporting tactile and multi-modality communication services” December 2021.

[7] O. Akgul et al., “Deliverable D3.3 Initial analysis of architectural enablers and framework”, Hexa-X II, April 2024.

[8] TTTech Auto Unveils Groundbreaking Scheduler to Revolutionize Complex Software Integration – “MotionWise Schedule”. Link available at: https://www.tttech-auto.com/newsroom/tttech-auto-unveils-groundbreaking-scheduler-revolutionize-complex-software-integration

[9] 3GPP TS 23.501 v19.2.1 , “System architecture for the 5G System (5GS)”, January 2025.

[10] O-RAN Use Cases and Deployment Scenario. Link available at: https://www.o-ran.org/

[11] D. Alanis et al., “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,” 6G-SHINE, March 2023”.

[12] 3GPP TS 38.300 v18.0.0, “NR; NR and NG-RAN Overall Description; Stage 2”, December 2023

[13] 3GPP TS 33.501 v18.5.0, “Security architecture and procedures for 5G system”, May 2024

[14] 3GPP TS 23.304 v18.5.0, "Proximity based Services (ProSe) in the 5G System (5GS)”, March 2024

[15] S. Giannoulis et al., “Deliverable D5.1: Description and methodology of evaluation of initial PoCs,” 6G-SHINE, February 2024.

[16] M. T. Lemes et al., "A Tutorial on Trusted and Untrusted Non-3GPP Accesses in 5G Systems—First Steps Toward a Unified Communications Infrastructure," in IEEE Access 2022.

[17] X. Lin, “Artificial intelligence in 3gpp 5g-advanced: A survey,” *ArXiv*, vol. abs/2305.05092, 2023.

[18] 3GPP TR 38.843 v18.0.0, “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface,” tech. rep., June 2022.

[19] 3GPP TSG RAN, “RP-213602: Artificial intelligence (AI)/machine learning (ML) for NG-RAN.” 3GPP TSG RAN Meeting #94e, Dec. 2021.

[20] Smart Networks in the Context of NGI, Technical Annex to Strategic Research and Innovation Agenda 2022-27, pp. 1–358, Technical Annex to SRIA 2024, v0.31 for consultation, 2024. Technical Annex to SRIA 2024 v0.31 for consultation.pdf.

[21] Hexa-X-II Deliverable D2.3, “Interim overall 6G system design,” June 2024.

[22] 3GPP TS 22.137 v19.1.0, “Service requirements for Integrated Sensing and Communication”, March 2024

[23] D. Abode, P. M. de Sant Ana, R. Adeogun and G. Berardinelli, "Communication-Aware Dynamic Speed Control for Interference Mitigation in 6G Mobile Subnetworks," *2024 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2024.