Beyond 5G Multi-Tenant Private Networks Integrating Cellular, Wi-Fi, and LiFi, Powered by Artificial Intelligence and Intent Based Policy

5G-CLARITY Deliverable D2.2

Primary System Architecture

Contractual Date of Delivery:	**October 31, 2020**
Actual Date of Delivery:	**October 31, 2020**
Editor(s):	**Jose Ordonez-Lucena (TID)**
Author(s):	**Daniel Camps-Mur, Hamzeh Khalili (I2CAT), Antonio Garcia (ACC), Alain Mourad, Ibrahim Hemadeh, Jani-Pekka Kainulainen (IDCC), Pablo Ameigeiras, Jonathan Prados-Garzon (UGR), Oscar Adamuz-Hinojosa (UGR), Tezcan Cogalan (UEDIN), Rui Bian (PLF), Erik Aumayr, Sven van der Meer (LMI), Carlos Colman, Shuangyi Yan, Hilary Frank, Amin Emami (UNIVBRIS), Jesús Gutiérrez, Vladica Sark (IHP), Mir Ghoraishi (GIGASYS)**
Work Package:	**WP2**
Target Dissemination Level:	**Public**

This document has been produced in the course of 5G-CLARITY Project. The research leading to these results received funding from the European Commission H2020 Programme under grant agreement No. H2020-871428. All information in this document is provided “as is", there is no guarantee that the information is fit for any particular purpose. The user thereof uses the information at its own risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors view.

Revision History

Revision	Date	Editor /Commentator	Description of Edits
0.1	14.04.2020	Jose Ordonez-Lucena (TID)	D2.2 ToC and contributors
0.2	21.04.2020	Jose Ordonez-Lucena (TID)	Subsection structure proposal for Chapters 5, 6, 7 and 8.
0.3	08.05.2020	Rui Bian (PLF) Antonio Garcia (ACC) Jani-Pekka Kainulainen (IDCC) Pablo Ameigeiras, Jonathan Prados-Garzon (UGR) Erik Aumayr (LMI) Vladica Sark (IHP) Jose Ordonez-Lucena (TID) Daniel Camps Mur (I2CAT)	PLF, ACC, IHP, LMI, UGR: 5G-CLARITY technical ecosystem (Section 2.1) I2CAT: 5G-CLARITY concepts (Section 2.2) TID: 5G-CLARITY Role Model (Section 2.3) PLF: integrate complete 1st draft Chapter 2
0.4	22.05.2020	Jose Ordonez-Lucena (TID) Jani-Pekka Kainulainen (IDCC)	TID: Architectural principles (Section 3.1) and 5G-CLARITY service delivery models (Section 3.3) IDCC: Baseline architecture design (Section 3.2) TID: Integrate complete 1st draft Chapter 3
0.5	01.06.2020	Daniel Camps Mur (I2CAT) Jose Ordonez-Lucena (TID) Pablo Ameigeiras, Jonathan Prados-Garzon (UGR) Tezcan Cogalan (UEDIN) Antonio Garcia (ACC)	I2CAT: Chapter 7 – design principles, NFV MANO and VIM, Multi-WAT non-RT-RIC, Slice Manager; Figure 7-1 added UGR: Chapter 7 – SDN transport controller TID: Chapter 7 – data semantics fabric, cloud native support functions (authentication and registration) UEDIN: Chapter 6 – core network: UPF, 5G SBA, RT-AT3S control; access network: N3IWF and TNGF, Figure 6-1 added ACC: Chapter 6 – access network: gNB-CU-UP
0.6	09.06.2020	Jose Ordonez-Lucena (TID)	Updates on D2.2 ToC: split original Chapter 2 into new Chapter 2 and 3; extend Chapter 4 scope Chapter 2 – 5G-CLARITY Technical innovations: only focus on technical innovations, moving 5G-CLARITY concepts and role model to Annex A and Chapter 3, respectively. E2E network slicing added as new technical innovation. Every technical innovation with content on SoTA + innovation. Chapter 3 – 5G-CLARITY Overview: 5G-CLARITY Business Ecosystem (Section 3.1), 5G-CLARITY services (Section 3.2) and 5G-CLARITY service delivery models (Section 3.3). Architectural principles moved to Chapter 4. Chapter 4 – 5G-CLARITY System Architecture: Architecture requirements (Section 4.1); baseline architecture design (Section 4.2)
0.6	21.06.2020	Jose Ordonez-Lucena (TID) Erik Aumayr (LMI)	TID: Chapter 1 - Introduction; Chapter 10 - Conclusions; Chapter 4 – design principles on infrastructure stratum, NF stratum and MO stratum. LMI: Chapter 4 – design principles on intelligence stratum
0.7	26.06.2020	Shuangyi Yan, Hilary Frank, Amid, Hamid (UNIVBRIS) Antonio Garcia (ACC) Erik Aumayr (LMI)	UEDIN: Chapter 5- compute nodes; private site GW, SDN-enabled Ethernet switches; Figure 5-1 added. ACC: Chapter 5 – 5G RAN nodes LMI: Chapter 8 – AI and intent engines
0.8	01.07.2020	Jose Ordonez-Lucena (TID) Daniel Camps Mur (I2CAT) Shuangyi Yan (UNIVBRIS) Antonio Garcia (ACC) Rui Bian (PLF) Pablo Ameigeiras, Jonathan Prados-Garzon (UGR) Jani-Pekka Kainulainen (IDCC) Vladica Sark (IHP) Tezcan Cogalan (UEDIN)	TID: Integrate 1st draft Chapters 3 and 4; Chapter 7 –data semantics fabric and cloud native support functions updated. I2CAT: Chapter 5 – Wi-Fi Access points; Chapter 6 – Wi-Fi/LiFi user and control plane functions; Chapter 7 – integrate complete 1st draft. UNIVBRIS: Chapter 5 – Edge computing cluster ACC: Chapter 5 – RAN cluster; Chapter 6 – NRT RIC xApps; PLF: Chapter 5 – LiFi Access Points; UGR: Chapter 5 – TSN nodes; Chapter 7 – SDN controller updated.

			IDCC: Chapter 6 – Multi-connectivity user plane; Chapter 7 – Data lake, cloud native support functions (distributed data storage), 5G-CLARITY Mediator Function. IHP: Chapter 6 – positioning server UEDIN: Chapter 6 – integrate complete 1st draft
0.9	05.07.2020	Rui Bian (PLF) Antonio Garcia (ACC) Oscar Adamuz-Hinojosa (UGR)	Updates on Chapter 2 sections, according to the new ToC in v0.6 PLF: LiFi technology section ACC: 5G/Wi-Fi/LiFi multi-connectivity UGR: End-to-end network slicing
0.91	08.07.2020	Shuangyi Yan (UNIVBRIS) Erik Aumayr (LMI)	UNIVBRIS: Integrate complete 1st draft Chapter 5 LMI: Chapter 8 – figure 8-1 added; integrate complete 1st draft Chapter 8
0.92	13.07.2020	Jose Ordonez-Lucena (TID)	TID: Chapters 1-8 integrated into the D2.2 master doc; acronym list added; resolved referencing issues for tables/figures; resolved bibliography related issues
0.93	19.07.2020	Jose Ordonez-Lucena (TID) Pablo Ameigeiras, Jonathan Prados-Garzon (UGR) Tezcan Cogalan (UEDIN)	TID: Chapter 9 – NPN-PLMN integration: management plane; Chapter 9 and Annex B integrated into the D2.2 master doc; 1st complete draft of D2.2 generated. UGR: Chapter 9 – deployment scenarios, NPN-PLMN integration: user and control plane; Annex B UEDIN: Chapter 6 – functional requirements of 5G-CLARITY network function stratum.
0.94	21.07.2020	Daniel Camps (I2CAT) Erik Aumayr (LMI) Jose Ordonez-Lucena (TID)	I2CAT: Chapter 7 – Table with requirements of 5G-CLARITY management and orchestration stratum LMI: Chapter 9 – Table with requirements of 5G-CLARITY intelligence stratum TID: Integrate I2CAT and LMI into D2.2 master doc
0.95	25.07.2020	Jose Ordonez-Lucena (TID)	Produce D2.2 1st full draft
0.96	31.08.2020	Jose Ordonez-Lucena (TID) Daniel Camps (I2CAT) Alain Mourad (IDCC)	Review of D2.2 1st full draft
0.97	25.09.2020	ALL Jose Ordonez-Lucena (TID)	ALL: Reviewers' comments addressed TID: Produce D2.2 2nd full draft
0.98	05.10.2020	Antonio Garcia (ACC) Jesús Gutiérrez (IHP) Mir Ghoraishi (GIGASYS)	Review of D2.2 2nd full draft
0.99	13.10.2020	ALL Jose Ordonez-Lucena (TID)	ALL: Reviewers' comments addressed TID: Produce D2.2 final draft
1.0	31.10.2020	Jesús Gutierrez (IHP) Mir Ghoraishi (GIGASYS)	Final version

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
5GNR	5G New Radio
5G-ACIA	5G Alliance for Connected Industries and Automation
5GC	5G Core
ADR	Angle Diversity Receiver
ADT	Angle Diversity Transmitter
AF	Application Function
AFC	Automated Frequency Control
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AMF	Access and Mobility management Function
APD	Avalanche photodiode
API	Application Programming Interface
ATS	Asynchronous Traffic Shaper
AT3S/AT3S	Access Traffic Splitting, Switching, Steering
AT3S-LL	AT3S Low Layer
B2B2X	Business-to-Business-to-X
B5G	Beyond 5G
BSI	Broadcast System Information
CaaS	Container as a Service
CAG	Closed Access Group
CBRS	Citizens Broadband Radio Service
CI/CD	Continuous Integration / Continuous Development
CNC	Central Network Controller
COTS	Commercial-Off-The-Shelf
CPE	Customer Premises Equipment
CUC	Central User Controller
CUPS	Control User Plane Separation
DCSP	Data Centre Service Provider
DL AoD	Downlink Angle-of-Departure
DNN	Data Network Name
DPDK	Data Plane Development Kit
DSA	Dynamic Shared Access
EDCA	Enhanced Distributed Channel Access
eMBB	enhanced Mobile Broadband
eMBMS	Evolved Multimedia Broadcast Multicast Services
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FDD	Frequency Division Duplexing
FPGA	Field Programmable Gate Array
FQDN	Fully Qualified Domain Name

FWA	Fixed Wireless Access
gNB	next-generation Node B
gNB-CU	gNB Central Unit
gNB-DU	gNB Distributed Unit
gPTP	generic Precision Time Protocol
GSA	Global mobile Supplier Alliance
GSMA	Global System for Mobile Alliance
HCF	Hybrid coordination function
HLS	High-Layer Split
IaaS	Infrastructure as a Service
IAB	Integrated Access backhauls
ICT	Information and Communication Technologies
IETF	Internet Engineering Task Force
IWSN	Industrial Wireless Sensor Network
LGA	Land Grid Array
LLS	Low-Layer Split
LSA	License Shared Access
LWA	LTE WLAN Aggregation
LWIP	Level Integration with IPSEC Tunnel
MEC	Multi-access Edge Computing
MF	Management Function
ML	Machine Learning
MN	Master Node
mMTC	Massive Machine-Type Communications
MOCN	Multi-Operator Core Network
MPTCP	MultiPath Transmission Control Protocol
MR-DC	Multi-Radio Dual Connectivity
N3IWF	Non-3GPP InterWorking Function
near-RT-RIC	Near-Real-Time RIC
NEF	Network Exposure Function
NF	Network Function
NFV	Network Functions Virtualization
NFVI	NFV Infrastructure
NFVO	NFV Orchestrator
NG-RAN	Next Generation Radio Access Network
NGMN	Next Generation Mobile Networks
non-RT-RIC	Non-Real-Time RIC
NPN	Non-Public Network
NRM	Network Resource Model
NSaaS	Network Slice as a Service
NSD	Network Service Descriptor
NSMF	Network Slice Management Function
NSSAAF	Network Slice Specific Authentication Authorization Function
NSSAI	Network Slice Selection Assistance Identifier

NSSF	Network Slice Selection Function
NWDAF	NetWork Data Analytics Function
OBSS	Overlapping basic service set
OFDM	Orthogonal Frequency Division Multiplexing
OOK	on-off-keying
OSS	Operation Support System(s)
OT	Operation Technology
OWC	Optical Wireless Communications
PAM	Pulse Amplitude Modulation
PCF	Policy Control Function
PDU	Packet Data Unit
PLMN	Public Land Mobile Network
PNF	Physical Network Function
PNI-NPN	Public Network Integrated NPN
PoC	Proof-of-Concept
PoT	Proof-of-Transit
PPM	Pulse Position Modulation
PRB	Physical Radio Block
QoS	Quality of Service
RGB	red, green and blue
RIC	RAN Intelligent Controller (RIC)
RRM	Radio Resource Management
RSS	Received Signal Strength
RU	Radio Unit
SAP	Service Access Point
SAS	Shared Access Spectrum
SBA	Service Based Architecture
SBMA	Service Based Management Architecture
SDN	Software Defined Networking
SDO	Standard Development Organization
SECF	Service Exposure Control Function
SLA	Service Level Agreement
SMF	Session Management Function
SN	Secondary Node
SNPN	Standalone NPN
SOA	Service Oriented Architecture
SON	Self Organized Networks
SPAD	Single-photon avalanche diodes
SRB	Signalling Radio Bearer
SSC	Session and Service Continuity
SotA	State-of-the-Art
SRP	Stream Reservation Protocol
TAS	Time-Aware Shaper
TDD	Time Division Duplexing

TDMA	Time Division Medium Access
TNAN	Trusted non-3GPP access networks
TNGF	Trusted Non-3GPP Gateway Function
ToF	Time-of-Flight
TSC	Time Sensitive Communication
TSN	Time Sensitive Networking
TTI	Transmission Time Interval
UDR	Unified Data Repository
UDSF	Unstructured Data Storage Function
UE	User Equipment
UICC	Universal Integrated Circuit Card
UL-AOA	Uplink Angle-of-Arrival
UNI	User Network Interface
UPF	User Plane Function
uRLLC	ultra-Reliable Low Latency Communications
VAF	Virtualized Application Function
VCSEL	Vertical-Cavity Surface-Entity Laser
VDU	Virtual Deployment Unit
VIM	Virtualized Infrastructure Manager
VISP	Virtualized Infrastructure Service Provider
VLC	Visible Light Communication
VNF	Virtualized Network Function
VNFD	VNF Descriptor
VPN	Virtual Private Network
WAT	Wireless Access Technology
WDM	Wavelength Division Multiplexing
YAML	YAML (Yet Another Markup Language) Ain’t Markup Language
YANG	Yet Another Next Generation

Executive Summary

The mission of 5G-CLARITY is to develop and demonstrate a Beyond 5G (B5G) system for private networks integrating multiple wireless access technologies including 5G, Wi-Fi and LiFi technologies, all operated through Artificial Intelligence (AI)-based autonomic networking. 5G-CLARITY brings forward the design of a system that addresses the wide variety of challenges identified today in private network environments, including spectrum flexibility, delivery of critical services, integration with public network infrastructures, and automated network management with built-in slicing. These features will be implemented, assessed and showcased through two innovative use cases: *i)* enhanced human-robot interaction demonstrated in M-Shed museum in Bristol; *ii)* wireless network slicing for production data exchange in Industry 4.0 services, and enhanced positioning of Automated Guided Vehicles (AGVs) for intralogistics process in BOSCH assembly plant in Spain. A detailed description of these use cases, including the specification of their functional requirements and target Key Performance Indicators (KPIs), is provided in the 5G-CLARITY D2.1 [1].

This document dives into the initial architecture design of the 5G-CLARITY system, which is architected into four different strata, with segregated scope and different technology pace each:

- **Infrastructure stratum** – it is formed of all the on-premise hardware and software resources building up the 5G-CLARITY substrate, including user equipment and a wide variety of compute, storage and networking fabric.
- **Network and application function stratum** – it conveys the 5G-CLARITY user, control and application plane functionality. The stratum includes all virtualized network and application functions that can be executed atop the 5G-CLARITY cloud infrastructure.
- **Management and Orchestration stratum** – it encompasses all the necessary functionality to deploy and operate the different 5G-CLARITY services (and associated resources) throughout their lifetime, from their commissioning to their de-commissioning. This includes provisioning functions (for lifecycle management), monitoring functions (for data collection and processing) and other supporting functions.
- **Intelligence stratum** – it hosts the ML models and related policies which provide AI-driven and intent-based operation capabilities to the overall 5G-CLARITY stratum. This stratum allows providing usage simplicity and zero-touch experience for 5G-CLARITY system consumers, specially Operation Technology (OT) actors (e.g. industry verticals), facilitating their access to the system behaviour for Service Level Agreement (SLA) assurance purposes.

Our architecture approach allows the 5G-CLARITY system to provide a rich set of capabilities in private networks. These capabilities can be flexibly adapted, combined and extended to support a wide variety of services for both public and non-public use, including infrastructural services and communication/digital services. The 5G-CLARITY service portfolio brings forward a business ecosystem where multiple actors are allowed to coexist and interact, with vertical (intra-domain) and horizontal (inter-domain) customer-provider relationships among them, empowering innovative service delivery models that go well beyond those defined in the 5G-PPP community so far.

The main contributions in this deliverable are the following:

- **5G-CLARITY technical innovations** (Section 2), which collectively define 5G-CLARITY system capabilities. For these innovations, we provide details on their individual scope in relation to state-of-the-art (SotA) solutions, and we motivate their choice for the 5G-CLARITY system.
- **5G-CLARITY business ecosystem, its actor-role model and the offered services** (Section 3), inspired by the definitions provided in 3GPP and 5G-PPP. In this ecosystem, business relationships and associated service delivery models are identified and analysed, with emphasis on the presence of

public and private administrative domains while using resources from different access technologies.

- **The requirements** (Section 4) of the 5G-CLARITY system. In order to get a future-proof system for a wide range of B5G use cases, these system requirements are derived based on, and beyond to, the use case requirements and KPIs specified 5G-CLARITY D2.1 [1].
- **The baseline architecture design** (Section 4) of the 5G-CLARITY system, with the four main strata of our system: infrastructure stratum, network and application function stratum, management and orchestration stratum, and intelligence stratum. The architecture principles guiding the design of these strata are also specified.
- **Initial design of the four 5G-CLARITY system strata** (Sections 5, 6, 7 and 8), diving into their internal architecture, including information on individual components and communication interfaces among them.
- **Insight into the integration between private and public networks** (Section 9), using 5G-CLARITY system. This integration allows for secure exchange of user and signalling traffic between private and public domains, as well as seamless interoperation of their management and orchestration systems, i.e. 5G-CLARITY system and 3GPP management system.

1 Introduction

With the standardization of 3GPP Rel-16 features, much more focused on providing necessary capabilities for the support of mission-critical services and ultra-reliable low-latency communications (uRLLC), 5G systems are rapidly gaining recognition as an all-inclusive critical communications platform for industry digitization. By providing authority over wireless coverage and capacity, private 5G network market ensures guaranteed and secured connectivity, while supporting a wide range of applications, ranging from push-to-talk group communications and real-time video delivery to wireless control and automation in industrial environments. This has motivated to a wide variety of industry verticals (e.g. militaries, utilities, railway and port operators, public safety agencies, manufacturers) to start making sizeable investments in private 5G networks. Apart from these actors, a number of independent neutral-host and wholesale operators are also stepping up with pioneering business models to provide 5G connectivity services to both mobile network operators and enterprises, particularly in indoor settings and locations like public venues (e.g. stadia, shopping malls, transportation hubs), where it is technically or economically not feasible for traditional operators to deliver substantial wireless coverage and capacity.

The success of private 5G networks will be judged by the level of utilization of its advanced capabilities by OT organizations to deliver new and innovative services that are stable enough to introduce into the market. The widespread adoption of private 5G networks will only become a reality if their operational costs are small, and a seamless interworking between 5G access and other industry technologies (e.g. wired Ethernet, Wi-Fi) is achieved. To that end, a deep involvement of the ICT organizations (e.g. mobile network operators, vendors, hardware supplier and software developers) is required. The mission of these organizations is to provide necessary network capabilities to make 5G private networks ready for use, allowing the OT organizations focus on their core business. Examples of these capabilities include spectrum management, multi-RAN and multi-tenancy support, SDN/NFV-powered infrastructure slicing, and data-driven network management. 5G-CLARITY project will focus on the design, integration and operation of some of these ICT industry provided capabilities in private networks, with the definition of a novel system solution.

The present deliverable provides an initial architecture design of the 5G-CLARITY system, positioning it as a future-proof solution for the provisioning and operation of B5G services in private networks. The 5G-CLARITY system, built upon a number of technical innovations, provides a rich business environment, with multiple stakeholders taking part in the entire value chain, interacting with each other following horizontal and vertical customer-provider relationships. These relationships unveil innovative service delivery models, with services spanning public and private administrative domains while using resources from different access technologies.

1.1 Scope of this document

This document constitutes the second deliverable of Work Package (WP2). This WP, entitled “Scenario Description, Architecture and Requirements”, aims to set the directions to conduct the technical work of the 5G-CLARITY project, setting up the overall requirements for the development and assessment of the complete system architecture towards the end product of the project, which are demonstration systems for two use cases: UC-I and UC-II. On the one hand, UC-I features enhanced human-robot interaction demonstrated in M-Shed museum in Bristol. On the other hand, UC-II features wireless network slicing for production data exchange in Industry 4.0 services, and enhanced positioning of AGVs for intralogistics process in BOSCH assembly plant in Spain. A detailed description of these use cases, including the specification of their functional requirements and target KPIs is provided in the 5G-CLARITY, was provided in 5G-CLARITY D2.1 [1].

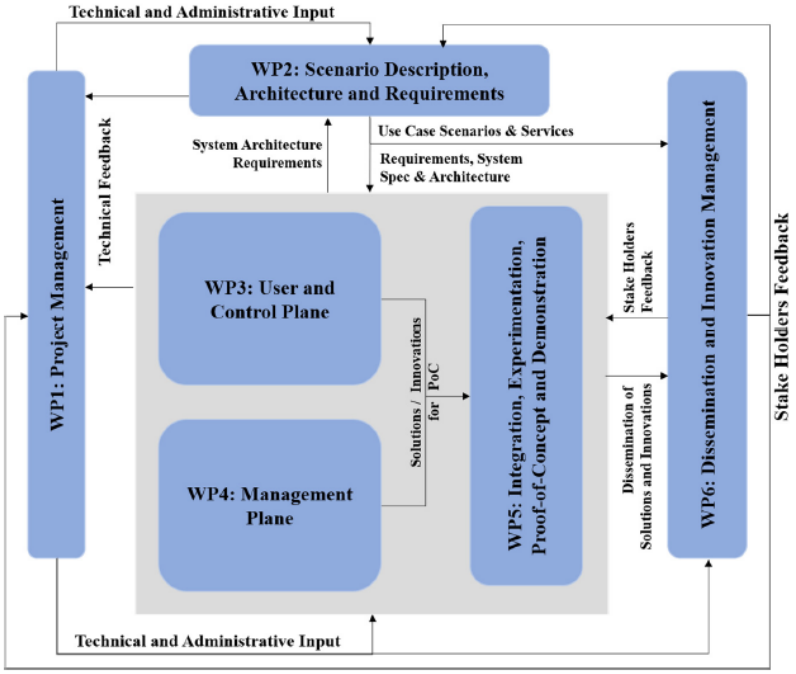

Figure 1.1: 5G-CLARITY Work Packages structure

The objective of deliverable D2.2 is to present the initial architecture design of the 5G-CLARITY system, diving into the functionality of individual components and the interfaces across them. This primary system architecture provides details at to the design solution framework, laying the groundwork for further work in other WPs, where necessary hardware and software solutions will be designed, implemented and integrated (WP3 and WP4) for UC-I and UC-II demonstration (WP5). These relationships are captured in Figure 1.1.

According to the above rationale, deliverable D2.2 constitutes the starting point of much of the work that is to be undertaken in the remainder of WP2 and in the other WPs. In this regard, D2.2 bears relationships with other deliverables from the 5G-CLARITY project, including both *input* (i.e. material from one past deliverable, used as input to D2.2) and *output* (i.e. material from D2.2 used as input to a future deliverable) relationships, as follows:

- **WP2**: *D2.1 "Use-case specifications and requirements" (input), D2.3 "Primary system architecture evaluation" (output), and D2.4 "Final system architecture and its evaluation" (output)*. D2.2 builds on D2.1 11 as a starting point for 5G-CLARITY service ecosystem and system architecture definition. 5G-CLARITY system architecture requirements specified in D2.2 will be evaluated in D2.3. This evaluation will qualify the behaviour and readiness of the different system components and architectural features under different contexts and scenarios. These contexts and scenarios will be reproduced following a well-defined methodology that will include theoretical analysis, simulations and (optionally) testing with Proof-of-Concept (PoC) prototypes. Finally, D2.2 outcomes will consolidate the basis for further architectural refinements (with likely extensions and enhancements) in D2.4.
- **WP3**: *D3.1 "State-of-the-art review and initial design of the integrated 5G/Wi-Fi/LiFi network frameworks on coexistence, multi-connectivity, resource management and positioning" (input)*, and *D3.2 "Design refinement and initial evaluation of the coexistence, multi-connectivity, resource management and positioning frameworks*" (output). On the one hand, state-of-the-art technologies presented in D3.1 [2] have been taken into consideration for the design of 5G-CLARITY infrastructure and network function strata. On the other hand, the design principles and architectural features of these strata, which will be presented in D2.2, will guide the refinements and performance evaluation in D3.2.

- **WP4**: *D4.1 "Initial design of the 5G-CLARITY SDN/NFV platform, interface design with 5G service platform, and initial definition of evaluation of ML algorithms (output)".* The D2.2 outcomes on the upper layer strata (i.e. 5G-CLARITY management and orchestration stratum, and 5G-CLARITY intelligence stratum), including offered management services and associated capabilities, will be used for solution set specifications and Machine Learning (ML) design in D4.1.

1.2 Objectives of this document

The specific objectives of the deliverable D2.2 are as follows:

- **OBJ-1: Specification of 5G-CLARITY system capabilities.** These capabilities capitalize on a set of technology enablers which are arranged into five technical innovations: *(i)* the use of LiFi technology, *(ii)* multi-access connectivity support, *(iii)* enhanced localization and device synchronization, *(iv)* AI-driven and intent-based network management, and *(v)* public-private networks integration.
- **OBJ-2: Description of 5G-CLARITY business ecosystem and service portfolio.** Spanning well beyond the project's duration, this ecosystem aims at laying the foundation for a rich technology market in private networks, with a flexible governance model that lowers barriers to entry for new actors, thus boosting service innovation. In this objective, we identify the different 5G-CLARITY stakeholders, the customer-provider relationships among them and the associated service delivery models.
- **OBJ-3: Elicitation of 5G-CLARITY system requirements**, including functional and non-functional requirements.
- **OBJ-4: Baseline architecture design of the 5G-CLARITY system**, with the definition of four separate 5G-CLARITY strata (i.e. infrastructure stratum, network and application function stratum, management and orchestration stratum, and intelligence stratum) and the specification of the architecture principles guiding their individual design.
- **OBJ-5: Summary of the internal architecture design of individual 5G-CLARITY strata,** diving into the functionality of different components and the interfaces among them.
- **OBJ-6: Identification, characterisation and analysis of 5G-CLARITY mechanisms for communication and interoperation between private networks** (i.e. on-premise network) and **public networks** (i.e. PLMN).

1.3 Document structure

The rest of this document is structured as follows:

- Section 2 covers **OBJ-1**, providing details on the individual technical innovations within the scope of the 5G-CLARITY system, capturing key technology advancements compared to the SotA. This SotA includes research activities and outcomes from standards development organizations (SDOs), industry alliances, research projects and open source initiatives.
- Section 3 covers **OBJ-2**, presenting an overall description of the stakeholders and services taking part in the 5G-CLARITY business ecosystem, together with an analysis of related service delivery models.
- Section 4 covers **OBJ-3**, analysing the functional and non-functional requirements that services and stakeholders impose onto the design of the system architecture. Based on this analysis, Section 4 also depicts the baseline architecture, mapping these requirements into architecture principles for the design of individual 5G-CLARITY strata, thus covering **OBJ-4**.
- Section 4, 5, 6 and 7 cover summarise the key functionalities and exposed capabilities of the four 5G-CLARITY strata, thus covering **OBJ-5**. This provides a complete view of the whole 5G-CLARITY system architecture, showing that the design of the different components is aligned with the architecture

principles outlined in Section 4.

- Section 9 covers **OBJ-6**, giving insight into the integration of private network and public network in the 5G-CLARITY ecosystem throughout different deployment scenarios, together with an analysis on secure means for their interoperation.
- Finally, Section 10 provides the concluding remarks.

2 5G-CLARITY Technical Innovations

This section aims to specify the technical innovations that define 5G-CLARITY project ambition, providing details on their individual scope in relation to SotA solutions. This overview includes research activities and outcomes from SDOs, industry alliances, research projects and open source initiatives.

2.1 LiFi technology

LiFi is a technology that uses part of the visible light or infrared portion of the electromagnetic spectrum to transmit information at very high speeds. As shown in Figure 2.1(a), LiFi technology is already having an impact on the connectivity and lighting landscapes, because it can provide the right ecosystem that tackle the current connectivity challenges and enable the billions of smart devices that are shaping our lives [3].

2.1.1 State-of-the-art

Modulation techniques

In principle, LiFi also relies on electromagnetic radiation for information transmission. Therefore, typically used modulation techniques in RF communication can be applied to LiFi with necessary modifications [4]. In single carrier modulation, the widely used modulation schemes are on-off-keying (OOK), pulse position modulation (PPM) and pulse amplitude modulation (PAM). Orthogonal frequency division multiplexing (OFDM) is one and perhaps the most common realisation of multi-carrier modulation in LiFi networks that could enhance the system performance significantly if combined with adaptive bit and power loading techniques [4]. Color shift keying (CSK) is a LiFi specific modulation technique where signals are encoded into color intensities emitted by red, green and blue (RGB) LEDs [4].

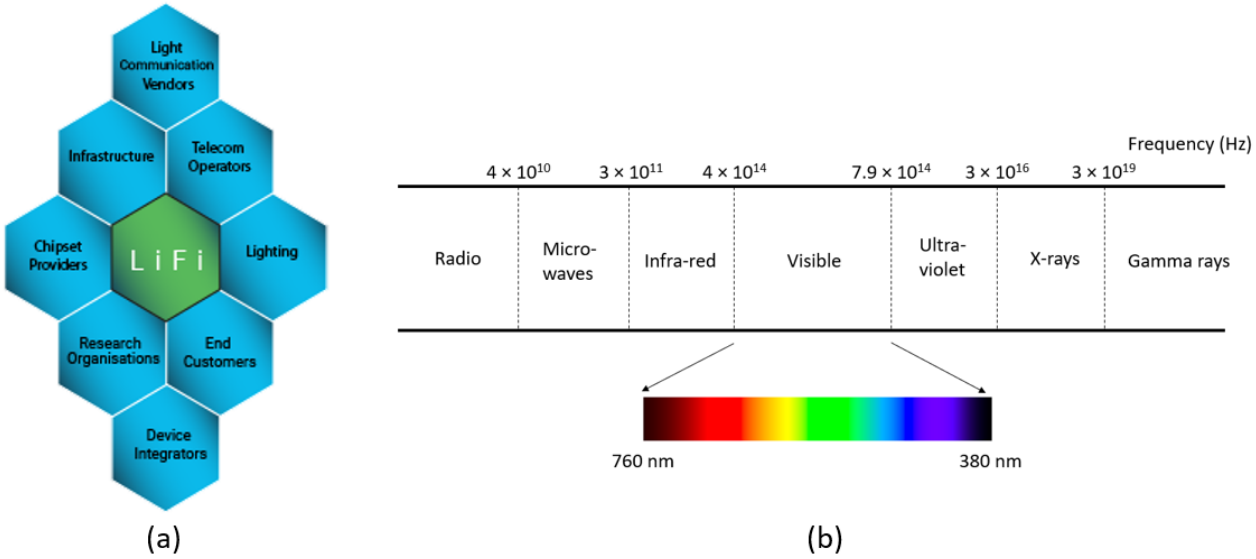

Figure 2.1: (a) Ecosystem of LiFi technology [3]; (b) The electromagnetic spectrum.

Devices and components

While the spectrum resource for LiFi is plenty, as shown in Figure 2.1(b), there are fundamental limitations to fully harness this huge amount of wireless transmission resource – especially when using LEDs as transmitters. This is because the bandwidth of typical LED devices is limited to 10s to 100s of MHz [5], as most of the LEDs are designed for illumination purposes rather than for wireless communications. However, with novel LiFi modulation techniques, over Gbps optical wireless communication (OWC) links can still be implemented using commercial LEDs. It was shown in [6] that a wavelength division multiplexing (WDM) implementation using four off-the-shelf LEDs offers an aggregate data rate of 15.7 Gbps at over 1.6 m link

distance despite the partial overlap of their emission spectra. Various laser devices such as vertical-cavity surface-emitting lasers (VCSELs) operating in the infrared spectrum have also been used to build high-speed OWC links. Such devices have higher bandwidth than LEDs. However, eye-safety standards need to be met for the optical systems while achieving quality of service (QoS) requirements.

On the receiver end, p-type, intrinsic region and n-type (PIN) photodiode and Avalanche photodiode (APD) are the most promising types of photodetector in LiFi. PIN PDs are capable of achieving a very high bandwidth while the APD can achieve much higher responsivity due to the avalanche multiplication. Single-photon avalanche diodes (SPADs) are also being investigated for LiFi studies as they are able to detect low intensity signals (down to the single photon) and to signal the time of the photon arrival with high temporal resolution (few tens of picoseconds). Solar cells have been proofed as a candidate for LiFi application for simultaneous data communication and energy harvesting. Work in [7] demonstrated an outdoor backhaul use case.

LiFi networks

LiFi is enabled by an ecosystem of multiuser techniques, resource allocation algorithms and security strategies. A LiFi network illustration is shown in Figure 2.2. A complete LiFi network includes downlink, uplink, and backhaul connections. In addition, the system should provide a handover function, mobility support, and multiple access capability [8]. LiFi networks were designed from the start to work seamlessly with radio frequency wireless networks, in the attempt to jointly enable efficient, opportunistic load balancing, and augmented capacity in heterogeneous networks [9]. Efficient load balancing in such a hybrid network is one of the main issues [10]. This challenge has been formulated as a mixed integer nonlinear programming problem in [11], where a joint optimisation algorithm and a separate optimisation algorithm have been proposed. In recent studies, the load balancing in LiFi/Wi-Fi hybrid networks in time-varying conditions with UE movement and rotation have been investigated [12]. Results of a real-world use cases where a LiFi network was deployed along with Wi-Fi network in a classroom have been presented in [8].

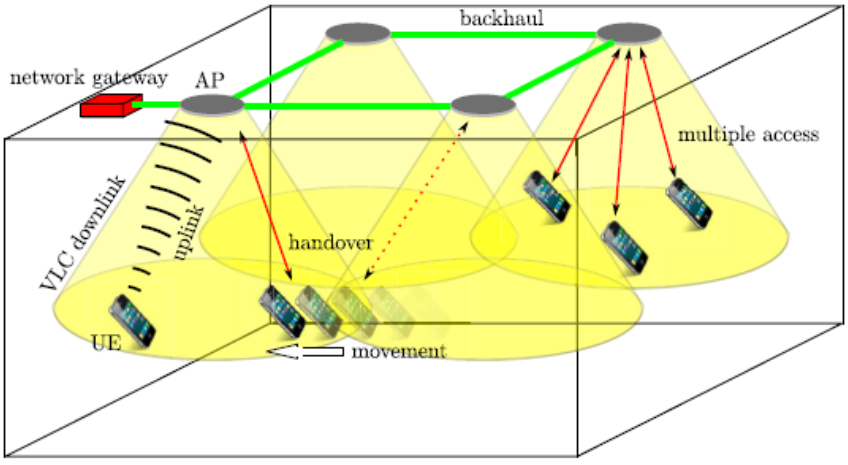

Figure 2.2: LiFi network illustration [6].

IEEE 802.11bb

The IEEE 802.11 Light Communication Amendment Task Group 'bb' is focused on introducing necessary changes to the base IEEE 802.11 standards to enable communication in the light medium. The general scope for the task group is [13]:

- Uplink and downlink operations in 380 nm to 5,000 nm band,
- All PHY modes of operation achieve minimum single-link throughput of 10 Mbps and at least one mode of operation that achieves single-link throughput of at least 5 Gbps, as measured at the MAC data service access point (SAP),

- Interoperability among solid state light sources with different modulation bandwidths.

This amendment specifies changes to the IEEE 802.11 MAC that are limited to the following:

- Hybrid coordination function (HCF) channel access,
- Overlapping basic service set (OBSS) detection and coexistence,
- Existing power management modes of operation (excluding new modes).

This task group will also address the security of the transition between the new Light Communication PHY and the existing 802.11 PHYs as well as the security implications in supporting Fast Session Transfer.

2.1.2 Innovation in 5G-CLARITY

A common misconception of LiFi is that it aims to replace RF wireless technologies. However, this is not true. In fact, LiFi is complementary to such RF wireless technologies while bringing significant connectivity advantages. LiFi is designed to work with other wireless technologies and complement them on those applications. 5G-CLARITY aims to bring a connectivity ecosystem including 5G, Wi-Fi and LiFi working co-ordinately to enable extremely high aggregate bandwidth and thus offer a high QoS for users.

In 5G-CLARITY we put forward a novel, beyond 5G, infrastructure for private networks. Regarding LiFi technology, the main innovations would be:

- Development of a seamless integration of LiFi with 5G and Wi-Fi technology, while Integrating LiFi with 5G slice management systems, in order to dynamically deploy slices that make use of LiFi resources.
- Incorporation of ML techniques to enable automatic management of the heterogeneous network, to optimize capacity and association of users across technologies.

2.2 5G/Wi-Fi/LiFi multi-connectivity framework

2.2.1 State-of-the-art

Multi-access based multi-connectivity and resource management

The concept of multi-connectivity was first introduced with the 3GPP intra-E-UTRA Dual Connectivity capabilities in Release 12 as a solution to improve the per-user throughput by utilising radio resources from separate eNBs. In Release 13, the concept of RAN-level spectrum interworking was introduced by aggregating the capabilities of LTE and WLAN in the downlink, and then for the uplink in Release 14. Subsequently, the concept of DC was generalised to Multi-Radio Dual Connectivity (MR-DC) in Release 15. This MR-DC concept was extended with the support of Integrated Access backhaul (IAB) in 3GPP Release 16 [14]. Figure 2.3 illustrates the standardization path from Release 12 to Release 15.

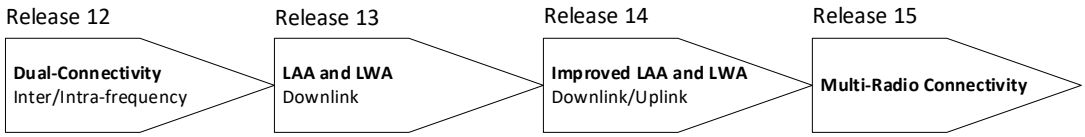

Figure 2.3: Standardization of DC in various 3GPP specification releases

The MR-DC introduced in Release 15 is a generalization of the intra-E-UTRA dual connectivity, where a multiple Rx/Tx capable UE can be configured to utilise resources provided by two different nodes connected via non-ideal backhaul, one providing 5GNR access and the other one providing either E-UTRA or 5GNR access. One node acts as the Master Node (MN) and the other as the Secondary Node (SN). The MN and SN are connected via a network interface and at least the MN is connected to the core network [15]. Figure 2.4 shows the C-plane architecture for the four different MR-DC modes.

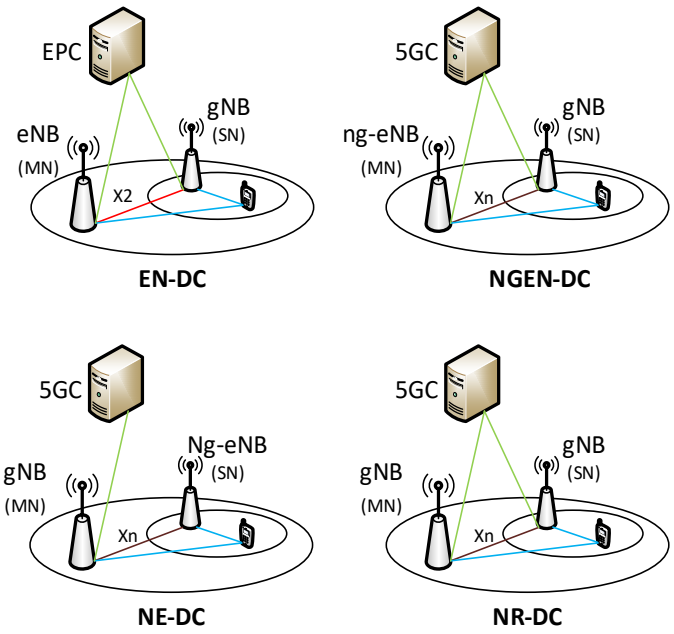

Figure 2.4: MR-DC four modes.

LTE WLAN aggregation (LWA) at the RAN level intended to provide an integration solution similar to the DC architecture between a master and a secondary eNBs. A key aspect in the design of integration architectures between WLAN and 3GPP, which was not present in the 3GPP multi-connectivity solutions, was the need for upgrades in an existent base of deployed WLAN APs. 3GPP came up with two separate solutions: *i)* 3GPP LWA, targeting greenfield deployments where Wi-Fi APs could embed new functions specified by 3GPP, and *ii)* 3GPP WLAN Radio Level Integration with IPSEC Tunnel (LWIP) targeting scenarios with legacy APs that could not be modified.

In 3GPP 5G specifications the approach to multi-access integration, whether wireless or wireline, is not done at the RAN level, but at the 5G Core (5GC) level. The 5GC supports both untrusted non-3GPP access networks and trusted non-3GPP access networks (TNANs). An untrusted non-3GPP access network is connected to 5GC network via Non-3GPP Interworking Function (N3IWF), whereas a trusted non-3GPP access network may be connected to the 5GC network via TNGF. Figure 2.5 and Figure 2.6 show non-roaming architecture for 5GC network with untrusted non-3GPP access and trusted non-3GPP access. Both the N3IWF and the TNGF interface with the 5GC network control plane and user plane functions via the N2 and N3 interfaces, respectively.

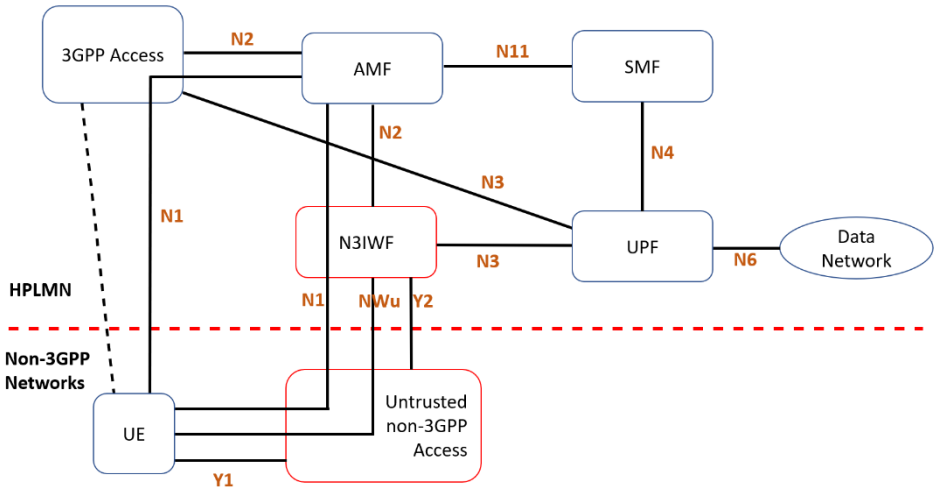

Figure 2.5: Non-roaming architecture for 5GC network with untrusted non-3GPP access

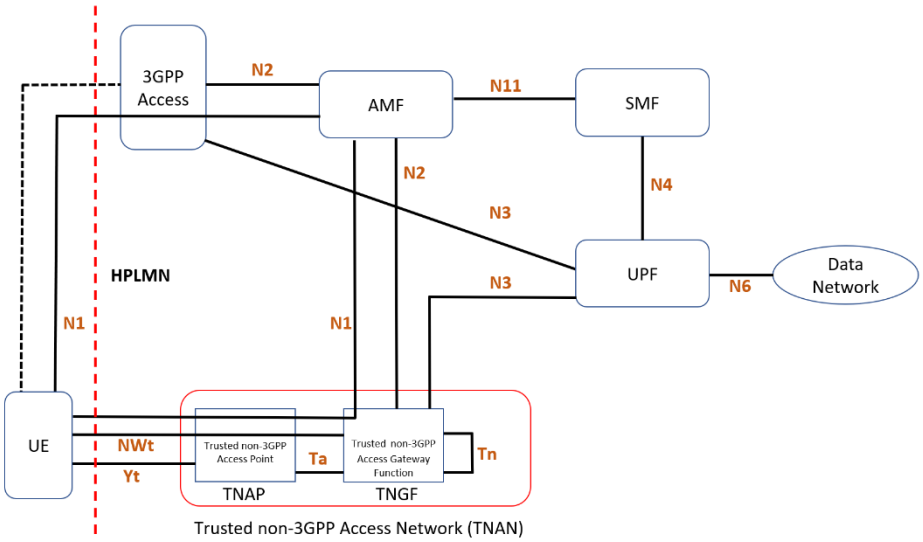

Figure 2.6: Non-roaming architecture for 5G Core Network with trusted non-3GPP access

A key cornerstone of the non-3GPP access integration into the 5GC is the Access Traffic Steering, Switching and Splitting function (ATSSS or AT3S) whose purpose is to support steering, switching and splitting functions between the 3GPP and non-3GPP access flows.

Figure 2.7 illustrates the AT3S architecture in a 5G network where UEs supporting AT3S functionality support one or more of the steering functionalities specified, e.g. multipath transmission control protocol (MPTCP) functionality and/or AT3S-Low Layer (AT3S-LL). Each steering functionality in the UE enables traffic steering, switching and splitting across 3GPP access and non-3GPP access, in accordance with the AT3S rules provided by the network. On the network side, the User Plane Function (UPF) may support MPTCP proxy functionality, which communicates with the MPTCP functionality in the UE by using the MPTCP protocol. The UPF may support AT3S-LL functionality, which is similar to the AT3S-LL functionality defined for the UE.

For more details on State-of-the-Art on multi-access based multi-connectivity and resource management, see D3.1 [2].

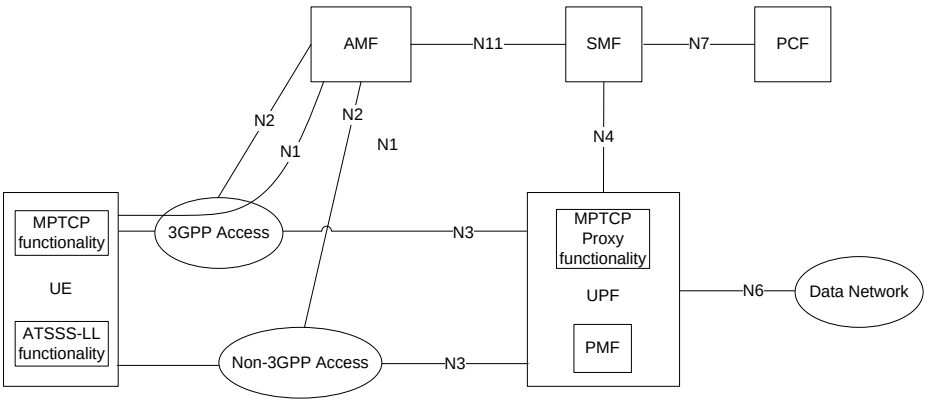

Figure 2.7: Non-roaming and Roaming with Local Breakout architecture for AT3S support

Shared Spectrum Access

As discussed in [16], regulators in a number of countries have authorized automated and even dynamic frequency coordination databases to manage real-time assignments in shared bands and to protect incumbent operations (including military and public safety systems) from harmful interference. While spectrum database coordination is nothing new, it has in recent years evolved from manual, to automated, to dynamic – adding automation and propagation modelling to static licensing data.

The European License Shared Access (LSA) was a database-assisted model that facilitated two-tier sharing between primary (incumbent) and secondary licensees. In this regulatory model analysed in [17] the regulator plays a direct role in managing the database of information by which primary and secondary licensees share the band although in a semi-static fashion.

The most representative country example of dynamic spectrum sharing is the US. In [18] and [19], the Federal Communications Commission (FCC) established the regulatory grounds for the Citizens Broadband Radio Service (CBRS), involving the shared commercial use of the 3.5 GHz. In CBRS, a novel three-tier sharing paradigm coordinates spectrum access among the incumbent military radars, satellite ground stations and temporarily protected Fixed Wireless Access (FWA) legacy stations and new commercial users as shown in Figure 2.8.

As described in [16], the CBRS framework enables the use of sensing network inputs to enable real time awareness of naval radars and allows dynamic interference protection managed by the Shared Access Spectrum (SAS) as shown in Figure 2.9.

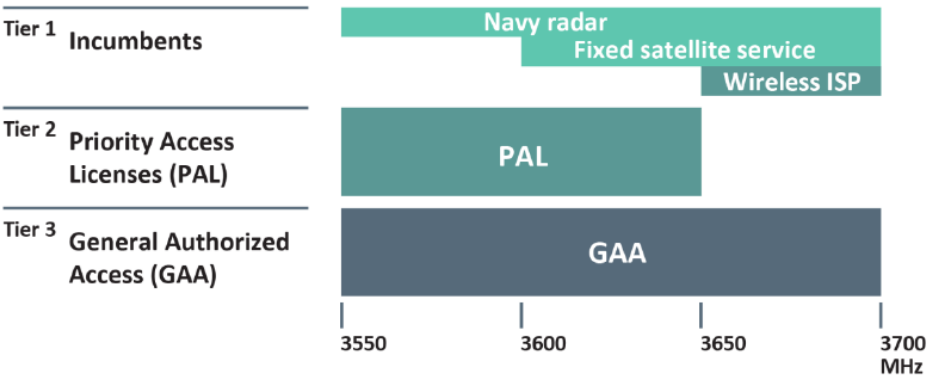

Figure 2.8: Three-tier coordination in CBRS

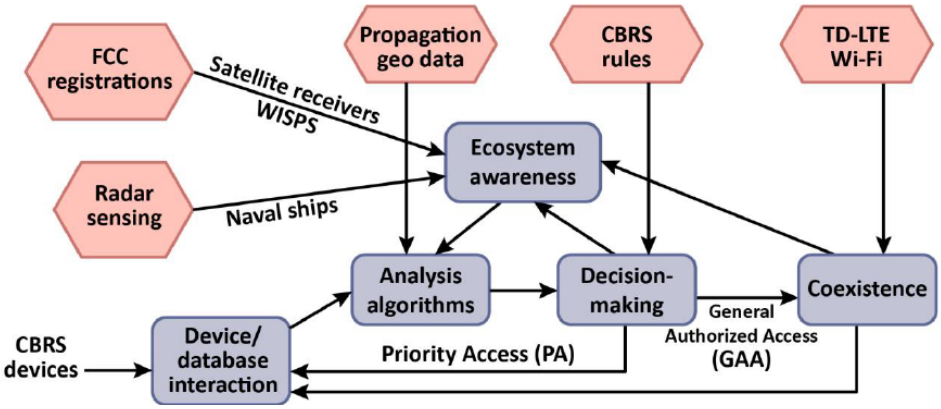

Figure 2.9: Admission control system architecture CBRS

On 23rd April 2020 FCC announced the adoption of rules that make 1200 MHz of spectrum in the 6 GHz band (5.925-7.125 GHz) available for unlicensed use in the US [20]. The FCC is authorizing two types of unlicensed operations in the 6 GHz band. One of them, authorizing unlicensed standard-power access points in the U-NII-5 and U-NII-7 bands is being done via the use of an Automated Frequency Control (AFC) system. This will permit operations at the same power levels already permitted in the 5 GHz U-NII-1 and U-NII-3 bands (5.150-5.250 GHz and 5.725-5.850 GHz bands, respectively) while, at the same time, protecting incumbents' users (6 GHz comprises allocations for Fixed Services, Mobile Services and Fixed Satellite Services). The AFC System Framework and Database proposed by FCC is similar to the CBRS system.

In [21], Ofcom enabled two new licence schemes to make it easier for a wider range of users in the UK to access radio spectrum on a shared basis and to improve wireless connectivity in enterprise sites as well as underserved areas. One of the two schemes is based on shared access licence, providing access to four spectrum bands which support mobile technology. The spectrum assignment is done currently statically, but Ofcom plans to make it fully dynamic in a similar fashion to CBRS.

