



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

5G-CLARITY Deliverable D2.4

Final System Architecture and Its Evaluation

Contractual Date of Delivery:	March 31, 2022
Actual Date of Delivery:	October 26, 2022
Editor(s): Author(s):	Anna Tzanakaki, Markos Anastasopoulos (UNIVBRIS/IASA) Jose Ordonez-Lucena (TID), Daniel Camps-Mur (I2CAT), Markos Anastasopoulos, Alexandros Manolopoulos, Viktoria Maria Alevizaki, Petros Georgiadis, Anna Tzanakaki (UNIVBRIS/IASA), Carlos Colman Meixner, Shuangyi Yang (UNIVBRIS), Hamada Alshaer (USTRATH), Antonio Garcia, Kiran Chackravaram (ACC), Jonathan Prados-Garzon, Lorena Chinchilla-Romero, Pablo Ameigeiras, Pablo Muñoz (UGR), Daniel González Sánchez, Ignacio Domínguez Martínez-Cas. (UPM), Rui Bian (PLF), Tezcan Cogalan (IDCC), Joseph McNamara (LMI), Navaneetha Channiganathota Manjappa, Ramya Vasist, Meysam Goodarzi, Jesús Gutiérrez, Vladica Sark (IHP), Mir Ghoraishi (GIGASYS), Akshay Jain (Neutron – In kind)
Work Package:	WP2
Target Dissemination Level:	Public

This document has been produced during 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

Rev	Date	Editor /Commentator	Description of Edits
0.01	01.10.2021	Jose Ordonez-Lucena (TID)	Master document created
0.05	15.10.2021	Markos Anastasopoulos (IASA)	ToC created
0.1	22.10.2022	Markos Anastasopoulos (IASA) Jose Ordonez-Lucena (TID)	ToC refined and work plan agreed (contributors and timing).
0.2	15.11.2021	Carlos Colman Meixner (UNIVBRIS) Kiran Chackravaram (ACC) Daniel Camps-Mur (I2CAT) Anil Yesilkaya (USTRATH) Joseph McNamara (IHP)	First draft of Section 2 available. UNIVBRIS: Section 2.1 USTRATH, ACC: Section 2.2 I2CAT: Section 2.2 LMI: Section 2.4
0.3	05.12.2021	Carlos Colman Meixner (UNIVBRIS) Jonathan Prados-Garzon, Pablo Ameigeiras (UGR) Daniel González Sánchez, Ignacio Domínguez Martínez-Casanueva (UPM)	First draft of Section 3 available. UNIVBRIS: Section 3.1 UGR: Sections 3.1 and 3.2 UPM: Section 3.3
0.4	07.01.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	First draft of Section 4 available UNIVBRIS/IASA: Section 4.1 UGR: Section 4.2 USTRATH: Section 4.3 IHP: Section 4.4
		Carlos Colman Meixner (UNIVBRIS) Kiran Chackravaram (ACC) Daniel Camps-Mur (I2CAT), Akshay Jain (Neutron) Anil Yesilkaya (USTRATH) Joseph McNamara (IHP)	Second draft of Section 2 available. UNIVBRIS: Section 2.1 USTRATH, ACC: Section 2.2 I2CAT, Neutron: Section 2.2 LMI: Section 2.4
0.5	14.01.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	First draft of Section 5 available UNIVBRIS/IASA: Section 5.1 UGR: Section 5.2 USTRATH: Section 5.3 IHP: Section 5.4
0.55	21.01.2022	Jose Ordonez-Lucena (TID) Daniel Camps-Mur (I2CAT)	Section 2 external review
		Carlos Colman Meixner (UNIVBRIS) Jonathan Prados-Garzon, Pablo Ameigeiras (UGR) Daniel González Sánchez, Ignacio Domínguez Martínez-Casanueva (UPM)	Second draft of Section 3 available. UNIVBRIS: Section 3.1 UGR: Sections 3.1 and 3.2 UPM: Section 3.3
0.6	28.01.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	Second draft of Section 4 available UNIVBRIS/IASA: Section 4.1 UGR: Section 4.2 USTRATH: Section 4.3 IHP: Section 4.4
0.7	04.02.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	First draft of Section 6 available UNIVBRIS/IASA: Section 6.1 UGR: Section 6.2 USTRATH: Section 6.3 IHP: Section 6.4
		Markos Anastasopoulos (IASA), Vladica Sark (IHP)	Section 3 external review
0.75	11.02.2022	Tezcan Cogalan (IDCC) Daniel González Sánchez, Ignacio Domínguez Martínez-Casanueva (UPM)	Section 4 external review
		Anna Tzanakaki (IASA/UNIVBRIS)	First draft of Sections 1 and 8
0.8	18.02.2022	Tezcan Cogalan (IDCC) Daniel González Sánchez, Ignacio Domínguez Martínez-Casanueva (UPM)	Sections 5 and 6 external review
		Carlos Colman Meixner (UNIVBRIS)	Final draft of Section 2 available.

		Kiran Chackravaram (ACC) Daniel Camps-Mur (I2CAT), Akshay Jain (Neutron – In kind) Anil Yesilkaya (USTRATH) Joseph McNamara (IHP)	UNIVBRIS: Section 2.1 USTRATH, ACC: Section 2.2 I2CAT, Neutron: Section 2.2 LMI: Section 2.4
0.85	25.02.2022	Carlos Colman Meixner (UNIVBRIS) Jonathan Prados-Garzon, Pablo Ameigeiras (UGR) Daniel González Sánchez, Ignacio Domínguez Martínez-Casanueva (UPM)	Final draft of Section 3 available. UNIVBRIS: Section 3.1 UGR: Sections 3.1 and 3.2 UPM: Section 3.3
		Jose Ordonez-Lucena (TID)	First draft of Section 7 available
0.9	04.03.2022	Jose Ordonez-Lucena (TID)	Sections 1 and 8 external review
		Mir Ghoraiishi (GIGASYS)	Section 7 external review
0.91	11.03.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	Final draft of Section 5 available UNIVBRIS/IASA: Section 5.1 UGR: Section 5.2 USTRATH: Section 5.3 IHP: Section 5.4
0.92	18.03.2022	Markos Anastasopoulos, Alexandros Manolopoulos (UNIVBRIS / IASA) Lorena Chinchilla-Romero, Pablo Muñoz (UGR) Hamada Alshaer (USTRATH) Meysam Goodarzi, Navaneetha Channiganathota Manjappa (IHP)	Final draft of Section 6 available UNIVBRIS/IASA: Section 6.1 UGR: Section 6.2 USTRATH: Section 6.3 IHP: Section 6.4
		Jose Ordonez-Lucena (TID)	Final draft of Section 7 available
0.93	25.03.2022	Anna Tzanakaki (IASA/UNIVBRIS) Jose Ordonez-Lucena (TID)	First draft of Sections 1 and 8 available.
0.94	01.04.2022	Jose Ordonez-Lucena (TID)	Executive summary available
0.95	09.04.2022	Markos Anastasopoulos (IASA/UNIVBRIS)	Merging final drafts into D2.4 master doc and editorial work
0.96	14.04.2022	Jose Ordonez Lucena (TID) Mir Ghoraiishi (GIGASYS)	First proof reading
1.0	09.05.2022	Jose Ordonez Lucena (TID) Mir Ghoraiishi (GIGASYS)	Final reading and finalization
1.1	25.10.2022	Markos Anastasopoulos (IASA/UNIVBRIS), ALL	Addressing recommendations from the Interim Review, Final reading and finalization
2.0	26.10.2022	Jesús Gutiérrez (IHP) Mir Ghoraiishi (GIGASYS)	Final revision and submission of the deliverable

Table of Contents

List of Acronyms	10
Executive Summary	12
1 Introduction	13
1.1 Scope and objectives of this document	13
1.2 Document structure	14
2 Revised Architecture	15
2.1 Infrastructure stratum	15
2.2 Network function and application stratum	16
2.3 Extensions to Management and Orchestration stratum	23
2.3.1 Impact of O-RAN management plane on 5G-CLARITY architecture	23
2.3.1.1 Review of O-RAN management architecture	23
2.3.1.2 Integration between 5G-CLARITY and O-RAN management stratum	24
2.3.2 Cloud native subsystem: scalability analysis of M&O stratum	27
2.3.2.1 Single-tenant vs multi-tenant analysis	27
2.3.2.2 Telemetry subsystem: Scalability of data storage requirements	28
2.4 Intelligence stratum	30
2.4.1 Intent engine operation	31
2.4.2 Communication with Data Lake	31
2.4.3 AI Engine to Data Lake Interface	31
3 Analysis of Advanced Features and Application Scenarios on NPNs	32
3.1 Architecture and operation of PNI-NPNs	32
3.1.1 PNI-NPN standardization	32
3.1.2 PNI-NPN deployment considerations	33
3.2 Multi-site NPNs	34
3.2.1 Example of 5G-CLARITY framework architecture in single- and multi-site NPNs	35
3.2.1.1 Single Site NPNs	35
3.2.1.2 Multi-site NPNs	36
3.2.2 Mobility in NPNs	37
3.2.3 Mobility in PNI-NPNs	38
3.2.4 Mobility in SNPNs	38
3.2.5 5G-CLARITY system enhanced roaming support	39
3.2.6 Example of flows for call and connection requests on Single and Multi-Sites NPN	41
3.3 Data aggregation in NPN-PLMN scenarios	43
3.3.1 Data semantic fabric management services	43
3.3.2 Use case: data aggregation in NPN-PLMN using DSF	44
4 Final 5G-CLARITY system evaluation	48
5 Mobility in public-private 5G industrial networks	50
5.1 End-to-end modelling	51
5.2 Scenario Details	54

5.2.1	Scenario setup	54
5.2.2	Scenario validation	58
6	Multi-WAT capabilities	62
6.1	End-to-end modelling tools	63
6.1.1	Multi-WAT offloading in industry 4.0	64
6.1.2	Mobility and traffic offload management in Wi-Fi/LiFi integrated networks	66
6.1.3	Smart network user mobility algorithm	68
6.2	Multi-WAT offloading in industry 4.0	69
6.2.1	Scenario setup	69
6.2.2	Scenario validation	71
6.3	Mobility and traffic offload management in Wi-Fi/LiFi integrated networks	74
6.3.1	Scenario setup	74
6.3.2	Scenario validation	75
6.4	Positioning using multi-WAT	79
6.4.1	TDoA/ToA model	80
6.4.2	TWR model	82
6.4.3	RSS based model	83
6.4.4	Geometric method model	84
7	Joint Synchronization and Localization	86
7.1	End-to-end modelling tools	86
7.2	Modelling of functional components	87
7.2.1	Introduction	87
7.2.2	NLoS identification and Angle of Arrival	88
7.2.2.1	NLoS identification and channel impulse response	88
7.2.2.2	Angle of Arrival	89
7.2.3	Particle Gaussian Mixture Filter	89
7.2.4	Scenario validation	90
8	Pathways Towards 6G	95
8.1	6G ecosystem	95
8.1.1	State-of-the-art	95
8.1.2	5G-CLARITY approach to 6G	96
8.2	6G technology drivers	97
8.2.1	State-of-the-art	97
8.2.2	5G-CLARITY approach to 6G	99
8.3	6G use cases	100
8.3.1	State-of-the-art	100
8.3.2	5G-CLARITY approach to 6G	102
8.4	6G Architecture Vision	103
8.4.1	State-of-the-art	103
8.4.2	5G-CLARITY approach to 6G	104

9	Conclusions.....	106
10	Bibliography.....	107

List of Figures

Figure 2-1 5G-CLARITY final system architecture	15
Figure 2-2 5G-CLARITY infrastructure stratum architecture (5G-CLARITY D2.2) [1]	16
Figure 2-3 5G-CLARITY network and application function stratum.....	17
Figure 2-4 Initial UP architecture (5G-CLARITY D3.1) [4].....	18
Figure 2-5 5G-CLARITY virtual testbed architecture (5G-CLARITY D3.2) [5]	19
Figure 2-6 5G-CLARITY CU/DU/RU integration architecture, Phase 1: F1 and NG integration (5G-CLARITY D3.2) [5]	20
Figure 2-7 5G-CLARITY CU/DU/RU integration architecture, Phase 2: NR-UU integration and testing with Research RU and commercial UE (5G-CLARITY D3.3)	20
Figure 2-8 5G-CLARITY CU/DU/RU integration architecture, Phase 3: commercial third-party RU and commercial UE integration	21
Figure 2-9 Testbed used for Wi-Fi and LiFi integrated L2 network (5G-CLARITY D3.2) [5]	21
Figure 2-10 5G-CLARITY eAT3S xApp environment (5G-CLARITY D3.2) [5]	22
Figure 2-11 5G-CLARITY localization server architecture (5G-CLARITY D4.2) [6]	23
Figure 2-12 O-RAN SMO architecture with R1 and A1 interfaces. Original image from [7]	24
Figure 2-13 5G-CLARITY management and orchestration stratum from [1]	25
Figure 2-14 5G-CLARITY intelligence stratum from [1]	25
Figure 2-15 Integration between O-RAN and 5G-CLARITY management stratums	26
Figure 2-16 Single tenant VS multi-tenant deployment of M&O stratum	27
Figure 2-17 Example of 5G-CLARITY RAN data exporter	29
Figure 2-18 Example of 5G-CLARITY edge cluster exporter.....	29
Figure 2-19 Data exporter scalability analysis: number of base stations (left), number of cells in 1 base station (middle), number of UEs connected to 1 cell (right)	30
Figure 2-20 Intelligence stratum with new interface	31
Figure 3-1 Telemeeting as a use case of multi-site NPNs.....	34
Figure 3-2 Multi-site NPN deployed as a PNI-NPN supported by the same PLMN	34
Figure 3-3 Multi-site NPN as a set of segregated SNPNs interconnected through an SD-WAN	35
Figure 3-4 Single site SNPN setup with 5G-CLARITY Framework components, enabling two services	36
Figure 3-5 Two NPN sites enabling PNI-NPN with 5G-CLARITY framework delivering 3 types of services	37
Figure 3-6 Configuration defined in 3GPP Rel-16 to access to PLMN services via SNPN	39
Figure 3-7 Procedure to perform a handover from an SNPN to a PLMN. The SNPN is seen as a non-3GPP access network	39
Figure 3-8 Roaming architecture considered for the 5G-CLARITY system	40
Figure 3-9 Roaming procedure call flow considered for the 5G-CLARITY system	41
Figure 3-10 Data aggregation in NPN-PLMN scenarios using the DSF	45
Figure 3-11 Sequence diagram about the data aggregation workflow between the management systems of the public and private domains.....	47
Figure 5-1 Joint user handover and VM migration problem to ensure service continuity in public/private 5G systems.	50
Figure 5-2 5G network with mobility support.....	53
Figure 5-3 Connectivity diagram (left) and experimental infrastructure used to host the 5G SA platform.....	55
Figure 5-4 Traffic generated at a UE (a) traversing the UPF (b) and terminated at the AS (c).	55
Figure 5-5 Signalling traffic generated by a) the AMF and b) the SMF.	56
Figure 5-6 CU measurements under different configurations (top-left): CPU utilization for SISO and 2x2 MIMO configuration, (top-right): Memory utilization, (middle-left): fronthaul traffic, (middle-right): CU processing distributed at different CPUs, (bottom left): disk read/write, (bottom -right): disk usage	56
Figure 5-7 Power consumption of the physical machines used to host the gNB (top, green) and the core (bottom, yellow)	57
Figure 5-8 a) Time series showing the traffic generated during VM migration from a source to a target VM. b) Correlation between background mobile network traffic per gNB and speed per UE, c) Impact of average UE throughput on 5GC computational resources [23]	57

Figure 5-9 Comparative analysis of different 5G network deployment strategies in terms of a) network, b) compute and c) total cost under and c) total costs under low mobility 60

Figure 5-10 Comparative analysis of different 5G network deployment strategies in terms of a) network, b) compute and c) total costs under high mobility 61

Figure 6-1 Components under evaluation of 5G-CLARITY system architecture for assessing multi-WAT capabilities 63

Figure 6-2 Illustration of distance from point D, marking the session request, in a LiFi or a Wi-Fi cell to point C, marking the UE exit on the cell 63

Figure 6-3 LoS path of the LiFi channel considered 65

Figure 6-4 Smart network user mobility algorithm 68

Figure 6-5 Industrial scenario layout with multi-WAT RAN 69

Figure 6-6 Average eMBB users’ throughput versus the 5G bandwidth allocated to eMBB users 72

Figure 6-7 URLLC slice packet loss ratio vs the traffic load 72

Figure 6-8 Average eMBB user achievable throughput per technology 73

Figure 6-9 Average eMBB users’ throughput versus the 5G bandwidth allocated to eMBB users for different separation values between LiFi APs 73

Figure 6-10 Average eMBB users’ throughput versus the separation values between LiFi APs 74

Figure 6-11 Network scenario and smart network user mobility architecture 75

Figure 6-12 SDN-enabled LiFi/Wi-Fi integrated network simulation environment 76

Figure 6-13 Experimental and simulation-based user handover performance from LiFi to LiFi and LiFi to Wi-Fi in SDN-enabled LiFi/Wi-Fi integrated network 77

Figure 6-14 Average Network throughput versus number of users 78

Figure 6-15 Average Network throughput versus user speed 78

Figure 6-16 CDF network average delay of URLLC users (service) 79

Figure 6-17 Extended position estimation model 79

Figure 6-18 DL/TDoA method 80

Figure 6-19 Range-based localization approach 81

Figure 6-20 TWR approach 82

Figure 6-21 Time diagram of the TWR process 82

Figure 6-22 LiFi positioning approach 83

Figure 6-23 Visible light positioning system 83

Figure 6-24 Visible light positioning geometrical model 84

Figure 7-1 Architectural components of 5G-CLARITY system architecture involved in joint synchronization and localization. 86

Figure 7-2 Positioning test scenario using multi-WATs 87

Figure 7-3 Deep neural network employed for NLoS-identification. It has $LH = 2$ hidden layers with nH neurons and two output neurons [41] 88

Figure 7-4 Example distribution of the ξ_i for a given timestamp measurement [41] 90

Figure 7-5 Performance comparison of the three joint synchronization and localization algorithms in terms of clock-offset estimation 91

Figure 7-6 Performance comparison of the three joint synchronization and localization algorithms in terms of position estimation. 92

Figure 7-7 Performance comparison of L-BRF and DePF when estimating the MUs’ clock offset [41] 93

Figure 7-8 Performance comparison of L-BRF and DePF when estimating the MUs’ position 93

Figure 7-9 Performance of joint sync&loc algorithm for different number of GDFs [41] 94

Figure 7-10 Position estimation performance of joint sync&loc algorithm with different timestamp accuracy [41] 94

Figure 8-1 6G Goals and target KPIs [47] 96

Figure 8-2 Technologies for 6G and relationship with 6G goals (requirements) and KPIs 98

Figure 8-3 Representative HEXA-X use cases, enabling services and mapping to KVis/KPIs [50] 101

Figure 8-4 In-scope HEXA-X use cases (indicated by red arrows) 102

Figure 8-5 In-scope Oulu 6G Flagship use cases (red circled) 102

Figure 8-6 6G Architecture Vision [47] 103

List of Tables

Table 2-1 5G-CLARITY Infrastructure Stratum Requirements	15
Table 2-2 Mapping of the D3.3 Implementations, the 5G-CLARITY Network Function and Application Stratum KPIs in D2.1, Requirements in D2.2, and KPIs in D2.1.....	17
Table 4-1 Extensions of 5G-CLARITY evaluation studies	48
Table 6-1 LiFi Channel Modeling Notation Summary	66
Table 6-2 Components included in the scenario modelling	70
Table 6-3 Configuration of the Main Simulation Parameters	70
Table 6-4 Configuration of LiFi APs in the Scenario.....	73
Table 6-5 Scenario setup components	74
Table 6-6 LiFi Attocellular and Wi-Fi Networks Simulation Parameters	76
Table 6-7 Services and Traffic Simulation Parameters	77
Table 6-8 Positioning Relevant Parameters for Different 5G-CLARITY Technologies.....	80
Table 7-1 Simulation parameters.	91
Table 7-2 Validation components.	91
Table 8-1 Drivers for 6G	95
Table 8-2 Key Technologies for 6G	98
Table 8-3 Hexa-X Families of Use Cases	100
Table 8-4 On the Relationship with In-Scope Technologies with Relevant 6G Use Cases.....	102
Table 8-5 5G-CLARITY Road Towards 6G: Architecture Evolution.....	104

List of Acronyms

AF	Application Function
AGV	Automated guided vehicle
AoA	Angle of Arrival
AP	Access Point
API	Application Programming Interface
ATSSS (AT3S)	Access Traffic Steering, Switching and Splitting
BRF	Bayesian Recursive Filtering
CAG	Closed Access Groups
CIR	Channel Impulse Response
CP	Control Plane
CPE	Customer Premises Equipment
CU	Central Unit
CUPS	Control User Plane Separation
DePF	Deep neural network assisted particle filter-based
DNN	Data Network Name
DRB	Data Radio Bearer
DU	Distributed Unit
GDF	Gaussian Density Function
GPU	Graphical Processing Unit
HPLMN	Home PLMN
IMSO	International Mobile Subscriber Identity
IOT	Internet of Things
ISAC	Integrated Sensing and Communications
KPI	Key Performance Indicator
KVI	Key Value Indicator
L3VPN	Layer 3 virtual private network
LADN	Local Area Data Network
LoS	Line of Sight
MEC	Mobile Edge Cloud
MF	Management Function
MPC	Model Predictive Control
MPTCP	Multi-Path TCP
NBI	North Bound Interfaces
NF	Network Function
NLLSQ	Nonlinear least squares
NPNS	Non-Public Networks
NR	New Radio
OAI	Open Air Interface
OCC	Optical Camera Communications
ONMP	One-Network-Multiple-AP
PDU	Protocol Data Unit
PGM	Particle Gaussian Mixture
PLMN	Public Land Mobile Network
PSA	PDU Session Anchor
RAN	Radio Access Network
RIC	RAN Intelligent Controller
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RT	Real Time
RU	Remote Unit
SDN	Software Defined Networks
SD-WAN	software-Defined Wide Area Network

SEPP	Security Edge Proxy Protection
SMO	Service and Management Orchestration
SNPN	Stand-alone NPNs
SNR	signal-to-noise ratio
SNUM	smart network user mobility
SSC	Session and Service Continuity
TA	Tracking Area
TDM	Time Division Multiplexing
TOA	Time of Arrival
UE	User Equipment
UL	Uplink
UL-CL	UL Classifier
UMP	User Mobility Profile
UP	User Plane
UPF	User Plane Function
URSP	UE Route Selection Policy
VLP	Visible Light Positioning
VM	Virtual Machine
WAT	Wireless Access Technology
XR	eXtended Reality

Executive Summary

This deliverable is a report on the final **5G-CLARITY** system architecture. It builds upon the solution design and validation work done over the initial **5G-CLARITY** system, refining it according to the lessons learnt from WP3 (use and control plane), WP4 (management plane) and WP5 (in-project pilots). In addition, this deliverable reports on **5G-CLARITY** pathways towards 6G, outlining the actionable items on **5G-CLARITY** capabilities to make system architecture evolve according to 6G expectations.

The deliverable also leverages on the topics and scenarios presented in D2.3, refining and extending them to address the technical challenges identified. This includes the need to:

- effectively coordinate different types of controllers and orchestration platforms to minimize service disruption time under mobility considerations.
- tighter integrate the multi-WAT segment of **5G-CLARITY** through the development of more accurate channel model techniques.
- further improve the performance of the joint synchronization and positioning module under adverse operational conditions such as non-line-of sight

Through extensive theoretical and experimental studies, it is proven that the **5G-CLARITY** architecture can effectively address these challenges as it provides the necessary tools to:

- optimally allocate compute and network resources under private/public mobile network deployments. This is achieved utilizing the functionalities offered through the intelligence stratum of the **5G-CLARITY** architecture to predict a variety of parameters (such as mobile traffic, UE trajectory patterns, evolution of system resources, etc.) and prescribe optimal service provisioning plans. To further improve the operation of the system and ensure service continuity under mobility, additional concepts, such as predictive migration of virtualized network functions, are introduced.
- manage traffic and mobility in the multi-WAT segment of the **5G-CLARITY** architecture in a more effective way. This is achieved through the introduction of more accurate channel models and smart mobility management techniques. These techniques enable allocation of dynamic routing paths for URLLC and eMBB services with e2e QoS guarantees
- perform more accurate synchronization and localization using multi wireless access technologies. This is achieved by replacing the previously used linear Bayesian recursive filter by particle filter to enhance the performance operating both in Line of Sight and Non Line of Sight setting.

1 Introduction

The high-level description of the main components of the platform has been described in 5G-CLARITY D2.2 [1]. Over this first system architecture version, 5G-CLARITY D2.3 [2] reported an initial evaluation of the key features of the 5G-CLARITY system in order to identify the main merits and limitations of the proposed architecture. However, not all the requirements reported in 5G-CLARITY D2.1 [3] were captured in this evaluation; in addition, there have been lessons learnt as the project made progress. Therefore, a final version of the system architecture together with feature evaluation is needed. This is the aim of the present deliverable.

1.1 Scope and objectives of this document

5G-CLARITY D2.4 describes the final 5G-CLARITY system architecture with respect to the architectural description provided in 5G-CLARITY D2.2 [1], taking into consideration knowledge and experience gathered acquired in 5G-CLARITY WP3, i.e., the implementation of ‘Network and Application Function stratum’, in 5G-CLARITY WP4, i.e., the works on ‘Management and Orchestration stratum and Intelligence stratum’, and in 5G-CLARITY WP5, i.e., setting up and execution of project pilots. The proposed architectural extensions cover both data and control planes associated with multiple wireless access technologies (WATs). 5G-CLARITY D2.3 investigated and reported performance enhancements, mobility capabilities, i.e., roaming from public to private network and reverse, traffic offloading from 5G NR to Wi-Fi/LiFi, interactions among different network providers (public or private), data management under scenarios involving multiple operators and exposure of infrastructure capabilities that can be achieved through the deployment of the 5G-CLARITY architectural structure and technology options. 5G-CLARITY D2.4 also provides a detailed architectural evaluation extending the previous work reported in 5G-CLARITY D2.3 and exploiting the functional elements and the end-to-end modelling tools reported in 5G-CLARITY D2.3. These elements and system modelling tools have been further extended and refined to better address the revised 5G-CLARITY architecture functionality and characteristics and enable more realistic and accurate performance evaluation of the 5G-CLARITY platform. Adopting the proposed evaluation methodology and modelling tools the 5G-CLARITY architecture was evaluated for several relevant use cases. The evaluation results produced for these use cases indicate that the 5G-CLARITY solution can offer clear benefits with respect to the relevant state-of-the-art (SotA), while relevant trade-offs are being discussed.

The objectives of this document are listed below:

- **D2.4-OBJ-1:** Specification of final version of 5G-CLARITY system architecture, and assessment of relevant functional aspects. This final release incorporates refinements over the original system architecture, based on the feedback of WP3, WP4, WP5 activities.
- **D2.4-OBJ-2:** Consolidating 5G-CLARITY prominent position for the support of advanced non-public network (NPN) scenarios, with analysis of architecture solutions allowing for this support. These scenarios capture use cases that were intentionally ruled out in 5G-CLARITY D2.2 due to their complexity and are now analysed in more details.
- **D2.4-OBJ-3:** Development of suitable models for the identified architectural functional elements relying on a mixture of theoretical and simulation tools as well as experimental profiling of specific architectural elements as appropriate.
- **D2.4-OBJ-4:** Specification of evaluation scenarios that will be considered for the assessment and associated input/parameters definition.
- **D2.4-OBJ-5:** System level evaluation execution integrating the developed functional element models to validate and benchmark the performance of the overall 5G-CLARITY architecture with respect to alternative state-of-the-art approaches.

- **D2.4-OBJ-6:** Outlining evolution of **5G-CLARITY** system to support in-scope 6G functional requirements, KPIs and use cases, according to the vision captured in the on-going flagship 6G projects. The pathways towards 6G shall guarantee evolution of **5G-CLARITY** assets and features are done with minimal integration efforts while ensuring backwards compatibility.

1.2 Document structure

The rest of the document is organized as follows:

- Section 2 covers **D2.4-OBJ-1**, by providing the main refinements done for final **5G-CLARITY** system architecture, in relation to the first version captured in **5G-CLARITY** D2.2 [1].
- Section 3 covers **D2.4-OBJ-2**, elaborating on advanced features and application scenarios on NPN.
- Section 4 provides an overview of the improvements/refinements for each scenario considered in the project over D2.3
- Sections 5, 6 and 7 describe the refined/extended scenarios under evaluation and present the results. These sections provide a description of the revised Use Cases with emphasis on the additional constraints introduced in the model to capture new features, refinements in the modelling tools and assessment of the performance of the overall **5G-CLARITY** solution, indicating clear benefits with respect to the relevant state-of-the-art as well as associated trade-offs.
- Section 8 covers **D2.4-OBJ-6**, outlining the pathways for **5G-CLARITY** evolution to 6G.
- Finally, Section 9 summarizes and concludes this document.

2 Revised Architecture

The final 5G-CLARITY system architecture is pictured in Figure 2-1. This section reports on this final system release, highlight the refinements and/or modifications made to the initial versions reported in 5G-CLARITY D2.2 [1].

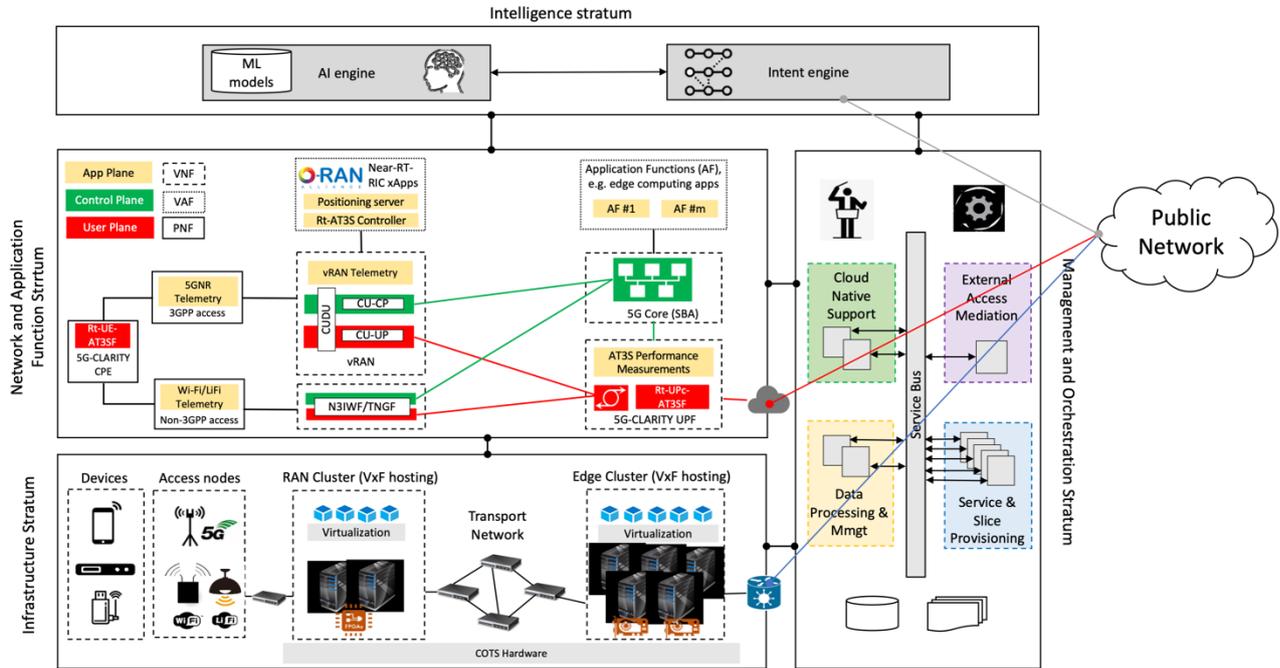


Figure 2-1 5G-CLARITY final system architecture

2.1 Infrastructure stratum

The 5G-CLARITY infrastructure stratum requirements identified in 5G-CLARITY D2.2 [1] are listed in Table 2-1 5G-CLARITY Infrastructure Stratum Requirements. In this section we introduce the final architecture of 5G-CLARITY infrastructure stratum designed to fulfil the given requirements. In this section we introduce the main components in process to be deployed.

Table 2-1 5G-CLARITY Infrastructure Stratum Requirements

Requirement ID	Requirement Description
CLARITY-INF-R1	5G-CLARITY system managed resources are restricted to on-premises resources, 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 (5G NR and non-3GPP Wi-Fi, 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 fronthaul, midhaul and backhaul 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.

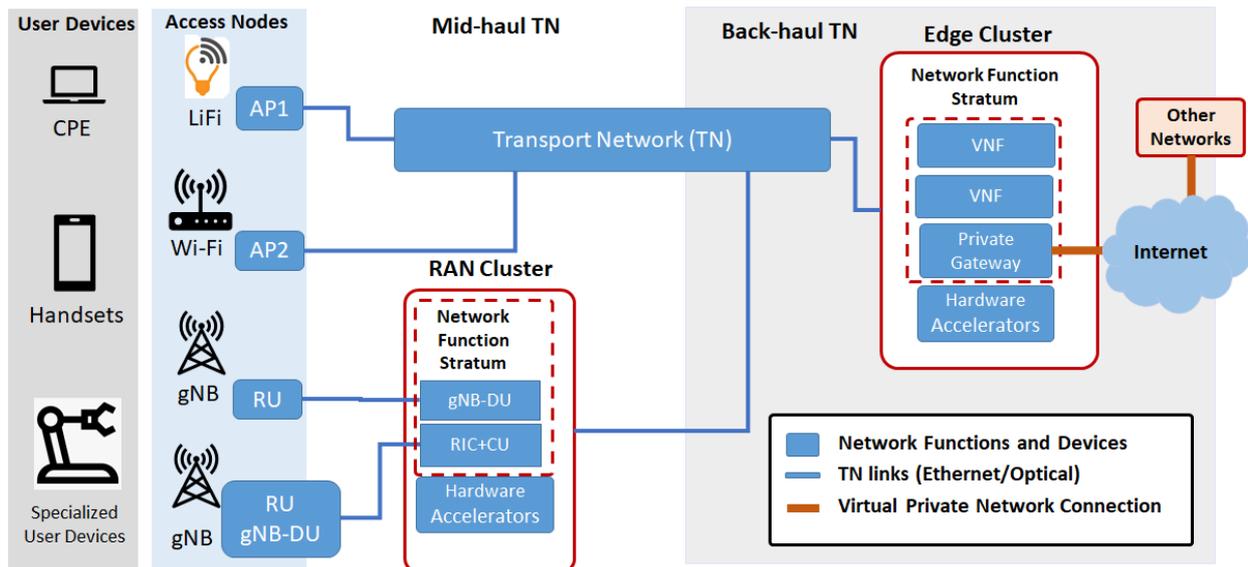


Figure 2-2 **5G-CLARITY** infrastructure stratum architecture (**5G-CLARITY D2.2**) [1]

The final architecture **5G-CLARITY** infrastructure stratum remains the ones introduced on **5G-CLARITY D2.2** [1]. This section summarizes the final architecture for the infrastructure stratum, presented in Figure 2-2, formed by:

- **User devices or equipment (UE)** enabling 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 standards.
- **Access nodes**, forming the **5G-CLARITY** multi-WAT connecting APs for LiFi and Wi-Fi technologies and a RU and DU for 5G technologies with **5G-CLARITY** RAN cluster.
- **Compute nodes**, hosting virtualized functions and applications from the different **5G-CLARITY** system strata (i.e., intelligence stratum, management and orchestration stratum and network and application function stratum), and providing an NFV infrastructure environment for their execution. The RAN cluster host 5G access nodes (i.e., gNB-DU, RIC, and RU) and the edge cluster to host core functions of others **5G-CLARITY** strata (e.g., 5GC, NFVO, AI engine).
- **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 backhaul.
- **An on-site private gateway**, to provide connectivity and reachability with external resources, e.g., PLMN resources or hyperscale's resources.

Figure 2-2 introduces the final architecture design being deployed for the UCs without important changes since **5G-CLARITY D2.2** [1] and D2.3 [2].

2.2 Network function and application stratum

The **5G-CLARITY** network function and application stratum is envisaged to be executed on top of the **5G-CLARITY** network functions virtualization infrastructure (NFVI) with its virtual network and vertical functions (VxFs). These VxFs will flexibly utilized to harvest computational resources for hosting, deploying, instantiating, and supporting one or more **5G-CLARITY** slices. During the lifecycle of VxF's, they will serve as a platform to realize a flexible, transparent, and replicable testing of an NFV ecosystem. The design of the **5G-CLARITY** architecture is based on the control and user plane separation (CUPS), cloud native VxFs, RAN functional splitting, integration of O-RAN alliance framework, multi-WAT protocol stack principles.

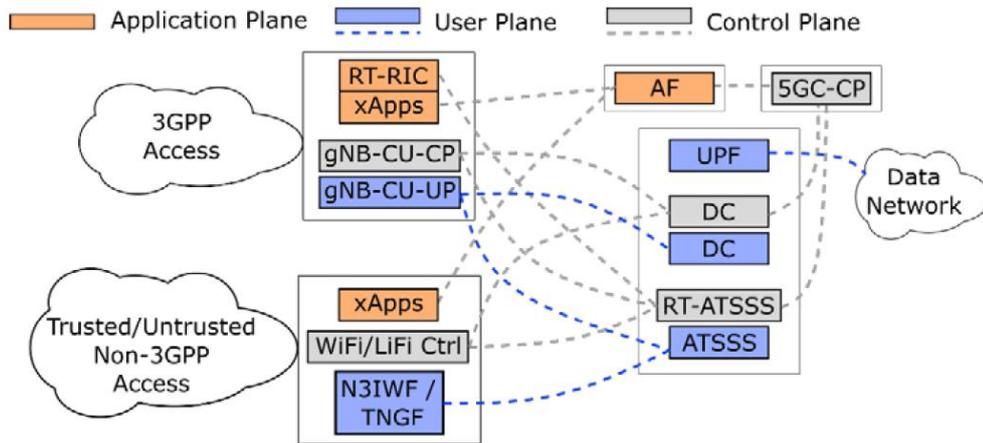


Figure 2-3 5G-CLARITY network and application function stratum

In this deliverable, D2.4, we will explain how the initial system design and task-specific system requirements are shaped the actual implementations, which are given in detail within deliverables 5G-CLARITY D3.1, D3.2 and D3.3 in WP3.

The original architectural design of the 5G-CLARITY Network Function and Application stratum were presented in 5G-CLARITY D2.2 [1], where the overall design is depicted in Figure 2-3. Accordingly, the 3GPP network functions, new radio access network (gNB-CU-UP), core network (5GC UPF), non-3GPP network functions, multi connectivity are the most important elements to implement user plane protocols and functionalities. Similarly, access network (gNB-CU-CP), core network (5GC CP), application functions and services, non-3GPP network functions and multi-connectivity functionalities in the control plane are also considered in the initial design. Lastly, the application plane functions such as Near-RT RAN intelligence control (RIC), 5G-CLARITY xApps that are designed to run on the Near-RT RIC and the localization server implementation, which exploits the 5G-CLARITY WATs are also investigated in the designed architecture.

Note that the 5G-CLARITY D3.1 [4] and D3.2 [5], have addressed most of the functional requirements mentioned in 5G-CLARITY D2.2, Table 6-1 [1]. The 5G-CLARITY D2.1 and D2.2 KPIs / technical requirements listed in Table 2-2 are planned to be addressed in the upcoming 5G-CLARITY D3.3 to comprehensively cover the initial system design and task-specific system requirements

Table 2-2 Mapping of the D3.3 Implementations, the 5G-CLARITY Network Function and Application Stratum KPIs in D2.1, Requirements in D2.2, and KPIs in D2.1

Requirement/KPI ID	Requirement Description	Addressing Component in D3.3
D22-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.	<ul style="list-style-type: none"> MPTCP testbed
D22-CLARITY-NFAS-R18	The 5G-CLARITY network function and application stratum shall allow controlling physical resources of 5G NR gNBs, Wi-Fi and LiFi APs.	<ul style="list-style-type: none"> Wi-Fi airtime-based scheduler LiFi airtime scheduling LiFi spectrum resource scheduling
D22-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.	<ul style="list-style-type: none"> dRAX (near/non-real-time RIC); Positioning server
D22-CLARITY-	The 5G-CLARITY network function and application stratum shall provide	<ul style="list-style-type: none"> dRAX (near/non-

NFAS-R20	necessary telemetry data to the hosted xApps.	real-time RIC)
D21-5GC.KPI-2	Reducing latency in the air interface < 1 ms for uplink and downlink through parallel access across various technologies	• MPTCP testbed
D21-5GC.KPI-3	Providing reliability of at least six 9s through smart interface selection.	• MPTCP testbed
D21-5GC.KPI-4	Supporting vertical handover between wireless technologies with handover times < 5 ms.	• MPTCP testbed
D21-5GC.KPI-5	Demonstrate aggregate system area capacity in relevant indoor scenarios > 500 Mbps/m ² through smart RRM algorithms and SDN control frameworks that fully exploit the capacity of the combined 5G/Wi-Fi/LiFi access	• Real-time traffic simulation platform
D21-5GC.KPI-8	Positioning to a peak accuracy < 1 cm, and availability of < 1 meter accuracy 99% of the time.	• Positioning server

In terms of the UP architecture and validation, an initial design for the 5G-CLARITY architecture was proposed in 5G-CLARITY D3.1 to investigate 5G NR, Wi-Fi and LiFi access networks with common core network, which is depicted in Figure 2-4. It is important to note that the proposed architecture will be supporting enhanced AT3S (eAT3S) functionalities.