For more details on SotA on spectrum sharing access refer to [2].

2.2.2 Innovation in 5G-CLARITY

The 5G-CLARITY multi-connectivity framework enables the integration of different access network technologies, namely 5GNR, Wi-Fi and LiFi.

5G brings a new radio interface (5GNR) based on a OFDM numerology with subcarrier spacing, symbol length and Transmission Time Interval (TTI) scalable flexibly to support operation in different bands (sub-6 GHz including pioneer 3.5 GHz capacity band and mmWave) in order to support different service requirements such as enhanced mobile broadband (eMBB), uRLLC and massive machine type communications (mMTC) articulated via end-to-end slicing across core, access and transport network resources. 5G also brings the support of unlicensed spectrum (namely 802.11 based technologies), whether standalone on its own right or integrated through RAN licensed assisted or core-based methods.

The 5G-CLARITY multi-connectivity framework will enable innovations and advancements in different fronts which span way beyond the simple combination of these access technologies, including smart radio resource management (RRM), co-existence mechanisms, new spectrum sharing paradigms and open RAN interfaces, all combined for an automated and interoperable near real-time (and non-real time) network and service management.

Two approaches are addressed in 5G-CLARITY multi-connectivity framework. The first approach with an associated implementation in the 5G-CLARITY pilots, is based on an enhanced AT3S at the core network level, which can be applicable to co-located and non-co-located access solutions. The second approach is based on

typical RAN side integration, and will be studied and analyzed without an actual implementation in the 5G-CLARITY pilots.

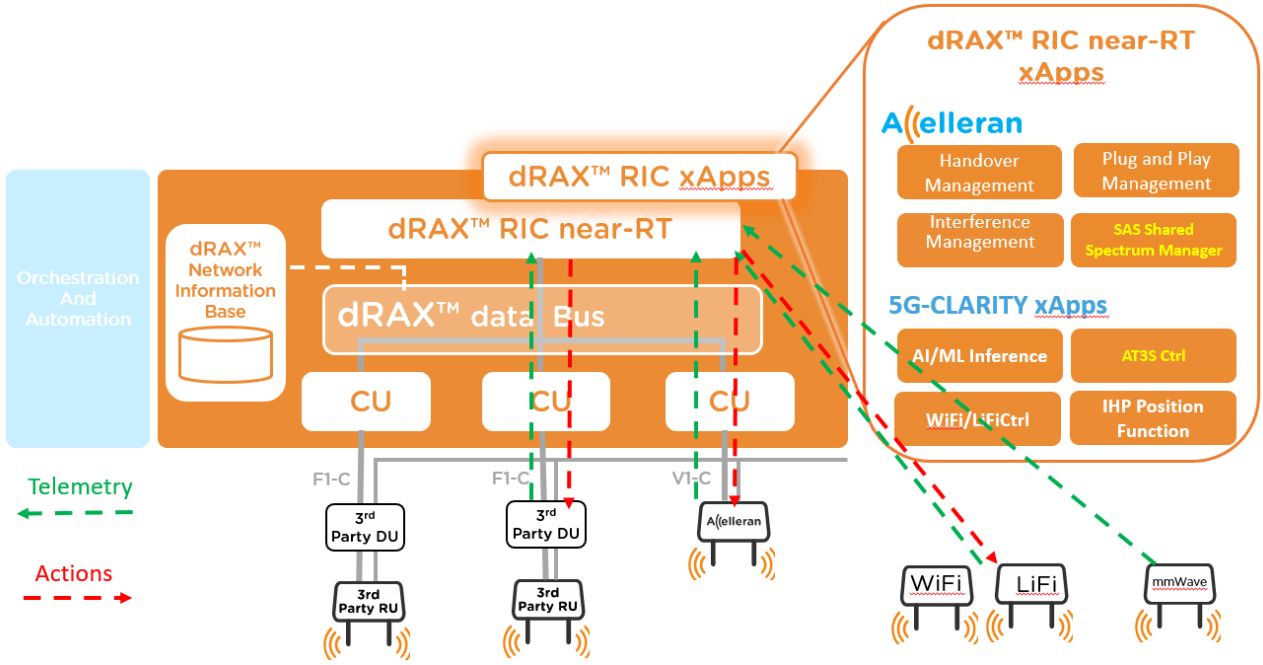

Figure 2.10: Accelleran dRAX multi-WAT reference architecture

The 5G-CLARITY solution implemented in the pilots will effectively be a hybrid of both the core side and RAN side integration approaches in the sense that the AT3S in the core side approach will be enhanced to incorporate near-RT (real-time) RAN control of the AT3S policies behind the UPF by using multi-WAT telemetry at RAN level in order to make more real time decisions driving AT3S. This is done using a near-RT RAN Intelligent control (RIC) and xApps O-RAN framework as illustrated in Figure 2.10 and further described in [2].

In 5G-CLARITY, the non-3GPP access network is considered as an integrated Wi-Fi-LiFi SDN L2 network. As the non-3GPP access network will be integrated with 3GPP access network via AT3S, how to route the traffic flows to 3GPP and non-3GPP access networks can be considered as a resource management problem. Moreover, as a 5G-CLARITY customer's premises equipment (CPE)/UE will be capable of using 5G, Wi-Fi and LiFi technologies, multi-user access to physical resources of those technologies can also be considered as a resource management problem. Therefore, resource management in 5G-CLARITY will be considered as a two-stage process namely traffic routing and gNB/AP-level resource scheduling. For the traffic routing telemetry and performance measurements will be used to route traffic flows to non-3GPP and 3GPP networks in real-time by eAT3S. Within 5G-CLARITY, a real-time AT3S traffic routing will be based on *(i)* comparing the predefined threshold values of network service related KPIs and the real-time performance measurements/telemetry data; and *(ii)* a machine learning algorithm that routes traffic onto different WATs by using the telemetry data and performance measurements. It is important to note that in the former routing method, the WAT-specific telemetry data is used along with the path performance measurements to decide traffic routing decision. Having a WAT-specific telemetry data and using it within AT3S routing leveraging O-RAN reference architecture is not considered/defined in the current 3GPP AT3S framework. Therefore, both noted routing methods are novel solutions proposed in 5G-CLARITY.

5G-CLARITY will also enable a Dynamic Spectrum Access (DSA) paradigm based on the use of CBRS SAS architecture. This innovative regulatory regime might be potentially replicated in other geographies besides the US. Its intrinsic advantage is that since it is based on a 3-tier approach, it can be mapped to regulatory

spectrum regimes spanning from usual traditional licensed ones to others where incumbent protection or local vertical licenses are granted to private network deployments. Figure 2.9 shows the logical CBRS architecture within the scope of 5GNR. In Figure 2.11 it is shown how this logical architecture will be supported in O-RAN context via a SAS client xApp within Accelleran dRAX framework.

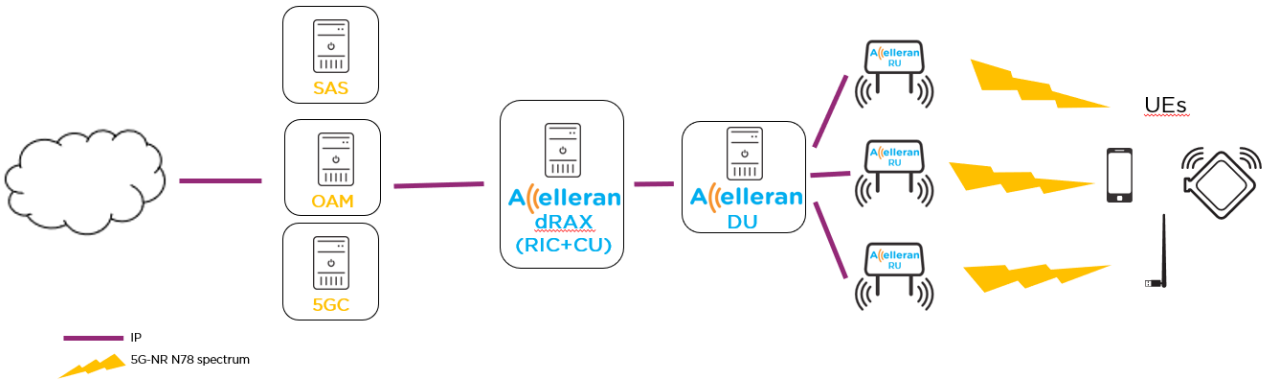

Figure 2.11: Overall CBRS logical architecture with 5G.

The external SAS acting as an automated and intelligent dynamic spectrum management coordination entity will be simulated/emulated to provide the RF spectrum parameters that the 5G-CLARITY 5G NR SAS client requires for operation. The current Accelleran integrated LTE SAS client used in Accelleran CBRS Small Cells will be implemented as an xApp and enhanced to support 5G NR within an O-RAN dRAX context for 5G-CLARITY.

2.3 Advanced localization and synchronization capabilities

2.3.1 State-of-the-art

The knowledge of an accurate and real-time location is a key requirement for many new services, either for indoor or outdoor scenarios. Particularly for indoor environments, given the inability to rely on a precise localization using GNSS/GPS systems, the use of measurements made by wireless devices on cellular radio signals is necessary to estimate the location or position of a device. A relevant example is the operation of private (non-public) 5G networks in industrial environments.

The location-based services that are enabled by the latest developments in wireless networks, cellular networks (5G), etc., are nowadays receiving significant attention. For example, the 5G Alliance for Connected Industries and Automation (5G-ACIA) has described the functional requirements for exposing the capabilities expected from non-public 5G systems to connected industries and automation application [22]. 3GPP carried out a study on positioning use cases for both outdoor and indoor environments, which analyses positioning use cases and complements existing work on 5G use cases involving positioning needs in order to identify potential requirements for 5G positioning services [23].

RF-based positioning has received a lot of attention in the literature, and the research and development work carried out in the last decades is still relevant for current communications systems. As an example, 5G targets a positioning accuracy that is one order of magnitude better than in current cellular networks (LTE) [24]. To that end, 5G embraces the concept of location-aware communication systems, where high-accuracy positioning will demand a level of accuracy less than 1 meter in more than 95% of service area, including communication in indoor environments [25]. Next 3GPP Releases are expected to further specify methods for sub-meter accuracy and low latency.

Location-aware applications in 5G are proved much more attractive, as the recent work leverages on the operation at higher carrier frequencies (millimeter waves) [26] and on the utilization of massive antenna

arrays, which provide degrees of freedom to improve the positioning accuracy compared to current localization systems [27][28].The use of these technologies entail additional challenges when considering mobile environments to ensure a continuous communication, e.g. need of large beam training overhead. Integration of features such as ranging and localization in wireless communications systems are becoming paramount to overcome these limitations [29]. As an example, pencil beams achieved via antenna arrays at these frequencies ensure a wide coverage area (distance), but at the expense of very high setup time due to the exhaustive searching process. Localization provides a promising solution for finding out the best mmWave beams within a small setup time.

3GPP claims the achievement of 10s of cm-accuracy in Rel-16, with the aim at enhancing these values (cm accuracy), promote the use of positioning for vehicle-to-everything (V2X) communications, leveraging 3D positioning (looking at vertical and horizontal), and with the goal of latency and reliability improvements [30]. Recent research [31] has shown some simulation results that prove that both Uplink Angle-of-Arrival (UL-AOA) and Downlink Angle-of-Departure (DL-AOD) can reach meter level accuracy in indoor scenarios

Known Received Signal Strength (RSS)-based and RF-based algorithms have been well developed and used in indoor visible light positioning systems, for which implementations have benefited from the penetration of LEDs in the lighting market. In an indoor Visible Light Communication system, the received signal power follows a channel model which includes the position of both transmitter and receiver. Random device orientation significantly affects the positioning accuracy performance of the RSS-based approaches. To achieve strict positioning accuracy requirements, angle diversity transmitter and angle diversity receiver are seen as the more relevant approaches, being independent from the environment, the device orientation and the AP/LED luminaire deployment. Optical Camera Communications (OCC) is also considered as a LiFi-specific positioning system, leveraging the widespread availability of CMOS cameras in smartphones. Indoor OCC schemes are the most prominent, as the majority of the intended applications among the use cases listed in the standardization are for indoor environments [32]

A more thorough review of the SotA in localization systems and requirements for 5G networks is provided in 5G-CLARITY [2].

One key requirement for localization in a wireless communication system is the synchronization of BSs and UEs. A reliable distribution of synchronization references throughout the network is a key requirement for various vertical sectors. For example, in industrial environments, AGVs need to know their position in cm-level accuracy to speed operations such as docking.

Previous generation networks only required frequency synchronization to keep signals aligned. Nowadays, the strict timing and sync requirements can be avoided by equipping each base station with a GNSS receiver, fact that is far to be realized successfully given the existence of tall buildings, trees and other obstacles present in urban canyons [33]. Synchronization would be unnecessary in case of two-way ranging (TWR) methods [34], where the time-of-flight is utilized to estimate the range and clock bias, or three-way ranging [34] and multi-way ranging [35] to additionally estimate higher-order artifacts such as clock drift and skew. However, such methods have not been evaluated for mmWave systems. Such systems possess different features, including highly sparse channels and directional transmission, making the estimation of the angles of arrival and departure as relevant to localization as the time-of-arrival (ToA).

Current synchronization protocols distribute a master clock across a set of slave devices, which is well supported in wired networks, but has a limited support among wireless technologies.

Synchronization below 1 us is required in several industrial processes, and synchronization in the tens of ns is required to support TSN services, or to synchronize base stations operating with a Cloud RAN architecture. In [36] some of the synchronization requirements are presented. A common consensus is that a synchronization precision of ~130 ns [37] should be satisfactory for most of the 5G use cases. This should be

possible between clusters of RUs belonging to same DU or collocated CU/DU. Nevertheless, the networks are dynamic and prone to failures. Therefore, in order to obtain protection, a single RRU can be connected to multiple DUs and DUs can be connected to multiple RUs. In this case it can become quite unclear which RU belongs to which cluster. Therefore, end-to-end synchronization becomes more prospective.

Table 2-1: Required Timing and Frequency Synchronization Precision

Requirement	Application	Precision
Time Synchronization Accuracy	ToA based precised indoor positioning	~ 1-5 ns
	CoMP type of coordinated RAN features OTDOA positioning of emergency services	<= 1 us
	Time Division Duplexing (TDD)	~ 1.5 us
	Carrier Aggregation (CA) Evolved Multimedia Broadcast Multicast Services (eMBMS)	~ 3-5 us
	UE timing	~ 10 us
Frequency Synchronization Accuracy	Frequency Division Duplexing (FDD).	50 ppb

A thorough revision of the state-of-the-art regarding localization and synchronization has been included in 5G-CLARITY D3.1 [2].

2.3.2 Innovation in 5G-CLARITY

The advanced localization system of the 5G-CLARITY project aims to unify the user positioning functionality, which is independent of the underlying wireless technology employed. This system will enable the use of different positioning technologies like Wi-Fi, LiFi, mmWave, OCC, etc., regardless of the physical layer they use and the precision they can deliver. It will leverage the integration of the available technologies to obtain the best possible location estimation based on user or application requirements.

In this framework, new advanced positioning methods will be used, enabling high precision indoor localization, where GNSS support is not available. These methods are also applicable for outdoor environments, which are usually less challenging for Sub-6 wireless systems but could impair heavily the communication when using mmWave systems in rainy conditions.

These methods use the larger bandwidths made available by some of the aforementioned technologies to increase the localization precision and accuracy in indoor environments, so as to reduce the ToF sensitivity to multipath propagation. Depending on the available channel bandwidth, these technologies will be able to achieve from sub-meter to cm-positioning precision.

5G-CLARITY will develop a joint radio, mmWave, LiFi, and OCC system for real-time high-resolution positioning. It will fuse all available positioning information to provide a beyond state-of-the-art performance in positioning and tracking of UE. ML algorithms would be developed for detection of NLoS scenarios as well as to combine the positioning information from different technologies optimally. A localization server would be developed in order to enable a unified positioning interface regardless of the underlying positioning technology.

The 5G-CLARITY multi-connectivity framework not only targets an advanced precise localization but also will offer support for including new wireless positioning technologies. These technologies can be used both in transport network and RAN devices to additionally improve the synchronization precision in the presence of

multipath propagation.

Value added services stemming from the work on positioning and synchronization will be adressed, taking into account the relation existing between the two in the framework of 5G-CLARITY.

The project 5G-CLARITY promises synchronization to the ns-level via wireless transport of clock distribution protocols. The IEEE 1588 protocol is being supported lately in wireless technologies, making it suitable for distribution of the clock in private networks, e.g. in the context of Industry 4.0 use cases. The proposed synchronization approaches developed in 5G-CLARITY will enable dynamic planning of the synchronization, depending on the available network configuration as well as the failures which can occur.

2.4 AI-driven and intent-based network management

2.4.1 State-of-the-art

Artificial intelligence (AI) has been used in telecommunication systems for a long time, for instance in form of expert systems [38][39]. Recent advances in ML in both theory and available tool chains, allow for a wide-spread introduction of self-learning machine processes. This is supported by a continued softwarization of telecommunication networks, which harmonizes hardware and substitutes it with software wherever possible.

The areas that are affected by introducing modern AI/ML techniques are manifold, as shown in Figure 2.12. The areas include network planning and management, QoS and network performance optimisation as well as incident management and self-healing. Traditionally, these areas have been tackled with human expert intervention and rule-based systems, using static rules that are derived from human expertise. Currently, modern techniques are finding their way into many of these areas of network management [41] [42][43][44], including some cutting-edge technologies such as deep learning. As these technologies are finding their way into the network management domain, telecoms standardisation organisations are keen to provide common approaches for AI in 5G, including ML architectures and pipelines as proposed by ITU [45], O-RAN [46] and ETSI ENI [47].

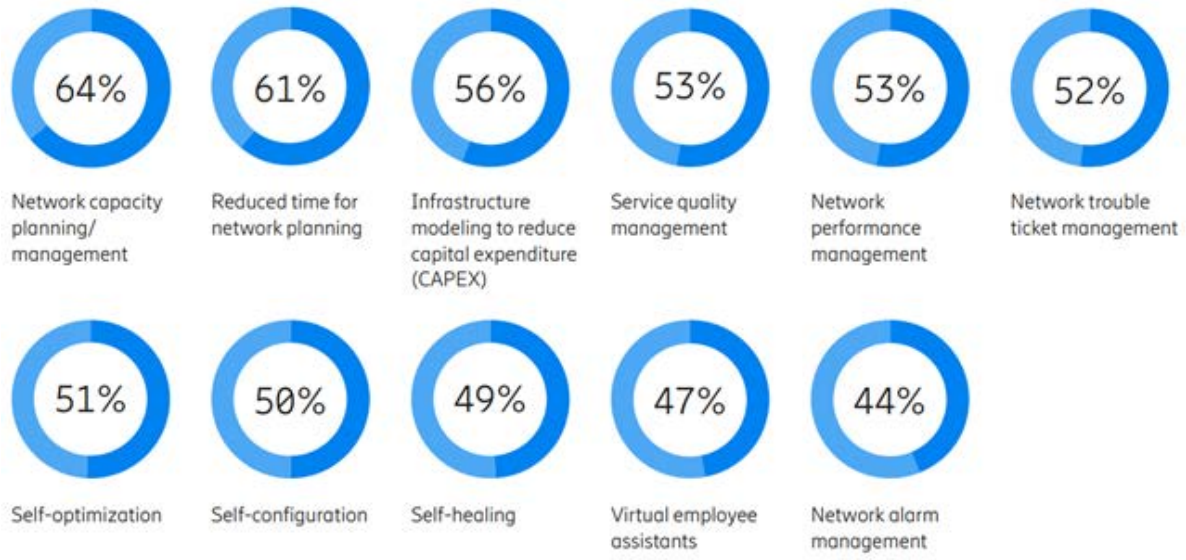

Figure 2.12: Focus areas for adopting AI by network providers (decreasing level of priority) [40].

In separation, AI and intent can help to reduce costs (capital and operational), optimize network performance, and improve revenue (or even create new revenue streams). In combination, both technologies have the potential to accelerate network automation across all domains, i.e. a massive shift of how networks are built, deployed, and operated. This is partially driven by new requirements from private 5G networks, a new and

growing market for telecommunication vendors). We have seen similar impacts when AI and intent were introduced in other areas, for instance navigation (personal and fleet) and logistics (transport, aviation), and personal assistants (e.g. Watson and Wolfram Alpha).

In network management, automation is often addressed using closed control loops [48]. Figure 2.13 shows an example with analytics, policy, and orchestration, similar to what Open Network Automation Platform (ONAP) and Open Source MANO (OSM) support and aligned with the ongoing work on ETSI ZSM 009. AI, particularly ML algorithms and expert systems, are already in use to solve specific problems of event processing and correlations, decision making, and distributed deployment and monitoring. This loop can be extended to a feedback loop (from orchestration to policy and further to analytics). Feedback can be used to signal problems inside the loop, unforeseen changes on the managed entity, or to optimise the loops behaviour.

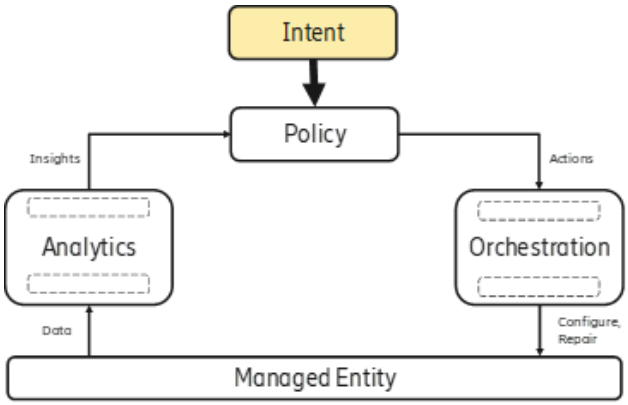

Figure 2.13: Control loop with intent goals

Intents can be used to address three open issues: *(i)* setting the frame and borders in which the loop is allowed to make autonomous decisions, *(ii)* to provide a domain-specific language to describe the loop's functionality and parameters, and *(iii)* to allow to change loop parameters in a secure and safe way without jeopardising the loop's autonomy (since an intent is invariant). With an intent directing the loop, ML can now employ contextual information unavailable before. Since the intent is providing the goal, decisions on how to reach the goal (policies in this loop) can be highly adaptive and even created at runtime. Learning can also extend to the overall loop, using culture (from operator business via intents), previous experience (learned at runtime, trained, templated; in any case changed over time), and new information (which the AI decides to collect or obtain).

2.4.2 Innovation in 5G-CLARITY

5G-CLARITY aims to drive innovation and standardisation of AI-driven intent-based network management with the specification, implementation and demonstration of an AI engine. The AI engine will follow 5G-CLARITY's service-based design principles and will provide ML models as containerised, cloud-native services. Furthermore, the proposed AI engine approach (see D4.1 [49] for details) will enable the decoupling of ML model development and ML model deployment, which allows ML designers to develop ML models without the need to take care of ML model deployment and hosting details. This will greatly increase the ease of use for ML model developers.

5G-CLARITY will also showcase concrete service-based ML use cases that will be hosted by the AI engine. 9 ML use cases will be developed that demonstrate different aspects of AI-supported 5G network management as scoped by the project goals. These ML models will tackle prediction and optimisation problems such as SLA violation, AT3S traffic routing and slice resource provisioning. The ML use cases are described in detail in 5G-CLARITY D4.1 [49].

The communication between the user, ML models and network components will be facilitated by an intent

interface. To this end, 5G-CLARITY will design and demonstrate an intent engine with associated intent language. An important goal of the intent engine is to provide communication abstraction between the 5G-CLARITY platform user and the platform itself, by allowing the user to use simplified language to run ML models that are hosted by the AI engine and to configure 5G-CLARITY slice. Another goal is to simplify communication between the ML models in the AI engine, the telemetry data provider and network functions for seamless execution of AI/ML use cases. 5G-CLARITY will investigate the practicality of this intent-based communication approach in real network scenarios. To enable this, 5G-CLARITY will provide integration between the intent engine, AI engine, network telemetry and various network functions.

2.5 Integration and interoperation of private and public networks

Although the roll-out of private LTE networks is a hot topic in operationally ready carrier networks, 3GPP specifications are looking ahead, studying the applicability of these private networks in 5G systems. Following 3GPP terminology, these private networks are referred to as Non-Public Networks (NPNs). Unlike 4G, wherein private LTE networks are usually deployed entirely isolated from the public network (PLMN), in the 5G era a much closer cooperation between NPNs and PLMN is expected. This cooperation, based on the deployment of hybrid (public-private) scenarios, enables new cross-domain scenarios and allows for a CAPEX and OPEX reduction for public and private actors. However, how this interaction among administrative domains can be effectively realized (and what future-proof business models can benefit from this interaction) is still far from being resolved.

2.5.1 State-of-the-art

The specifications related to the second phase of 5G networks (3GPP Release 16 and beyond) enables the support of NPNs [50]. As defined in [51], NPNs are 5G systems intended for the exclusive use of a private entity such as an enterprise, and might be deployed through a variety of settings, using both virtual and physical entities. According to [50], NPNs are classified into the following two types:

- **Stand-alone NPNs (SNPNs),** defined as those NPNs that do not rely on network functions provided by a PLMN. SNPNs are identified by a combination of a PLMN ID and a Network ID (NID). The UEs registered in an SNPN might access PLMN services, if required, by carrying out another registration with a PLMN using the SNPN User Plane, while the SNPN is playing the role of an untrusted non-3GPP access. A symmetric scenario is allowed to access SNPN services from a PLMN. Rel-16 specs do not include support for interworking with Evolved Packet System (EPS), i.e. the 4G core network, and emergency services in SNPNs as well as roaming and handover between SNPNs.
- **Public Network Integrated NPNs (PNI-NPNs),** defined as those NPNs whose deployment is supported by a PLMN. The PNI-NPNs might be provided by a PLMN as dedicated Data Network Names (DNNs) or as network slices. In the first case, the PLMN defines a mobile pipe to convey NPN traffic to a dedicated mobile gateway (UPF), in charge of dispatching traffic towards a non-public data network. In the second case, the PLMN provisions a dedicated slice, with resources allocated for the exclusive execution of non-public services. Typically, the private UEs will only have access to the PNI-NPN within a limited coverage area. However, Network Slicing does not include any mechanism for limiting the PNI-NPN footprint. To overcome this issue, 3GPP standards specify the Closed Access Group (CAG), which defines a list of subscribers who are allowed to access the cells associated with it. In other words, CAGs serve to apply access control to the UEs depending on their geographic position, thus preventing UEs from accessing the network in areas where they are not permitted to access the Network Slice implementing the PNI-NPN.

Network sharing is a pivotal element to facilitate and lower the cost of the NPNs deployments. 5G Rel-16 specs only support 5G Multi-Operator Core Network (MOCN) sharing architecture, where only the RAN

segment (including RAN infrastructure, functionality, and spectrum) is shared among multiple independent core networks owned by different network operators. The NG-RAN sharing functionality has been extended in Rel-16 to support MOCN scenarios involving NPNs. Specifically, each Cell Identity might be associated with one of the following options: one or several SNPNs, one or several PNI-NPNs (with CAG), or one or several PLMNs. Then, the NG-RAN is configured to radiate the PLMN IDs and PLMN IDs and NIDs of the available PLMNs and SNPNs, respectively, through the Broadcast System Information (BSI) for selection by UEs (see Figure 2.14). The BSI also includes additional parameters per PLMN, such as cell ID and Tracking Areas.

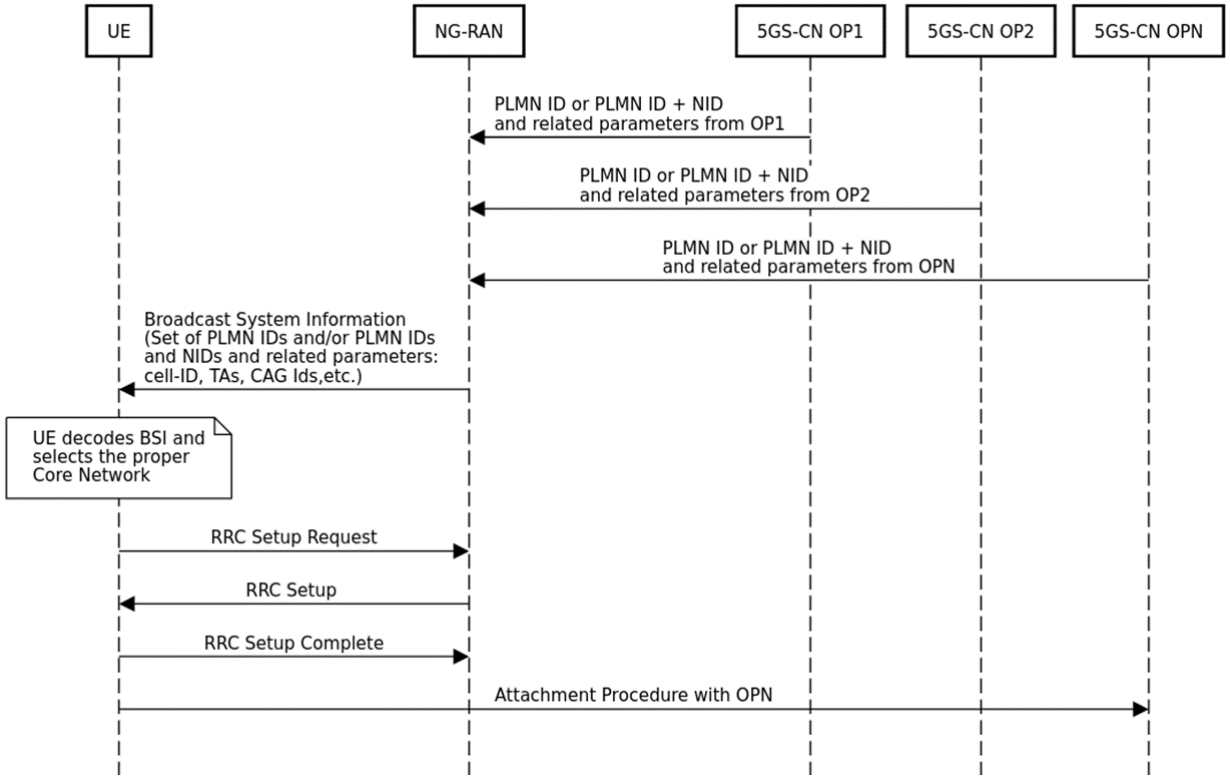

Figure 2.14: Sequence diagram of the available networks (PLMNs and/or SNPNs) broadcasting for selection by UEs [50].

The 5G-ACIA, in its white paper [52], has identified the following deployment options targeted to industrial 5G NPN (see Figure 2.15):

- **Standalone NPN** (isolated deployment). It corresponds with a 3GPP SNPN in which there is no network sharing.
- **NPNs in conjunction with public networks**. It refers to 3GPP NPNs that are fully or partially supported by PLMNs. This category is subdivided into the following three deployment options:
    - Shared Radio Access Network. This deployment scenario corresponds with an SNPN, in which the RAN is shared with a PLMN by means of MOCN.
    - Shared Radio Access Network and Control Plane. It corresponds with a PNI-NPN, in which there is a dedicated UPF within the private premises for the private services.
    - NPN hosted by the public network. This deployment scenario corresponds with a PNI-NPN where all the network functions are shared between the public and private services.

The scenarios listed above are further analysed and extended in [53]-[55]. They offer different trade-offs between performance, security, and cost to cover the necessities of the distinct industrial scenarios.

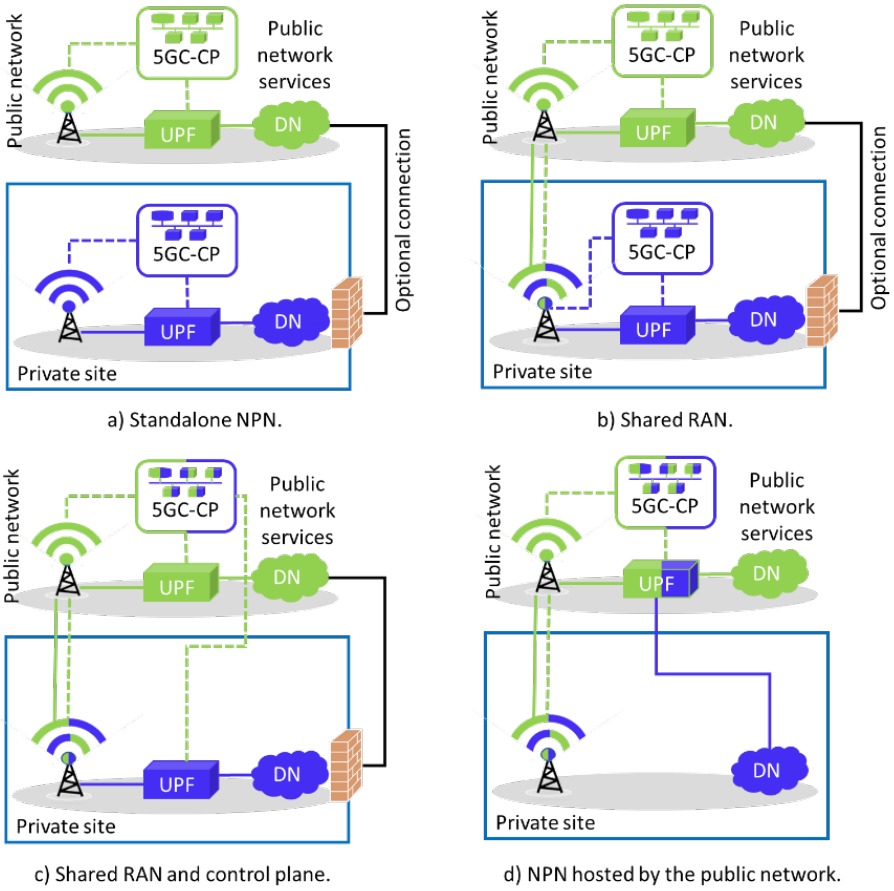

Figure 2.15: 5G-ACIA deployment options for industrial 5G non-public networks.

2.5.2 Innovation in 5G-CLARITY

The key contribution of 5G-CLARITY regarding NPNs and their integration with PLMNs is the proposal of technical solutions, covering user, control and management planes for the following three scenarios:

- **On-premise RAN sharing through MOCN**. A 5G RAN segment, providing coverage within the private premises and operated by a single network operator, is shared among one or several NPNs and one or several PLMNs by means of MOCN. The MNOs of the PLMNs leverage this RAN sharing for extending their footprint within the private premises easily and affordably.
- **PNI-NPN deployed as a slice of a PLMN**. An MNO provides one network slice from its PLMN to realize a PNI-NPN. The MNO might offer the possibility to manage the NPN slice externally upon agreement. The NPN service provider might optionally manage the network slice upon agreement with the MNO. The PNI-NPN coverage does not extend outside the private premises, thanks to the use of CAGs. The PLMN RAN within the private premises might be shared between the PNI-NPN and other network slices for supporting network services. Optionally, the PNI-NPN may have a dedicated on-premise UPF for supporting private ultra-low latency services.
- **Mobility between SNPNs and PLMNs**. There are a set of private UEs accessing public services from an SNPN and need seamless movement between the SNPN and outside. The PLMN provides the private UEs with the RAN coverage out-of-premises. Examples of these services are a set of UAVs that need to exit the private premises to collect some data or the tracking of products equipped with RFIDs since its production until their delivery. In this context, the innovation of 5G-CLARITY is centered around the proposal of distinct methods to ensure session and service continuity of private UEs between SNPN and outside.

3 5G-CLARITY Business Ecosystem and Offered Services

3.1 5G-CLARITY business ecosystem

This section identifies the stakeholders in the 5G-CLARITY ecosystem, building on the models already suggested in the literature, and taking into account two main concrete aspects related to this ecosystem, namely: the presence of both public and private administrative domains, and the combined use of resources from different WATs.

3.1.1 Stakeholder roles in the 5G ecosystem

Different role models have been proposed in the literature so far, most of them based on the primal works from 3GPP [56] and NGMN [57]. For example, 5G-PPP Phase I/II projects have leveraged on the stakeholders (as well as the customer-provider relationships across them) defined in the models proposed by 3GPP and NGMN for the specification of their business ecosystems, extending them as necessary. Based on the outcomes of these projects, along with the feedback provided by the new Phase III projects (e.g. ICT-17 projects), the 5G-PPP Architecture Working Group has proposed an up-to-date actor role model for 5G ecosystem [58]. This model, illustrated in Figure 3.1, is based on the definition of a set of core roles. These roles, listed below, correspond to the stack of roles agreed by the 3GPP SA5 community:

- **Data Centre Service Provider (DCSP)**: provides data centre services. Instances of this role design, build and operate their data centres. These data centres can be of any nature, from large-scale public cloud infrastructures to highly specialized edge sites. The multi-site, multi-cloud capability of current network orchestration platforms allows for an implicit aggregation without a concrete entity playing that role.
- **Virtualized Infrastructure Service Provider (VISP)**: provides virtualized infrastructure services. Instances of this role design, build and operate the virtualization infrastructure(s) under their management scope. This role is equivalent to the VISP role as originally defined in [59], but extended to also include the provisioning of data forwarding services across WAN infrastructure (what is usually referred as transport services in 3GPP documents). This means that VISP managed infrastructure comprises both computing resources (i.e. VISP-C managed infrastructure) and transport resources (e.g. VISP-T managed infrastructure). Unlike a DSCP, which typically offers simple services for consumption of "raw resources" (i.e. host servers) located in a centralized location (datacentre), a VISP-C offers access to a variety of virtual resources by aggregating multiple technology domains and making them accessible through a single, unified Application Programming Interface (API).
- **Network Operator (NOP)**: in charge of orchestrating resources, potentially from multiple virtualized infrastructure providers (VISP). The NOP uses aggregated virtualized infrastructure services to design, build and operate network services that are offered to the upper service provider(s).
- **Service Provider (SP)**: offers services through own/leased/brokered network to one or multiple Service Customers (SC). It designs, builds, and operates these services using aggregated network services. Depending on the service offered to SCs, different categories can be defined for a SP, including Communication Service Provider (CSP), Digital Service Provider (DSP), and Network Slice as a Service (NSaaS) provider. The last role represents a particularization of the CSP/DSP, whereby the communication/digital service is provided in the form of a slice.
- **Service Customer (SC)**: represents the consumer of communication/digital services. This role ranges from an end-user, to an SME doing business on a specific vertical or to a large service provider that

offers online application services. This role is also commonly referred to as tenant.

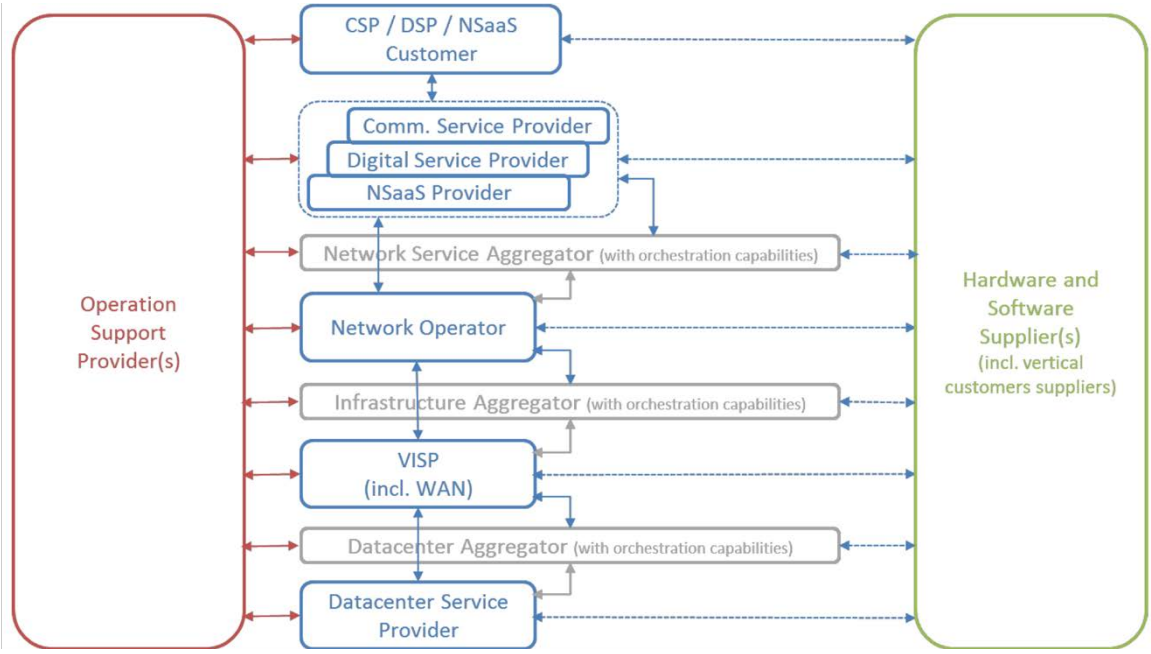

Figure 3.1: Stakeholder roles in the 5G ecosystem [58].

Apart from the core roles listed above, the diagram includes a number of ancillary roles, relevant for the provisioning of services but not considered a core part of the 5G ecosystem. They are related with supply, support and aggregation activities.

- **Suppliers of hardware and software**, required for running 5G infrastructure and executing experiments on it. These suppliers are grouped into a single, general role to abstract all the potential vertical and transversal relationships among them and with the different core roles. 3GPP identifies three individual supplier roles, namely Network Equipment Provider, NFVI supplier and Hardware supplier.
- **Operation supporters**, in many cases connected to the suppliers discussed above, and focused on the integration and control of specialized elements. These supporters also provide the different core roles with advanced (e.g. AI-based) mechanisms for an intelligent and zero-touch provisioning of their services. The role of operation support provider has been identified in some earlier projects (e.g. 5GEx project [60]), and it is essential in any realistic 5G network service provisioning scenario.
- **Aggregators**, in charge of providing means for federation. Originally defined in some recent projects (e.g. SLICENET project [61]), the role of aggregators is essential for considering federation schemas, especially when federated services are provided by the aggregator themselves to consumers, abstracting their individual services of the federation components. The diagram depicts aggregator roles above all core roles, except the top one, directly connected to the customers.

3.1.2 5G-CLARITY actor role model

The diagram presented in Figure 3.1 is aligned with the scope of 5G-PPP ICT-17 projects [62], focused on building up a pan-European large-scale 5G test platform for advanced vertical experimentation. Leveraging on 5G public resources (i.e. 3GPP Rel-15/16 5G systems deployed across multiple EU cloud nodes) and the use of SDN/NFV technologies, this platform aims at demonstrating that the key 5G KPIs can be met for a myriad of vertical industry use-cases. The verticals (represented by the 5G-PPP ICT-19 projects [62]) will access and use the ICT-17 provided platforms to test and validate their use cases through the execution of a set of trials.

The above reasoning shows the limited scope of the current 5G-PPP role model, focused on the **sole use of 3GPP 5G access** solutions for the provision of **public services**. 5G-CLARITY represents a significant leap forward, considering the **combined use of multiple wireless access technologies** for the provision of both **public and private services**. For the E2E service delivery, 5G-CLARITY makes use of an on-premise infrastructure, formed of a set of compute and network resources owned by a private entity. This private infrastructure is stand-alone, though interoperation with the PLMN is also supported, for the cases where services span beyond the perimeter of the defined premises. Interoperation between private and public network can happen at the infrastructure layer (e.g. data plane connectivity between private site and the PLMN's ingress node) and at the network function layer (e.g. a private network function offloaded to the PLMN, a public network function executed on the on-premise infrastructure). The latter assumes network functions are deployed virtualized.

To reflect the novelties above, a new role model needs to be defined. The result is the 5G-CLARITY role model, depicted in Figure 3.2. The key principles of this diagram are described below.

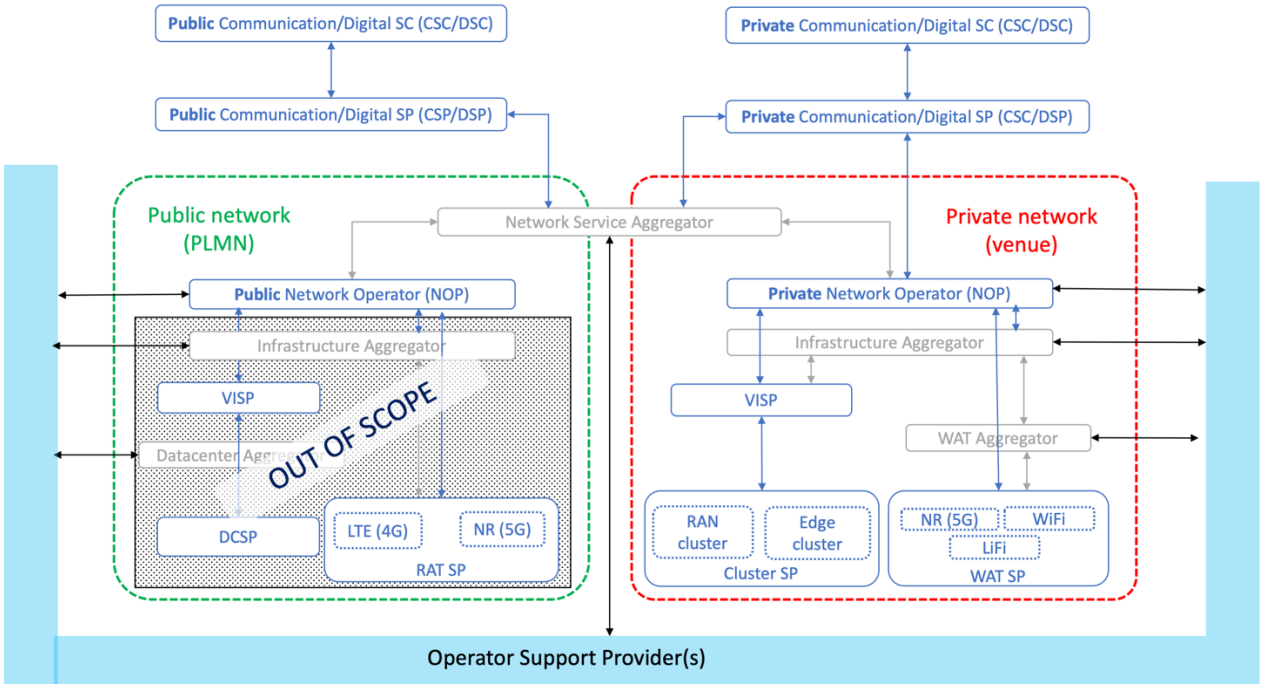

Figure 3.2: 5G-CLARITY actor role model.

On the one hand, the private and public roles are decoupled, to keep in-house management and orchestration separated from provisioning activities executed on the PLMN. For example, the management scope of a private NOP is limited to on-premise resources, while the public NOP can only act on PLMN resources. This decoupling ensures the private network can be operated independently of the PLMN, facilitating the realization of SNPNs. For PNI-NPN scenarios, the network service aggregator oversees providing the (necessary) means for the public-private integration.

On the other hand, additional roles are defined. These new roles, belonging to the private administrative domain, allow for dealing with the on-premise operational aspects intentionally omitted with the 5G-PPP model (which is only focused on the PLMN). A brief summary of these roles is as follows:

- **WAT service provider**: provides services related to one or more wireless access technologies, including 5GNR, Wi-Fi and LiFi.
- **WAT aggregator:** allows federating different WATs together, for a unified and consistent management of wireless resources when used in conjunction (e.g. for bandwidth aggregation,

enhanced reliability, etc.).

- **Cluster service provider**: provides integrated infrastructural services in local environments. Instances of this role design, build and operate the cluster(s) under their management, including RAN clusters and edge clusters. These clusters are built out of small-scale servers, sized for local executions and typically provisioned with hardware acceleration assets (e.g. GPUs, FPGAs/ASICs) that provide enhanced, line-rate data processing capabilities on the network path for uRLLC support. This constitutes a key difference with respect to the commodity servers in the data centres, establishing a clear demarcation point between the roles of cluster service provider and DCSP.

The scope of 5G-CLARITY is the **provision of both public and private services using a private infrastructure**. This means that although the diagram presented Figure 3.2 provides a complete actor role model for both public and private administrative domains, we will mainly focus on private roles, considering their interaction with the public part through the network service aggregator role. This means that the roles which are placed below the public NOP are out of scope of 5G-CLARITY. These are assumed as defined in the 5G-PPP role model in Figure 3.1.

3.2 5G-CLARITY services

5G-CLARITY builds an innovative service platform able to provide B5G network capabilities over a private infrastructure. This rich set of capabilities can be flexibly adapted, combined, and extended to support a wide variety of 5G-CLARITY services. In this section, we provide an overview of the different services which are within the scope of the 5G-CLARITY ecosystem. This includes customer-facing services (Section 3.2.1) and resource-facing services (Section 3.2.2), with slicing (Section 3.2.3) establishing necessary mapping mechanisms across them.

3.2.1 5G-CLARITY customer-facing services

5G-CLARITY customer-facing services represent advanced communication / digital services that can be provisioned using the 5G-CLARITY system. As specified in Table 3-1, these services are intended for private use (Industry 4.0 services, banking services, automotive service) or public use (e.g. mobile broadband experience, smart city service). In the first case, the service is provided by private CSPs/DSPs to private CSCs/DSCs. In the second case, the CSP/DSP and CSC/DSC are both public.

Table 3-1: 5G-CLARITY Customer-Facing Services.

5G-CLARITY Role		5G-CLARITY service for private use	5G-CLARITY service for public use
CSP/DSP[1]		Industry vertical (a company from a vertical market)	MNO (telecom service provider)
CSC/DSC	**Internal[2]**	Industry vertical employees (e.g. factory worker, in industry 4.0)	X
	External[3]	Private subscribers (e.g. end-users using immersive applications from a tourism company; drivers using V2X apps from an automotive company; customers attending a sports event).	Public subscribers (e.g. end-users using mobile broadband services everywhere, including private venues like stadiums or museums; citizens using smart city applications).

[1] As shown in Figure 3.2, a public CSP/DSP can provide communication/digital services using the private infrastructure. The resources from this on-premise infrastructure are managed by the private NOP.
[2] The Internal attribute means that the CSC/DSC and the CSP/DSP belong to the same organization.
[3] The External attribute means that the CSC/DSC does not belong to the organization taking the role of the CSP/DSP.

3.2.2 5G-CLARITY resource-facing services

5G-CLARITY resource-facing services are on-premise infrastructure services that are deployed using 5G-CLARITY private resources. These resources include a set of wireless access nodes (i.e. gNBs, Wi-Fi/LiFi access points), transport network devices (i.e. Ethernet switches, and optionally TSN equipment) and compute servers (i.e. RAN and edge clusters).

Unlike a customer-facing service, managed by a CSP/DSP (either public or private), a resource-facing service is always operated by a private NOP. In 5G-CLARITY system, there exists three types of resource-facing services: 5G-CLARITY wireless service, 5G-CLARITY compute service and 5G-CLARITY transport service.

A 5G-CLARITY **wireless service** is the configuration that needs to be set on one or more wireless access nodes, to make them operationally ready. This configuration requires the definition of network identifiers on individual nodes, and the association of these identifiers with resource quotas, so that nodes can allocate wireless resources as needed. Access nodes from different WATs require the definition of separate 5G-CLARITY wireless services: Wi-Fi/LiFi services and LTE/NR services. A Wi-Fi/LiFi service consists in defining a given Service Set Identifier (SSID) over one or more Wi-Fi/LiFi access points, specifying the resource quota corresponding to this SSID. An LTE/NR service consists in defining a given tuple {PLMN ID, NSSAI} over one or more physical gNBs, specifying the resource quota corresponding to this tuple. Both PLMN ID and Network Slice Selection Assistance Identifier (NSSAI) are 3GPP signalling identifiers used in gNBs. For more information about these identifiers, see Annex A – 5G-CLARITY Concepts.

A 5G-CLARITY **compute service** is a composition of Virtual Deployment Units (VDUs) which are executed on 5G-CLARITY clusters, including RAN and edge clusters. These VDUs may host software images corresponding to network functions (NFs) and application functions (AFs), resulting in Virtualized Network Functions (VNFs) and Virtualized Application Functions (VAFs), respectively. From a deployment viewpoint, a 5G-CLARITY compute service can be modelled as a fully virtualized ETSI NFV network service [63]

Finally, a 5G-CLARITY **transport service** represents the configuration that needs to be set on the transport network devices, in order to make them deliver the traffic from a 5G-CLARITY wireless service into a 5G-CLARITY compute service. For both Ethernet (and optionally TSN switching) devices, a 5G-CLARITY transport service may be signalled using an IEEE 802.1Q VLAN tag.