In 5G-CLARITY D3.2, the initially designed architecture has been implemented by a virtual testbed environment, which consists of three virtual machines (VMs) and supports eAT3S functionalities by utilizing multipath TCP (MPTCP) capability. Accordingly, the first VM in Figure 2-5, namely *mptcpUe*, represents 5G-CLARITY customer-premises equipment (CPE) with Wi-Fi and LiFi interfaces. The VM2, which is named as *free5gc*, is the implementation of a standalone (SA) 5G Core (5GC). Lastly, the VM3 with the name *mptcpProxy* is utilized to terminate the OpenVPN over MPTCP connections coming from the 5G-CLARITY CPE.

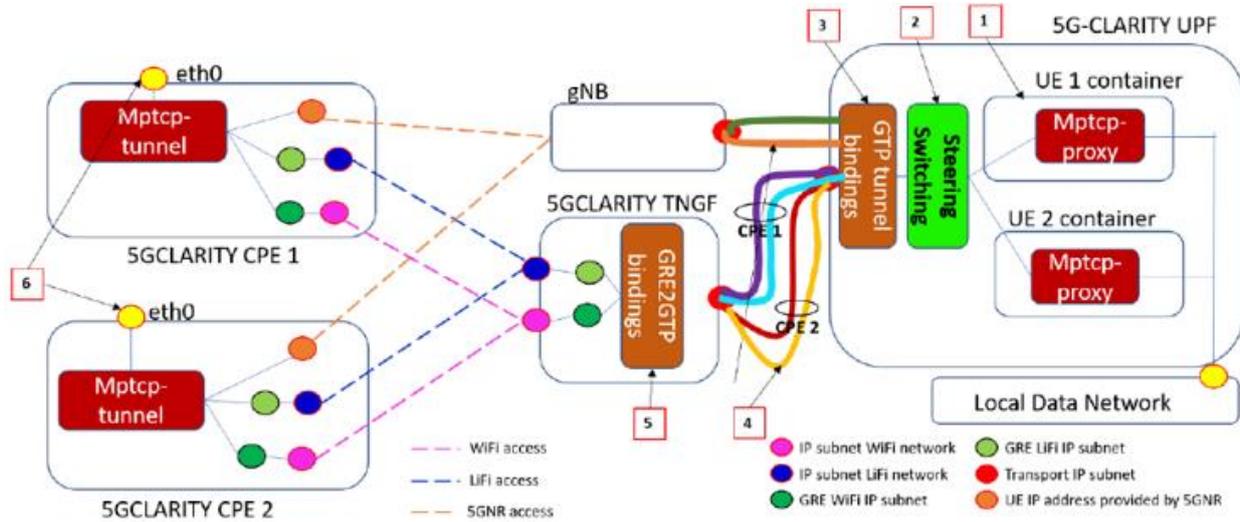


Figure 2-4 Initial UP architecture (5G-CLARITY D3.1) [4]

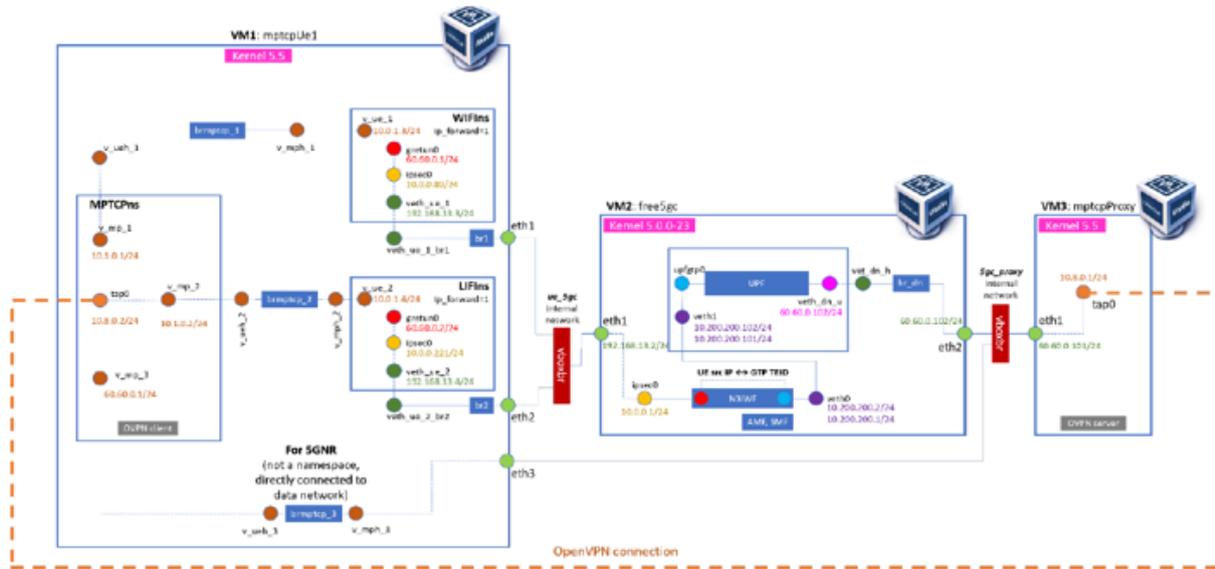


Figure 2-5 5G-CLARITY virtual testbed architecture (5G-CLARITY D3.2) [5]

The proposed MPTCP testbed addresses the following requirements raised in 5G-CLARITY D2.2:

- Realization of a 5G SA architecture with UPF deployed in the edge computer cluster.
- Realization of the integration of non-3GPP systems, i.e., Wi-Fi and LiFi, through non-3GPP interworking function (N3IWF) function.
- The 5GC UPF functions such as smart traffic management/radio resource management (RRM) with devices that support multi-connectivity and control of the traffic flow in proposed eAT3S via MPTCP
- Bridging a transparent L2 traffic between the 5G-CLARITY CPE and the MPTCP proxy function sitting behind the core network, which corresponds to a control plane (CP) application function (AF).
- Traffic flow control/management via various MPTCP schedulers.

Note that the dual connectivity (DC), which was briefly mentioned in both 5G-CLARITY D3.1 and D2.2, is not planned to be implemented within the virtual testbed based 5G-CLARITY architecture evaluation, since there is no commitment to showcase the DC and the related functionalities in WP5 UC implementations.

In terms of the 3GPP network functions (NFs) in both the user and control planes, the integration of virtualized DUs and CUs are realized by the O-RAN reference architecture. The 5G NR CU/DU/RU integration is presented by three phases in 5G-CLARITY D3.2. Accordingly, three phases for the CU/DU/RU integration and their details could be summarized as follows:

- **Phase 1:** as shown in Figure 2-6, F1 and NG are integrated with virtual UE. This initial phase was completed to integrate the developed CU components together with the third-party DU over F1 using the virtual UE.
- **Phase 2:** the NR-UU testing with the Research SDR RU and commercial UE integration and testing was presented. This integration phase was based on the setup of Phase 1, where the DU component was integrated and connected with National Instruments B20 SDR RU. Both Phase 1 and 2 are depicted in Figure 2-7.
- **Phase 3,** as presented in Figure 2-8, NR-UU testing with commercial third-party RU and commercial UE will be tested and will be reported in 5G-CLARITY D3.3. In other words, the NI B210 will be replaced with a commercial RU to align with 5G-CLARITY pilots in Phase 3.

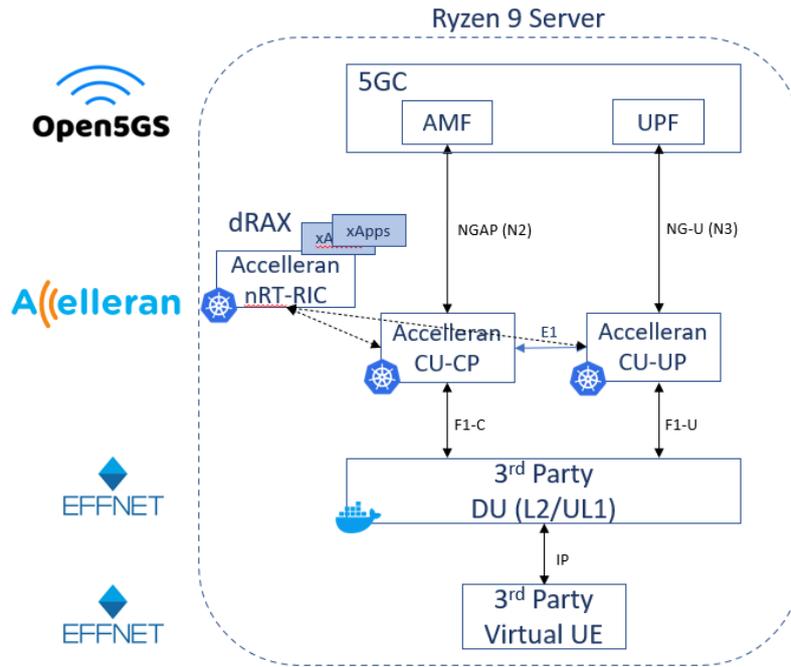


Figure 2-6 5G-CLARITY CU/DU/RU integration architecture, Phase 1: F1 and NG integration (5G-CLARITY D3.2) [5]

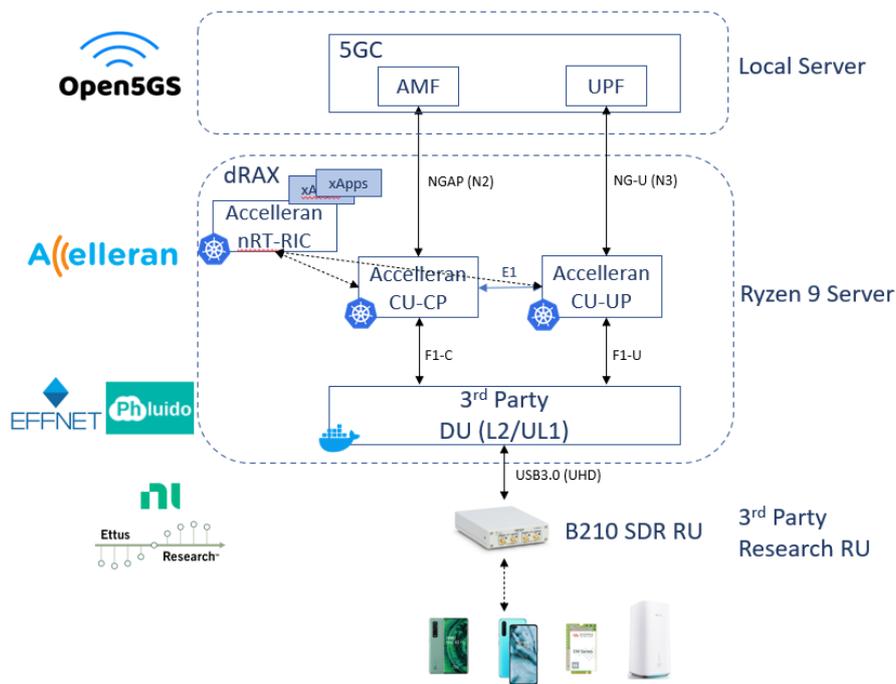


Figure 2-7 5G-CLARITY CU/DU/RU integration architecture, Phase 2: NR-UU integration and testing with Research RU and commercial UE (5G-CLARITY D3.3)

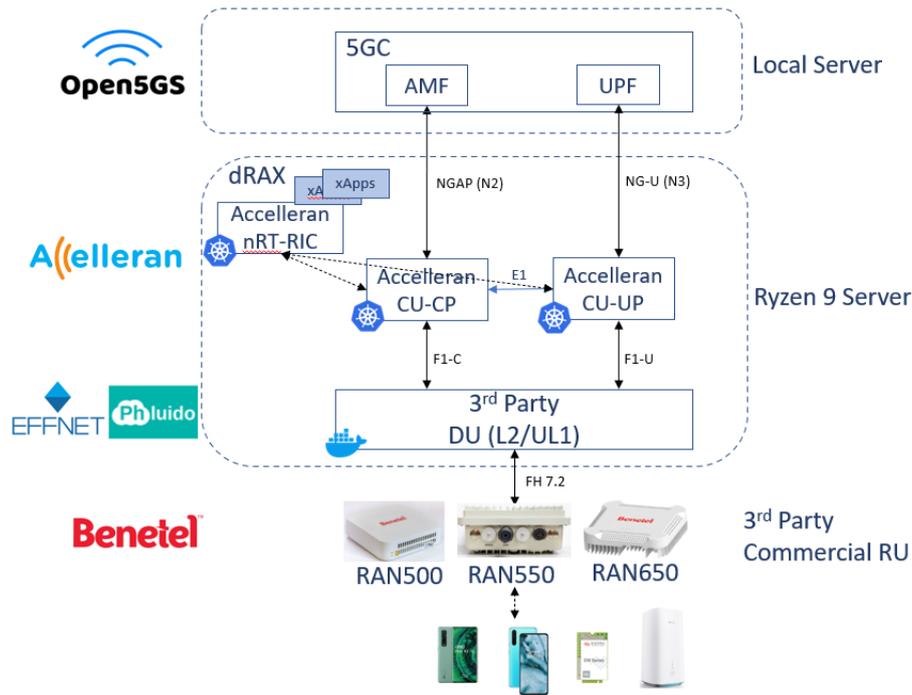


Figure 2-8 5G-CLARITY CU/DU/RU integration architecture, Phase 3: commercial third-party RU and commercial UE integration

Since the non-3GPP consists of Wi-Fi and LiFi technologies in 5G-CLARITY, the integration/aggregation of Wi-Fi and LiFi is investigated as a first step for “Non-3GPP Integration” concept mentioned in 5G-CLARITY D2.2. Accordingly, the designed testbed, which is depicted in Figure 2-9, is utilized to integrate L2 network with Wi-Fi and LiFi WATs via software defined networking (SDN).

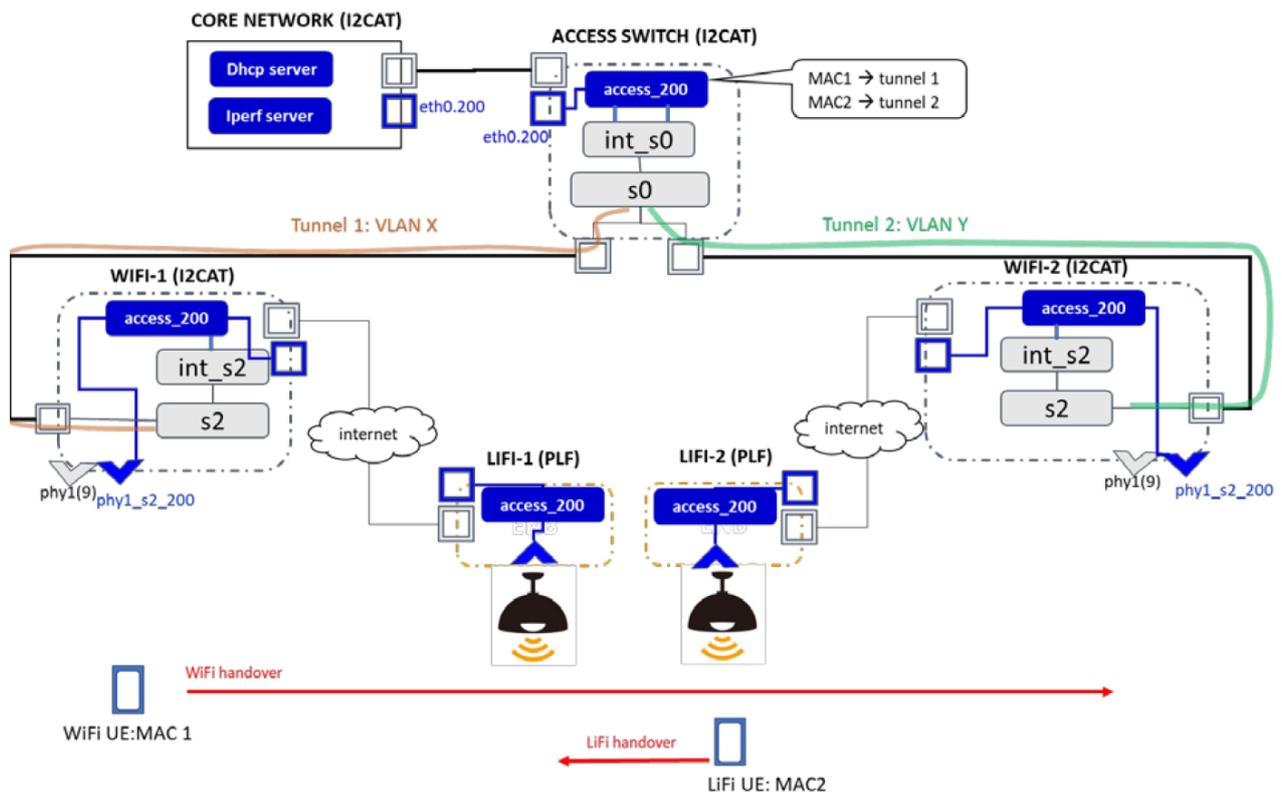


Figure 2-9 Testbed used for Wi-Fi and LiFi integrated L2 network (5G-CLARITY D3.2) [5]

To integrate Wi-Fi and LiFi networks to 5GC via N3IWF/trusted network gateway function (TNGF) functionalities, the following test has been conducted in the proposed testbed:

- Test 1: Session continuity across Wi-Fi technology, where the UE executes a handover between the APs Wi-Fi-1 and Wi-Fi-2.
- Test 2: Session continuity across LiFi technology, where the UE executes a handover between the APs LiFi-1 and LiFi-2.

Consequently, the developed prototype includes the design of an L2 access network and anchor which integrates non-3GPP technologies along with the mobility support. Moreover, the realization of such network by using SDN controller and its functionalities is also important since this was another architectural element mentioned in 5G-CLARITY D2.2.

The proposed 5G-CLARITY multi-connectivity framework implements the eAT3S in near RT-RIC UP and CP functionalities, which are provided in 5G-CLARITY D2.2, in a hybrid manner. The proposed eAT3S solution, controls the data flow to steer, switch and split the traffic. Thus, both 3GPP and non-3GPP user and control plane resources could be utilized for enhanced performance. It is important to note that the steering, switching, and splitting functionalities are realized by a high-layer MPTCP protocol in our developed testbed in 5G-CLARITY D3.1 and D3.2, where the low layer functionalities, eAT3S-LL, will be captured by 5G-CLARITY D3.3. In our eAT3S implementations in both 5G-CLARITY D3.1 and D3.2, the eAT3S policies are incorporated as an xApp within the O-RAN RIC framework behind the UPF, which will use the multi-WAT telemetry in decision-making. As previously mentioned, the proposed structure also complies with the O-RAN reference architecture. 5G-CLARITY eAT3S xApp environment, which is detailed in 5G-CLARITY D3.2, is depicted in Figure 2-10. Accordingly, the core based eAT3S integration is validated and tested via both the MPTCP testbed and computer simulations.

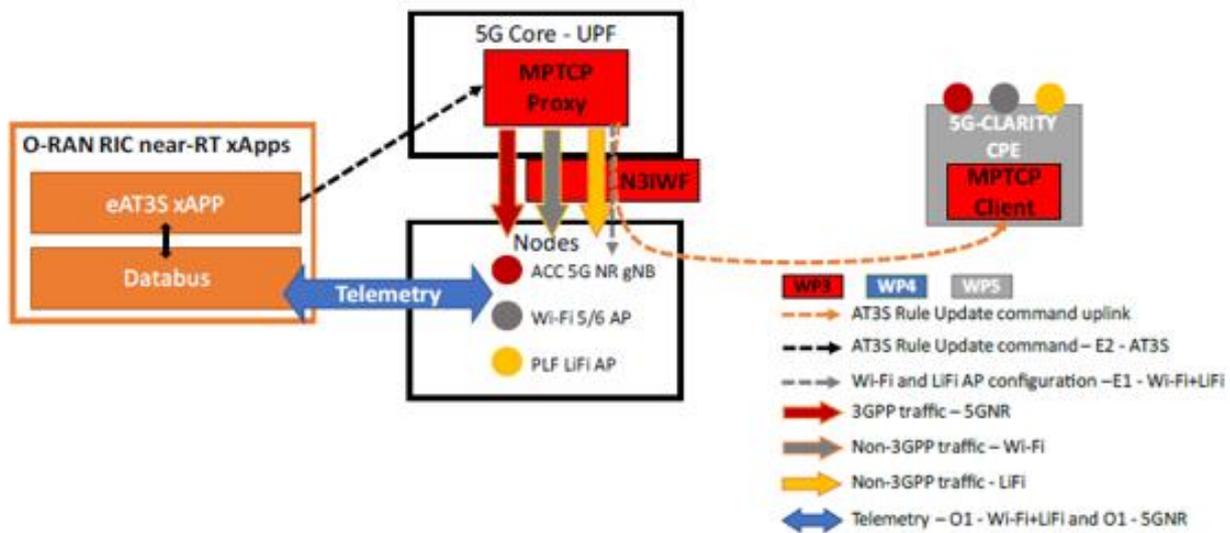


Figure 2-10 5G-CLARITY eAT3S xApp environment (5G-CLARITY D3.2) [5]

To implement smart decision making in traffic flow management in eAT3S via machine learning (ML) and artificial intelligence (AI), various implementations of eAT3S and utility-based resource scheduling techniques will also be presented in 5G-CLARITY D3.3 to optimize the system performance while minimizing the latency in decision-making.

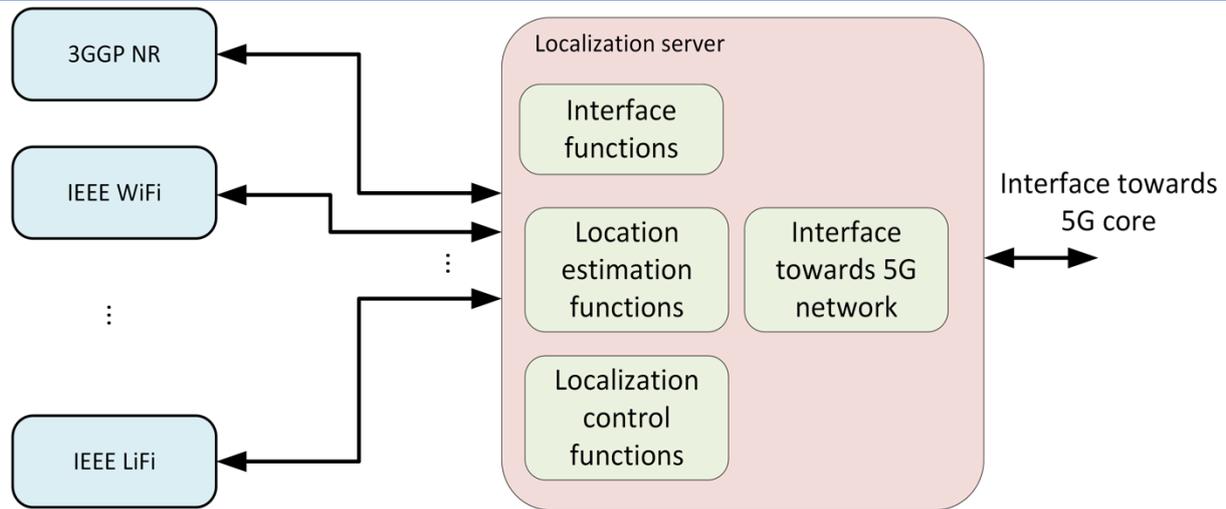


Figure 2-11 5G-CLARITY localization server architecture (5G-CLARITY D4.2) [6]

Lastly, the multi-connectivity framework, which employs three different WATs, features enhanced localization capabilities. Specifically, four proposed localization WATs; 60 GHz mmWave (3GPP NR), sub-6 GHz (Wi-Fi), LiFi and optical camera communications (OCC) are being investigated in 5G-CLARITY D3.2 and under investigation in 5G-CLARITY D3.3. To realize the multi-WAT positioning server concept in 5G-CLARITY D2.2 as depicted in Figure 2-11, the localization performances of each WAT are investigated in 5G-CLARITY D3.2 independently and they will be integrated as a whole system in 5G-CLARITY D3.3.

It is reported in 5G-CLARITY D3.2 that the maximum RMSE of 2.2 cm is achieved via 60 GHz WAT for 1.5 m AP separation. Similarly, 90% of the measurement errors are less than or equal to 15 cm and almost all measurement errors are less than 20 cm for sub-6 GHz WAT. For LiFi, mobile terminal could be located with a coarse AP-level positioning accuracy of 1 m via RSSI, which will be enhanced even further to exploit the advantages of using optical atto-cellular structure. For OCC, 96% of the computed localization errors are reported to be less than or equal to 10 cm, when 6 lights are employed. Moreover, 90% of the localization errors are less than or equal to 22 cm, when 4 lights are utilized. The average orientation error of 2.87 degrees is also reported.

In a nutshell, there won't be any diversions during the implementation of the initial architectural design, which is proposed in 5G-CLARITY D2.2. The work in WP3 will be focusing on addressing rest of the network function and application stratum requirements via computer simulations and testbed implementations.

2.3 Extensions to Management and Orchestration stratum

2.3.1 Impact of O-RAN management plane on 5G-CLARITY architecture

Since the original architectural design of the 5G-CLARITY Management and Orchestration (M&O) stratum and Intelligence stratum presented in 5G-CLARITY D2.2 [1], the O-RAN Alliance has progressed in the definition of the O-RAN Service and Management Orchestration (SMO). The latest O-RAN proposal is relevant to the 5G-CLARITY M&O and intelligence strata, and therefore in this section we describe how the 5G-CLARITY architecture can incorporate O-RAN compliant networks.

2.3.1.1 Review of O-RAN management architecture

Figure 2-12 depicts the architecture of the O-RAN SMO, as described in [7]. Within the SMO, the non-RT RAN RIC is the component in charge of controlling the Near-RTRICs, which in turn control the radio infrastructure nodes (gNBs). Key interfaces between the O-RAN SMO, non-RT RIC and the gNBs are the following:

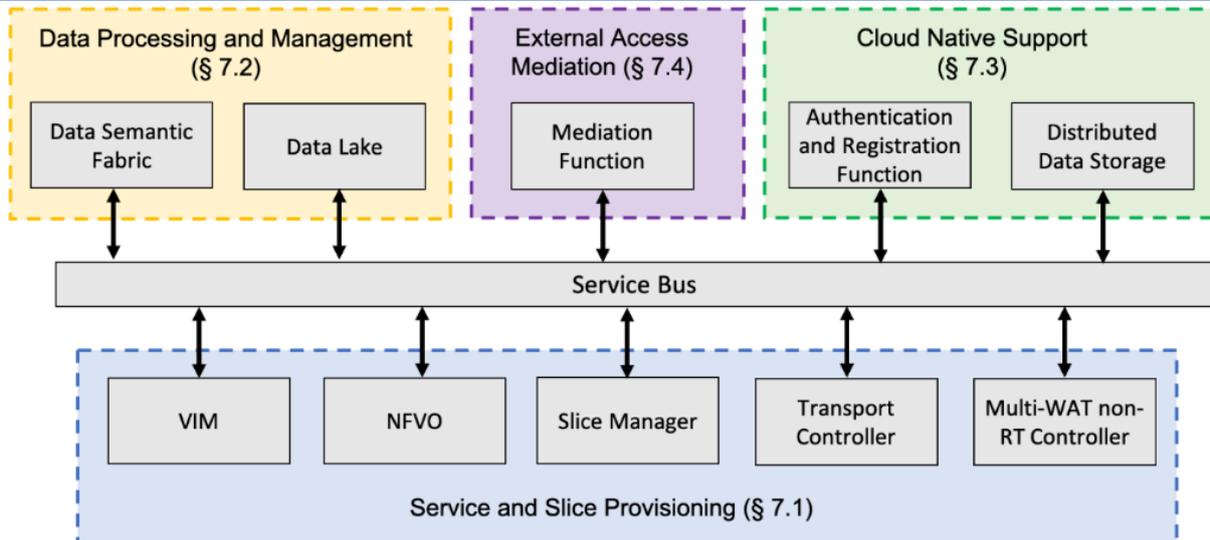


Figure 2-13 5G-CLARITY management and orchestration stratum from [1]

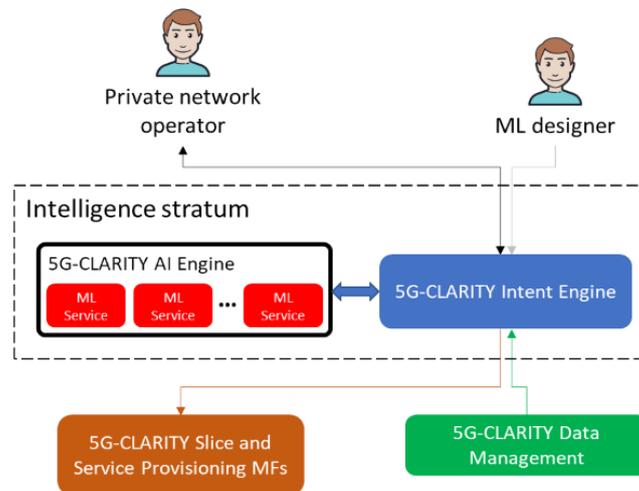


Figure 2-14 5G-CLARITY intelligence stratum from [1]

The 5G-CLARITY and O-RAN management architectures are related in the following way:

- multi-WAT Non-RT Controller.** This component can be seen as the O1 termination component from the O-RAN Architecture. The 5G-CLARITY multi-WAT non-RT RIC implements a NETCONF client to manage a Near-RT RIC (Accelleran’s dRAX) and the E2 components (Accelleran’s Centralized Unit function). In addition, extending the O-RAN architecture, the 5G-CLARITY multi-WAT non-RT RIC also uses NETCONF to manage IEEE 802.11 Wi-Fi and LiFi Access Points [6].
- Slice Manager.** This component exposes a north-bound interface (NBI) that can be used to configure radio infrastructure and compute infrastructure. In an O-RAN context the consumers of this north-bound interface would be the rApps. Therefore, one can see the Slice Manager’s NBI as partially overlapping with the O-RAN R1 interface, although the 5G-CLARITY Slice Manager scope is not limited to the RAN.
- 5G-CLARITY AI Engine.** This is an execution environment for ML models, which in the context of O-RAN can be understood as rApp services. Hence, in the O-RAN context the 5G-CLARITY AI Engine could be understood as the component of the SMO hosting the rApps. However, notice that the scope of 5G CLARITY is wider than the scope of O-RAN, and 5G-CLARITY ML models are not restricted

to the RAN. The interested reader is referred to 5G-CLARITY D4.2 [6] for a description of the 5G-CLARITY ML models.

- 5G-CLARITY Intent Engine.** In the context of O-RAN stack, the 5G-CLARITY Intent Engine can be seen as a mediator between the rApps (i.e., the ML models in the AI Engine), and the raw R1 interface offered by the Slice Manager, or other O-RAN implementations. So, in the O-RAN context we can see the Intent Engine as offering a “higher level R1 interface”, which is based on Intents, e.g. the natural language based interface described in 5G-CLARITY D4.2 [6].

There is however some O-RAN functionality that is not explicitly covered by the 5G-CLARITY architecture provided in [1], which is the A1 interface. There reason is that neither the 5G-CLARITY multi-WAT RT-RIC nor the 5G-CLARITY Slice Manager support A1 functionalities. However, being the Management and Orchestration stratum of 5G-CLARITY based on a service-based management architecture (SBMA), A1 capabilities would be easy to integrate in 5G-CLARITY in the following way:

- Including additional management functions (MFs) providing A1 services to the 5G-CLARITY Management and Orchestration stratum. For example, a “multi-WAT A1 policy” MF could be used to provision consistent policies across different technologies, e.g. (UE-ID X cannot connect to cell A / MAC Y cannot connect to AP B), hence mimic A1-P interface capabilities.
- Define 5G-CLARITY Intent Engine as service consumers of abovementioned MFs. The Intent Engine will consume A1-P interface capabilities from these MFs and transform supported policies into high-level intents. These intents will be available for consumption through Intent Engine’s NBI. Refer to [6] for a detailed description of how to integrate providers to the 5G-CLARITY Intent Engine.

Figure 2-15 illustrates how the O-RAN management plane could be integrated within the 5G-CLARITY management architecture.

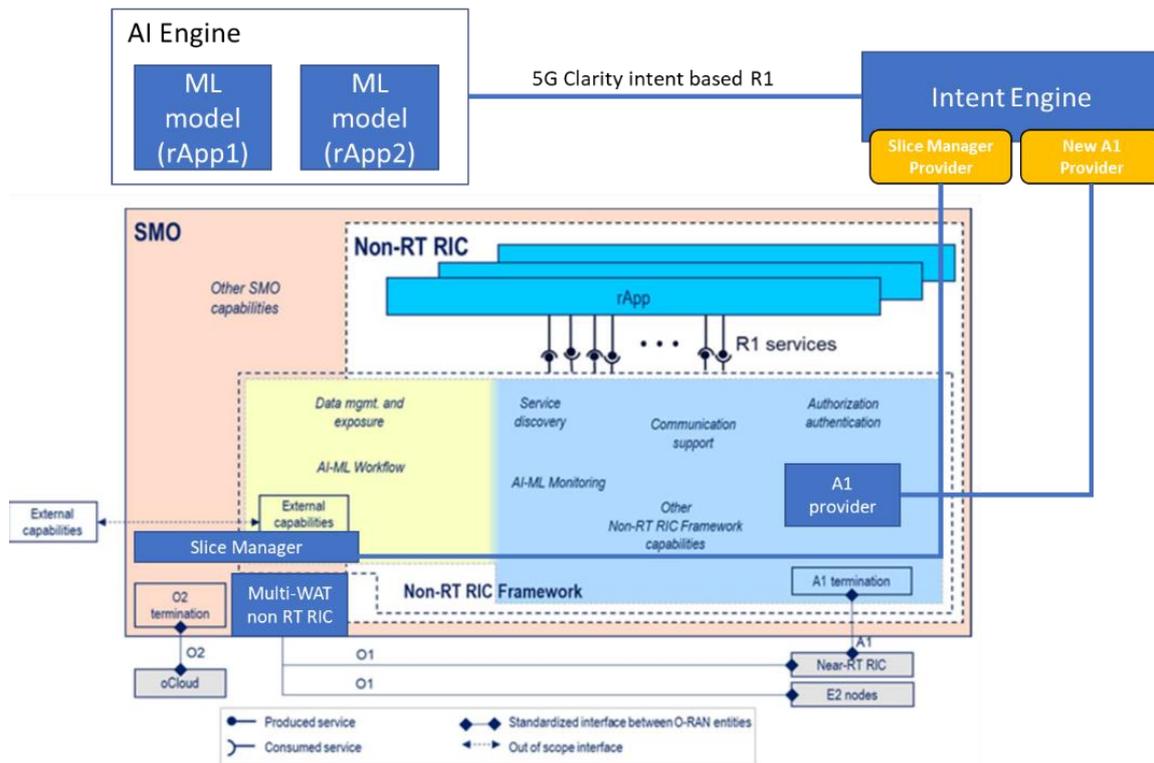


Figure 2-15 Integration between O-RAN and 5G-CLARITY management stratum

2.3.2 Cloud native subsystem: scalability analysis of M&O stratum

In this section we study the scalability of a cloud-based implementation of the 5G-CLARITY management and orchestration stratum. Our goal is to understand the expected OPEX costs associated to a cloud-based deployment and suggest enhancements that can reduce those costs.

2.3.2.1 Single-tenant vs multi-tenant analysis

In this section we evaluate two different strategies to deploy the slice and service provisioning subsystem of 5G-CLARITY using a cloud-native approach. The two considered approaches are:

- Single Tenant M&O: Consisting of deploying a separate instance of the service and slice provisioning subsystem for each tenant, i.e., for each 5G-CLARITY private network.
- Multi-Tenant M&O: Consisting of deploying a single instance of the service and slice provisioning subsystem shared across all tenants, i.e., each 5G-CLARITY private network.

The motivation of studying the two potential deployment mechanisms becomes relevant when a single 5G-CLARITY M&O stratum instance can be used to manage multiple private networks. For example, a mobile network operator (MNO) can use the 5G-CLARITY M&O stratum to manage multiple PNI-NPNs offered to their enterprise customers. Likewise, a large industry vertical (e.g., BOSCH) could use a single 5G-CLARITY M&O stratum instance to operate the private networks deployed across multiple factories. The two considered deployment mechanisms are described in Figure 2-16.

As previously mentioned, a natural option to deploy the 5G-CLARITY M&O stratum is using a cloud provider, thereby, in principle, alleviating the costs required to run the management plane. Obviously an on-premises deployment of the M&O stratum is also possible, but this section will focus on the former case. The three main available cloud providers, i.e., AWS, Azure, and Google Cloud, offer equivalent service offering options and pricing, which are relevant for our study. Thus, without lack of generality we will focus our study on AWS.

In 5G-CLARITY WP4 we have developed a prototype implementation of the slice and service provisioning subsystem that is used to manage a typical private network deployment, namely the private network deployment that will be used to support the BOSCH use case in WP5. Thus, we will use our current prototype as baseline for our analysis. Based on our initial prototypes, and based on the options made available by the AWS EC2 service [9], we can deploy the 5G-CLARITY service and slice provisioning subsystem using two instances of the following type:

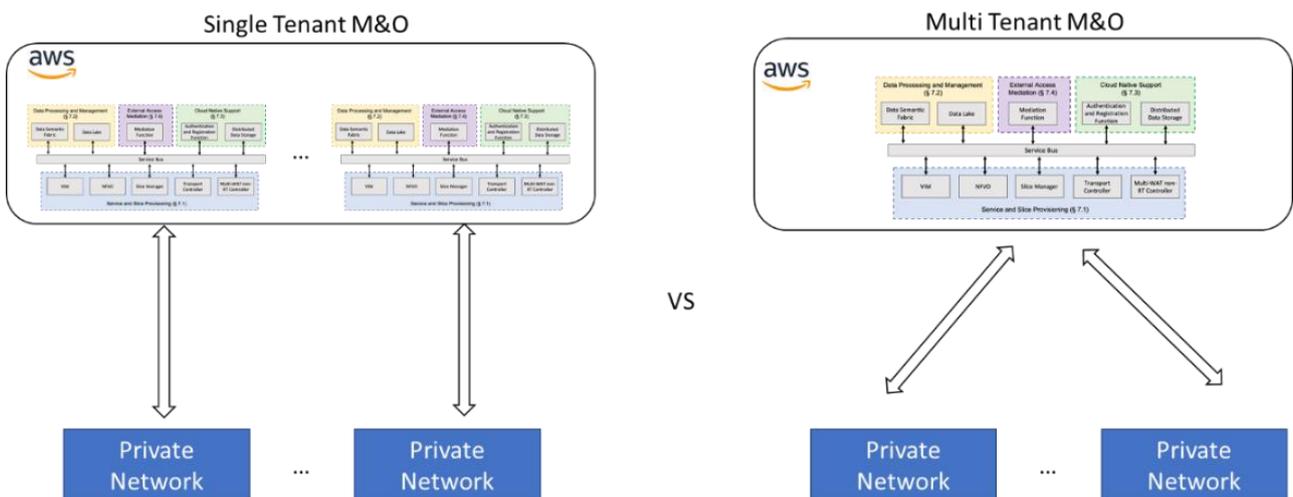


Figure 2-16 Single tenant VS multi-tenant deployment of M&O stratum

- Instance 1:
 - Service and Slice provisioning components: Open-Source MANO and Slice Manager
 - AWS EC2 type: m5.xlarge: {4 vCPU, 16 GB, BW: < 10Gbps}
 - Pricing: \$0.224/hour¹
- Instance 2:
 - Service and Slice provisioning components: multi-WAT non-rt controller
 - AWS EC2 type: m5.large: {2 vCPU, 8 GB, BW: < 10Gbps}
 - Pricing: \$0.112/hour

Thus, the OPEX per year associated to the **5G-CLARITY** service and slice provisioning subsystem to maintain one private network using the single tenant approach would be 2.943,36€/year. This means that a **5G-CLARITY** provider managing 100 private networks, which seems a reasonable number, through the **5G-CLARITY** service and slice provisioning subsystem would incur in more than 200K\$/year of operational expenses². This motivates the design of a multi-tenant service and slice provisioning subsystem, which we discuss next.