3.2.3 5G-CLARITY slicing

Slicing technology allows splitting 5G-CLARITY infrastructure into a number of segregated logical resource partitions referred to as 5G-CLARITY slices. The use of this technology in 5G-CLARITY differs from the one in 3GPP community: while for 5G-CLARITY is **multi-tenancy support** (the ability to deliver separate resource chunks for different tenants), for 3GPP is **multi-service support** (the ability to deliver tailored functionality for different services). The different goals pursued by 5G-CLARITY and 3GPP also makes their slicing concepts differ. On the one hand, 3GPP slicing concept results in the definition of **network slices**, i.e. are dedicated set of core network functions that are customized for service differentiation. On the other hand, 5G-CLARITY slicing concept results in the definition of **infrastructure slices**, i.e. a set of infrastructure resources that are segregated for their delivery to separate tenants.

A 5G-CLARITY slice is a **logical partition of the** 5G-CLARITY **infrastructure that provides an isolated execution environment for a particular tenant**. This execution environment is formed of composition of 5G-CLARITY resource-facing services, including wireless services, compute services and transport services. Figure 3.3 illustrates an example of two 5G-CLARITY slices. As seen, the individual slices are formed each of two 5G-CLARITY wireless services (Wi-Fi/LiFi service and 5GNR service), one compute service and one transport service. Apart from the use of different signalling identifiers (SSID-A and {PLMNID-A, NSSAI-A} for the red slice, and SSID-B and {PLMNID-B, NSSAI-B} for the green slice), the main difference between slices is on their

compute service. While the green slice only has one VDU (VDU-B1), the red slice has two VDUs (VDU-A1 and VDU-A2). In the latter case, the compute service can be modelled as a multi-VDU network service, or as two separate single-VDU network services.

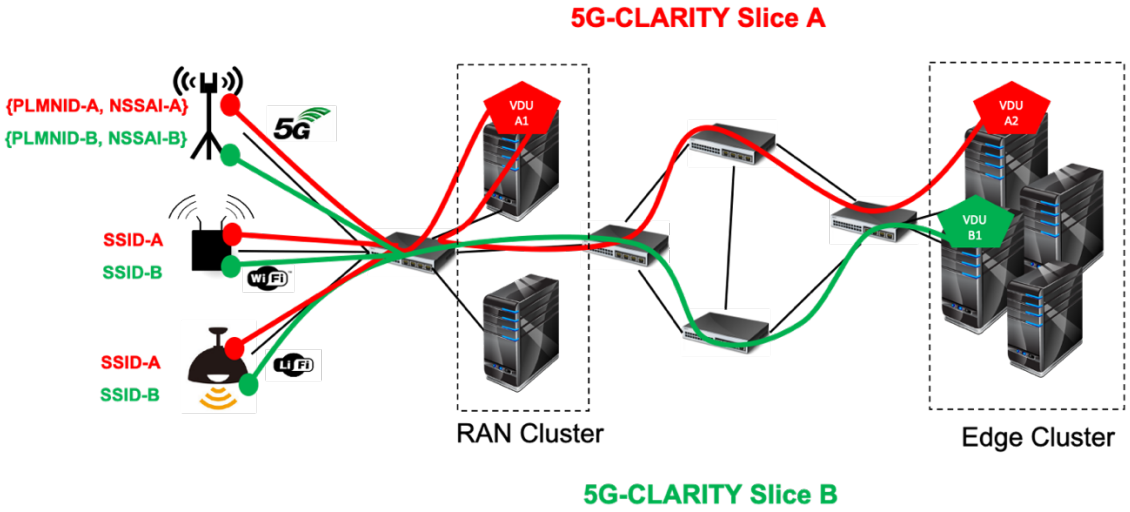

Figure 3.3: Examples of 5G-CLARITY slices

Based on the above rationale, it is clear that a 5G-CLARITY is an on-premise infrastructure slice, built out of resources managed by the private NOP. Depending on the intended use of individual slices, a tenant could be either:

- **CSP/DSP, either public or private**. It uses the slice as received from the private NOP, without further modifications. The CSP/DSP simply builds communication/digital services on top of the slice. With this setup, the services are entirely provisioned within the boundaries of the private venue.
- **Public NOP**. It uses the slice received from the private NOP to define a new slice, with further functionality. The public NOP can then re-sell this slice to public or private CSP/DSPs, fostering Business-to-Business-to-X (B2B2X) partnerships. The process of defining a new slice out of the 5G-CLARITY slice can be done using two different approaches. On the one hand, the public NOP can deploy public VNFs on the 5G-CLARITY infrastructure, using the in-house resources managed by the private NOP, and attach these VNFs to the original 5G-CLARITY slice. Note that the resulting slice is also on-premise slice. On the other hand, the public NOP can extend the original 5G-CLARITY slice to the PLMN, where the MNO can do further processing, e.g. aggregating public VNFs, and extend coverage. Unlike in the first scenario, the resulting slice is not an on-premise slice, as it includes PLMN resources. These off-premise resources are beyond the management scope of private NOP, and therefore of 5G-CLARITY system.
- **Hyperscaler**. It performs an equivalent tenant role as the public NOP, with the only difference that off-premise resources are not PLMN resources, but private cloud resources (e.g. Google Cloud, Amazon Web Services, Microsoft Azure).

In 5G-CLARITY, the focus will be on the above first two tenants.

Finally, it is worth noting that 5G-CLARITY slices are deployed and operated in the form of slice instances. Multiple instances can be created from the same 5G-CLARITY slice, each representing a particular realization of that slice on the 5G-CLARITY physical infrastructure. The resources allocated to a particular slice collectively define the 5G-CLARITY slice quota, consisting of three individual resource quotas:

- 5G-CLARITY wireless quota: it is a set of wireless resources in each gNB or Wi-Fi/LiFi AP allocated to one 5G-CLARITY slice instance. These resources are provided by the WAT SP (see Figure 3.2). The

implementation of a 5G-CLARITY wireless quota depends on the underlying wireless technology. For example, it could be expressed as Physical Radio Blocks (PRBs) in 5GNR, airtime in Wi-Fi, and wavelengths or airtime in LiFi.

- 5G-CLARITY compute quota: it is the set of computing resources (e.g. CPUs), storage (e.g. RAM) and networking resources (i.e. NIC) allocated to one 5G-CLARITY slice instance. These resources are provided by the cluster SP (see Figure 3.2). A 5G-CLARITY compute quota may be implemented using OpenStack.
- 5G-CLARITY transport quota: It is a set of transport resources allocated to one 5G-CLARITY slice instance. The available transport resources depend on the underlying transport technology and could be expressed in terms of data-rate, latency or buffer space.

3.3 5G-CLARITY service delivery models

The 5G-CLARITY ecosystem allows the definition of multiple models for service delivery across different administrative domains. Table 3-2provides a summary of these models, specifying the actors taking part in the service delivery and the provider-customer relationships established between them.

The first three models involve the interaction of actors from public and private domains, required for the deployment and operation of PNI-NPNs. The last model is orthogonal to the previous ones; indeed, it illustrates how 3rd party actors can provide both public and private actors with advanced operation support capabilities for an intelligent orchestration of their managed networks. Individual descriptions of these service delivery models are provided in Annex B – 5G-CLARITY Service Delivery Models .

Table 3-2: 5G-CLARITY Service Delivery Models

Service Delivery Model	Provider → Customer
WAT as a Service	Private NOP → Public NOP
NFV Infrastructure as a Service	Private NOP → Public NOP
	Public NOP → Private NOP
Slice as a Service	Private NOP → Public CSP/DSP or private NOP
	Public NOP → Private NOP
Intelligence as a Service	Operation Support Provider → Private NOP
	Operator Support Provider → Public NOP

4 5G-CLARITY System Architecture

This chapter defines and describes the reference architecture for the 5G-CLARITY system. Section 4.1 specifies the non-functional and functional requirements that the 5G-CLARITY system needs to satisfy. Section 4.2 introduces the proposed 5G-CLARITY system architecture, presenting a high-level overview of the different strata and introducing their key design principles. The details of every stratum, including a description of the individual functional blocks and their offered capabilities, are provided in this deliverable in the following Chapters 4, 5, 6 and 7.

4.1 Architecture requirements

Technical requirements on the 5G-CLARITY system and the different use cases have been already defined in [1]. These requirements have focused on properties of specific industry service and associated KPIs. However, the 5G-CLARITY system aims at providing a future-proof execution environment for a much wider variety of B5G use cases. This requires the definition of much more ambitious system requirements, to open the 5G-CLARITY architecture up for further service innovation. Table 4-1 and Table 4-2 list the 5G-CLARITY system architecture requirements. As seen, these requirements are classified into two categories:

- Functional (F) requirements – define what the 5G-CLARITY system must do or what processing actions it is to take, meaning the expected inputs and outputs. Furthermore, functional requirements describe calculations, technical details, data manipulation and processes.
- Non-Functional (NF) requirements – specify quality attributes of the 5G-CLARITY system. These requirements define the properties that the functions must have, such as performance, usability and data security needs.

Table 4-1: Functional Requirements of the 5G-CLARITY System Architecture

Requirement ID	Requirement Description
CLARITY-SYST-F-R1	The 5G-CLARITY system shall be able to support one or more communication/digital services from a given customer
CLARITY-SYST-F-R2	The 5G-CLARITY system shall be able to support long-live and short-lived communication/digital services.
CLARITY-SYST-F-R3	The 5G-CLARITY system shall expose appropriate interfaces to external customers to allow them to consume capabilities offered by the 5G-CLARITY system.
CLARITY-SYST-F-R4	The 5G-CLARITY system shall support functionality to provide data to the customer according to the customer's requirements (e.g. relevant data, relevant time, relevant form).
CLARITY-SYST-F-R5	The 5G-CLARITY system shall provide a mechanism to perform customer accounting. This information should be available internally and externally (for the 5G-CLARITY customer).
CLARITY-SYST-F-R6	The 5G-CLARITY system shall support integration of both new and legacy functions.
CLARITY-SYST-F-R7	The 5G-CLARITY system shall be able to provision functions using resources of different technology domains, including LTE/5GNR/Wi-Fi/LiFi resources in the access network infrastructure, Ethernet/TSN resources in the transport network infrastructure, and IT hardware/software resources in the compute network infrastructure.
CLARITY-SYST-F-R8	The 5G-CLARITY system shall be able to combine resources from 3GPP technologies (i.e. LTE/5G) and non-3GPP technologies (i.e. Wi-Fi/LiFi) for enhanced bandwidth aggregation and advanced multi-technology features in the access network infrastructure.
CLARITY-SYST-F-R9	The 5G-CLARITY system shall be able to incorporate SDN programmability in the transport network infrastructure.

CLARITY-SYST-F-R10	The 5G-CLARITY system shall be able to provide NFV support in the compute network infrastructure, allowing the deployment and operation of some network/application functions as VNFs/VAFs, when applicable.
CLARITY-SYST-F-R11	The 5G-CLARITY system shall support network slicing.
CLARITY-SYST-F-R12	The 5G-CLARITY system shall support ML assisted decision-making mechanisms.
CLARITY-SYST-F-R13	The 5G-CLARITY system shall support intent-driven network management.
CLARITY-SYST-F-R14	The 5G-CLARITY system shall support functionality that enables collecting and storing up-to-date data[4].
CLARITY-SYST-F-R15	The 5G-CLARITY system shall support confidentiality and integrity of data at rest, in transit and in use.
CLARITY-SYST-F-R16	The 5G-CLARITY system shall support the capability to ensure availability of data, resources, functions and services, in so far as security measures to handle availability threats are concerned.

Table 4-2: Non-functional Requirements of the 5G-CLARITY System Architecture

Requirement ID	Requirement Description
CLARITY-SYST-NF-R1	The 5G-CLARITY system shall be reliable (as carrier class component providing 5 nines availability).
CLARITY-SYST-NF-R2	The 5G-CLARITY system shall support different kinds of customers, including CSPs (e.g. MNOs) and DSPs (e.g. AR/VR service providers, content service providers, IoT service provider, etc).
CLARITY-SYST-NF-R3	The 5G-CLARITY system shall be open and extensible to support integration with hyperscalers (e.g. Amazon Web Services, Google Cloud, Microsoft Azure) in the long term.
CLARITY-SYST-NF-R4	The 5G-CLARITY system shall provide multi-tenancy support, allowing the concurrent execution of communication/digital services from different customers with isolation guarantees in terms of performance and management.
CLARITY-SYST-NF-R5	The 5G-CLARITY system shall keep responsiveness for customer requests.
CLARITY-SYST-NF-R6	The 5G-CLARITY system resources shall include private resources (i.e. on-premise resources), with the optional support of public resources (i.e. PLMN resources).
CLARITY-SYST-NF-R7	The functions built upon 5G-CLARITY system resources shall include service functions (i.e. network functions and application functions) and management functions.
CLARITY-SYST-NF-R8	The 5G-CLARITY system shall support the capability to achieve high data availability.
CLARITY-SYST-NF-R9	The 5G-CLARITY system shall allow the provisioning of NPNs in either of their forms, including SNPNs and PNI-NPNs.

[4] This data include telemetry data (related to infrastructure nodes and functions), logs (related to the system and application software entities) and labelled training for machine learning.

4.2 Baseline architecture design

The 5G-CLARITY system is designed to support multiple combinations of stakeholders (see Section 3.1) over a heterogeneous infrastructure with advanced capabilities (e.g. multi-WAT support, network slicing, high-precision localization) to provide a myriad of B5G services for non-public use. The creation of simplicity out of this complex environment requires applying the principles of abstraction and separation of concerns into the architecture design. The result is an architecture structured into different strata, each with a confined scope that can evolve independently from the rest of strata. In 5G-CLARITY, we envision a system of four major strata: infrastructure stratum, network and application function stratum, management and orchestration stratum, and intelligence stratum. Figure 4.1 captures the logical structure of these strata into the 5G-CLARITY system architecture.

In following sections, the design principles of individual strata are to be specified.

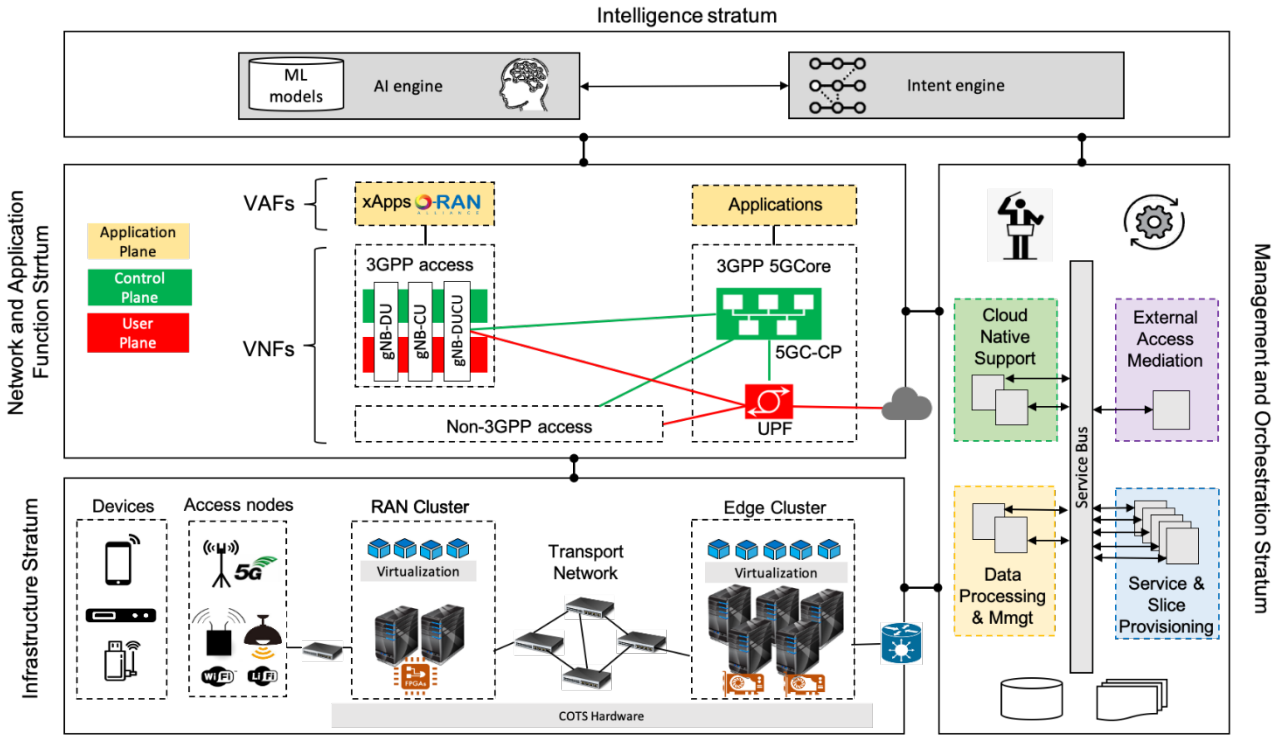

Figure 4.1: 5G-CLARITY system architecture

4.2.1 Infrastructure stratum

The 5G-CLARITY infrastructure stratum defines all the hardware and software resources building up the 5G-CLARITY substrate. This includes user equipment and a wide variety of compute, storage and networking fabric. This fabric can be used to implement bespoke PNFs (e.g. physical access nodes), or to provide a virtualized execution environment for VNF/VAF hosting. VNFs and VAFs are collectively referred to as VxFs.

The design principles that will guide the specification of this 5G-CLARITY architecture stratum are detailed below.

Support for multiple types of devices

This principle represents a significant leap forward for end-user connectivity. As of today, the Global mobile Supplier Association (GSA) has reported that the number of announced 5G devices has broken the 300 barriers of which at least 100 are commercially available [64]. These devices are categorized into multiple

form factors[5], depending on their performance profile and usability (see Figure 4.2). The rate at which new devices are being announced and the diversity of form factors points to continue rapid deployment and uptake of new services in 5G and beyond. Proof of this is the on-going design of devices able to make a combined use of 5G, Wi-Fi and LiFi technologies. These devices, which are not currently available, are envisioned for B5G systems, and thus are within the scope of 5G-CLARITY.

The 5G-CLARITY system will natively have the ability to accommodate a wide variety of devices into the designed architecture. For the sake of simplicity, 5G-CLARITY project will validate this multi-device support feature considering three different form factors: handheld terminals (e.g. smartphones), CPEs and dongle-cards. The combination of these form factors provides a representative sample for showcasing purposes.

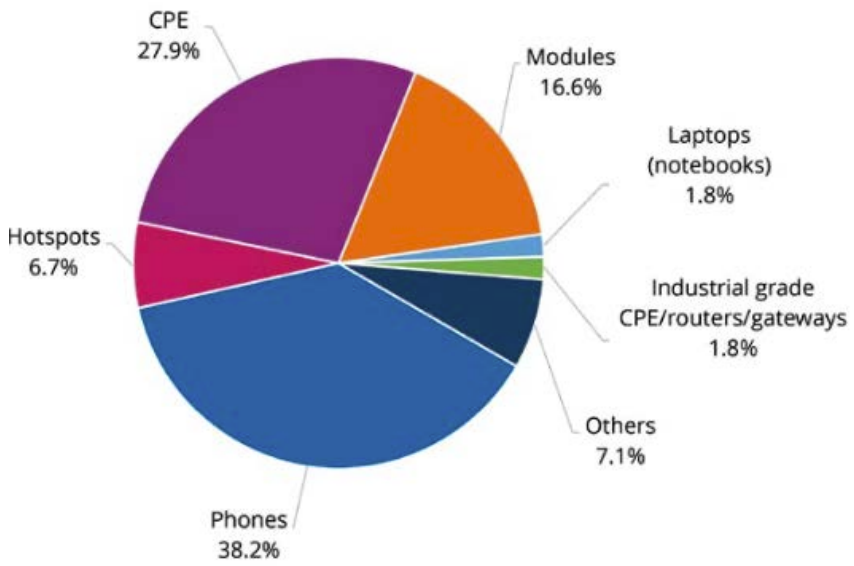

Figure 4.2: Announced 5G devices, by form factor [64].

Coexistence of purpose-built and commercial off-the-shelf (COTS) hardware

This principle is based on the seamless integration of both purpose-built and Commercial-Off-The-Shelf (COTS) hardware platforms to build an interoperable end-to-end infrastructure substrate for private venues. 5G-CLARITY substrate includes two differentiated parts: *physical access nodes*, which provide air interface connectivity to devices using WAT-specific protocol stack (e.g. 3GPP 5GNR for 5G technology, IEEE 802.11 for Wi-Fi and LiFi technologies); and *x86-based commodity servers*, which provide a cloud-ready execution environment for the execution of NFV based services and applications. On the one hand, the access nodes are built using custom-made hardware, typically tightly coupled with the vendor-specific software executed atop. These purpose-built devices suffer from traditional vendor lock-in, which offers operators very limited capabilities, usually through a proprietary interface enabled only for remote configuration purposes. On the other hand, the x86-based commodity servers lie on the use of affordable COTS hardware, with different built-in virtualization capabilities (e.g. VM-based, container-based) that are combined together to provide an enriched NFV Infrastructure (NFVI) for B5G. This NFVI promises to fully capture the promise of cloud and NFV, based on the delivery of Container as a Service (CaaS) functionality along with traditional Infrastructure as a Service (IaaS). Figure 4.3 shows a stepwise journey for the transition towards this future-proof NFVI in 5G-

[5] Up to sixteen different form factors have been identified by the GSA: phones, head-mounted displays, hotspots, indoor CPE, outdoor CPE, laptops/notebooks, modules, snap-on dongles/adapters, industrial-grade CPE / routers / gateways, drones, robots, tablets, TVs, switches/modems and vending machines.

CLARITY system.

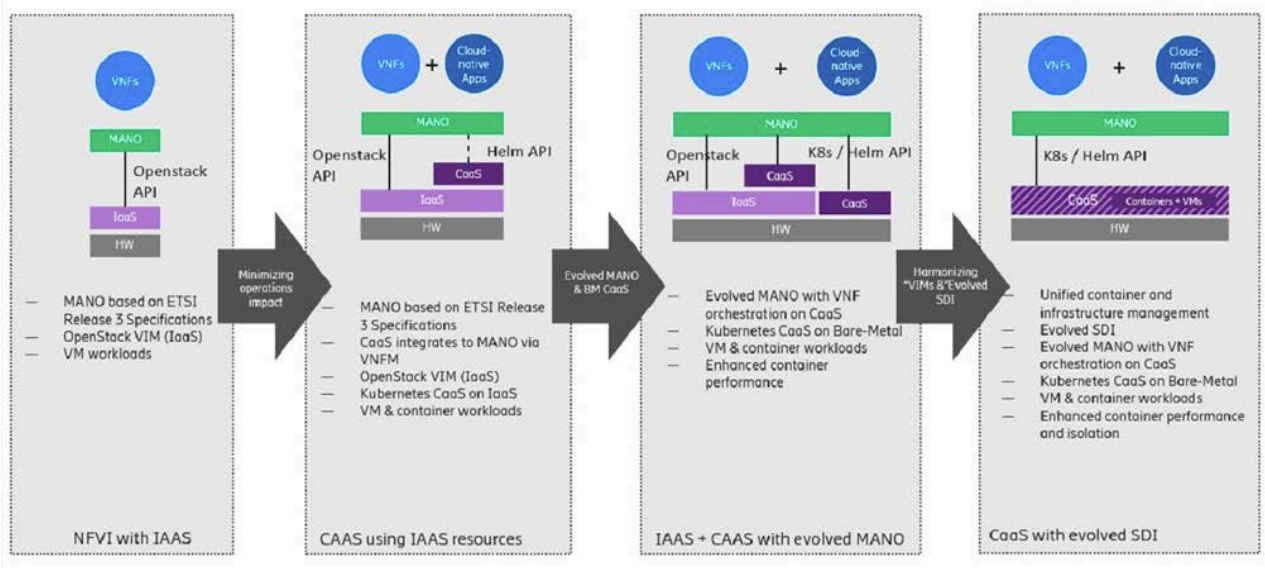

Figure 4.3: NFVI evolution [65].

In-network computing

This principle is about augmenting an NFVI based on general-purpose CPUs with technologies that accelerate compute intensive VxFs. Although COTS servers are well-suited to handle many use cases, there are network functions and workloads which require a high level of QoS (e.g. throughput, latency, jitter), predictable performance and security, especially those dealing with data plane processing in uRLLC scenarios. When implementing such VxFs on standard hardware, operators often expect parity with middleboxes (i.e. purpose-built hardware devices), but what they perceive is a performance gap. In fact, replacing NFs middleboxes with VxFs may have a detrimental effect on their packet-processing performance, such as loss of throughput and/or nondeterministic latency. This is mainly due to the technology limitations imposed by virtualization overheads from multiple layers of packet processing needed as traffic flows from network interface cards (NICs) to VxFs.

According to the above rationale, matching the performance of middleboxes will be one of the key challenges faced by operators in the future with regards to the widespread NFV adoption. This challenge, which has caused NFV to reach a productivity plateau as technology [66], has led to a recent interest in hardware-acceleration techniques for VxFs using externally connected hardware devices, e.g. Graphic Processing Units (GPUs), Field Programmable Gate Arrays (FPGAs), Smart NICs, Network Processing Units (NPUs), etc. Hardware accelerators and CPUs can be used in conjunction such that CPU-intensive tasks (e.g. security, packet processing) can be offloaded from VxFs to hardware accelerators, and the rest of the VxF operations can be performed by the CPU of general-purpose hardware (COTS servers). As a consequence, an improvement in the overall performance can be achieved, freeing up also more CPU cores that can be now dedicated to host new VxFs. These hardware acceleration solutions can be complemented with software-centric accelerators as shown in Figure 4.4.

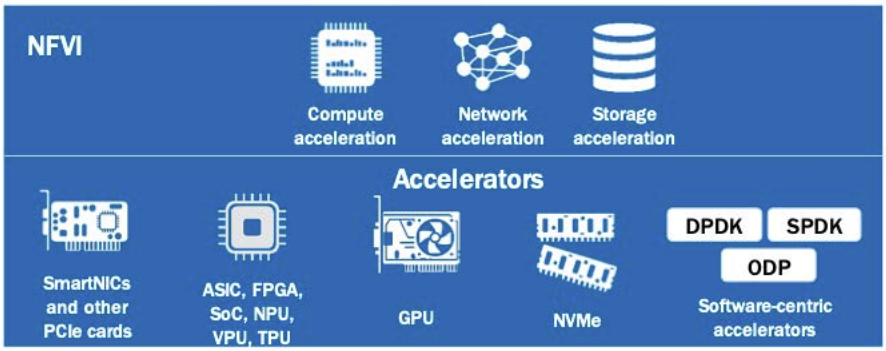

Figure 4.4: Acceleration technologies [67].

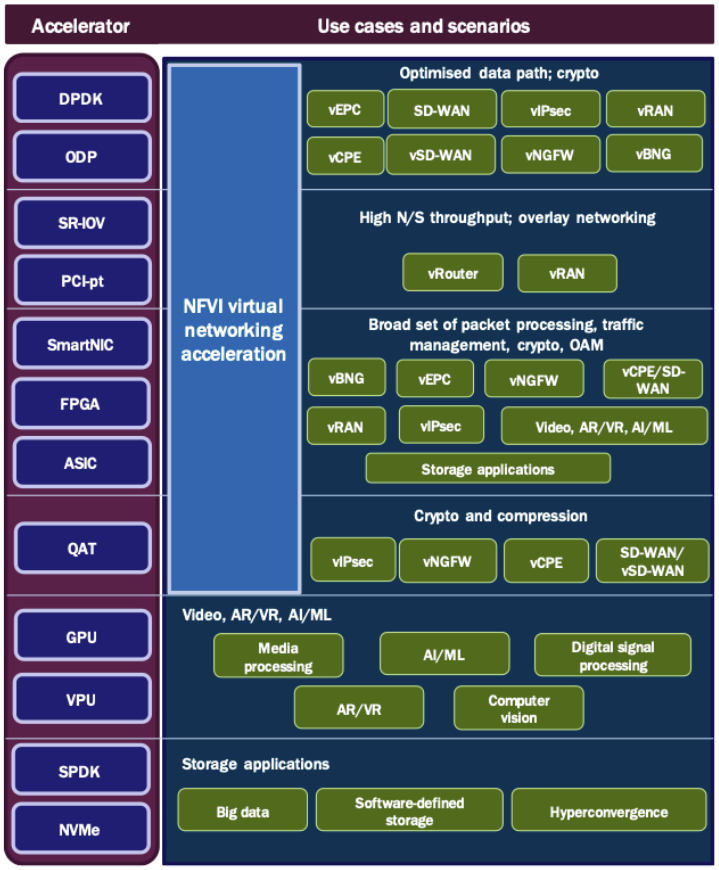

Figure 4.5: Acceleration technologies and use cases [66], [67].

Figure 4.5 provides a landscape of the most promising hardware and software acceleration solutions for a performant NFVI, relating them to specific use cases and scenarios. 5G-CLARITY will leverage on these acceleration solutions to support the execution of delay-critical VxFs on top of the NFVI. For a summary on the applicability of these solution in B5G use cases, see [67].

Clustered NFVI

This principle allows having separate execution environments within the same NFVI, by defining different resource zones. A resource zone is a set of NFVI resources logically grouped according to physical isolation and redundancy capabilities, or to certain administrative policies for the NFVI.

5G-CLARITY will leverage on the above-referred principle to define two separate resource zones: RAN cluster and edge cluster. The definition of these clusters allows 5G-CLARITY to logically separate the management

and VxFs related to RAN from the rest of network domains. While the RAN cluster is for the exclusive use of VxFs providing RAN functionality, the edge cluster provides an on-premise VxF execution environment to host any other (no RAN-related) functionality. Examples of VxFs that can be deployed on the edge cluster include core network functions and needed service applications.

Industry-ready transport network

This principle defines the ability of having a private transport network, able to cope with the new breed of Industrial IoT applications and visionary B5G approach depicted by ITU Focus Group Technologies for Network 2030 (FG NET-2030) in [68]. This requires combining legacy transport solutions from optical and IP/Ethernet technology domains with novel solutions like eCPRI, Flexible Ethernet (Flex-E), Deterministic Networking (DetNet) and Time Sensitive Networking (TSN). These technologies, with great applicability in fronthaul and mid-haul segments, allows providing zero-perceived latency and high availability in (although not limited to) private industrial environments.

5G-CLARITY will take a snapshot of this multi-technology environment to define a convergent transport network based on the combination of Ethernet-based L2 switches and a TSN fabric.

4.2.2 Network function and application stratum

The 5G-CLARITY network and application function stratum includes all VxFs that can be executed on top of the 5G-CLARITY NFVI. These VxFs can be flexibly combined into multiple 5G-CLARITY compute services, each hosting the computing resources of one or more 5G-CLARITY slices. The resource provisioning and allocation of the individual VxFs depends on the needs of the associated 5G-CLARITY slices.

The design of this 5G-CLARITY architecture stratum will be done according to the principles detailed below.

Control and User Separation (CUPS)

Following this SDN principle, in 5G-CLARITY architecture the user plane VxFs are decoupled from control plane VxFs for completely independent capacity scaling, maximum topology flexibility and decoupled technical evolution.

Cloud-native VxFs

This principle is about applying existing DevOps practices and microservices design in VxF on-boarding and operation, making them ready to run on container-based NFVI, i.e. NFVI with built-in CaaS capabilities (see Figure 4.6). A B5G, cloud-native VxF requires the following characteristics: *(i)* the decomposition and modularization of VxF functionality into loosely coupled components; *(ii)* the need for lightweight orchestration, using container technology; *(iii)* the use of service mesh topology; and *(iv)* the reliance of Continuous Integration / Continuous Development (CI/CD) pipeline, and consequently the need for appropriate DevOps toolchain.

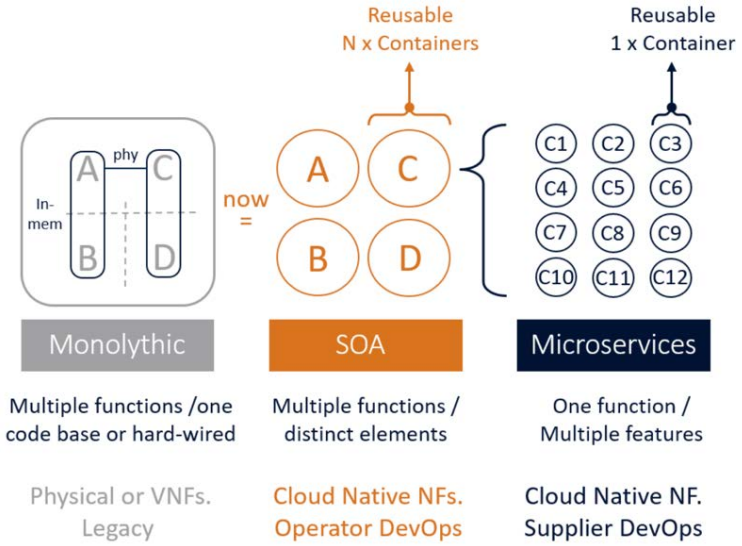

Figure 4.6: Cloud-native VNFs.

This new style allows rearchitecting traditional VxFs, rather monolithic and with some limitations:

1. High resource consumption: VxF consumes a huge amount of hardware to be highly available.
2. Tight coupling with the underlying NFVI: VxFs are developed, configured and tested to run on specific NFV hardware infrastructure.
3. Lack of standard processes for complete management, from development to deployment and monitoring: VxFs developed by different vendors typically have different methodologies for VxF operation in existing NFV environment. The existence of different (vendor-specific) operational model raises the need for manual VxF on-boarding (installation, i.e. day-0 operation), deployment (instantiation + configuration + activation, i.e. day-1 operation) and run-time supervision (i.e. day-2 operation).
4. Lack of automation: this prevents VxFs to be automatically scaled or (re)configured, e.g. in order to serve the sudden spike in demand for resource utilization.
5. No multi-tenancy support: VxFs cannot be easily shared for infrastructure reuse.

Building cloud-native VxFs overcomes the previously discussed limitations and provides the following benefits:

6. Provisioning of APIs for self-management purposes: automated installation and configuration; automated scaling; self-healing and fault tolerance; automated monitoring and analysis of VxFs for capacity management, performance assurance and fault supervision; and automated upgrading and updating of VNFs, for applying new releases and patches.
7. Standard and simplified management: this enables lower power consumption, reducing unnecessary allocated resources.
8. Multi-tenant VNFs: reusability and sharing of processes within VxFs can be easily achieved, in NFV environments.

The journey from today's monolithic VxFs to cloud native VxFs is illustrated in Figure 4.7. For more information on this evolution, see [69] and [70].

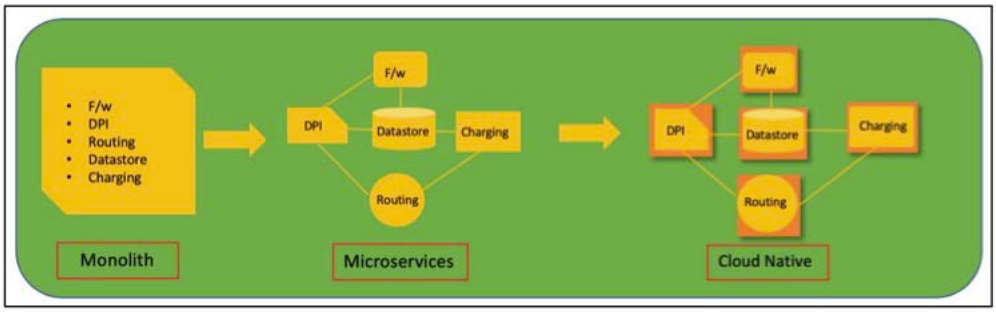

Figure 4.7: Evolution to cloud-native VNFs [69].

The applicability of cloud-native approach in telco environments requires the adoption of a proper architecture. This architecture shall be based on the domain driven design [71], which is required for an efficient (and flexible) VNF composition and communication. In this regard, there exist three main candidate solutions: microservice architecture, Software Oriented Architecture (SOA) and Service Based Architecture (SBA). Table 4-3 provides a comparative analysis among these three architecture solutions. As seen, SBA represents hybrid approach between microservice architecture (fine-grained approach) and SOA (coarse-grained approach). This trade-off solution makes SBA the best architecture solution for the operation of cloud native VxFs.

Table 4-3: Comparison of Architectures for Cloud-Native VNFs

Criterion	Microservice architecture	SOA	SBA
Agility	High	Low	Medium
Deployment	High	Low	Medium
Testability	High	Low	Medium
Scalability	High	Medium	Medium
Performance	Medium	Low	Medium
Simplicity	Medium	Low	Medium

5G-CLARITY system will adopt cloud-native design for the VxFs which are executed on the NFVI's edge cluster, and an SBA for their interaction. This results in disaggregated, containerized ETSI NFV network services, with individual VxFs attached to a common software bus. This bus allows VxFs to exchange information between them using service-based interfaces, typically conveying JSON encoded data over the HTTP2/TCP protocol. Much of this work is already done, since 5G stand-alone (SA) solution leverages on SBA for the realization of 5GC control plane functionality.

RAN functional splitting

To take advantage of the RAN virtualization benefits in terms of scalability and centralization, 3GPP 5GNR allows to split the gNB functionality into three logical modules: Radio Unit (RU), provisioned with RF circuitry; the Distributed Unit (gNB-DU), hosting gNB real-time functions; and the Centralized Unit (gNB-CU), hosting gNB non-real-time functions. The distribution of gNB functionality (see Figure 4.8) across these three modules has been in discussion within 3GPP and the industry in general for some time. At this stage, eight split options have been identified. The critical question here is where to make the necessary splits between gNB-CU, gNB-DU and RU: *High-Layer Split (HLS)*, between gNB-CU and gNB-DU, and *Low-Layer Split (LLS)*, between gNB-DU and RU. This is the question that the RAN functional splitting is addressing.

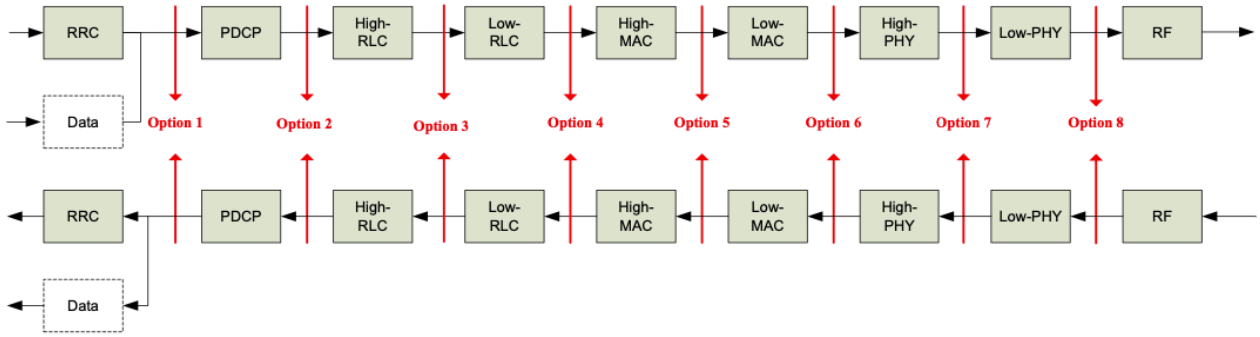

Figure 4.8: RAN decomposition options. RRC and data (SDAP) provide control and user plane functionality, respectively.

If all the system and software in cloud RAN architecture are supplied by a single vendor, then the splits can be dictated by the vendor. This approach does not support open architectures and therefore makes it difficult for carriers to source systems and software from multiple suppliers. In this line, the industry has coalesced around split option 2 for the HLS. This is now the standard 3GPP F1 mid-haul interface defined to connect the gNB-CU and gNB-DU. There is less agreement on the LLS between gNB-DU and RU, although many deployments will use split option 7-2, a split variant of split option 7.

5G-CLARITY will follow the above industry recommendations, providing the gNB decomposition as illustrated in Figure 4.9. On the one hand, the RU will always correspond to the 5G physical access node, and therefore will be deployed as PNF in the 5G-CLARITY infrastructure stratum (Section 5.2). On the other hand, with regards to gNB-DU and gNB-CU, different variants can be selected in 5G-CLARITY:

- Both gNB-DU and gNB-CU executed atop the NFVI's RAN cluster but deployed separately. In this scenario gNB-DU and gNB-CU can be modelled as (monolithic) VNFs.
- Both gNB-DU and gNB-CU executed atop the NFVI's RAN cluster but deployed as a single (monolithic) VNF. The co-location described in this scenario corresponds to classical C-RAN, with this VNF providing BaseBand Unit (BBU) functionality.
- gNB-DU co-located with the RU, and gNB-CU running on the NFVI's RAN cluster. In this scenario, gNB-CU is deployed as a (monolithic) VNF, and the gNB-DU and RU are integrated into a single PNF.

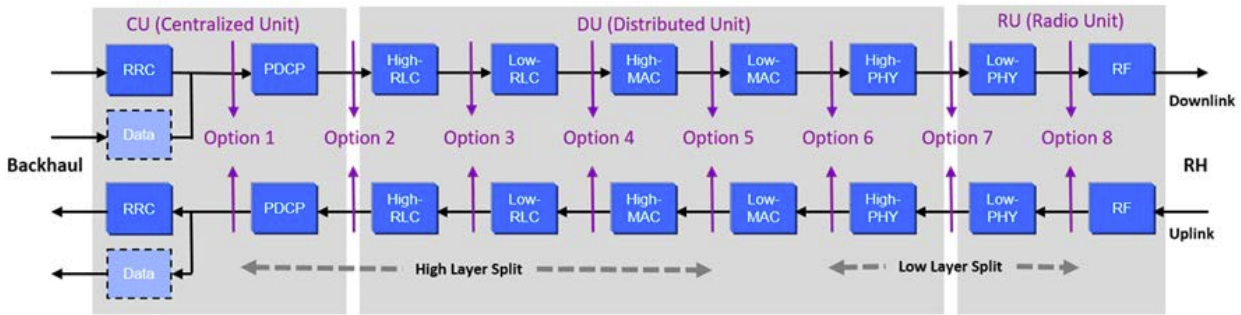

Figure 4.9: HLS and LLS in 5G-CLARITY system

Integration of O-RAN alliance framework

The O-RAN Alliance [72] is a world-wide community of more than 170 mobile network operators, vendors and research and academic institutions founded to accelerate the adoption of virtualized RAN on white box hardware, with embedded AI-powered radio control and SDN/NFV mechanisms. The mission of O-RAN is to re-shape the industry towards more intelligent, open, virtualized and fully interoperable mobile networks,

therefore enabling a more competitive and vibrant RAN supplier ecosystem with faster innovation to improve user experience. To that end, O-RAN has defined an architecture based on well-defined, standardized interfaces in full support of and complementary to standards promoted by 3GPP and other industry standards organizations (e.g. ITU-T, Small Cell Forum). Figure 4.10 provides a simplified view of O-RAN architectural framework. As it can be seen, the O-RAN architecture follows the 3-split solution captured in Figure 4.9, with Open Fronthaul implementing the option 7-2 for LLS. Indeed, the definition of this option 7-2 is one of the key outcomes attributed to the O-RAN alliance.

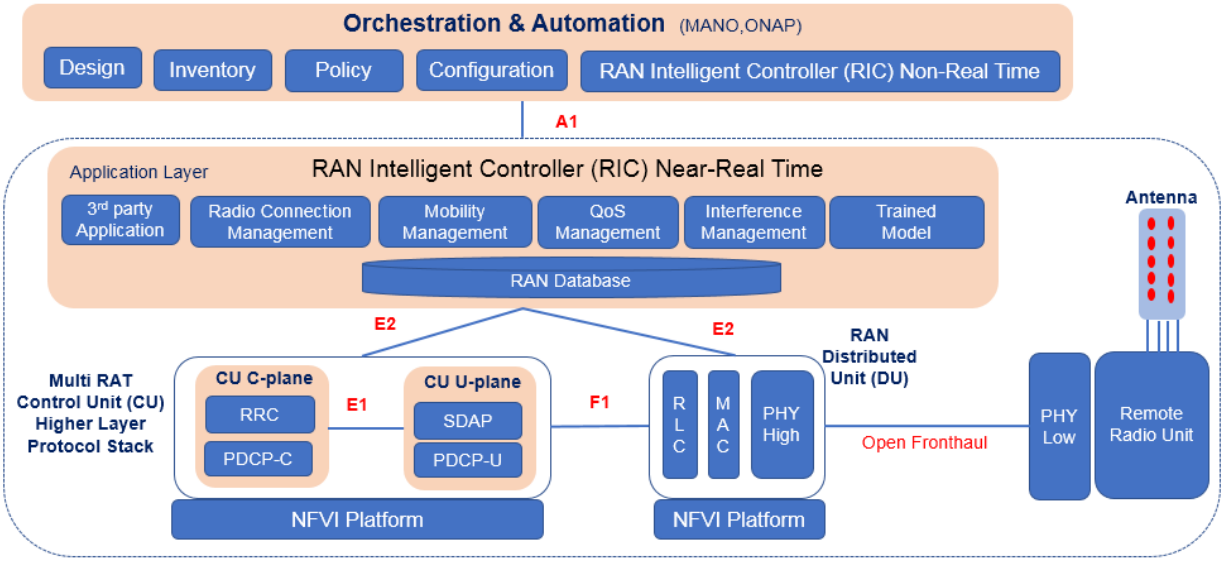

Figure 4.10: O-RAN alliance architecture framework.

5G-CLARITY system will incorporate relevant (beyond 3GPP) O-RAN architectural components, including both RAN Intelligence Controller (RIC) types: near-Real Time RIC (near-RT-RIC) and non-Real-Time RIC (non-RT-RIC). On the one hand, the near-RT-RIC provides ms-level RRM functionality with embedded intelligence, including per-UE controlled load-balancing, radio bearer management, interference detection and mitigation, QoS management, connectivity management and seamless handover control. It also delivers a robust, scalable and secure platform that allows for flexible on-boarding of 3rd party control-applications. On the other hand, non-RT-RIC provides non-time-critical functionality, including service and policy management, RAN analytics and model-training for the near-RT-RIC.

In 5G-CLARITY, near-RT-RIC is imbued in the 5G-CLARITY network function and application stratum, while non-RT-RIC is logically positioned in the 5G-CLARITY management and orchestration stratum.

Multi-WAT protocol stack

This principle is about ensuring a unified traffic processing when coming from multiple WATs, each having a different protocol stack. Different WATs have traditionally been used in a disjoint manner. Proof of this is Wi-Fi and cellular ecosystem, which have followed their own development paths. The latest version of each technology has greatly enhanced capability compared with early offerings, with the Wi-Fi 6 and 3GPP 5G technology. As society increasingly depends on fast and reliable data connectivity, an important capability for the industry is the convergence between 5G and Wi-Fi, so that unique and complementary capabilities of both WATs are leveraged to provide innovative network services. Figure 4.11 illustrates an example of this 5G-Wi-Fi convergence. New set of 5G use cases may require combined resources from both 3GPP and Wi-Fi networks in providing cost-effective solutions that meet diverse sets of requirements on throughput, latency, connection density, coverage and reliability.

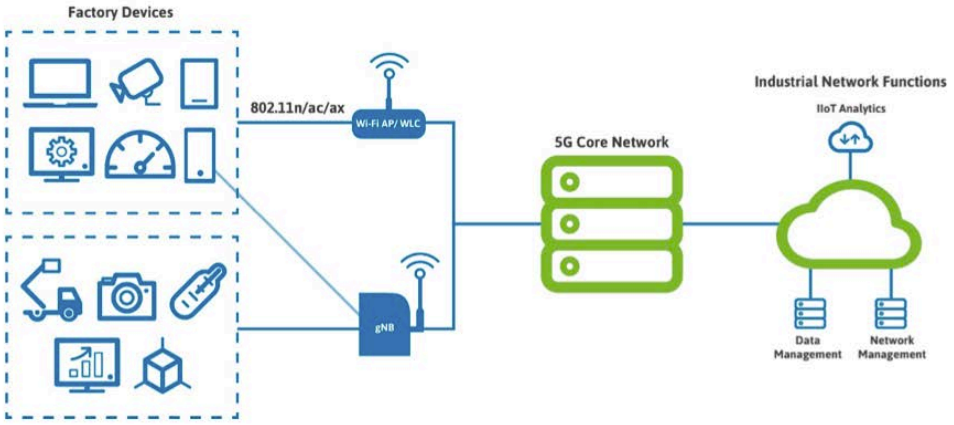

Figure 4.11: Combined use of 3GPP 5G and IEEE Wi-Fi technology [72].

5G-CLARITY will leverage the principle of multi-WAT protocol stack to design an architecture solution based on the convergence of the two WATs mentioned above (i.e. 5G and Wi-Fi), along with LiFi. The combined use of these three WATs allows for enhanced data rates (e.g. by means of bandwidth aggregation) and improved reliability (e.g. by setting up back-up sessions). To that end, advanced 3GPP mechanisms based on the use of specific VNFs will be considered in the proposed 5G-CLARITY architecture design.

4.2.3 Management and orchestration stratum

The 5G-CLARITY management and orchestration stratum encompasses all the necessary functionality to deploy and operate the different 5G-CLARITY services (and associated resources) throughout their lifetime, from their commissioning to their de-commissioning. This includes not only provisioning (i.e. instantiation, scaling, modification and other lifecycle management operations), but also monitoring activities related to performance assurance and fault supervision.

The design principles that will guide the specification of the management and orchestration stratum are detailed below.

Service-based management architecture

This principle, based on extending the SBA principles to the management plane, represents a significant leap forward in OSS digital transformation, required to handle operational complexity that B5G technology will bring for operators. This complexity resides in the need to manage and orchestrate a wide variety of services across all the network segments, with an end-to-end perspective. The specificities of these segments, with different pace of technology evolution each and with solutions from different vendors, unveils significant integration issues for operators. This is exacerbated as the number of hosted services increases, some of them with very different KPIs.

The above reasoning requires operators to transform their current Operation Support Systems (OSS), adopting novel architectural approaches that allow addressing these integration and scalability challenges. And Service Based Management Architecture (SBMA) is one of them. This architecture style means migrating from functional blocks exposing telecom-style protocol interfaces (e.g. Network Managers / Element Managers providing 3GPP Itf-N interfaces) to management services exposing APIs based on web-based technology. This change of paradigm facilitates a rapid evolution of management and orchestration capabilities in compliance with the innovation of the underlying network, by simply adding or updating APIs using libraries and other enablers (e.g. development tools, specification tools, code generators, security mechanisms) which are broadly available. This approach allows service innovation with minimal integration effort. Different SDOs have already captured the benefits of having a SBMA in their specifications. For

example, 3GPP SA5 [74] and ETSI ISG ZSM [75] define their architectural frameworks based on SBMA. Even ETSI ISG NFV, which originally chose an interface-centric approach for the design of the NFV Management and Orchestration (MANO) framework, has now decided to migrate towards a SBMA from NFV Release 4 on [76].

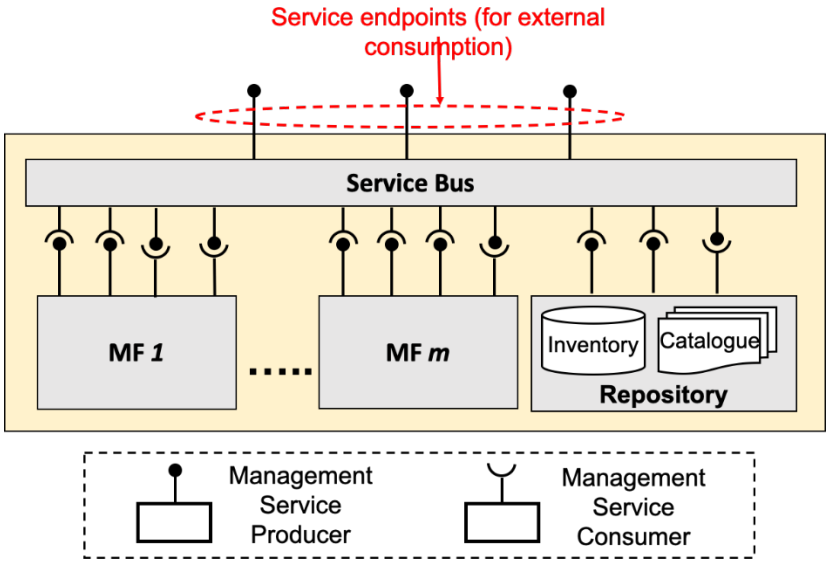

Figure 4.12: Blueprint of a baseline SBMA.

The 5G-CLARITY system will also adopt a SBMA for the design of the management and orchestration stratum, following the principles stated above. Figure 4.12 illustrates an archetypical SBMA. As it can be seen, it is formed of a set of management services which are federated together based on the definition of three novel entities:

- One or more Management Functions: A Management Function (MF) is a management entity playing the roles of management service producer and/or management service consumer. A SBMA consists of different MFs, each typically producing/consuming management services that are used to manipulate instances of the same network entity. NFV Orchestrator (NFVO) and Network Slice Management Function (NSMF) are examples of different MFs. The first is focused on the deployment and operation of instances of NFV services, while the second deals with instances of slices.
- One repository, which is a datastore that provides a single integrated catalogue and inventory for the entire SBMA.
- Finally, the service bus, which allows interoperation and communication between the MFs taking part in the SBMA, including their interaction with the repository. The functionalities of this software bus (e.g. service registration and discovery, message routing) are equivalent to the one described for the SBA. Indeed, the application/transport layer protocols (HTTP2/TCP) and serialization protocol (JSON) remain the same.

For more details on a SBMA, please see Annex C – Implementation Details on the SBMA.

Extensibility

This principle is based on augmenting the functional scope of a SBMA by adding new MFs or upgrading the existing ones. In 5G-CLARITY, this can be easily done by simply defining corresponding APIs. The integration of these APIs in the common service bus can be easily done using any of the web-based technology toolkits which are available for this end.

MF statelessness

This principle is based on designing MFs where processing (i.e. MF workload) is separated from associated data (i.e. MF state). This compute-storage separation not only allows the containerization of individual MFs, but also facilitates:

- **MF scalability**. To scale out a stateless MF, all you need to do is to create more instances out of that MF, replicating this process as much as needed. For scale-in, running instances no longer needed can be gracefully terminated once they are done with their current work.
- **MF roll-back**. In case you come across a bad deployment, stateless MF instances can be easily rolled back, as you can dismiss them and plan to launch instances of the old version.
- **MF load balancing**. This is much easier in stateless MF instances, as any instance can process any request. Apart from providing means for autonomous scaling, load balancing is also used to increase fault tolerance by applying corresponding 1:N / N:M resiliency models.

5G-CLARITY will apply the statelessness design principle on the different MFs taking part in the management and orchestration stratum.

Model-driven operation

This principle is about performing the management of services and resources through the use of information models that capture the definition of managed entities (e.g. VxFs, NFV services, transport nodes) in terms of attributes and supported operations. These models are defined in a protocol and technology neutral way (e.g. UML language), independent from the implementation of the managed network entities in order to facilitate portability, reusability and to allow vendor-neutral resource and service management.

First information models were defined in NFV ecosystem, with the definition of Network Service Descriptors (NSDs) and VNF Descriptors (VNFDs). In 3GPP, the information models for NG-RAN, 5GC and slice are defined in the Network Resource Model (NRM). For the transport network nodes, novel models for L2 and L3 service delivery (e.g. L2SM and L3SM) have been recently defined in the IETF.

5G-CLARITY plans to incorporate some of the above models for resource and service provisioning, but also for data ingestion, leveraging model-based telemetry solution. This approach allows the collection and aggregation of data from multiple sources based on the use of individual information models, one for each source. Every model declares what type of information (e.g. performance measurements, fault alarms) can be gathered from the corresponding data source, and how an authorized consumer can gain access to it.

It is worth noting that to be usable, the above information model definitions need to be mapped to a specific protocol definition, resulting in data models. In the 5G-CLARITY system, YANG may be chosen as data modelling language.

Reproducibility

It defines the ability to replicate the management and orchestration stratum across multiple sites, including public and private sites. This design principle will allow 5G-CLARITY system to create two instances of this stratum: one for private network operator, and other for the public network operator. In other words, 5G-CLARITY assumes that both types of operators have very similar strata, at least from a functional viewpoint. In terms of resource capacity, public network operator's management and orchestrator stratum will be considerably larger than the private network operator's one.

Service capability exposure

Service capability exposure is the ability to configure which management services from the SBMA can be made externally visible, so other systems can consume them via well-defined service endpoints. These service endpoints are represented in the upper part of Figure 4.12. The importance of service exposure has

been already highlighted in [77], where the GSMA claims it to be one of the three key value-added features in the network slicing value chain. The idea is even reinforced by the NGMN Alliance in [78], where secure exposure of 5G capabilities to customer and trusted third parties is discussed.

Service capability exposure is a required design principle in 5G-CLARITY system, especially for the interoperation and communication between public and private network domains. For example, in PNI-NPN scenarios, the private network operator may need to expose management services to the public network operator, allowing him to take some degree of control and management of the private resources for an effective cross-domain slice delivery.

4.2.4 Intelligence Stratum

The 5G-CLARITY intelligence stratum includes all the assets that drive automation and data-based intelligence on 5G-CLARITY service and network operation. This includes an AI engine, which deals with the execution and maintenance of deployed ML models, and an intent engine in charge of providing customer-facing abstraction functionality to facilitate the external usability of 5G-CLARITY system.

The design principles driving the development of the 5G-CLARITY intelligence stratum revolves around the flexibility and ease of use, key for a timely deployment and execution of ML models.

Cloud native and service oriented

The cloud native design principle is in line with the other 5G-CLARITY strata. It enables the modularisation of function blocks using Docker containers. In the intelligence stratum, containerised functions include intent engine, AI engine and individual deployed ML models. These ML models shall be deployed as modular and natively scalable services that are accessible through well-defined APIs, e.g. RESTful interfaces, through which they can communicate with the network to retrieve data and forward network configurations, and to the end user who is a consumer of the offered functionalities. The cloud native design at the intelligence stratum allows for:

- Enriched features of ML models, including modular model packaging (for easy pushing into RIC near-RT), smart model training (on-premise, i.e. on private site's edge cloud, and off-premise, i.e. on an external cloud site) and easy lifecycle management of ML models.
- Service exposure, for which a service registry will be used to discover what ML services are available for use.

Language/ML framework independence

The ML models created by the ML designer can come from multiple sources using different programming languages and/or libraries. In 5G-CLARITY, ML algorithms may be implemented in Python, R and MATLAB, using libraries such as PyTorch and TensorFlow (Python), nnet and RSNNS (R) or Deep Learning Toolbox (MATLAB). The ML designer shall have freedom of choice of the framework that implements the ML algorithm and it shall work in a plug-and-play fashion within the intelligence stratum.

The cloud native design enables experimentation with cutting edge algorithms that may not yet be part of state-of-the-art libraries that are provided by popular ML frameworks.

Ease of use

The intelligence stratum aims to facilitate human operation and supervision of the 5G-CLARITY system. A typical use case is the setup of a new slice, which shall be made as simple as possible for the human operator. In this use case, an Intent engine offers an abstraction of the required configurations and parameters of slice provisioning in such a way that a non-expert network/venue operator gains quick and easy access to setting up a slice without the need to know the ins and outs of slice provisioning.

5 Infrastructure Stratum Design

The 5G-CLARITY infrastructure stratum requirements which reflect the general system requirements and were identified in 5G-CLARITY D2.1 are listed in Table 5-1.

Table 5-1: 5G-CLARITY Infrastructure Stratum Requirements

Requirement ID	Requirement Description
CLARITY-INF-R1	5G-CLARITY system managed infrastructure resources are restricted to on-premises physical equipment, NFVI, etc., i.e. resources that are present/deployed within the logical perimeter of the private venue.
CLARITY-INF-R2	5G-CLARITY system managed resources include wireless resources, compute resources (i.e. computing and storage nodes) and connectivity resources (i.e. links and forwarding devices).
CLARITY-INF-R3	5G-CLARITY system managed wireless resources shall include resources from two or more wireless access technologies, including 3GPP (5GNR) and non-3GPP (Wi-Fi and LiFi) technologies.
CLARITY-INF-R4	5G-CLARITY system managed compute resources shall have in-built virtualization capabilities to allow the execution of VNF instances.
CLARITY-INF-R5	5G-CLARITY system managed connectivity resources span across different network segments, providing front-haul, mid-haul and back-haul capacity.
CLARITY-INF-R6	5G-CLARITY system managed connectivity resources shall provide QoS-assured data plane connectivity across deployed network functions, including PNFs and VNF instances.
CLARITY-INF-R7	5G-CLARITY system managed compute and connectivity resources shall be able to interact with MNO provided PLMN resources for the realization of public network integrated NPNs.

To incorporate the requirements of Table 5-1 and to support efficiently other stratums of the 5G-CLARITY framework an infrastructure stratum architecture is introduced on the Figure 5.1. The design will include:

- **User devices**, referencing to user equipment (UE) with capability for multiple Wireless Access Technologies (Multi-WAT) such as an innovative customer's premises equipment (CPE) for robotic devices and IoT devices for industry 4.0, and other standard devices such as handheld terminals or dongle cards or modems (see section 5.1).
- **Access nodes**, as building blocks of multi-WAT including APs for LiFi and Wi-Fi technologies and RUs and DUs for 4G or 5G technologies interconnected to extended edge servers forming the RAN cluster. 5G-CLARITY will deploy two types of access nodes to enable multi-WAT capabilities; (I) novel APs for LiFi and for Wi-Fi technologies, and (ii) Virtualized and splitable gNBs for 5GNR (see section 5.2).
- **Compute nodes**, including extended edge servers and edge servers hosting functions of the RAN cluster and forming the edge cluster of virtualized functions and applications of various stratums of 5G-CLARITY framework (i.e., intelligence stratum, management and orchestration stratum and network and application function stratum). Both clusters can deploy Network Function Virtualization Infrastructures (NFVIs) and deploy software and hardware accelerators required by other stratums (e.g., GPU, FPGA). The RAN cluster deploys essential networks functions for the 4G and 5G access nodes (i.e., gNB-DU, RIC, and RU) and the edge cluster creates an edge NFVI to host core functions of 5G-CLARITY stratums (e.g., 5GC, NFVO, AI engine). See section 5.3 for further details.
- **A network infrastructure**, connecting the access nodes with the RAN cluster in the mid-haul and with the edge cluster and on-site private gateway in the back-haul. The network infrastructure can enable

various communication technologies to support network slicing, different RAN technologies and services (e.g., Time Sensitive Networks (TSN)) (see Section 5.4).

- **An on-site private gateway**, to provide connectivity and reachability with external resources, e.g. PLMN resources or hyperscale's resources. The on-site private gateway will enable routing and multiple levels of traffic isolation to protect tenant's privacy and allow a reliable interconnection regardless geographic locations or coverage. An example of the usage of a private gateway is introduced in the section 5.4.3.

In the next sub-sections, we describe each components and elements of the stratum to be deployed by the 5G-CLARITY project.

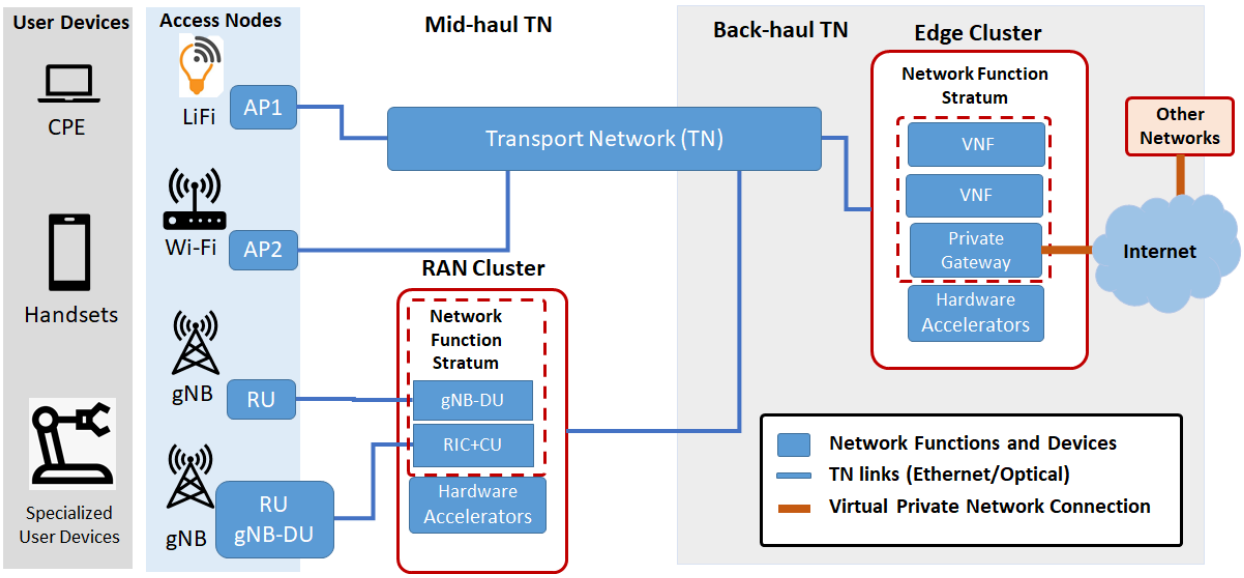

Figure 5.1: 5G-CLARITY infrastructure stratum architecture.

5.1 User devices

User devices refer to end-user equipment (UE) designed to provide network and radio connectivity to user's services and applications. There are two main groups of UEs, fixed, which includes desktop computers, fixed sensors or IoT devices and others fixed network terminals, and mobile UEs, including handsets (e.g., smart-phones, iPhones), mobile computers, tablets, mobile sensors or IoT devices (e.g., body sensors, headphones), and others. Mobile user devices (UEs) designed for LTE and 5GNR networks follow form factors and characteristics [64] aligned to 3GPP standards in terms of size, processing capability, power consumption, transmission power, antenna gain, user interface capabilities, and external interfaces. Similarly, user devices to be able to operate with others wireless technologies as Wi-Fi and LiFi follows the standards of IEEE 802 task forces to work properly.

However, Industry 4.0 is deploying massive industrial wireless sensor network (IWSN) use cases and requirements beyond uRLLC use cases included in the 3GPP user devices road maps (TR 22.804, TS 22.104, TR 22.832, and TS 22.261). As a result, to innovate in industry 4.0 and enable multi-WAT capability, Time Sensitive Communication (TSC) services, and multiple RAN technologies 5G-CLARITY will develop, deploy, and demonstrate on its infrastructure stratum three types of user devices:

1. A novel **5G-CLARITY CPE** to enable multi-WAT communication for specialized mobile systems such a robotic device by adding the benefits of CPEs in terms of promising performance in extreme scenarios and with the possibility to enable and combine new wireless technologies, I.e., 5GNR, LTE, Wi-Fi, and LiFi and more advance management capacity.

2. **Handsets or end-user mobile equipment (UE)**, with capacity to connect at least one radio access technology (5GNR, LTE, Wi-Fi, and LiFi) and enable services for mobile users.
3. **Alternatives mobile devices or adapters** to add multi-WAT capacity to any user devices without strict performance requirements.

The following sections describe further the user devices to be deployed as part of the infrastructure stratum in the 5G-CLARITY test beds.

5.1.1 5G-CLARITY CPE

5G-CLARITY considers three different WATs to bring network connectivity to the UE, i.e., 5GNR, Wi-Fi, and LiFi. Since each of these WATs makes use of a different wireless frequency and many specialized user devices (e.g., moving robotic devices) require adaptability to changes in the radio conditions a`s well as performance which is not possible with access to a single access node and frequency or spectrum. The study and development of novel solutions to allow specialized user devices to get the benefits from multi-WAT deployments are essential for new 5G services and verticals. One promising solution is the design on an 5G-CLARITY CPE embedding multiple technology-specific wireless interfaces for Wi-Fi, LiFi, and 5GNR with two main functionalities, First, with the ability to switch and route the traffic between different WATs, either manually or automatically (e.g. based on predefined algorithms), second with the ability to aggregate all traffic routes passes through different WATs into the specialized device. For example, by using the predefined algorithms, the CPE can automatically switch from Wi-Fi to the LiFi if the availability of LiFi light sources can provide better performance and coverage, or switch to 5GNR based on the received signal level and/or low congestion.

An example of the application of 5G-CLARITY CPE is presented in Figure 5.2, a humanoid robot (part of one of the 5G-CLARITY use cases) which need multiple WAT to adapt to the changes in traffic and signal conditions. Because It Is expected that the humanoid robot will move across different locations to interact with the people by demanding permanent high throughput and low latency network connectivity to remain operational.

Based on the above considerations, the 5G-CLARITY CPE will be attached to the humanoid robot and would act as part of it to bring network connectivity to the robot. The connection between the CPE and the robot would be a 1 Gbps Ethernet link, and it would be enough to even passing the aggregated traffic between all the WAT RATs to the robot and to near UEs. In this scenario, the CPE location is not fixed and would change by the robot movement, so the CPE can switch to different RAT based on the algorithms mentioned above or predefined scenarios. Figure 5.2 shows the connectivity of the robot to the network using the CPE.

There are also several IP cameras in the use case to track and report the accurate robot location. These cameras send the raw images (which require high throughput link) to the edge application and the output of the application which is the accurate location of the robot will send back to the robot for path tracking and precise location purposes (required low latency communication). To enable these IP cameras to access the network using multiple WATs the CPE will stablish a multiple communication in parallel deciding what WAT to use based on the location to preserve the synchronization and performance required by the system to be operations.

In this example, the 5G-CLARITY CPE will be equipped with a USB dongle/M.2 socket 5G modem to access 5G networks, an LiFi USB dongle for LiFi connectivity, and with a Wi-Fi module for Wi-Fi 5 and/or Wi-Fi 6 (2.4 GHz and 5 GHz). The connection between the CPE and specialized device will be a standard 1 Gbps Ethernet link. More details of the humanoid robot will be introduced in next deliverables.

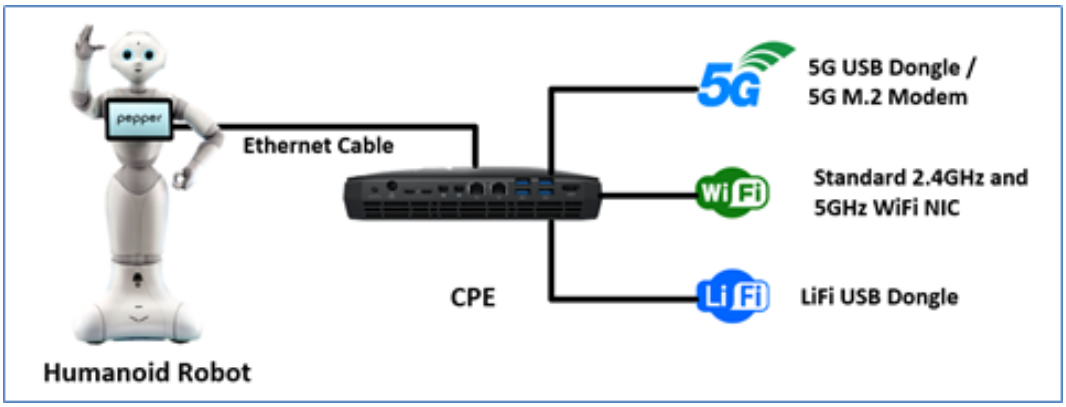

Figure 5.2: CPE and Humanoid Robot integration.

5.1.2 Handheld or handset terminals

To introduce the handheld or handset terminals to be used in 5G-CLARITY infrastructure stratum we briefly revise the Figure 5.3, which presents the roadmap for user devices for 3GPP Release 15 and beyond. Most of the chipset and devices produced in 2020 does not support TSC and Massive IoT (mIoT), which are essential for Industry 4.0. Indeed, TSC must be supported in the same network besides eMBB, mMTC, and uRLLC use cases. As a result, in 5G-CLARITY we study the integration of TSC and mIoT for industry 4.0.

Following the Figure 5.3 trend uRLLC features were introduced in 3GPP R15 for both LTE and 5GNR, and 5GNR uRLLC is further enhanced in 3GPP R16 for Industrial IoT work items by introducing support for TSC together with 5G integration. The 3GPP R17 and beyond 5G UE can be considered to support variety of use cases from industrial wireless sensor networks to 5GNR connected to smart city video cameras. It is also expected that eMBB, mMTC, uRLLC and TSC use cases may all need to be supported in the same network [66]. With regards to the use of 5GNR UEs there are currently handheld equipment and wireless cards supporting sub-6GHz frequency bands with some of them including mmWave frequency bands. The status of form-factors for NSA and SA models are discussed in [77]. Hence, 5G-CLARITY will support UE complying with 3GPP R15 or R16 in SA mode as well as Wi-Fi and LiFi. To demonstrate this flexibility with multi-WAT the project will implement a CPE besides handsets installed in a mobile robotic device.

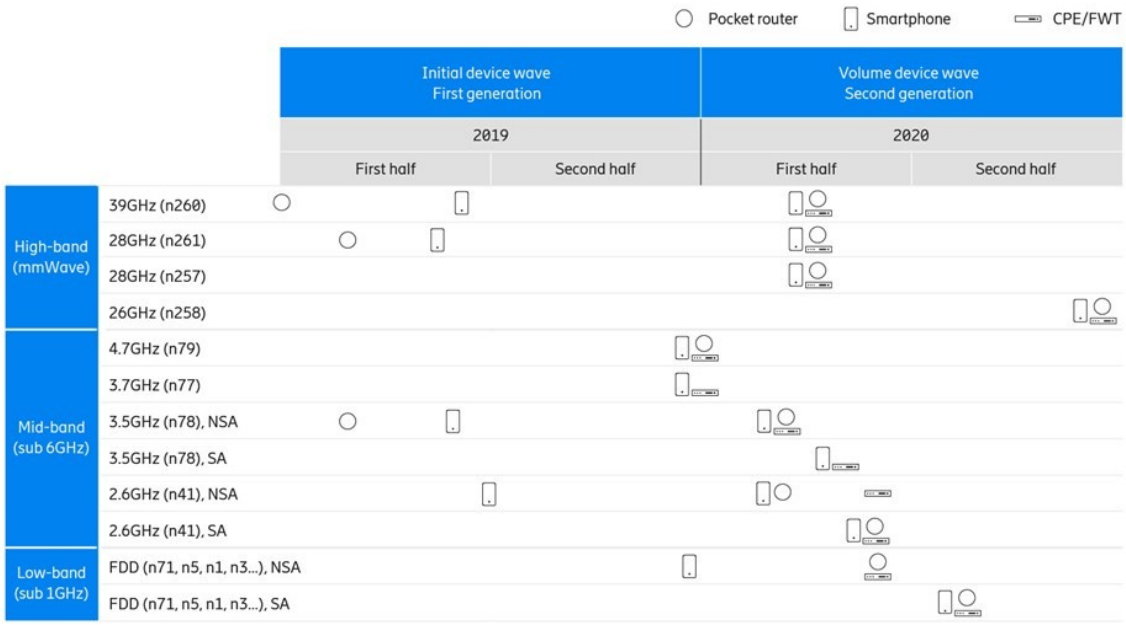

Figure 5.3: 5G device availability [80].

5.1.3 USB dongles/wireless cards

USB dongles or M.2 PCI express cards are small form factor UEs without user display capabilities that can be used either integrated (via PCI express interface into main board) or externally plugged (via USB connector) into computing devices (e.g. desktops, laptops, etc.). The M.2 form factor cards (see an example of commercial M.2 card in Figure 5.4) can be used together with a M.2 enclosure with antennas, Universal Integrated Circuit Card (UICC) slot and USB 3.0 connector interface. There is an additional commercial 5G module with Land Grid Array (LGA) form factor. For practical purposes only the M.2 card scan be considered as candidates for the 5G-CLARITY CPE. It is expected that the market will soon also have USB dongles supporting 5GNR.

Figure 5.4: Sierra Wireless EM9199 M.2 Card.

5.2 Access nodes

5.2.1 5GNR nodes

The NG-RAN decomposition and HLS/LLS options were discussed in Section 4.2.2. The RU will always correspond to the 5G physical access node and will therefore be deployed as PNF in the 5G-CLARITY infrastructure stratum.

With regards to gNB-DU and gNB-CU, different variants can exist in the 5G ecosystem according to Figure 5.5:

1. Both gNB-DU and gNB-CU are executed atop the NFVI's RAN resource zone but deployed separately. In this scenario, gNB-DU and gNB-CU can be modelled as (monolithic) VNFs.
2. Both gNB-DU and gNB-CU executed atop the NFVI's RAN resource zone but deployed together as a single (monolithic) VNF. The co-location described in this scenario corresponds to classical C-RAN, with this VNF providing base band unit BBU functionality.
3. gNB-DU combined with the RU, and gNB-CU running on the NFVI's RAN resource zone. In this scenario, gNB-CU is deployed as a (monolithic) VNF, and the combined gNB-DU and RU are integrated into a single PNF
4. Fully integrated gNB with CU, DU and RU combined functions running as a single PNF (traditional fully integrated small cell with NG backhaul interface)

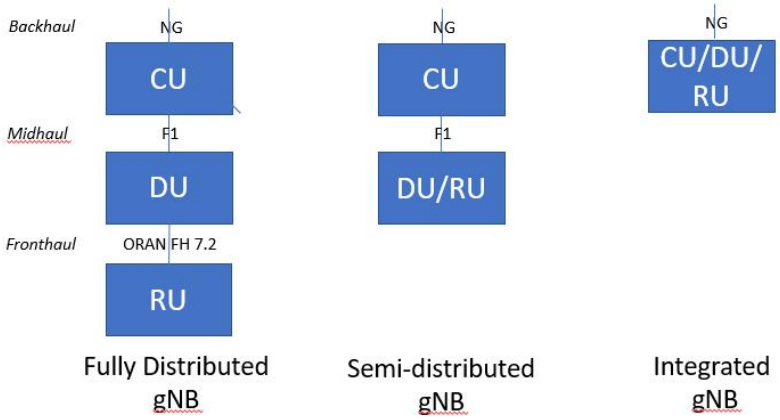

Figure 5.5: gNB deployment options.

For 5G-CLARITY either variants 1/2 or variant 3 will be used. In general variants 1/2 are more suitable for big deployments (high number of RUs) where scalability is achieved via hierarchical topologies where a gNB-CU can control a certain number of gNB-DUs (ex. 8 gNB-DUs per gNB-CU) with each gNB-DU controlling a certain number of RUs (ex. 8 RUs per gNB-DU) which at the same time support a certain number of RUs depending on the computing and transport characteristics. For smaller deployments (low number of RUs) a simplified approach based on variant 3 would be sufficient and less complex to deploy.

5.2.2 5G-CLARITY Wi-Fi nodes

Within the context of 5G-CLARITY we consider a Wi-Fi node to be a network function with the components depicted in Figure 5.6. These components are:

- A host platform with an ARM based processor driven by a Linux based OS.
- One or more COTS wireless modules, including:
    - PHY technologies operating at the 2.4 GHz or 5GHz bands defined by one of the following technologies: IEEE 802.11n, IEEE 802.11ac or IEEE 802.11ax. The interested reader is referred to D3.1 [2] for a detailed description of these technologies
    - An AP MAC layer function defined in IEEE 802.11m, characterised by:
        - A contention-based channel access function known as Enhanced Distributed Channel Access (EDCA)
        - A Basic Service Set (BSS) defined by an Access Point function and the connected Wi-Fi stations identified by a SSID.
    - Each wireless module is controllable by software allowing to: *i)* select the operating channel, and *ii)* allowing to instantiate multiple AP MAC layer functions concurrently operating on the same module but advertising different services.
- One or more Ethernet modules, which will serve three main functions:
    - Serving as access interface for other 5G-CLARITY radio nodes, for example LiFi nodes
    - Connecting Wi-Fi nodes with each other forming an extended backhaul network
    - Connecting the Wi-Fi nodes to the Ethernet network segment inside the private network.

Armed with the previous components, the 5G-CLARITY Wi-Fi node goes beyond offering only radio access functions. Indeed, connected through the Ethernet interfaces the 5G-CLARITY Wi-Fi nodes can form a layer 2 (L2) network that can be used to provide access to additional devices like the LiFi AP, thus forming a tightly integrated Wi-Fi and LiFi network. A control plane deciding how to forward packets across this backhaul is required, which is why we consider the inclusion of an SDN agent running the host in Figure 5.6.

In 5G-CLARITY WP3 a control plane for this L2 network will be developed, which will be part of the network function stratum. This control plane will decide how traffic flows within the L2 network to reach the appropriate Wi-Fi or LiFi access point when communicating with a given customer. Figure 5.7 depicts a logical model for the Wi-Fi/SDN switch where we can distinguish access interfaces, which can be Wi-Fi or Ethernet, and backhaul interfaces controlled by the SDN agent, which can also be wireless and wired.

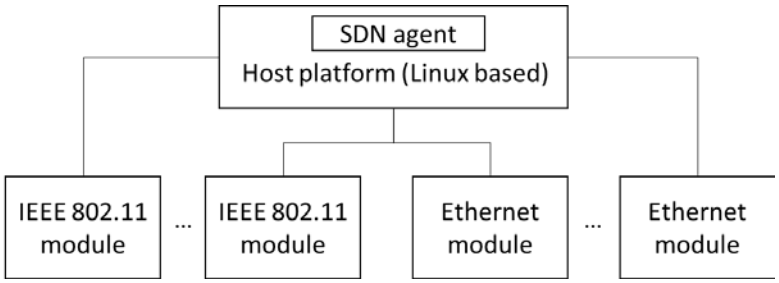

Figure 5.6: Components of a 5G-CLARITY Wi-Fi AP

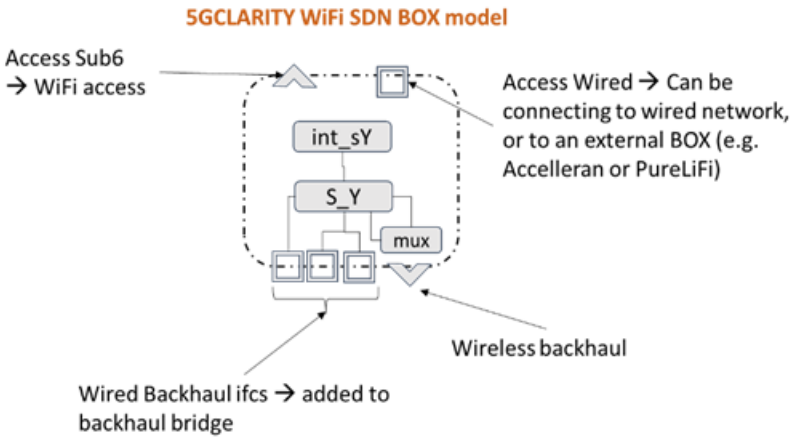

Figure 5.7: 5G-CLARITY integrated Wi-Fi SDN box model.

5.2.3 5G-CLARITY LiFi nodes

Within the context of 5G-CLARITY we consider a LiFi node to be a set of devices that could convert and deliver data packets in the form of light. These devices are:

- LiFi-AP, which includes:
  - A Physical layer implementation based on 802.11 OFDM PHY. The interested reader is referred to deliverable D3.1 [2] for a detailed description of these technologies
  - Digital-to-analogue and analogue-to-digital convertors which convert the digital signal to analogue signal for downlink transmission, and covert the received analogue signal to digital signal for uplink decoding.
  - A MAC layer interface between the PHY and the upper layers, with functions defined in IEEE 802.11. The implemented MAC could be modified to provide full-duplex operation, high protocol efficiency and multiuser support.
- Transmitter driver, for driving the light luminaire.
- LiFi transmitter, which acts as both luminaire for illumination and antenna for optical wireless signals. It is usually an LED lamp for general use cases, and it can be within either visible light spectrum or infrared spectrum.

The LiFi node system structure is shown in Figure 5.8(a) and Figure 5.8(b) shows an example LiFi AP device from the LiFi-XC product. Figure 5.8(c) shows a logical model for the LiFi node. As introduced in previous section, the LiFi node can be connected to the 5G-CLARITY Wi-Fi node shown in Figure 5.8(c), thus forming a tightly integrated Wi-Fi/LiFi network.

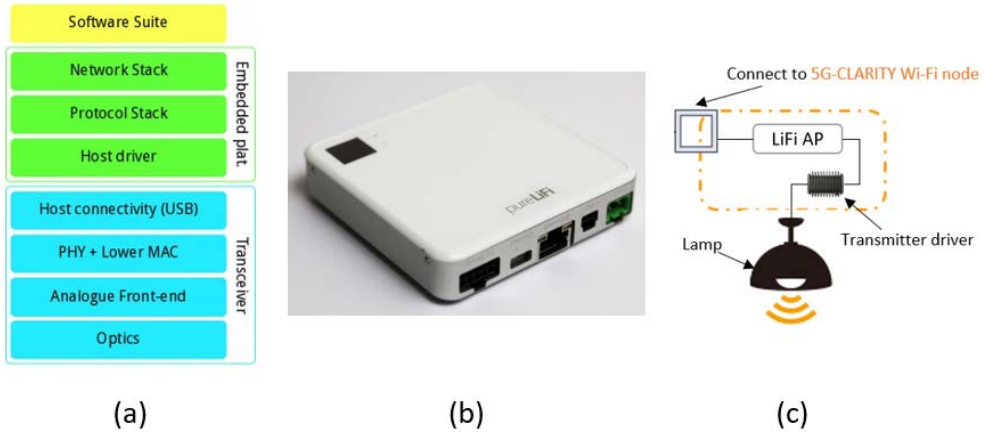

Figure 5.8: (a) LiFi node system structure; (b) LiFi-XC AP; (c) LiFi node logical model.

5.3 Compute nodes

One of the key design principles of 5G-CLARITY infrastructure stratum is the clustering of NFVIs introduced in section 4.2.1 to allow separate execution environments within the same NFVI. Two clusters of NFVIs are planned in the infrastructure stratum: RAN cluster and edge cluster. In this section, we detail the main differences between compute nodes from RAN cluster and from edge cluster.

The 5G-CLARITY-RAN cluster is an extension of the NFVI formed by a set of compute nodes near by the access nodes (I.e., gNB-RU or RU) for the proper deployment of RAN functions or eventually VNFs. This will allow the deployment of:

- gNB-DUs and gNB-CU functions as VNFs to meet the strict latency constraints imposed between RAN functions.
- 5GNR or LTE core network functions (5GC/vEPC) for local gateway breakouts near user devices for delay sensitive applications.
- Hardware accelerators (e.g., GPU, FPGA) to support compute and delay sensitive services (e.g., dynamic spectrum sharing and computing intensive radio optimisation algorithms).

A combination of the mentioned deployments converts 5G-CLARITY-RAN cluster as essential components of the infrastructure stratum to deliver industry 4.0 use cases. Figure 5.9(a)(b) and (c) present an example of three possible 5GNR RAN deployments with the 5G-CLARITY-RAN cluster.

The 5G-CLARITY edge cluster is defined as the largest NFVI between both clusters to instantiate VNFs requiring larger compute and storge resources but with less delay constraints (e.g., 5G/LTE core network functions (5GC/vEPC)). Like the 5G-CLARITY RAN cluster, it will include hardware accelerators, a private gateway, and software enablers. Figure 5.9(d)(e) shows two types of 5GC deployment with services in combination with any 5GNR RAN split scheme supported by the 5G-CLARITY RAN cluster. In addition, to support Industry 4.0 and other critical use cases, 5G-CLARITY clusters will be designed with the following criteria:

- COTS design criteria to have the capability to meet strict space constraints as well as to operate in dusty, less well maintained, and less temperature-regulated environment, commonly fine in most industry.

- Enable *hardware accelerators* as specialized hardware (such as FPGAs) to perform some function faster than executing the same function on a general-purpose central processing unit (CPU) or on a traditional networking (or other I/O) device (such as network interface controller (NIC), switch, storage controller, etc.).
- Enable *software acceleration* such as Data Plane Development Kit (DPDK) providing a set of one or more optional software layers that are selectively added within elements of an NFV deployment (e.g. Compute, Hypervisor, VNF, etc.) to bring improved capabilities (e.g. increased network throughput, reduced operating overhead) which result in measurable improvements over standard, un-accelerated implementations.
- Use *control plane* specific hardware including X86 CPU with virtualization capabilities, high volume of RAM and specialist FPGA boards Data plane specific hardware including Smart NIC cards with IPSEC Specific hardware for security protocols, and Quantum key distribution systems, high performance Disks, and GPU processing cards to support ML/AI applications and Graphic rendering services such as AR/VR.

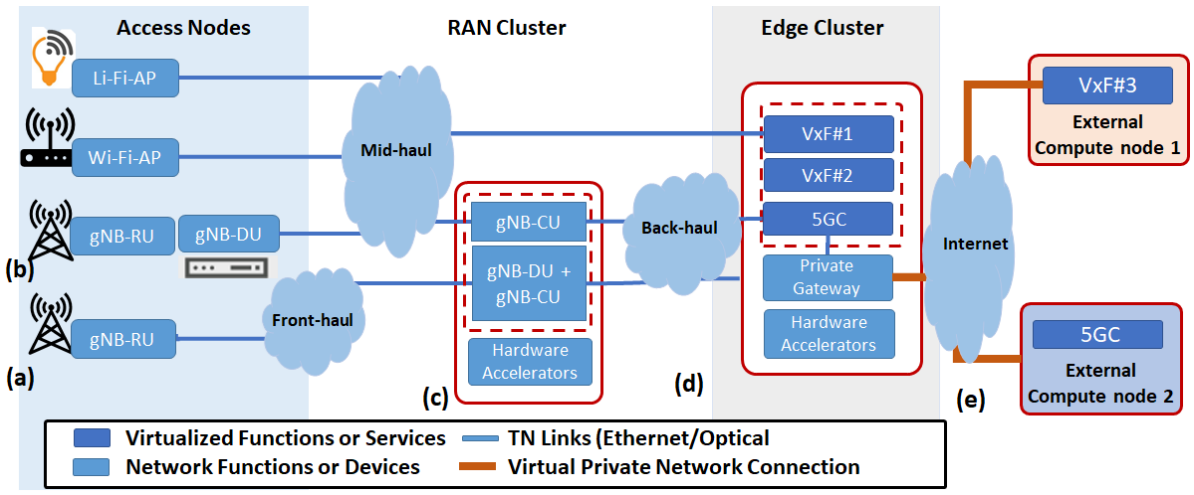

Figure 5.9: **Example of 5G-CLARITY clusters architecture supporting two types of 5GNR RAN splitting options (a), (b) and (c) and 5GC splitting and options (d) and (e).**

Figure 5.9 presents an example of how 5G-CLARITY infrastructure stratum can implements two types of RAN function splitting, virtualization, and distribution. The example of Figure 5.9(a)(c) presents the lower layer RAN functional split in which the RU relies on neutrality and proximity of the RAN cluster for efficient interchange of delay constrained physical layer traffic (e.g., CPRI) with the upper layer functions gNB-DU and gNB-CU. Then the example of Figure 5.9 (b)(c) shows how the RAN cluster can host the gNB-CU function to provide mid-level split while gNB-DU is working in a RU or in a specialized equipment connecting to single or multiple RUs. In both cases the RAN cluster provides benefits for these two types of RAN function splitting and also allows the virtualization of upper layer RAN functions as well as the execution of VxFs near the users for ultra-low latency services and capacity by enabling hardware and software accelerators to increase the performance for large RAN deployments Figure 5.9(d)(e) presents two 5GC deployment and three services, one 5GG in edge cluster to create a local gateway breakdown for services at the edge and 5GC in an external compute node. In similar way services can benefit of clusters.

5.4 Network infrastructure

In previous sections we introduced the user devices and compute nodes to be deployed by 5G-CLARITY

infrastructure stratum and how tenants can deploy a diverse type of services and RAN and core implementations. The network infrastructure is formed by three main parts: (I) Transport Network (TN) connecting the segments of mid-haul and back-haul and enabling SDN/IP-links and layer 2 (L2) switches and nodes, (II) Time sensitive network (TSN) segments in the front-haul formed by L2 switches, and nodes supervised by a TSN controller, and (iii) 5G-CLARITY - private gateway deployed as virtualized and/or bare metal function or routing device as part of the edge cluster.

In this section we describe the three parts of the network infrastructure that would make possible the optimal communication between Multi-WAT access nodes, the 5G-CLARITY RAN cluster, the 5G-CLARITY edge cluster, and remote compute nodes and networks as introduced on the Figure 5.1 and the Figure 5.9, Figure 5.10 illustrates the overall network architecture formed by TSN and SDN enabled transport network technologies, three essential network and compute node controllers from 5G-CLARITY M&O stratum and finally the 5G-CLARITY private gateway. The Figure 5.10(a) presents an example of TSN nodes and L2 switches deployed to extend the 5GNR/LTE front-haul. Then, the Figure 5.10(b) shows an example of Multi-WAT mid-haul TN connectivity trough SDN enabled L2 switches. And finally Figure 5.10(c), (d), and (e) present the control plane implemented by 5G-CLARITY M&O stratum for TSN and SDN enabled elements as well as distributed the NFVIs (i.e., 5G-CLARITY RAN cluster and edge cluster) forming a VIM domain. In the next sub sections, we extend the description the all concepts introduced so far with an example of NPN Implementation using 5G-CLARITY - infrastructure stratum architecture.

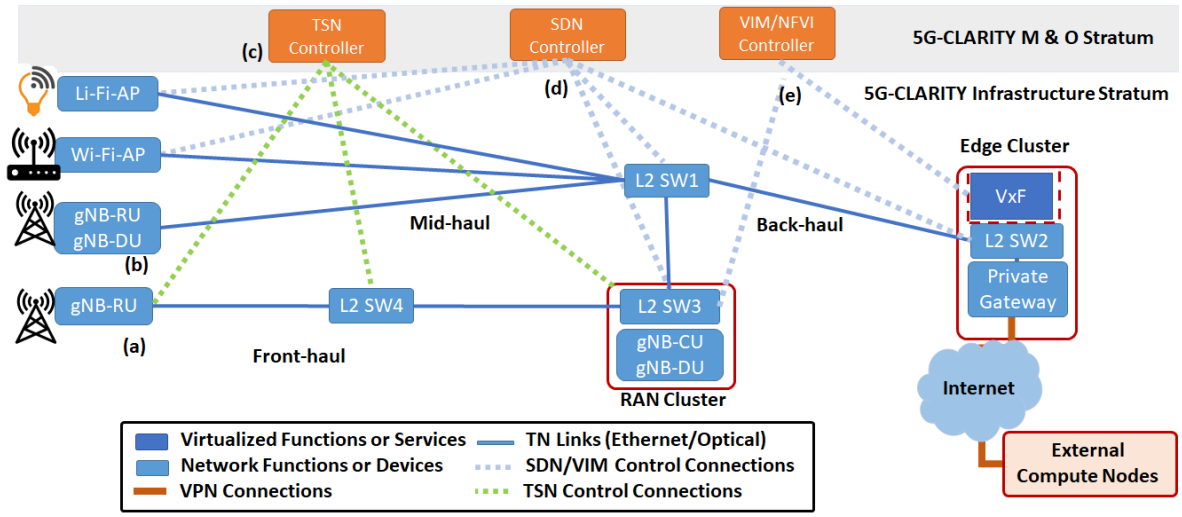

Figure 5.10: **Example 5G-CLARITY network infrastructure**

5.4.1 Transport network

The transport network (TN) covers two key network segments of the infrastructure, the mid and the back-haul of the RAN (Figure 5.10(b)). To support multiple tenants and services, the TN must offer optimal performance, flexibility, and traffic isolation (i.e., 5G-CLARITY slicing) by enabling VLAN, SDN, and NETCONF protocols in Layer 2 (L2) links and switches. Each segment has its characteristics.

The mid-haul interconnects access nodes with the RAN clusters through L2 switches and various transmission technologies to support multiple RAN technologies, e.g., Ethernet of 1 and 10 Gbps, Wi-Fi and LiFi (see Figure 5.1). The backhaul network segment can interconnect multiple mid-hauls with one edge cluster and with the private gateway through L2 switches and 10 Gbps Ethernet links for traffic aggregation. The usage of Virtual LANs allows each tenant to encapsulate E2E traffic in a set of VLAN-IDs (Tags) and to assign different IP addresses or IP networks or sub IP networks (subnets) with various QoS or bandwidth.

Depending on the deployment and availability of technology, both TN segments can combine optical links with Ethernet and wireless technologies to provide flexibility, scalability, resiliency, and cost-efficiency. Enabling multiple VLAN tagging and SDN technology in TN segments will allow traffic isolation and programmability, essential for the 5G E2E slicing and other, 5G-CLARITY solutions and use cases. SDN switches would support OpenFlow protocol or other means of service flow control protocols such as Open Daylight and Floodlight. Another control plane protocol considered in the TN architecture is NETCONF which can be deployed with SDN to combine user devices with IoT devices in the control plane.

5.4.2 TSN nodes

TSN is a converged L2 technology able to provide paths with deterministic QoS in terms of latency, jitter, frame loss, and reliability. Unlike standard Ethernet networks, TSN can convey critical streams with stringent performance constraints, while enabling their coexistence with non-performance sensitive traffic. A TSN stream is a unidirectional unicast or multicast data flow uniquely identified within a TSN network. Figure Figure 5.11 shows the primary TSN features to enable the support of ultra-reliable low-latency (uRLLC) streams together with the IEEE 802.1 amendments in which they are specified.

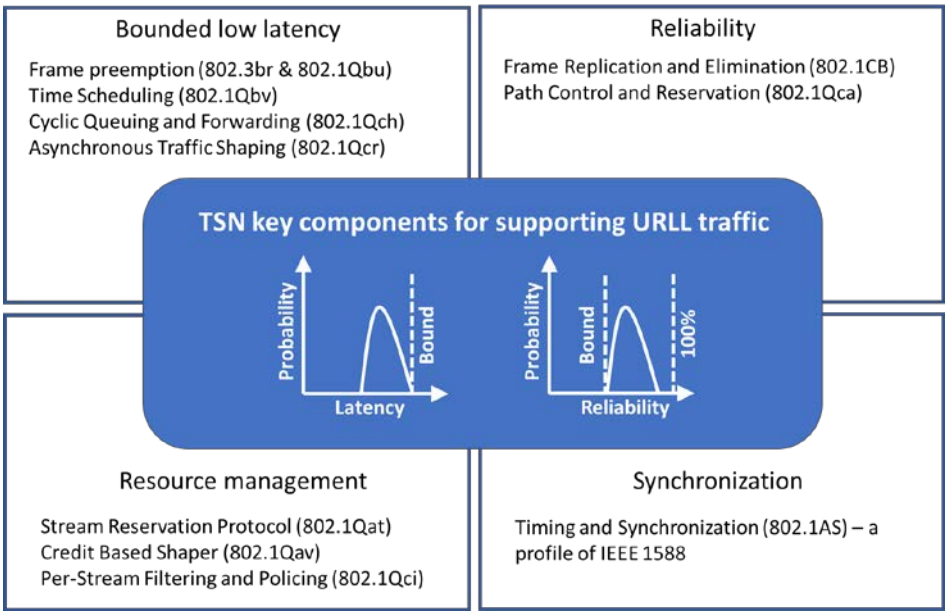

Figure 5.11: TSN key components for supporting critical flows[82].

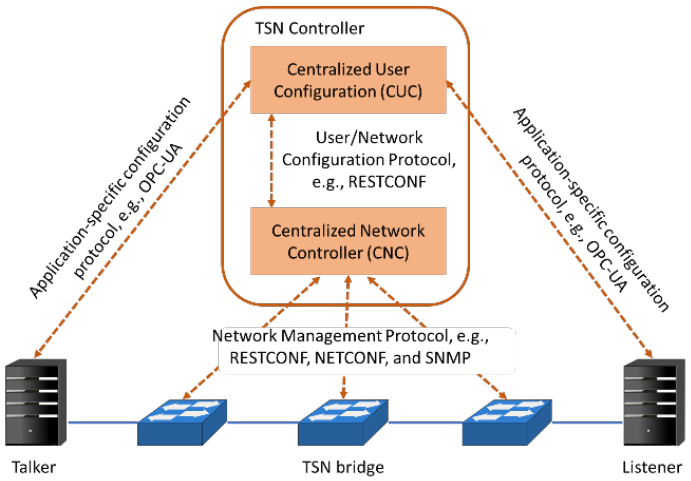

Figure 5.12: TSN fully centralized architecture.

In the context of 5G-CLARITY system, TSN is regarded as the L2 technology to provide QoS-assured connectivity across deployed network functions (refer to 5G-CLARITY infrastructure requirement CLARITY-INFRA-R6 in Table). 5G-CLARITY system allows gNB-DUs deployed as VNFs at the RAN cluster. In this scenario, TSN technology addresses the stringent performance requirements (one-way delays around 100 us) of the fronthaul network, i.e., the transport network segment that provides connectivity between RUs and gNB-DUs. In fact, IEEE TSN Task Group has defined the IEEE 802.1CM standard that specifies the use of TSN for fronthaul to enable the deterministic transport of 4G CPRI and 5G eCPRI streams. In the same way, TSN serves to realize deterministic mid-haul and back-haul networks for conveying the uRLLC streams, thus enabling 5G-CLARITY system to provide an end-to-end QoS-assured.

Figure 5.12 illustrates the TSN fully centralized architecture, which adopts the SDN principles. The control plane consists of the following two components:

- Centralized Network Controller (CNC): Application provided by the vendor that is responsible for determining the required TSN bridges configuration to ensure the QoS requirements of the incoming and ongoing flows. If the CNC finds a feasible configuration for the incoming flow, it will apply that configuration to the TSN bridges through a network management protocol. Otherwise, the incoming flow will be rejected.
- Centralized User Configuration (CUC): Vendor-specific application in charge of gathering and composing the stream requirements and communicate them to the CNC via a user/network configuration protocol. Besides the communication requests with specific requirements, the CUC might issue other queries to the CNC like the TSN network's physical topology discovery.

Regarding the TSN data plane, we can distinguish the following three elements:

- Talker: The end-station that acts as the source of a stream.
- Listener: The end-station that is the destination of a stream.
- TSN node or bridge: Network device within a TSN enabled network that conforms to the mandatory or optional features defined in TSN standards [78] [79].
- Last, it is worth mentioning that TSN data plane can be directly integrated and work with non-TSN Ethernet, but deterministic QoS can only be ensured inside the TSN network. In this way, we might have hybrid transport networks combining TSN and SDN-enabled standard Ethernet data planes. The TSN network could be primarily used for transporting URLL streams, whereas standard Ethernet for non-performance sensitive traffic. The TSN and SDN controllers might remain fully segregated. If it is required that part of the best-effort traffic traverses both the TSN and standard network, there might be applications functioning as virtual listeners/talkers running on top the SDN controller for issuing the required requests to the TSN controller.

5.4.3 Private site gateway

In 5G-CLARITY system, the private site gateway is responsible for providing L2/L3 connectivity to external data networks, typically PLMNs. This connectivity allows the private site to gain access to the PLMN, either to consume public mobile network services, or to reach other private sites (thereby enabling multi-site deployments). Figure 5.13 presents a deployment scenario where both options are covered. As it can be seen, VPNs are used for the provision of required L2/L3 connectivity services.

As can be expected, there are many different categories of the data that can be conveyed through the established connectivity. This includes user data, service control data and network management data. A 5G-CLARITY private site gateway must consider the security and policy control associated with individual data streams, to ensure their constituent flows, behave as expected, in terms of data privacy and QoS.

In the example of Figure 5.13 we present two tenants sharing industrial premises in Barcelona, in which the multi-WAT is used by three robots and the gateway breakdown is done in the Figure 5.13(c) by hosting the 5GC and VxFs of both tenants to minimize the delay in the control. Then the private gateway, Figure 5.13(d) allows the customers of both slices to interact with the industrial facility. In this way we see the advantage of 5G-CLARITY clusters. Next section we describe more in details the network function and application stratum.

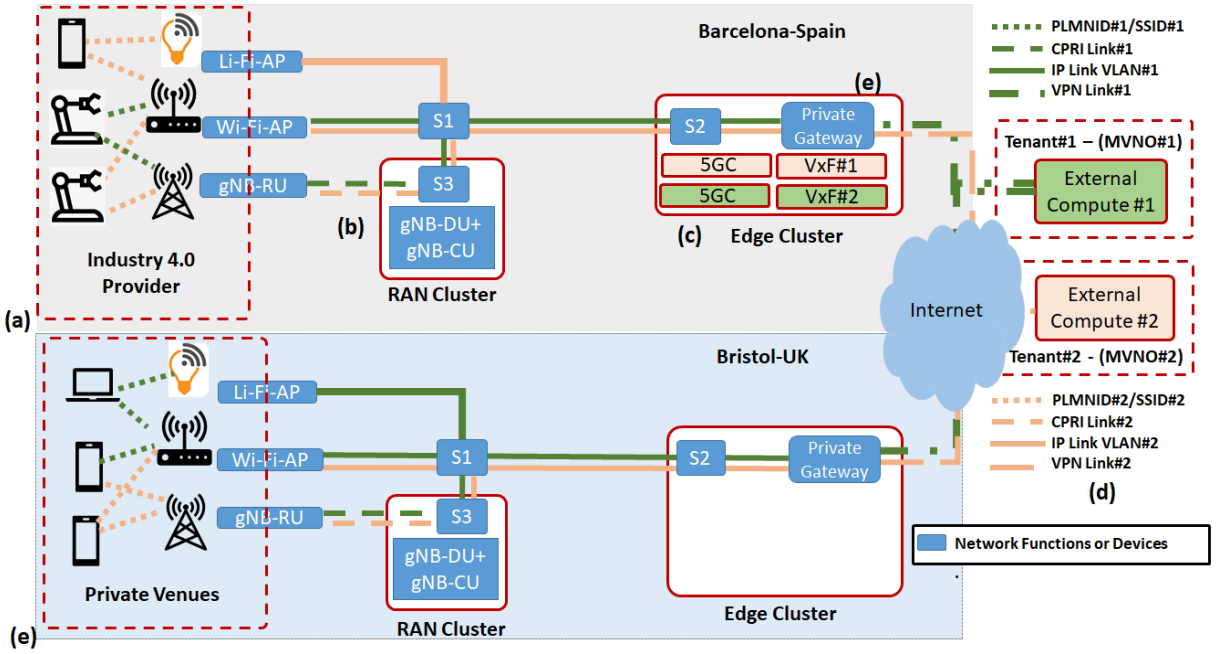

Figure 5.13: Example of two NVNOs or tenants.