The following aspects need to be considered to add multi-tenancy capabilities to the service and slice provisioning subsystem:

- A user/role Management Function (MF) needs to be added to manage authorize different private network users operating on the same multi-tenant Slice Manager or multi-WAT non-rt controller. Standard authorization frameworks used by Web applications such as Keycloak [10] can be reused for this purpose.
- A Network Registration MF is required, whereby each private network registers its network components, i.e., edge cluster, RAN cluster and Wi-Fi/LiFi APs to a particular tenant ID.
- The NFVO needs to support multiple VIM instances, which would correspond to the Clarity edge and RAN clusters from each private network. This is already the case in OSM, which natively supports multiple VIM accounts.
- The Slice Manager and multi-WAT non-rt controller MFs need to be updated for multi-tenancy. Given the design of these management functions described in **5G-CLARITY** D2.2 [1], where the multi-WAT non-RT Controller configures each network device in a stateless manner using NETCONF and the Slice Manager interacts with the multi-WAT non-rt Controller through RESTful interface, adding multi-tenancy capability is straightforward. First, a notion of a tenant ID needs to be added to Slice Manager and multi-WAT non-rt Controller. The tenant ID will then be used to interact with the external database service where the configuration state of each network device is maintained, thus filtering the relevant topology in each endpoint according to the provided user ID.

2.3.2.2 Telemetry subsystem: Scalability of data storage requirements

In this section we present a case study of the scalability of the **5G-CLARITY** telemetry subsystem. As described in WP4, part of the data processing subsystem of **5G-CLARITY** can be implemented as a Data Lake implemented using several AWS services [6]. It is therefore important to understand the operational costs incurred by the **5G-CLARITY** data processing subsystem as the size of the network grows.

¹ Pricing available in AWS web page when writing this report

² This cost does not include other potentially expensive components such as the telemetry subsystem



Figure 2-17 Example of 5G-CLARITY RAN data exporter

For this purpose, we have developed two data exporters, using Prometheus [11], which export relevant metrics of the 5G-CLARITY RAN cluster³ and the 5G-CLARITY edge cluster.

The two developed exporters generate the following data:

- RAN cluster exporter: Per UE/cell counters, including throughput, packets, signal level, etc.
- Edge cluster exporter: CPU, memory usage per OpenStack project

An exemplary visualization of the data generated by the 5G-CLARITY RAN and edge exporters is depicted respectively in Figure 2-17 and Figure 2-18.

The data generated by these data exporters is periodically polled, e.g., every 10-15 seconds, and pushed towards the 5G-CLARITY Data Lake where is stored using the AWS S3 storage service, and thus made available to the Intent and AI engines. It is therefore important to understand the impact of the proposed data pipeline architecture in terms of the OPEX incurred by the AWS S3 storage service.

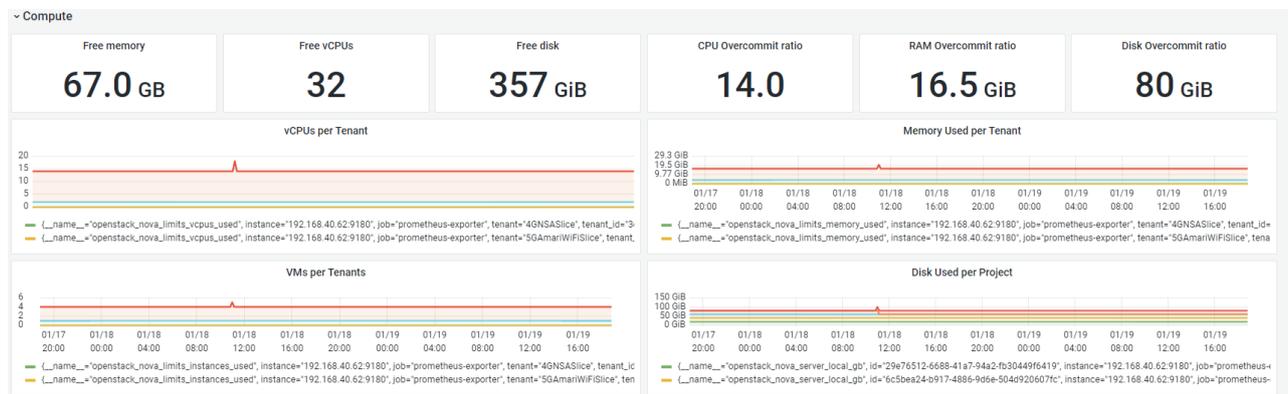


Figure 2-18 Example of 5G-CLARITY edge cluster exporter

³ The RAN technology used in our evaluation is Amarisoft [59], as the Accelleran dRAX technology used in WP5 was not yet available at the time of writing this deliverable. Data volumes generated by the two technologies are assumed to be similar, since the same notion of per UE, per cell, or per base station counters is used across different vendors.

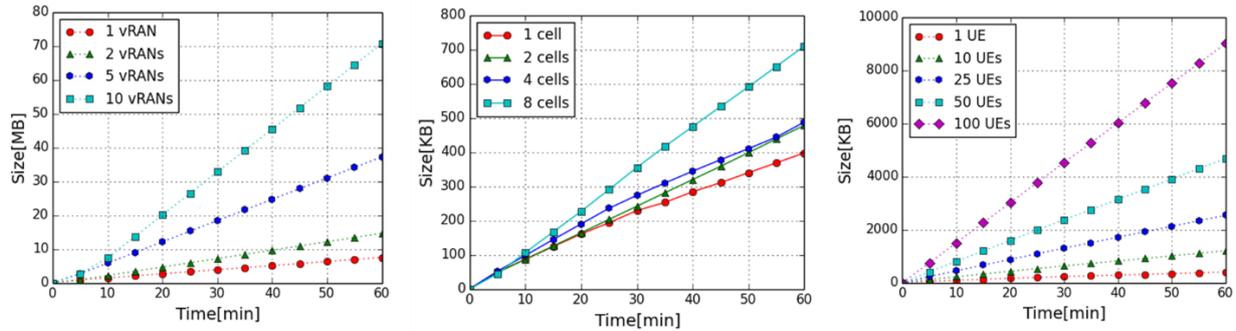


Figure 2-19 Data exporter scalability analysis: number of base stations (left), number of cells in 1 base station (middle), number of UEs connected to 1 cell (right)

To understand how [5G-CLARITY](#) data storage requirements scale, Figure 2-19 depicts the data requirements generated by the RAN data and vRAN server exporters over time, depending on the number of base stations (vRANs) in the private network (left graph), the number of cells connected to 1 base station (middle graph), with one UE in each cell, and the number of UEs connected to 1 cell (right graph).

Based on Figure 2-19 we can see that a typical [5G-CLARITY](#) private network featuring two base stations with 1 cell each, and an average of 10 connected UEs in each cell would generate data requirements around 16 MB/hour, which translates into a monthly storage requirements of ~11.5GB/month. The above results only consider the RAN telemetry, so to consider a full estimate of the overall storage data requirements for the private network let's conservatively estimate that the same amount of data is required to monitor the edge cluster, and the core network function inside the edge, resulting in an overall monthly requirement of ~34.5GB/month. Consequently, a [5G-CLARITY](#) provider managing 100 private networks would have storage requirements of 3.4 TB/month.

The [5G-CLARITY](#) data lake architecture proposed in [5G-CLARITY](#) D4.3 [12] uses AWS S3 buckets as the basic storage service to power the models in the AI engine. Looking at the pricing model of AWS S3 buckets at the time of writing this deliverable [13] we can see the following pricing model:

- First 50TB/month – \$0.023/GB
- Next 450TB/month – \$0.022/GB
- Over 500TB/month – \$0.021/GB

Thus, a [5G-CLARITY](#) provider serving 100 private networks would only incur a monthly cost of 79,35\$, which proves the scalability of the [5G-CLARITY](#) telemetry collection subsystem.

2.4 Intelligence stratum

This section details a new interface from the AI Engine to the Data Lake. This interface will allow the ML modeler to query the Data Lake directly from within the model. Large standardisation efforts have been seen in relation to intent and its role in autonomous networks. The ETSI Zero-touch network and Service Management (ZSM) group aims to accelerate the definition of required architecture and solutions for full end-to-end automation of network and service management [14]. In December 2020, a study began to investigate intent as a key enabler in autonomous network and service management within the ZSM framework. Work item ZSM-011 is currently in early draft but aims to provide guidelines on utilising intent-driven management interfaces between framework consumers and management domains. The Network Management Research Group (NMRG) has included intent-based networking as one of their three priorities for investigation for a five-year period between 2017 and 2022 [15].

2.4.1 Intent engine operation

As presented in 5G-CLARITY D2.2 [1] the Intent Engine is the only front-end interface at the 5G-CLARITY intelligence stratum. Any message from/to the AI Engine must pass through the Intent Engine. The Intent Engine is designed to match the request field of the intent message with the function descriptions of registered components. This is the same operation when receiving messages from outside the Intelligence Stratum (NBI services from 5G-CLARITY Management & Orchestration stratum) and inside (AI Engine). For a component to integrate its functionality, it must register with the Intent Engine and provide structured API information. The comparison of the intent request with detailed functionality influences the selection of actions triggerable by the Intent Engine. This is a bottom-up approach that allows for the dynamic incorporation of components through registration and execution of actions, both through a straightforward abstracted interface. The Data Lake does not have the same role as the provisioning subsystems, which benefit from this bottom-up approach.

2.4.2 Communication with Data Lake

The Data Lake provides data storage, analytics, and services to the platform. In a scenario where an ML model requires information to perform a task, the ML model would generate an intent for the information. This intent would then be sent to the Intent Engine where it could be translated into an action. In the translation phase metrics such as a similarity score are generated from the intent request and the function description. While this approach is suitable for requesting actions you would like to perform in an abstract manner, requesting specific information for the execution of a model does not benefit from the Intent Engine operation.

2.4.3 AI Engine to Data Lake Interface

A REST interface between the AI Engine and Data Lake is proposed to allow ML modelers to directly access their information during the execution of their model. ML modelers can generate a request within the model to query the Data Lake for the information. This will require ML modelers to know the location of data relevant to their model at the authoring stage. On successful receipt of the requested data the model can execute as normal.

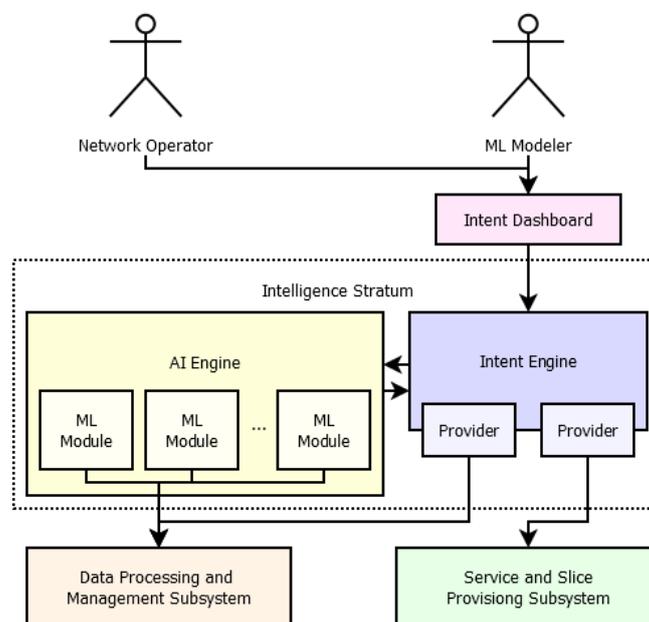


Figure 2-20 Intelligence stratum with new interface

3 Analysis of Advanced Features and Application Scenarios on NPNs

3.1 Architecture and operation of PNI-NPNs

3.1.1 PNI-NPN standardization

According to 3GPP standards, a PNI-NPN is an NPN deployed with the support of a PLMN. NPN devices must have a subscription to the PLMN to access the PNI-NPN. According to [16], a PNI-NPN may be provided by a PLMN by means of a dedicated Data Network Name (DNN) or by deploying network slices allocated for the NPN. Some details are as following:

- *Provision as a DNN:* In this case, the PNI-NPN is provided as a data network, which is used for hosting the NPN services and applications. The DNN identifies the data network, and whenever the subscriber executes the NPN application, the UE triggers the establishment of a Protocol Data Unit (PDU) session to the NPN DNN. As typically NPNs provide services within a limited coverage area, the 3GPP has standardized the concept of Local Area Data Network (LADN), which enables access to the DNN in each area (e.g., stadium or museum), but not outside. The LADN service area is defined as one or several Tracking Areas (TAs). A TA is a group of cells where a user can move around without updating the Access and Mobility Management Function (AMF). When the UE is inside the LADN service area, it can request a PDU session establishment for the LADN DNN, and the network will grant such PDU session. The PLMN-Op can use the UE Route Selection Policy (URSP) rules to control the PDU session request from the UE when this is inside (or outside) the LADN service area. These rules indicate to the device how application traffic shall be sent over the network, e.g., which data session, slice, SSC mode etc. shall be used for a certain application [17]. When an application starts in the device, the URSP can be used to determine if an existing session can be used or a new session is needed with an appropriate SSC mode, etc.
- *Provision as a network slice:* Network slicing is a technological solution that provides isolated logical networks with diverging performance requirements over a common network infrastructure. A 5G network slice is composed of the 3GPP 5GS network functions (e.g., gNBs, AMF, UPF, Session Management Function (SMF), etc.), it is identified by a Single Network Slice Selection Assistance Information (S-NSSAI), and it consumes a certain amount of radio resources in each cell. A PLMN-Op can use network slicing to provide public network services, or NPN services, i.e., a PNI-NPN. The PLMN-Op can deploy one or several dedicated network slices for the PNI-NPN, if NPN isolation or specific QoS treatment is desired. The customer can consume the received slice directly, or optionally extend it with additional features (e.g., device on-boarding, secondary authentication). Using network slicing for the PNI-NPN allows to control the access to the NPN because the subscriptions to the dedicated S-NSSAIs can be restricted to the NPN devices. In PNI-NPNs, the UE needs to be pre-configured with the S-NSSAI to access the slice. The PLMN-Op can also use the URSP rules for this purpose.

A relevant requirement of a NPN is that it can control the access of NPN devices to the network in areas in which they are not permitted to. However, as in the case of LADN service area, network slices are set on a per TA basis [16]. That is, neither LADN nor network slicing allow the possibility to prevent UEs from automatically selecting and accessing specific cells within a TA. Closed Access Groups (CAG) may optionally be used in NPNs for this purpose. A CAG defines a list of subscribers who are allowed to access a CAG cell associated with it. A CAG cell is a cell that only UEs supporting CAG can access. Hence, CAG can be used in PNI-NPNs to prevent unauthorized UEs to access specific CAG cells inside a private venue (e.g., stadium or museum). Please note that CAGs are independent from any network slice.

3.1.2 PNI-NPN deployment considerations

PNI-NPNs represent a reduced OPEX/CAPEX deployment option compared to SNPNs as they may leverage the MNO managed infrastructure, owned spectrum, and know-how. As described in the previous subsection, the PLMN-Op may provide the PNI-NPN by means of a DNN or a dedicated network slice.

The implementation of the PNI-NPN presents several issues:

- The on-premises 5G NR connectivity: the gNBs deployed in-house can be owned by the enterprise customer (e.g., purchased directly to the network equipment provider) or made available by the MNO.
- The ability to dedicate and customize the PNI-NPN: the MNO can configure the PNI-NPN in terms of functionality and capacity according to the enterprise customer's needs by provisioning network and application functions specifically dedicated and adjusted to the NPN requirements. For example, the MNO may deploy a customer-tailored, lightweight 5GC that includes only the network functions (UDM, AMF, SMF, Network Repository Function [NRF], UPF) and with the specific capacity as required by the private services.
- The location of the PNI-NPN functions: some NPN scenarios require the network functions to be executed on the customer premises, either for performance or privacy reasons. For example, the UPF may be deployed onsite to reduce the latency. The UDM may also be executed on-premises to keep subscription data locally stored.
- The UE access control: the MNO can enforce the access control by means of the CAG, LADN, and network slicing mechanisms as described in the previous subsection.

In PNI-NPNs, there are scenarios in which the customer enterprise wants to retain control and management of some specific parts of the network. In such a case, hybrid solutions can be defined, with MNOs taking the main control and management activities, while exposing needed capabilities to the enterprise customer. These capabilities can be of two types:

- Configuration related capabilities: this group of capabilities defines the ability of an enterprise customer to modify the parameters of certain network functions and infrastructure nodes. To that end, the PLMN-Op needs to characterize the permissions (i.e., isReadable, isWritable, isInvariant, isNotifiable) associated to these parameters accordingly.
- Assurance related capabilities: this second capability group defines the ability of an enterprise customer to subscribe to certain performance measurements and fault alarms, so that the customer can consume them in the format it sees more appropriate according to its business needs (e.g., for performance management, batches vs streaming).

To make capabilities available for consumption by the enterprise customer, the MNO shall have a business support system (BSS) hosted integration fabric in charge of mediating the request-response messages between the customer and MNO. It is important for the MNO to expose these capabilities in a controlled, secure, and auditable way. To that end, the solution design for this integration fabric will require the implementation of an API gateway, together with mechanisms for token-based authentication and non-repudiation. However, how to build this solution is still unclear, and much work ahead is agreed in the telco industry community. On the one hand, it is still not clear for enterprise customers the capabilities they need to consume for their business processes; this is mostly due to their lack of knowledge/expertise with telco and networking issues. On the other hand, the MNOs need to think about the implementation of this BSS hosted integration fabric, with a particular focus on:

- The control, security, and auditability implications of exposing these capabilities to the customer,

especially considering multi-tenancy environments, where multiple customers will request the PLMN-Op to consume (potentially) different capabilities.

- The mapping of customer requests into network actions, and the API transformation behind this. In this regard, the MNO shall define mechanisms to map customer-facing, service APIs into low-level, internal network APIs. Examples of these customer-facing service APIs will be captured in 5G-CLARITY D4.3 (in preparation).

3.2 Multi-site NPNs

A multi-site NPN scenario represents a deployment use case whereby NPN provisioning aims at serving a given enterprise customer whose facility includes two or more sites, e.g., branch offices. Here, we focus on use cases for multi-site NPNs involving non-critical services that might generate large volumes of data traffic such as UHD video streaming (e.g., telepresence in council meetings, see Figure 3-1) and XR video experience (e.g., AR assisted supervision on a remote factory). These services consist of streaming video traffic among branch offices, harnessing traffic casting (e.g., unicast/multicast/broadcast) mechanisms as needed.

Considering the services mentioned above, we might primarily find two multi-site NPN scenarios: (i) the private 5G network of the different sites belong to the same 5GS, sharing the 5GC-CP, and (ii) the private 5G network of the different sites are supported by different 5GS instances. An example of the first case includes one or several PNI-NPNs supported by the same PLMN for the entire facility (see Figure 3-2). In the second case, we might find dedicated SNPNs (one per site) relying on a fully managed WAN connectivity services (e.g., Software-Defined Wide Area Network [SD-WAN]) for interconnecting them (see Figure 3-3) or dedicated sets of PNI-NPNs, each set per site and provided by different PLMNs.

For the first category, it is assumed there exists a single 5GS for the entire facility. The traffic across facility sites is entirely protected under the umbrella of 3GPP 5G security framework, and the 5GS is now partially hosted by the PLMN. Typical layouts in this category consists in having lightweight branch offices, keeping user plane on premises and offloading 5GC-CP complexity towards a PLMN-Op’s edge node. The resulting deployment scenario is formed of a set of branch offices, each hosting a CP-free 5GS (i.e., RAN and UPF), and a PLMN-Op’s edge node, which hosts 5GC (i.e., 5GC-CP and UPF).

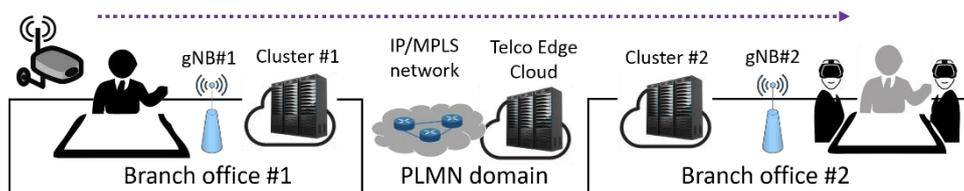


Figure 3-1 Telemeeting as a use case of multi-site NPNs

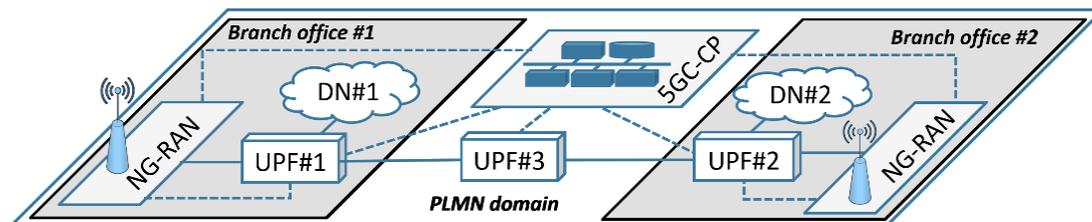


Figure 3-2 Multi-site NPN deployed as a PNI-NPN supported by the same PLMN

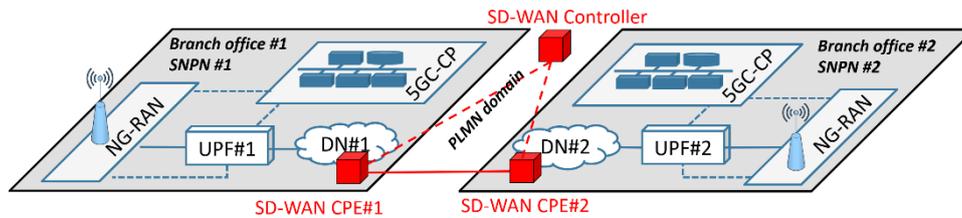


Figure 3-3 Multi-site NPN as a set of segregated SNPNs interconnected through an SD-WAN

Regarding the traffic handling on the scenario described above, the on-premises UPF from source branch office, which performs UL Classifier (UL-CL) functionality, receives incoming IP packets corresponding to video service. Grouped in a PDU session, these IP packets are encapsulated in a GTP tunnel before their delivery to the PLMN hosted UPF. This UPF, deployed at PLMN-Op's edge node and performing the PDU Session Anchoring (PSA) functionality, receives the encapsulated sessions and applies necessary traffic casting policy to route them towards end branch offices, where local UPFs proceed with the GTP tunnel decapsulation, so IP packets can reach end users. In the overall process, participant UPFs are in charge of keeping 5QI-to-DSCP mapping (i.e., translation of 3GPP 5G QoS indicators into IP QoS indicators), so the QoS can be assured along the IP/MPLS substrate which connects the different branch offices.

The second category is typical for Industry 4.0 enterprises, where independent 5G-enabled manufacturing tasks are executed at individual branch offices. The branch offices need to communicate between them to only exchange industry-specific data (e.g., file exchange, database accessibility); this means that no signalling/data plane 5G traffic is exchanged among individual SNPNs. For this communication, a plausible solution is to set up an overlay connectivity service (e.g., SDWAN) atop the MNO's IP/MPLS substrate. In these setups, the enterprise customer demands the data to be protected when travelling across sites through the transport network. Now that the 3GPP 5G in-built security features no longer apply in this inter-site communication (see Section III-F), the transport network operator needs to find workarounds. One solution is to use LxVPN services (e.g., L2VPN or L3VPN) with IPsec in the underlay, which ensures confidentiality (avoids external visibility on exchanged traffic) and integrity (prevents payload modification, e.g., Denial of Service (DoS) and fraud). This carrier-grade solution, widely used in today's enterprise connectivity services, may remain valid for quite a large portion of customers in the multi-NPN category.

3.2.1 Example of 5G-CLARITY framework architecture in single- and multi-site NPNs

This subsection introduces illustrative examples of the 5G-CLARITY framework architecture enabling single and multi-site NPNs based on the 5G-CLARITY use case 1 (UC1) 'Human and Robot Interaction' set-up proposed on 5G-CLARITY D5.1 [18]. Furthermore, this example develops the stand-alone NPN (SNPN) and PNI-NPN implementations of the UC1 to enable the services delivery models introduced in 5G-CLARITY D4.2 [6].

3.2.1.1 Single Site NPNs

Figure 3-4 presents the 5G-CLARITY UC1 set-up as a single site SNPN scenario enabling services examples of NFVaaS and WATAaaS. The provision services for this setup are:

- Services 1: Third-party robot assistance, advertising, and video services virtual network function/physical network functions (VNFs/PNFs) hosted in the NFVI of the 5G-CLARITY Edge cluster, as an example of NFVaaS.
- Service 2: MNO or VMNO sharing 3GPP and non-3GPP wireless/mobile infrastructure of the sNPN to extend their coverage and enable a dedicated UPF as VNF, an example of WATAaaS.

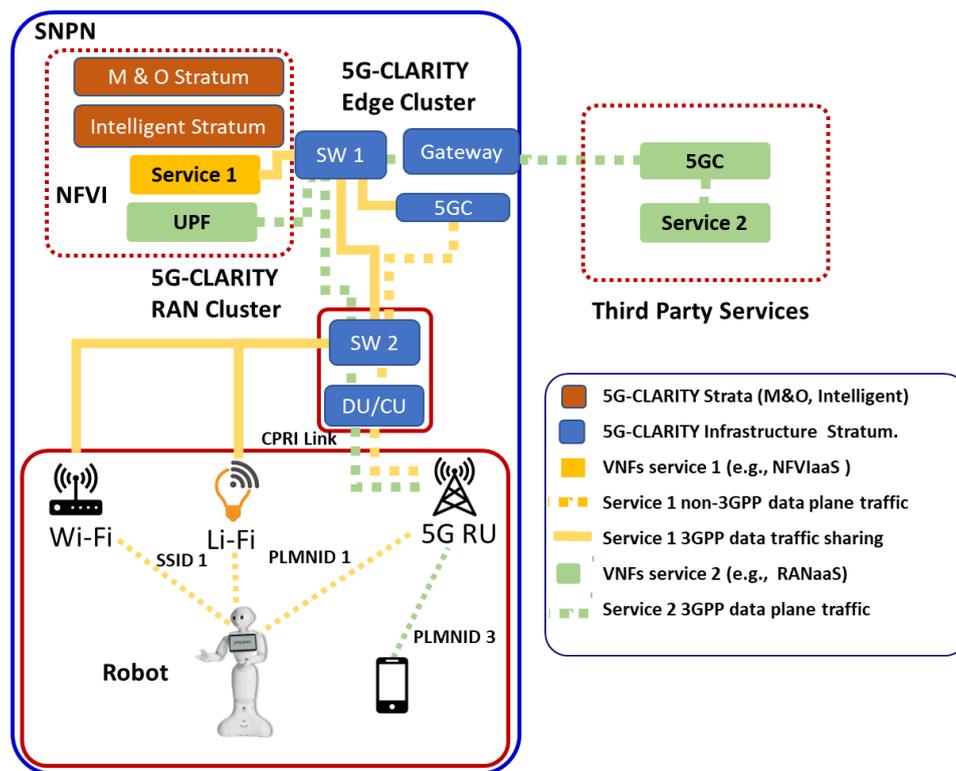


Figure 3-4 Single site SNPN setup with **5G-CLARITY** Framework components, enabling two services

Service 1 (yellow) is sharing the **5G-CLARITY** multi-WAT network to deploy robot assistance applications. When the robot moves, the capability to handover between different WATs and to aggregate multiple connections will provide resilience and adaptability to changes in the network. The service will use an internal or NPN PLMNID 1 of the 5G NR connections and SSID 1 for Wi-Fi and Li-Fi. On the other hand, Service 2 (green) represents an MNO or MVNO extending service footprint through the premises of the SNPN without sacrificing the security and privacy (exposure) of the private network and the MNO and MVNO traffic using their PLMNID 3 (Public). In terms of connectivity, the gateway can only connect a dedicated link through a transport network provider running VPNs or MPLS protocols or through a dedicated fibre optics or optical link. The multi-WAT capability of **5G-CLARITY** framework will allow the aggregation between non-3GPP wireless technology provided by the NPN and 3GPP mobile technology managed by VMNOs and MNOs (VMNO and MNO).

3.2.1.2 Multi-site NPNs

Figure 3-5 introduces multi-site NPNs with a PNI-NPN scenario by extending the UC1 deployment described in Figure 3-4. In this example, we describe three services in multi-sites NPNs:

- Service 1 – Robot assistance (**5G-CLARITY** UC1 ‘Human Robot Interaction’) and third-party advertising and video services introduced in Figure 3-4, extended between multi-site NPNs integrated by a public network as an example of extended NFVaaS scenario.
- Service 2 - MNO or VMNO deployed between multi-site NPNs to extend their coverage through two NPNs, extended example of WATaaS.
- Service 3 - Introduced as ultra-high definition (UHD) video and immersive services that require high performance and processing and low latency only provided by an E2E Slice or SLaaS scenario.

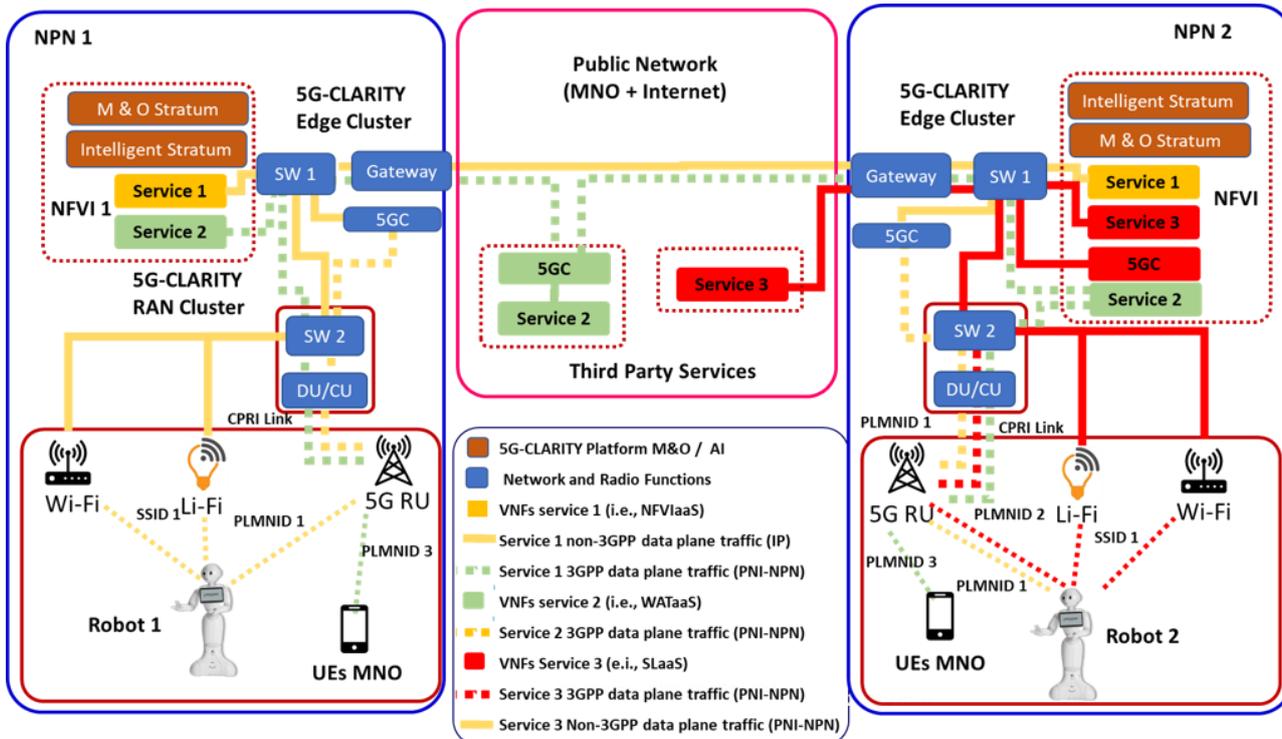


Figure 3-5 Two NPN sites enabling PNI-NPN with 5G-CLARITY framework delivering 3 types of services

In Figure 3-5, Service 1 (yellow) is primarily deployed in a single site VNF to provide advertising, web services and robot assistance services (Robot 1) through multi-WAT. Then the third-party company request an extension of their services for a short period or event and in a specific location of the NPN 2. Then, a VNF is onboarded in the NFVI of the NPN 2 to host a local advertising server interconnected with the NPN 1 through a dedicated connection between two public gateways. The services provided in NPN 1 cover the whole premise (Robot 1), while the multi-site extension areas are covered only by 5G technology. This is due to the higher independence from the control of Robot 2 compared to Robot 1.

In the extension of service 2 (green), the mobile operator already sharing the RAN deploying PLMNID 3 on NPN 1 decides to extend its coverage. As a result, NPN 2 setups a dedicated connection with the 5G control plane and data plane via the public gateway of NPN 2. Furthermore, because some applications require edge access, a UPF is also deployed in the NPN 2 edge cluster to serve the PLMNID 3 (UEs of MNOs) in both sites, or in the NPN. Therefore, the WATAaS scenario presented is a PNI-NPN setup where MNOs or VMNOs can benefit from the NPNs to extend their coverage by sharing 3GPP and non-3GPP technologies deployed by the NPN.

Service 3 (Red) represents UHD and Immersive services requiring enormous computational and network resources from NPN 2. For service 3, Graphical Processing Unit (GPU)s and extensive storage in the Edge and communication capability must be reserved and isolated from other services. 5G-CLARITY Infrastructure Stratum enabling hardware accelerators and slicing capabilities (SLaaS) will create an E2E slice with its own 5G control and user plane (5GC/UPF) with dedicated storage and GPU at the Edge Cluster. Indeed, the service 3 example covers a single NPN site. However, it is still a scenario of PNI-NPN, given that the service requires public network connectivity for the service to be operational.

3.2.2 Mobility in NPNs

There are numerous use cases for 5G NPNs in which the devices eventually leave and enter the private premises or even move between remote sites of the same private entity. Examples of these use cases include the use of drones for monitoring of crop growth in agriculture and package delivering in logistics. In the

context of **5G-CLARITY** use cases, for UC2.1 [3] real-time tracking of goods might be needed when they are conveyed between Bosch manufacturing and the distribution center. These use cases require a PLMN providing wireless connectivity out-of-premises and mechanisms to ensure a given Session and Service Continuity (SSC) mode. 3GPP standards support three SSC modes: SSC mode 1, SSC mode 2, and SSC mode 3. SSC mode 1 ensures IP address continuity, i.e., the same IP address is preserved, during and after the handover procedure to handle UE mobility. In SSC modes 2 and 3, there is a UPF PDU Session Anchor (PSA) relocation, which means that IP address continuity is not enabled. In contrast to SSC 1 and SSC 3 that have a “make-before-break” operation, SSC mode 2 follows a “break-before-make” approach, then, the connectivity is interrupted during the handover. The SSC mode to be provided might depend on the service characteristics. Here, we consider services requiring a seamless experience when the private UE moves between different mobile networks premises, thus we focus on SSC modes 1 and 3-based mobility solutions for NPNs. Please note that, overall, SSC mode 1 applies only to intra-NPN mobility. For inter-NPN, inter-PLMN or NPN to PLMN mobility, SSC mode 3 is used to ensure seamless connectivity for the private UE. Please note that SSC mode 3 does not provide IP address continuity, thus the application/service oversees handling the IP address change without much impact to the user experience [17].

The 3GPP capabilities and mobility mechanisms employed to ensure service continuity in NPNs depend on the 5G NPN deployment option. For instance, for SNPNS the solution to defined in 3GPP standards to access public services from an SNPNS (and the other way around) can be leveraged for seamless mobility. In contrast to SNPNS, for PNI-NPNs, the UE might not need to move out of the coverage area of the PLMN supporting the respective PNI-NPN when it leaves the private premises. In this case, ordinary intra-PLMN handover procedures may be used. Also, the UE capabilities, e.g., whether the UE is equipped with Dual Subscriber Identity Module (SIM) or not, can have an impact on the available solutions for mobility.

Next, we will summarize the mobility mechanisms available in 3GPP standards for each deployment option. **5G-CLARITY** system is 3GPP standard compliant, and therefore can leverage all these mobility support features. Following that, the roaming solution considered in the **5G-CLARITY** system is described. This solution enables UEs equipped with a single SIM to achieve seamless mobility between several PLMNs and SNPNS.

3.2.3 Mobility in PNI-NPNs

In the case of PNI-NPNs, the same PLMN supporting their deployment can provide radio access both on-premises and out-of-premises. Under this consideration, the intra-PLMN handover procedures defined in 3GPP standards can be used to provide SSC when UE crosses the private site borders.

The UE might also move out of the coverage area of the PLMN supporting its PNI-NPN, hereinafter referred to as home PLMN (HPLMN). In this case, the UE needs to roam onto a different PLMN, referred to as to visited PLMN (VPLMN), for broadband wireless connectivity. In contrast to 3GPP Rel-15, Rel-16 enables a new feature: PDU Session continuity support for roaming. The roaming use cases in scope include, among others, mobility between VPLMNs, mobility from HPLMN to VPLMN and the other way around, and intra-VPLMN mobility between two visited Session Management Function (V-SMF) entities. This is possible because 5G also enables the insertion, change, and removal of V-SMF entities, which facilitates the seamless connectivity. In Subsection 3.2.5, the enhanced roaming solution considered in the **5G-CLARITY** system is described. This solution can be used in PNI-NPNs to ensure a seamless experience in roaming scenarios.

3.2.4 Mobility in SNPNS

3GPP Rel-16 specifications define a configuration to access PLMN services through an SNPNS (see Figure 3-6). This configuration enables a private UE registered in a SNPNS to perform another registration with the PLMN. To that end, there is a two-step procedure defined in 3GPP standards.

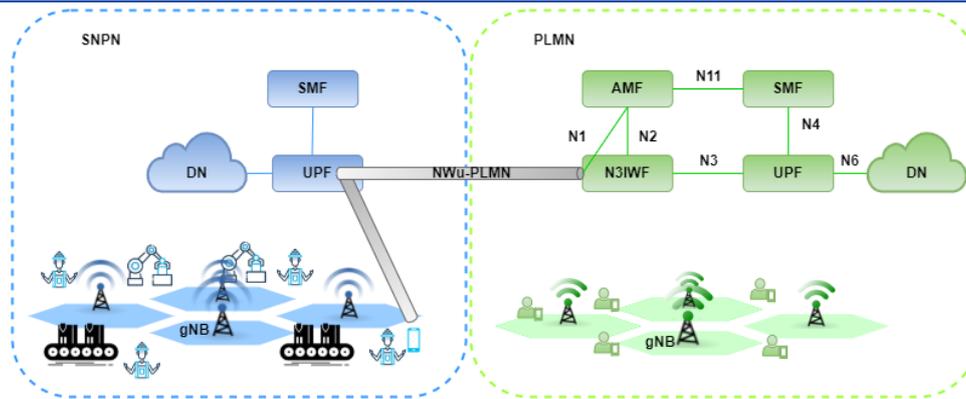


Figure 3-6 Configuration defined in 3GPP Rel-16 to access to PLMN services via SNPN

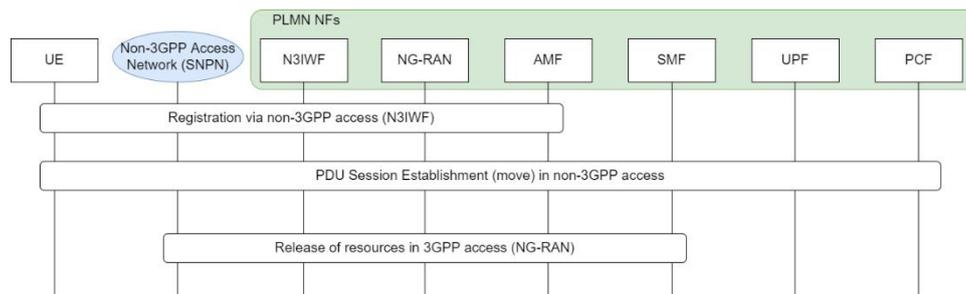


Figure 3-7 Procedure to perform a handover from an SNPN to a PLMN. The SNPN is seen as a non-3GPP access network

First, the private UE harnesses the SNPN subscription to get external access. Second, the UE employs the PLMN subscription to connect to the 5GC of the PLMN. The SNPN acts as an untrusted non-3GPP access point from the PLMN viewpoint. An IPsec tunnel might be established between the UE and the public N3IWF. The standards also define a symmetric configuration to access SNPN services from a PLMN.