6 Network Function and Application Stratum Design

This section describes the network function and application stratum of 5G-CLARITY. Before describing the details of the network function components, the 5G-CLARITY network function stratum requirements which reflect the general system requirements given in Chapter 4 and task-specific requirements given in 5G-CLARITY D3.1 [2] are listed in Table 6-1.

Table 6-1: 5G-CLARITY Network Function and Application Stratum Requirements.

Requirement ID	Requirement Description
CLARITY-NFAS-R1	The 5G-CLARITY network function and application stratum shall include necessary VxFs for the provisioning of E2E services. These VxFs provide access network, core network and multi-connectivity functionality for the entire 5G-CLARITY system.
CLARITY-NFAS-R2	The 5G-CLARITY network function and application stratum shall be managed by the 5G-CLARITY management and orchestration stratum.
CLARITY-NFAS-R3	The 5G-CLARITY network and application function stratum shall support a multi-WAT network based on the combination of 3GPP and non-3GPP technologies.
CLARITY-NFAS-R1	The 5G-CLARITY network function and application stratum shall support a converged, stand-alone core network that is able to process traffic from both 3GPP and non-3GPP access networks.
CLARITY-NFAS-R5	The 5G-CLARITY network function and application stratum shall include NG-RAN VNFs for 5G mobile access network. This access network is referred to the 3GPP access network.
CLARITY-NFAS-R6	The 5G-CLARITY network function and application stratum shall include IEEE 802.11 VNFs for Wi-Fi and LiFi access networks. These individual networks are referred to non-3GPP access networks.
CLARITY-NFAS-R7	The 5G-CLARITY network function and application stratum shall include VNFs providing means for the integration of Wi-Fi and LiFi technologies into a single unified non-3GPP network.
CLARITY-NFAS-R8	The 5G-CLARITY network function and application stratum shall include 5GC VNFs for the converged core network.
CLARITY-NFAS-R9	The 5G-CLARITY network function and application stratum shall include VNFs providing means for multi-access connectivity support for 5G, Wi-Fi and LiFi access technologies.
CLARITY-NFAS-R10	The 5G-CLARITY network function and application stratum shall include VNFs providing means for the integration of non-3GPP access network to the converged core network.
CLARITY-NFAS-R11	The 5G-CLARITY network function and application stratum shall support RAN functional splitting options 2 for high-layer (gNB-CU and gNB-DU) splitting and 7-2 for low-layer (gNB-DU and RU) splitting.
CLARITY-NFAS-R12	The 5G-CLARITY network function and application stratum shall incorporate O-RAN reference architecture and interfaces.
CLARITY-NFAS-R13	The 5G-CLARITY network function and application stratum shall support the use of existing O-RAN interfaces to *i)* provision/configure network functions; *ii)* provide policies to network functions; and *iii)* receive telemetry from network functions.
CLARITY-NFAS-R14	The 5G-CLARITY network function and application stratum shall define new interfaces to *i)* provision/configure network functions; *ii)* provide policies to network functions; *iii)* receive telemetry from network functions; and *iv)* provide the gathered telemetry to another stratum such as the 5G-CLARITY management and orchestration stratum.

CLARITY-NFAS-R15	The 5G-CLARITY network function and application stratum shall support effective usage of available access networks by access traffic steering, splitting and switching.
CLARITY-NFAS-R16	The 5G-CLARITY network function and application stratum shall decouple downlink and uplink transmissions and shall have the capability to schedule downlink and uplink traffic to different WATs.
CLARITY-NFAS-R17	The 5G-CLARITY network function and application stratum shall allow non-real time and real-time modifications for access traffic steering, splitting and switching for both downlink and uplink transmissions by leveraging O-RAN architectural components.
CLARITY-NFAS-R18	The 5G-CLARITY network function and application stratum shall allow controlling physical resources of 5GNR gNBs, Wi-Fi and LiFi APs.
CLARITY-NFAS-R19	The 5G-CLARITY network function and application stratum shall support hosting xApps to provide value added services such as spectrum access system, localization server, real-time access traffic controller, integrated Wi-Fi/LiFi network controller, etc.
CLARITY-NFAS-R20	The 5G-CLARITY network function and application stratum shall provide necessary telemetry data to the hosted xApps.

This 5G-CLARITY stratum adopts the user and control plane separation principle of 3GPP and leverages O-RAN reference architecture, along with its interfaces. 5G-CLARITY network and application function stratum also adopts the use of 5GC in 5G Stand-alone (SA) mode. The rationale behind using 5G SA is that the private spectrum licenses in EU mostly cover n78 and n77 bands. These bands are auctioned/reframed for pure 5G use that is without the need to anchor to a 4G band [2], and can enable higher throughput and lower latencies than with NSA mode 4G anchor. In addition to that, another reason for using 5G SA is to provide *i)* a seamless integration of 5G, Wi-Fi and LiFi WATs; and *ii)* multi-connectivity by employing 5G-specific functions such as TNGF/N3IWF and AT3S.

Figure 6.1 shows the overall network function and application stratum design of 5G-CLARITY. This stratum is composed of VxFs from the three networking planes, including:

- User plane functions (Section 6.1). Represented as blue boxes, these VxFs are used to carry user data through the network function stratum.
- Control plane functions (Section 6.2). Represented as grey boxes, these VxFs are used to control sessions and the connection between the UE and network and/or access networks and core network including requesting the service, controlling different transmission resources, handover, subscription management, data management, etc.
- Application plane functions (Section 6.3). Represented as orange boxes, these VxFs include O-RAN functions in charge of collecting data and providing value-added services.

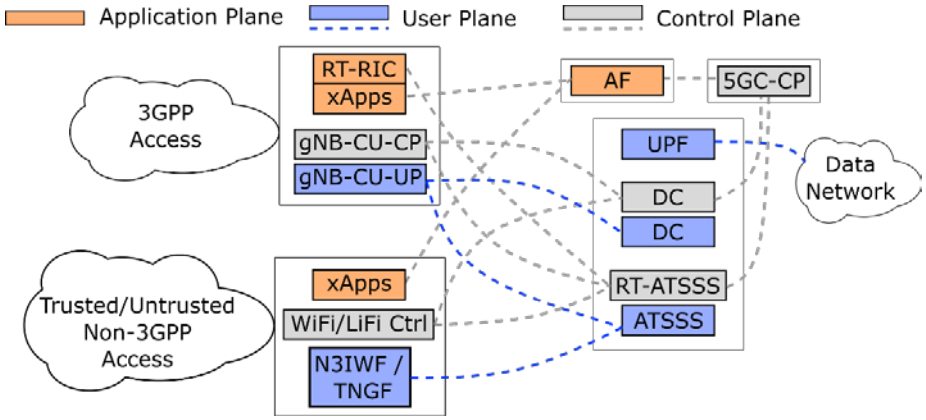

Figure 6.1: 5G-CLARITY network and application function stratum

6.1 User plane functions

6.1.1 3GPP network functions

The user plane protocols implement the actual PDU Session service. This service, based on carrying user data through the 3GPP Access Stratum, allows the exchange of PDUs between an UE and a Data Network [50]. Figure 6.2 shows the protocols on the Uu and the NG interfaces that linked together provide this PDU Session Resource service.

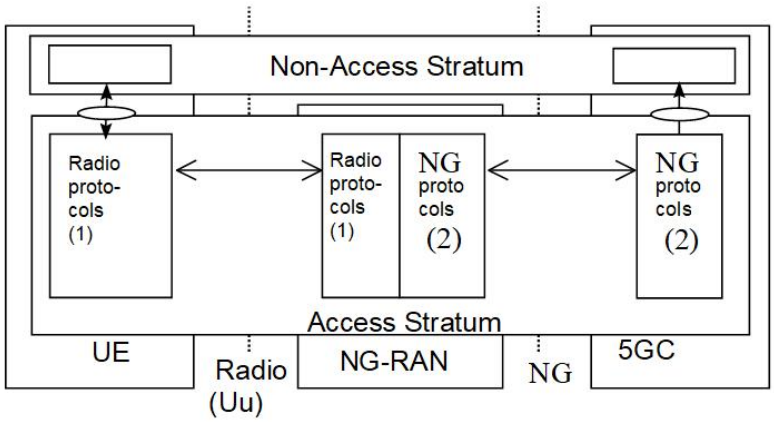

Figure 6.2: End-to-end user plane protocols.

6.1.1.1 Access network: NG-RAN gNB-CU-UP

The gNB-CU User Plane (gNB-CU-UP) is a logical node hosting the user plane part of the Packed Data Convergence Protocol (PDCP) protocol of the gNB-CU for an en-gNB, and the user plane part of the PDCP protocol and the Service Data Adaptation Protocol (SDAP) protocol of the gNB-CU for a gNB. In 5G-CLARITY, only gNBs will be used and not en-gNBs. The gNB-CU-UP terminates the E1 interface connected with the gNB-CU-CP and the F1-U interface connected with the gNB-DU. The overall user plane protocols are depicted in Figure 6.3.

- PDCP, RLC and MAC sublayers (terminated in gNB on the network side) perform the following functions: ROHC, Security, Segmentation/ARQ, Scheduling/Priority Handling, UE Multiplexing and HARQ as per CU CP but specific for the transport of user plane.

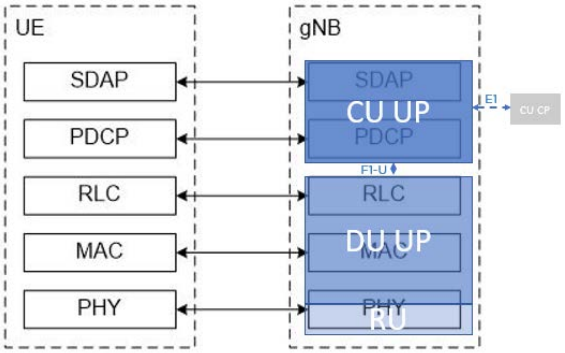

Figure 6.3: CU-UP in Uu User Plane protocol stack.

6.1.1.2 Core Network: 5GC UPF

The UPF supports features and capabilities to facilitate user plane operation in the 5GC. It handles packet processing/transmission from the gNB-CU-UP to the Data Network (upstream) and the other way around (downstream), providing functionality such as packet routing and forwarding, PDU session processing, QoS policy enforcement and data buffering. The definition of UPF represents a natural evolution of the CUPS architecture originally introduced for the EPC, whereby SGW-U and PGW-U are now merged into a unified user plane functionality.

UPF uses the following interfaces:

- **N4 interface**: to communicate with the 5GC control plane (Section 6.2.1.2). This communication is used to get policy enforcements. Based on the rules pushed by SMF via N4 interface, UPF performs packet routing and forwarding, identifies user plane traffic flows, allocates IP address to UE, responds to ARP requests and handles per-flow QoS. It also reports traffic usage to SMF.
- **N3 interface**: to connect 3GPP access network via gNBs and non-3GPP access network via N3IWF or TNGF (see Section 6.1.2 for more details) to the 5GC.
- **N6 interface**: to have a connection with the data network. It can be considered as the interconnect point between the mobile infrastructure and the data network, and it handles encapsulation and decapsulation of GTP-U. N6 interface can also be used to deploy VAFs.
- **N9 interface**: to interconnect multiple UPFs, in there exist multiple UPFs deployed.

As UPF performs packet routing and forwarding, it can be considered as the PDU session anchor point for providing mobility within and between RATs, as explained in more detail in Section 6.1.3. In addition to the PDU session anchor point functionality, UPF can take the role of uplink flow classifier, known as UL-CL and branching point function. In UL-CL, classification of flows based on source or destination IP address is enabled. Traffic classification rules (also known as traffic filter rules) are applied to steer the UE traffic to an appropriate data network, whereas, the branching point function forwards uplink traffic based on the rules related with the prefix of the IPv6 multi-homed PDU session.

6.1.2 Non-3GPP network functions

The integration of non-3GPP networks to 5GC is enabled via N3IWF and TNGF functions for untrusted and trusted non-3GPP access, respectively. Trusted non-3GPP networks can be considered as MNO managed networks, whereas untrusted non-3GPP networks can be considered as 3rd party managed networks. N3IWF and TNGF are considered as the termination point for non-3GPP networks to access the 5GC.

6.1.2.1 Access network: N3WIF

As noted, N3IWF is responsible for interworking between the untrusted non-3GPP networks and the 5G core. Once a UE establishes a connection to a non-3GPP access, the UE sets up a tunnel against N3IWF, which is then mapped to the N2 and N3 interfaces against the 5G core. In other words, N3IWF terminates N2 and N3 interfaces to 5G core network for control plane and user plane (UPF), respectively.

6.1.2.2 Access network: TNGF

Similar to N3IWF, TNGF is responsible for interworking between the trusted non-3GPP networks and the 5G core. It also terminates the user plane and control plane interfaces N2 and N3, respectively and uses tunnels against UE.

The following figure represents the user plane protocol stack for N3IWF which is also the same for TNGF. The control plane for non-3GPP access via N3IWF or TNGF also follows a similar protocol stack, where instead of UPF in Figure 6.4 AMF is the function that uses N2 interface for the control plane to N3IWF or TNGF, and instead of PDU layer transmission NAS signalling is transmitted.

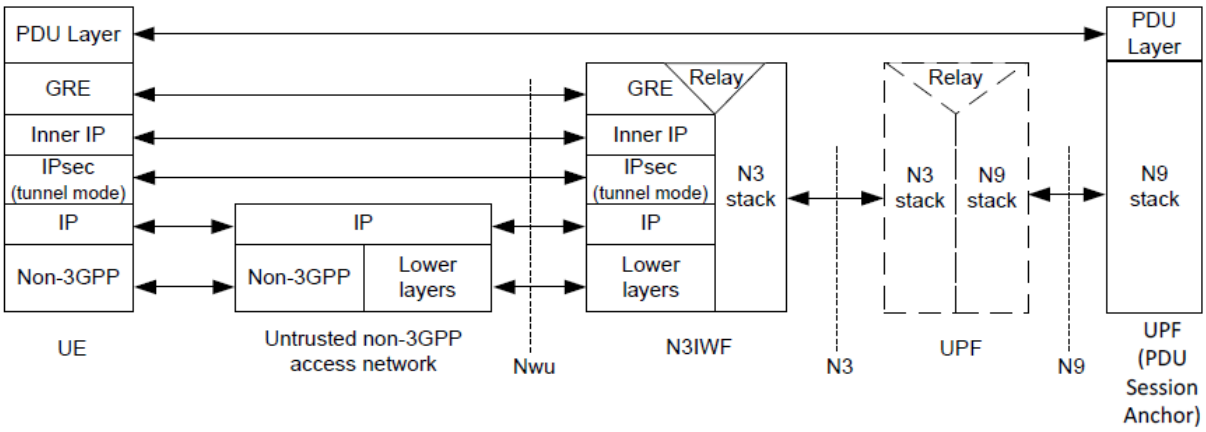

Figure 6.4: User plane for non-3GPP access [83].

6.1.3 Multi-connectivity user plane

6.1.3.1 Access network: dual connectivity

Defined as a natural evolution of LTE Dual Connectivity [15], Multi-Radio Dual Connectivity utilises radio resources provided by two schedulers, located in two different radio access network nodes: One of the nodes provides 5GNR access and the other one providing either LTE or 5GNR access. 5GNR multi-radio dual connectivity is based on LTE dual connectivity, where terminals with multiple transmitter-receivers may be configured to utilise resources provided by two LTE nodes.

6.1.3.2 Core network: AT3S

Access traffic steering, switching and splitting (AT3S) enables multi connectivity service for multi-access PDU (MAPDU) sessions [2] that can simultaneously utilize user plane resources of both 3GPP and non-3GPP access networks:

- **Traffic steering** is a functionality that selects an access network for a new data flow and transfers the traffic of this data flow over the selected network, which can be 3GPP or non-3GPP access network. Steering functionality can be used to balance the load between the two access networks or prioritize one of the networks for data flow.

- **Traffic switching** is a functionality that moves all ongoing data traffic from one access network to another. It can be used to provide data traffic continuity.
- **Traffic splitting** is a procedure that splits the data traffic across 3GPP and non-3GPP access networks. It can be used to utilize user plane resources of 3GPP and non-3GPP networks at the same time.

The steering, switching and splitting functionalities can only be used when there are 3GPP and non-3GPP networks available. It is not possible to use these functionalities for only 3GPP or only non-3GPP access network. The steering, switching and splitting functionalities can be based on a high-layer protocol, which is MPTCP and a low-layer protocol based on AT3S function, which is AT3S-LL.

As 5G systems have various service types, the required KPIs and SLAs can be different from one UE to another. Therefore, multi-connectivity policies that can achieve the required KPIs/SLAs should be on a per-UE basis. In the user plane of AT3S, UPF can use MPTCP proxies for each user. On the UE side, each access technology can have its IP address provided by 5G, Wi-Fi and LiFi networks and these IP addresses can be used to reach UE from a single source, which is the MPTCP proxy used for each user as a container inside the UPF. As captured in Figure 6-13 from D3.1 [2], from UPF to UE, per UE per WAT routes are used. For 3GPP access, this route is from UPF to gNB via either a null tunnel or a GTP tunnel, and then gNB to UE. The former is used for an IP flow, whereas the latter is used for any type of traffic. For non-3GPP access, the route is from UPF to N3IWF/TNGF first and then from N3IWF/TNGF to UE with the same tunnelling options. This one tunnel per UE and per WAT can be considered as N3 interfaces and a binding function can be used within N3IWF/TNGF to bind tunnel IP to tunnel ID, and vice versa for each per UE per WAT route. A steering and switching function can be used within the UPF to route the traffic flow from/to tunnel bindings to/from each UE container.

6.2 Control plane functions

The control plane protocols control the PDU Sessions and the connection between the UE and the network from different aspects, including service request, controlling different transmission resources, handover etc. They also provide a mechanism for transparent transfer of 3GPP Non-Access Stratum (NAS) messages [84] for Connection Management (CM) and Session Management (SM) functions in the Non-Access Stratum. Figure 6.5 shows the control plane (signalling) protocol stacks on NG and Uu interfaces.

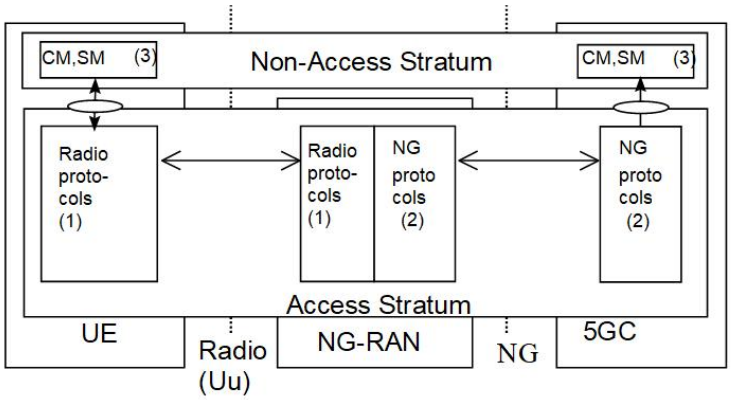

Figure 6.5: E2E control plane protocols.

6.2.1 3GPP network functions

6.2.1.1 Access network: NG-RAN gNB-CU-CP

The gNB-CU-CP is a logical node hosting the RRC and the control plane part of the PDCP protocol of the gNBCU for an en-gNB or a gNB. In 5G-CLARITY only gNBs will be used. The gNB-CU-CP terminates the E1 interface connected with the gNB-CU-UP and the F1-C interface connected with the gNB-DU. The overall control plane

protocols are depicted in Figure 6.6.

- PDCP, RLC and MAC sublayers (terminated in gNB on the network side) perform the following functions: ROHC, Security, Segmentation/ARQ, Scheduling/Priority Handling, UE Multiplexing and HARQ as per CU UP but specific for the transport signalling.
- RRC (terminated in gNB on the network side) performs the following functions: Broadcast of System Information related to AS and NAS, Paging initiated by 5GC or NG-RAN, Establishment, maintenance and release of an RRC connection between the UE and NG-RAN, Security functions including key management, establishment, configuration, maintenance and release of Signalling Radio Bearers (SRBs) and Data Radio Bearers (DRBs), Mobility, QoS management functions, UE measurement reporting and control of the reporting, Detection of and recovery from radio link failure and NAS message transfer to/from NAS from/to UE.
- NAS control protocol (terminated in 5GC on the network side) performs the functions such as authentication, mobility management, security control, etc.

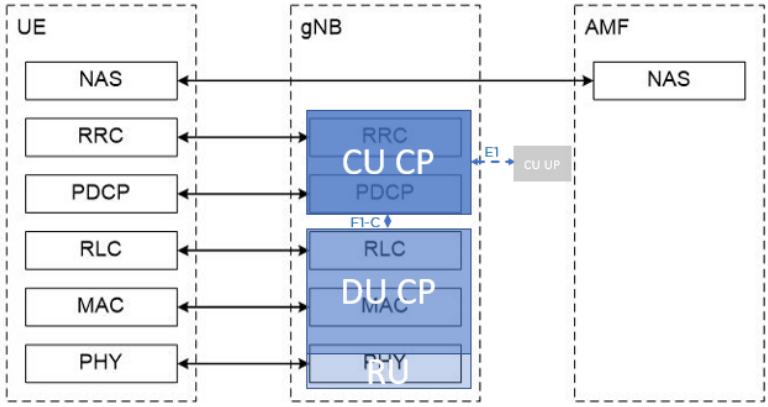

Figure 6.6 CU-CP in Uu control plane protocol stack.

6.2.1.2 Core network: 5GC CP

The following table summarizes the relevant 5GC network functions that follow cloud-native principles as noted in Chapter 4. These network functions are modular and connected using service-based interfaces that result in an SBA as shown in Figure 6.7.

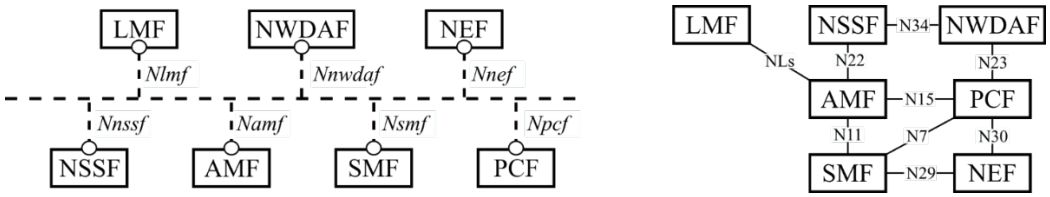

Figure 6.7: 5GC SBA with (left) service-based interfaces; (right) reference point representation

These NFs are being considered in 5G-CLARITY control plane architecture and are represented by 5G-CP (grey) box in Figure 6.1.

Table 6-2: 5GC network functions [50]

Network Function	Description
AMF	The Access and Mobility Function (AMF) receives all connection and session related information from UE and responsible for registration, connection, mobility and reachability.

	gNB-CU-CP also terminates here using N2 reference point.
SMF	The Session Management Function (SMF) is responsible for interacting with the decoupled data plane, creating, updating and removing PDU sessions. It also manages session context with the UPF by defining traffic steering parameters and ensuring appropriate routing of packets. Communication between SMF and UPF is done through N4 reference point.
PCF	The Policy Control Function (PCF) provides policy rules for control plane functions such as network slicing, roaming and mobility management and dynamically controls the policy and charging behaviour at the SMF.
NSSF	The Network Slice Selection Function (NSSF) assists the AMF on selecting the network slice instances and providing the requested network slice information.
NWDAF	The Network Data Analytics Function (NWDAF) supports data collection from NFs, application functions (AFs) (see 6.2.1.3) and OAM, and it is responsible for providing network analysis information to NFs upon request.
NEF	The Network Exposure Function (NEF) provides a mechanism to securely expose services and features of the 5GC by masking network and user sensitive information to external AFs (see 6.2.1.3). It also translates the information received from an AF to the one sent to internal core NFs.
LMF	The Location Management Function (LMF) is responsible for managing overall location determination. Coordinates resource scheduling required for the location determination of a UE, calculates a final location for a UE, and estimates the achieved accuracy.

6.2.1.3 Application function

AFs exist for different application services that can be owned by the network owner or third parties. AFs interact with the 5GC NFs such as NWDAF, PCF either directly or indirectly through the NEF to provide services such as traffic routing, policy control, etc. For example, for traffic routing, an AF may send requests to influence SMF routing decisions which may also influence UPF as well as DN selections. At this point, the AF firstly interacts either directly with PCF or via NEF where NEF performs several mappings to create the requests. Then, SMF may send user plane management notifications to the AF either directly or via NEF.

6.2.2 Non-3GPP network functions

6.2.2.1 Access Network: Wi-Fi – LiFi control plane

Figure 6.8 depicts the 5G-CLARITY concept of an integrated Wi-Fi-LiFi SDN based L2 network. This network provides a self-contained L2 network that interfaces on one side with the user devices that can be equipped with Wi-Fi interfaces, LiFi interfaces, or both, and on the other side with a standard 802.1 Ethernet network segment. The main advantage of the proposed architecture, as compared to a standard 802.1 Ethernet segment, is that this architecture enables the 5G-CLARITY control plane to decide how packet flows are steered across the L2 network, which is a critical feature to support slicing in the access and transport domains, and could not be achieved with a standard Ethernet forwarding model based on autonomous learning bridges.

In the proposed architecture, user devices are identified using 48-bit IEEE compliant MAC addressed, since both Wi-Fi and LiFi conform to the 802.11 protocol stack. If a device has both a Wi-Fi and a LiFi interface, then the device will have one MAC address in each interface, and the L2 network will perceive it as two different devices. It is up to the multi-connectivity solution of the 5G-CLARITY system to reconcile the fact that the two MAC addresses belong to the same device.

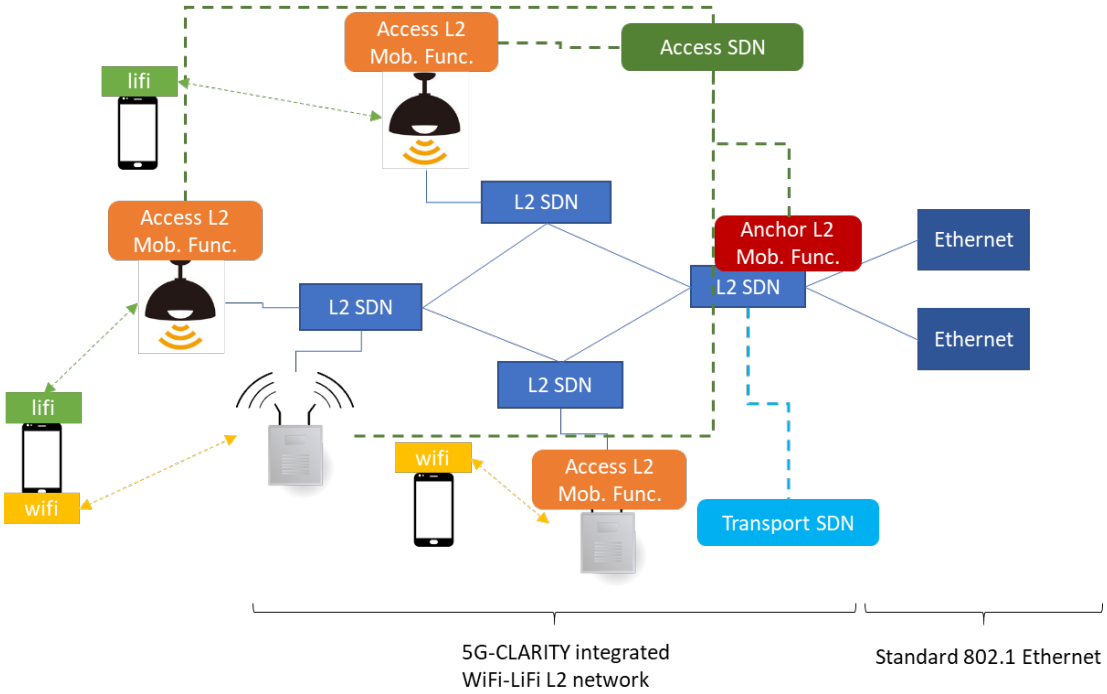

Figure 6.8: 5G-CLARITY integrated Wi-Fi-LiFi L2 network.

The proposed integrated L2 network provides mobility services, whereby a user device can roam through the network having its Wi-Fi or LiFi interfaces independently executing handovers across access points, and the L2 network will readjust the forwarding paths to redirect the user packets to the appropriate access point. However, only "break-before-make" handover will be supported, since "make-before-break" would require a custom control plane between the network and the user devices, which would limit the adoption of the solution.

The proposed design is composed of the following types of functions:

- Physical network functions, including Wi-Fi access points, LiFi access point and SDN enabled L2 switches.
- Software-based user plane functions, including the Access L2 mobility function and the Anchor L2 mobility function.
- Virtual Control Plane functions including the Access SDN controller.

Table 6-3 describes the role of the software-based user plane functions and the virtual control plane functions. The physical network functions have already been described in Chapter 5.

Table 6-3: Network Functions for the Integrated L2 Wi-Fi-LiFi Network.

Network Function	Description
Anchor L2 mobility function	This function maintains the bindings between the MAC addresses of the user devices and the LiFi or Wi-Fi access points where this user device is attached.
Access L2 mobility function	Collocated with the LiFi and Wi-Fi access points, this function detects a new attachment from a user device and updates the bindings for the MAC address of this device in Anchor L2 mobility function.
Access SDN Controller	An SDN controller which communicates with the Access L2 mobility functions, in order to detect new user attachments. It also interfaces with the LiFi and Wi-Fi physical access points to gather wireless specific telemetry.

6.2.3 Multi-connectivity control plane

5G-CLARITY enhances 3GPP AT3S function by introducing a real-time control function to the AT3S. This section focuses on control plane function of the enhanced AT3S. Control plane of the 3GPP multi-radio dual connectivity is the same as 3GPP describes in [84].

6.2.3.1 Core Network: RT-AT3S control

Based on 3GPP specification, AT3S policies are provided by 5GC CP to UPF. As these rules are derived based on what is provided by 5GC in the first place, any change on the AT3S rules should also be instantiated by 5GC based on the path performance measurements from UE and UPF. In virtue of CUPS in 5G SBA, the functions that derive AT3S policies can be located apart from UPF, which can be located in the network edge such as in private premises. Therefore, in order to incorporate instant changes on the network status, having policy related functions apart from UP introduces latency on generating AT3S rules which by the time of the arrival of new AT3S rules, the network state may have changed again. In order to cope with this, a core-based enhanced AT3S (eAT3S) solution is proposed in 5G-CLARITY. The proposed solution enables the use of multi-WAT telemetry at RAN level via near RT RIC and xApps O-RAN framework, which are introduced in Section 6.3, to decide AT3S rules in real time. The main motivation of the proposed eAT3S solution is to react to sudden changes on the network status in order to efficiently use 3GPP and non-3GPP access networks. A new steering mode named “real-time” is introduced to initiate eAT3S solution along with near-RT RIC and multi-WAT telemetry.

Accordingly, in the control plane of the proposed eAT3S solution, a real-time controller is used to gather telemetry data from 5GNR, Wi-Fi and LiFi along with AT3S performance measurements from UPF and modifies AT3S rules in real time. Once the steering mode is chosen as “real-time” mode, the modified rules are listed before the AT3S rules that are pushed by SMF. If a UE does not support eAT3S functionality, then it can still use the previously generated rules. Several eAT3S-based control plane functions are described for eAT3S and their roles are summarized in Table 6-4 and shown in Figure 6.9.

In addition to the network functions described in Table 6-4, “Access SDN Controller” defined in Table 6-3 is used to provide seamless L2 mobility and to gather access-specific telemetry for Wi-Fi and LiFi networks.

Table 6-4: Network Functions for eAT3S.

NF	Description/Role
Rt-UPc-AT3SF	Enables adaptive modifications of steering/splitting parameters within the UPF.
Rt-AT3S Controller	Collects telemetry data from 5GNR, Wi-Fi and LiFi as well as path measurement results from the UPF. Based on these data, it generates real-time AT3S rules and pushes them to the UPF.
RT-UE-AT3SF	Gets network status related updates/modifications from Rt-AT3S Controller and routes uplink traffic accordingly.
5GNR telemetry	Provides 5GNR telemetry data to the Rt-AT3S Controller.

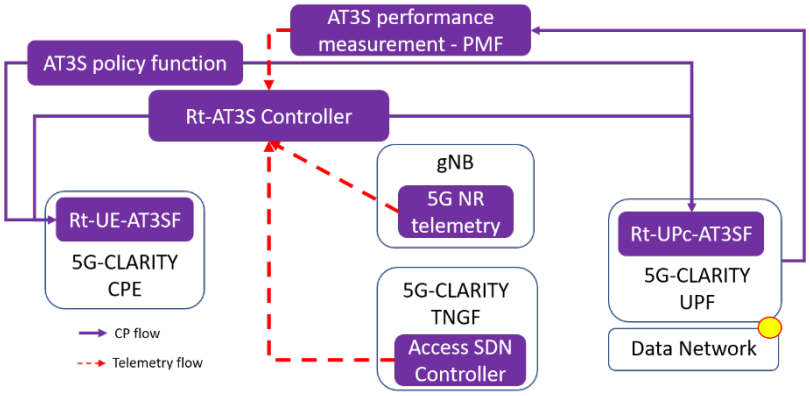

Figure 6.9: Diagram for eAT3S functions.

6.3 Application plane functions

These 5G-CLARITY functions include near-RT-RIC and non-RT-RIC xApps. These xApps, to be used in the O-RAN Alliance framework, are enabled by the Acceleran dRAX solution.

6.3.1 near-RT RIC

As defined by the O-RAN Alliance reference architecture, the near-RT RIC is a logical function that enables near real-time control and optimization of E2 nodes (e.g. gNB-CU-CP, gNB-CU-UP, gNB-DU), functions and resources via fine-grained data collection and actions over the E2 interface with control loops in the order of 10 ms-1s.

Table 6-5 near-RT RIC Functionality.

Functionality	Description
near-RT Data Collection	Exposes fine grained RAN data from E2 nodes via the E2 interface
near-RT Actions	Enables actions on RAN functions towards E2 nodes via the E2 interface (e.g. monitor, suspend/stop, override, or control the behaviour of the E2 node)
xApps hosting	Hosts one or more xApps that use E2 interface to collect near real-time information (e.g. on a UE basis or a Cell basis) and provide value added services
non-RT Policy steering	Enables steering of the E2 nodes via the policies and the enrichment data provided via A1 from the non-RT-RIC
RRM/SON	Enables RRM/SON functions between the non-RT-RIC and the E2 node as exposed in the E2 Service Model (including cluster wide RRM/SON functions)

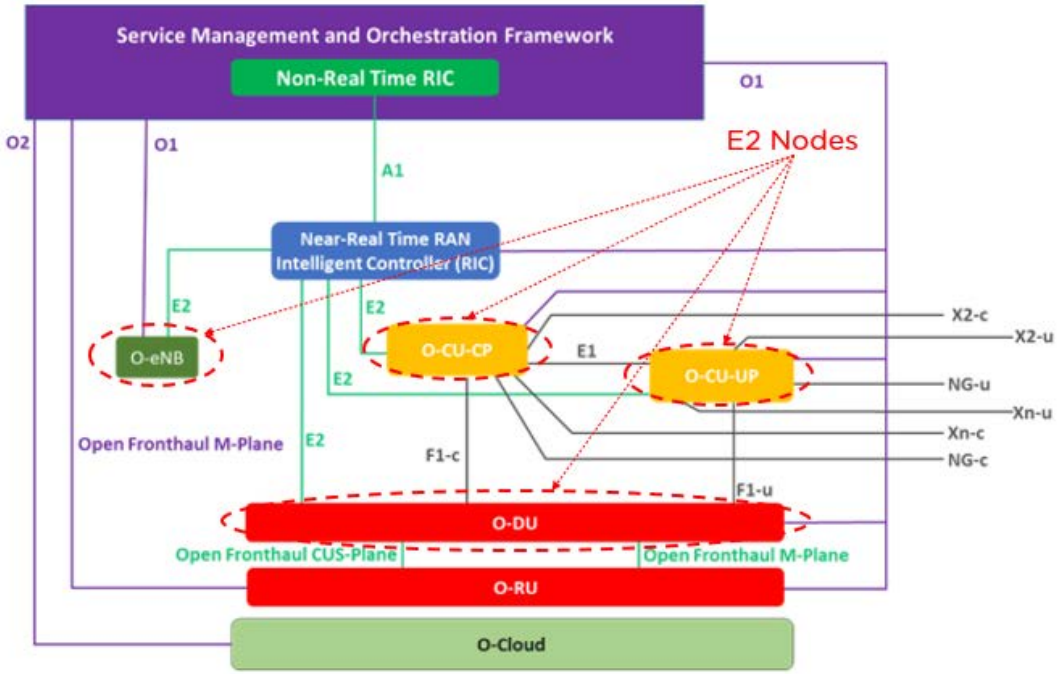

Figure 6.10 O-RAN E2-nodes.

6.3.2 Accelleran xApps and 5G-CLARITY xApps

An xApp is an application designed to run on the near-RT RIC. Such an application is likely to consist of one or more microservices and at the point of on-boarding will identify which data it consumes and which data it provides in an open and interoperable manner. The application is independent of the near-RT RIC and may be provided by any third party. The E2 enables a direct association between the xApp and the RAN functionality.

The overall Accelleran dRAX solution is illustrated in Figure 6.11, with xApps executed atop the RIC nR-RT. The Accelleran near-RT-RIC can host different xApps that have access to the Accelleran databus to collect near real-time information and provide value added services. Typical default Accelleran xApps relate to usual network functions associated to handling a cluster of 5GNR small cells such as Plug and Play, Interference Management, Handover Management, etc. As part of 5G-CLARITY, the Accelleran dRAX will be enabled with multi-WAT telemetry data from 5GNR, Wi-Fi and LiFi which will be exposed via the Accelleran data bus the 5G-CLARITY AT3S controller multi-WAT xApp. Accelleran will enable a 5GNR-enhanced SAS clients an xApp and potentially many others related to AI/ML inference, Position Calculation, etc.

Table 6-6: Tentative 5G-CLARITY xApps.

5G-CLARITY xApp	Description/Role
SAS Shared Spectrum Manager (enhanced SAS CBRS client)	This xApp implements SAS client functionality that will be used together with simulated external SAS shared Access Controller to enable flexible 3-tier Shared Access Spectrum Management including: *i)* Interferece Coordination Coexistence group; *ii)* 5GNR SA.
AT3S Controller	This xApp implements real-time AT3S control of multi-WAT access based on multi-WAT telemetry

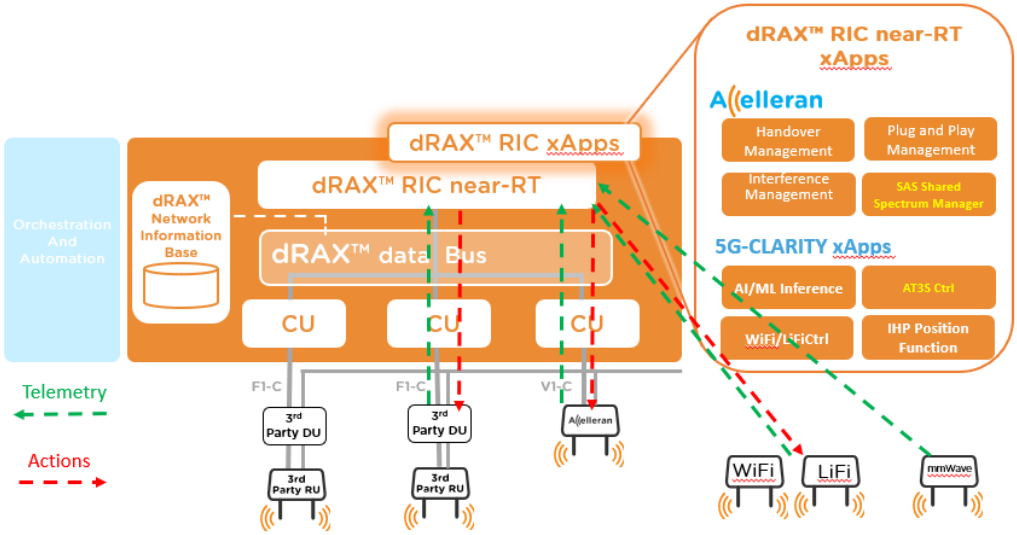

Figure 6.11: Accelleran dRAX with multi-WAT telemetry.

6.3.3 Localization server

The multi-connectivity environment available in 5G-CLARITY comprises different WATs, each of them featuring localization capabilities that can be used for estimating the position of a UE. These WATs, depending on the underlying technology they use, leverage a diverse range of localization approaches and strategies. Some of these approaches are capable of direct position estimation, but some of them would provide only intermediate data which can be used further for position estimation of UEs.

The APs for different WATs may feature different synchronization and additional capabilities necessary for the localization functionality. Additionally, position estimates for a given UE can be obtained by multiple WATs, supported by the UE. These multiple position estimates can be combined in order to obtain a better position estimate. These functionalities will be integrated in a so-called localization server. The general architecture of the localization system is shown in Figure 6.12.

The localization server provides different interfaces towards the different WATs. These interfaces accept the WAT specific localization data. This data is later combined in order to estimate the position of a given UE. Additionally, when a UE can access multiple WATs, the position estimates from different WATs can be merged and combined in order to improve the localization accuracy. Additionally, the localization server controls the localization process. It selects the APs that are used for location estimation based on multiple parameters, like the received signal strength of the UE signal, LOS/NLOS condition, etc.

The functionality of each of the methods depicted in the figure are the following:

- Acceptance of services requests, and trigger responses to the received requests via the interface towards the different WATs.
- Merging of the localization data from different technologies. This can be performed in different ways, i.e. subject to the localization method employed and the underlying technology. Kalman filtering or least squares methods are possible implementations in this regard.
- Control of the positioning functions, which takes care of identifying which (and how many) technologies are involved in the localization process for fulfilling the localization requirements.
- Perform the position estimation of a specific target UE.

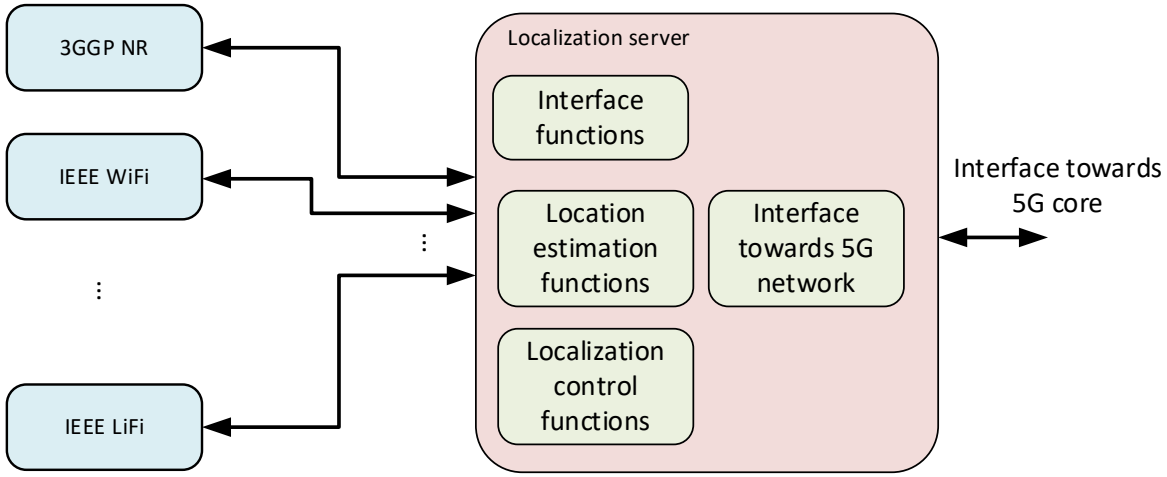

Figure 6.12: Localization server architecture

Depending on the configuration and the requirements needed to be implemented in the localization server, its functionality or part of its functionalities can be implemented as a xApps. This means that the localization server can be implemented as an near-RT xApp, since a slight delay can be tolerated in the localization system, or an xApp can use data from the localization server to implement some (or additional) localization functions and services.

7 Management and Orchestration Stratum Design

Table 7-1 summarizes the requirements to be addressed by the Management and Orchestration stratum in 5G-CLARITY.

Table 7-1: 5G-CLARITY Management and Orchestration Stratum Requirements

Requirement ID	Requirement Description
CLARITY-MOS-R1	The 5G-CLARITY management and orchestration stratum shall be architected following the SBMA principles, with a set of MFs providing/consuming management services through a service bus.
CLARITY-MOS-R2	The 5G-CLARITY management and orchestration stratum shall allow for the provisioning of 5G-CLARITY resource-facing services (i.e. 5G-CLARITY wireless, compute and transport services).
CLARITY-MOS-R3	The 5G-CLARITY management and orchestration stratum shall keep a resource inventory, with information on the on-premise resources that can be used for the provision of 5G-CLARITY resource-facing services. This includes information on: *i)* the resource capacity of deployed wireless access nodes, including Wi-Fi/LiFi APs and physical gNBs; *ii)* the compute nodes available in the clustered NFVI (RAN cluster and edge cluster), and related computing/storage/networking resources; *ii)* the capacity and topology of deployed transport network.
CLARITY-MOS-R4	The 5G-CLARITY management and orchestration stratum shall store a catalog of VxFs/NSDs.
CLARITY-MOS-R5	The 5G-CLARITY management and orchestration stratum shall support to create, retrieve, update and delete VxFDs/NSDs
CLARITY-MOS-R6	The 5G-CLARITY management and orchestration stratum shall allow to create several instances of the same VxF/NFV service.
CLARITY-MOS-R7	The 5G-CLARITY management and orchestration stratum shall allow VxF / NFV service scaling. This scaling includes the scaling-in and scaling-out the resources of deployed VxF / NFV service instances.
CLARITY-MOS-R8	The 5G-CLARITY management and orchestration stratum shall allow for the provisioning of 5G-CLARITY slices, by defining separate resource quotas when allocating individual 5G-CLARITY resource-facing services.
CLARITY-MOS-R9	The 5G-CLARITY management and orchestration stratum shall maintain information regarding the mapping between 5G-CLARITY slices, constituent 5G-CLARITY resource-facing services and allocated resources.
CLARITY-MOS-R10	The 5G-CLARITY management and orchestration stratum shall allow resource elasticity and AI-assisted placement optimization as part of the 5G-CLARITY slice lifecycle management.
CLARITY-MOS-R11	The 5G-CLARITY management and orchestration stratum shall provide means for model-based data aggregation, with the ability to collect and process management data (e.g. performance measurements, fault alarms) from different sources in an automated and scalable manner.
CLARITY-MOS-R12	The 5G-CLARITY management and orchestration stratum shall be able to correlate aggregated data with deployed 5G-CLARITY slices and services instances, providing input to the intelligence engine for AI assisted operation of these instances.
CLARITY-MOS-R13	The 5G-CLARITY management and orchestration stratum shall provide necessary cloud-native capabilities for MF service production/consumption across the entire stratum.
CLARITY-MOS-R14	The 5G-CLARITY management and orchestration stratum shall allow individual 5G-CLARITY customers (e.g. MNOs) to securely access and consume MF services.

CLARITY-MOS-R15	The 5G-CLARITY management and orchestration stratum shall provide the means to expose capabilities with appropriate abstraction levels to individual 5G-CLARITY customers
CLARITY-MOS-R16	The 5G-CLARITY management and orchestration stratum shall provide isolation among customers' workflows and request

Following the design principles described in Section 4.2.3 and considering the previous requirements, Figure 7.1 illustrates the MFs that compose the 5G-CLARITY Management and Orchestration stratum, all connected through a service bus using a SBMA. Although not explicitly shown in the figure, it is assumed that there is an underlying cloud that supports the instantiation and management of resources building up these MFs. Notice that this cloud is logically independent of the RAN and edge clusters described in Section 5.3[6].

The MFs composing the Management and Orchestration stratum can be arranged into four main functional groups:

- **Service and Slice Provisioning**: Grouping all the MFs that are involved in the provisioning of 5G-CLARITY slices, composed of wireless, transport and compute resources, as they have been defined in Section 3.2.2.
- **Data Processing and Management**: Grouping all the MFs that collect telemetry data from the physical and the Network Function stratums, in order to make the data available to the Intelligence stratum or to the external entities interacting with the management plane.
- **Cloud native support:** These are MFs required to enable a cloud native deployment of the 5G-CLARITY management and orchestration stratum. For example, enabling the deployment of stateless MFs that can be scaled dynamically in a cloud environment.
- **External Access Mediation**: Including MFs that police the interactions with external entities, such as the 5G-CLARITY intelligence stratum (Section 8) or customers' management systems (e.g. MNO's 3GPP management system).

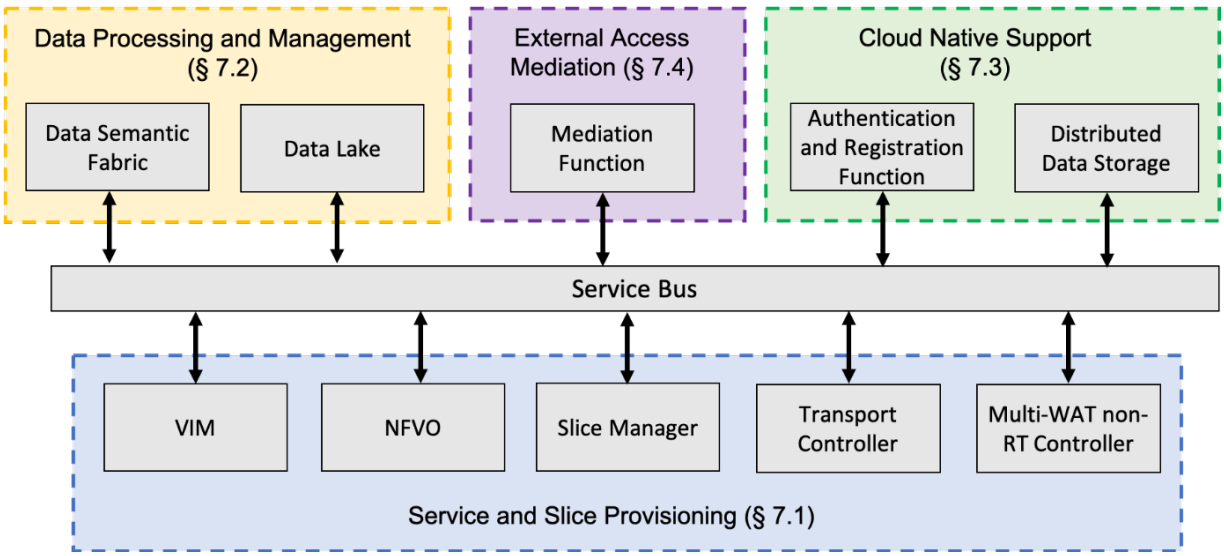

Figure 7.1: 5G-CLARITY management and orchestration stratum: an SBMA approach.

[6] Resources from RAN and edge cluster are operated by the 5G-CLARITY management and orchestration stratum. Therefore, it is needed another cloud to host MFs from this stratum.

In the following subsections, we introduce the individual MFs envisioned within the 5G-CLARITY Management Plane and present the main services exposed by each MF. For an initial design of each MF, the interested reader is referred to 5G-CLARITY D4.1 [49].

7.1 Service and slice provisioning functions

Figure 7.2 depicts the relations of the MFs dealing with Service and Slice Provisioning group, highlighting the elements of the Infrastructure and Network Function Stratums that each MF interacts with, as well as potential reference interfaces between MFs. Notice though that the reference diagram is provided to illustrate the role of each MF, as in principle the services offered by a MF can be accessed by any other MF through the shared message bus.

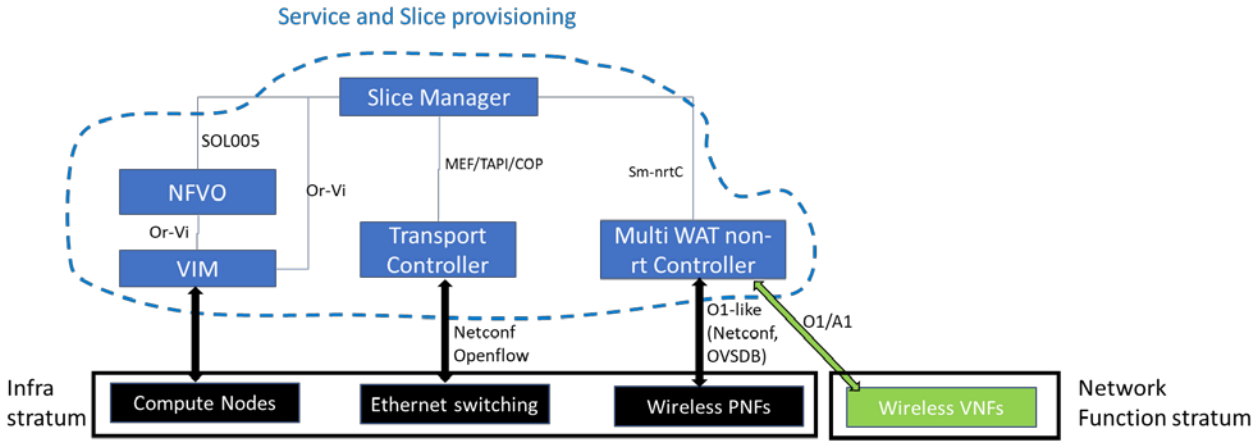

Figure 7.2: Service and slice provisioning MFs.

7.1.1 NFV MANO

The first two MFs, the NFVO and the VIM represent an ETSI NFV MANO [85] stack contained within the 5G-CLARITY management plane. The goal of this software stack is to deploy and operate network services on virtualized execution environments, i.e. NFVI nodes. A virtualized network service, either partially (some network functions are VNFs and other are PNFs) or fully (all network functions are deployed as VNFs) is called an NFV service. Figure 7.3 depicts an example of a typical NFV service. As it can be seen, it includes multiple VNFs connected with each other using virtual links (VL).

The ETSI NFV MANO the ETSI NFV MANO makes use of deployment templates called NFV descriptors to manage individual NFV services (and their components) throughout their lifecycle, from instantiation to termination. These descriptors describe resource requirements for the managed instances in a technology-agnostic manner, allowing reusability in different execution environments (i.e. the same NSD can be used to deploy two different NFV services, one running on VMWare and the other on OpenStack) and facilitate a model-driven automation (i.e. day-1 and day-2 configuration scripts and allowed scaling levels are embedded in the descriptors). The most relevant NFV descriptors for 5G-CLARITY system will be NFV Service descriptors (NSDs), VNF Descriptors (VNFDs), PNF Descriptors (PNFDs) and Virtual Link Descriptors.

The NFVO offers a set of services through its northbound interface, the so-called Os-Ma-nfvo reference point [86]. The RESTFul implementation of these services is specified in the ETSI NFV-SOL005 specification [87].

Table 7-2 describes the services offered by the NFVO in 5G-CLARITY, specifying for service a service ID, a service name, a high-level description, and a reference to the standard specifications where these services are defined.

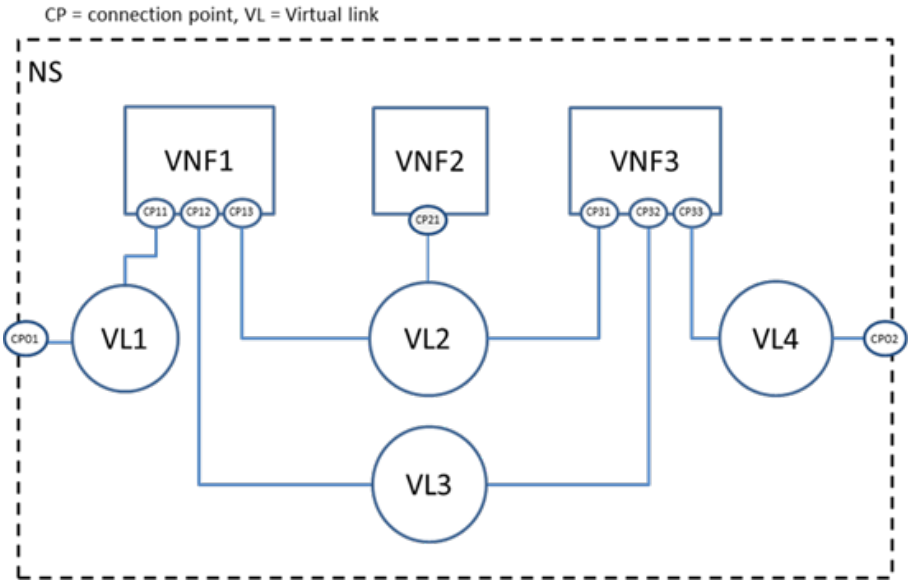

Figure 7.3: Example of an NFV network service.

Table 7-2: NFVO Services.

MF service ID	MF service Name	Description	Reference specifications
Nfvo_Nsd_Mgmt	**NSD management**	This service allows the management of NSDs and PNFDs, offering means for their individual on-boarding, updating, deletion and fetching.	SOL005
Nfvo_Ns_Lcm	**NS Lifecycle Management**	This service allows manipulating NFV service instances throughout their entire lifecycle, from their commissioning to decommissioning. This include NFV service instantiation, update, scaling (in/out), healing and deletion operations.	SOL005
Nfvo_Perf_Mgmt	**NS Performance Management**	This service allows the creation, deletion of performance measurement jobs for NFV services and constituent VNF instances [88]. It allows to obtain performance reports.	SOL005
Nfvo_Fault_Mgmt	**NS Fault Management**	This service allows processing of alarms set up in the virtual infrastructure manager or the PNFs	SOL005
Nfvo_Vnf_Mgmt	**VNF package management**	This service allows creation, uploading, deletion of VNF resources. Enables onboarding of VNFDs	SOL005

The other 5G-CLARITY MF that along with NFVO constitutes the ETSI NFV stack is the VIM, which offers services standardized by ETSI NFV under the Or-Vi reference point [89]. The following table describes the main services offered by the VIM that are relevant to the 5G-CLARITY Management and Orchestration stratum.

Table 7-3: VIM Services

MF Service ID	MF Service Name	Description	Reference Specifications

Vim_Compute	**Virtualised Compute service**	This service offers virtualization of compute resources, e.g. in the form of virtual machines or containers	Or-Vi
Vim_Network	**Virtualised Network service**	This service offers virtualization of network resources, e.g. setting up virtual network connecting virtual machines or containers	Or-Vi
Vim_Storage	**Virtualised Storage service**	This service offers virtualization of storage resources, e.g. offering a dedicate volume that can be accessed from a virtual machine or container	Or-Vi
Vim_Fault_Mgmt	**Virtualised Resource Fault Management service**	This service allows providing alarms from the VIM resulting from the faults related to the virtualised resources.	Or-Vi
Vim_Perf_Mgmt	**Virtualised Resources Performance Management service**	This service allows the subscription to performance management information related to the virtualised resources controlled by the VIM	Or-Vi
Vim_Res_Rsrv	**Virtualised Resource Reservation service**	This service allows an authorized MF to perform operations on virtualised compute resources reservations. The service includes operations for creating, querying, updating and terminating reservations on virtualised compute resources	Or-Vi
Vim_Res_Quota	**Virtualised Resource Quota service**	This service allows an authorized MF to perform operations on virtualised compute resources quotas available to the consumer functional block. The interface includes operations for creating, querying, updating and terminating quotas on virtualised compute resources.	Or-Vi
Vim_Host_Resrv	**Compute Host Reservation Management service**	This service allows an authorized MF to create, query, update, and terminate compute host reservation operations.	Or-Vi
Vim_Cpct_Mgmt	**NFVI Capacity Management service**	This service allows an authorized MF to request operations related to capacity and usage reporting, including: available, used, reserved and total capacity of the physical compute resources managed by a VIM instance, globally or per resource zone, and ii) utilization of the capacity, both on VIM global level but also per resource zone.	Or-Vi
Vim_Plc_Mgmt	**Policy Management service**	This service allows the NFVO to manage subscriptions to notifications sent by the VIM which inform about changes of a policy and about any detected policy conflicts. It allows the VIM to provide such notifications to the subscriber (e.g., NFVO)	Or-Vi

7.1.2 SDN transport controller

The SDN Transport Controller is responsible for abstracting the control plane functions of the 5G-CLARITY

transport network, which provides connectivity among the different 5G network functions and services. In the 5G-CLARITY management plane, this manager shall operate Layer 2 bridged transport networks comprising standard Ethernet nodes and/or TSN bridges. It might be realized as a recursive hierarchy of transport controllers. Table 7-4 includes the services exposed by the SDN Transport Controller.

Table 7-4: SDN Transport Controller Services

MF Service ID	MF Service Name	Description	Reference Specifications
Tc_Topo	**Topology service**	This service enables to access topological information about the transport network (e.g., top-level network topology, per link details, per node and per layer details).	T-API, COP
Tc_Conn	**Connectivity service**	This service enables setting up the connectivity between two service end-points of the transport network. The service consumer is allowed to retrieve connectivity information and issue connectivity service requests. The lifecycle management (LCM) operations of a connectivity service are: create, update, and delete. A connectivity service can be created either by expressing the source and destination paths, or by specifying the full list of nodes to be traversed.	T-API, COP
Tc_Conn_Inventory	**Inventory service**	This service returns a data structure listing all the currently active connectivity services, including source and destination end-points as well as the service identifier.	T-API, COP
Tc_Path_Comp	**Path computation service**	This service allows the computation of paths fulfilling a set of constraints (e.g., a minimum capacity, a maximum path cost) given a set of sources and destinations. The service may optionally be externalized to e.g. the intelligence stratum, which can compute a path on its own and request a connectivity service.	T-API
Tc_Virt_Net	**Virtual network service**	This service enables the realization of virtual networks (VNs) that interconnects multiple service-end-points of the transport network. In contrast to the connectivity service, the service consumer might change, control and operate the VN dynamically if allowed by the service agreement. The service consumer can retrieve the ongoing virtual network (VN) information and carry out VN's LCM operations of (e.g., create, modify, and delete). It supports two formats for requesting a virtual network: • The request includes the full specification of the VN setup. • The request provides a top-level description of the VN and constraints. For instance, QoS requisites (e.g., latency, jitter, packet loss, reliability), end points to be connected, traffic matrix, and traffic demands characterization. This service returns a control end-point that allows to manage the virtual network elements, for example an OpenFlow datapath endpoint.	T-API, Custom (REST)
Tc_Notif	**Notification**	This service enables the subscription to receive notifications of	T-API

	service	events of interest (e.g., alarms, performance threshold crossing, state change, etc). This service requires: • Notification types discovery. • Procedures for the notification subscription LCM (e.g., create, modify, delete, suspend, resume).	
Tc_Tsn_Conf	**TSN configuration service**	This service enables the application of global configurations to the TSN forwarding plane (e.g., per port priority-to-traffic class mapping, time synchronization, etc.).	Custom (REST)

7.1.3 Multi-WAT non-RT RIC

The multi-WAT non-RT-RIC MF extends the O-RAN defined Non-rt RIC functionality. The new MF can now manage the following network functions described in Sections 5 and 6:

- PNFs including individual RU+gNB-DU, as well Wi-Fi and LiFi APs.
- VNFs including gNB-CUs, O-RAN RT RIC and Wi-Fi-LiFi real time controllers.

The main goal of the multi-WAT non-RT RIC is to enable the provisioning of the WAT services (i.e. 5GNR, Wi-Fi and LiFi services) defined in section 3.3.2, while reserving the required resources for each service. The multi-WAT non-RT-RIC will also offer services to configure the various target PNFs, and to interact with the real time controllers of the 5GNR, Wi-Fi and LiFi wireless access technologies. The multi-WAT non-RT-RIC will offer to the other MFs in the 5G-CLARITY management plane the services listed in Table 7-5. Custom interfaces based on REST APIs will be defined in WP4 to expose the mentioned services.

Table 7-5: Multi-WAT non-RT RIC Services

MF Service ID	MF Service Name	Description	Reference Specifications
Mwat_Plmn_Lcm	**PLMNID lifecycle management**	This service allows instructing a gNB-DU to radiate a PLMN ID in a set of gNB-DUs under the control of that gNB-CU. In addition, it provisions the IP end-point associated with that PLMN ID	Custom (REST)
Mwat_Snssai_Lcm	**S-NSSAI lifecycle management**	This service allows instructing a gNB-CU to radiate a S-NSSAI in a set of DUs under the control of t gNB-CU	Custom (REST)
Mwat_Wi-Fi_Ssid_Lcm	**Wi-Fi SSID lifecycle management**	This service allows instructing one or more Wi-Fi APs to radiate a SSID with a specific set of security credentials	Custom (REST)
Mwat_Lifi_Ssid_Lcm	**LiFi SSID lifecycle management**	This service allows instructing one or more LiFi APs to radiate a SSID with a specific set of security credentials	Custom (REST)
Mwat_Plmn_Res_Rsrv	**PLMNID resource reservation service**	This service allows allocating a percentage of Physical Resource Blocks (PRBs) in a set of gNB-DU under the control of the target gNB-CU, in order to carry the traffic of a given PLMN ID	Custom (REST)
Mwat_Snssai_Res_Rsrv	**S-NSSAI resource reservation service**	This service allows allocating a percentage of Physical Resource Blocks (PRBs) in a set of gNB-DUs under the control of the target gNB-CU, in order to carry the traffic of a given S-NSSAI	Custom (REST)
Mwat_Wi-	**Wi-Fi SSID resource**	This service allows allocating a percentage of airtime	Custom (REST)

Fi_Ssid_Res_Rsrv	**reservation service**	resources in one or more APs to carry the traffic of the provided SSID	
Mwat_Lifi_Ssid_Res_Rsrv	**LiFi SSID resource reservation service**	This service allows allocating a percentage of airtime resources in one or more APs to carry the traffic of the provided SSID	Custom (REST)
Mwat_5gnr_Cell_Conf	**5GNR Cell configuration service**	This service enables the configuration of a set of 5GNR cells under the control of a gNB-CU including parameters such as carrier frequency, cell identifier, transmission power and neighbour lists	Custom (REST)
Mwat_Wi-Fi_Ap_Conf	**Wi-Fi AP configuration service**	This service enables the configuration of one or more APs including parameters such as the operating channel and bonding mode, the Wi-Fi mode (e.g.VHT, HT, a/b/g) and the transmission power	Custom (REST)
Mwat_Lifi_Ap_Conf	**LiFi AP configuration service**	This service enables the configuration of one or more APs including parameters on device info, WLAN configuration and Lamp configuration	Custom (REST)
Mwat_5gnr_Topo	**5GNR topology service**	This service allows providing 5GNR physical topology including list of cells connected to a given gNB-DU instance, and list of gNB-instances connected to a given gNB-CU instance	Custom (REST)
Mwat_Wi-Fi_Topo	**Wi-Fi topology service**	This service allows providing Wi-Fi topology information including list of physical AP appliances and the capabilities of the physical radios included in each AP	Custom (REST)
Mwat_Lifi_Topo	**LiFi topology service**	This service allows providing a list of the available physical LiFi AP appliances	Custom (REST)
Mwat_Ric_Mgmt	**5GNR rtRIC management service**	This service allows providing a list of rt-RIC instances controlled by the mWAT Non-rt Controller. It enables operations such as policy management, and xAPP management	O-RAN based (REST)
Mwat_Wi-Fi_Ric_Mgmt	**Wi-Fi-LiFi rt controller management service**	This service allows providing a list of Wi-Fi/LiFi real time controllers managed by the mWAT Non-rt Controller. Enables deployment of policies into Wi-Fi/LiFi real time controllers	Custom (REST)
Mwat_Inventory	**Inventory Service**	This service allows returning a data structure containing a list of the currently active wireless services including the service identifier (PLMN ID, S-NSSAI, and SSID), their resource quota and the nodes where the service is active	Custom (REST)

7.1.4 Slice Manager

The Slice Manager is the MF in charge of deploying and operating end-to-end 5G-CLARITY slices. As defined in Section 3.2.3, these end-to-end slices integrate compute, transport and wireless services. The Slice Manager has two main functional blocks: *Slice Lifecycle Management* and *Slice Repository*. The Network Slice Lifecycle Management is responsible for managing services and underlying infrastructures, including create and allocate slice, deploy network service and provision network functions for the tenant. The Slice

Repository is responsible for storing the information of slices, such as the amount of physical and virtual resources that allocated to a slice.

Table 7-6 lists the services that the Slice Manager exposes to the other MFs of the 5G-CLARITY management plane.

Table 7-6: Slice Manager Services

MF Service ID	MF Service Name	Description	Reference Specifications
Sm_Sl_Rsrv	**5G-CLARITY Slice reservation service**	This service receives as input: • Target VIM nodes and compute resource quota • A set of gNB-DUs and PLMN ID to be instantiated • A set of Wi-Fi and LiFi APs and SSIDs to be instantiated Based on this input the Slice Manager configures the resource reservation in the VIM, and instantiates the wireless services in the target WATs	Custom (REST)
Sm_Sl_Actv	**5G-CLARITY Slice activation service**	This service receives as input an NSD identifier, which has been previously on-boarded on to the NFVO, and a 5G-CLARITY slice identifier, and instantiates the target NSD on the VIM resources provisioned for that slice	Custom (REST)
Sm_Sl_Inv	**5G-CLARITY Slice inventory**	This service returns a data structure containing the list of slices currently deployed in the system along with the corresponding slice descriptors	Custom (REST)

7.2 Data processing and management

5G-CLARITY Data Processing Management defines a framework that allows retrieving, transforming and storing data generated at the 5G-CLARITY system, including telemetry and monitoring data, state data and other type of persistent data (e.g. policy, subscription, profiling data). These data are streamed from multiple sources into this framework, from where they can be observed and acted upon by data consumers, including Service and Slice Provisioning MFs (see Section 7.1) and AI/intent engines belonging to the Intelligence Stratum (see Chapter 8). 5G-CLARITY Data Processing and Management concept is based on two building blocks: data semantic fabric (Section 7.2.1) and data lake (Section 7.2.2). Figure 7.4 captures the interaction between both blocks.

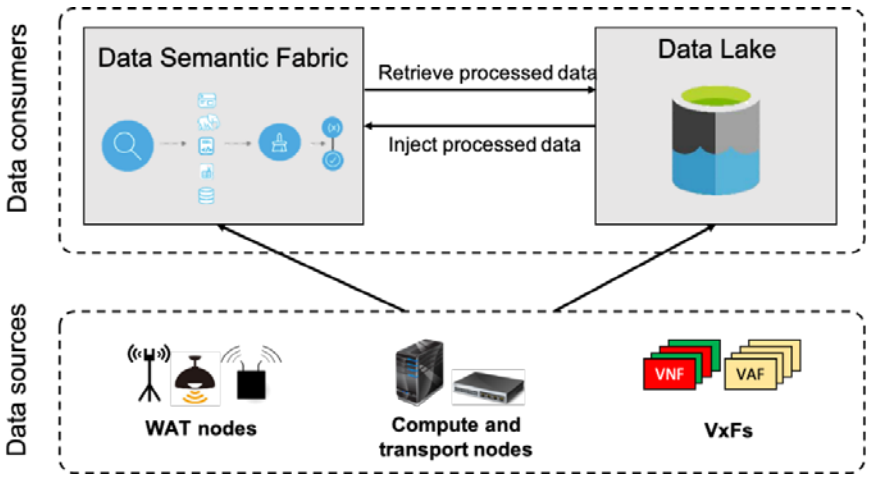

Figure 7.4: MFs from data and processing and management group.

7.2.1 Data semantic fabric

The data semantic fabric builds up the 5G-CLARITY data pipeline, which constitutes the core block of the 5G-CLARITY data processing and management framework. It allows consolidating data from a wide variety of *sources* and turn them into useful information for *consumers*, by applying necessary processing on the collected data before their transmission and storage. For an effective data ingestion, the 5G-CLARITY data pipeline includes a set of logical entities, each with a well-defined functionality: collector, aggregator and dispatcher.

The **collector** is the pipeline node responsible for data harvesting. Examples of sources from which collector can retrieve data include infrastructure equipment (e.g. network devices, compute nodes), instances of manageable network entities (e.g. network slices/services, VNFs, probes) and databases. Every data source has associated a well-defined class, which provides a complete description of this source. A class consists of two types of information:

- Endpoint of the source, including the URI (e.g. path, host, port) and the corresponding credentials (e.g. user, password, method). This information allows the collector to establish a subscription with the source.
- The data items (e.g. in-octets, in-pks, in-unicast-pkts, MTU size) that can be retrieved from this source by a subscribed collector. These data items, which conveys the semantics of the data source, are typically structured into a standards-based YANG model. YANG is a data modelling language based on the structure of management information, next-generation system, being used to model semantics and organization of configuration and state data manipulated by the NETCONF protocol. To retrieve data from a source, using either pulling or push-based (streaming) mechanisms, the collector subscribes to specific data items it needs, by using the YANG model embedded in the class.

The **aggregator** is the pipeline node in charge of manipulating and combining individual data collected together, making them available for their consumption. This processing is done according to some rules (e.g. arithmetic operations, filtering, thresholding).

Finally, the **dispatcher** is the pipeline node which sends aggregated data out to the target destination. This destination is where the data is stored for their consumption. In 5G-CLARITY, this destination is the data lake (see Section 7.2.2).

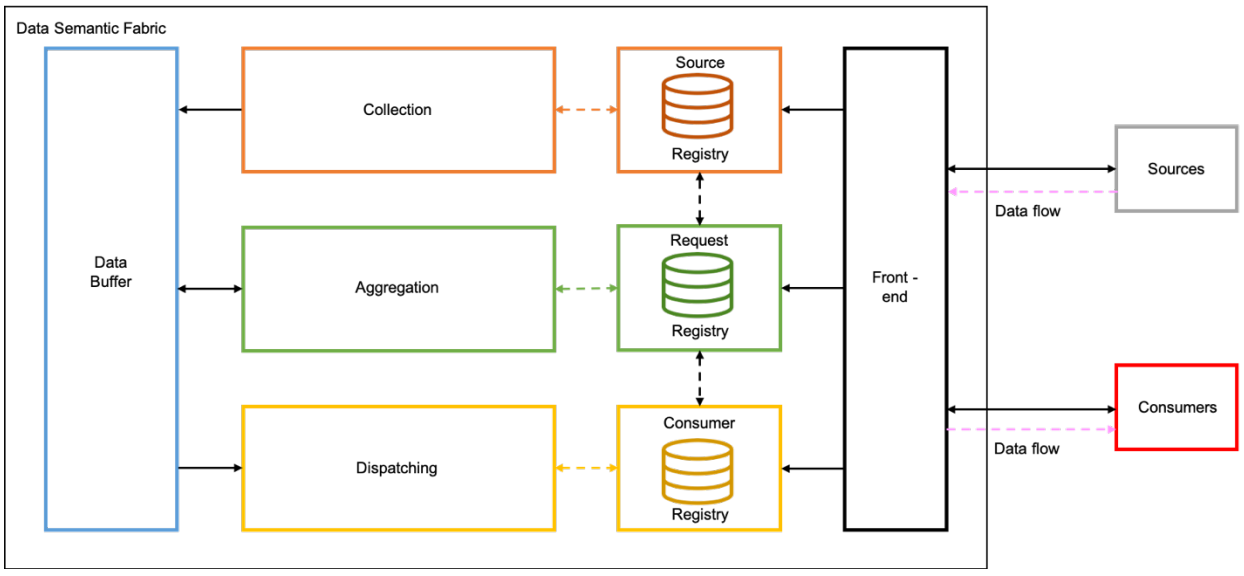

Figure 7.5: Data semantic fabric.

Figure 7.5 illustrates the data semantic fabric. In addition to the three pipeline modules, the fabric components include a data buffer, a front-end and a collection of registries. The **data buffer** is used to temporarily store collected data while it is being processed by the aggregation module. In essence, this buffer allows the temporary collection and storage of data before it is moved towards the data lake. The **front-end** defines a single-entry point for data consumers. It allows the data consumers to issue service requests towards the data semantics fabric. Finally, there are three registries including one (data) source registry, one (data) consumer registry and one request registry:

- The **source registry** keeps a directory with all the data sources the collector is subscribed to, and their associated classes.
- The **consumer registry** keeps information about the different consumers, i.e. the fabric customers that are authorized to issue service requests through the front-end. The information stored for every consumer includes:
    - An UUID, which uniquely identifies the consumer in the data semantics fabric.
    - An endpoint, e.g. URL, which specifies where the consumer is reachable. This endpoint allows the fabric to keep communication open with the consumer, e.g. for dispatching aggregated data.
    - Data schema, e.g. JSON, which specifies the format the consumer expects to receive aggregated data in.
- The **request registry** stores all the service requests that the fabric receives from the registered consumers. Every request conveys the following artifacts described below.
    - Selector (mandatory): it allows the consumer to specify the raw data it needs from individual sources. This is done by selecting the data items from the corresponding classes, stored in the source registry. The selector artifact triggers corresponding actions on the collection module.
    - Processing (optional): it allows the consumer to request for the aggregation of selected data, specifying the set of rules to be applied for this aggregation. This artifact triggers corresponding actions on the aggregation module. It is worth noting that the rules made available for selection depends on the aggregator capabilities.
    - Target destination (mandatory): it allows the consumer to specify where the aggregated data shall be transmitted and stored, for their further consumption. This artefact triggers corresponding actions on the dispatching module.

Besides allowing a better traceability in the system (for auditability purposes), keeping consumer-triggered requests aside in a separate registry boosts efficiency in terms of data collection and storage. Indeed, only requested data is managed in the data semantics fabric. This allows curbing malpractices, consisting in collecting and storing high amount of data, most of them down the drain, as it is not related to what the consumers really want / need.

Table 6-2 describes the management services exposed by the data semantics fabric.

Table 7-7: Data Semantic Fabric Services

MF Service ID	MF Service Name	Description	Reference Specifications
DSF_Src_Mgmt	**Source management**	This service allows manipulating (create, update, read, delete) the entries in the source registry. Each entry contains the class corresponding to a different data source. The start (termination) of a subscription	Custom (REST)

		- with a data source deployed on the 5G-CLARITY system - translates into the insertion (removal) of a new (existing) subscription into (from) the source registry.	
DFS_Cns_Mgmt	**Consumer management**	This service allows manipulating (create, update, read, delete) the entries in the consumer registry. Each entry contains the information corresponding to a different data consumer.	Custom (REST)
DFS_Pipe_Prov	**Data pipeline provisioning**	This service allows issuing individual service requests for the creation of data pipelines and retrieving information about their state throughout their lifecycle.	Custom (REST)
DFS_Cap_Ex	**Capability exposure**	This service allows providing file version system capabilities. Examples of these capabilities include in-built rules for data aggregation. It is worth noting that different (more recent) versions may bring different (more advanced) set of capabilities.	Custom (REST)

7.2.2 Data lake

A data lake is an architectural pattern of large data repository that allows the storage of both structured and unstructured data and is scalable based on user or organization needs. Data Lake combines a large-scale storage repository with variety of high-performance processing engines that are usually virtualized. Large scale and independent scalability of processing and storage allows a data lake to store data as it is produced without pre-processing and processed upon-access. This approach allows flexible usage of produced data and it enables organizations run different types of analytics from dashboards and visualizations to big data processing, real-time analytics, and ML.

The data lake allows running analytics without the need to move data to a separate analytics system. It also allows data-based access controls that enable multiple roles within the same organization or external organizations to gain access to a specific data for a specific time. For instance, in one organization data scientists, data developers, and business analysts are able to access the same underlying data with their preferred tools and frameworks. This includes open source frameworks such as Apache Hadoop [90], Presto [91], and Apache Spark [92], and commercial products from data warehouse and business intelligence vendors.

5G-CLARITY uses data lake to enable prototyping and application development in the intelligence stratum. Figure 7.6 presents basic functional building blocks for data lake architecture. The 5G-CLARITY Intelligence Engines are consumers of the 5G-CLARITY data sources through data lake architecture. Telemetry originating from the network telemetry bus is steered to the storage via data dispatcher component. The data originating from the network may be either structured or unstructured from the data lake perspective, which means that it may have schema association, or it may be object data from network nodes. The dispatcher function is directing the data to a correct pre-defined data storage location.

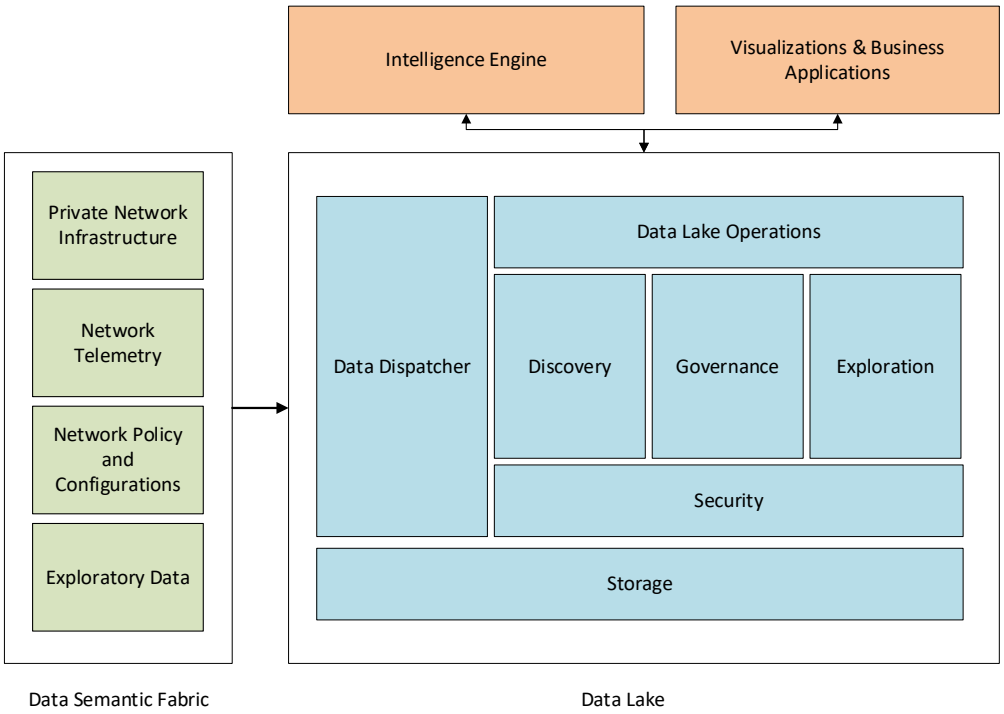

Figure 7.6: Functional architecture of the data lake.

The Intelligence Stratum may use the data lake for exploration of the stored data, or it may use computational capabilities of the data lake to pre-process the data on access. The AI-engine and data lake may be either in the same network / cloud domain or they may be in different network domains.

Table 7-8 describes the management services exposed by the data lake.

Table 7-8: Data Lake Services

MF Service ID	MF Service Name	Description	Reference Specifications
DataLake_Ingress_Service	**Data Lake Ingress Service**	This service allows ingestion of structured and unstructured data to data lake.	Custom
DataLake_Exposure_Service	**Data Lake Exposure Service**	This service exposes data lake storage to authenticated users.	Custom
DataLake_Data Security_Service	**Data Lake Data Security Service**	This service maintains data access policies of the data lake.	Custom
DataLake_Discovery_Service	**Data Discovery Service**	This service allows data discovery queries to the data lake metadata.	Custom
DataLake_Exploration_Service	**Data Exploration Service**	Data Exploration Service Allows User to gain access to specified data in data lake in order to explore it.	Web service (e.g. Jupyter Notebooks)

7.3 Cloud native support

7.3.1 Authentication and registration function

The Authorization and Registration Function (AuRF) is key in the SBMA. It supports the following functionality:

- **MF registration and de-registration**. The AuRF maintains up-to-date information of available MF instances and their supported services. Every MF instance registers at the AuRF the list of MF services it exposes. The MF instance may register this information with the AuRF when the MF instance becomes operative for the first time. During the registration operation, the MF instance provides a MF profile that is maintained in the AuRF. The MF profile consists of general parameters of the MF Instance (e.g. MF Instance ID, MF Type, Fully Qualified Domain Name (FQDN) / IP address, identification of stored data/information) and also the parameters of the different services exposed by the MF instance. The MF instance should also de-register from the AuRF when it is about to (gracefully) shut down or disconnect from the network in a controlled way. If an MF instance becomes unavailable or unreachable due to unplanned errors (e.g. MF crashes), an authorized entity should deregister the MF instance from the AuRF.
- **MF discovery**. This functionality enables one MF to discover a set of MF instances fulfilling certain criteria (i.e. instances providing a given service, instances of a target MF type). The AuRF receives a discovery request from a MF instance and provides the information of the available MF instances compliant with the criteria. Based on this information, the requester MF can select the MnF instance(s) to which he wants to communicate.
- **MF authorization**: to ensure the MF consumer is authorized to access one or more services from another MF (service producer).

Table 7-9: AuRF Services

MF Service ID	MF Service Name	Description	Reference Specifications
Aurf_Mf_Mgmt	**MF Management**	This service allows a MF instance to*: i)* register, update or deregister its profile in the AuRF; *ii)* subscribe to be notified of registration, deregistration and profile changes of MF instances along with their MF services; *iii)* retrieve a list of MF instances currently registered in the AuRF, or the MF profile of a given MF instance.	Custom (REST)
Aurf_Mf_Disc	**MF Discovery**	This service allows a MF instance to discover services offered by other MF instances, by querying the AuRF. In this query, the IP address(es) or FQDN of the MF instance(s) or service(s) matching certain input criteria are provided.	Custom (REST)
Aurf_Mf_Auth	**MF Access Token**	This service provides OAuth2 Access tokens (based on [93]) to a MF consumer for authorization purposes. The MF consumer can subsequently use the access token to a MF service producer that is authorized to use the service.	Custom (REST)

Existing cloud services like Hashicorp Consul [94] can be used to implement the AuRF. Consul consists of an agent device that needs to be embedded into each MF, and one or more agents selected to act as Consul servers. The Consul agents check the API services offered by each MF, including the ability to perform health checks, and communicate the available services to the Consul servers. A Consul agent can then discover all services in the network by issuing a DNS query to the Consul server or another Consul agent. For example, in a cloud native deployment of the 5G-CLARITY management stratum there could be at any time multiple instances of the Slice Manager MF. Other MFs would issue a DNS query to the Consul server for the Slice Manager which would return the list of IP address of the current Slice Manager instances. In this way Consul

can also act as a load balancer function across multiple instances of the same function, as well as performing authentication functions.

7.3.2 Distributed data storage

In 3GPP systems, the 5GC supports a Data Storage architecture. The Unified Data Repository (UDR) is the master database. The Unstructured Data Storage Function (UDSF) is introduced to store dynamic state data. The distributed data storage needs to support storing of dynamic state of the management function, which allows it to be stateless and scale dynamically by the underlying cloud management system. Any network function within 5GC can use UDSF to store and retrieve unstructured data, i.e., data that is not defined in 3GPP specifications. The UDSF is deployed in the same network where the control plane is located and the same UDSF may be shared by all the NFs in the PLMN to store/retrieve their respective data or an NF may have its own UDSF depending on operator configuration. The UDSF supports UDSF Data Repository Service API that is specified in [95].

The distributed data storage supports storage of both dynamic and static data storage. Below are things to consider in the design of 5G-CLARITY storage:

- Support of large sets of unstructured data
- Use of adequate storage schemes (e.g. Object, NoSQL, Distributed SQL) for enabling:
    - Scalability
    - Analytics support
    - Fast data retrieval
    - Cost and resource
- Supports different deployment options such as:
    - Private network RAN edge
    - Network edge (behind UPF)
    - Operator data centre inside operator network
    - Integration with data lake object storage (i.e. cloud / private cloud / regional cloud)
- Supports searching requirements for static/dynamic data.

7.4 External access mediation

7.4.1 Mediation function

The Mediation Function provides a single-entry point for external consumers. The logic of this MF provides means for secure communication between internal MFs (i.e. MFs belonging to the 5G-CLARITY Management and Orchestration stratum) and external MFs (i.e. MFs from the intelligence stratum or MFs from the 3GPP management system of a public NOP). From a conceptual viewpoint, the mediator function acts as a combination of UDR and NEF functionalities as defined in the 5GC.

According to the above rationale, external MFs do not have direct access to all the MF services, but are mediated by the MF Exposure Function that policies what each external entity is allowed to access and maybe can provide higher level abstraction APIs for external services. The 5G-CLARITY mediator function provides the following capabilities:

- Authentication and Security
- Monitoring – Mediator Function tracking
- Dynamic Routing - dynamically routing requests to different computing nodes

- Load Shedding - allocating capacity for each type of request and dropping requests that go over the limit.
- Static Response handling – mediator function may process some responses directly and not redirect them

The 5G-CLARITY Mediator Function may be implemented by using API gateway technologies such as Zuul [96] could be used to implement the exposure function if underlying communication between Management Functions is REST based. If underlying communication is based on message broker, then REST to message broker gateways, e.g. [97] for NATS, could also be implemented so that external client see a REST end-point and the gateway is connected to the message bus.

8 Intelligence Stratum Design

The 5G-CLARITY intelligence stratum includes all the assets that drive automation and data-based intelligence on 5G-CLARITY service and network operation. This stratum provides an amalgam of AI- and intent-based services, which are registered and published into a unified marketplace. Components from both the 5G-CLARITY Network and Application Function stratum (Chapter 6) and the 5G-CLARITY Management and Orchestration stratum (Chapter 7) can gain access to this marketplace, browsing available services, subscribe to those they are interested in and consume them on-demand, following Platform as a Service consumption patterns. According to the Intelligence as a Service model introduced in Section 3.3 and detailed in Annex B, VxFs and MFs from any NOP (either public NOP or private NOP) can subscribe and invoke these AI-and intent-based services.

The requirements for the intelligence stratum are listed in Table 8-1.

There are two core components of the 5G-CLARITY intelligence stratum:

- AI engine, which provides hosting and management of ML services
- Intent engine, which provides point of contact to and from the AI engine as well as a layer of abstraction towards the consumer of the AI functionalities.

An overview of the intelligence stratum with AI engine and Intent engine is shown in Figure 8.1, along with a high-level view on the north and southbound interfaces.

Table 8-1: 5G-CLARITY Intelligence Stratum Requirements

Requirement ID	Requirement Description
CLARITY-INTS-R1	The 5G-CLARITY intelligence stratum can be hosted by the private site or by a 3rd party site (e.g. hyperscaler cloud site).
CLARITY-INTS-R2	The 5G-CLARITY intelligence stratum can be offered with PaaS capabilities, so both public and private NOPs can create instances of intelligence stratum services on their corresponding administrative domains.
CLARITY-INTS-R3	The 5G-CLARITY intelligence stratum shall leverage Machine Learning (ML) models to support intelligence management of NOP managed functions, including VxFs (i.e. functions from the network and application function stratum) and MFs (i.e. functions from the management and orchestration stratum).
CLARITY-INTS-R4	The 5G-CLARITY intelligence stratum shall provide a point of access for ML services to consume data from the NOP managed functions, and forward recommended configurations to corresponding functions.
CLARITY-INTS-R5	The 5G-CLARITY intelligence stratum shall provide ML designers a process or interface to manage the lifecycle of ML models
CLARITY-INTS-R6	The 5G-CLARITY intelligence stratum shall expose a communication interface towards the end user that simplifies the management of the 5G-CLARITY platform using intents, including intent-based network configuration and intent-based usage of available ML services.
CLARITY-INTS-R7	The 5G-CLARITY intelligence stratum shall expose an intent management interface through which the intent lifecycle can be controlled, including creation and removal.

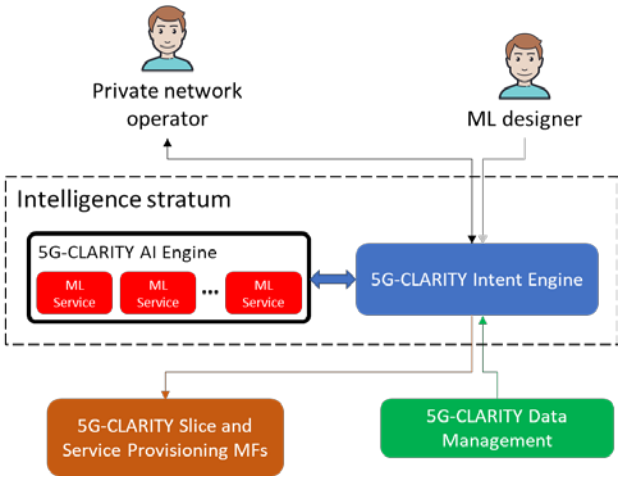

Figure 8.1: Architectural overview of the 5G-CLARITY intelligence stratum with its two main components, the AI-engine and the Intent engine. For the sake of simplicity, only private NOP is captured as consumer.

8.1 AI-engine

The primary task of the AI engine is to host ML models. These models provide ML functionality that can be pushed towards subscribed NFs and MFs, assisting these functions in their day-to-day operation. Designed by an operation support provider (e.g. software engineer), created models need to be individually trained before their on-boarding into the AI engine. In this regard, two use cases can be envisioned:

A. Training of individual ML models takes place outside the AI engine, either in the operation support provider's premise or at a 3rd party cloud computing environment, e.g. hyperscaler cloud site. The first case corresponds to a local model training, while the second one corresponds to a remote model training.

B. Training of individual ML models takes place inside the AI engine, making use of data available in this engine. This data is collected from the 5G-CLARITY data management framework (Section 7.2), specifically from the data lake or data semantics engine.

ML models that are trained outside the AI-engine (use case A) require manual intervention to their lifecycle management, as these standalone models typically do not have access to training data. On the other hand, ML models that are trained inside the AI engine (use case B) can run in an online training mode and recurrently monitor and update themselves. For the models that require lifecycle management, the ML Lifecycle Manager enables ML models to be deployed, updated and removed through exposed services (Table 8-2).

The deployed ML models take data from the 5G-CLARITY data management framework (ML input), process them according to model logic (ML processing) and derive insights (ML output). These insights consist in predictions and configurations that are ultimately forwarded to subscribed NFs and MFs. These NFs/MFs and 5G-CLARITY data management framework shall expose services that enable the retrieval of data and forwarding of network configurations. These services will be accessed by ML models through the interface provided by the Intent engine (Section 8.2).

Finally, the deployed ML models that are ready for use will be registered in the ML Service Registry, from where they will be discoverable from outside the AI engine, and executed by the network operator or by other network services that want to utilise ML functionalities that are hosted by the AI engine. Most of the communication to and from the AI engine is envisioned to go through the Intent engine (Section 8.2) that provides a common point of communication and an abstraction towards the consumers of the AI engine.

Table 8-2 lists the available services offered by the AI engine, mainly supporting the lifecycle management of ML models.

Table 8-2: AI Engine Services.

MF Service ID	MF Service Name	Description	Reference Specifications
AIEngine_Ml_Mod_Mgmt	**ML Model Management service**	This service allows the deployment, update and removal of trained ML models by the ML designer. The ML model shall be provided in containerised form. It will consume input data from the Data Management and send output to various network functions (e.g. Slice Manager). Deployed ML models can be updated after they have been retrained, e.g. when additional data has become available. At the end of their life span, deployed ML models can be removed from the AI engine.	Custom (REST)
AIEngine_Ml_Mod_Register	**ML Model Registry service**	The deployed ML model can be registered in the ML Service Registry. In the case of pre-trained models, this service will allow the intent engine to discover ML services and connect them to a given intent.	Custom (REST)
AIEngine_Ml_Mod_List	**ML Model List service**	Lists the ML models that are currently deployed and ready for execution.	Custom (REST)
AIEngine_Ml_Mod_Run	**ML Model Execution service**	Runs a deployed ML model that is available in the ML Service Registry. Once executed, the deployed ML model may run as a once-off or recurrently.	Custom (REST)
AIEngine_Push_XApp	**xApp Pushing service**	Certain (real-time ready) ML models can be pushed down into the near-RT-RIC as xApps.	Custom (REST)

8.2 Intent engine

The Intent engine fulfils the role of communication interface from and to the AI engine and offers a level of abstraction towards the 5G-CLARITY intelligence stratum users, which are public and private NOPs. A typical use case for the intent engine is to translate a NOP query to create a new 5G-CLARITY slice (e.g. written in structured English language) with specific parameters into a Slice Manager configuration, perhaps with the help of an ML model to optimize the resource usage.

The intent engine provides an intent-based interface to and from the AI engine and implements an intent-driven process. This process takes an intent, translates it into an operational context (configurations and/or actions), monitors the operational context against the original intent, and reports on the intent execution. The operational context is represented by Intent Providers, being part of the process. These components can either implement all required functions (e.g. a full access to a database or repository) or use proxies to communicate with a managed resource (e.g. a proxy that communicates with the Slice Manager to manage network slices).

Intents and providers can be added to and removed from the engine at runtime. Both artefacts come with metadata that assists the engine in matching an intent to a provider that is able to execute it. 5G-CLARITY intelligence stratum users or automated processes can request lists and descriptions of intents and providers

in different levels of details and several representations.

Levels of details range from simple (e.g. simple list, overview description) to advanced (detailed lists with filters, detailed descriptions of properties). Available representations include structured English (natural language-like), translated (aiming to provide English sentences), several JSON and YAML structures (using different schemas), tree structures (as plain tree or XML), hash maps (key/value maps), or raw (the internal representation inside the engine). Table 8-3 lists the services offered by the Intent engine.

Some examples of services that the Intent engine may consume from the Slice and Service Provisioning MFs include Nfvo_Ns_Lcm, Nfvo_Vnf_Mgmt (Section 7.1.1), Tc_Virt_Net, Tc_Tsn_Conf (Section 7.1.2), Sm_SlR_Rsc, Sm_Sl_Actv and Sm_Sl_Inv (Section 7.1.4).

Table 8-3: Intent Engine Services

MF Service ID	MF Service name	Description	Reference Specifications
IntentEngine_NtntProv_Mgmt	**Intent Provider Management service**	For an intent to be fulfilled, it needs to be matched to an intent provider. This service allows the addition, updating and removal of intent providers for the intent engine to match intent descriptions against. Update intent providers for when intent providers change. Many AI models are intent providers, which may be updated and replaced from time to time. Intent providers may be removed when they are outdated or no longer desired.	Custom (REST)
IntentEngine_ NtntProv _List	**Intent Provider List service**	Lists currently registered and available providers. The list can be simple (names) or extended (provider names and supported intents).	Custom (REST)
IntentEngine_ NtntProv _Describe	**Intent Provider Description service**	Returns a description of an intent provider. This description can be simple (overview of the provider's functionality), enhanced (description with overview of supported intents), or detailed (description of a particular intent and how the provider supports it).	Custom (REST)
IntentEngine_ NtntProv _Run	**Intent Execution service**	Runs a received intent on a provider. The provider can be explicitly named. When no provider is named, the intent engine will try to match the provided intent description with an existing intent provider, who executes the intent description. It may then return a result or run recurrently.	Custom (REST)
IntentEngine_Ntnt_Terminate	**Intent Terminations service**	Terminates an intent that is currently executed on the engine.	Custom (REST)
IntentEngine_Ntnt_Status	**Intent Status service**	Returns the status of a running intent.	Custom (REST)
IntentEngine_Ntnt_Get_Telemetry	**Telemetry Data Retrieval service**	Get telemetry data from the Telemetry Collector. This service allows ML models to retrieve network data.	Custom (REST)
IntentEngine_Ntnt_Push_Config	**Network Configuration Forwarding**	Forward network configurations to various network elements (e.g. Slice Manager). This service allows ML models to push network configurations into the network.	Custom (REST)

	service		
IntentEngine_Ntnt_Mgmt	**Intent Management service**	Allows for the deployment of intents on the engine that are ready to be executed (run) and to remove intents from the engine. If the intent is currently executed, it will be terminated before removal.	Custom (REST)
IntentEngine_Ntnt_List	**Intent List service**	Lists currently deployed intents. The list can be simple (intent names) or extended (intent names and matching providers).	Custom (REST)
IntentEngine_Ntnt_Describe	**Intent Description service**	Describes an intent. The description can be simple (overview of the intent) or detailed (explaining all details of the intent)	Custom (REST)

8.2.1 Intent specification

An intent is a declaration of an operational goal, with optional required properties, specified in a declarative manner. The goal can be stated as "who wants what" in form of an action sentence. For example: "The NOP wants to create a new network slice". Required properties can then be added to the goal as required. For instance: "with Wi-Fi and LiFi, but without RAN". We call the goal the *primary part* of the intent and additional requirements the *secondary part* of the intent.

The primary part is not negotiable and can furthermore be used to categorise and name intents. The name of an intent is then the verb (e.g. "create") and the object (e.g. "network slice"). Providers can then state which intents they do support by stating intent names (e.g. "create—network-slice"). This method allows the engine to select a provider based on the intent name.

The secondary part of an intent contains all further requirements. It can be negotiable as a whole or in parts. We can use the pattern of 5WH (six questions as who, when, where, what, why, and how) for the secondary part. Additional language constructs (e.g. "with" or "without", adverbs) can also be used. A provider can then state which of the secondary parts of an intent it supports. This method allows the engine to select a provider based on the intent name *and* the providers' capabilities.

For each intent, we can define a template with all possible (or permitted) requirements (secondary part), including potential dependencies (e.g. using logical operations like OR and AND on requirements). Such templates can be used by a user interface (graphical or otherwise) to specify (create) intents.

A formal representation of an intent is then a set of two hash maps: one for the primary part and one for the secondary part. The map of the secondary part can be flat (a list of requirements) or tree-like (structured and also dependent requirements). The intent can therefore be formally represented as either a set of key/value pairs (simplest form) or a composite hierarchical structure. In the simple form, the semantic of keys have to be overloaded with separators to allow a translation into a hierarchy. This method allows to represent an intent in any language that has a data type or other means to declare a hash map or key/value map.

8.2.2 Providers and proxies

A provider implements the execution of an intent. They are used in the intent engine to allow for flexibility and easy extensibility (by simply adding or removing providers). Each provider must state which intent(s) it can support (primary part) and for each supported intent which requirements it supports (secondary part). A provider can support any number of intents in any form.

Proxies can be used by a provider to further decouple the internal mechanism of the intent engine from the underlying network. They are only visible to the provider, i.e. not visible to the engine. For instance, a

provider for intents that manage data might use several proxies to access different data types and sources.

8.2.3 Intent life cycle

An intent can be authored, deployed, and executed. An intent is specified in the authoring phase. This can be done using the dashboard or some other authoring application provided. Authoring can be supported by intent templates. The final intent (specification) can then be stored and if required also analysed or validated outside the engine.

The next step is to take an intent specification and deploy it in an intent engine instance. The engine will take the intent, validate it (against the general formalisation of primary and secondary part), and prepare it for execution.

The next step is to execute the deployed intent in the engine. The intent can be executed on a specified provider or the engine can select an available provider that best matches the intent specification. The intent is then handed over to a provider, which will translate the intent specification into its operational context, create configurations or actions, deploy the configurations or run the actions or forward the actions to another component, evaluate the results, and report back on the intent execution. This process can be a one-off (execute, evaluate, report) or a closed loop continuously monitoring the operational context, validating it against the original intent, and running repair actions if required.

The final steps allow to terminate an intent (i.e. stop its execution but have it still deployed on the engine) and undeploy an intent (remove it from the engine).

9 Enablers for NPN-PLMN Interworking

So far, 5G-CLARITY system architecture description has focused on private environments under the exclusive management of a private NOP, i.e. SNPNs. However, a key feature of 5G-CLARITY system is also the ability to facilitate public-private network integration, thereby allowing a seamless operation of PNI-NPNs in which both public and private NOPs are involved. In this chapter, we give an insight on the integration of public and private networks in 5G-CLARITY system (Section 9.1), throughout different deployment scenarios (Section 9.2) and an analysis on their interoperation at the management and orchestration plane (Section 9.3). This integration needs to be accompanied with security mechanisms (Section 9.4) that ensure privacy and trustworthiness between the corresponding administrative domains.

9.1 Overview of deployment options

Table 9-1 summarizes the main NPN deployment options, which have been extracted from 3GPP specification [50], and the scenarios that could most benefit from them.