In the configurations mentioned above, the Handover of a PDU Session procedure from untrusted non-3GPP to 3GPP access can be used for mobility. Figure 3-7 shows the call flow of this procedure. Assuming the private UE is equipped with dual SIM and has a subscription with the PLMN, the handover procedure can be performed without service interruption. The server-side app (DN hosted AS) is responsible for handling the connections and traffic when the UE is simultaneously connected to both the SNPN and the PLMN.

The solution described above requires the private UE has a subscription with the target PLMN. Then, the UE has to be equipped with dual SIM, embedded SIM (eSIM) or multi-International Mobile Subscriber Identity (IMSI) support. In any case, the number of networks the private UE might roam onto is limited. Furthermore, this solution is not valid for a UE equipped with single SIM. The roaming solution for 5G-CLARITY system described in Section 3.2.5 might overcome these issues.

3.2.5 5G-CLARITY system enhanced roaming support

In this section, we address the enhancements proposed in 5G-CLARITY project for roaming support. There exist many private use cases that include devices equipped with a single SIM [19], such as Internet of Things (IoT) sensors. Under this consideration, roaming procedures are suitable for these private use cases when the UE moves between many different PLMNs or NPNs. Although 3GPP Rel-16 defines inter-PLMN mobility procedures for both local breakout and home-routed roaming scenarios, some considerations and enhancements have to be taken into account in order to use them when a private UE roam onto a visited NPN or PLMN. Specifically, some requirements must be fulfilled. First, there must be direct communication

between the 5GC-CP instances of the SNPN and the PLMN. Second, the public UDM shall perform an onboarding of the UE subscription data by requesting them through the private UDM. Last, network functions from the two 5GC instances must interact with other, to perform the respective handover procedures. Nonetheless, the roaming architectures defined in 3GPP standards require enhancements to cover the necessities of the NPNs use cases, as pointed out in [20].

Figure 3-8 shows an enhanced roaming architecture for the 5G-CLARITY system based on some of the requirements and extensions provided in Section 3.2.5. First, to ease the addition of new allowed visited networks in the NPN, the ‘visited network discovery and selection function’ is included in the control plane [16]. This function provides the private UE with the list of available networks in its vicinity. Similarly, the visited network might include a Home NRF to store data (e.g., IP addresses to reach the targeted 5GC network functions of the home networks) related to the networks it has roaming agreement with. The UPF and the SEPP have also to be extended for executing the required operations to provide the required security and isolation levels in NPNs. The management and orchestration solution of the visited network shall verify and validate the data and control planes connectivity between it and the 5G-CLARITY system. To that end, the stack of the visited network might interact with the SDN controller of the SD-WAN providing the connectivity. Similarly, the management and orchestration stratum of the 5G-CLARITY system (see Section 2.3) oversees ensuring that the networks interconnecting the private UE with the cloud hosting the application VFs will fulfil the QoS requisites of the specific service through leveraging the capability exposure functionality. Last, the roaming process could benefit from the 5G-CLARITY system intelligence stratum. For instance, AI-assisted UE mobility prediction might help to decide among different visited networks candidates. Also, UE mobility prediction might be harnessed to proactively carry out some needed configurations in the target visited network in advance, thus speeding up the roaming procedure. On the other hand, the expected UE traffic load profile via predictive analytics functions running of the IS might be needed by some network domains to perform resource allocation for the UE.

A tentative call flow for a roaming procedure leveraging architecture presented above is depicted in Figure 3-9. First, the UE gets the list of candidates visited networks from the VNDSF.

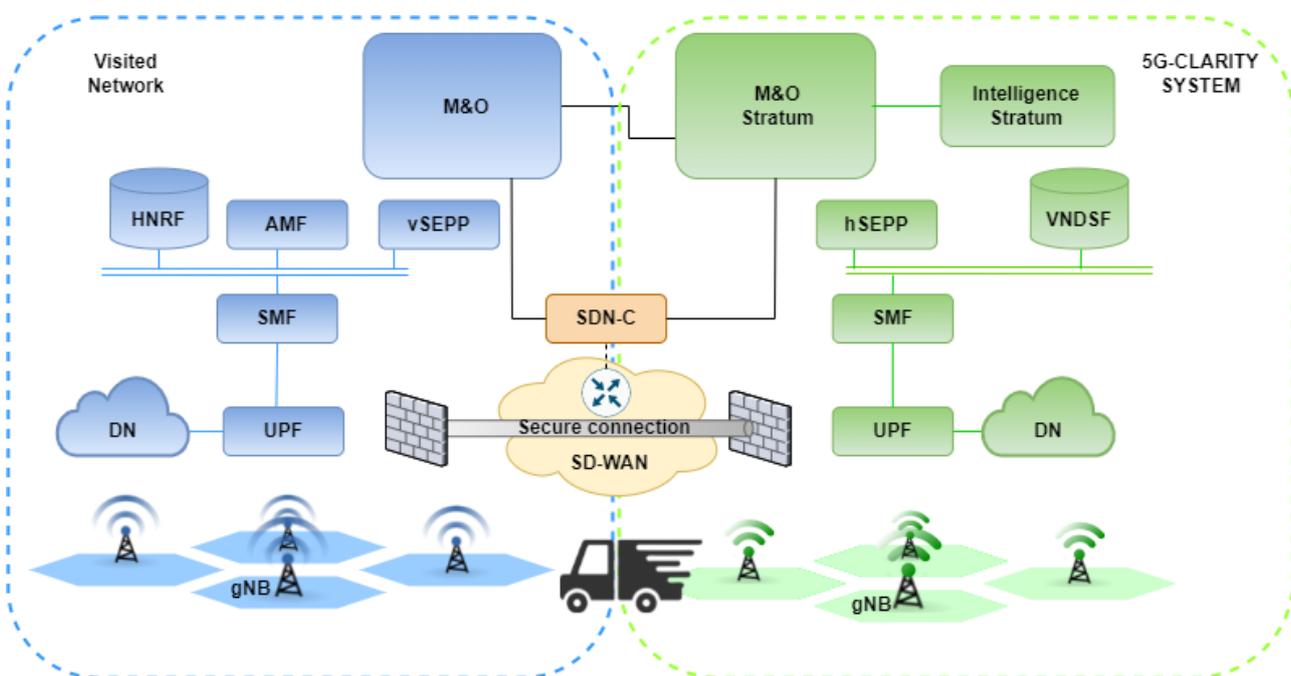


Figure 3-8 Roaming architecture considered for the 5G-CLARITY system

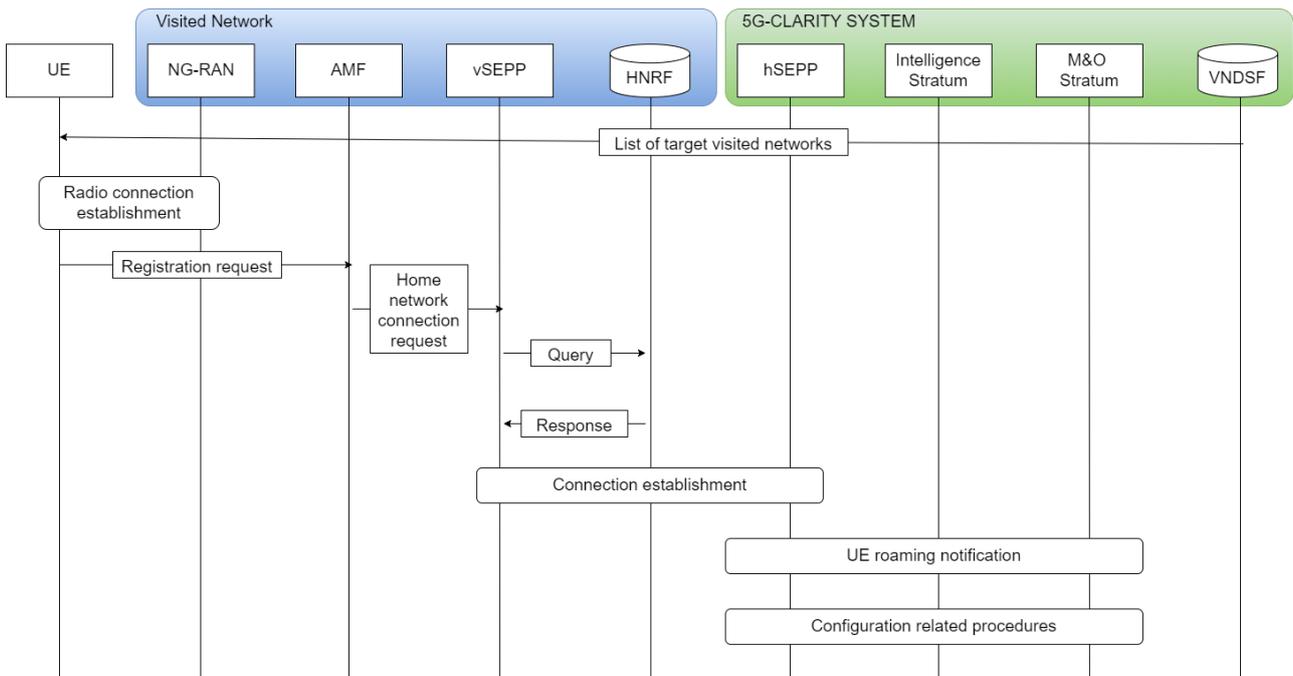


Figure 3-9 Roaming procedure call flow considered for the 5G-CLARITY system

Then, it performs a radio connection establishment with the NG-RAN of the chosen visited network and sends a registration request to the AMF of the selected visited network. The AMF checks whether the UE is local or not. If not, the AMF forwards the request to the vSEPP. Once the vSEPP receives the request, it verifies whether there is already an established connection with the home network of the UE. If not, it queries the HNRF for the required information to establish the connection with the home network of the UE as described in Figure 3-9. Lastly, the hSEPP might notify the mediation function and 5G-CLARITY Intelligence stratum to intervene in the setup of the UE connectivity and ensure the requisites of its ongoing communications.

In order to enable low latency, the CAFs might be migrated to clouds in the vicinity of the UE as it moves through the premises of different networks. In roaming scenarios, this solution is used in local breakout setup for low latency. Again, this process might be assisted by the 5G-CLARITY intelligence stratum. In this deliverable, the benefits in terms of performance brought by this solution are quantitatively evaluated for intra-NPN mobility scenarios in the 5G-CLARITY system. The requirements and details to realize this feature are included in Section 4. The modelling, experimental setup, and evaluation results are reported in Sections 4, 5.1 and 6.1, respectively.

3.2.6 Example of flows for call and connection requests on Single and Multi-Sites NPN

Based on the call flows introduced before we present an example of call flows and connection requests required by the services introduced in Figure 3-4 and Figure 3-5. Five flows are presented:

Flow 1 - multi-WAT UE attach and connection requests in a SNPN: In this example the CPE represent a UE requiring 5G, Wi-Fi and LiFi connectivity to provide control for a guide robot (UC1 scenario).

- 5G port/modem: CPE sends to the 5G gNB a registration request for the SIM-1 of the SNPN PLMNID. And the when the request is approved, the 5G port/modem setups a PDU session for the data plane.
- Wi-Fi and LiFi ports select the SSID 1 and complete the authentication.
- The network connection between Robot, CPE and UC1 service is setup.

Flow 2 – Service 1 deployed as NFVlaaS using the already established multi-WAT network between SNPN and third-party services hosted or connected through the MNO (5G-CLARITY UC1 scenario).

- Third party service request NFVlaaS to the 5G-CLARITY Intelligent stratum function Intent Engine.
- Intent engine requests to the 5G-CLARITY M & O stratum function, NFVO to on board a service or VNF and setup a VPN connection between Gateway and third-party content server.
- VPN Connection between GW and Content Service is provisioned and VNF Video & Ad Services is onboarded by the NFVO.
- Finally, the Third-party video and ad service is updated in the NFVI and a connection to the UE is established.

Flow 3 – Service 2 deployed as WATaaS example using a UE with MNO or Public Network SIM/PLMNID.

- As initial step as Flow 2 the MNO/VMNO requests WATaaS from the NPN by setting a connection between their 5GC to the GW of the NPN and requesting a VNF for their edge UPF function.
- Once the connectivity is setup and UPF is onboarded the 5GC establish the connections N9 and N4 to the UPF and N2 to the gNB.

Flow 4 – Service 2 UE call or connection request flow.

- Flow 3 completed, UE with SIM of the MNO connected can request radio connectivity and registration similarly to Roaming process as introduced before.

Flow 5 – Service 3 UE call or connection request flow in a slice.

- 5G port/modem Sends of the CPE send radio connection request to the 5G gNB and send the Registration Request for the SIM- 1 of the SNPN PLMNID. Finally, the 5G port/modem setups a PDU Session for the data plane.
- Wi-Fi port select the SSID 1 and complete the authentication.

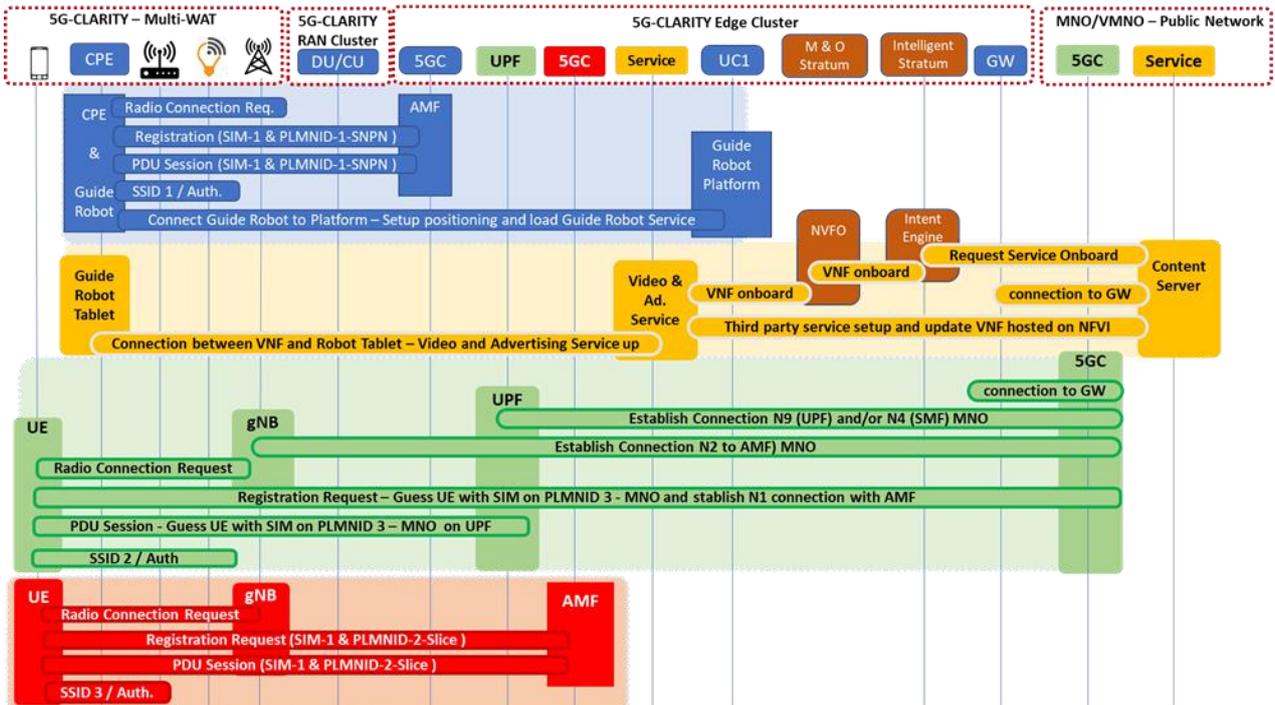


Figure 3- 10 5G-CLARITY UC1- single-site (SNPN) and multi-site (PNI-SPN) scenario flows – single and multiple SIMs

3.3 Data aggregation in NPN-PLMN scenarios

For the interoperability between private and public networks in NPN-PLMN scenarios, to ensure a unified operation of these public and private domains, it is required that the management systems of the private and public network operators interact with each other, exchanging trusted and verifiable messages between them.

In order to facilitate the interaction between the management systems in NPN-PLMN scenarios, namely the 5G-CLARITY management system and the PLMN management system (e.g., 3GPP management system), the mediation function component plays a key role. The mediation function is the single-entry point to the 5G-CLARITY management system, providing a controllable capability exposure to external management systems (e.g., PLMN management system). To that end, the mediation function implements API registry/discovery, access control (e.g., authentication, authorization) and associated services such as auditability, traceability, and logging.

Within the monitoring functions of the 5G-CLARITY management system, the data semantic fabric (DSF) framework constitutes one of the main blocks for the data processing and management subsystem. The DSF is a semantic, model-driven monitoring framework that enables data collection, data transformation, and data aggregation from different monitoring elements, and coordinates the flow of these data among a set of heterogeneous data sources and data consumers. By using formal data models, the DSF adapts data collected from the available sources into the formats suitable for consumers. In this regard, the DSF can be used to aggregate data from heterogeneous sources in multi-domain scenarios, and in particular to aggregate data from external data sources available in a public domain. Taking advantage of the capability exposure feature from the Mediation Function, the DSF can register data sources from the public domain to be able to collect and aggregate the regarding telemetry data inside the 5G-CLARITY infrastructure system.

In the following subsections, the management services exposed by the DSF framework are explained to cover the data aggregation process for telemetry data in NPN-PLMN scenarios.

3.3.1 Data semantic fabric management services

In order to detail how to apply data aggregation mechanisms in NPN-PLMN scenarios, it is necessary to review and detail the management services exposed by the DSF framework, which were already defined in the previous deliverables. The detailed information of these management services exposed by the DSF is as follows:

- **Source Registry:** Service that allows managing the registration of data sources. Registration must include required information for the DSF to set up a connection to the data source, namely, IP address, port, and authentication credentials. Depending on the authentication mechanism these credentials could range from basic username-password tuple to *OAuth* token. The resulting connection will be used by the DSF to ingest data from the registered data source, but additionally, the same connection may be leveraged to retrieve context information associated with the data source itself, e.g., discovery of objects and buckets available in AWS S3. This dynamic discovery of additional capabilities from data sources is covered in the Capability Registry management service. It is important to highlight that the Source Registry service allows the registration of heterogeneous sources of different nature and multiple domains, allowing, among other cases, to cover scenarios such as NPN-PLMN. In this multi-domain scenario, the DSF must address both the registration of internal sources that belong to the 5G-CLARITY infrastructure domain (i.e., on-premises workloads), and external sources that belong to the PLMN domain (i.e., off-premises workloads). For external data sources, the PLMN management system could trigger the Source Registry service through the Mediation Function component of 5G-CLARITY infrastructure.

- **Consumer Registry:** Enables the registration of data consumers. This service is similar to the Source Registry with the difference that the configured connection will be used by the DSF to deliver data to the registered data consumer. Regarding context information associated with the data consumer, most of the information will be known in advance by the DSF as it has control over what data are produced and where, e.g., data sent to a Kafka topic. Therefore, in this sense, the connection with the data consumer may be needed only to collect complementary context information from the data consumer, e.g., metadata generated by the consumer such as internal identifiers. As with data sources, the dynamic discovery of additional capabilities from data consumers is covered in the Capability Registry management service.
- **Data Pipeline Provisioner:** This service enables building data pipelines within the DSF that coordinate the exchange of data between data sources and data consumers. The composition of a data pipeline can be divided into three main blocks:
 - **Data Collection:** Selection of data sources amongst those registered in the DSF – through the Source Registry service – to ingest data from. The configuration of this step depends on the capabilities exposed by the chosen data source. For example, a configuration to ingest data from a REST-based would require specifying an endpoint URL along with an interval to periodically request for new data.
 - **Data Aggregation:** Block that configures transformations to be applied on the data ingested in the Data Collection block. This block supports defining chains comprising multiple data transformations. The configuration of each step in the chain depends on the type of the chosen data transformation. Data transformations can consist of data filtering, combining telemetry data from different sources, or aggregating data for calculating valuable KPIs, among other things.
 - **Data Dispatch:** Selection of data consumers amongst those registered in the DSF – through the Consumer Registry service – to deliver data to. Aggregated data resulting from the Data Aggregation block are adapted into a format suitable for the destination data consumers. In the same fashion as the Data Collection block, configuration details needed to deliver data to the chosen consumers must be provided. Such configuration depends on the nature of the data consumers. For instance, when a data consumer is a MySQL database, the destination table must be indicated in the configuration.
- **Capability Registry:** Service that publishes capabilities of data sources and data consumers that have been registered in the DSF by making use of the Source Registry and Consumer Registry services respectively. The concept of capabilities refers to the additional context information associated with the data sources or data consumers themselves. This additional capability information could be, for example, the list YANG models supported by a network device, or the list of objects and buckets available in AWS S3 service for the Data Lake solution. Depending on the nature of the source or the consumer, related capabilities could be either dynamically discovered by the DSF or provided by an external component that accounts for the related source or consumer. In the latter scenario, an external component would register those discovered capabilities in the DSF by invoking the Capability Registry service.

3.3.2 Use case: data aggregation in NPN-PLMN using DSF

This subsection details the application of the 5G-CLARITY DSF framework in NPN-PLMN scenarios with a familiar approach. Let's consider a possible scenario where the same AWS cloud-based Data Lake solution targeted in 5G-CLARITY is deployed in the PLMN domain.

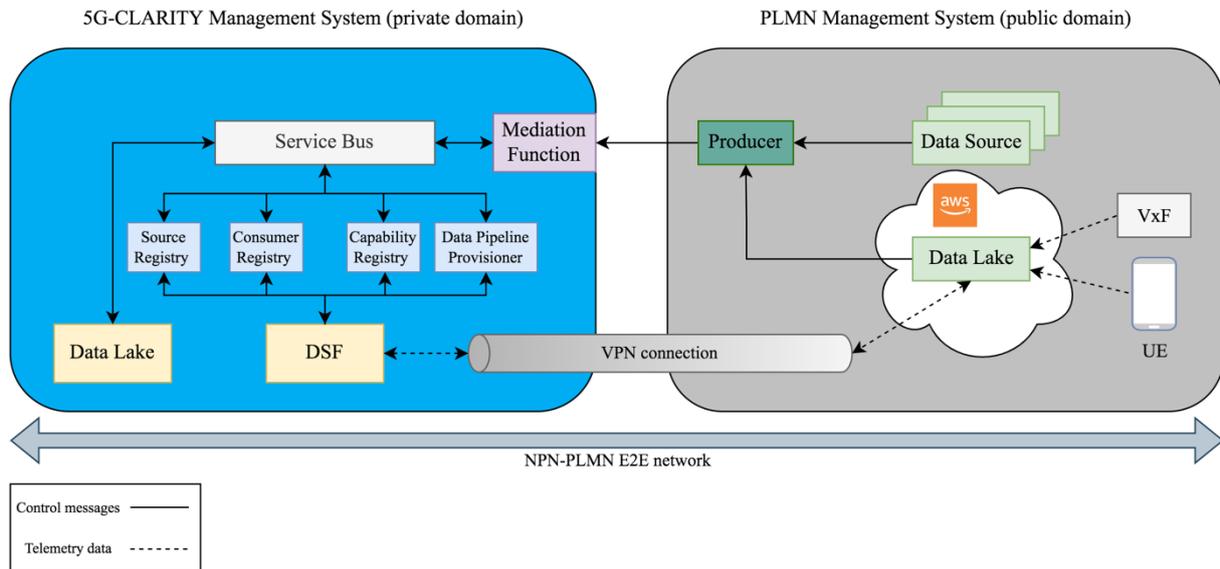


Figure 3-10 Data aggregation in NPN-PLMN scenarios using the DSF

Therefore, suppose that the same Data Lake solution is deployed in a public cloud, which is controlled by the PLMN NOP, and the related telemetry data it stores wants to be aggregated from the 5G-CLARITY infrastructure.

For this approach, let's think about a specific mobility use case in which a particular user equipment (UE) moves beyond the private coverage area and thus into the public coverage area. To maintain service continuity, the UE is registered in the PLMN domain, and its telemetry data is stored in the off-premises Data Lake (i.e., the Data Lake belonging to the PLMN domain). On the other hand, let's think about another use case in which a private workload, meaning a VxF, needs to be hosted by the PLMN domain due to lack of resources in the 5G-CLARITY infrastructure. Then, the telemetry data related to the private workload that was collected in the on-premises Data Lake (i.e., the Data Lake belonging to the 5G-CLARITY infrastructure) will now be collected by the off-premise Data Lake. In these particular use cases, it is needed to aggregate the related telemetry data stored in the off-premises Data Lake from the DSF framework of the 5G-CLARITY infrastructure. The Figure 3-10 shows an overview of the scenario introduced, with the main entities involved in the management systems of the public and private domains and the interactions between them.

The interaction workflow between the management systems of the public and private domains for this data aggregation scenario is described step-by-step below (the sequence diagram with each of the control steps associated with the interaction workflow is depicted in Figure 3-11):

1. In order to aggregate data from a workload located in the public domain, the management system of the PLMN NOP needs to register the external data source in DSF. For that purpose, an external system that accounts for the related data sources in the public domain (i.e., the producer entity depicted in the Figure 3-10) can perform the registration process by consuming the source registry service through the mediation function. In this sense, the mediation function of 5G-CLARITY controls that the authorized system with the role of producer can securely access the Source Registry management service of the DSF framework. As specified in the description of the Source Registry service in the previous subsection, the registration must include required information for the DSF to set up a connection to the data source. In the case of the off-premises Data Lake working as a data source for the DSF, the registration process must provide information such as the endpoint URL associated with the API Gateway of the AWS S3 service, the API key credentials, and the target regions for the related S3 buckets in which the telemetry data are stored. Additionally, as part of the source registration process, the PLMN NOP could consider providing a VPN connection to allow

communication between the public and private domains to enable collecting the telemetry data.

2. However, there could be scenarios where additional capability information of registered data sources might be needed. For instance, when the off-premises Data Lake works as a data source, the DSF registers important context information such as the list of objects and buckets available in AWS S3 service that are allowed to be accessed from the private domain. For the registration of these capabilities, there are two possible scenarios:
 - a. Capability discovery process from the public domain side: The Producer system that accounts for the related data sources in the public domain implements the logic to discover their capabilities and publish them by consuming the Capability Registry service through the Mediation Function. Considering the particular case of the Data Lake working as a data source in the public domain, there are some services such as AWS Glue⁴ that enable a data catalogue to store metadata associated with AWS-based solutions, which could be used by the *Producer* system to obtain this type of additional capability information from the Data Lake.
 - b. Capability discovery process from the private domain side: The DSF implements additional mechanisms that dynamically discover the capabilities of data sources. These capabilities are then registered in the DSF directly via its Capability Registry service. In this sense, for the DSF to automatically discover the capabilities associated with data sources belonging to the public domain, the PLMN management system must implement an API Gateway-based solution like the 5G-CLARITY Mediation Function. This solution should enable the PLMN domain to securely expose those management functions and services related to the data sources to the 5G-CLARITY management system. Considering the case of the Data Lake working as a data source in the public domain, the DSF can interact with the Data Lake's particular API Gateway, which is based on AWS solution, for the discovery of some interesting associated metadata.
3. For those systems interested in consuming telemetry data that comes from data sources belonging to the public domain, the Consumer Registry service provided by DSF must be used. In this sense, an operational user within the 5G-CLARITY infrastructure could register a data consumer in the private domain. Considering the particular case of the on-premises Data Lake working as a data consumer in the private domain, the consumer registration process would be very similar to the source registration process. In this case, information such as the URL associated with the API Gateway, the API key credentials, and the target regions for the related S3 buckets where the private Data Lake will store the incoming telemetry data will be required.
4. Once the registration processes of data sources, data consumers and their related capabilities are covered, the exchange of data between them can be triggered by building data pipelines within the DSF through its data pipeline provisioner service. In this sense, in the same way that happened with the consumer registration, an operational user within the 5G-CLARITY infrastructure could build data pipelines in the private domain. On the one hand, the data pipeline can define a direct forwarding of telemetry data from the source to the consumer, e.g., ingest telemetry data from the off-premises Data Lake and deliver it to the on-premises Data Lake. On the other hand, the data pipeline may include a data aggregation phase in which additional transformations are performed. For instance, different telemetry data belonging to the off-premises Data Lake may be combined to perform transformations based on stream processing applications in the DSF and deliver the resulting data to the on-premises Data Lake.

⁴ AWS Glue - <https://aws.amazon.com/glue/>

In summary, the aggregation process for telemetry data has been demonstrated in NPN-PLMN scenarios. To address this problem, a use case is presented where telemetry data related to the public domain is seamlessly aggregated within the private domain. As a result, the capabilities of the DSF framework are shown to aggregate data from telemetry sources in the public domain, applying each of the management services involved.

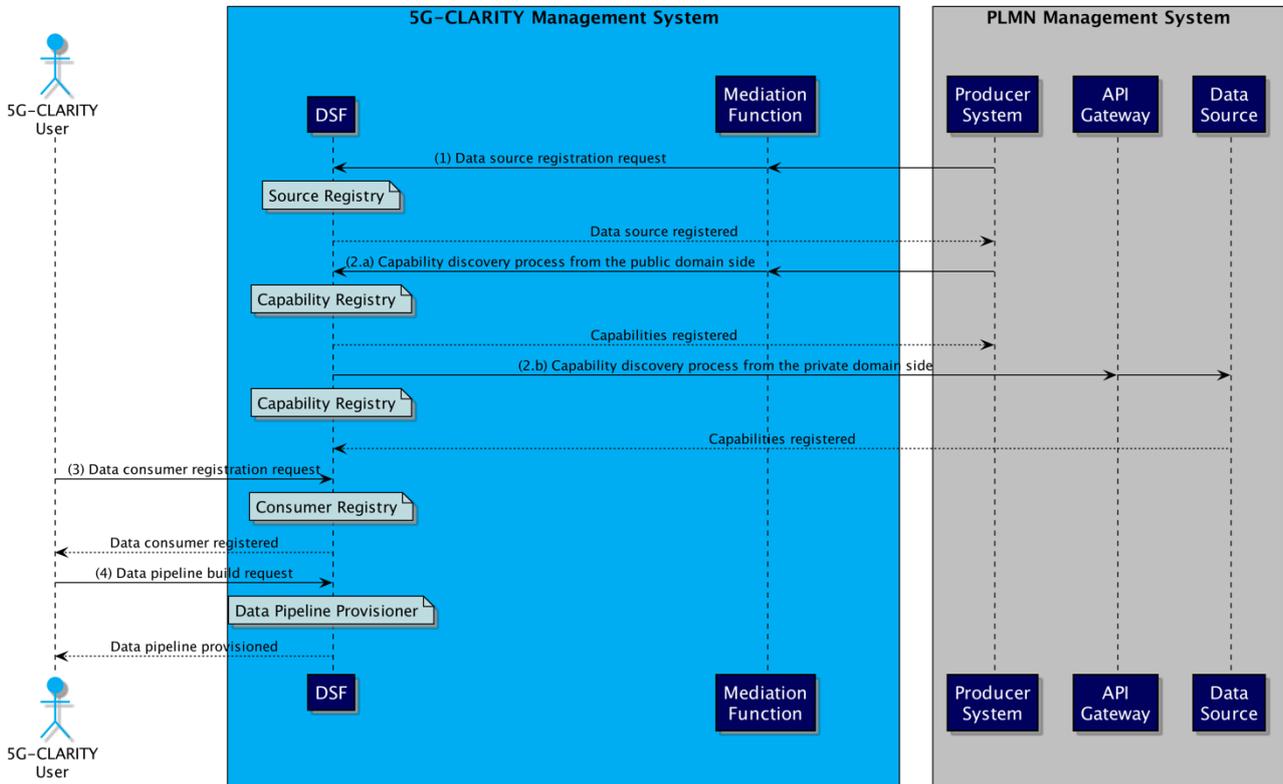


Figure 3-11 Sequence diagram about the data aggregation workflow between the management systems of the public and private domains

4 Final 5G-CLARITY system evaluation

This deliverable aims to leverage the topics and scenarios presented in D2.3, refining and extending them to address the technical challenges identified, including:

- the need to effectively coordinate different type of controllers and orchestration platforms to minimize service disruption time under mobility. This also requires the development of predictive and prescriptive analytics schemes that can predict traffic and UEs trajectory and optimize use of resources.
- the need to tighter coordinate the multi-WAT segment of 5G-CLARITY through the development of more accurate channel model techniques and integration of multi-WAT with slicing.
- the need to further improve the performance of the joint synchronization and positioning module under harsh conditions such as non-line-of sight.

To address these challenges, the scenarios under considerations have been purposely extended, following the guidelines for next steps captured in D2.3, Section 7. A summary of these refinements/extensions is provided in Table 4-1.

Table 4-1 Extensions of 5G-CLARITY evaluation studies

Feature	Enablers	Scenario	Refinements/modifications in relation to D2.3
Feature #1: Mobility in public-private 5G network scenarios.	<p>5G-CLARITY system managed compute and connectivity able to interact with MNO provided PLMN resources for the realization of public network integrated NPNs.</p> <p>D22-CLARITY-NFAS-R20: The 5G-CLARITY network function and application stratum shall provide necessary telemetry data to the hosted xApps.</p>	Multi-service provisioning in public-private 5G industrial networks	<p><u>Enhanced modelling tools:</u> YES, Modelling of the optimal service provisioning problem under mobility using mathematical optimization models running in the orchestrator.</p> <p><u>Refined setup configuration:</u> YES, Application Service (AS) migration functionality from public to private, configuration and testing with multiple gNB configurations.</p> <p><u>Change of validation environment:</u> No</p>
Feature #3: Multi-WAT capabilities	<p>The 5G-CLARITY enablers in which this evaluation scenario puts its focus are the design of the infrastructure stratum and the functionality provided by the Network Function and Application Stratum that allow the management and control of wireless resources (see requirement CLARITY-INF-R2 in Table 2-1 of section 2.1, and KPI D22-CLARITY-NFAS-R18 in Table 2-2 of section 2.2)</p>	Wi-Fi offloading in industry 4.0	<p><u>Enhanced models:</u> YES, the LiFi communication channel model has been integrated in the system-level industrial RAN simulator used in D2.3.</p> <p><u>Refined setup configuration:</u> YES, LiFi technology deployment and related configurations.</p> <p><u>Change of validation environment:</u> YES, besides Wi-Fi and 5G NR, LiFi is also considered.</p>
	This runs within the	Mobility and traffic	<u>Enhanced models:</u> YES, LiFi and WiFi

	<p>infrastructure Stratum. It has the following Requirements</p> <p>CLARITY-INF-R3 5G-CLARITY system managed wireless resources shall include resources from two or more wireless access technologies, including Wi-Fi and LiFi technologies.</p>	<p>load management in Wi-Fi/LiFi integrated networks</p>	<p>channel models are integrated in the system model. Also, a smart network user mobility (SNUM) scheme is proposed to manage network user mobility.</p> <p><u>Refined setup configuration:</u> YES, SNUM scheme running in the SDN controller to manage user mobility. A regular LiFi network topology deployment in an indoor environment. Dynamic routing paths allocation for URLLC and eMBB services with e2e guarantees to their quality-of-service requirements.</p> <p><u>Change of validation environment:</u> YES, an SDN-enabled integrated network composed of 16 LiFi APs and WiFi AP is considered in the simulation. The SNUM scheme performance is compared to the standard user mobility approach in terms of network throughput and average delay.</p>
	<p>It addresses the infrastructure stratum requirements mentioned in CLARITY-INF-R3 in Table 2-1 of section 2.1. Furthermore, it is the key enabler of the application stratum KPI D21-5GC.KPI-8 and the network function D22-CLARITY-NFAS-R19 in Table 2-2 of section 2.2</p>	<p>Positioning Server</p>	<p><u>Enhanced Models:</u> YES, the access technologies have been modelled to facilitate the implementation of the positioning server. In particular, models of sub-6 GHz, mmWave, LiFi, and VLP/OCC technologies have been discussed and used as representatives of the actual technologies as they were not accessible in practice in the initial phase of the project.</p> <p><u>Refined setup configuration:</u> NO</p> <p><u>Change of validation environment:</u> NO</p>
<p>Feature #4: Synchronization and localization</p>	<p>This can be embedded as the key enabler of the localization server, residing in its core and utilizing the measurements from multiple technologies to perform localization (CLARITY-INF-R3 in Table 2-1 of section 2.1, and KPI D22-CLARITY-NFAS-R19 and D21-5GC.KPI-8 in Table 2-2 of section 2.2).</p>	<p>Joint synchronization and localization using multi wireless access technologies</p>	<p><u>Enhanced models:</u> YES, the previously used linear Bayesian recursive filter has been replaced by particle filter to enhance the performance</p> <p><u>Refined setup configuration:</u> NO.</p> <p><u>Change of validation environment:</u> YES, in the new environment the links can be LoS or NLoS and the new model can distinguish and drop the NLoS links.</p>

5 Mobility in public-private 5G industrial networks

We consider a 5G-CLARITY architecture hosted in virtualized compute systems supporting mobile UEs. This use case has been developed to evaluate mobility in PNI-NPNs as described in Sec 3.2.3. In this environment, users request the establishment of end-to-end connections between their applications hosted in their UEs and the Application Servers (AS) hosted in cloud compute facilities.

As shown in Figure 5-1, users located in the coverage of the public network have access through the virtualized gNB and to UPF to the AS hosted in the public cloud server. However, as they move towards the coverage area of the private network, the end-to-end connections need to be reconsidered and new tunnels need to be established interconnecting UEs with the local private AS. Therefore, a main issue that needs to be resolved is how network and compute resources are allocated when a user leaves the area of coverage of private network and enters the area served by public network (and the reverse).

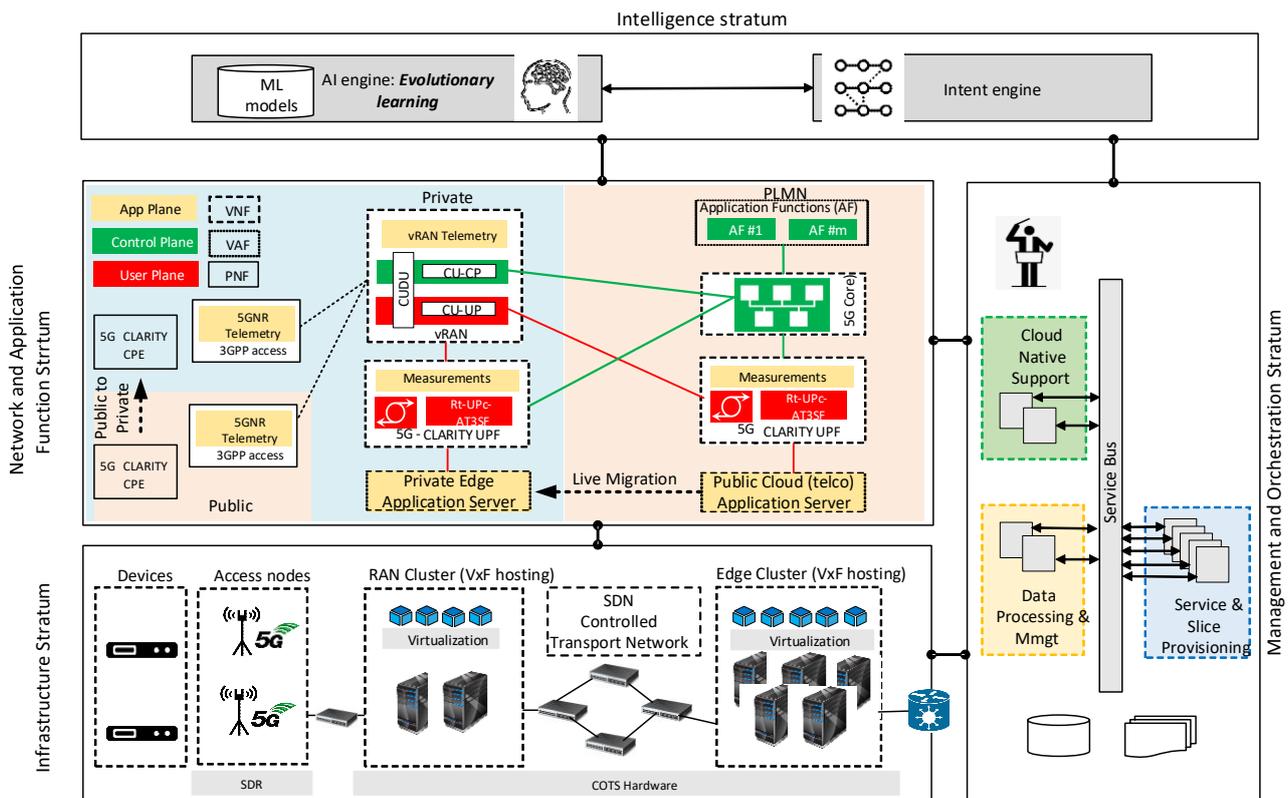


Figure 5-1 Joint user handover and VM migration problem to ensure service continuity in public/private 5G systems.

Another challenging aspect is associated with the reservation of sufficient resources across all elements of the 5G system (i.e., RAN, Core and transport network providing connectivity between these) to support mobility. As users move from one gNB to another, PDU sessions with the required QoS Flow Identifier should be established. To address this challenge, the present study considers the adoption of joint user handover and AS migration to ensure service continuity in MEC-assisted 5G environments supporting advanced transport network connectivity. As an example of the supported functionality, consider the scenario shown in Figure 5-1 where a mobile UE moves from a source public gNB to a target private gNB.

It is obvious that to ensure service continuity for MEC-assisted 5G services, a complex chain of several processes needs to be performed ensuring efficient allocation of connectivity between the UEs and the MEC nodes [21]. To successfully complete all these processes in a timely manner in order to reduce a service disruption, several issues need to be considered during the service provisioning phase including allocation of: (i) sufficient network resources for the establishment of the necessary connections between the 5G RAN

and the 5G core elements, (ii) sufficient computational resources to host not only the virtualized 5G functions (CU, DU, UPF, etc.) but also end user applications and ii availability of network resources for the interconnection of servers to perform live migration.

In response to this, a multi-stage framework is developed where at each stage of the optimization process a decision related to the placement of each virtualized function to the appropriate servers is taken. The objective of the proposed framework is to minimize the network cost for the provisioning of the services to the end users with the required KPIs. This cost function considers the weighted average of the utilization of the network and compute elements, as well as a penalty when service latency increases. The analysis is based on realistic statistics for network traffic and user mobility patterns, as well as actual measurements for the VM/AS migration process overheads and the traffic generated when new user sessions are created over the 5G network.

The previous study reported in [5G-CLARITY D2.3 \[2\]](#) considered as an unknown variable only the UPF location where PDU sessions are terminated. However, this model extends our previous work in several ways:

1. It solves jointly the end-to-end 5G service provisioning problem and identifies the optimal location where gNB-DU, gNB-CU, UPF and DN elements can be placed under mobility considerations,
2. It introduces the option of migration for the AS hosted in the DNs. This technique is employed to reduce the distance of the UEs from the AS and, therefore, the associated end-to-end latencies. Migration overheads for the AS are evaluated using extensive experimental results,

The analysis uses realistic statistics for network traffic based on actual UEs.

5.1 End-to-end modelling

We evaluate the scenario assuming a network interconnecting a variety of general and specific purpose compute/storage and network elements applying the concepts of hardware programmability and network softwarisation to support a variety of 5G-RAN deployment options considered in [5G-CLARITY](#). To achieve this, the transport network can provide the necessary interfaces to enable: (i) disaggregation of Base Station nodes and, ii) CUPS. Through Base Station disaggregation, functions building up the 5G NR protocol stack (including RU, DU, CU) can be physically separated and hosted at different locations. Connectivity between the different protocols can be provided over the optical transport through interfaces such as the O-RAN FH interconnecting the RU with the DU and the F1 interface interconnecting DU with CU. Separation of the CU Control Plane (CU-CP) and User Plane (CU-UP) enables flexibility in network deployment and operation, as well as cost efficient traffic management.

The user plane has the role to provide connectivity between the different network elements. This includes connectivity of the UE and the NG-RAN over the radio access technology, connectivity of the AN to the UPF over the N3 interface, connectivity between UPFs with different roles via the N9 interface, and finally connectivity from the UPF towards the external DN over the N6 interface. User plane data that travel over the N3 and N6 interfaces are carried in GPRS Tunnelling Protocol User plane (GTP-U) tunnels.

The establishment of end-to-end connectivity in 5G systems requires reservation of specific resources to set up appropriate data radio bearer (DRB) tunnels between the UE and the gNBs, and N3 GTP-U tunnels between the gNB and the UPFs. In addition to the PDU sessions, for services requiring access to a specific data network (hosting targeted MEC server) N6 tunnels should be established between the UPFs and the MEC and maintained for the whole duration of the connection of the mobile user. In addition to networking resources, compute resources need to be also allocated for the 5G system elements. Therefore, a critical decision that needs to be taken by the SMF is when and over which elements these sessions should be established to ensure service continuity.

To mathematically describe this concept, we consider a 5G network modeled as an undirected graph $G(\mathcal{N}, \mathcal{E})$ where \mathcal{N} is the set of nodes and \mathcal{E} the set of links. This 5G system comprises both RAN and core elements. The 5G-RAN segment comprises a set $\mathcal{R}, \mathcal{R} \subseteq \mathcal{N}$, of R gNBs used to provide connectivity services for a set \mathcal{U} of U mobile users. For gNBs, the concept of functional split is adopted according to which the RUs, DUs and CUs are separated. The processing requirements of each gNB r has been calculated using OpenAirInterface (OAI) [22]. As OAI supports split option 7.2., the computational requirements denoted as $C_{rti}, i = DU, CU$ (measured in giga operations per second – GOPS) can be determined for wireless access traffic load, element $i, i: \{CU, DU\}$ supporting RU r at time t, w_{rt} of RU r at time t .

To extract, CU and DU computational/network requirements OAI has been deployed in a containerized environment running as VNFs in commercial off-the-shelf MEC servers. Beside 5G-RAN elements, MEC servers also host core elements, i.e., UPF, which are necessary for the establishment of the appropriate services and ultimately on the provisioning of the end-user services.

To establish a PDU Session, the UE must be registered with the ‘access and mobility management function’ (AMF). PDU session establishment also involves communication between the UE the AMF and the SMF. At the end of the process, the SMF selects a suitable available UPF and the DN that will be used to support the requested service. Now let $\mathcal{N}_f \subseteq \mathcal{N}$ be the set of UPF nodes and \mathcal{S} the set of servers. A UE connection is then established after the setup of the resources to interconnect the UE with the RAN, the RAN with the UPF and the UPF with the DN. To ensure service continuity, necessary resources need to be reversed across the travelling paths of all users. To achieve this, accurate modeling of mobility patterns of the UEs is necessary.

To model user mobility, we assume that for each user u , there is a specific origin/destination (O/D) pair defined through the vector $z_u = \{z_u^{(o)}, z_u^{(d)}\}$ with $z_u^{(o)}$ being the origin and $z_u^{(d)}$ the destination locations, e.g. coordinates of UEs. The departure/arrival time for the O/D pair z_u is known in advance for every time period $t \in \mathcal{T}$, with \mathcal{T} being the horizon (or number of times) where the control policy is applied. Each period t has a duration equal to \hat{t} . As regards to the timing events, we assume that the arrival of user at the origin location $z_u^{(o)}$ occurs at time instant $\tau_u \in [t, t + \hat{t}]$, $t \in \mathcal{T}$. Based on the O/D pair z_u and using standard route planner software, the traveling path p_u for every user $u \in \mathcal{U}$ can be determined. Wireless connectivity across traveling path p_u is provided through a set $\mathcal{R}_{p_u} \subseteq \mathcal{R}$ of RUs. Each RU $r \in \mathcal{R}_{p_u}$ is responsible for providing network connectivity for a segment of the route p_u .

As mobile users travel across p_u , the average time spent at every segment of the route can be readily determined based on history road traffic statistics. To model this parameter, Δ_{urt} is introduced to capture the residence time of user $u \in \mathcal{U}$ within the path segment covered by RU $r \in \mathcal{R}_{p_u}$ during time period t . The corresponding normalized residence time (fraction of time spent is the area covered by RU r over the duration the time period \hat{t}) is given by $\hat{\Delta}_{urt} = \Delta_{urt}/\hat{t}$.

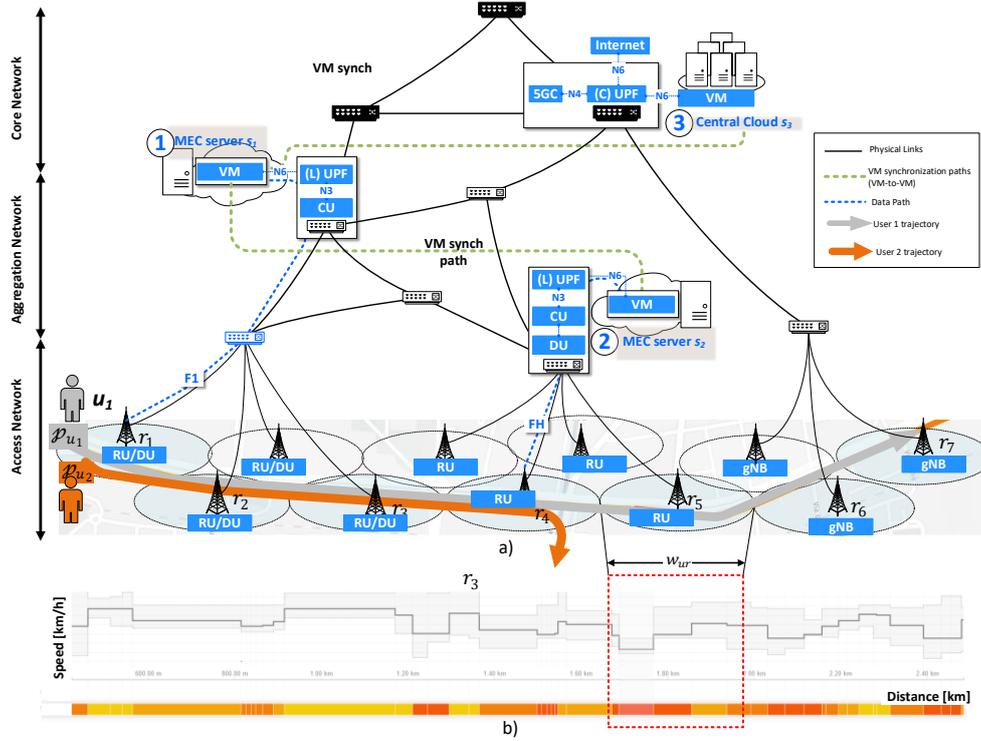


Figure 5-2 5G network with mobility support

Based on Δ_{urt} and τ_u , the time periods where each user is supported by the corresponding RUs in its trip can be determined. This is modeled through the binary coefficient ℓ_{urt} taking value equal to 1 if user u is located in the segment of the route covered by RU r at time t , 0 otherwise.

In their turn, mobile users should be interconnected with the remote AS's. ASs can be hosted at a set S of S servers that can be placed either on-premises (within 5G-CLARITY admin domain) or at the telco edge cloud node (in the public network, and thus out of 5G-CLARITY admin domain). Connectivity between RUs and compute resources is provided through an optical transport network [23]. Based on the 5G deployment option adopted, this transport network can be used to support:

- The requirement of the FH protocol for the RU-DU interconnection
- the requirements of the F1 interface for DU-CU connectivity
- the requirements of the N3, N6 and N9 interfaces for CU – UPF, UPF-MEC, and UPF-UPF, respectively, connectivity.

At this point it should be mentioned that to provide services with QoS guarantees, specific KPIs (measured in terms of network throughput, processing and transport network delay) across the entire traveling path p_u should be satisfied. The transport/network, computational/processing and service delay requirement of services requested from user $u \in \mathcal{U}$ are denoted as h_u , π_u and d_u , respectively.

Based on the above-mentioned modeling assumptions, an optimization problem is formulated that tries to minimize a cost function of the form over a time period T

$$\min_{x_0 \in \mathcal{X}_0} f_0 + \sum_{t=t_1}^T \mathbb{E} \left[\inf_{x_t \in \mathcal{X}_t} f_t \right] \tag{5-1}$$

where

$$f_t = \sum_{s \in S} \mathcal{N}_s (C_{st} + C'_{st}) + \sum_{e \in E} \mathcal{L}_e (C_{et} + C'_{et}) \quad (5-2)$$

subject to a set of service chaining, routing, network capacity and processing constraints. In (5-2), C_{st} denotes the capacity of server s at time period t for the operation of the 5G system and C'_{st} is the additional capacity needed to support AS migration. Similarly, C_{et} is network capacity of a link e at time period t for the operation of the 5G system and C'_{et} is additional capacity needed to support Live AS migration.

5.2 Scenario Details

5.2.1 Scenario setup

To solve the problem of joint VM migration and mobility management in 5G systems, a 5G testbed has been deployed over a virtualized cloud environment allowing the accurate estimation of network and compute resources consumed during the establishment of new UE sessions. These measurements are coupled with actual network traffic and user mobility statistics collected over an operational mobile network system. To quantify this cost, the 5G standalone version of the OAI [22] has been deployed in the private cloud infrastructure shown in Figure 5-3.

In this environment, we consider a private 5G platform (shown in the blue box) comprising a set of softwarized RAN and core elements. These are hosted at physically separated servers. An AS has been also deployed at a different server. 5G UEs based on the Rel.16 RM500Q sub-6GHz module have been also considered requesting access to the AS. The RUs has been implemented using N310 and B210 USRPs. All compute nodes are physically interconnected with an Aruba JL323A. A public 5G SA platform has been also deployed having some basic 5G functionality (for the experimental purposes we assume only the presence of AMF/SMF/UPF). These elements are hosted in physically separated machines through an Aruba JL255A. Both switches are interconnected using a 10G point to point link. Network/compute resource utilization metrics have been collected through Prometheus and visualized using Grafana.

Typical samples of the traffic generated by a UE, reaching the UPF in VM4 and, finally terminated at the AS in VM 3 is shown in Figure 5-4. The patterns of the periodic signalling traffic transmitted by the AMF and the SMF is shown in Figure 5-5 a) and Figure 5-5 b) respectively.

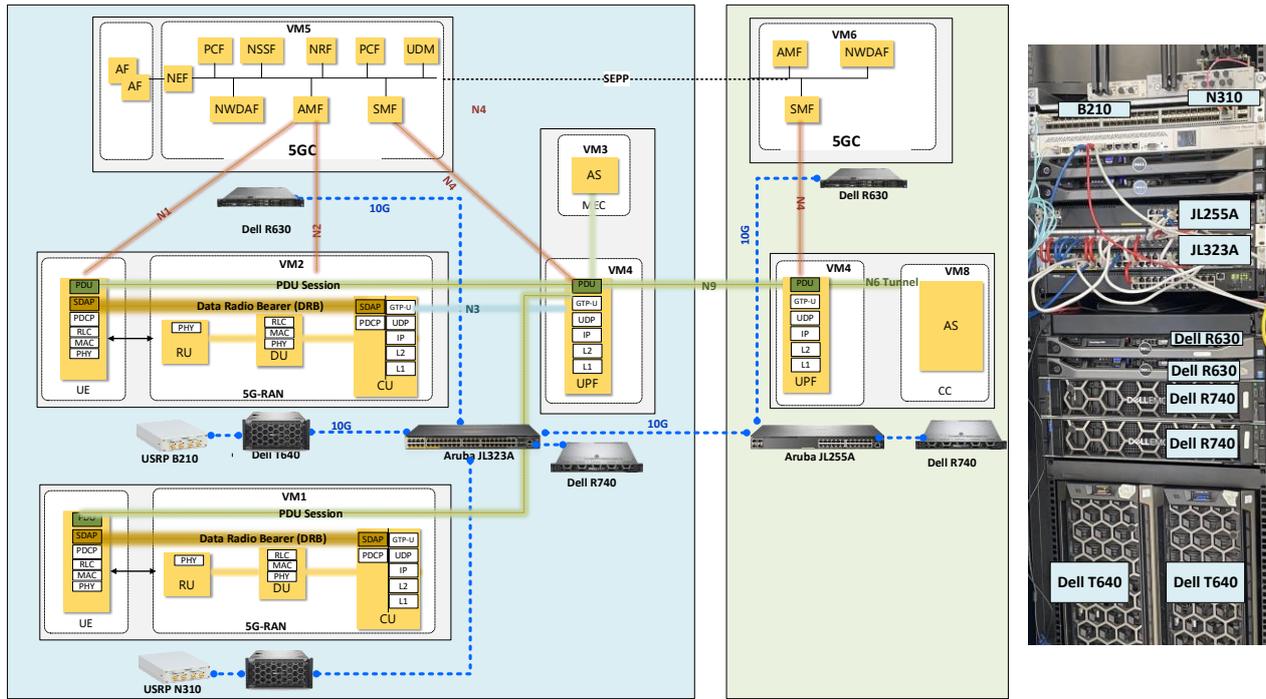
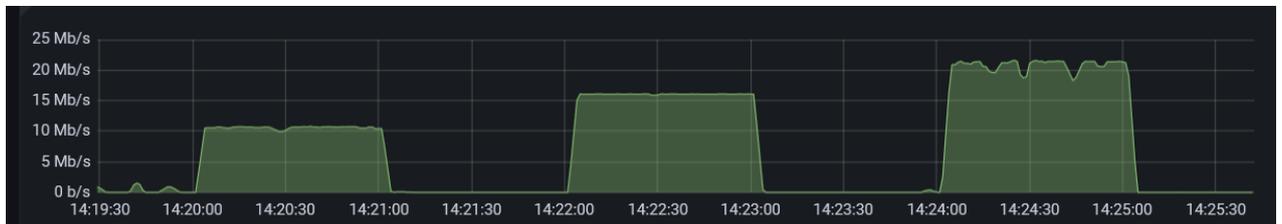
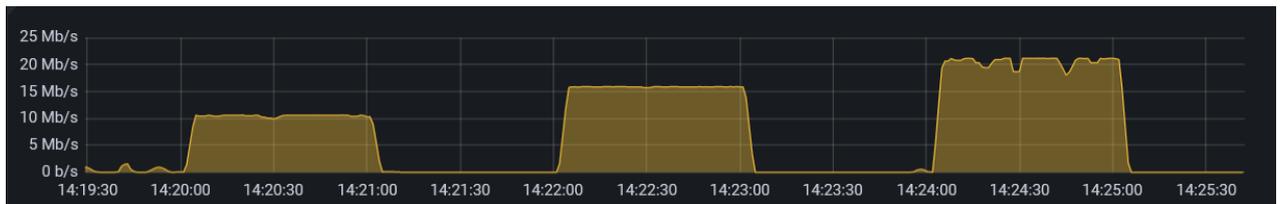


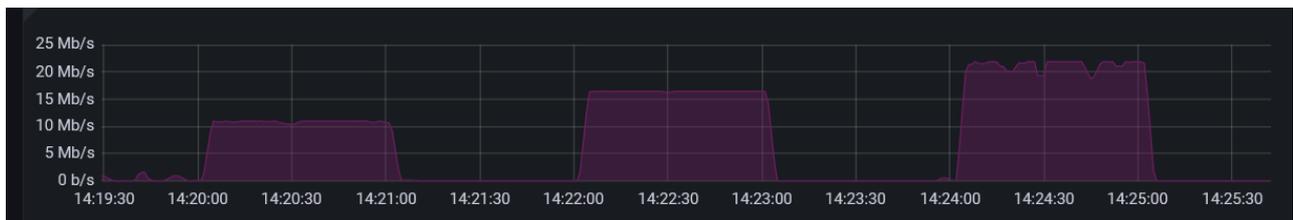
Figure 5-3 Connectivity diagram (left) and experimental infrastructure used to host the 5G SA platform



a)



b)



c)

Figure 5-4 Traffic generated at a UE (a) traversing the UPF (b) and terminated at the AS (c).

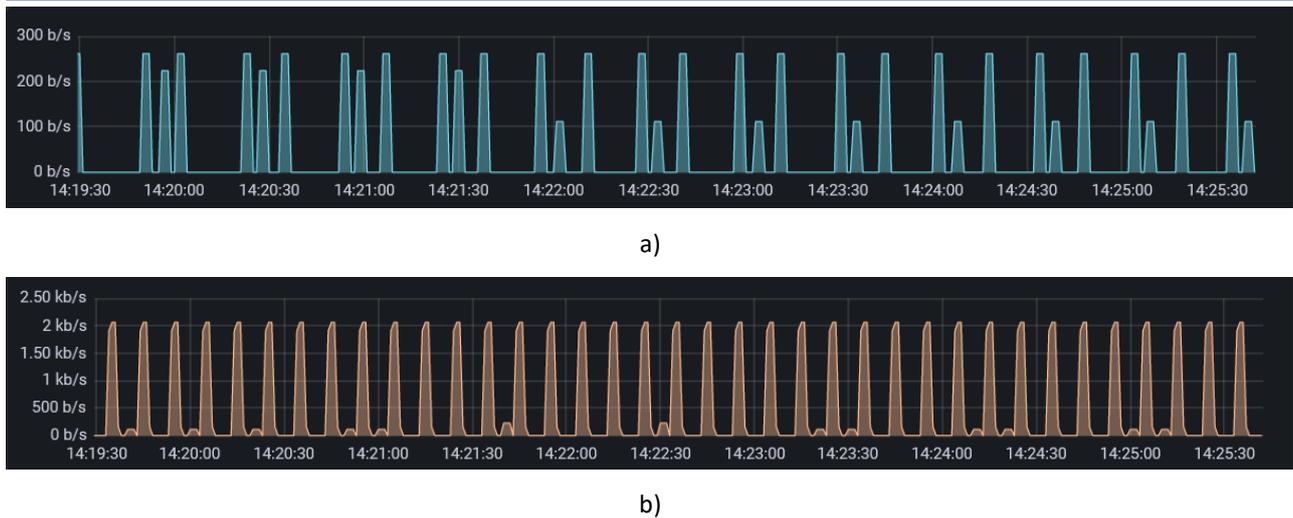


Figure 5-5 Signalling traffic generated by a) the AMF and b) the SMF.

Figure 5-6 shows CPU, memory, disk and network traffic measurements reaching the CU under different configurations of the gNB.

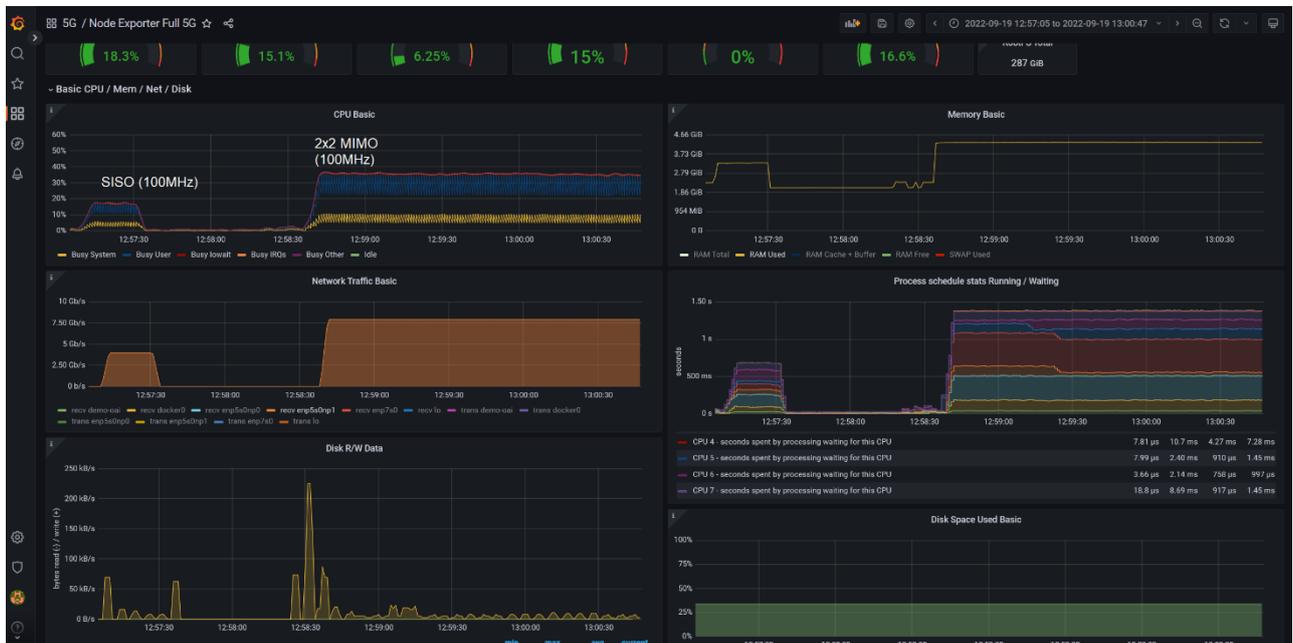


Figure 5-6 CU measurements under different configurations (top-left): CPU utilization for SISO and 2x2 MIMO configuration, (top-right): Memory utilization, (middle-left): fronthaul traffic, (middle-right): CU processing distributed at different CPUs, (bottom left): disk read/write, (bottom -right): disk usage

The power consumption of the physical machines hosting the gNB and core elements is shown in Figure 5-7. At this point it should be noted that both entities are hosted in identical machines. It is observed that due to the match higher processing requirements of the gNB, more than double power consumption compared to server hosting the core segment has been observed. This is expected as x86 servers are used to perform DU processing functionalities need specialized configuration that increases overall power consumption.



Figure 5-7 Power consumption of the physical machines used to host the gNB (top, green) and the core (bottom, yellow)

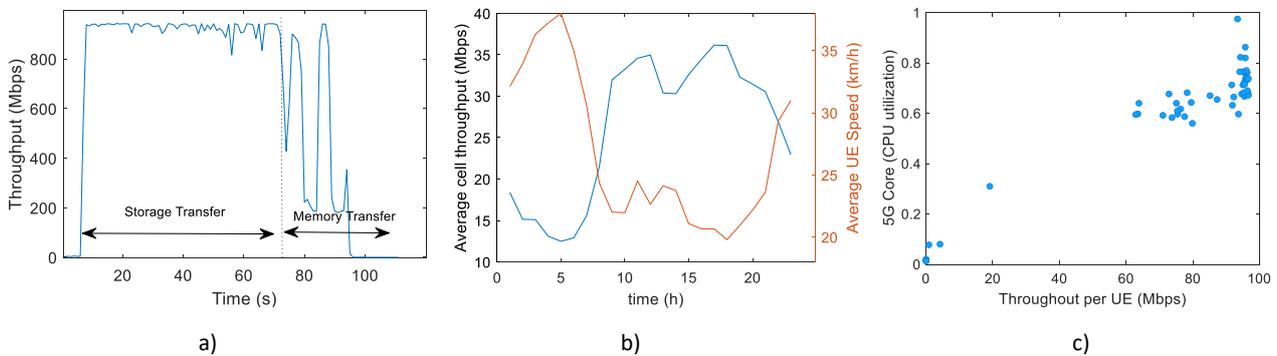


Figure 5-8 a) Time series showing the traffic generated during VM migration from a source to a target VM. b) Correlation between background mobile network traffic per gNB and speed per UE, c) Impact of average UE throughput on 5GC computational resources [23]

Figure 5-8 shows an example of the network traffic generated during the migration from a source MEC server to a target MEC server, measured in the testbed. In this example, the AS has been configured to hosts a 4K streaming video server. During this live service migration process, the memory and disk state of the VM is transferred from the source host (VM8) to the destination host (VM3). Storage transfer is performed through a steady throughput, while memory transfer is done through multiple synchronization iterations. As mentioned above, a prerequisite for the success of the AS migration process is the availability of network and compute resources during the storage and memory copy phase. The availability of these resources depends on the area where UEs move and the background network traffic. Higher background network traffic is observed in densely populated areas (e.g., city centers) where the speed of the mobile UEs is also lower. The interrelation between the average mobile traffic per gNB and the average speed per UE within the area covered by the gNB is shown in Figure 5-8b. The relevant traces have been captured from an operational mobile environment, whereas average speed statistics have been collected from GPS trackers. The impact of the mobile network traffic on CPU utilization of the virtualized 5G platform is shown in Figure 5-8 c. As expected, the average traffic per UE increases the CPU utilization of the platform used to host the virtualized 5G system. It is concluded that possible migrations associated with a user moving from a gNB covering a sparsely populated region to a densely populated region should be treated carefully as service

disruptions may occur.

Components	Background	Extensions in CLARITY	Justification
5G-RAN	OAI 5G-RAN	Deployment in containerized environment monitored using node exporters and orchestrated using MANO	Part of the end-to-end service chain of the scenario under evaluation
5GC	OAI 5G-Core	Deployment in containerized environment monitored using node exporters and orchestrated using MANO	Part of the end-to-end service chain of the scenario under evaluation
UPF	OAI UPF	Deployment in containerized environment monitored using node exporters and orchestrated using MANO Interfaces with SDN to orchestrate transport network operation	Part of the end-to-end service chain of the scenario under evaluation
AF	New developed application	AI-assisted operation of AF component dynamically adjusting operations based on the status of the 5G platform	Part of the end-to-end service development
Transport Network Controller	OpenDayLight (ODL)	Extensions of ODL to interact with the UPF and perform statistical traffic multiplexing of the SDN controlled switches	Application of queuing policy at the transport network for service differentiation
Smart NIC (Netronome P4)	OAI UPF	Acceleration to perform deep packet processing	Reduce delay at UPF
AI Engine	ML platform (ML repository hosting trained ML models)	Develop ML models for performing profiling of the 5G elements. Interaction with the 5G data monitoring system to gather datasets and train the corresponding models	Extract the cost models for the multi-agent reinforcement learning model
Orchestration	OSM (Open Source MANO)	Extensions to interact with the 5GC	Apply policies to deploy the full 5G platform

Parameter	Value
gNB	100MHz bandwidth (273 RBs), 400Mps measured throughput per cell (TCP connection with Cubic)
Transport Network Nodes	4 x SDN Switches with 10G capacity for the access and 40G for the aggregation part

5.2.2 Scenario validation

A comparative analysis of different 5G network deployment options in terms of network cost (total capacity in Gbps), compute cost (number of vCores) and monetary costs is shown in Figure 5-10 a), Figure 5-10 b) and Figure 5-10 c), respectively. For the comparisons we have considered four different optimization objectives:

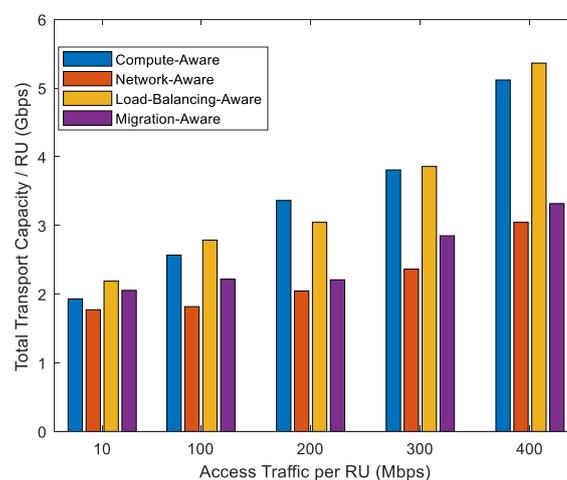
- The 'Network-Aware' policy that considers in the optimization process only the network resources required for the system to operate.
- The 'Compute-Aware' policy that considers in the optimization process only the compute resources required for the system to operate.
- The 'Balancing-Aware' policy that tries to uniformly allocate all compute tasks to the servers. This policy the average utilization per server can be reduced.
- The 'Migration-Aware' that tries to minimize the VM migration overhead under mobility scenarios.

For the monetary costs the values reported in [13] have been considered for the compute and network resources. For the network resources we have assumed a price of 0.33€/hour/Gbps⁵ while for the compute we have assumed an on demand hourly price of €0.11 per vCore. The results have been extracted under different mobility scenarios. The MEC has been configured on a single 1 thread per core while central cloud supports two-threads per core.

Under low mobility scenario, we observe from Figure 5-10 a) that the policy that minimizes network resources has the least network cost as in this case all virtualized entities are placed close to the RUs. The relevant results have been extracted assuming that the end-to-end delay (measured in terms of number of hops) from the UE to the AS is 2. We also observe that the policy that minimizes migrations has higher network cost as in this case the AS is placed in more centralized locations to reduce the number where the migration process is triggered. High network costs are also observed for the policies that minimize the compute costs and balance traffic across servers. By these policies it is tried to minimize the number of active servers in the 5G system satisfying at the same time KPI requirements. This results in an overall higher network resource requirements as longer end-to-end function chains are established.

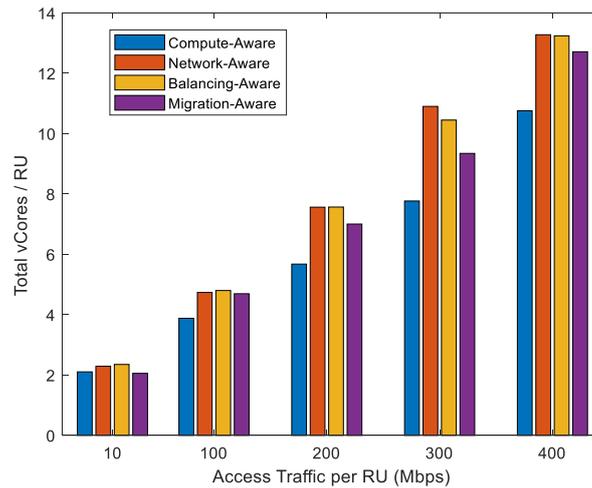
Figure 5-10 b) shows the compute cost under different optimization strategies. The compute-aware policy achieves the best performance as it optimally allocates functions to the appropriate compute nodes increasing centralization gains (i.e., a single UPF hosted in a more centralized location can support multiple RUs). As expected, the compute cost for the other placement strategies is higher as all entities of the service chain are placed close to the RUs requiring the instantiation of additional 5G core and RAN elements. A good compromise is observed in the migration-aware strategy requiring less compute resources to operate compared to the network or balancing aware policy.

The combined network and compute cost in monetary values per RU is shown in in Figure 5-10 c). Given that the network connectivity costs are higher compared to compute costs, the most efficient deployment option is to place all elements close to the edge. It is also observed that under low mobility, the migration-aware strategy has almost the same deployment costs with the network -aware policy as the higher network costs in the ‘migration-aware’ policy are counterbalanced by the savings in compute costs.

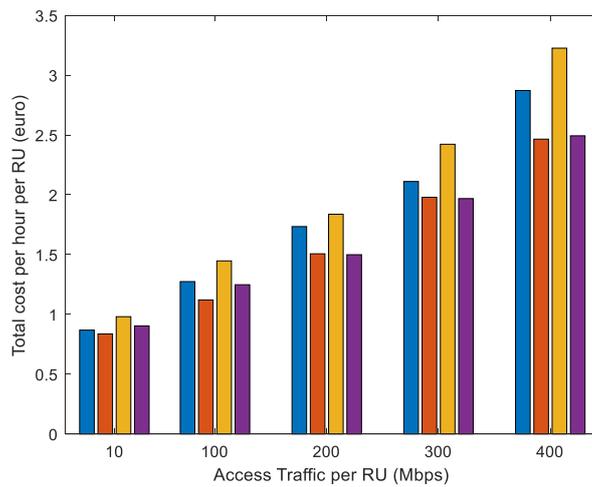


a)

⁵ <https://aws.amazon.com/directconnect/pricing/>



b)



c)

Figure 5-9 Comparative analysis of different 5G network deployment strategies in terms of a) network, b) compute and c) total cost under and c) total costs under low mobility

The performance of the different deployment strategies under high mobility is shown in Figure 5-10. Under high mobility the overall compute and network costs increase for all strategies. Under high mobility additional resources need to be allocated across all segments of the system to ensure seamless handovers. This includes additional resources for the UPFs to handle traffic steering, as well as network and compute capacity to support VM migration tasks. However, the relevant cost increase for the ‘migration-aware’ strategy is smaller. By predicting UEs trajectories, the migration-aware policy places 5G elements to the appropriate positions and size their capacities according reducing relevant overheads. The relevant results for the network, compute and total costs are shown in in Figure 5-10a), Figure 5-10 b) and in Figure 5-10 c), respectively. Therefore, under high mobility we observe that the policy that minimizes network cost places AS close to the UEs. In this case, the migration process is triggered frequently as the AS will follow mobility pattern of the users: every time a user moves to a new gNB, the AS will be placed to a closely located server. The migration overhead is also very high and independent on the end-to-end service delay. On the other hand, the policy that minimizes the number of migrations results in high overheads when end-to-end delays are strict. By relaxing these constraints, the scheme predicts the future position of the UEs and optimally places the ASs to minimize the associated migration overhead.

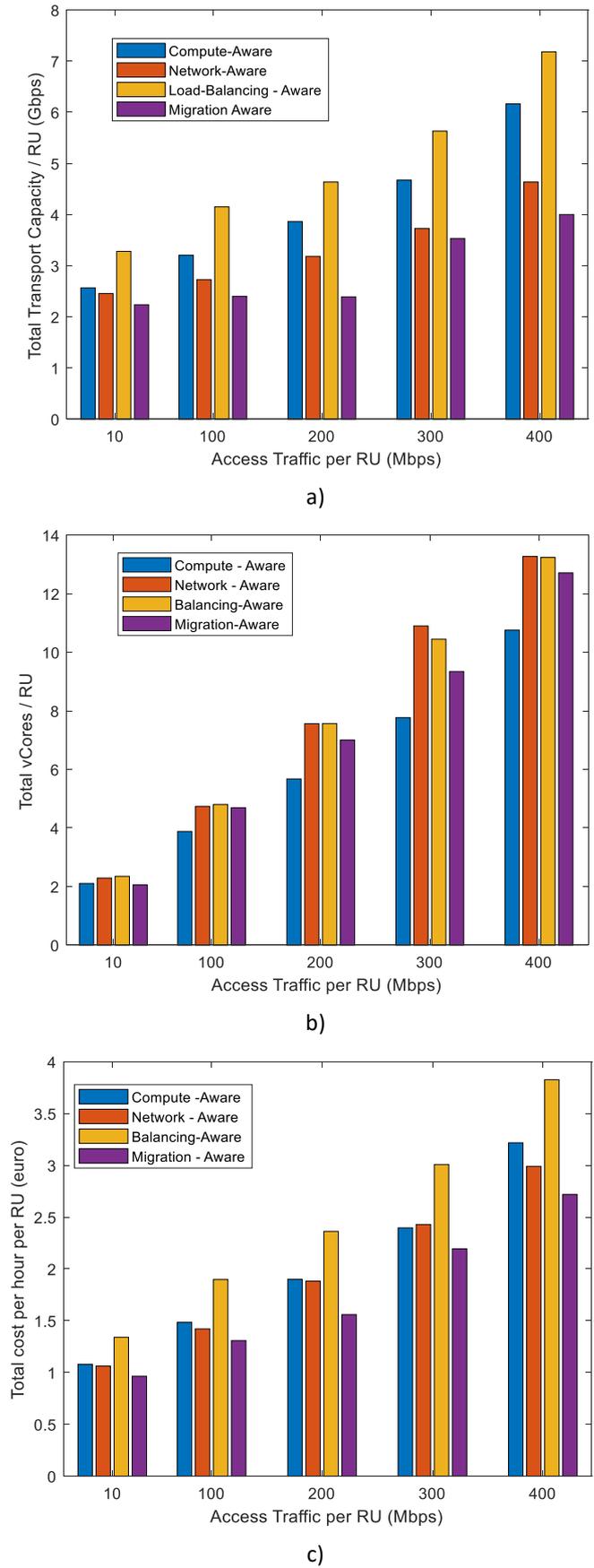


Figure 5-10 Comparative analysis of different 5G network deployment strategies in terms of a) network, b) compute and c) total costs under high mobility

6 Multi-WAT capabilities

The deployment of multiple private wireless access networks is an opportunity for service providers to offer services to multiple user profiles and support diverse enterprise and indoor use cases. The multi-WAT empowers heterogeneous RANs with advanced wireless connectivity capabilities, which can support multiple network topology deployment scenarios for provisioning reliable and real-time Industry 4.0 services. However, Multi-WAT also requires new user mobility models that can leverage the heterogeneity of wireless technologies to provide comprehensive insights regarding the performance of user mobility and services provisioning. To assess the multi-WAT capabilities of 5G-CLARITY system architecture, in D2.3 [2] we proposed the evaluation of two different scenarios in which multi-WAT was used for different purposes in private networks (see Section 5 of D2.3 for the details of these scenarios). In this deliverable, we further extend these scenarios with more network components and add more performance evaluation scenarios.