Table 9-1: NPN Deployment Options and Their Primary Target Scenarios.

Deployment option		Description	Primary Target Scenarios
SNPN		NPN that does not rely on network functions provided by a PLMN.	Large-sized enterprises requiring wireless access with stringent security, performance, and reliability constraints.
PNI-NPN	**DNN**	NPN whose deployment is supported by a PLMN using Data Network Naming (DNN) mechanism.	Small and medium-sized enterprises requiring support for private services with lenient security, performance, and reliability constraints, though some 5GC NFs might be deployed on-site to reduce latency.
	Slicing	NPN whose deployment is supported by a PLMN using 3GPP network slicing mechanism. It offers a higher isolation between private and public services than the use of dedicated DNNs.	Small and medium-sized enterprises requiring support for private services with lenient security, performance, and reliability constraints, though some 5GC NFs might be deployed on-site to reduce latency.
MOCN in NPNs	**SNPN(s) and PLMN(s)**	In-house RAN shared through MOCN between one or multiple SNPN(s) and one or multiple PLMNs.	Private venue of interest for public network operators to extend their footprints. Private site supporting private services that need to be deployed on segregated SNPNs because they require strong isolation.
	SNPN(s) and PNI-NPN(s)	RAN shared through MOCN between one or several SNPN(s) and one or several PNI-NPN(s).	Industrial states, in which parts of a public RAN is shared among several enterprises, some of them enabling private services via SNPNs and others via PNI-NPNs.
	PNI-NPN(s) and PLMN(s)	RAN shared through MOCN between one or several PNI-NPN(s) and one or several PLMN(s).	Big venues of interest for public network operators to extend their footprints. Big venues hosting several enterprises or where several actors enable private services via PNI-NPNs and need to share parts of the RAN
SNPN + PLMN		SNPN accessing the PLMN services as an untrusted/trusted non-3GPP network and vice versa.	SNPNs and PLMNs requiring access to public and private services, respectively. Mobility services / access to public services

9.2 NPN-PLMN integration: user and control plane

For the analysis of this integration in terms of data forwarding (user plane) and signalling (control plane) mechanisms, a characterization on the location and ownership of participant NFs, as well as on the key technology enablers are needed. The governance rules of these features in the different deployment options specified earlier are summarized in Table 9-2.

Table 9-2: User and Control Plane Analysis of NPN Deployment Options

Deployment Option		NF Location			NF Nature			Key Enablers
		RAN	5GC-CP	UPF	RAN	5GC-CP	UPF	
SNPN		On-premise	Both valid	Both valid	Both valid	private	private	E1
PNI-NPN	DNN	On-premise	Off-premise	Both valid	Public	Public	Public	E2; E3; E6
	Slicing	On-premise	Off-premise	Both valid	Public	Public	Public	E2; E3; E5
MOCN in NPNs	SNPN(s) and PLMN(s)	On-premise	S: Both valid P: Off-premise	S: Both valid P: Off-premise	Both valid	S: Private P: public	S: Private P: Public	E1; E4; E8
	SNPN(s) and PNI-NPN(s)	Off-premise	S: Both valid P: Off-premise	S both valid P: Off-premise	Public	S: Private P: Public	S Private P: Public	E1; E2; E3; E4; E5 or E6; E8
	PNI-NPN(s) and PLMN(s)	On-premise	Off-premise	out-of-premises	Public	Public	Public	E2; E3; E4; E5
SPN + PLMN		S: On-premise P: Off-premise	S: Both valid P: Off-premise	S: Both valid P: Off-premise	SNPN: P P: Public	S: Private P: Public	S: Private P: Public	E1; E7

NF location refers to the place where each NF is deployed. According to this criterium, a NF can be classified as an on-premise NF, when the NF is placed within the logical perimeter of the private site, or as an off-premise NF. When both options are enabled, the NF placement depends on each specific use case. Although the NG-RAN (labelled as "RAN" in Table 9-2) and control plane of the 5GC (labelled as "5GC-CP" in Table 9-2) have several constituent NFs, for simplicity, we consider the NFs of each to share the same location in Table 9-2. However, this does not preclude that, in real scenarios, we might find them geographically separated, if required.

NF nature specifies the public or private origin of individual NFs. A public (private) NF is a NF that is (not) provided by the PLMN. When both options are possible, it means the NF nature depends on the specific scenario. Again, for simplicity, we consider that all the constituent NFs of the NG-RAN and 5GC-CP share the same nature.

Key enablers, providing means for realizing the NPN deployment options. The following enablers apply:

E1. 3GPP NPN support: all features included in 3GPP Rel-16 and beyond specifications for supporting SNPNs.

E2. 3GPP PNI-NPN support: all features included in 3GPP Rel-16 and beyond specifications for supporting PNI-NPNs.

E3. CAGs: 3GPP mechanism to prevent a group of UEs from accessing certain cells within a PNI-NPN. For instance, CAGs might be used to apply mobility restrictions to a particular set of users within a given

network slice's coverage area.

E4. MOCN: Network sharing architecture, in which only the RAN (infrastructure, network functions, and spectrum) is shared in 5GS. It enables multiple 5GCs from different 5G networks to share the NG-RAN.

E5. 3GPP network slicing: 3GPP feature whereby multiple logical network partitions can be provisioned for multiple services / tenants using a single 5GC.

E6. Data Network Name (DNN): it is equivalent to the Access Point Name (APN) in the Evolved Packet System (EPS) architecture. In the context of NPNs, it is used for realizing a PNI-NPN as a group of dedicated PDU sessions, conveying their traffic flows towards a dedicated DN by applying appropriate QoS policy.

E7. Untrusted/Trusted Non-3GPP access: functionality to integrate different WATs in a 5GS. In the context of NPNs, it is used for enabling SNPN access to the PLMN services and vice versa.

E8. 5G-CLARITY slicing: a mechanism whereby a multi-technology physical infrastructure is segregated into a number of logical resource partitions, for their delivery to different tenants. This differs from 3GPP network slicing, where segregation is not applied at the infrastructure layer (for multi-tenancy support), but at the network function layer (for multi-service support).

In the following, selected scenarios scoping NPN-PLMN integration are described. Other scenarios within the scope of 5G-CLARITY, but focused on SNPN category, are presented in Annex D – 5G-CLARITY Slicing Enabled SNPNs.

9.2.1 In-house NPN RAN sharing through MOCN

An NG-RAN providing on-premises coverage and operated by the site owner is shared between one or several NPNs and one or several PLMNs by means of MOCN, as illustrated in Figure 9.1. The 5GC from the SNPN(s) are deployed on-premises, whereas the 5GC belonging to PLMN(s) are allocated beyond the private site. The PLMNs benefit from RAN sharing to extend their footprint within the private premises easily and affordably.

Exemplary use cases for this scenario are the following:

- A 5G smart stadium broadcasting on-the-fly multimedia content through an SNPN to the local audience for an enhanced match-viewing experience (e-g. beyond 4K resolution, 360-video experience, director choice). The private RAN might be shared with multiple PLMNs from different MNOs, allowing the audience to access the public services, and a PNI-NPN provided as dedicated DNNs by an MNO, enabling an over-the-top service provider to stream the game live.
- A factory of the future, with two separate SNPNs for different services, typically consumed by distinct enterprise departments. For instance, an SNPN might provide 5G access to the IT department and another to the OT department.
- An airport, in which there might be a separate SNPN per airline (e.g., Iberia, Lufthansa, and Vueling). The shared infrastructure, including the RAN, is managed by an infrastructure service provider (e.g., Aena).

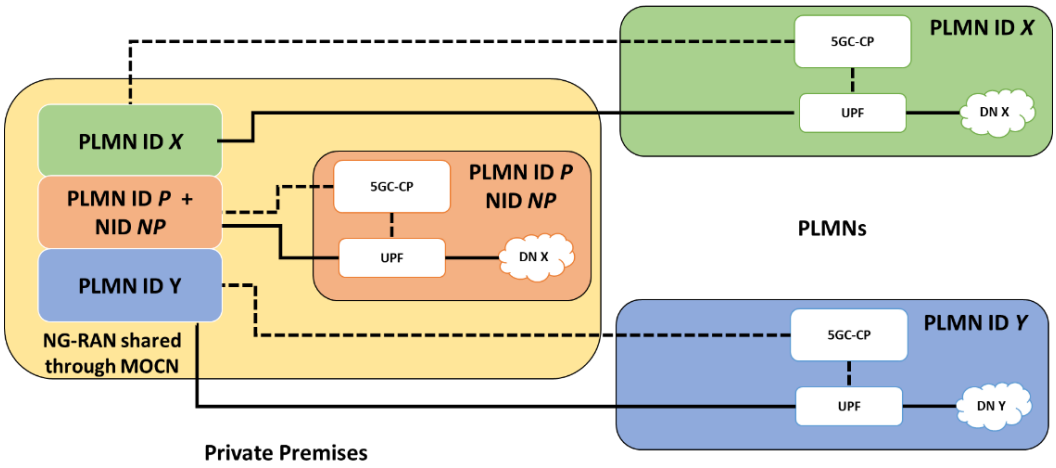

Figure 9.1: Example scenario, in which two PLMNs and a single SNPN are sharing the on-premise RAN using MOCN.

9.2.2 PNI-NPN as a slice of a PLMN

The NPN is deployed as a slice within a PLMN and provided by an MNO. The MNO can give the NPN service provider the possibility to manage the slice. CAGs might be used to restrict the access of some UEs to certain cells within the slice. MNO might deploy additional slices to offer public services for its subscribers within the private boundaries.

Figure 9.2 shows a possible realization of the scenario described above. Specifically, the example includes two slices, one (S-NSSAI Y) for providing access to the public services within the private venue and the other (S-NSSAI X) for realizing the PNI-NPN. All the network functions are shared between the two slices. The whole 5GC is deployed out-of-premises at any point of presence of the PLMN. For example, as considered in the 5G-CLARITY UC1 pilot, a museum could leverage this deployment option. The museum tourism company might offer AR applications to the visitors for an enhanced museum tour. At the same time, there is a coexisting slice that provides visitors with access to public services.

Figure 9.3 depicts an alternative deployment option for the PNI-NPN. For this case, the MNO has deployed a UPF instance within the private premises. In this way, the latency to access private services is reduced. For the same purpose, the MNO also deploys public services on the private DN accessible through the in-house UPF. The out-of-premises UPF is only destined for public services. Then, the private data traffic is kept within the private premises. The in-house UPF might act as both uplink classifier and PDU Session Anchor to properly divert the public traffic to its destination (e.g., the private DN or the UPF out-of-premises). Please note that some 5GC-CP network functions (e.g., AMF and SMF) could also be instantiated in-house to offload signalling traffic from the 5GC-CP out-of-premises, for instance.

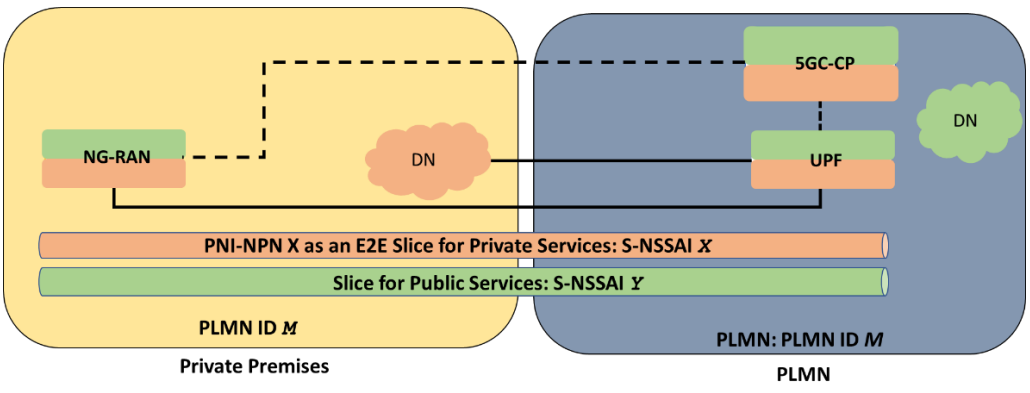

Figure 9.2: PNI-NPN provided as a dedicated end-to-end slice within a PLMN.

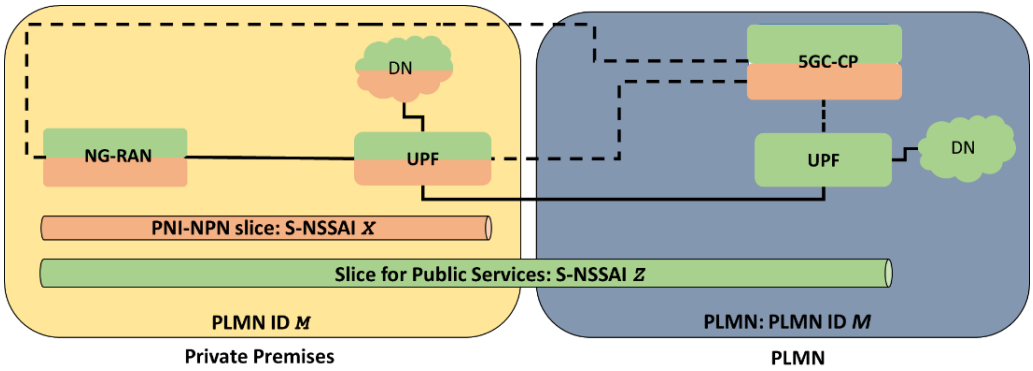

Figure 9.3: PNI-NPN deployed as a slice within a PLMN. A UPF deployed on the private premises is used in order to reduce the latency.

9.2.3 Mobility between SNPN and PLMN

SNPNs supporting private services that involve UEs moving from inside the private premises to the outside and vice versa. The SNPN RAN coverage area does not extend beyond private premises. Then, a PLMN has to provide 5G connectivity to the UE as soon as it exits the private venue. Examples of these services are a set of UAVs that need to exit the private premises to collect some data or the tracking of products equipped with RFIDs since its production until their delivery (see Figure 9.4).

These services require the connectivity is kept all the time, even when the UE moves from the private premises to the outside. Both Session and Service Continuity (SSC) mode 1 and SSC mode 3 meet that requirement. SSC mode 1 also allows to maintain the PDU session anchor and, thus, the IP address. In 3GPP standards, the handover of PDU session procedure from untrusted non-3GPP to 3GPP access may be considered to achieve seamless service continuity when the UE is leaving the private premises.

A UE registered with an SNPN might carry out another registration with a PLMN. To that end, first, a secured tunnel is established between the UE and the PLMN's N3IWF through which the UE and PLMN will exchange the signalling and data traffic. This tunnel is labelled as "NWu-PLMN" in Figure 9.5. From the PLMN point of view, the UE is connecting to the network through a non-3GPP access technology. After the establishment of the NWu-PLMN tunnel, the UE initiates the registration procedure with the PLMN's AMF to have access to the public services.

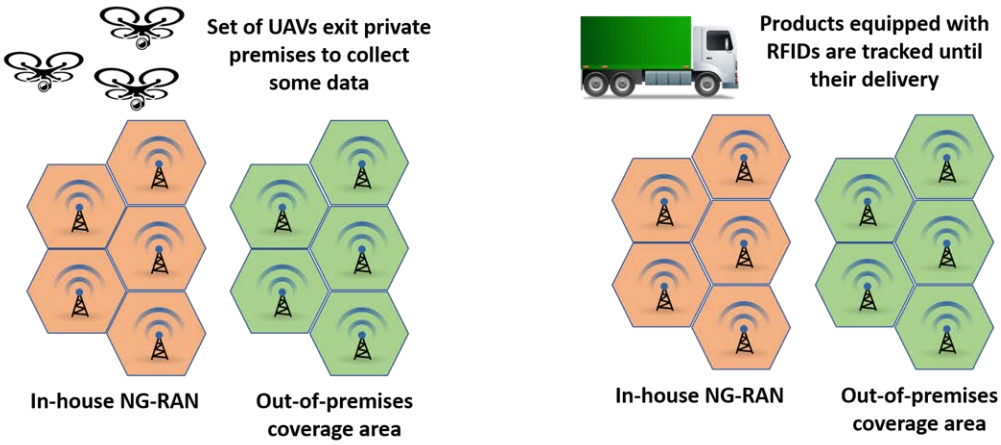

Figure 9.4: Example use cases for mobility scenarios in NPNs

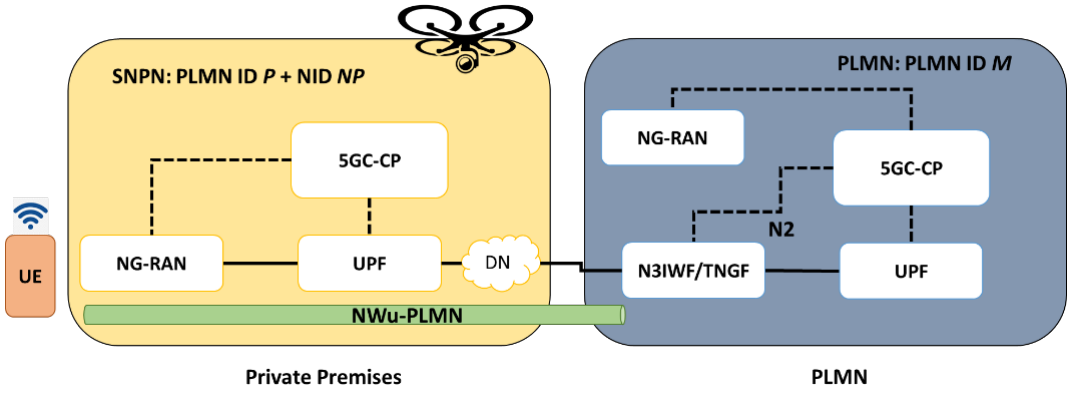

Figure 9.5: Access to PLMN services via SNPN.

When the UE is under indoor coverage, it is needed to move the PDU session established with the PLMN from non-3GPP access to 3GPP access. First, the UE needs to register in the public NG-RAN if it did not do it previously. Then, the UE issues a PDU session establishment request explicitly indicating the AMF the request refers to an ongoing PDU session. This procedure enables to preserve the PDU session anchor UPF and the PLMN IP address. To ensure a seamless communication during the procedure, the UE needs to keep connected both the SNPN and the PLMN RANs until the User Plane resources and PDU session context in N3IWF are released.

In the same way, SNPN services might be accessed from a PLMN using similar principles as described above, but now the PLMN playing the role of untrusted non-3GPP access from the SNPN perspective (see Figure 9.6). Also, a similar procedure as described above is valid in this scenario to assure a seamless service continuity when the UE moves from a PLMN to an SNPN.

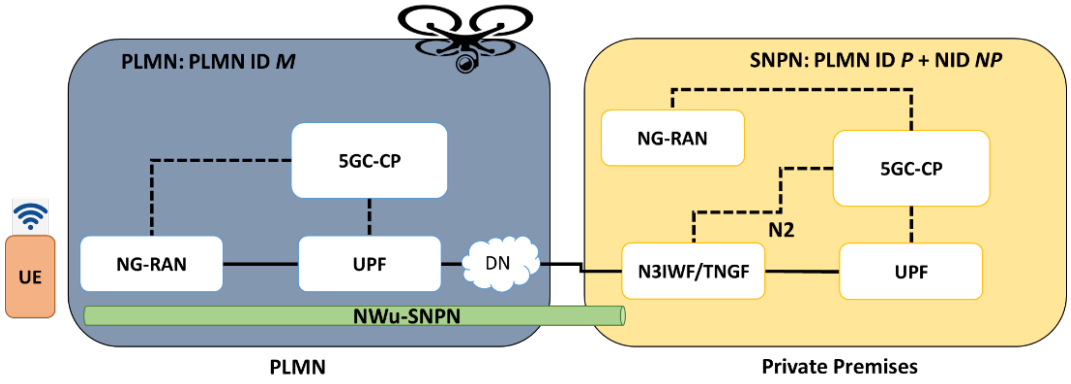

Figure 9.6: Access to SNPN services via PLMN.

9.3 NPN-PLMN interaction: management plane

As presented in Sections 9.1 and 9.2, a PNI-NPN is an E2E network composed of two sub-networks: one *private*, consisting of VxFs that are provided by the private NOP; and *one public*, consisting of public VxFs that are provided by the public NOP. While private VxFs are typically executed on 5G-CLARITY infrastructure[7], public VxFs can be deployed using either 5G-CLARITY resources (on-premise resources) or PLMN resources (off-premise resources).

Based on this rationale, it is clear that both the private NOP - responsible for providing private VxFs and

[7] The possibility of executing private VxFs on the MNO's telco cloud is also considered in 5G-CLARITY. For more details, see Annex B.

managing 5G-CLARITY resources - and the public NOP - responsible for providing public VxFs and managing PLMN resources - play a key role for the provisioning of a PNI-NPN. To ensure a unified operation of this E2E network, it is thus required that the management systems of both NOPs interact with each other, exchanging trusted and verifiable messages between them. To facilitate the interaction between these systems, namely the 5G-CLARITY management and orchestration stratum (managed by the private NOP) and the 3GPP management system (managed by the public NOP), the network service aggregator role provides mechanisms leveraging on two key functionalities: capability exposure (Section 9.3.1) and auditability (Section 9.3.2).

9.3.1 Capability exposure

Capability exposure can be defined as the ability of a NOP to securely expose capabilities from their managed functions towards one or more authorized tenants. These functions include:

- **VxFs**. The ability of exposing the capabilities from one or more VxFs (including the NFV network services resulting from their composition), making them available for external consumption, is referred to as *network capability exposure*. First solutions on this exposure were specified for 3GPP core networks, with the definition of the Service Capability Exposure Function (SCEF) in the EPC+ and Network Exposure Function (NEF) in the 5GC. In the RAN segment, the O-RAN fostered principles of RAN disaggregation and openness facilitate the exposure of gNB-DU and gNB-CU offered capabilities through the E2 interface, so that RIC and associated xApps (including RIC built-in xApps as well as over-the-top xApps) can consume and manipulate them.
- **MFs**. The ability of exposing the services from one or more MFs, making these services available for external consumption, is referred to as *management capability exposure*. This category, unprecedented so far and much less mature than network capability exposure, has raised attention in industry community with the use of novel XaaS model, specially NaaS. In this regard, different surveys that can be found in acclaimed technology analyst reports (e.g. Gartner, Heavy Reading, Analysis Manson, etc.) state that NaaS, and in particular NSaaS, open up opportunities for service innovation, with a win-win solution for both slice providers and slice tenants. NSaaS allows the provider to make slice instances available to different tenants, allowing the latter to consume them at their own. Every tenant can use the provided slice instance as an isolated service platform, deploying their digital/communication services atop for their own customers. To allow for these B2B2X models, the different analyst reports recognize the need for the slice provider to open its management systems to individual tenants, allowing them to get involved in the operation of received slice instance beyond passive monitoring (e.g. performance assurance, fault supervision) activities. By regulating this openness and exposure, the slice provider can define the degree of control the tenant can take over the slice.

Management capability exposure in multi-tenant environments like 5G-CLARITY brings some implications, particularly considering that different tenants may want to have different levels of management over their serving 5G-CLARITY slices. This makes it necessary to define different levels of exposure. Some proposals have been suggested so far to address this problem. For example, in [98] Telefónica proposes a exposure model with two types of slices (see Figure 9.7): *i) provider-managed slices*, meaning that the provider keeps the full control and management of the slice, while the customer can merely use the network resources of the provided slice, without any further capability of managing or controlling them; and *ii) tenant-managed slices*, implying that the tenant has access to a (limited) set of operations and/or configuration actions, and the provider just segregates the infrastructure as required for that purpose.

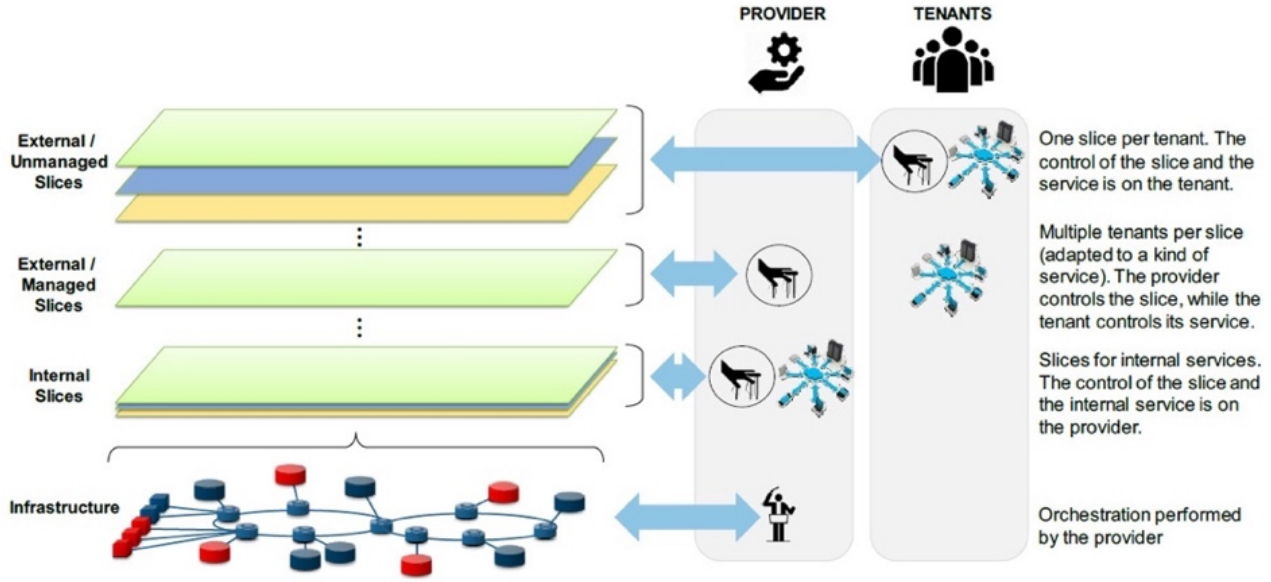

Figure 9.7: Different capability exposure when slicing [98].

The 5G-CLARITY system leverages on these state-of-the-art solutions to design a future-proof capability exposure mechanism, which is to be detailed in deliverables D4.1 and D4.2. Enforced by the network service aggregator role, this mechanism shall provide means to establish a clear demarcation point between a public NOP and a private NOP, defining their individual management scopes in the operation of a PNI-NPN. Specifically, this mechanism needs to:

- Ensure the public NOP is able to retain control over the public VxFs, when these are deployed on the 5G-CLARITY infrastructure. This means that, apart from leading VxF configuration management, the public NOP should be able to trigger lifecycle management operations (e.g. scaling) over those VxFs, when required. To that end, the private NOP (acting as provider) needs to expose necessary capabilities to the public NOP (acting as tenant). For example, the private NOP can authorize the public NOP to access the 5G-CLARITY management and orchestration stratum, so that the 3GPP management system can consume 5G-CLARITY NFVO services.
- Ensure the public NOP is able to configure and manage connectivity between the private site and the PLMN, in case public VxFs are deployed on the PLMN. This requires the private NOP (acting as provider) to communicate external-facing connectivity information towards the public NOP (acting as tenant), so that the latter can setup VPN services across the private and public sites. Examples of this information include IP addresses and ports of the on-premise gateway.

For a fine-grained control of this capability exposure across public and private NOPs, 5G-CLARITY system may make leverage on the use of token-based authentication. When the public NOP register into the private NOP admin domain for the first time, the first one is granted with a unique access token that specifies the set of management services it can consume at operation time. The logic behind this token-based authentication is as follows: every time a public NOP invokes a management service from a MF, the MF checks the permissions imbued in the token assigned to the public NOP. If these permissions include the requested management service, then the MF authorizes the corresponding API invocation. The process for the token verification by the MF (API producer) depends on the token format and the associated metadata. This definition, as well as the token request and generation profile, is up to the 5G-CLARITY mediator function (see Section 7.4.1)

9.3.2 Auditability

Auditability aims at making the 5G-CLARITY system a verifiable and trustable stack in multi-tenant environments, where multiple actors from different administrative domains, i.e. private NOP and individual public NOPs, interact with each other. To that end, non-repudiation mechanisms complemented with advanced security methodologies need to be defined and integrated into the platform.

The non-repudiation principle means that each pair of actors taking part in any interaction (typically, a requestor and a responder) can demonstrate that a certain request or response message has been effectively generated by the other one, relating them to previous relevant messages, and associating them with a consistent temporal line. To achieve this, both actors shall save their (equivalent) evidence of the exchanges on a private trusted store. These evidences will provide means for external verifiability of all messages exchanges, allowing system auditability, and further applications enabled by it. The integration of non-repudiation mechanisms will allow the 5G-CLARITY platform to keep audit trails for traceability purposes, by storing a trusted record of all the request-response interactions exchanged on the platform. These mechanisms shall be compatible with the protocol(s) selected for these exchanges.

According to the above reasoning, 5G-CLARITY shall allow the generation of available, clear, original (unchanged), verifiable and accurate audit trails for the interplay of private venue with the PLMN, ensuring traceable and secure interactions across them, at both control and orchestration levels. The messages exchanged between these two administrative domains, including request-response and subscribe-notify messages, need to support non-repudiation.

9.4 NPN-PLMN: security considerations

The capability exposure and the exchange of information with public networks will increase the openness of 5G-CLARITY system and hence increase the attack surface, making the system more vulnerable to security and privacy threats. To avoid this, it would be recommendable to include mechanisms for device authentication and privacy, remote attestation and Proof-of-Transit. It is worth noting that these mechanisms will not be developed or integrated in 5G-CLARITY system pilots, since security is formally out of the project' scope at this first stage. The reason for describing these mechanisms is to provide guidelines on what security solutions should be provisioned to 5G-CLARITY system beyond project's lifetime, to make the system a reliable and security framework in production-ready environments.

Device authentication

Authentication is fundamental to 5G in order to build trust between the user equipment (UE) and the network. The authentication fundamentals are described below:

- Access security is managed in a unified manner, for which the network function Authentication Server Function (AUSF) has been defined.
- Authentication methods mandatory to support by a PLMN are 5G-AKA and EAP-AKA'. EAP-TLS is optionally defined. Any method can be used to authenticate the UE over both access types.
- The home network gets confirmation if the UE was successfully authenticated in the serving network. Upon confirmation receipt, the home network provides to the serving network the security anchor key for generating further key material between the UE and the serving network. During the authentication process, the serving network sends a challenge to the UE, which the UE needs to respond to. If the response by the UE is equal to an expected response provided earlier by the home network to the serving network, then the serving network will allow the UE to access its network.
- There is binding of the serving network ID into the authentication request in order to prevent fraud, e.g. a serving network attempting to register a UE that is not presented in the visited network

With the introduction of NPNs from 3GPP Release 16 onwards, new device authentication related issues have been emerged, especially for NPNs [99]. For PNI-NPNs, since they will be connecting to PLMNs, the use of 5G-AKA and EAP-AKA' authentication methods are mandatory, together with the presence of USIMs in the UE for PLMN authentication. Depending on deployment scenarios, additional authentication needs to be made.

On the one hand, when the PNI-NPN is deployed as a slice, slice-specific authentication using EAP-based authentication [100] can be optionally performed after primary authentication between the UE and the network. As shown in Figure 9.8, when the UE is authenticated for network access, the serving network and the UE receives a list of permitted network slices, indicated by their NSSAIs. The permitted network slices may further require slice-specific authentication by the Network Slice Specific Authentication Authorization Function (NSAAF). This additional slice-specific authentication is indicated by the UE subscription information. If it is the case, the AMF in the serving network triggers the EAP authentication procedure for slice-specific access. This additional slice-specific authentication gives much more control to the NPN slice tenant in managing access to the slice instead of solely relying on the PLMN operator for access control.

The combination of primary authentication and slice-specific authentication shall be considered for future implementation in 5G-CLARITY system, beyond project's lifetime, as long as Rel16 components are available for use.

On the other hand, when the device has credentials and subscription for both PLMN and SNPN independently, there may situations such as the one described in Section 9.2.3 where the device needs to access PLMN services while camping in SNPN RAN (and viceversa). In such a case, the UE can leverage the IP connectivity offered by the SNPN (PLMN) to establish an IPSec tunnel to the SNPN (PLMN). Then, the UE registers with the SNPN (PLMN) using the credentials to access the SNPN (PLMN) and obtain access to corresponding PLMN (SNPN) services.

The integration of the feature above into the 5G-CLARITY system is reasonably straightforward with today's 3GPP 5GC and device capabilities.

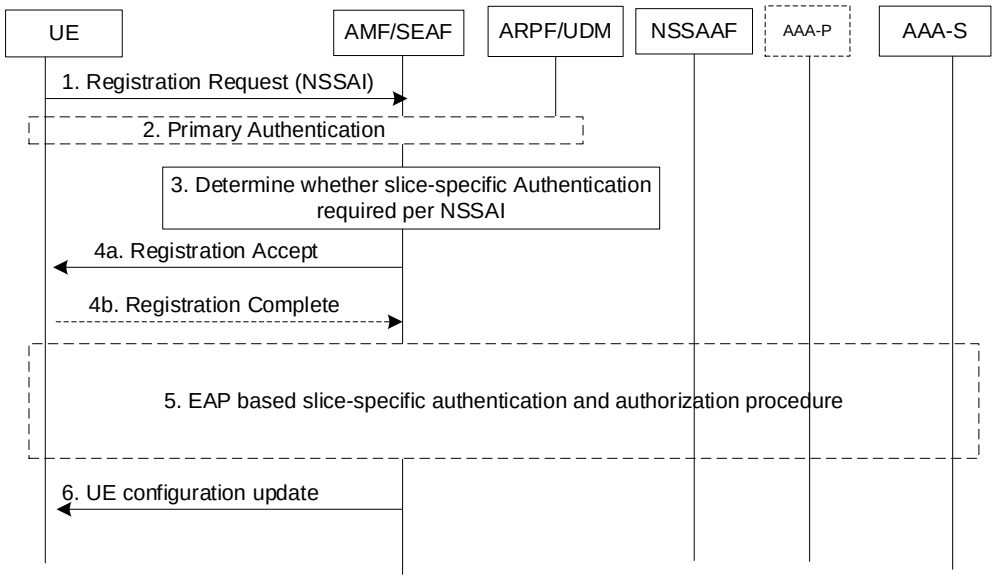

Figure 9.8: Relationship between primary authentication and slice-specific authentication [101]

Remote attestation for NFVI and VxFs.

Remote attestation is a technique based on the principle of Chain of Trust (CoT), a transitive mechanism that ensures the trustworthiness of an execution platform via a step-by-step extension process. Each element of the CoT is responsible for measuring and storing the integrity of the next element, so that the whole chain can be verified by a third party. This means that remote attestation extends the CoT outside of the execution platform to involve a trusted third party, who verifies that the conditions are still valid.

This technique has gained momentum NFV environments because it generates trust and liability for the NFVI and VxFs. Indeed, this technology has been standardized by ETSI NFV-SEC group as a clear statement of intentions to be adopted. Figure 9.9 shows the general concept, where the *trust assessor* is in possession of a set of good known values or "golden values", which are Processor Capacity Reservation (PCR) registers stored in a database of the *target platform*. *Remote verifier*, i.e. trusted third party, triggers the remote attestation to check the integrity and trust of the platform and upper layers (hypervisor and VNFs). This is as simple as requesting an integrity measurement report to the *target platform* and comparing the values obtained with the golden values. If there is no match, the *remote verifier* will lose the trust in the platform and software. The role of *remote verifier* can be delegated or taken by several actors, from the NFVI provider to the tenant of the VxFs.

One of the most attractive aspects for remote attestation technology is that it is based on the use of Trusted Platform Module (TPM) standard. The TPM is a device acting as a secure cryptoprocessor capable of storing keys, secrets, identities and measurements of the platform integrity. Integrity measurements are protected by the TPM's PCRs. PCRs can only be updated by the TPM itself, using an internal secure hash function, via the "extend" operation: at each step, the current value of the PCR is concatenated with the new measurement and the digest of the resulting messages is stored in the PCR[8].

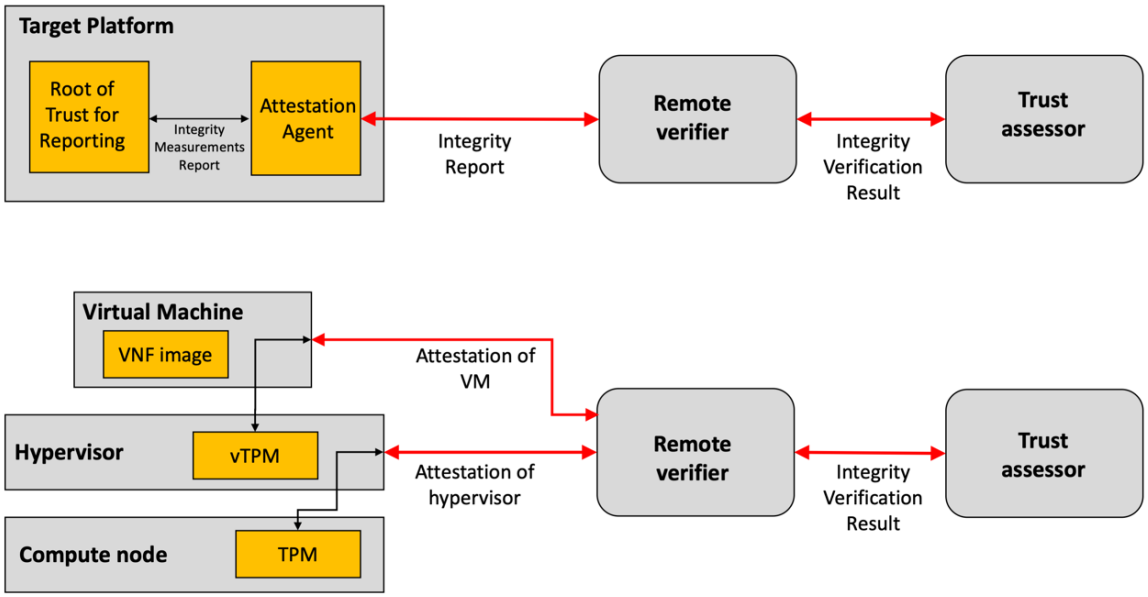

Figure 9.9: Remote attestation over NFVI and VNFs

[8] This mechanism ensures that unless the platform is rebooted, no PCR-stored measurement can be erased, thus software-based attacks cannot hide execution of untrusted binaries.

The PCRs' value can be accessed by a remote verifier by challenging the TPM with a nonce; using a hardware-protected key (i.e. only the TPM can use the private key for signing), the TPM protects the integrity of the PCRs' with a signature which includes the challenge nonce for freshness. Using the prior knowledge of all the platform's TPM public key used for attestation, the remote verifier can verify the genuineness of the signature – which also validates the hardware entity, as well as the content of the logged software events.

In 5G-CLARITY, the execution of remote attestation for NFVI and VxFs is important for both private and public NOPs. For example, when the private NOP offers the on-premise edge NFVI to the public NOP for VxF hosting (i.e. public NOP operated VxFs are executed on this NFVI), it is important for the private NOP to ensure the trust and liability of this NFVI. Otherwise, public NOPs will be reluctant to deploy VxFs there, considering that this NFVI is beyond their control, thereby preventing 5G-CLARITY system to be used for multi-tenancy purposes.

Proof-of-Transit (PoT)

Another important security aspect for 5G-CLARITY is proof of traffic isolation or processing points. Furthermore, regulatory obligations or a compliance policy require NOPs to prove that all packets that are supposed to follow a specific path are indeed being forwarded across and exact of pre-determined nodes. Solutions that provide PoT for packets traversing a specific path are being investigated for that purpose. The method relies on adding PoT data to all packets that traverse a path. The added PoT data allows a verifier node (potentially, an egress node) to check whether a packet traversed the identified set of nodes on a path as expected or not. In the proposed scheme, security mechanisms are natively built into the generation of the PoT data to protect against misuse and compromises (e.g. configuration, mistakes).

In 5G-CLARITY, PoT allows two NOPs, one private and other public, to verify that data-plane packets exchanged between them follow the agreed forwarding path. This is especially important when packets stepped out their administrative domains, as it happens in the transport network connecting the private site with the PLMN.

10 Conclusions

This deliverable provided an overview of 5G-CLARITY ecosystem, introducing the stakeholders and associated service delivery models, as well as a series of requirements that have ultimately driven the architectural design of 5G-CLARITY system. The design of the 5G-CLARITY overall functional architecture followed the baseline requirements and related KPIs of the two targeted use cases detailed in deliverable D2.1.

The proposed 5G-CLARITY architecture is structured into four separate strata, namely infrastructure stratum, network and application function stratum, management and orchestration stratum, and intelligence stratum. Based on the "separation of concerns" principle, this architecture design allows independent evolution of individual strata, with different technology pace each.

Chapter 3 captured the functional and non-functional requirements to be satisfied by the 5G-CLARITY system and specified the principles that will guide the system architecture design. A complete characterization of individual strata, including details on their functional components and offered capabilities, are provided in Chapters 4, 5, 6 and 7, respectively. Finally, Chapter 9 provided insights on the integration of public and private networks in 5G-CLARITY system, through different deployment scenarios and an analysis on their interoperation at the management and orchestration plane with security considerations.

This deliverable serves as a baseline for upcoming implementation activities within the project, particularly in WP3 and WP4. We note, however, that the network capabilities made available by the different functional components and their design presented here might be subject to future extensions and modifications, depending on feedback obtained during the development phase.

11 Bibliography

[1] 5G-CLARITY Deliverable D2.1, "Use Cases and Requirements", March 2020.

[2] 5G-CLARITY Deliverable D3.1, "State-of-the-art review and initial design of the integrated 5GNR/Wi-Fi/LiFi network frameworks on coexistence, multi-connectivity, resource management and positioning", September 2020.

[3] PureLiFi, ''Ecosystem of LiFi'' [Online]. Available: https://purelifi.com/lifi-technology/ecosystem-of-lifi-pp/ [Accessed June 2020]

[4] H. Haas, J. Elmirghani, I. White, ''Optical wireless communication''. Phil. Trans. R. Soc. A 378: 20200051, 2020.

[5] O'Brien DC, Zeng L, Le-Minh H, Faulkner G, Walewski JW, Randel S. ''Visible light communications: challenges and possibilities'', *IEEE 19th Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Cannes, France, 15–18 September 2008, pp. 1–5. IEEE.

[6] R. Bian, I. Tavakkolnia, I. and H. Haas, ''15.73 Gb/s visible light communication with off-the-shelf LEDs'', *Journal of Lightwave Technology*, 2019, 37(10), pp.2418-2424.

[7] British Standards Institution (BSI), ''Safety of laser products. Equipment classification and requirements'', Std. 2014, BS EN 60 825-1.

[8] Haas, H., Yin, L., Chen, C., Videv, S., Parol, D., Poves, E., Alshaer, H. and Islim, M.S., ''Introduction to indoor networking concepts and challenges", *LiFi Journal of Optical Communications and Networking, 2020*, 12(2), pp. A190-A203.

[9] T. D. C. Little and M. Rahaim, "Network topologies for mixed RFVLC HetNets," *IEEE Summer Topicals Meeting Series (SUM*), 2015, pp. 163–164.

[10] W. Ma and L. Zhang, "QoE-driven optimized load balancing design for hybrid LiFi and Wi-Fi networks," *IEEE Communication Letters*, vol. 22, no. 11, pp. 2354-2357, Nov. 2018

[11] Y. Wang, D. A. Basnayaka, X. Wu, and H. Haas, "Optimization of load balancing in hybrid LiFi/RF networks," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1708–1720, April 2017.

[12] Y. Wang and H. Haas, "Dynamic load balancing with handover in hybrid LiFi and Wi-Fi networks," *Journal of Lightwave Technology,* vol.33, no. 22, pp. 4671–4682, Nov. 2015.

[13] IEEE, "IEEE P802.11 - LIGHT COMMUNICATION (LC) TASK GROUP (TG) - MEETING UPDATE''. [Online] Available: http://www.ieee802.org/11/Reports/tgbb_update.htm [Accessed July 2020]

[14] 3GPP TS 38.300, "Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2"

[15] 3GPP TS 37.340, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage 2"

[16] Dynamic Spectrum Alliance (DSA), "Automated Frequency Coordination - An established tool for modern spectrum management", March 2019.

[17] European Conference of Postal and Telecommunications Administrations (CEPT), "ECC Report 205 - Licensed Shared Access (LSA)", Feb. 2014

[18] Federal Communications Commission (FCC), "Report and Order and Second Further Notice of Proposed Rulemaking", April 2015

[19] Federal Communications Commission (FCC), "Order on Reconsideration and Second Report and Order", May 2016

[20] Federal Communications Commission (FCC), "Report & Order & Further Notice of Proposed Rulemaking: Unlicensed Use of the 6 GHz Band - Expanding Flexible Use in Mid-Band Spectrum Between 3.7 and 24 GHz", GN Docket No. 17-183, 2020

[21] Ofcom, "Enabling wireless innovation through local licensing: Shared access to spectrum supporting mobile technology", July 2019

[22] 5G-ACIA White Paper "Exposure of 5G Capabilities for Connected Industries and Automation Applications", May 2020 [Online]. Available: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_Exposure_of_5G_Capabilities_for_Connected_Industries_and_Automation_Applications/5G-ACIA_Exposure_of_5G_Capabilities_Download.pdf [Accessed Sept 2020]

[23] 3GPP TR 22.872, "Technical Specification Group Services and System Aspects; Study on positioning use cases; Stage 1"

[24] A. Yassin et al., "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1327-1346, Q2 2017.

[25] 3GPP TR 22.862, "Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Market Technology Enablers for Critical Communications; Stage 1"

[26] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo and G. Seco-Granados, "Survey of Cellular Mobile Radio Localization Methods: From 1G to 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124-1148, Q2 2018

[27] C. Fiandrino, H. Assasa, P. Casari and J. Widmer, "Scaling Millimeter-Wave Networks to Dense Deployments and Dynamic Environments," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 732-745, April 2019

[28] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436-453, April 2016

[29] F. Lemic *et al.*, "Localization as a feature of mmWave communication," *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Paphos, 2016, pp. 1033-1038

[30] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson and H. Wymeersch, "Location-Aware Communications for 5G Networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102-112, Nov. 2014

[31] Y. Wang, Z. Shi, Y. Yu, S. Huang and L. Chen, "Enabling Angle-based Positioning to 3GPP NR Systems," *2019 16th Workshop on Positioning, Navigation and Communications (WPNC)*, Bremen, Germany, 2019, pp. 1-7.

[32] C. Danakis, M. Afgani, G. Povey, I. Underwood and H. Haas, "Using a CMOS camera sensor for visible light communication," *2012 IEEE Globecom Workshops*, Anaheim, CA, 2012, pp. 1244-1248.

[33] J. Lin, "Synchronization Requirements for 5G: An Overview of Standards and Specifications for Cellular Networks," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 91-99, Sept. 2018

[34] Z. Sahinoglu, S. Gezici, and I. Guvenc, "Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols", Cambridge University Press, 2008

[35] V. Sark, E. Grass and J. Gutiérrez, "Multi-Way ranging with clock offset compensation," *2015 Advances in Wireless and Optical Communications (RTUWO)*, Riga, 2015, pp. 68-71

[36] K. Boyle, "5G is all in the timing" [Online]. Available: https://www.ericsson.com/en/blog/2019/8/what-you-need-to-know-about-timing-and-sync-in-5G-transport-networks [Accessed June 2020]

[37] H. Li, L. Han, R. Duan and G. M. Garner, "Analysis of the Synchronization Requirements of 5g and Corresponding Solutions," *IEEE Communications Standards Magazine*, vol. 1, no. 1, pp. 52-58, March 2017.

[38] C. Catal and B. Diri, "A systematic review of software fault prediction studies", *Expert Systems with Applications*, vol. 36, Issue 4, pp. 7346-7354, May 2009.

[39] A.A. Covo, T.M. Moruzzi, E.D. Peterson. "AI-assisted telecommunications network management", *1989 IEEE Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond'*, Dallas, TX, USA, 1989, pp. 487-491, vol. 1

[40] Ericsson report, "AI techniques to enhance returns on 5G network investments" [Online]. Available: https://www.ericsson.com/49b63f/assets/local/networks/offerings/machine-learning-and-ai-aw-screen.pdf [Accessed June 2020]

[41] C. Zhang, Y. Ueng, C. Studer and A. Burg, "Artificial Intelligence for 5G and Beyond 5G: Implementations, Algorithms and Optimizations", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no.2, pp. 149-163, June 2020

[42] F. Hussain, S.A. Hassan, R. Hussain and E. Hossain, "Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges", *IEEE Communication Surveys & Tutorials*, vol. 22, no. 2, pp. 1251-1275, Q2 2020.

[43] X. Wang, Y. Han, V.C.M. Leung, D. Niyato, X. Yan and X. Chen, "Convergence of Edge Computing and Deep Learning: A Comprehensive Survey", *IEEE Communication Surveys & Tutorials*, vol. 22, no.2, pp.869-904, Q2 2020.

[44] N. C. Luong et al. "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey", IEEE Communication Surveys & Tutorials, vol. 21, no. 4, pp. 3133-3174, Q4 2019.

[45] International Telecommunications Union (ITU), "Y.3172 : Architectural framework for machine learning in future networks including IMT-2020", 2019 [Online]. Available: https://www.itu.int/rec/T-REC-Y.3172-201906-I/en [Accessed Sept 2020]

[46] O-RAN Alliance, "AI/ML workflow description and requirements", 2019 [Online]. Available: https://static1.squarespace.com/static/5ad774cce74940d7115044b0/t/5de7b47d1c05887efd8e62cb/1575466123773/ORAN-WG2.AIML.v01.00.pdf

[47] ETSI GS ENI 005 "Experiential Networked Intelligence (ENI); System Architecture", 2019 [Online]. Available: https://www.etsi.org/deliver/etsi_gs/ENI/001_099/005/01.01.01_60/gs_ENI005v010101p.pdf

[48] L. Fallon, J. Keeney, M. McFadden, J. Quilty and S. van der Meer, "Using the COMPA autonomous architecture for mobile network security", *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, Lisbon, 2017, pp. 747-753.

[49] 5G-CLARITY D4.1, "Initial design of the SDN/NFV platform and identification of target 5G-CLARITY ML algorithms", October 2020.

[50] 3GPP TS 23.501, "System Architecture for the 5G System (5GS); Stage 2

[51] 3GPP TS 22.261, "Service requirements for the 5G system"

[52] 5G-ACIA, "5G Non-Public Networks for Industrial Scenarios," July 2019. [Online]. Available: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_5G_for_Non-Public_Networks_for_Industrial_Scenarios/WP_5G_NPN_2019_01.pdf [Accessed June 2020]

[53] Harrison J. Son, "7 Deployment Scenarios of 5G Private Networks," NETMANIAS Tech-Blog Post. October 2019. [Online]. Available: https://www.netmanias.com/en/post/blog/14500/5g-edge-kt-sk-telecom/7-deployment-scenarios-of-private-5g-networks [Accessed June 2020]

[54] J. Ordonez-Lucena, J. F. Chavarria, L. M. Contreras and A. Pastor, "The use of 5G Non-Public Networks to support Industry 4.0 scenarios," *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*, Granada, Spain, 2019, pp. 1-7.

[55] A. Rostami, "Private 5G Networks for Vertical Industries: Deployment and Operation Models," *2019 IEEE 2nd 5G World Forum (5GWF)*, Dresden, Germany, 2019, pp. 433-439.

[56] 3GPP TR 28.801, "Telecommunication management; Study on management and orchestration of network slicing for next generation network "

[57] NGMN Alliance, "NGMN 5G White Paper", February 2015. [Online]. Available: https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf

[58] 5G-PPP White Paper, "View on 5G Architecture", Architecture Working Group, Version 3.0, June 2019 [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2019/07/5G-PPP-5G-Architecture-White-Paper_v3.0_PublicConsultation.pdf [Accessed June 2020]

[59] 3GPP TS 28.530, "5G; Management and Orchestration; Concepts, use cases and requirements"

[60] 5G-EXchange (5GEx), "H2020 5GEx project"

[61] SLICENET, "H2020 SLICENET project" [Online]. Available: https://slicenet.eu [Accessed September 2020]

[62] 5G-PPP Phase 3 projects [Online]. Available: https://5g-ppp.eu/5g-ppp-phase-3-projects [Accessed September 2020]

[63] ETSI GS NFV-003, "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV"

[64] Global mobile Supplier Alliance, "5G Device Ecosystem Report Executive Summary", April 2020.

[65] Ericsson blog, "Why you should be taking note of cloud native NFVI". [Online]. Available: https://www.ericsson.com/en/blog/2019/8/why-you-should-be-taking-note-of-cloud-native-nfvi [Accessed June 2020]

[66] "Hype Cycle for Digital Government Technology", Gartner, 2019.