- “Scenario 1: Multi-WAT offloading in industry 4.0”: In 5G-CLARITY D2.3 [2], the multi-WAT benefits were evaluated in terms of the data rate offered in a RAN consisting of 5G-NR and Wi-Fi technologies providing wireless access to eMBB and URLLC services. It was analyzed how eMBB traffic offloading from 5G NR to Wi-Fi can result in an increment of data rate for eMBB traffic, and thus, an increase of the capacity of URLLC services served by 5G NR. The offloading mechanism was basically based on the SINR level that eMBB users received from both technologies. So that, eMBB users with less 5G SINR level were offloaded to the nearest Wi-Fi access point. The study was carried out in an industrial scenario that tried to resemble a Bosch factor floor. In this deliverable, we extend the industrial network scenario by adding the LiFi access network to the multi-WAT 5G NR and Wi-Fi access networks.
- “Scenario 2: Mobility and traffic offload management in Wi-Fi/LiFi integrated networks”: To address the issue of providing QoS guarantees for UEs in the SDN-enabled LiFi/Wi-Fi integrated network in 5G-CLARITY D2.3 [2], a virtual connection tree scheme is proposed to support the QoS guarantees for UEs. In this scheme, adjacent cells are grouped into a cell cluster in a static fashion. Upon a session request admission control, the scheme preestablishes a session between a root switch and each AP in the cell cluster. This may result in an unnecessary resource overloading that may underutilize the network resources and cause severe congestion. In this deliverable, a smart network user mobility (SNUM) scheme is proposed to support seamless user mobility across the network. It enables the SDN controller to view the LiFi and Wi-Fi networks as a single cluster, in which the UEs are offered their requested services according to the available network and their context information.
- “Scenario 3: Position server for Multi-WAT”: As mentioned in 5G-CLARITY D2.3 [2], the positioning server relies on the measurements coming from different technologies to provide a precise estimation of the AGV position. Here, not having access to the actual technologies, in the initial phase of the project, to facilitate the implementation of the architecture and to be able to simulate the overall system in order to optimize it, we provide a model of each technology. In particular, we provide the models for sub-6 GHz, mmWave, LiFi, and VLP/OCC technologies. The real positioning system will be employed in the final demonstration in the Bosch factory.

Figure 6-1 depicts the architectural components of the 5G-CLARITY system architecture involved in the evaluation scenarios, which are highlighted in purple boxes.

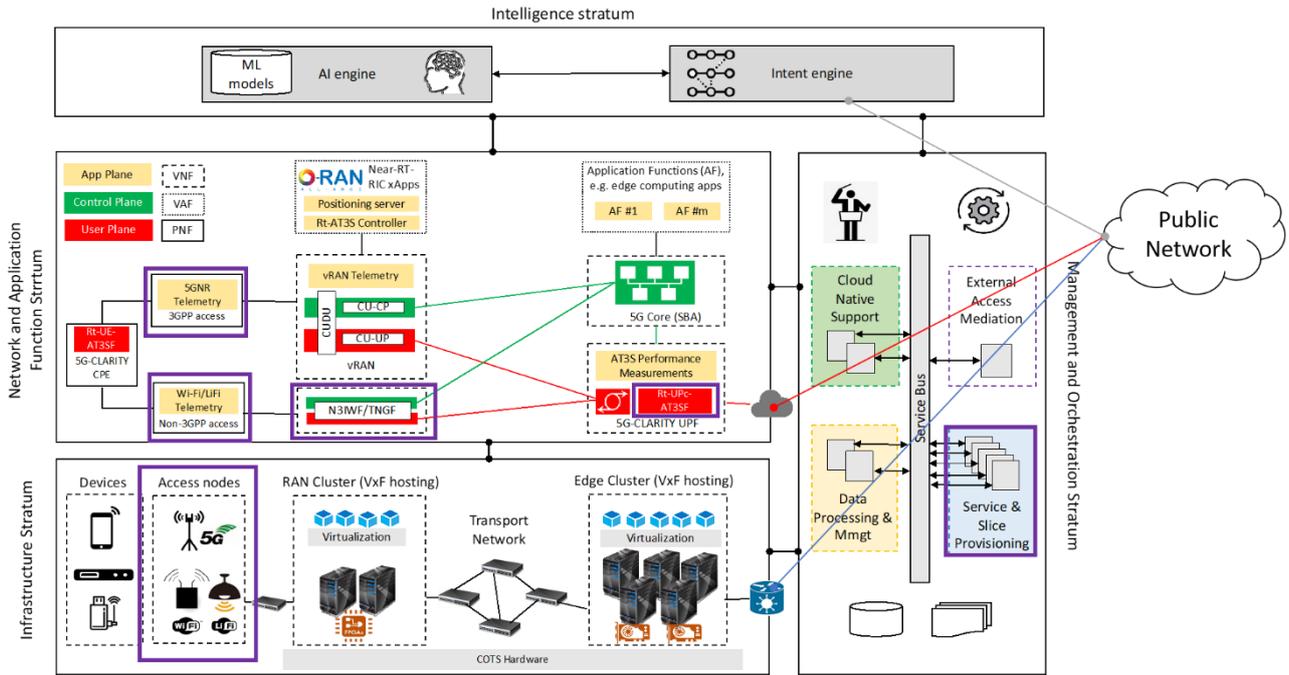


Figure 6-1 Components under evaluation of 5G-CLARITY system architecture for assessing multi-WAT capabilities

6.1 End-to-end modelling tools

Before discussing the proposed SNUM scheme, we introduce some key equations that can be used in developing probabilistic analysis regarding the user residence time, handoff rate and failure.

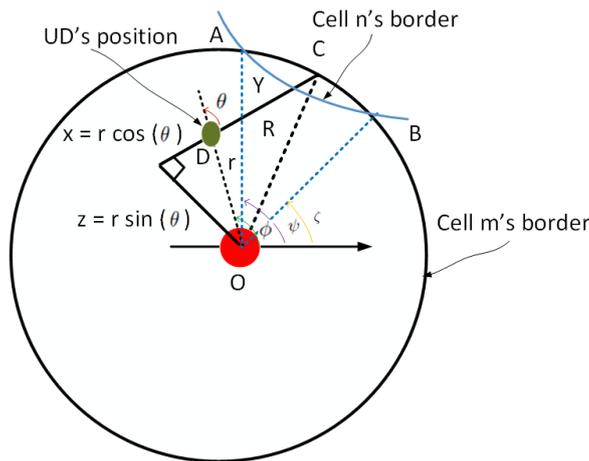


Figure 6-2 Illustration of distance from point D, marking the session request, in a LiFi or Wi-Fi cell to point C, marking the UE exit on the cell

The probability distribution function (pdf) of user distance in a cell, where a session request is originated, can be obtained by using the standard methods [24] as follows:

$$f_Y(y) = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2 - \left(\frac{y}{2}\right)^2}, & \text{for } 0 \leq y \leq 2R \\ 0, & \text{O.W} \end{cases} \tag{6-1}$$

The residence time random variable is defined as the time (duration) that a UE resides in the cell in which the session is originated. So, the pdf of user residence time can be derived by dividing the random user

distance variable by the random user speed variable in the cell. When the UE is about to arrive at the exit point C on the cell shown in Figure 6-2, the handover time starts to be counted in the next cell where it should be handed off. So, the handover time is defined as the time the UE resides in the cell to which their session is handed off. However, during this time, new session requests can be blocked in the next cell as well as the handed off sessions request may be terminated or dropped. The handoff failure can be studied by investigating the cumulative distribution function (CDF) of the user distance in the next cell. If we assume that the UE moves in any y direction with equal probability in the cell shown in Figure 6-2, then the CDF of $F_Y(y)$ can be expressed as follows [24]:

$$F_Y(y) = \begin{cases} 0, & \text{for } y \leq 0 \\ 1 - \frac{2}{\pi} \arccos\left(\frac{y}{2R}\right), & \text{for } 0 \leq y \leq 2R \\ 1, & \text{for } y \geq 2R \end{cases} \quad (6-2)$$

While these provide further analytical analysis to enrich the mathematical framework developed in 5G-CLARITY D2.3 [2], the SNUM scheme is developed to leverage the network and user information to mitigate their dynamics in providing seamless user mobility and support 5G URLLC and eMBB services.

6.1.1 Multi-WAT offloading in industry 4.0

The goal of the evaluation carried out in 5G-CLARITY D2.3 [2] was the assessment of the benefits of having 5G NR and Wi-Fi as wireless access technologies in an industrial scenario, in which eMBB and URLLC slices were deployed (see 5G-CLARITY D2.3, Section 5.5 [2]). In this deliverable we propose the extension of the evaluation of 5G-CLARITY D2.3 by including LiFi as an additional wireless access technology. LiFi is becoming an appealing radio access technology in private sites due to its optical radiation characteristics. It provides reliable and secure data transmissions as they do not interfere with other electromagnetic waves. These features make from LiFi a good candidate wireless technology for network scenarios with traffic heterogeneity to serve those data-hungry applications. Thus, in the evaluation carried out in this deliverable we assess the impact of having LiFi technology, in addition to Wi-Fi and 5G NR, in the same industrial private 5G network as the one considered in D2.3. In this regard, we have included the LiFi communication channel model to the industrial RAN simulator that was developed for the evaluation carried out in D2.3. The description of this specific model is included below.

Optical signals travelling through a visible light communication channel suffer from some effects that may lead to different time delays for the arriving signals at the receiver. The primary channel component consists of the line of sight (LoS) transmission, which is the straight line between the LiFi access point (AP) and the user's photodiode (PD). Additionally, other components that may add to the signal through the LoS path is non-line of sight (NLoS) paths. Even though the signal power from NLoS paths is not negligible in certain scenarios [25], in this evaluation scenario only the LoS path is going to be considered due to the specific characteristics of the industrial scenario layout under study.

In order to estimate the LiFi communication channel capacity, we use the Shannon formula as it has been proven to be a good upper bound approximation [25]:

$$c = \frac{1}{2} W \log_2 \left(1 + \frac{\sigma_0^2 \|H_0\|^2}{\sum_{i \in I} \sigma_i^2 \|H_i\|^2 + WN_0} \right) \quad (6-3)$$

In the expression above W refers to the modulation bandwidth, σ_i^2 refers to the electrical signal variance from the AP i , $\|H_i\|^2$ represents the channel gain from AP i to the specific user, I refers to the set of APs using the same transmission resource, and N_0 is the double-sided noise power spectral density.

Additionally, it shall be noted that the rational term of the capacity expression stands for the signal-to-

interference-plus-noise-ratio (SINR).

Specifically, the SINR experienced by the user u having AP i as the transmitter access point is denoted by $\Upsilon^{i,u}$, and its expression is the following [26]:

$$\Upsilon^{i,u} = \frac{(R_{pd} H_{LiFi}^{i,u} P_{opt} / \zeta)^2}{N_{LiFi} B_{LiFi}} \quad (6-4)$$

Where:

- R_{pd} is the detector responsivity.
- $H_{LiFi}^{i,u}$ represents the LiFi channel gain.
- P_{opt} is the transmitted optical signal power.
- ζ represents the ratio of P_{opt} to the optical signal power.
- N_{LiFi} stands for the Power Spectral Density (PSD) of noise at the receiver.
- B_{LiFi} denotes the LiFi AP bandwidth.

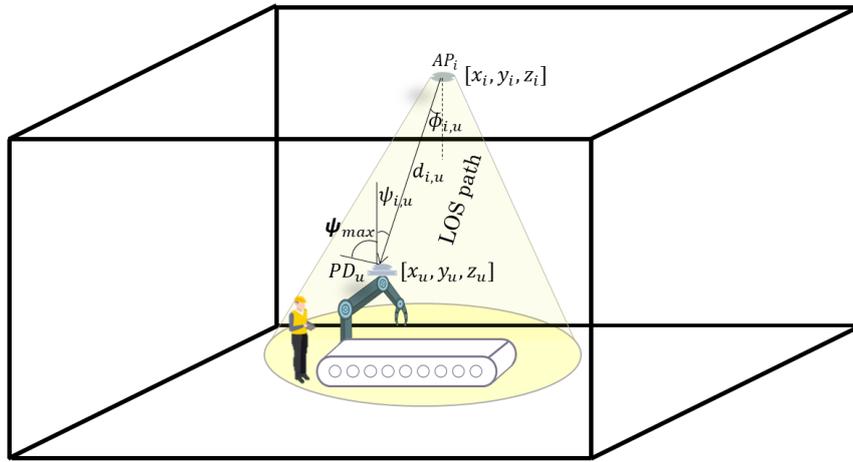


Figure 6-3 LoS path of the LiFi channel considered

In the considered scenario, the LiFi channel gain between AP i and user u ($H_{LiFi}^{i,u}$) corresponds to the channel gain of the LoS path ($H_{LoS}^{i,u}$):

$$H_{LiFi}^{i,u} = H_{LoS}^{i,u} \quad (6-5)$$

Based on the works in [26], [27], [28], the basic optical communication system under consideration comprises on a transmitter fixed on the ceiling of the factory, which is directed perpendicularly downward towards the floor. The receiver is a PD located in the user and directed upward towards the transmitter. Figure 6-3 illustrates the LoS path of the LiFi channel considered.

Let i stand for the AP and u denote the user. The channel gain of the LoS path, represented by $H_{LoS}^{i,u}$, is given by the following expression:

$$H_{LoS}^{i,u} = \frac{(m+1)A_{PD}}{2\pi d_{i,u}^2} \cos^m(\phi_{i,u}) g_f g_c(\psi_{i,u}) \cos(\psi_{i,u}) \quad (6-6)$$

Where:

- $m = -\ln(2)/\ln(\cos\phi_{1/2})$ denotes the Lambertian emission order, being $\phi_{1/2}$ the angle of half intensity.

- $\phi_{i,u}$ represents the irradiance angle.
- $\psi_{i,u}$ denotes the incidence angle.
- A_{PD} stands for the physical area of the PD.
- $d_{i,u}$ represents the Euclidean distance between the transmitter AP i and receiver's PD u .
- g_f is the gain of the optical filter.
- $g_c(\psi_{i,u})$ is the gain of the optical concentrator, given by the following equation:

$$g_c(\psi_{i,u}) = \begin{cases} \frac{n^2}{\sin^2(\Psi_{max})}, & 0 \leq \psi_{i,u} \leq \Psi_{max} \\ 0, & \psi_{i,u} > \Psi_{max} \end{cases} \quad (6-7)$$

Where n is the refractive index, and Ψ_{max} represents the semi-angle of the field of view (FOV) of the PD.

For the sake of readability, Table 6-1 includes the notation of the main parameters used along the equations' descriptions.

Table 6-1 LiFi Channel Modeling Notation Summary

Symbol	Parameter
$H_{LiFi}^{i,u}$	LiFi channel gain between AP i and user u
$H_{LoS}^{i,u}$	Channel gain of the LoS path between AP i and user u
$\phi_{1/2}$	LED half-intensity angle
m	Lambertian emission order
$\phi_{i,u}$	Irradiance angle
$\psi_{i,u}$	Incidence angle
$d_{i,u}$	Euclidean distance between AP i and PD u
g_f	Gain of the optical filter
$g_c(\psi_{i,u})$	Gain of the optical concentrator
n	Refractive index
ψ_{max}	FOV semi-angle of the PD
A_{PD}	Physical area of a PD
$\Upsilon^{i,u}$	SINR experienced by user u with the transmitter AP i
R_{PD}	PD responsivity
P_{opt}	Transmitted optical power
ζ	Ratio of P_{opt} to the optical signal power
N_{LiFi}	PSD of noise at the receiver
B_{LiFi}	Bandwidth LiFi AP

6.1.2 Mobility and traffic offload management in Wi-Fi/LiFi integrated networks

The SNUM scheme runs in an SDN controller that manages a LiFi and Wi-Fi integrated network, which comprises a set of LiFi APs and a single Wi-Fi AP. The pathloss model of Wi-Fi channel is calculated as follows [29]:

$$P_L(d) = 40.05 + 20 \log_{10} \left(\frac{f_c}{2.4} \right) + 20 \log_{10}(\min(d, 10)) + (d > 10) 35 \log_{10} \left(\frac{d}{10} \right) \quad (6-8)$$

Where, f_c is the center frequency of Wi-Fi signal in GHz, d is the distance between the UE and the Wi-Fi AP, and 10 refers to the cell radius. The Wi-Fi channel gain of user k associated with Wi-Fi AP n can be expressed as follows [30] [29]:

$$G_{n,k} = \sqrt{|h_{n,k}|^2 10^{\frac{-P_L(d)+Z_{SF}}{10}}} \quad (6-9)$$

Where $h_{n,k}$ represents the small-scale fading gain of user k with Wi-Fi AP $n = 0$, which follows independent identical Rayleigh distribution with an average magnitude 2.46 dB [30]. The shadow fading (SF) loss, Z_{SF} , is modelled by a log-normal Gaussian distribution with zero mean, which is taken to be 4 dB. The SNR of user k associated with Wi-Fi AP $n = 0$ is expressed as follows:

$$\eta_{n,k} = \frac{G_{n,k}P_n}{N_o B_w} \quad (6-10)$$

The data rate of user k associated with Wi-Fi AP $n = 0$, is calculated as follows

$$\gamma_{n,k} = B_n \log_2(1 + \eta_{n,k}) \quad (6-11)$$

The signal-to-Interference-plus-Noise Ratio (SINR) of user k associated with LiFi AP $n > 0$ is calculated as follows [31]:

$$\eta_{n,k} = \frac{(\Gamma|h_{n,k}|^2 P_n)}{N_0 B_n + \sum_{n \neq i} (\Gamma|h_{i,k}|^2 P_i)} \quad (6-12)$$

The data rate of user k associate with AP n , $n > 1$ is calculated as follows [31]:

$$\gamma_{n,k} = \frac{B_n}{2} \log_2 \left(1 + \frac{e}{2\pi} \eta_{n,k} \right) \quad (6-13)$$

Where Γ denotes the optical to electric conversion efficiency of LED in LiFi AP; and e is the Euler's constant; B_n denotes the bandwidth of the downlink channel of AP.

To study the evolution process of the user k data transmission time, let us denote the size of data packet arrives at time slot t for user k by $\zeta_k(t)$. The time a user k , $T_k(t)$, requires for transmitting the integrity of data packet can be calculated as follows:

$$T_k(t) = \frac{\zeta_{k_r}(t) + \zeta_k(t)}{\eta_{n,k}}, \quad (6-14)$$

Where $\zeta_{k_r}(t)$ is the remaining of data packets left for transmission from the previous slots assigned to user k . Here there are two main cases can emerge:

- Case # 1: $\zeta_{k_r}(t) + \zeta_k(t) = 0$, which indicates that the user k has successfully to transmit all the data packets in the previous time slots. It is currently idle, and their time slots can be allocated to other uses.
- Case # 2: $T \leq T_k(t)$, which means that the transmission time of user k is longer than the current slot duration time T . Thus, the remaining data packets are transmitted in the next time slots.

Let us assume that the user plane packets arrive at a Wi-Fi AP and a LiFi AP follow Poisson distribution functions with rate λ_w and λ_l , respectively. The user packets are assumed to have an equal size, following an exponential service time distribution with parameters for Wi-Fi and LiFi APs as, μ_w^{-1} and μ_l^{-1} , respectively. The Wi-Fi and LiFi APs are modelled by an M/M/1 queueing system. The average delay (latency) experienced by a packet serviced by the Wi-Fi AP and the LiFi AP is given, respectively, by [32] as follows:

$$\widetilde{D}_w = E[D_w] = \frac{1}{\lambda_w - \mu_w} \quad (6-15)$$

$$\widetilde{D}_l = E[D_l] = \frac{1}{\lambda_l - \mu_l} \quad (6-16)$$

6.1.3 Smart network user mobility algorithm

SNUM scheme, which is alternatively called One-Network-Multiple-AP (ONMA), moves the user-AP association table of each AP and UD's IP resolution and configuration to the SDN controller. It is designed to support seamless mobility, enabling the integrated LiFi and Wi-Fi network to efficiently support URLLC and eMBB services. The scheme assigns APs to UDs based on their signal strength and traffic load of APs they can associate with. A use case such as virtual classrooms, in which SDN-enabled LiFi attocellular and Wi-Fi networks are integrated, can be a great enabler for delivering the teaching materials of the next generation education system. Both the students and teachers can get their teaching materials and the courses contents independent of their QoS requirements and location in the classrooms or elsewhere.

Algorithm 1: Smart network user mobility algorithm.

Input : $\widetilde{D}_w, \widetilde{\gamma}_w$: predicted average delay and throughput of WiFi AP $n, n = 0$
 $\widetilde{D}_l, \widetilde{\gamma}_l$: predicted average delay and throughput of LiFi AP $n, n > 0$
 $\eta_{n,th}$: LiFi AP SNR threshold

Output: user AP association

```

1  $n_p^* = \underset{n}{\operatorname{argmax}} \widetilde{\eta}_n, \forall n \in \mathcal{N}_{re}$ , SNR requirement
    $n_D^* = \underset{n}{\operatorname{argmin}} \widetilde{D}_n, \forall n \in \mathcal{N}_{re}$ , delay requirement
    $n_{\widetilde{\gamma}}^* = \underset{n}{\operatorname{argmax}} \widetilde{\gamma}_n, \forall n \in \mathcal{N}_{re}$ , throughput requirement
2 if  $\widetilde{\eta}_{n_p^*}[t] > \eta_{n,th}$ , begin
3   switch  $k = 1$  do
4     URLLC user  $k$ 
5     if  $\widetilde{D}_{n_D^*}[t] < \widetilde{D}_w[t]$ 
6        $k \leftarrow n_D^*$ , LiFi AP else
7          $k \leftarrow n_w^*$ 
8     WiFi AP
9   switch  $k = 2$  do
10    eMBB user  $k$ 
11    if  $\widetilde{\gamma}_{n_{\widetilde{\gamma}}^*}[t] > \widetilde{\gamma}_w[t]$ 
12       $k \leftarrow n_{\widetilde{\gamma}}^*$ , LiFi AP else
13         $k \leftarrow n_w^*$ 
14    WiFi AP
15 else
16    $k \leftarrow n_w^*$ 
17 WiFi AP

```

Figure 6-4 Smart network user mobility algorithm

The SNUM procedure is described in Algorithm depicted in Figure 6-4. It receives as inputs the available network and user's context-aware information. This mainly includes the UEs' SNR (SINR) measurements and service requirements (delay and throughput), the available spectrum resource units, average delay, and throughput of each network AP. This information is processed by the SDN controller on regular basis to support seamless UEs association to network APs that can support their requested URLLC or eMBB service classes. A cell is not considered as a candidate to a UE, if it provides SNR lower than a specific threshold. The

SNUM algorithm is further discussed in section 6.3.1 and its performance evaluation is discussed in Section 6.3.2.

6.2 Multi-WAT offloading in industry 4.0

6.2.1 Scenario setup

As pointed out in the introduction of this section, the objective of this scenario evaluation in this deliverable is to extend the evaluation performed in 5G-CLARITY D2.3 [2] by additionally including LiFi as wireless access technology in an industrial private 5G network. Specifically, the goal is to assess the amount of 5G radio resources that can be released from serving eMBB users and use them to serve URLLCs since eMBB users will be served by Wi-Fi and LiFi technologies.

Figure 6-5 represents the setup of the considered scenario for the evaluation study. The scenario tries to resemble the industrial scenario envisioned in the 5G-CLARITY UC2.1 [18]. The services considered are typical eMBB (e.g., VR/ AR, video-streaming and mobile broadband access required by workers) and URLLC (e.g., motion control).

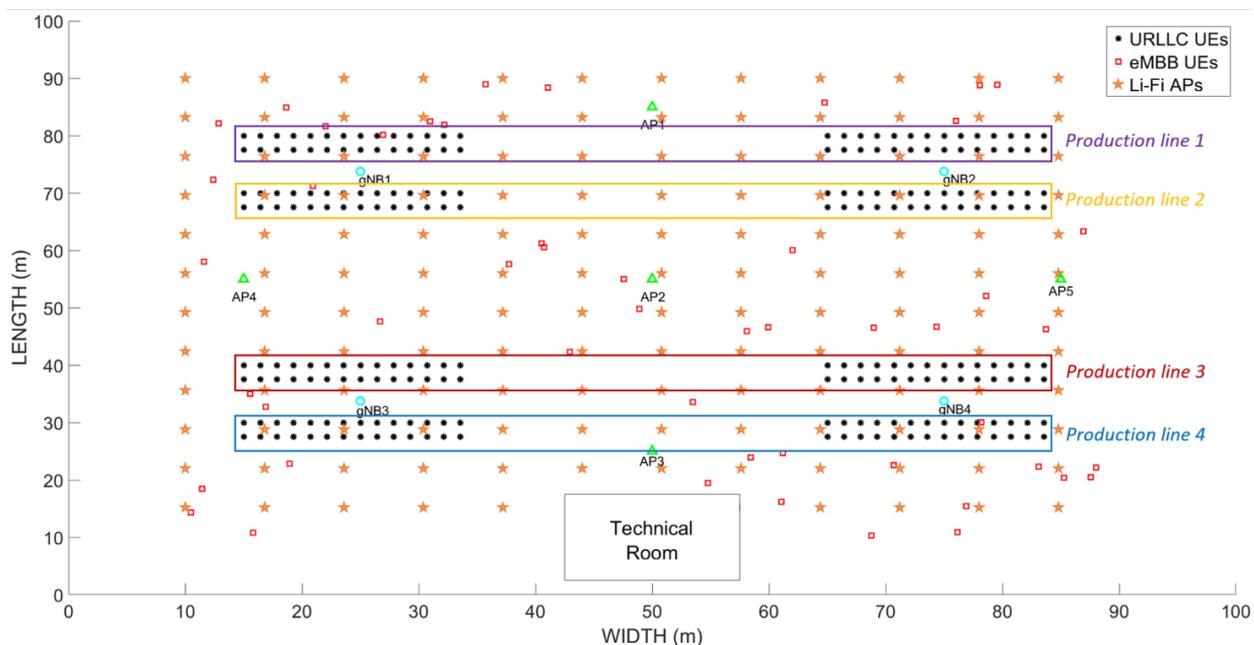


Figure 6-5 Industrial scenario layout with multi-WAT RAN

The industrial private site occupies a geographical area with dimensions 100 m x 100 m. The scenario includes four 5G gNBs and five Wi-Fi access points. Additionally, the number of LiFi access points varies depending on the percentage of LiFi coverage desired in the scenario. There are two types of users depending on the requested service:

- eMBB users, uniformly distributed along the factory floor (e.g., workers using Augmented Reality (AR) and Virtual Reality (VR) applications).
- URLLC users, distributed in the four production lines of the industrial premises.

In order to conduct the evaluation study, we will compare the performance in terms of throughput in two different configurations of the scenario depicted in Figure 6-5. In the first configuration considered, dubbed as baseline scenario, 5G NR is the only WAT deployed. In this case, both types of traffic eMBB and URLLC are served by 5G NR. In contrast, 5G NR, Wi-Fi and LiFi are the WATs deployed in the second configuration. In this second situation, part of the eMBB users is served through Wi-Fi or LiFi, depending on the SINR level

they receive from the nearest access points of each technology. The bandwidth consumed by eMBB users in the second situation will be compared with the one consumed in the baseline scenario, in which there are no Wi-Fi or LiFi access points deployed. In this way, the 5G radio resources freed up by Wi-Fi and LiFi technologies can be estimated, being these 5G resources available to be used to serve URLLC processes, or even they can be released by the private operator saving costs associated with the licensed spectrum acquisition.

Table 6-2 specifies the components used for this scenario evaluation. The corresponding configuration issued by the Service and Slice Provisioning system was used as input to setup the system-level industrial RAN simulator. Moreover, Table 6-3 includes the specific configuration of the main parameters of the setup.

Table 6-2 Components included in the scenario modelling

Components	Background	Extensions in CLARITY	Justification
Multi-WAT RAN (5G NR, Wi-Fi and LiFi)	System-level RAN simulator for small cells	RAN simulator extended for an industrial network with multiple wireless access technologies (5G NR and Wi-Fi)	Private industrial networks are a key use case scenario in 5G-CLARITY

Table 6-3 Configuration of the Main Simulation Parameters

Parameter	Configuration
eMBB UEs guaranteed bitrate	5 Mbps
URLLC UEs bitrate	1.55 Mbps
URLLC delay requirement	1 ms
Number of eMBB UEs	50
Number of URLLC UEs	224 distributed in 4 production lines
Cell type	Femtocells, Wi-Fi, and LiFi cells
Number of femtocells	4
Number of Wi-Fi cells	5
Number of LiFi cells	Variable in the range [144 - 841]
Direction of transmission	Downlink
eMBB traffic distribution	Uniform
Path loss model for femtocells	Indoor Hotspot (InH)
Path loss model for Wi-Fi cells	Indoor Hotspot (InH)
Antenna height in femtocells	6 m
Antenna height of Wi-Fi APs	4 m
Transmission power in femtocells	30 dBm
Transmission power in Wi-Fi cells	20 dBm
UE height	1.5 m
UE thermal noise	-174 dBm/Hz
Noise figure	9 dB
Carrier frequency in femtocells	3.5 GHz
Carrier frequency in Wi-Fi APs	2.4 GHz
Bandwidth of 5G gNBs	100 MHz
Bandwidth of Wi-Fi APs	40 MHz
Frequency reuse	1
URLLC packet size	80 bytes
Sweep of number of URLLC UEs to estimate the load for the URLLC model	From 1 to 28 URLLC UEs
Number of eMBB slices	1

Number of URLLC slices	4 (one per production line)
LiFi Parameters	
LED half-intensity angle ($\phi_{1/2}$)	60°
Field of View (FOV) semi-angle of the PD (ψ_{max})	85°
Physical area of a PD (A_{PD})	10 ⁻⁴ m ²
PD responsivity (R_{PD})	0.53 A/W
Transmitted optical power (P_{opt})	10 W
Ratio of P_{opt} to the optical signal power (ζ)	1
PSD of noise at the receiver	10 ⁻²¹ A ² /Hz
Bandwidth of LiFi APs	20 MHz
Height of LiFi APs	3 m

6.2.2 Scenario validation

As in the evaluation conducted in the deliverable D2.3, we have used simulation tools for the scenario evaluation purposes. Specifically, we have conducted several simulations in an industrial RAN simulator developed in particular for this project. The reason behind that, as stated in D2.3, is because of the flexibility and scalability offered by these tools to rapidly modify the configuration of most of the parameters. Additionally, this simulator includes analytical expressions that have been proved to fairly model the reality.

One of the benefits of having a multi-WAT access network composed of the above-mentioned technologies (i.e., 5G NR, Wi-Fi, and LiFi) is the throughput gain. Here, we assume that the deployed versions of the Wi-Fi and LiFi technologies do not include URLLCs support. Thus, these technologies are used only to perform the offloading of 5G technology from eMBB traffic. Therefore, the saved 5G radio resources can be dedicated, for instance, to increase URLLC throughput, as shown in the results discussed below.

The first metric assessed and shown in Figure 6-6 was the mean attainable throughput for eMBB UE as a function of the 5G bandwidth allocated for eMBB traffic. Three configurations of the scenario were considered to perform this evaluation: (i) RAN with only 5G NR (baseline scenario), (ii) RAN with 5G NR and Wi-Fi; (iii) RAN with 5G NR, Wi-Fi, and LiFi. Full coverage was considered for LiFi technology (i.e., there is a separation of 2.8 m between LiFi APs). Remarkably, the obtained results show an over 100 times increase in the mean attainable eMBB UE throughput for the setup considered due to the inclusion of LiFi technology. LiFi provides very high data rates that perfectly fits data-hungry applications in reduced geographical areas in which users usually have fixed positions (e.g., classrooms, or videoconference rooms). In the context of industry, an exemplary use case could be eXtended Reality (XR)-assisted workers at fixed or quasi-static locations. Interestingly, as can be observed in the light blue line represented in Figure 6-6, LiFi technology outperforms the throughput the eMBB users can achieve compared to the one they can reach when they are served by 5G and Wi-Fi. This is the reason why we see a decrease in the eMBB UE achievable throughput when the 5G bandwidth allocated to eMBB slice increases. The more 5G bandwidth allocated to this kind of traffic, the higher the number of users served by 5G technology. Consequently, fewer eMBB users are served by LiFi, what implies a reduction in the mean reachable throughput. As a concluding remark, Wi-Fi and LiFi technologies have been proved to be useful to offload eMBB traffic from 5G NR. In this way, all the 5G radio resources can be released and destined for services with the most stringent performance requisites, thus cheapening the deployment and operational costs of the industrial private radio access network.

Figure 6-7 shows the packet loss ratio (PLR) considering a delay constraint of 1 ms of the URLLC slice versus the slice load for the three scenarios mentioned above.

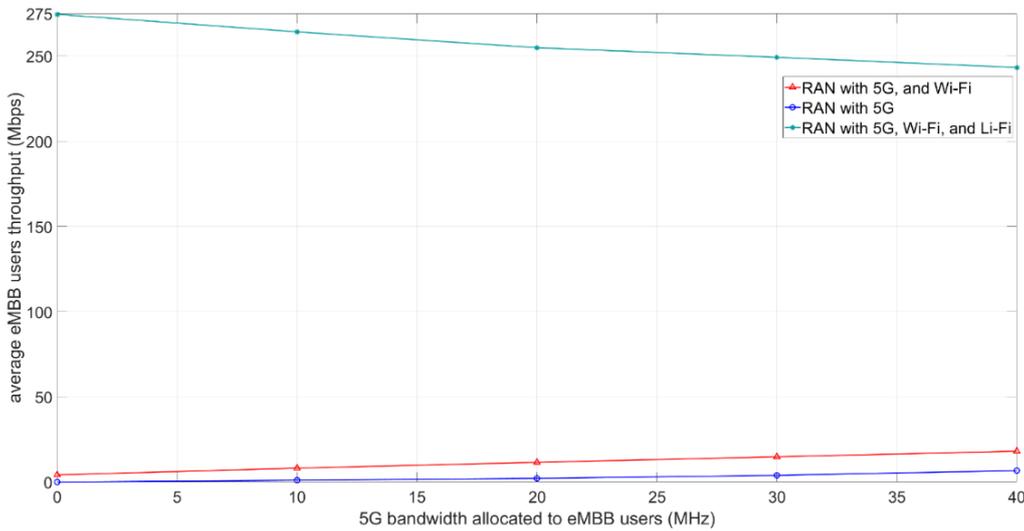


Figure 6-6 Average eMBB users' throughput versus the 5G bandwidth allocated to eMBB users

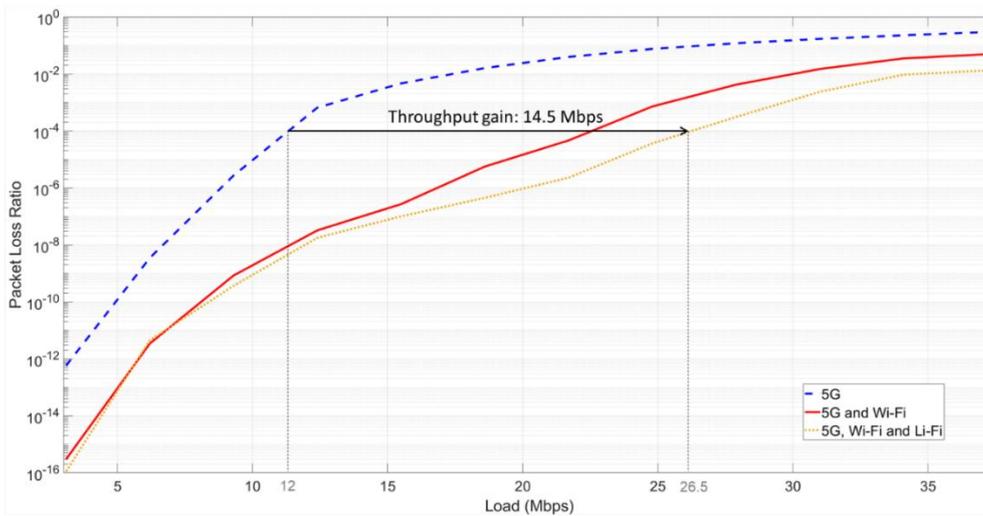


Figure 6-7 URLLC slice packet loss ratio vs the traffic load

The analytical model for URLLC traffic performance in 5G NR and described in Section 5.2.3 of the 5G-CLARITY deliverable D2.3 was used to generate these results. As observed, the released 5G radio resources due to eMBB traffic offloading to LiFi and Wi-Fi can be harnessed to increase the URLLC slice throughput while preserving the same PLR and delay constraints. Specifically, the inclusion of Wi-Fi and LiFi allows for an increase in the URLLC slice throughput more than twice over while preserving the same performance in PLR and delay. By way of illustration, in our setup we observe a gain of 14.5 Mbps of throughput for the URLLC slice when Wi-Fi and LiFi technologies are present, since all the 5G bandwidth (i.e., 50 MHz per URLLC slice) is allocated to URLLC services. In other words, the 5G radio interface (NR-Uu) can additionally support 14.5 Mbps per URLLC slice while guaranteeing the same PLR and delay requisites (10^{-4} and 1 ms, respectively). Please, note that the limitations in the throughput gain of URLLC users are due to the maximum available bandwidth for a URLLC slice being 50 MHz.

On the other hand, Figure 6-8 shows the average throughput an eMBB user can achieve depending on the technology they are served with for a scenario configuration in which the 5G available bandwidth for the eMBB slice is 40 MHz, and there is a LiFi coverage percentage of 80.91% (see Table 6-4). It is worth highlighting that LiFi technology overcomes the performance provided by 5G and Wi-Fi in terms of the throughput eMBB users can reach for a setup with a dense LiFi deployment, such as the one considered in this study.

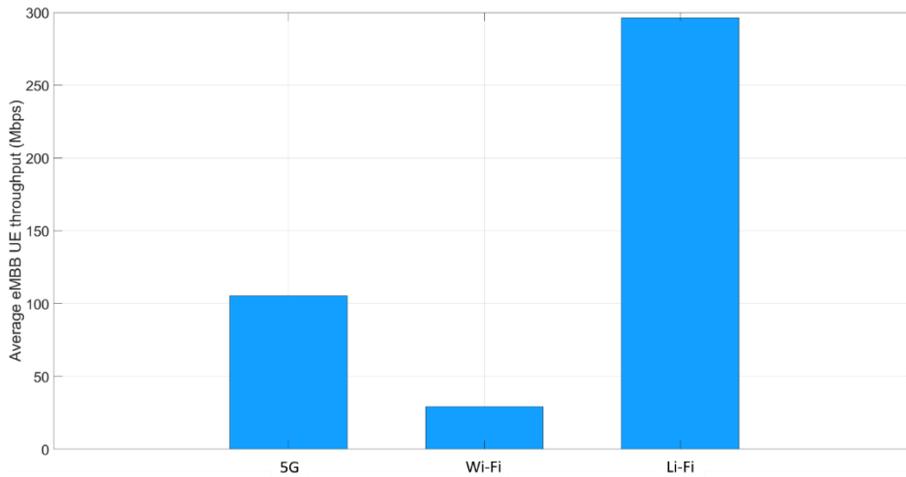


Figure 6-8 Average eMBB user achievable throughput per technology

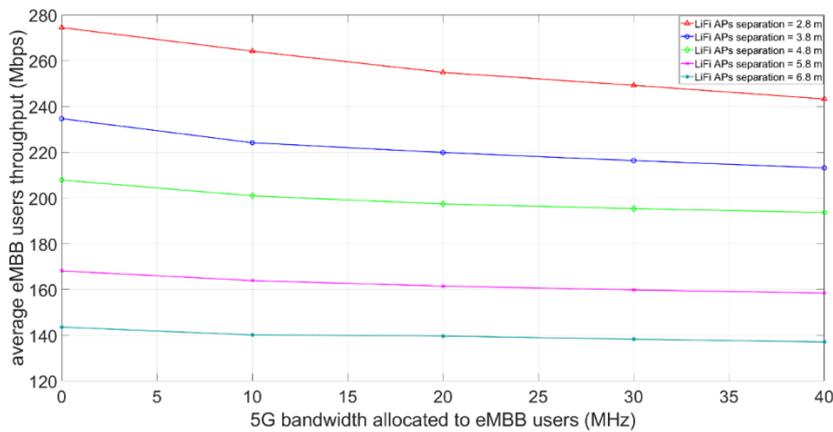


Figure 6-9 Average eMBB users' throughput versus the 5G bandwidth allocated to eMBB users for different separation values between LiFi APs

Last, we conduct a study to see how the percentage of LiFi coverage impacts the eMBB user achievable throughput. To that end, we have swept the inter-site distance between LiFi APs in steps of 1 meter, from 2.8 m to 6.8 m. Figure 6-9 depicts the average eMBB users' achievable throughput versus the 5G bandwidth allocated to eMBB users. As observed, the mean throughput that eMBB users can achieve decreases with the increment of the LiFi APs separation, as this implies a reduction in the number of APs deployed in the scenario. Similarly, in Figure 6-10 the average throughput of eMBB users is represented, but in this case as a function of the separation between LiFi APs. As can be observed in the figure, the mean eMBB user achievable throughput is roughly linearly dependent on the separation between LiFi APs. As expected, the throughput that eMBB users can achieve decreases with the percentage of LiFi coverage in the scenario. The number of LiFi APs and the coverage area for the values of the LiFi APs separation are gathered in Table 6-4.

Table 6-4 Configuration of LiFi APs in the Scenario

LiFi Aps Separation (m)	No of LiFi APs	Coverage (%)
2.8	841	80.91
3.8	441	42.43
4.8	289	27.81
5.8	196	18.86
6.8	144	13.85

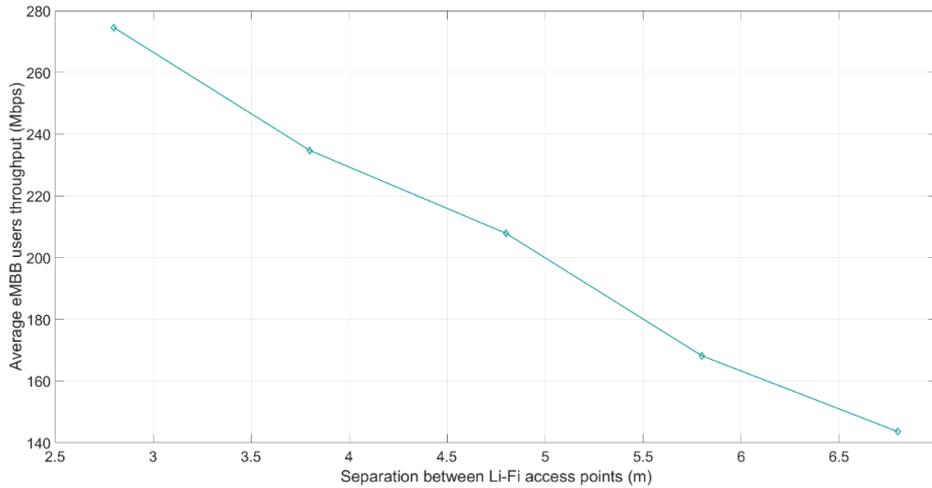


Figure 6-10 Average eMBB users' throughput versus the separation values between LiFi APs

6.3 Mobility and traffic offload management in Wi-Fi/LiFi integrated networks

6.3.1 Scenario setup

An SDN-enabled network composed of 16 LiFi APs and a single Wi-Fi AP is considered, as shown in Figure 6-11. The separation distance between two direct neighbouring LiFi APs is set to 3 m, and the maximum vertical distance between the users and LiFi APs is set to 2.20 m. Two baselines are considered: the standard (STD) and the trajectory-based handover approaches. The STD handover scheme assigns a user to an AP that provides the best signal-to-noise ratio (SNR). The trajectory-based handover approach assigns APs to users located along their network path. The handover process of users is independent from each other, though it affects the total network performance.

Table 6-5 Scenario setup components

Components	Background	Extensions in CLARITY	Justification
Controller	A Centralized unit for network management	Smart network user mobility scheme	Support seamless user mobility across the SDN-enabled LiFi/Wi-Fi integrated network
MATLAB Communications Toolbox	Understand some of the IEEE802.11 WLAN MAC models, specification, and call functions	Develop the user mobility and service provisioning functions in the LiFi/Wi-Fi network simulation files	Provide a simulation environment to integrate the LiFi network with IEEE802.11 WLAN network

The components which are involved in the SNUM scheme process and scenario evaluations are described in Table 6-5.

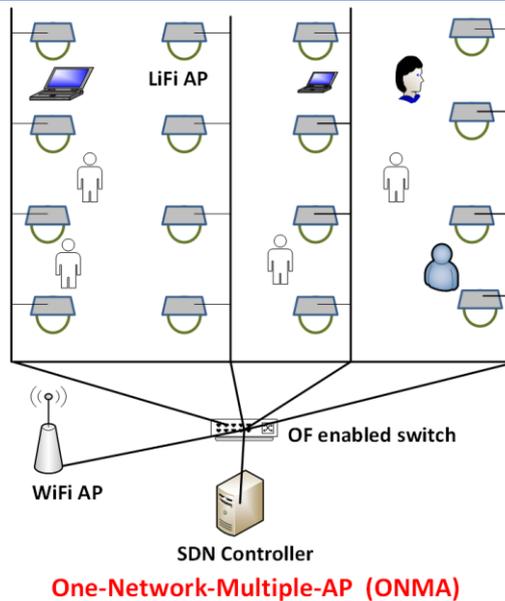


Figure 6-11 Network scenario and smart network user mobility architecture

The performance of user mobility management schemes is challenged by the number of users and their speed, which require a comprehensive evaluation for their impact on the integrated network throughput. Multiple scenarios are conducted, considering several users with different speeds and classes of traffic loads, to compare the performance of SNUM scheme to baselines in a simulation LiFi/Wi-Fi integrated network environment. The performance of SNUM scheme is evaluated in terms of the network throughput and guarantees of users' QoS requirements. The cumulative distribution functions of the network throughput and delay of these UDs are evaluated to show the reliability of the offered services in the integrated network.

6.3.2 Scenario validation

A discrete time simulation environment has been developed in MATLAB to evaluate the proposed smart network user mobility (SNUM) scheme in the SDN-enabled LiFi/Wi-Fi integrated network depicted in section 5. This is comprised of 16 LiFi APs and a single Wi-Fi AP, which provides services to UDs distributed in an indoor room of size $10 \times 10 \times 3$ ($W \times L \times H$), as shown in Figure 6-12. Users are classified into fixed and mobile URLLC and eMBB users. Two fixed URLLC and eMBB users are set under each LiFi AP. Other mobile users move within the room, following a random waypoint model with two speed uniform distributions that take values from a low-speed range $[0.1, 2]$ m/s or a higher-speed range $[3, 5]$ m/s. The main parameters of the LiFi attocellular network and Wi-Fi network are summarized in Table 6-6. The traffic and services parameters are summarized in Table 6-7. The UDs generate a periodic URLLC traffic and eMBB multimedia traffic packets according to the statistical traffic parameters introduced in Table 6-7.

An experimental handover scheme was implemented to support seamless user mobility in a hybrid LiFi/Wi-Fi network, which is managed by an SDN controller, as explained in Section 5.c of [33] and in 5G-CLARITY D2.3 [2]. It was observed that the handoff time between LiFi APs is shorter than that between LiFi and Wi-Fi APs, as shown in Figure 6-13 [33]. The main source of handoff delay in both handover events is attributed to the fact that a target AP requires to create a new entry for each user newly associated in their user association table. This contributes to the link switching delay budget. The SNUM scheme is designed to mitigate this source of handoff delay by centralizing the association table in the SDN controller, while maintaining a temporary virtual reference to users in their selected network APs. The controller links the user reference identifier to their history table that records the characteristics of their mobility, received services, and network APs states during their mobility across the network.

The experimental user mobility scenario, which is explained in Section 5 in D2.3 [2] and shown in Figure 8 in [33], is evaluated via simulation, in which the user x streams the measured video samples along a replicated path LiFi (16) - LiFi (15)- Wi-Fi(17) marked in the SDN-enabled LiFi/Wi-Fi integrated network simulation environment shown in Figure 6-12. As a result, the SNUM scheme could noticeably reduce the handoff time between LiFi APs, and LiFi and Wi-Fi APs, as shown in Figure 6-13. It enables the UE to smoothly move among the different APs, while keeping the video streaming rate at an acceptable level.

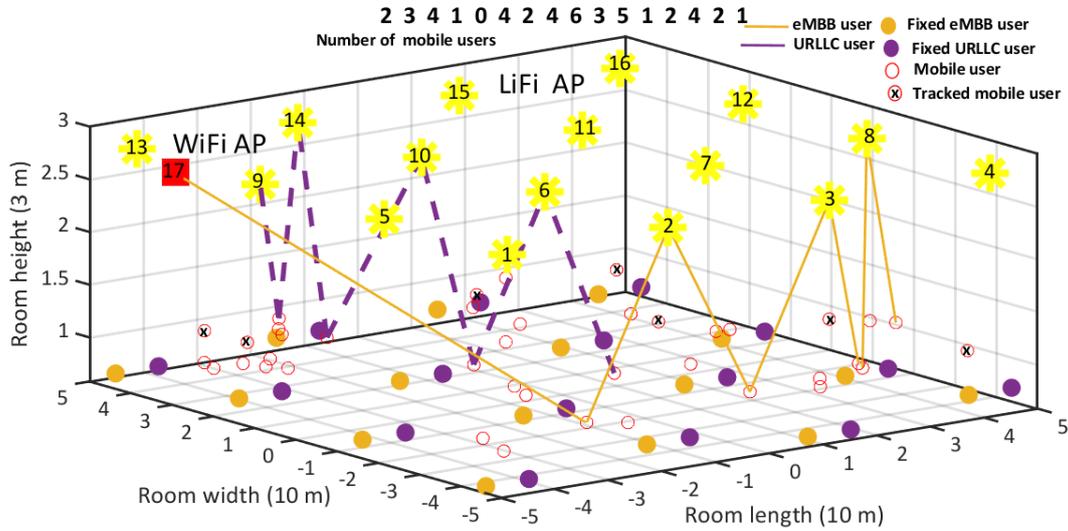


Figure 6-12 SDN-enabled LiFi/Wi-Fi integrated network simulation environment

Table 6-6 LiFi Attocellular and Wi-Fi Networks Simulation Parameters

Parameter	Description	Value
$\Phi_{1/2}$	LED half power semi angle	60°
η_w	Reflectivity factor of wall	0.8
g_f	Gain of optical filter	1
A_R	Physical area of photo detector (PD)	10 mm^2
η_f	Reflectivity factor of floor	0.8
N_0	Noise power spectral density	$10^{-21} \text{ A}^2/\text{Hz}$
B	Modulated bandwidth	20 MHz
η_c	Reflectivity factor of ceiling	0.8
Ψ_c	Receiver field of view (FOV)	90°
P_t	Transmission power	9 W
ζ	Refractive index	1.5
n_r	PD's orientation vector	[0, 0, 1]
R_{PD}	PD responsivity	0.5 A/W
n_t	AP's orientation vector	[0, 0, -1]
$ \mathcal{S} $	DCO-OFDM sub-carriers	512
$ \mathcal{N} $	Number of LiFi APs and Wi-Fi AP	17
f_{c_w}	Wi-Fi central frequency	2.4 GHz
B_w	Wi-Fi modulated bandwidth	20 MHz
P_w	Wi-Fi AP transmission power	20 dBm
N_{0_w}	Wi-Fi noise power spectral density	-174 dBm/Hz

Table 6-7 Services and Traffic Simulation Parameters

Parameter Description	URLLC Requirement	eMBB Requirement
D (delay) (ms)	[0.5, 8]	[10, 50]
γ (Throughput) (Mbps)	[5, 12]	[5,40]
λ (Packets arrival rate)	30	40
L_p (Packet length) (bytes)	250	1450

It improves the data rate of user while moving between LiFi APs, and LiFi and Wi-Fi APs. This demonstrates the advantage of centralizing the association tables of all APs in the SDN controller to support seamless reliable services to the users. The SDN controller engineer user to APs association according to their location in the network and their service requirements. It effectively demonstrates its capability to reduce the connection outage during handover from LiFi to LiFi and LiFi to Wi-Fi network, as shown in Figure 6-13.

The performance of SNUM scheme is compared to the standard (STD) and trajectory-based handover approaches. The STD approach associates a user to an AP that provides the best signal strength or signal-to-noise ratio (SNR). The trajectory-based approach selects APs to users along their trajectory path towards their destination. To realise the trajectory-based handover approach, eMBB and URLLC users follow flexible planned paths, as shown in the network simulation environment depicted in Figure 6-12. Mobile URLLC users follow the network APs path 6-10-14-9; and mobile eMBB users follow the network APs path 8-3-2-17. While these APs shape the trajectory paths for both URLLC and eMBB users, the users can be associated to other neighbouring APs providing a better SNR along their path towards their destination location.

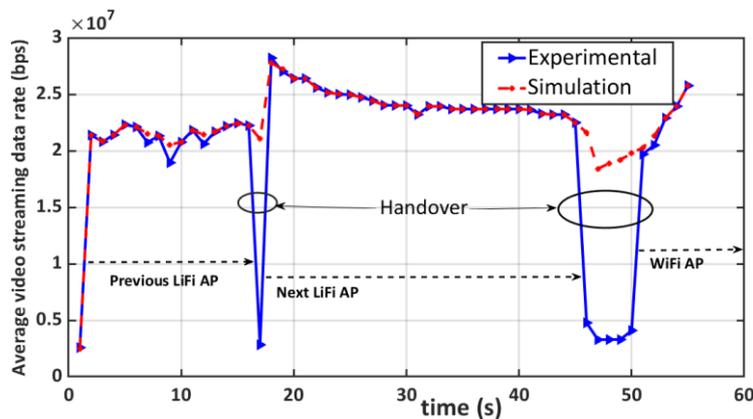


Figure 6-13 Experimental and simulation-based user handover performance from LiFi to LiFi and LiFi to Wi-Fi in SDN-enabled LiFi/Wi-Fi integrated network

Furthermore, they are offered a priority service by the APs, comparing to the STD approach users. Multiple scenarios are run on the network simulation environment shown in Figure 6-12, to demonstrate the impact of the number of users, user speed, delay, and throughput requirements on the performance of SNUM, STD and trajectory-based handover mobility approaches.

In this scenario the performance of the different approaches is evaluated in terms of the total network throughput under the influence increase in the number of users while fixing their speed to 0.3 m/s. The total network throughput is expressed as the total throughput of network APs. As the number of users increases, the SNUM scheme provides significant improvement over the other approaches in terms of the total network throughput, as shown in Figure 6-14. The total throughput improvement of SNUM scheme over trajectory-based and STD approaches is calculated as 26.15% and 41.3721%, respectively. This significant gain can be attributed to the capability of SNUM scheme to admit more users by associating them to LiFi APs corresponding better to their data rate or delay requirements. But also, when the LiFi APs are congested, the users can be connected to the Wi-Fi APs.

More network throughput gains are expected at higher users' speed and APs intensities, as SNUM scheme is carefully designed to increase the long-term network average network throughput.

In this scenario, the number of active fixed and mobile users is set to 80. The performance of the different approaches is evaluated as the speed of users is gradually increased from the lowest to the highest speed. As a result, the average network throughput continues to decrease in terms of the increase in the users' speed, which reflects the cost paid in terms of the reduced received average throughput during network user mobility. Note that the interference among neighbouring LiFi APs and higher users speed may enforce all the different handover approaches to associate the affected users to Wi-Fi AP.

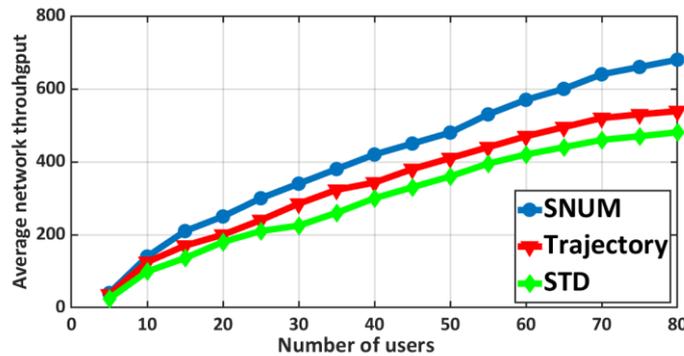


Figure 6-14 Average Network throughput versus number of users

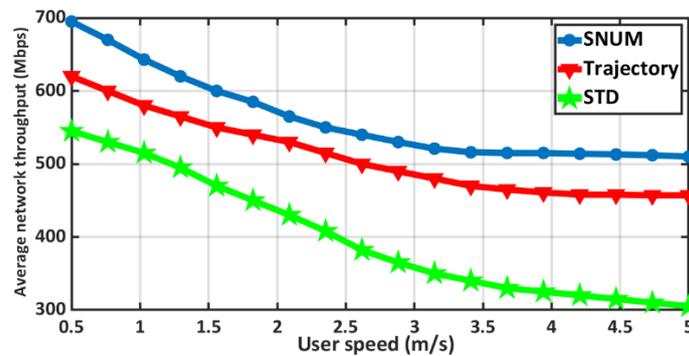


Figure 6-15 Average Network throughput versus user speed

In some cases, the trajectory-based approach sacrifices the best AP association to connect to the Wi-Fi AP to reduce the number of handovers, particularly at higher user speed.

The SNUM scheme achieves, at the lowest user speed, up to 12.09 % more total network throughput gains with respect to (w.r.t.) trajectory-based approach and 27.53 % w.r.t. STD approach. It also shows, at the highest user speed, improvement gain up to 11.16 % w.r.t. trajectory-based approach and 67.21% w.r.t. STD approach. This refers to the fact that the SNUM scheme decreases the handover rate for different user speeds. The different handover approaches are sensitive to different user speeds, where a user is transferred to the Wi-Fi AP at different user speed thresholds. This decreases the handover rate but may decrease the network throughput.

The different handover approaches are evaluated in terms of ensuring the strict delay and reliability requirement of URLLC services. This scenario evaluates the cumulative distribution function (CDF) of the average delay experienced by the URLLC users in the network. The arrival rate of URLLC user is increased at a double rate of λ_u . The mobile users move with a random speed selected from the low and high uniform speed ranges. The network APs have different numbers of URLLC users and traffic loads. The SNUM scheme achieves the best URLLC service reliability, comparing to the other approaches, as shown in Figure 6-16. This can be attributed to the fact that the URLLC users are associated to the least average delay network APs that serve the URLLC users with a shorter average delay.

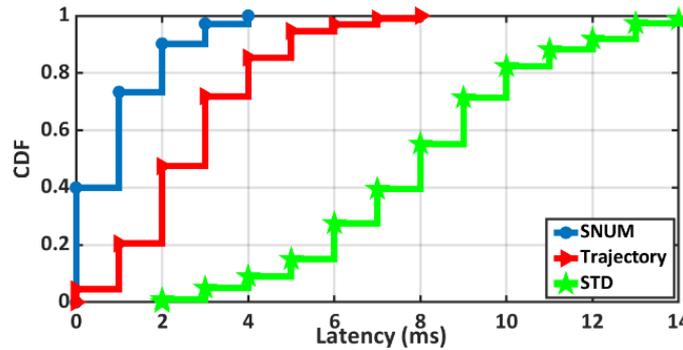


Figure 6-16 CDF network average delay of URLLC users (service)

Furthermore, the SNUM scheme ensures that the network APs allocate a number or resource units that can guarantee the service rate per timeslot to URLLC users. This determines in turn the maximal queue length that indicates the bound for URLLC packet delay guarantee.

6.4 Positioning using multi-WAT

Within the 5G-CLARITY project a system for precise UE positioning was developed. This system leverages the available WATs to estimate the position of a given UE. This is one of the key innovations in this project. The main advantage of the proposed positioning system is that it can leverage many different WATs in order to offer best possible position estimate. The positioning relevant data is collected in a positioning server which deploys a fusion algorithm to improve the position estimate. With this approach, if multiple positioning capable WATs are available, the UE position estimate can be additionally improved. Nevertheless, the availability of multiple WATs can also improve the coverage of the positioning system.

In order to enable optimal deployment of the WATs and to enable highest possible positioning precision a modelling tool was developed. This modelling tool was presented in 5G-CLARITY D2.3 [2]. Within the modelling tool each WAT is represented with a separate model. The WAT models receive the UE path, as well as the position of the available access points (AP) for each of the WAT. Based on the position of the UE and the APs used for the current position estimate, the position of the UE is calculated, and the corresponding noise is added to emulate the real position estimate.

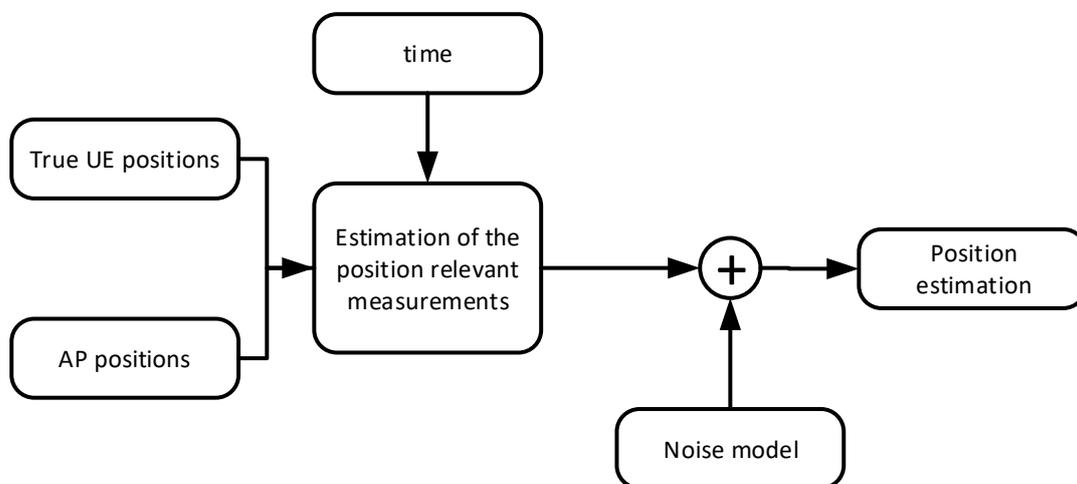


Figure 6-17 Extended position estimation model

Table 6-8 Positioning Relevant Parameters for Different 5G-CLARITY Technologies

Positioning Method	Associated Technology	Measurement Parameters
Time difference of arrival (TDoA)/Time of arrival (ToA)	Sub-6 GHz positioning	Time of arrival of a frame at the receiver.
Two-way ranging (TWR)	mmWave positioning	Time of arrival of a frame at the receiver. This is performed twice in a TWR approach.
RSSI based	LiFi positioning	RSSI value of the received light
Geometrical method	VLP/OCC positioning	Positions of a light sources in a photo taken with a camera

In Figure 6-17, the true UE and AP positions are supplied to the model. The position of the UE is changing in time and, therefore, the supplied positions are time stamped. Between two timestamped positions, the UE position is linearly interpolated. When the position of the UE is requested in a simulation scenario, the true position of the UE is calculated based on the supplied data for the UE movement as well as the time at which the position is estimated. Further, based on the AP positions, the positioning relevant measurements are calculated. These so-called positioning relevant measurements are different for different technologies. These measurements are the actual parameters which are measured by a positioning technology and are used for position estimation. These parameters can be for example RSSI values, time of arrival, positions of lights, etc.

In order to model each technology precisely, a few different classes of position estimation methods are defined. They represent the different WATs used in 5G-CLARITY. Furthermore, each of the defined class in Table 6-8 is modelled separately and the noise is added in the parameters that are being measured in a realistic case.

It is worth mentioning that the models in this section are representative of the technologies available in the project and are expected to be deployed in the Bosch factory demonstration. In the next section, we take a step beyond the existing technologies and rely on the time-stamp capability, which is anticipated to be embedded in the most of the future communication nodes under the standard Fine Time Measurements (FTM), and develop a more sophisticated localization algorithm.

6.4.1 TDoA/ToA model

Positioning using TDoA/ToA approach is performed by measuring pseudo-distances and performing trilateration. The localization process is performed by transmitting timestamped frames from synchronized localization APs. The timestamped frames are received by the UEs.

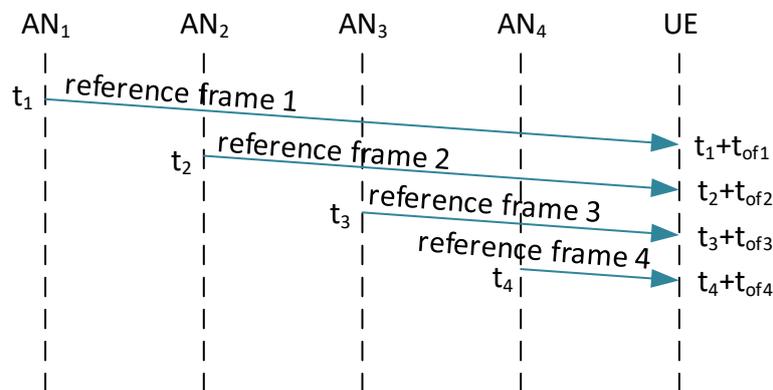


Figure 6-18 DL/TDoA method

The UEs estimate the time of arrival and estimate their position. This is called downlink TDoA (DL-TDoA). A

slightly different approach is called uplink TDoA (UL-TDoA). In this case the UE transmits a frame which is received by synchronized localization capable APs. From computational perspective both approaches are the same. The only slight difference is the entity in which the position is being estimated. In the DL-TDoA case the UE estimates its own position, while in UL-TDoA, the position of the UE is usually estimated in the localization server.

The basic principle of DL/UL-TDoA is shown in Figure 6-18. The anchor nodes AN1, AN2, ..., ANn are transmitting time stamped reference frames. These frames are received at the UE and the time of arrival of these frames are estimated. The position of the UE is estimated using the estimated ToAs.

The position estimation of the UE is performed using trilateration. The basic principle is shown in Figure 6-17. The estimated distances between the anchor nodes and the UE describe circles which intersect in a single point.

In order to simulate the DL-TDoA approach, in the developed tool, the positions of the ANs are known and also the path of the UE is known and available for the tool. Also, the path is timestamped, and the position of the UE is known in each period of time. The within the model of this positioning system, first the distances between the ANs and the UE are calculated as:

$$r_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \quad (6-17)$$

where, $(x_i; y_i)$ are the coordinates of the i -th anchor node, $(x; y)$ are the coordinates of the UE and r_i is the distance between the AN and the UE.

In order to simulate the process of arrival of the reference frames at the UE, the time of flight is calculated as $t_{of} = r \cdot c$ where c is the speed of light. The reference frames from different ANs are sent in time division multiplexing (TDM) fashion. Nevertheless, the time between sending the different frames is known and can be considered. For this reason, it can be assumed that the reference frames from all the ANs were sent at the same time, without loss of generality.

The ToA is used to measure the distance between the ANs and the UE. The obtained distances would describe circles which would intersect in a single point, in ideal case, i.e., the UE position. In the case of DL-TDoA, the estimated distances would describe a circle defined with the equation:

$$r_i = r_{ti} - d_{offset} = \sqrt{(x_i - x)^2 + (y_i - y)^2}, \quad (6-18)$$

where, r_{ti} is the estimated distance between the anchor point and the UE, and d_{offset} is the distance error due to the clock offset between ANs and the UE. The clocks of the UE and the AN are not synchronized since the synchronization process is too complex. Instead, the problem is solved by using one additional AN.

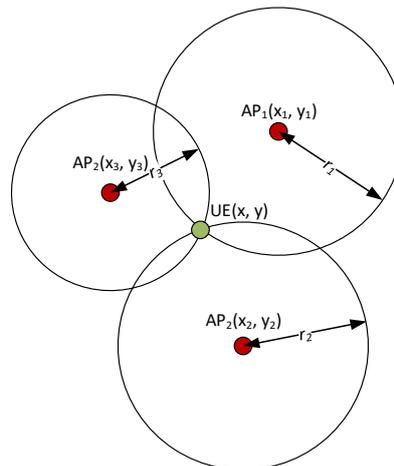


Figure 6-19 Range-based localization approach

The position is obtained by solving a system of equations consisting of the equations describing the circles around each node. In a 2-dimensional (2D) case, a total of 4 equations, i.e., ANs, would be needed since there are 3 unknowns, i.e., x, y, and offset, and the equations are quadratic. Consequently, in a 3-dimensional (3D) case, a total of 5 equations would be needed. By solving the system of equations, the position of the UE, as well as the time offset between the ANs and the UE would be obtained. The latter can be used for time synchronization of the UE to the ANs.

In a real scenario, the equations describing intersecting circles would include noise and they can be written as:

$$r_i = r_{ti} - d_{offset} + e_i = \sqrt{(x_i - x)^2 + (y_i - y)^2}, \tag{6-19}$$

where, e_i is a random variable representing a zero mean normal distribution with variance σ , i.e. $e_i \sim \mathcal{N}(0, \sigma)$.

The value of the variance depends on the received SNR at the receiver, i.e. on the distance between the receiver and the transmitter.

Since the instantons value of the random variable e_i is not known, the equations describing circles around ANs would be only approximate and the described circles would not intersect in a single point. Therefore, the system of equations would not have an exact solution. In order to overcome this issue, a nonlinear least squares (NLLSQ) approach is used to find a solution with a minimum quadratic error. There are a few different approaches for solving NLLSQ problems, and in this case a Levenberg-Marquardt algorithm is used.

6.4.2 TWR model

The TWR approach is shown in Figure 6-20. The TWR is performed between a single AN and a UE. The AN sends a frame to the UE which responds with another frame to the AN. These frames are timestamped and the roundtrip time at the AN as well as the reply time at the UE are measured.

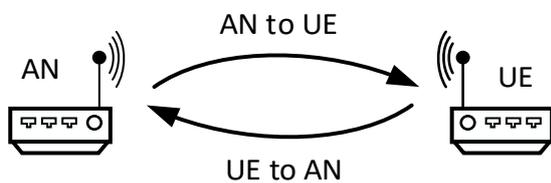


Figure 6-20 TWR approach

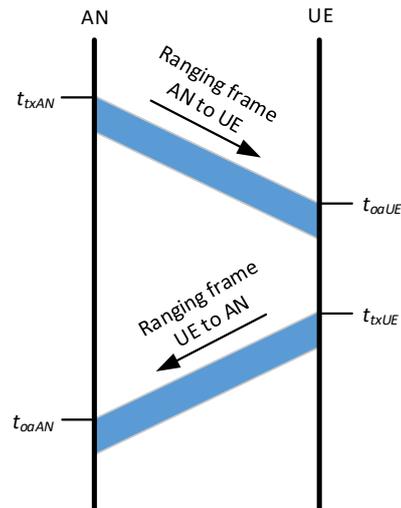


Figure 6-21 Time diagram of the TWR process

The time diagram of the TWR exchange is shown in Figure 6-20. The time of flight of the frames can be calculated as”

$$t_{of} = \frac{(t_{oaAN} - t_{txAN}) - (t_{txUE} - t_{oaUE})}{2}, \tag{6-20}$$

where, t_{of} is the time of flight, t_{oaAN} and t_{oaUE} are the time of arrival of the frames at the AN and the UE

respectively and t_{txUE} and t_{txAN} are the transmitting times of the frames at the UE and at the AN respectively. Having the time of flight, the distance between the AN and the UE can be estimated by multiplying with the speed of light.

In order to perform a 2D positioning, the TWR is performed between minimum of 3 ANs. A total of 3 distances between the ANs and UE would be obtained. These distances can be used as radii of 3 circles intersecting in the position of the UE. This is illustrated in Figure 6-19.

In the model used for simulation, it is expected that the time of arrivals are estimated with some noise. This noise would be translated into distance estimation noise. It is a white Gaussian noise and depends on the SNR of the received signal. Having a stronger signal, the distance estimation noise would be lower. Therefore, on all the distance estimates a noise is added. This noise is a Gaussian noise with variance σ^2 , i.e., $\mathcal{N}(0, \sigma)$.

6.4.3 RSS based model

The received signal strength approach is used in the LiFi positioning. The LiFi access points (AP) are used as an anchor node. They radiate the light from the ceiling to the floor. The approach is shown in Figure 6-22.

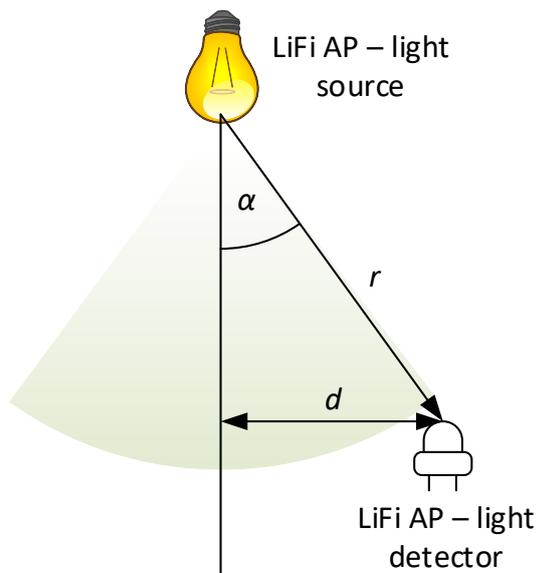


Figure 6-22 LiFi positioning approach

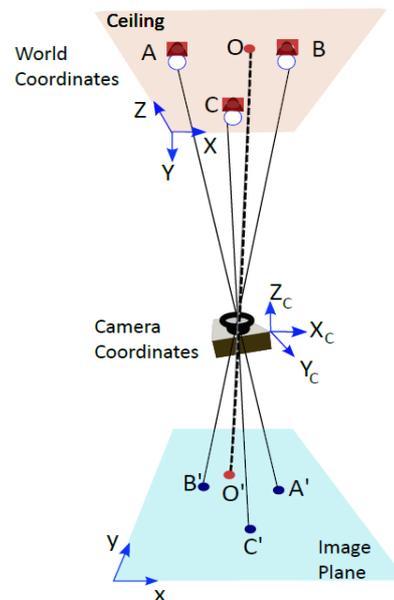


Figure 6-23 Visible light positioning system

The intensity of the light received at the photodetector depends on the distance between the light and the photodetector as well as the angle between the light emitter and light detector. The RSS value would therefore be:

$$RSS \sim \frac{1}{r^2} \cos \alpha. \tag{6-21}$$

In the model used for simulation, it is assumed that the light emitters would be pointed from the ceiling to the floor radiating perpendicular to the floor and that the light detector would be pointed perpendicular towards the ceiling. This would mean that the light incidence angle at the detector would be the same as the angle alpha. Having a fixed height between the emitter and the receiver, would enable estimation of the distance d .

Additionally, to further extend the model, an additional noise is added in the estimated RSS value.

6.4.4 Geometric method model

The schematic block diagram of visible light positioning (VLP) system is as shown in Figure 6-23 [34]. The model contains three LEDs which are evenly distributed on the ceiling, transmitting the codes and a complementary metal oxide semiconductor (CMOS) camera as a receiver. The light IDs received by the camera are then decoded by the receiver and 2-D image coordinates are extracted. From the known relationship between the 2-D image coordinates and 3-D real world coordinates of the fixtures, position of the camera can be estimated [5].

The transformation of 3-D real coordinates to camera coordinates and 2-D image points is as shown in Figure 6-23 [35]. The relationship between 2-D image points (x, y) to the real-world coordinates (X, Y, Z) is given by the equation:

$$w \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \cdot C \quad (6-22)$$

Where, w is a scaling factor, using which the camera coordinates can be scaled up to the image depth [5], C is the Camera matrix constituting extrinsic and intrinsic parameters. The rotation and translation parameters are known as extrinsic which are used to convert real world points into camera coordinates in 3-D. Which are then mapped to image points using intrinsic parameters such as focal length, the optical centre or principal point, and the skew coefficient [36].

The camera matrix is then given by the following equation:

$$C = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot [R | t] \quad (6-23)$$

Where:

- $[c_x \ c_y]$ is the optical center in pixel,
- $(f_x \ f_y)$ is focal length in pixel,
- $f_x = F/\rho_x$, $f_y = F/\rho_y$
- F is the focal length in mm,
- $(\rho_x \ \rho_y)$ is the size of the pixel,
- $s = f_x \tan \alpha$ is the skew coefficient which is non-zero if the image axes are not perpendicular.

By substituting (6-23) into (6-22), we can estimate the R and t parameters of the camera.

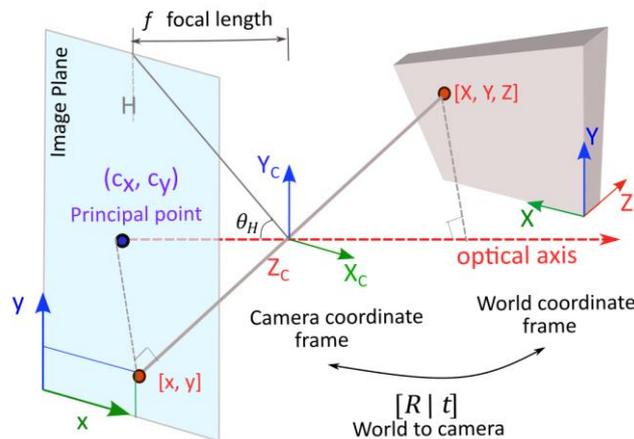


Figure 6-24 Visible light positioning geometrical model

Now if we know the rotation $[R]_{3 \times 3}$ matrix and translation $[t]_{1 \times 3}$ vector, then we can find the location of point $P (X_C, Y_C, Z_C)$ in camera coordinate system using the following equation [25]:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \Rightarrow$$

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = [R \mid t] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (6-24)$$

With this estimation, we can re-project and find the re-projection error of the LEDs. The R and t values are calibrated iteratively to find better estimates reducing the re-projection errors. This method can be further optimised using Levenberg-Marquardt optimization [35] [36].

In real scenario, the camera lens is affected by radial and tangential distortions. The light rays bend more at the edges of the lens than at its principal point resulting in radial distortion. The distorted points of undistorted (x, y) are calculated by the formula [36]:

$$x' = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (6-25)$$

$$y' = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (6-26)$$

- k_1, k_2, k_3 is radial distortion coefficients,
- $r^2 = x^2 + y^2$.

Tangential distortions occur due to divergent lens and image planes. The distorted points can be calibrated by the formula [36]:

$$x' = x + [2p_1 xy + p_2(r^2 + 2x^2)] \quad (6-27)$$

$$y' = y + [2p_2 xy + p_1(r^2 + 2y^2)] \quad (6-28)$$

- p_1, p_2 –tangential distortion coefficients of lens.
- $r^2 = x^2 + y^2$.

7 Joint Synchronization and Localization

As mentioned in the previous section, the goal of this section is to propose a solution for the localization problem in the context of 5G-CLARITY project by drawing on the time-stamp capability that is anticipated to be embedded in the communication nodes under the FTM standard [37]. This section is the continuation of Section 6 of 5G-CLARITY D2.3 [2]. The components related to this solution are highlighted in Figure 7-1. While the solution relies on the time-stamp exchange between the devices and the access nodes, it can be also integrated into the core of the positioning server where the algorithms for data fusion lie. In other words, this section presents algorithms which can reside in the core of the positioning server, i.e. localization server to perform accurate localization. A fundamental component of the solution proposed here is the ML-based NLoS identification developed in WP4, which serves as an assistant to the joint synchronization and positioning algorithm. Nevertheless, the implementation and evaluation remain in the simulation as the cornerstone of the methods utilized here is the time-stamp exchange capability which is yet not available in all the devices.

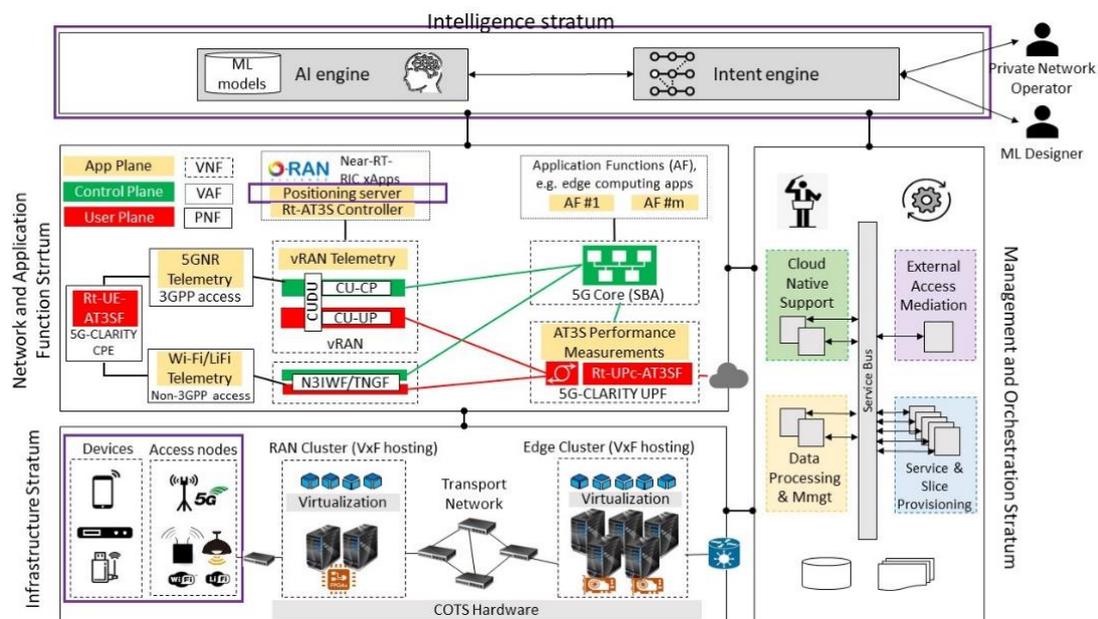


Figure 7-1 Architectural components of 5G-CLARITY system architecture involved in joint synchronization and localization.