[67] Analysis Manson, "Acceleration technologies: realizing the potential of network virtualization", White Paper, June 2019.

[68] FG-NET-2030, "Network 2030: A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond". [Online]. Available: https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf . [Accessed June 2020].

[69] 5G Americas, "5G and the Cloud", White Paper, December 2019. [Online]. Available: https://www.5gamericas.org/wp-content/uploads/2019/12/5G-Americas_5G-and-the-Cloud.pdf [Accessed June 2020]

[70] 5G PPP, "From Webscale to Telco, the Cloud Native Journey", White Paper, July 2018. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2018/07/5GPPP-Software-Network-WG-White-Paper-23052018-V5.pdf [Accessed June 2020]

[71] Eric Evans, 'Domain-Driven Design: Tackling Complexity in the Heart of Software', Book, 20 August 2003.

[72] O-RAN Alliance [Online]. Available: https://www.o-ran.org [Accessed June 2020]

[73] NGMN Alliance and WBA, "RAN Converge Paper", White Paper, August 2019. [Online]. Available: https://www.ngmn.org/wp-content/uploads/Publications/2019/190903-RAN-Convergence-Paper.pdf. [Accessed June 2020].

[74] 3GPP TS 28.533, "5G; Management and Orchestration; Architecture framework"

[75] ETSI GS ZSM 002, "Zero-touch network and Service Management (ZSM); Reference Architecture"

[76] ETSI NFV web site, "Standards for NFV". [Online]. Available: https://www.etsi.org/technologies/nfv

[77] GSM Alliance (GSMA); "An Introduction to Network Slicing", White Paper, 2017. [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf [Accessed June 2020]

[78] NGMN Alliance, "Security Aspects of Network Capabilities Exposure in 5G", v1.0, September 2018. [Online]. Available: https://www.ngmn.org/wp-content/uploads/Publications/2018/180921_NGMN-NCEsec_white_paper_v1.0.pdf [Accessed June 2020]

[79] GSM Alliance (GSMA); "The 5G Guide: A reference for operators", White Paper, 2019. [Online]. Available: https://www.gsma.com/wp-content/uploads/2019/04/The-5G-Guide_GSMA_2019_04_29_compressed.pdf [Accessed July 2020].

[80] Ericsson Technology Review, "Leveraging LTE and 5GNR networks for fixed wireless access", Aug. 2018. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/leveraging-lte-and-5GNR-networks-for-fixed-wireless-access [Accessed July 2020]

[81] IEEE Std 802.1Q-2018 *(Revision of IEEE Std 802.1Q-2014),* "IEEE Standard for Local and Metropolitan Area Network--Bridges and Bridged Networks," pp.1-1993, July 2018.

[82] A. Nasrallah *et al*., "Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 88-145, Q1 2019

[83] 3GPP TR 23.793, "Study on access traffic steering, switch and splitting support in the 5G system architecture"

[84] 3GPP TS 24.301: "Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3"

[85] ETSI GS NFV-MAN 001, "Network Functions Virtualisation (NFV); Management and Orchestration"

[86] ETSI GS NFV-IFA 013, "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Os-Ma-nfvo reference point – Interface and Information Model Specification"

[87] ETSI GS NFV-SOL 005, "Network Functions Virtualisation (NFV) Release 2; Protocols and Data Models; RESTful protocols specifications for the Os-Ma-nfvo Reference Point"

[88] M. Paule Odini, "Telco Cloud NFV Metrics and Performance Management", *IEEE Softwarization*, May 2017 [Online]. Available: https://sdn.ieee.org/newsletter/may-2017/telco-cloud-nfv-metrics-and-performance-management [Accessed Sept 2020]

[89] ETSI GS NFV-IFA 005, "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Or-Vi reference point – Interface and Information Model Specification"

[90] Apache Hadoop: "An open-source software for reliable, scalable, distributed computing" [Online] https://hadoop.apache.org/docs/current/ [Accessed July 2020]

[91] Presto: "A high performance, distributed SQL query engine for big data" [Online]. Available: https://prestosql.io/docs/current/ [Accessed July 2020]

[92] Apache Spark: "An unified analytics engine for large-scale data processing" [Online]. Available: https://spark.apache.org/docs/latest/ [Accessed July 2020]

[93] D. Hardt, "The Auth 2.0 Authorization Framework", IETF RFC 6749, Oct. 2012

[94] HashiCopr Consul, "A networking tool for a fully featured service-mesh control plane and service discovery" [Online]. Available: https://github.com/hashicorp/consul [Accessed July 2020]

[95] 3GPP TS 29.598, "Technical Specification Group Core Network and Terminals; 5G System; Unstructured Data Storage Services; Stage 3"

[96] Zuul: "A gateway service for dynamic routing, monitoring, resiliency and security" [Online]. Available: https://github.com/Netflix/zuul [Accessed July 2020]

[97] Restgate: "A Go project implementing a real-time API gateway for the REST protocol with NATS servers as messaging system" [Online]. Available: https://github.com/resgateio/resgate [Accessed July 2020]

[98] L. M. Contreras, D. Lopez, "A Network Service Provider Perspective on Network Slicing", *IEEE Softwarization*, Jan. 2018 [Online]. Available: https://sdn.ieee.org/newsletter/january-2018/a-network-service-provider-perspective-on-network-slicing [Accessed July 2020]

[99] A. Jerichow et al. "3GPP Non-Public Network Security", January 2020

[100] IETF RFC 3748, "Extensible Authentication Protocol (EAP)"

[101] 3GPP TS 33.501, "Security architecture and procedures for 5G System"

[102] https://www.cbronline.com/news/underground-fibre-4g, 2018 [Accessed September 2020]

[103] 3GPP TR 28.807, "Study on management aspects of Non-Public Networks (NPN)"

[104] 5G Core (5GC) Characteristics [Online]. Available: http://www.techplayon.com/5g-core-5gc-characteristics/ [Accessed September 2020]

[105] B. Chatras, "Applying a Service-Based Architecture Design Style to Network Functions Virtualizations", *2018 IEEE Conference on Standards for Communications and Networking (CSCN)*, Paris, 2018, pp. 1-4.

[106] RabbitMQ: "An open-source message broker" [Online] https://www.rabbitmq.com/getstarted.html [Accessed July 2020]

[107] ZeroMQ: "An open-source universal messaging library" [Online] https://zeromq.org [Accessed July 2020]

[108] NATS: "A simple, secure and high performance open-source messaging system" [Online] https://docs.nats.io [Accessed July 2020]

12 Annex A – 5G-CLARITY Concepts

12.1 State-of-the-art concepts

Evolved Packet System (EPS): it represents the 3GPP system architecture for 4G mobile technology. The EPS consists of the User Equipment (UE), the Evolved Universal Terrestrial Access Network (E-UTRAN) and the Evolved Packet Core (EPC).

Evolved Universal Terrestrial Access Network (E-UTRAN): it is the 3GPP access network of the EPS. The E-UTRAN consists of one or more eNBs providing individual UEs with 4G wireless access connectivity, based on the use of Long-Term Evolution (LTE) technology.

evolved Node B (eNB): it is the mobile access node in the E-UTRAN. The eNB represents a logical node providing LTE user plane and control plane termination towards individual UEs, and connected via the S1-MME and S1-U interfaces to the EPC.

Evolved Packet Core (EPC): it is the 3GPP core network of the EPS. First introduced in 3GPP Release 8, it represents a flat, all-IP architecture, so that data traffic can be handled efficiently from performance and costs perspective. In EPC, few network nodes (HSS, Serving GW, Packet Data Network GW and MME) are involved and protocol conversion is avoided.

5G System (5GS): it represents the 3GPP system architecture for 5G mobile technology. The 5GS consists of the User Equipment (UE), the 5G Access Network (AN) and the 5G Core (5GC).

5G Access Network (AN): it is the access network of the 5GS. The 5G AN is formed of a set of access technologies, including 3GPP access technologies (i.e. 5GNR, LTE) and non-3GPP access technologies (e.g. WiFi, fixed).

Next-Generation Radio Access Network (NG-RAN): contained in the 5G AN, it represents the 3GPP access network of the 5GS. The NG-RAN consists of one or more gNBs providing individual UEs with 5G wireless access connectivity, based on the use of 5G New Radio (5GNR) technology.

Next-generation Node B (gNB): it is the mobile access node in the NG-RAN. The gNB represents a logical node providing NR user plane and control plane termination towards individual UEs. The gNB can be connected to the EPC (via the S1-MME and S1-U interfaces) and/or the 5GC (via the N2 and N3 interfaces). In the first case, the result is a 5G Non-Stand Alone (NSA) scenario. In the second case, the result is a 5G Stand Alone (SA) scenario.

NOTE: To take advantage of the 5G RAN virtualization benefits in terms of scalability and centralization, gNB functionality can be optionally be split into three logical modules: the Radio Unit (RU), provisioned with RF circuitry; the gNB Distributed Unit (gNB-DU), hosting gNB real-time functions; and the gNB Centralized Unit (gNB-CU), hosting gNB non-real-time functions.

5G Core (5GC): it is the 3GPP core network of the 5GS. First introduced in 3GPP Release 15, the 5GC leverages on five main principles that go well beyond the EPC design: *(i)* converged core, with the ability to support multiple access technologies from the 5G AN, including 3GPP and non-3GPP technologies ; *(ii)* control user plane separation, allowing independent scalability and evolution of control plane (CP) and user plane (UP) functions; *(iii)* an generic user plane function (UPF) with capabilities that can be programmed by the CP; *(iv)* a SBA for the CP, leveraging on modular NFs with compute and storage separation; *(v)* network slicing support.

Radio Unit (RU): it represents the front-end and antenna subsystems of a gNB. This physical node provides 5G NR air interface to UEs, and typically offers a CPRI or digital Radio over Ethernet (RoE) interface towards the gNB-DU.

gNB Distributed Unit (gNB-DU): it is a logical node hosting RLC, MAC and PHY-High layers of the gNB. One gNB-DU supports one or multiple cells. One cell is supported by only one gNB-DU. The gNB-DU terminates the F1 interface connected with the gNB-CU.

gNB Central Unit (gNB-CU): it is a logical node hosting RRC, SDAP and PDCP protocols of the gNB. One gNB-CU is connected to one or more gNB-DUs, terminating their individual F1 interfaces. The CU is typically virtualized and decomposed into the gNB-CU-Control Plane (gNB-CU-CP) and gNB-CU-User Plane (gNB-CU-UP).

gNB-CU-Control Plane (gNB-CU-CP): a logical node hosting the RRC and the control plane part of the PDCP protocol of the gNB-CU. The gNB-CU-CP terminates the E1 interface connected with the gNB-CU-UP and the F1-C interface connected with the gNB-DU.

gNB-CU-User Plane (gNB-CU-UP): a logical node hosting the user plane part of the PDCP protocol and the SDAP protocol of the gNB-CU. The gNB-CU-CP terminates the E1 interface connected with the gNB-CU-CP and the F1-U interface connected with the gNB-DU.

PLMN Identifier (PLMN ID): it is a unique number identifying a public mobile network used to cover a given geographical area. It consists of a Mobile Country Code (MCC) and a Mobile Network Code (MNC).

Multi-Operator Core Network (MOCN): it is a 3GPP defined mechanism for network sharing, allowing core networks from different Mobile Network Operators (MNOs) to share the same RAN, while being kept separate. MOCN is the most resource efficient solution as it gives the MNOs the opportunity to pool their respective spectrum allocations, resulting in improved trunking efficiency. With current 3GPP specifications, MOCN allows a single physical eNB to connect to up to 6 different EPC instances, and a gNB to connect to up to 12 different 5GC instances.

3GPP slice: it is a logical network that provides specific network capabilities and network characteristics. Each 3GPP slice is uniquely identified by the Single Network Slice Selection Assistance Information (S-NSSAI).

NOTE 1: Currently, 3GPP specifications only allow network slicing in the 5GC. There is an on-going study item on Rel-17 to assess the applicability of network slicing in the 5G AN.

NOTE 2: 3GPP slices are deployed in form of 3GPP slice instances. A 3GPP slice instance is a set of network function instances and the required resources (e.g. compute, storage and networking resources) which collectively form a deployed 3GPP slice. Each 3GPP slice instance is uniquely identified by the Network Slice Instance (NSI ID). The NSI ID identifies the core network part of a 3GPP slice instance when multiple instances of the same 3GPP slice are deployed, and there is a need to differentiate among them (e.g. for multi-tenancy purposes).

Single Network Slice Selection Assistance Information (S-NSSAI): it is the unique identifier for a 3GPP slice. An S-NSSAI is comprised of: *(i)* a Slice/Service type (SST), which refers to the expected Network Slice behaviour in terms of features and services; *(ii)* Slice Differentiator (SD), which is an optional information that complements the Slice/Service type(s) to differentiate amongst multiple Network Slices of the same Slice/Service type. An S-NSSAI can have standard values (i.e. such an S-NSSAI is only comprised of an SST with a standardized SST value, and no SD) or non-standard values (i.e. such an S-NSSAI is comprised of either both an SST and an SD, or only an SST without a standardized SST value and no SD). An S-NSSAI with a non-standard value identifies a single 3GPP slice within the PLMN with which it is associated.

NOTE: Based on the operator's operational or deployment needs, an S-NSSAI can be associated with one or more 3GPP slice instances. Multiple network slice instances (each identified by a unique NSI ID) associated with the same S-NSSAI may be deployed in the same or different Tracking Areas. When multiple 3GPP slice instances associated with the same S-NSSAI are deployed in the same Tracking Areas, the AMF instance serving the UE may logically belong to (i.e. be common to) more than one 3GPP slice instance associated with

this S-NSSAI.

Network Slice Selection Assistance Information (NSSAI): it is a collection of S-NSSAIs. Currently, 3GPP allows up to eight S-NSSAIs in the NSSAI sent in signalling messages exchanged between the UE and the 5GC. This means that a single UE may be served by at most eight network slices at a time.

NOTE: 3GPP specifications defines three NSSAI types: allowed NSSAI, pending NSSAI and configured NSSAI. For more information on these NSSAI types, see [50].

Wi-Fi/LiFi Access Points: they are physical network functions which are used to provide Wi-Fi and LiFi radio services. The Wi-Fi/LiFi access point is conceptually equivalent to RU+gNB-DU in the case of 5GNR.

Service Set Identifier (SSID): it is a unique number identifying the IEEE 802.11 wireless service offered by a Wi-Fi/LiFi AP. SSID is conceptually equivalent to the PLMN ID.

Ethernet switch: it is a physical network function providing L2 Ethernet switching services, including VLAN tagging.

Time Sensitive Networking (TSN) switch: it is an Ethernet switch used for guaranteeing deterministic behavior (e.g. zero jitter, upper bounded latency) over Ethernet networks. This switch supports the IEEE 802.1CM standards.

NFV service: it is a composition of network functions deployed in NFV-based environments, where at least one of these network functions is a Virtualized Network Function (VNF). This composition is materialized through a set of virtual links, according to one or more VNF Forwarding Graphs (VNFFG). Virtual links are abstractions of physical links that logically connect network functions together. To specify how these connections are made along the entire NFV service, one or more VNFFG are used, each including forwarding rules to describe how traffic shall flow between VNFs defined in the in-VNFFG topology.

Virtual Infrastructure Manager (VIM): in the ETSI NFV MANO framework, it is the management function responsible for controlling and managing the NFV Infrastructure (NFVI) resources, usually within one operator's infrastructure domain. The VIM manages the storage, connectivity and compute resources building up the NFVI, keeping an inventory of the allocation of virtual resources to physical resources, thereby allowing for the orchestration of the allocation, upgrade, release and reclamation of NFVI resources and the optimization of their use. The VIM makes these virtualized resources management capabilities available for consumption towards authorized clients, referred to as VIM tenants. An example of a VIM tenant is the NFVO.

NOTE: VIM can be designed for VM-based orchestration (e.g. OpenStack), although recent designs also allows for container-based orchestration (e.g. Kubernetes).

OpenStack project: it is a collection of VIM resources to be allocated for a given OpenStack tenant. Resources from different OpenStack projects are segregated by means of resource quotas.

NFV Orchestrator (NFVO): in the ETSI NFV MANO framework, it is the management function responsible for managing the lifecycle of NFV services, usually within one operator's administrative domain. Instances of NFV services can be deployed over one VIM (single-site NFV service) or more VIMs (multi-site NFV service). The NFVO manages the instances throughout the entire lifetime, from their instantiation to their termination, keeping track of their running resources and governing their behavior according to the rules and requirements described in the corresponding Network Service Descriptors (NSDs). The NFVO makes NFV service management capabilities available for consumption towards authorized clients, referred to as NFVO tenants. An example of a NFVO tenant is the Slice Manager or any other upper layer service orchestrator.

12.2 5G-CLARITY concepts

5G-CLARITY venue: it is a venue, either private (e.g. factory, stadium) or public (e.g. transportation hub), that

provides an execution environment for 5G-CLARITY system realization. 5G-CLARITY services are provisioned using resources from the 5G-CLARITY venue.

5G-CLARITY physical infrastructure: it is the collection of wireless access nodes (i.e. eNBs and gNBs, Wi-Fi/LiFi access points), transport nodes (i.e. ethernet and TSN switches) and compute nodes (i.e. RAN and edge clusters), together with the external-facing gateway, that are deployed within the 5G-CLARITY venue. The resources building up the 5G-CLARITY physical infrastructure are also referred to as on-premise resources.

5G-CLARITY resource-facing services: on-premise infrastructure services that are deployed using 5G-CLARITY physical infrastructure resources. 5G-CLARITY resource-facing services can be categorized into 5G-CLARITY wireless, transport and compute services.

5G-CLARITY customer-facing services: they represent advanced communication / digital services that can be provisioned using the 5G-CLARITY system. These services, hosted on individual 5G-CLARITY slices, can be intended for private use (e.g. industry 4.0 services, banking service) or public use (e.g. mobile broadband experience, smart city service).

5G-CLARITY Virtualized x Function (VxF): it represents a Virtualized Network Function (VNF) or Virtualized Application Function (VAF) executed on a 5G-CLARITY physical infrastructure compute node.

5G-CLARITY Management Function (MF): it is a management entity offering a well-defined set of capabilities in a SBMA. The SBMA building up the 5G-CLARITY management and orchestration stratum consists of different MFs, each producing/consuming management services to/from other MFs.

5G-CLARITY operator: also known to as private NOP, it is the administrative entity that operates the 5G-CLARITY physical infrastructure and software functions built atop, including VxFs and MFs. The private NOP makes use of the on-premise 5G-CLARITY system to provision 5G-CLARITY resource-facing services as well as 5G-CLARITY slices.

5G-CLARITY wireless service: it is the configuration that needs to be set on one or more wireless access nodes, to make them operationally ready. Access nodes from different WATs require the definition of separate 5G-CLARITY wireless services: Wi-Fi/LiFi services and LTE/NR services. A Wi-Fi/LiFi service consists in defining a given SSID over one or more Wi-Fi/LiFi access points, specifying the resource quota corresponding to this SSID. An LTE/NR service consists in defining a given tuple {PLMN ID, NSSAI} over one or more physical gNBs, specifying the resource quota corresponding to this tuple.

5G-CLARITY compute service: it is a composition of Virtual Deployment Units (VDUs) which are executed on the 5G-CLARITY physical infrastructure compute nodes, including RAN and edge clusters. These VDUs may host software images corresponding to VxFs. From a deployment viewpoint, a 5G-CLARITY compute service can be modelled as a fully virtualized NFV service.

5G-CLARITY transport service: it represents the configuration that needs to be set on the transport nodes, in order to make them deliver the traffic from a 5G-CLARITY wireless service into a 5G-CLARITY compute service. For both Ethernet (and optionally TSN switching) devices, a 5G-CLARITY transport service may be signalled using an IEEE 802.1Q VLAN tag.

5G-CLARITY slice: it is a logical partition of the 5G-CLARITY physical infrastructure that provides an isolated execution environment for a 5G-CLARITY tenant. This execution environment is formed of the composition of 5G-CLARITY resource-facing services, including wireless services, compute services and transport services. The resource allocated to a particular slice collectively defined the 5G-CLARITY slice quota, consisting of three individual resource quotas: 5G-CLARITY wireless quota, 5G-CLARITY compute quota and 5G-CLARITY transport quota.

NOTE 1: Unlike a 3GPP slice, which is a PLMN defined network slice, a 5G-CLARITY slice is an on-premise

infrastructure slice.

NOTE 2: 3GPP slicing is a mechanism for multi-service support, while 5G-CLARITY slicing is a mechanism for multi-tenancy support.

5G-CLARITY wireless quota: the set of wireless resources in each gNB or Wi-Fi/LiFi AP which are allocated to one 5G-CLARITY slice instance. The implementation of a 5G-CLARITY wireless quota depends on the underlying wireless technology. For example, it could be expressed as namely Physical Radio Blocks (PRBs) in 5GNR, airtime in Wi-Fi, and wavelengths or airtime in LiFi.

5G-CLARITY compute quota: the set of computing resources (e.g. CPUs), storage (e.g. RAM) and networking resources (i.e. NIC) which are allocated to one 5G-CLARITY slice instance. These resources are provided by the cluster SP (see Figure 3.2). A 5G-CLARITY compute quota may be implemented using OpenStack.

5G-CLARITY transport quota: the set of transport resources allocated to one 5G-CLARITY slice instance. The available transport resources depend on the underlying transport technology and could be expressed in terms of data-rate, latency or buffer space.

5G-CLARITY tenant: it is the administrative entity that consumes a 5G-CLARITY slice. Examples of 5G-CLARITY tenants include communication/digital service providers, either public or private, MNOs (i.e. public NOPs) and hyperscalers (i.e. large-scale cloud providers). A 5G-CLARITY tenant can use the received 5G-CLARITY slice to provision one or more 5G-CLARITY customer-facing services atop.

NOTE: 5G-CLARITY operator and 5G-CLARITY tenant follow provider-customer relationships as follow: the 5G-CLARITY operator behaves as 5G-CLARITY slice provider, while the 5G-CLARITY tenant behaves as 5G-CLARITY slice customer.

5G-CLARITY User Equipment (UE): it is a device that has one or more wireless interfaces of type 5GNR, LTE, Wi-Fi or LiFi and has been provisioned with the credentials required to access a 5G-CLARITY slice.

5G-CLARITY CPE: A 5G-CLARITY CPE is an especial type of 5G-CLARITY UE that contains at least one 5GNR, one Wi-Fi and one LiFi interface to connect to a 5G-CLARITY slice. A 5G-CLARITY CPE also includes an Ethernet interface to receive traffic from non-5G-CLARITY devices located inside the private venue. Likewise, a 5G-CLARITY CPE can provide non 5G-CLARITY devices access into a 5G-CLARITY slice.

13 Annex B – 5G-CLARITY Service Delivery Models

This section delves into the 5G-CLARITY service delivery models that were originally presented in Table 3-2.

13.1 WAT as a service

A pre-requisite for any MNO (public NOP + public CSP/DSP) in B5G scenarios is to provide ubiquitous connectivity to his public subscribers, even when they are in indoor venues that are far beyond the MNO's coverage area. Examples of these venues include stadiums, museums, or transportation hubs (e.g. airports, underground stations). In such venues there are dense concentrations of end-users that demand advanced communication/digital services, including mobile broadband experience enhanced with VR/AR technology. However, appropriately covering these venues is quite expensive, hence it is difficult to define a viable business model from the perspective of a single MNO. Instead, an alternative business model is emerging based on the figure of a **neutral host**, whereby the venue owner invests in on-premise 5G network infrastructure, which are used for its own purposes, and leased to different MNOs. A relevant example is a recent tender for the London Underground where the contract winner will secure a 20-year concession to provide a neutral host network in the tunnels, stations and platforms [102].

Neutral hosting model is beneficial for both the private venue owner and the MNOs. On the one hand, the private venue owner (neutral host) monetizes the in-house infrastructure by selling wholesale and mobile localized coverage solutions to the different hosted MNOs[9]. On the other hand, the MNO can increase its coverage area without the need to invest in on-premise equipment, thus expanding its service footprint at a much more reduced cost.

WAT as a Service (WATaaS) is seen as a future-proof realization of the neutral hosting model in 5G-CLARITY. In essence, it is a service delivery model whereby a private NOP makes 5G-CLARITY wireless services available for public use, allowing any public NOP to set up connectivity services to on-premise access points, including gNBs and Wi-Fi/LiFi access points. The MNO (taking the role of NOP) leverages the allocated quotas to extend his coverage are to up the access nodes where the public subscribers are connected to. This allows the MNO (taking the role of public CSP/DSP) to serve its subscribers within the private venue, e.g. providing them with 5G voice and data services.

Within the WATaaS model, different variants can be found, including:

- NR as a Service: The 5G-CLARITY wireless service instance offered to a public NOP includes a set of dedicated PRBs from one or more gNBs.
- Wi-Fi as a Service: The 5G-CLARITY wireless service instance offered to a public NOP includes a set of dedicated airtime frames from one or more Wi-Fi access points.
- LiFi as a Service: The 5G-CLARITY wireless service instance offered to a public NOP includes a set of dedicated airtime frames and/or wavelengths from one or more LiFi access points.
- Multi-WAT as a Service: Any combination of the three above, e.g., LiFi-Wi-Fi as a Service, 5GNR-Wi-Fi-LiFi as a Service.

The WATaaS allows 5G-CLARITY to deliver neutral hosting functionality to MNOs, together with the benefits from the 'aaS' approach, including dynamicity (i.e. the MNO can acquire wireless resources from the neutral host on-demand) and 'pay-as-you-go' (i.e. the MNO pays for the use of these resources when used). A typical

[9] The set of neutral host's managed resources collectively act as a shared resource pool to the hosted MNOs. The neutrality aspect in this context does not imply strict quality between hosted MNOs, as the resources offered to each MNO are subjected to commercial agreement between the neutral host and the MNO, and policy-based management for resource dispatching may be applied.

example is an MNO that want to book resources in the metro station next to the soccer stadium only the match day, and so it would only pay for this use.

13.2 NFV Infrastructure as a service

NFV Infrastructure as a Service (NFVIaaS) is a service delivery model whereby a virtual infrastructure is made available for VNF hosting. These VNFs can be from the NFVIaaS provider itself, but also from external network providers. For the NFVIaaS provider, this service provides for economies of scale. The infrastructure is sized to support the provider's own needs for deploying VNFs and extra capacity that can be sold to one or more NFVIaaS customers. The provider dispatches this residual capacity to the different customers in the form of dedicated compute resource quotas, leveraging on existing multi-tenancy solutions like those available in OpenStack. One solution in this regard is the provision of separate OpenStack projects for different customers. Every NFVIaaS customer uses the resource quota provided in the allocated project to deploy and execute his own VNFs with great agility and flexibility.

In 5G-CLARITY ecosystem, NFVIaaS can be provisioned in both directions: from the private to the public administrative domains (i.e. private NOP → public NOP), or the other way around (i.e. public NOP → private NOP).

In the first case, which will be the main focus of 5G-CLARITY project, it is the private NOP who plays the role of NFVIaaS provider, making the on-premise capacity available for public use. This means that (part of) the resources from the edge and RAN cluster deployed in-house can be used to host the VNFs from one or more public NOPs, who behave as individual NFVIaaS customers. It is worth noting that this NFVIaaS scenario constitutes another realization of the neutral hosting model. Unlike WAT as a Service, what the neutral host provides to the different MNOs are compute resource quotas for the execution of public VNFs, i.e. VNFs used for serving public subscribers

In the second case, it is the public NOP who plays the role of NFVIaaS provider. It offers (part of) the capacity available on the telco cloud[10] for the execution of private VNFs from one or more private NOPs. For this end, the public NOP makes use of IaaS capabilities in a similar way as hyperscalers have traditionally done[11]. Examples of VNFs that are typically offloaded to the telco cloud include resource-hungry VNFs which do not process delay-sensitive traffic, nor store sensitive data.

13.3 Slicing as a service

Slicing as a Service (SlaaS) is a service delivery model whereby a business entity (taking the role of slice provider) provides an slice instance to another entity (taking the role of slice customer). It represents a leap beyond WATaaS and NFVIaaS. Unlike the two la, where service offerings are individual resource quotas, either wireless (as in WATaaS) or compute (as in NFVIaaS), SlaaS allows providing operationally isolated execution environment that the customer can use for their own purposes. This innovative service delivery model can be used in both directions: from the private to the public administrative domain, or the other way around.

In the first scenario, the slice provider is the private NOP, and the offered slice is an on-premise 5G-CLARITY slice, i.e. infrastructure slice. In such a case, the slice customer could be any of the 5G-CLARITY tenants described in Section 3.2.3: a CSP/DSP, a public NOP or an hyperscaler.

[10] The telco cloud is the set of Points of Presence (PoPs) which are under within the management scope of a public NOP. Examples of these PoPs include edge and regional data centers.

[11] The ability of hyperscalers to behave as NFVIaaS providers in 5G-CLARITY will be discussed in the second phase on the project.

In the second scenario, the slice provider is the public NOP, and the offered slice is a 3GPP slice, i.e. network slice. The customers of this slice could be a CSP/DSP, an hyperscaler or a private NOP. The first two are out of scope of 5G-CLARITY, as they do not involve the use of private infrastructure. When private NOP is the slice customer, the SlaaS are used to building a PNI-NPN. Figure 13.1 shows an example of this scenario. In this example, the PNI-NPN is an E2E network composed of two differentiated segments: one *private*, including nodes and VNFs deployed using on-premise resources; and *one public*, consisting of VNFs built upon PLMN resources. The public NOP delivers public segment to the private NOP in the form of a 3GPP slice instance, so that private NOP can aggregate the private segment. For more information on this scenario, named to as Network Slice as a Service (NSaaS) in 3GPP terminology, see [103].

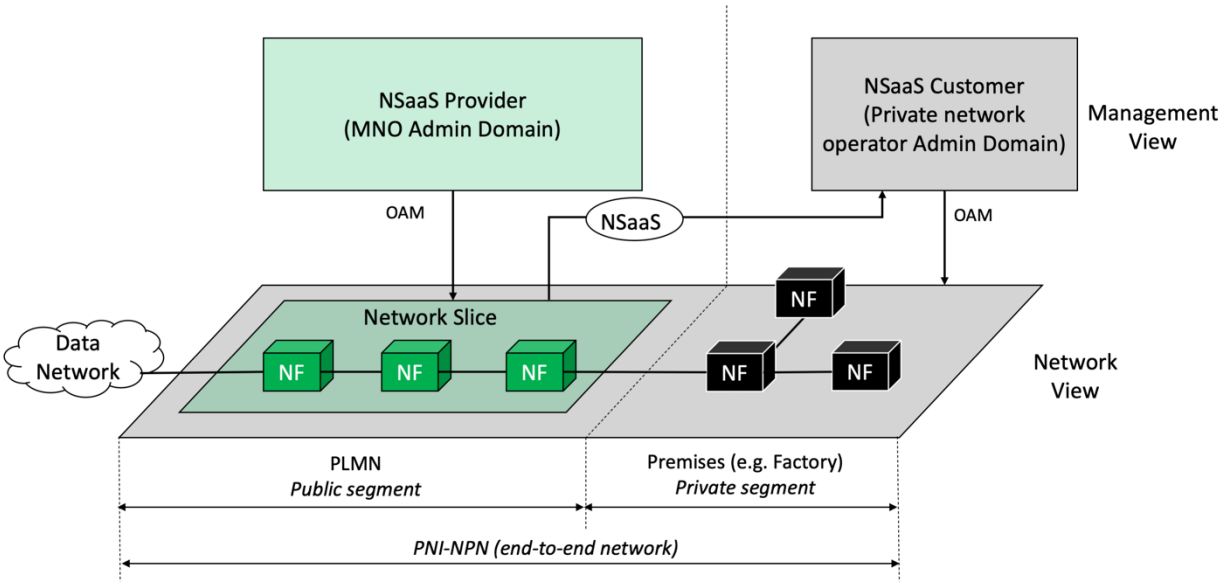

Figure 13.1: NSaaS for PNI-NPN provisioning

13.4 Intelligence as a service

In this fourth service delivery model, operation support providers (typically system integrators) provide different NOPs with ready-to-use toolkits imbued with AI and intent services. The built-in intelligence of these toolkits can help NOPs to simplify the day-to-day operation of their networks, increasing their automation as well. To facilitate the provisioning of this intelligence, operation support providers register the developed toolkits into a unified marketplace which is accessible to any NOP, including private and public NOPs. Every NOP can browse this marketplace and subscribe to the AI and intent services it may need. With this Intelligence as a Service, the NOP does no longer need to hire experienced professional for developing, implementing, and maintaining AI-based functionality for their business logic. Indeed, they only need to subscribe to needed services and invoke these on-demand, leveraging on the Platform as-a-Service consumption patterns.

14 Annex C – Implementation Details on the SBMA

SBA is an architectural style that places emphasis on the services provided by individual architectural components, rather than on the relationships between pre-defined pairs of architectural components. SBA is expected to enable flexible, rapid development and deployment of 5G services, as it becomes possible for a new architectural component to connect to existing components without introducing specific new reference points. Its adoption in standard-based network solutions is already a reality. Proof of this is the design of 5GC control plane (5GC-CP), whereby 3GPP has decide to migrate from traditional point-to-point network architecture towards a SBA. This represents a paradigm shift, based on the four key ideas:

- SBA replaces network elements (NE) with network functions (NFs). Unlike NEs, traditionally executed on purpose-built hardware equipment tied to the claws of vendor lock-in, NFs can benefit from NFV technology.
- Cloud-native NFs, whereby NF can be containerized and scaled independently.
- Individual NFs provide services to other NFs using a Service-Based Interface (SBI). This SBI exposes NF services, making them available for external consumption using RESTful HTTP-based APIs.
- Reference Point interfaces are replaced with a common bus to connect all NFs.

The above ideas make communication among NFs to be like a service mesh functions rather than serial chaining, which contributes to reduction in dependency between each interface and helps in independent scaling of each function (see Figure 14.1). As a result, the agility of having new features and services across network functions is increased.

When SBA style is applied to management and orchestration systems, the result is a SBMA. An example can be found in [105], where the author proposed a solution on how to migrate ETSI NFV-MANO framework from traditional interface-centric specifications towards a SBMA.

As discussed in Section 4.2.3, 5G-CLARITY system will follow SBMA solution for the design and development of the management and orchestration stratum. Figure 4.12 provides the baseline solution of SBMA for 5G-CLARITY, which is defined around the concept of management service.

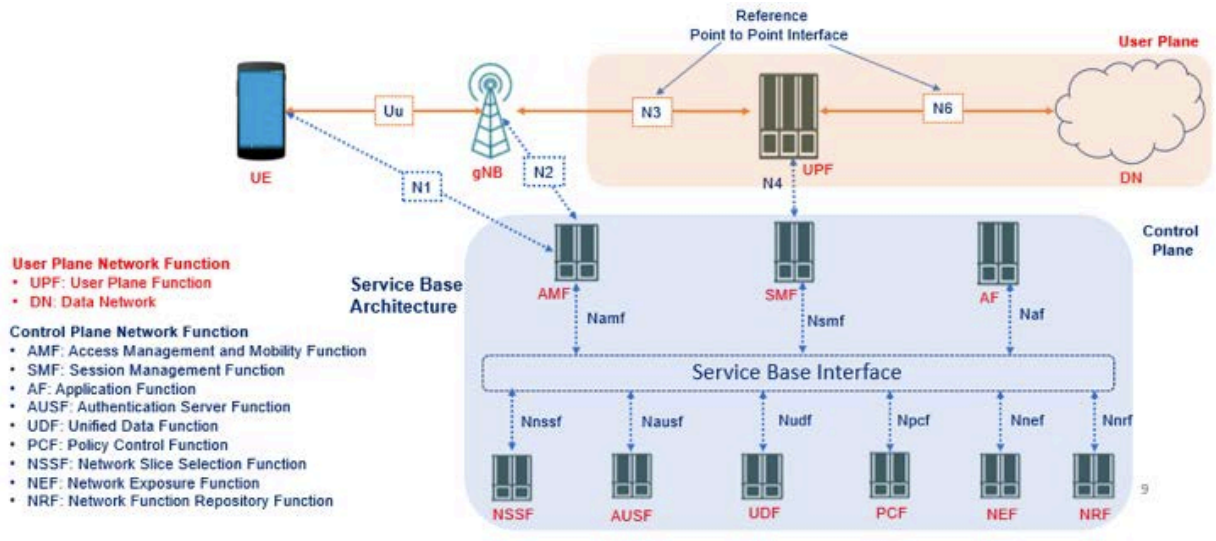

Figure 14.1: 3GPP 5GC [104].

A Management Service is the fundamental building block of any SBMA. It is a standalone service that offers a set of management capabilities for innovation and communication purposes within a well-defined environment. The scope of this environment typically covers a *single management aspect* (e.g. provisioning, performance assurance, fault supervision) on *a single network entity* (e.g. slice, network service, etc.).

Every management service is provided by a *management service producer* and can be consumed by one or multiple *management service customers.* The capabilities of a given management service are accessed by management service via a standard service interface, which conveys the following two artifacts:

- A group of management operations (e.g. create, read, update, delete, subscribe/unsuscribe) and/or notifications, providing primitives to view and manipulate objects according to the management aspects the management service is designed for. These primitives are network-agnostic, in the sense they do not include information about the semantics of the management objects. The implementation of these primitives is typically based on RESTful HTTP-based APIs, although other protocols (e.g. RESTCONF) can also be used
- An information model, specifying which network entity is managed using the management service. This information model describes the semantics of the class representing that network entity. This semantics (relationships, constraints) allows associating objects with instances of that network entity. Information model definitions, typically specified using protocol-agnostic language like UML, are mapped into data model definition used for implementation. A data model is the realization of an information model using a given protocol solution. Yet Another Markup Language (YAML) or YANG are examples of data modelling languages that can be used to that end.

To make analogy to SBA, where the concepts of "NF" and "NF service" are used, the SBMA also makes use of "MF" and "MF service". A MF service represents a management service that is exposed by a MF (taking the role of management service producer) to authorized MFs (taking the role of management service consumer) though a service-based interface. An MF can expose one or more services to other MFs . Similarly, a MF can consume one or more services from other MFs. Figure 4.12 illustrates this scenario.

Figure 14.2 provides a simplified view of the internal composition of a MF. As it can be seen, a MF is composed of one or MF services (see Figure 14.2a), each offering a group of model-driven primitives (i.e. management operations and/or notifications pinned to a specific data model) through a service based interface (see Figure 14.2b).

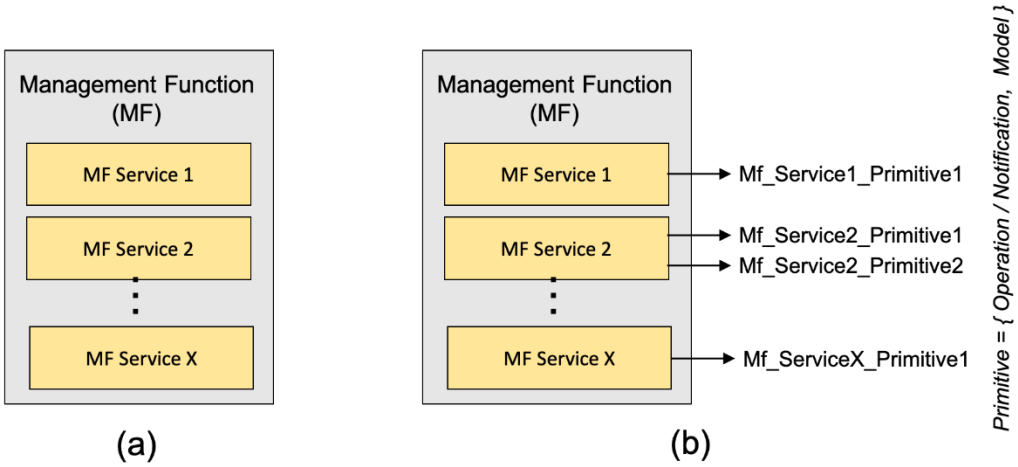

Figure 14.2: MF and MF service concepts.

The introduction of SBMA brings two additional characteristics compared to interface-based management systems: dynamic service registration and discovery, and the use of a persistent data storage service.

On the one hand, dynamic registration and discovery relies on a registry where services available in a MF are registered and can be discovered by other services in the same or different MFs. Figure 14.3 illustrates these actions. When a MF instance is deployed, the MF services it provides and the mechanisms to invoke them are registered. Similarly, if a MF instance is removed from the network, the corresponding entries are deleted

from the registered. A MF that requires a particular management service can then discover and select a MF instance that provides this service. Such a registry performs a similar functionality as the 3GPP Network functions Repository Function (NRF) do in the SBA for 5GC-CP.

On the other hand, the use of persistent data storage service allows for having a common data layer for the entire management and orchestration system, thereby allowing for stateless MFs. SBMA, as a design style, does not require common data storage approach, though goes well with it. Some MFs taking part in the SBMA can expose a data storage service, which other services can use to store any kind of data, including state information.

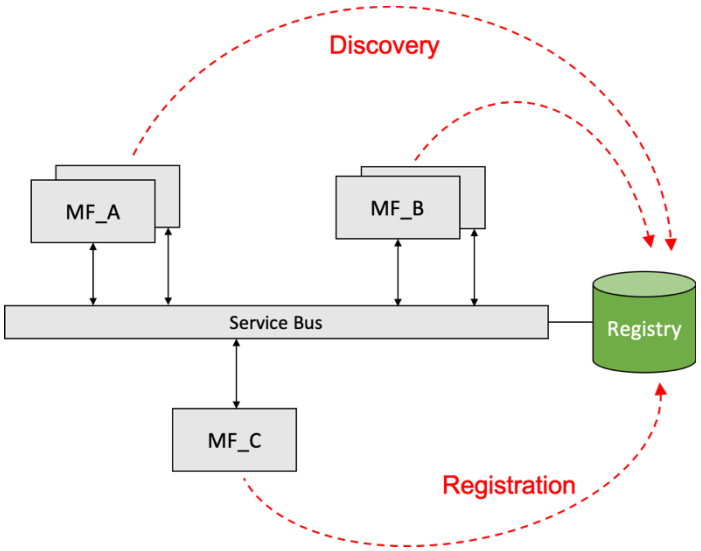

Figure 14.3: Service registration and discovery.

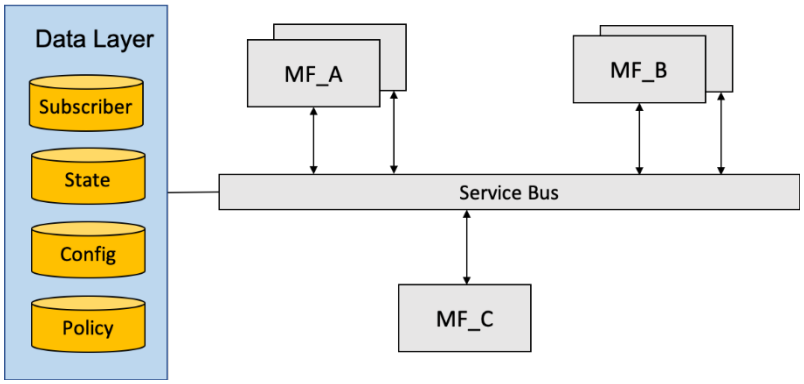

Figure 14.4: Common data layer.

Separation between processing and data means that all data used by services are stored in a logically centralized data repository, as shown in Figure 14.4. This is intended to increase resilience of the overall system, allowing easy scalability (i.e. the ability to deploy multiple instances of the same MF) and facilitating data sharing across multiple services (E.g. services from instances of the same MF). Such a repository performs a similar functionality as the 3GPP Unified Data Repository (UDR) and the Unstructured Data Storage Function (UDSF) do in the SBA for 5GC-CP.

Once the SBMA principles are clarified, the next step is to undestand how MFs can interact with each other. The interactions across MFs in the SBMA can follow two mechanisms:

- Request-Response (Figure 14.5a): MF_B (management service producer) is requested by another MF_A (management service consumer) to provide a certain MF service. MF_B provides a MF service

based on the request by MF_A. To fulfill the request, MF_B may in turn consume MF services from other MFs. For the Request-Response mechanism, communication is one-to-one between consumer and producer. A one-time respone from the producer to a request from the consumer is expected within a certain timeframe.

- Subscribe-Notify (Figure 14.5b): MF_A (management service consumer) subscribes to a MF service offered by another MF_B (management service producer). One or more MFs may subscribe to the same MF service. MF_B notifies all subscribed MFs about the results of this MF service. The subscription request normally includes: *(i)* the notification endpoint, e.g. the notification URL, of the MF service consumer; *(ii)* the need of receiving periodic updates; *(iii)* events that are interests for notification, e.g. the content of information requested has changed, reaches certain thresholds, etc.

All the above interactions across MFs are performed through the service bus. The service bus interconnecting the MFs and the management services they provide is an evocative expression, used in analogy with a computer hardware bus [105]. The SBMA does not mandate a particular type of communication bus. In a rather primitive form, the communication services provided by the bus can be limited to basic IP routing (e.g. when the communication bus is implemented as a layer 3 VPN overlaid on the network of a data center). In such cases a MF willing to consume a management service must have the processing logic to discover and select a MF instance providing this service. This typically involves retrieving a list of candidate instances and selecting one of them according to a load balancing algorithm and querying the Domain Name Service (DNS) to determine its IP address. More advanced forms of the service bus can reduce complexity at the client side, by providing application-layer message routing or by distributing messages according to a publish-subscribe pattern. The aforementioned registry functionality is them embbeded together with the communication service, in the service bus functionality; thereby enabling client MFs to offload the selection of a target management service and the determination of its IP address. Advanced communication buses may even provice fast failover functionality and message transformation capabilities to enabling connectiong non-compatible clients and servers.

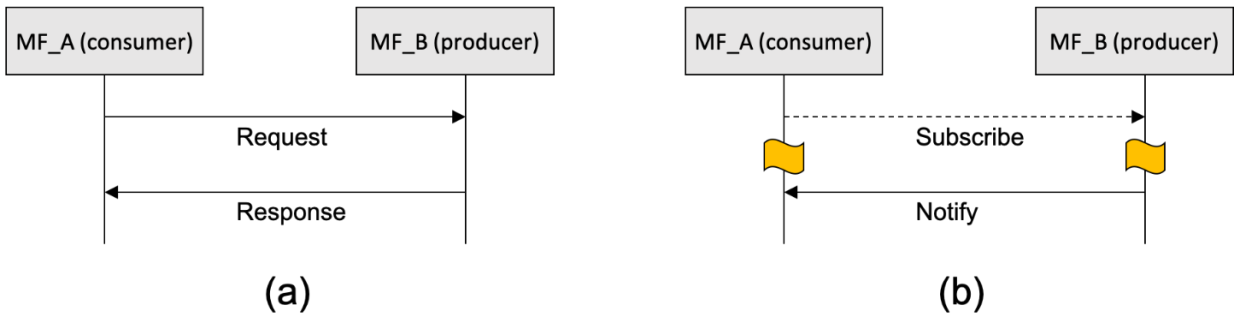

Figure 14.5: Illustrations for "Request-Response MF service" and "Subscribe-Notify MF service" interactions.

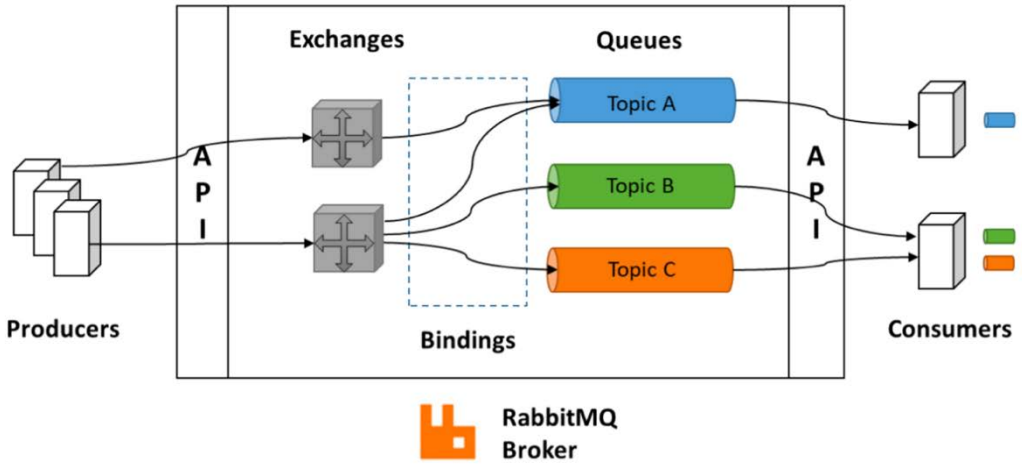

Figure 14.6: RabbitMQ Message Broker Concept (with topic-based queues).

For SBMA, a good candidate technology for service bus is message broker. Relevant message brokers are RabbitMQ [106], ZeroMQ [107] or NATS [108]. Message brokers provide message validation, transformation and routing functionality, as well as advanced mechanisms for failover management. For an example on RabbitMQ utilization, see Figure 14.6.

15 Annex D – 5G-CLARITY Slicing Enabled SNPNs

This annex presents the usability of 5G-CLARITY slicing solution in SNPN scenarios. Unlike the scenarios presented in Section 9, where slicing is intended to provide means for PNI-NPN provisioning, the use of slicing for SNPNs allows supporting the heterogeneity of on-premise private services (in terms of functionality, traffic and performance demands) that can coexist inside the private venue. Figure 15.1 captures the essence of this scenario.

Figure 15.2 sketches an industrial SNPN for providing wireless access within an industrial Operational Technology (OT) domain, in which there might be multiple industrial applications with diverse and sometimes opposing requirements. In the specific case of the scenario shown in Figure 15.2, note that here are three simultaneous slices, namely, X, Y, and Z.

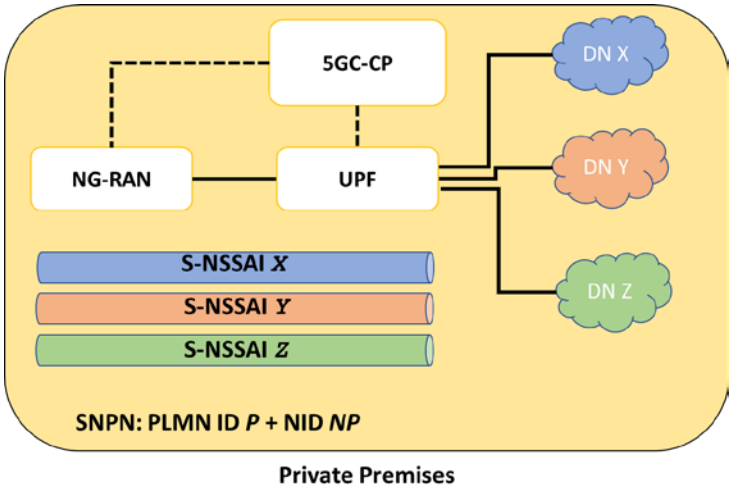

Figure 15.1: Slicing enabled SNPN.

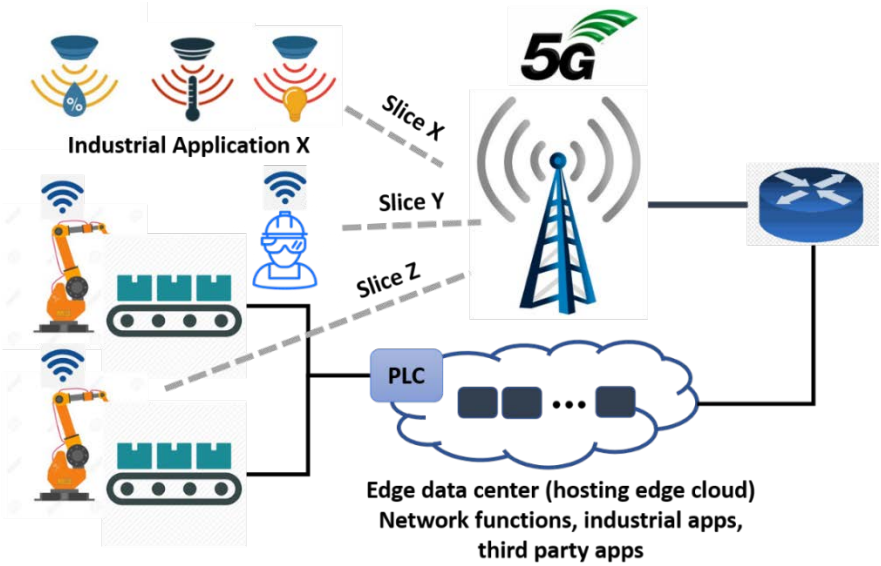

Figure 15.2: Industrial scenario leveraging multiple slices within a SNPN for accommodating different industrial applications with heterogeneous demands in terms of network functionality and performance constraints.