7.1 End-to-end modelling tools

This section remains as it was explained in 5G-CLARITY D2.3 [2]. That is, we perform joint synchronization and localization with the aid of time-stamp exchange among the access points and the AGV. In particular, and AGV moves in areas where coverage is provided by different WATs with FTM capability as depicted in Figure 7-1. To boost the localization accuracy, we then tackle the synchronization and localization problem jointly. As mentioned before, the modelling here take a step beyond what was explained in the previous section and assumes FTM capability in all the nodes. The algorithm developed here can potentially reside in the core of the localization server where the fusion is performed.

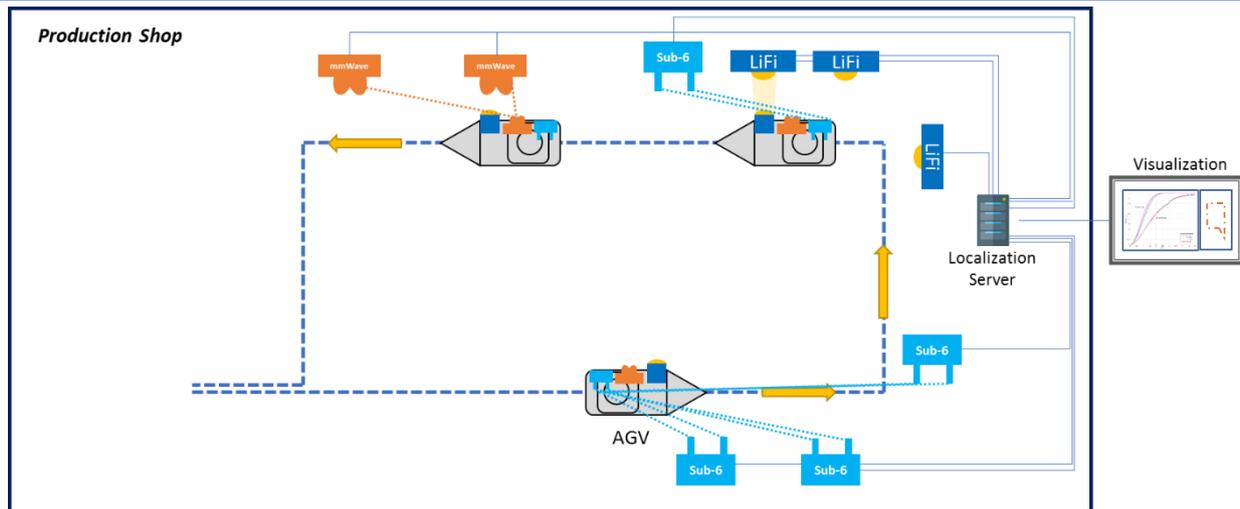


Figure 7-2 Positioning test scenario using multi-WATs

7.2 Modelling of functional components

7.2.1 Introduction

In [5G-CLARITY D2.3 \[2\]](#), we presented the principles of hybrid network synchronization, which paves the way for an accurate Mobile User (MU) joint synchronization and localization (sync&loc). In particular, hybrid synchronization enables precise inter Access Point (AP) synchronization, thereby permitting MU time-based localization and synchronization at the edge of communication networks. A Bayesian Recursive Filtering (BRF) based on joint MU sync&loc approach is developed, where Taylor expansion to linearise the non-linear relationship between the measurements and the position parameters is utilised. While Linearized BRF (L-BRF) can partially mitigate the destructive impact of non-linearities in the measurements, in addition to the covariance matrix underestimation, they are likely to diverge if a reliable estimate of the initial state is not available [38]. A promising approach, on one hand, to avoid such shortcomings of L-BRF and, on the other hand, to increase the accuracy of position estimation, is estimating the distributions, i.e., prediction, measurement likelihood, and posterior, by means of Particle Gaussian Mixture (PGM) filters introduced in [39]. Specifically, in this approach instead of a single Gaussian function, each distribution is approximated with a sum weighted of Gaussian functions, or, alternatively, Gaussian mixtures [40]. Nevertheless, the problem that immediately arises when using PGM filters is dimensionality, rendering the approach computationally expensive for multi-variable estimations. To overcome this drawback, we resort to a hybrid parametric and particle-based approach where we capitalize on the linear relationships in the measurements to reduce the dimensionality.

In this deliverable, based on [41], we propose a deep neural network assisted Particle filter-based (DePF) joint sync&loc algorithm which draws on the Channel Impulse Response (CIR) to estimate the Angle of Arrival (AoA) (using MUSIC algorithm [42]) and to determine the link condition – i.e., LoS or NLoS, using a pre-trained deep neural network thereby excluding the faulty measurements to enable a more precise parameter estimation. It then estimates the joint probability distribution of MU's clock and position parameters using the PGM filter. The dimension of the PGM filter is then reduced by revealing and exploiting the existing linear sub-structures in the measurements, thereby addressing the dimensionality problem.

However, there are several preliminaries for the PGM filter to return an accurate estimation of the MU's clock and position parameters. In addition to the times-tamp exchange mechanism explained in [5G-CLARITY D2.3 \[2\]](#), as mentioned above, AoA using MUSIC algorithm and deep neural network-based NLoS identification are the prerequisites for the DePF algorithm. We note that, in [5G-CLARITY D2.3 \[2\]](#), AoA

estimation was addressed using the Cramer-Rao bound, while for the link condition the assumption was that it is always LoS. In the following subsections, we briefly discuss these two prerequisites.

7.2.2 NLoS identification and Angle of Arrival

In this section, two algorithms whose outputs are necessary for PGM to return a reliable estimation are presented. Firstly, the MU-AP CIRs is fed into a deep neural network to identify the link condition. Then, the same CIRs are fed into the AoA estimation algorithm to estimate the direction of arrival. The former prevents the PGM from divergence while the latter increases the accuracy of position estimation.

7.2.2.1 NLoS identification and channel impulse response

The capability to estimate CIR is highly ubiquitous among the APs. Therefore, relying on the CIR to develop a localization algorithm appears to be a realistic approach. The AP-MU CIR is a rich source of information about the condition of the communication link, e.g., LoS or NLoS, and the location of the MU. More precisely, the former is crucial to know when estimating the latter as the accuracy of the distance, or time, and AoA measurements significantly drop declines if conducted under NLoS condition.

Figure 7-3 shows the architecture of the deep neural network deployed for NLoS-identification. We have used such a network in 5G-CLARITY D4.2 [6] to identify the link condition in indoor environments. The input layer has one channel fed with N samples, i.e., the magnitude of the CIR. The number of hidden layers and neurons in each hidden layer is set to l_H and n_H , respectively.

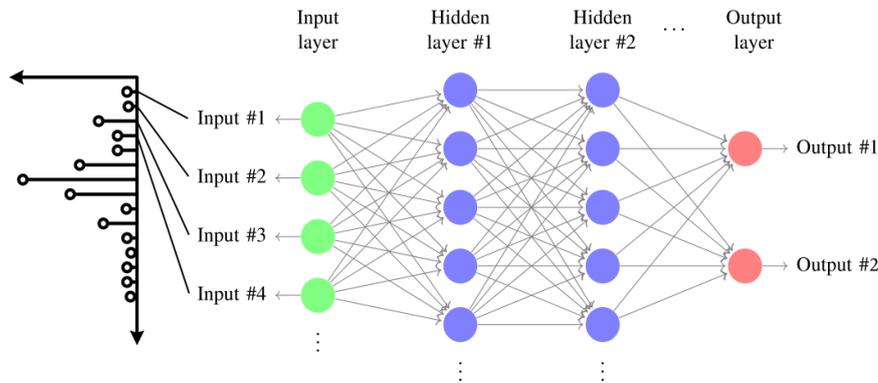


Figure 7-3 Deep neural network employed for NLoS-identification. It has $l_H = 2$ hidden layers with n_H neurons and two output neurons [41]

The rationale to rely on when selecting these numbers is that, according to [43], any classifier function can be realized by two hidden layers, i.e., currently there is no theoretical reason to use more than two. However, the lack of evidence does not imply that the deep neural networks with more hidden layers do not improve the accuracy of classification, it rather suggests that the number of required hidden layers does not follow a well-established logic and is mostly determined by a trial-and-error process. Therefore, for the algorithm proposed in this work, we empirically determine the l_H that delivers the best performance. Furthermore, as a rule of thumb, the number of neurons is suggested to be between the number of inputs and the number of outputs to prevent underfitting or overfitting. Let the output probability vector of the deep neural network be $[1 - \hat{p}_{nlos}, \hat{p}_{nlos}]$ where \hat{p}_{nlos} denotes the probability of the CIR being corresponded to an NLoS link. For the NLoS-identifier, we seek to train the deep neural network such that the output probability vector is as close as possible to the $[1,0]/[0,1]$ for the LoS/NLoS CIRs. In other words, from the optimization point of view, we aim to design a loss function whose output is small when the deep neural network returns the correct vector and is large otherwise. It turns out that the function that possesses the above-mentioned property is the logarithmic function [44]. Mathematically, the loss function is given by the following

expression:

$$L = -\frac{1}{M_c} \sum_i p_{nlos}^i \log(\hat{p}_{nlos}^i) + (1 - p_{nlos}^i) \log(1 - \hat{p}_{nlos}^i) \quad (7-1)$$

Where p_{nlos}^i denotes the true label corresponding to the i -th CIR sample in the data set, and is one if the CIR corresponds to an NLoS link and zero otherwise. Furthermore, M_c represents the total number of CIRs in the training set. The above formulation is also known in the literature as binary cross-entropy loss function. The goal of training is then to adjust the weights of the neurons such that the above loss function is minimized. Finally, when the trained deep neural network is employed in the context of joint sync&loc algorithm, the decision on the link condition is fed into the algorithm using the parameter ζ_i , which is set to one when $\hat{p}_{nlos}^i > 0.5$ and zero otherwise. In the sequel, we present the principles of the AoA estimation algorithm, which draws on the CIRs employed for NLoS identification.

7.2.2.2 Angle of Arrival

The CIR fed into the deep neural network to identify the link condition can be treated as an input signal to the MUSIC algorithm to obtain the AoA. We present the principles of AoA estimation for Uniform Linear Arrays (UPAs) based on [42]. The estimated AoA is given by the following expression:

$$(\varphi_{ij}, \alpha_{ij}) = \underset{\varphi, \alpha}{\operatorname{argmax}} \frac{1}{\mathbf{a}_n(\varphi, \alpha)^H \mathbf{N}^H \mathbf{N} \mathbf{a}_n(\varphi, \alpha)} \quad (7-2)$$

Where φ_{ij} and α_{ij} are the azimuth and elevation AoA of the signal received from the MU i at AP j , respectively. Parameter $\mathbf{a}_n(\varphi, \alpha)$ is the signal vector rotation on the n -th subcarrier. Matrix \mathbf{N} is constructed by $N_{ant}^2 - 1$ most right columns of the eigenvectors obtained when performing eigen decomposition of the covariance matrix of the received signal. That is,

$$\mathbf{R} = \mathbf{V} \mathbf{A} \mathbf{V}^H \quad (7-3)$$

where matrices \mathbf{A} and \mathbf{V} contain the eigenvalues and eigenvectors, respectively. Furthermore,

$$\mathbf{R} = \frac{1}{N_s} \sum_n \mathbf{x}_n \mathbf{x}_n^H \quad (7-4)$$

where the vector \mathbf{x}_n is of dimension $N_{ant}^2 \times 1$ and represents the n -th elements of Fast Fourier Transform (FFT) of the CIRs. Furthermore, N_s denotes the number of subcarriers, or alternatively, the size of FFT. It is worth mentioning that, when constructing \mathbf{N} , the eigen decomposition is assumed to sort the eigenvalues in decreasing order. Furthermore, each AP is assumed to have N_{ant}^2 CIRs at its disposal. Knowing the AoA and the channel condition, we can introduce the PGM filter based on which the DePF algorithm is designed. In the following subsections, we describe the details of the filter.

7.2.3 Particle Gaussian Mixture Filter

The idea underpinning PGM filters is to approximate the posterior Probability Distribution Function (PDF) defined as,

$$p(\boldsymbol{\xi}_i^k | \mathbf{c}_{ij}^{1:k}, \varphi_{ij}^{1:k}, \zeta_{ij}^{1:k}) = p(\mathbf{c}_{ij}^k, \varphi_{ij}^k, \zeta_{ij}^k | \boldsymbol{\xi}_i^k) p(\boldsymbol{\xi}_i^k | \mathbf{c}_{ij}^{1:k-1}, \varphi_{ij}^{1:k-1}, \zeta_{ij}^{1:k-1}) \quad (7-5)$$

by the sum of weighted Gaussian density functions (GDFs) [40]. Leveraging this idea, we can write the posterior expression as,

$$p(\boldsymbol{\xi}_i^k | \mathbf{c}_{ij}^{1:k}, \varphi_{ij}^{1:k}, \zeta_{ij}^{1:k}) = \sum_f w_f^k N(\boldsymbol{\xi}_i^k | \boldsymbol{\mu}_f^k, \boldsymbol{\Sigma}_f^k) \text{ with } \sum_f w_f^k = 1, w_f^k \geq 0 \forall f, \quad (7-6)$$

where $\boldsymbol{\mu}_f^k = [\boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}_i)_f^k, \boldsymbol{\mu}(\mathbf{p}_i)_f^k]^T$ and $\boldsymbol{\Sigma}_f^k = \text{diag}(\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_i)_f^k, \boldsymbol{\Sigma}(\mathbf{p}_i)_f^k)$ denote the mean vector and covariance matrix in the k -th round of estimation, respectively. Parameter F represents the total number of GDFs. Furthermore, $\boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}_i)_f^k / \boldsymbol{\mu}(\mathbf{p}_i)_f^k$ and $\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_i)_f^k / \boldsymbol{\Sigma}(\mathbf{p}_i)_f^k$, represent the mean vector and covariance matrix corresponding to the vector variable $\tilde{\boldsymbol{\theta}}_i / \mathbf{p}_i$, respectively. Considering the timestamp mechanism introduced in D2.3, we can conclude that $\tilde{\boldsymbol{\theta}}_i^k$, on one hand, is linearly dependent on the timestamps, and, on the other hand, does not depend on \mathbf{p}_i . This suggests that, although the $p(\mathbf{c}_{ij}^k, \varphi_{ij}^k, \zeta_{ij}^k | \boldsymbol{\xi}_i^k)$ is not Gaussian distributed in general, it is indeed Gaussian across the $\tilde{\boldsymbol{\theta}}_i$ axis. We capitalize on the linear Gaussian substructures in the model to keep the state dimensions low.

Consequently, the GDFs can be employed only across the \mathbf{p}_i axis transforming the structure of the above equation into the multiplication of a single GDF across $\tilde{\boldsymbol{\theta}}_i$ and sum weighted of multiple GDFs across \mathbf{p}_i (represented in Figure 7-4). Such a structure not only lays the ground for the hybrid parametric and particle-based implementation of BRf-based joint sync&loc estimation but also dramatically reduces the computational burden. Given above, the posterior expression can be simplified as,

$$p(\boldsymbol{\xi}_i^k | \mathbf{c}_{ij}^{1:k}, \varphi_{ij}^{1:k}, \zeta_{ij}^{1:k}) = N(\boldsymbol{\xi}_i^k | \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}_i)^k, \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_i)^k) \sum_f w_f^k N(\boldsymbol{\xi}_i^k | \boldsymbol{\mu}(\mathbf{p}_i)_f^k, \boldsymbol{\Sigma}(\mathbf{p}_i)_f^k) \quad (7-7)$$

We note that when $\boldsymbol{\Sigma}(\mathbf{p}_i)_f^k$ approaches 0, the term $N(\boldsymbol{\xi}_i^k | \boldsymbol{\mu}(\mathbf{p}_i)_f^k, \boldsymbol{\Sigma}(\mathbf{p}_i)_f^k)$ tends towards $\delta(\mathbf{p}_i - \boldsymbol{\mu}(\mathbf{p}_i)_f^k)$ where $\delta(\cdot)$ denote the Dirac impulse function. Such a function forms the basis of the classical particle filter.

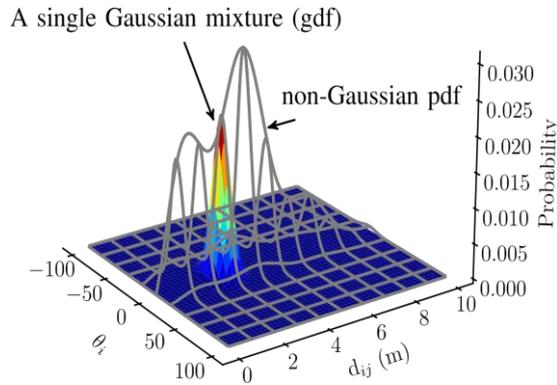


Figure 7-4 Example distribution of the $\boldsymbol{\xi}_i$ for a given timestamp measurement [41]

7.2.4 Scenario validation

We perform analysis for the scenario shown in Section 6.2 of 5G-CLARITY D2.3 [2], which is considered in [46] as a challenge. An agent starts its journey by accelerating until it reaches a speed of 14 m/s (= 50km/h). It continues moving with constant speed and decelerates upon approaching an intersection until it completely stops (e.g., due to the red light). The same is repeated between the two intersections. At the second intersection, it begins moving, then takes the turn, and continues to accelerate to 14 m/s limit until it exits the map. All the turns, as well as acceleration coefficients, are chosen randomly. During its journey, at each joint sync&loc round k , the MU exchanges timestamps with a fixed number of APs, the link to each of which is LoS/NLoS with the probability of 0.8/0.2. A further assumption is that at each joint sync&loc period T , $N_{\text{ant}} \times N_{\text{ant}}$ CIRs are available at each AP connected to the MU. In our simulations, the CIRs are obtained using the QuaDRiGa channel model. More explicitly, at each round k , knowing the true MU-AP distance and the link condition, i.e., LoS or NLoS, the CIRs are generated using the QuaDRiGa channel model. Moreover, the root mean squared errors (RMSEs) obtained by [46] serve as the baseline to our approach. The second scheme

with which we compare our proposed algorithm is the L-BRF filtering proposed in [47]. The aforementioned approaches are the most relevant as they draw on the same inputs as our proposed method does. We initialize all the clock offsets from the uniform distribution $U(-10^3, 10^3)$ nanoseconds (ns). The initial skews of all the clocks are drawn from the uniform distribution $U(1-10^{-4}, 1+10^{-4})$, which corresponds to skew values between 0 and 100 *part-per-million* (ppm). The covariance of clock process noise $Q_n(\tilde{\theta}_i)$ is set to $diag(10^{-5}, 100)$ to account for the residual errors from the previous iterations as well as the external noises on the clock skew and offset. The covariance of position process noise $Q_n(\mathbf{p}_i)$ amounts to $diag((14T)^2, (14T)^2)$ to account for every possible movement of the MU. A summary of simulation parameters can be found in Table 7-1. Furthermore, Table 7-2 indicates the components employed to perform the simulations and scenario validation.

Table 7-1 Simulation parameters.

Parameter	Value
# of simulations	1000
Random delay/skew	$U(-10^3, 10^3) / U(1-10^{-4}, 1+10^{-4})$
Max user velocity	14 m/s
Distance traversed by the user	600m
Period of joint sync&loc	100 ms
# of DNN hidden layers/neurons	2/50
Optimizer	Adam

Table 7-2 Validation components.

Components	Background	Extensions in 5G-CLARITY	Justification
Python-based Hybrid synchronization	Developed in 5G-PICTURE [45]	Combined with new algorithm to perform end-to-end synchronization and user localization	Lays the ground for an accurate localization
Python-based linear Bayesian mobile user joint synchronization and Localization	N/A	Developed from scratch	The first step towards a more accurate synchronization and localization
Python-based particle filter driven mobile user joint synchronization and Localization	N/A	Developed from scratch	It can model the realistic aspects of the scenario accurately

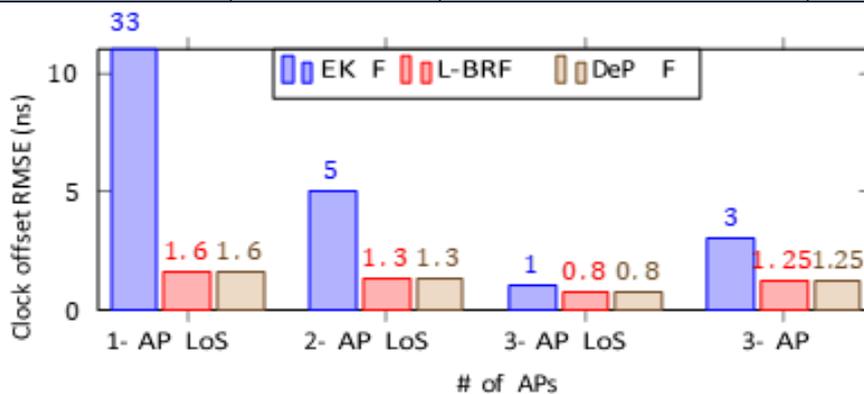


Figure 7-5 Performance comparison of the three joint synchronization and localization algorithms in terms of clock-offset estimation

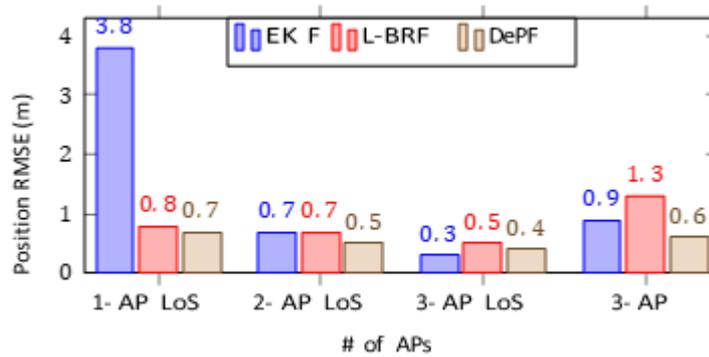


Figure 7-6 Performance comparison of the three joint synchronization and localization algorithms in terms of position estimation.

Figure 7-5 shows the RMSE of clock offset estimation for three joint sync&loc algorithms. The DePF algorithm is compared with two linear Bayesian methods, i.e., EKF and L-BRF, in multiple scenarios. The RMSEs in three scenarios is computed, with the number of LoS APs ranging from 1 to 3. In another additional scenario, we consider the MU being connected to three APs, where each MU-AP link condition is set to LoS with the probability of 0.8. As can be seen, for all the LoS scenarios the L-BRF and DePF deliver an identical performance, which is expected as they rely on the same approach to estimate the clock parameters. On the other hand, the performance of the EKF falls behind as it does not explicitly draw on the synchronization signals to estimate the clock offset. Moreover, the synchronization algorithm scheme utilized to synchronize the APs, i.e., hybrid BP-BRF network synchronization, leads to a more precise inter-AP synchronization and, consequently, lowers the MU clock offset estimation error. In the last case, the L-BRF and DePF that are drawing on deep neural network based NLoS identification outperform the EKF-based method where the NLoS links are identified by means of Rice factor of the incoming signal strength.

Figure 7-6 depicts the RMSE of position estimation for three joint sync&loc algorithms. The DePF algorithm is compared with two linear Bayesian methods, i.e., EKF and L-BRF, in the same scenarios as in Figure 620. As can be seen, for almost all the scenarios the DePF algorithm delivers superior performance. In particular, since the DePF employs a higher number of GDFs, rather than only one, to approximate the posterior distribution, it can estimate the position more accurately. Furthermore, DePF stands out when dealing with NLoS links. This is straightforward to notice as the RMSE of position estimation is lower for DePF in the 3-AP scenario where the L-BRF employs the same NLoS identifier as DePF. Additionally, unlike EKF and L-BRF, DePF does not need any initialization, which is of crucial importance in practice as initialization would require the APs to request position estimation from the MUs, which may not be always possible. Overall, considering 2-AP LoS, 3-AP LoS, and 3-AP scenarios, EKF and L-BRF perform close to DePF when a reliable initialization and MU-AP links with known LoS conditions are available. Nevertheless, such assumptions are questionable in practice, rendering the EKF-based and L-BRF algorithms futile in real-world scenarios.

Hereafter, all the simulations have been carried out assuming that there is always at least one LoS MU-AP link. Figure 7-7 presents the CDF of clock offset estimation error when the MU is connected to multiple APs. The estimation accuracy always remains below 2 ns and increases as both L-BRF and DePF utilize more measurements to estimate the clock offset and skew. Since the APs are precisely synchronized, collecting timestamps from each additional AP does provide additional information about the statistics of MU's clock parameters and, therefore, increases the accuracy of the estimation. Such precision is necessary if the location of the MU is to be accurately estimated.

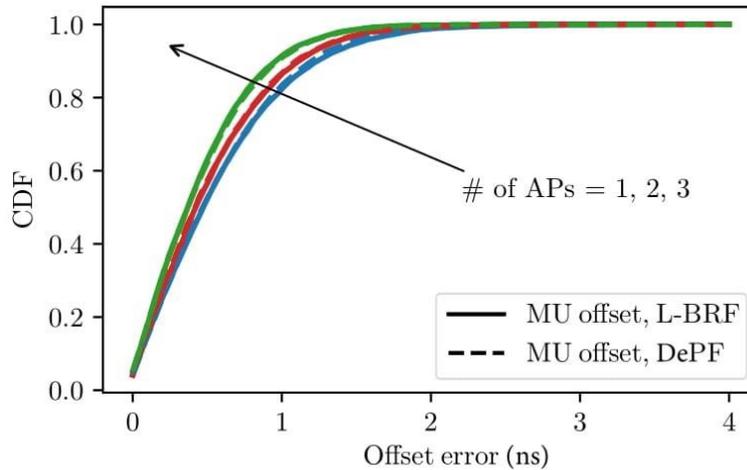


Figure 7-7 Performance comparison of L-BRF and DePF when estimating the MUs' clock offset [41]

It is noted that each single ns inaccuracy maps to 0.3 m distance measurement error and, consequently, worsens the location estimation. Furthermore, the performance of both schemes is identical as they draw on the same approach to estimate the clock parameters. That is, both schemes model the clock parameters with a single GDF.

Figure 7-8 presents the CDF of position estimation error when the MU is connected to multiple numbers of APs. As can be seen, the position estimation error is less than 1 meter in 90% of the cases for the DePF algorithm. We observe that DePF significantly outperforms the L-BRF, especially for 2- and 3-AP scenarios. Unlike the L-BRF that approximates the posterior with a single Gaussian distribution, in DePF the approximation is based on multiple GDFs. Consequently, the approximated posterior is closer to the true one, leading to a more precise position estimation. Another observation is that, although the position estimation error decays with the growth in the number of APs, increasing the number of APs from 2 to 3 only slightly improves the performance. In fact, the third AP is normally far away from the MU, leading to a poorer (AoA and timestamp) measurement accuracy compared to that of the first two APs. Hence, it does not provide substantial further information about the posterior distribution of the MU's location.

Figure 7-9 indicates the CDF of position estimation for multiple numbers of GDFs. It can be noticed that the position estimation improves with the increase of the number of GDFs. This is expected as in PGM filters the posterior distribution is approximated by multiple GDFs. Consequently, the more GDFs we employ, the more accuracy we achieve, albeit with higher computation time. Nevertheless, the error reduction is decreasing when increasing the number of GDFs, suggesting that a proper balance needs to be struck between the number of GDFs and the localization accuracy. In the scenarios presented in this work, one can achieve satisfactory performance even with 500 GDFs.

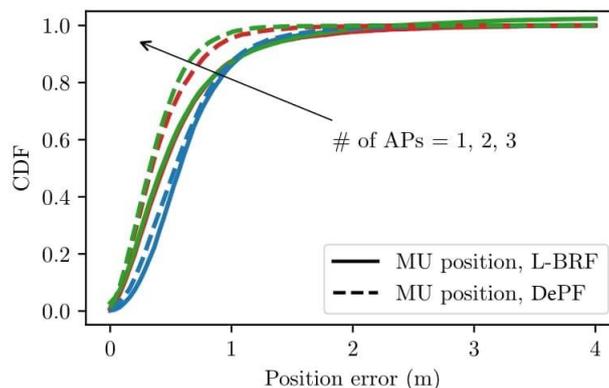


Figure 7-8 Performance comparison of L-BRF and DePF when estimating the MUs' position

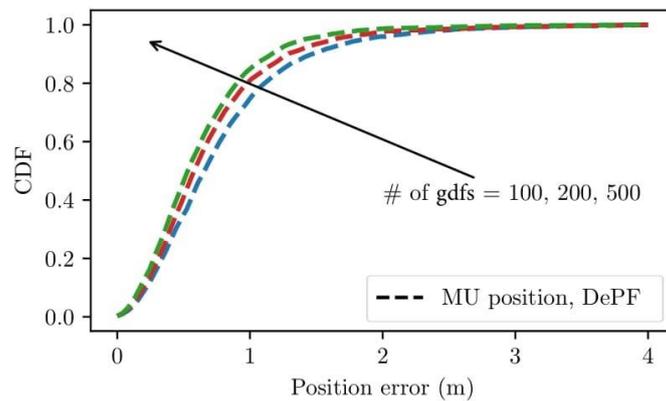


Figure 7-9 Performance of joint sync&loc algorithm for different number of GDFs [41]

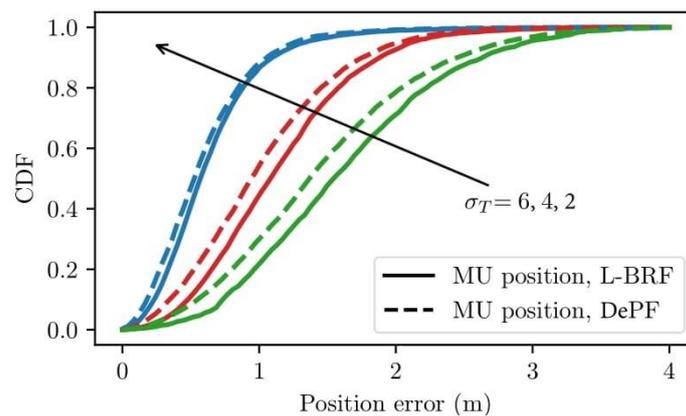


Figure 7-10 Position estimation performance of joint sync&loc algorithm with different timestamp accuracy [41]

Figure 7-10 shows the CDF of position estimation conducted by a single AP for different timestamp accuracies. It can be noticed that the position estimation accuracy deteriorates with the growth in the timestamp uncertainty. Specifically, the growth in uncertainty results in more erroneous distance measurements and offset estimations, which, consequently, worsens the position estimation accuracy. Nevertheless, it can be readily seen that DePF is more successful in mitigating the destructive effect of the timestamp uncertainty. Moreover, for both DePF and L-BRF, employing more APs can alleviate the negative impact of large timestamp uncertainty. Furthermore, it can be noticed that σ_T plays a decisive role in the outcome of the estimation algorithm, which also reveals the importance of hardware components in the design of a robust and precise joint sync&loc algorithm. In practice, such uncertainty in the commercial off-the-shelf devices is expected to be below 5 ns.

In summary, deep neural networks can play a decisive role by facilitating accurate decision making in simple but crucial tasks, such as NLoS identification. Furthermore, it can be noticed that in the case when we have multiple LoS links available the performance of the EKF-based and L-BRF approaches both in terms of clock offset and position is close to that of DePF. Nevertheless, in the absence of LoS condition, DePF demonstrates more competence in estimating the clock and position parameters by employing only a few hundred GDFs. Another point worth mentioning is that timestamp exchange has a high potential to be employed for performing joint synchronization and localization. In particular, the current communication devices can perform timestamp exchange up to 5 ns accuracy, fertilizing the ground for precise offset and distance measurements, which are the basis for precise joint synchronization and localization.

8 Pathways Towards 6G

In this section, we outline how 5G-CLARITY system should evolve beyond project’s lifetime to meet 6G expectations, with a clear focus on the ‘network of networks’ concept. This concept pushes well beyond the multi-connectivity mechanisms (e.g., AT3S) and private/hybrid network solutions (e.g., PNI-NPN) envisaged for 5G/B5G, with the definition of a novel solution suite enabling limitless connectivity and global coverage support, backed with a collection of networking, security, and orchestration capabilities. This new 6G approach will enable the integration of *a) Non-Terrestrial Networks*, i.e., satellite networks; *b) combined cell/multi-point transmission and seamless handover support across legacy/novel access technologies*; *c) L1/L2-mobility, Distributed MIMO and Multi-Transit Points*; *d) Device-to-Device communications (D2D) and mesh topologies*; *e) fully connected private-public networks fabric, with the possibility of integrating locally created ad-hoc networks to existing PNI-NPNs, making the latter recursive and scalable, following plug-and-play approaches.*

The analysis done in this section follows the principles of technological prospecting [46], and touches on different 6G facets, including ecosystem (Section 7.1), technology drivers (Section 7.2), use cases (Section 7.3) and architecture (Section 7.4). The goal is to demonstrate that 5G-CLARITY system lays the foundation for a robust and scalable service enablement platform allowing for ‘network of networks’ scenarios in 6G.

8.1 6G ecosystem

8.1.1 State-of-the-art

Like what was done in METIS when 5G technology started to be discussed, now it is time to carry out a similar exercise for 6G, by defining *a) societal, policy and business drivers* (and key value indicators, KVIs); *b) intended goals*; and *c) KPIs*. The table below provides a summary of the different drivers for upcoming generation, according to the guidelines outlined in [47].

Table 8-1 Drivers for 6G

Category	Description
Societal drivers	<ul style="list-style-type: none"> • European Green Deal • Human centric approach • End-user engagement • All-inclusive, contributing towards the reduction of digital drivers • Trustworthy AI Train core skills on STEM
Policy drivers	<ul style="list-style-type: none"> • A federated data infrastructure for Europe (GAIA-X) • Strengthen position of EU actors in ICT and OT industries Promote EU leadership in fabrication technology for silicon assets, to reduce dependencies on external markets.
Business drivers	<ul style="list-style-type: none"> • Broaden ‘deskless’ workers concept in the ‘enterprises of the future’. • Universal digital replications of real-world entities, leveraging digital twinning. Healthy ecosystem, backed with a fair regulation for all the stakeholders taking part in the value chain.

Figure 8-1 shows an overview of *i) the main functional requirements that 6G based service platforms shall provide to take the challenges we will face in 2030 and beyond*; and *ii) target KPIs, comparing them with the ones set for 5G*. Note that numbers resulting from this comparative analysis are estimated based on YoY growth in mobile data traffic volume and studies led by relevant analyst reports.

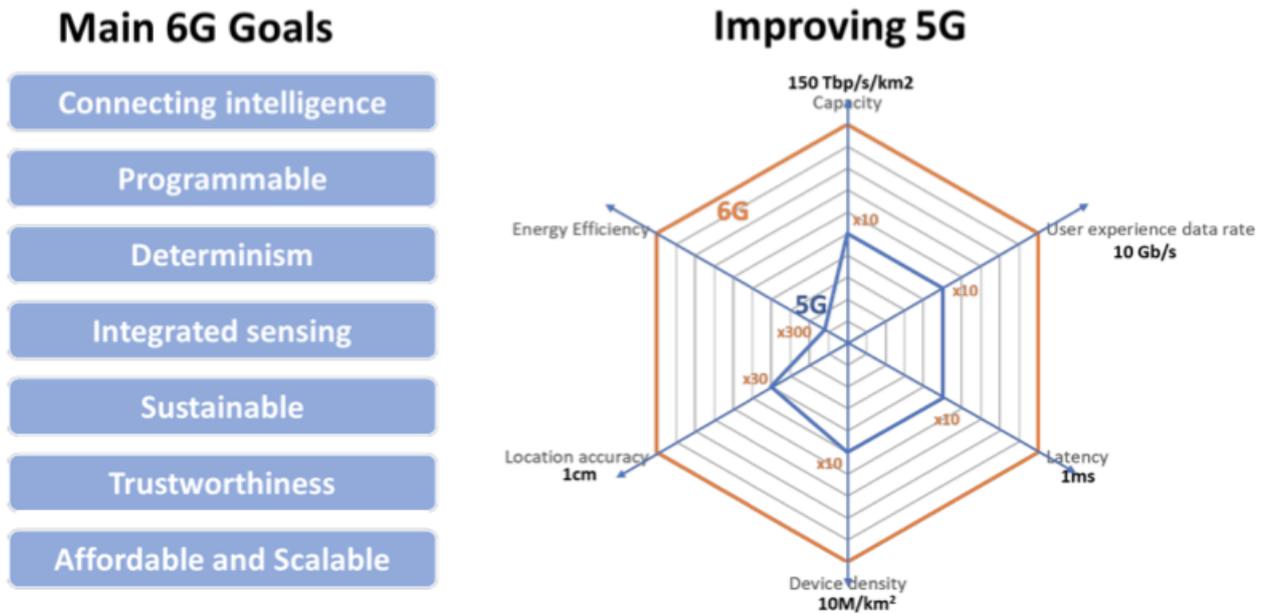


Figure 8-1 6G Goals and target KPIs [47]

8.1.2 5G-CLARITY approach to 6G

To make the 5G-CLARITY system compliant with 6G expectations, it is important to:

- Keep track of regulation activities, with a special attention to everything related to spectrum management, network neutrality, data privacy and management, all of them quite relevant for private/hybrid 6G networks. It can be foreseen that at some point, in the 6G timeframe, AI agents of regulators are embedded in networks. It is expected to keep a close eye on the progress of all these regulation aspects, to make sure we can adapt 5G-CLARITY assets and services accordingly.
- Push work towards the ICT-OT industry convergence in 6G, with an IoT-to-edge-to-cloud continuum and a single security management fabric that brings together telco (e.g., 3GPP) and OT security solution into a unified, coherent framework.
- Prioritize work on digital twinning scenarios and underlying technologies, as a key enabler to optimize business process and speed up rollout of services, especially when these call for low latency and high reliability requirements.
- Evolve original business models and actor-roles relationships (see 5G-CLARITY D2.2 [1]) to shape a healthy ecosystem, building on a regulation-compliant governance model that *i*) promotes service innovation in the network of network ecosystem lowering entry barriers for SMEs or other incumbent actors; and *ii*) distribute revenue streams fairly, in such a manner that stakeholders can use part of these revenues to keep investing in new assets.

In terms of in-scope functional requirements, 5G-CLARITY evolution should focus on:

- **Connected intelligence**, building up a data and connectivity infrastructure supporting cooperation of trusted AI functions from different networks (different type, different scale, and different ownership). This will require further elaboration and more solid basis on the ideas on AI distribution and cooperation in PNI-NPNs, captured in 5G-CLARITY D4.2 and D4.3.
- **Programmability**, turning networks from connectivity platforms to service enablement platforms, by applying service-based architecture patterns (RESTful-based HTTP APIs) across all layers and domains. In 5G-CLARITY, this will require an evolution of network and function stratum, and further

refinements in the Mediation Function.

- **Determinism**, with end-to-end perspective, from the end device to the last application function (beyond 3GPP scope of work). In 5G-CLARITY, this will require evolving TSN capabilities with Rel-18+ features and integrating model predictive control (MPC) solutions for Tactile Internet applications.
- **Integrating sensing and communication**, to achieve (i) cm-accuracy localization, critical in indoor environments for some I4.0/I5.0 applications; and (ii) context awareness, critical when having multiple networks connected to each other, following the ‘network of networks’ approaches, with multiple nodes and data sources that have impact on each other. The issue pointed in (i) will require an improvement of the capabilities of 5G-CLARITY localization server, while issues pointed in (ii) will take the insights and conclusions gathered from Section 3.4 (where the use of DSF for public-private networks is presented) as a starting point.
- **Sustainability**, aiming at a more environmental-friendly overall system to comply with the target goals documented in the European Green Deal. From 5G-CLARITY system viewpoint, we see data collection as a potential target for energy optimization, shifting from today’s practices (store everything into a data lake for later usage) towards a more selective, fine-grained approach based on the use of adaptable telemetry models. This optimization will allow for significant reduction in resource consumption for assurance activities (5G-CLARITY management and orchestration stratum) and AI/ML model training (5G-CLARITY intelligence stratum).
- **Trustworthy**, with a special focus on defining a security framework that allows protection of data in software environments, achieved with a continuous verification of all infrastructure assets, across all layers, to minimize surface attacks. This is a major gap to work out in 5G-CLARITY, especially considering the evolution of network of networks, with multiple non-trusted stakeholders that should operate under liability models.
- **Affordability, scalability**. 5G-CLARITY is designed from scratch to be a scalable solution, thanks to the SBMA approach. However, work is still to be done in this regard to meet expectations of 6G, based on ease replicability of solutions following the ‘design-once-deploy-everywhere’ approach. New provisioning and activation patterns need to be worked out to turn 5G-CLARITY into a system that can be set up in a matter of minutes with one-click.

8.2 6G technology drivers

8.2.1 State-of-the-art

Figure 8-2 depicts the envisaged key technologies for 6G, and their relationship with the 6G goals and target KPIs that were commented in Section 7.1.1. Table 8-2 provides a brief description of the scope of these technologies, based on the concepts elaborated in [47], [48].

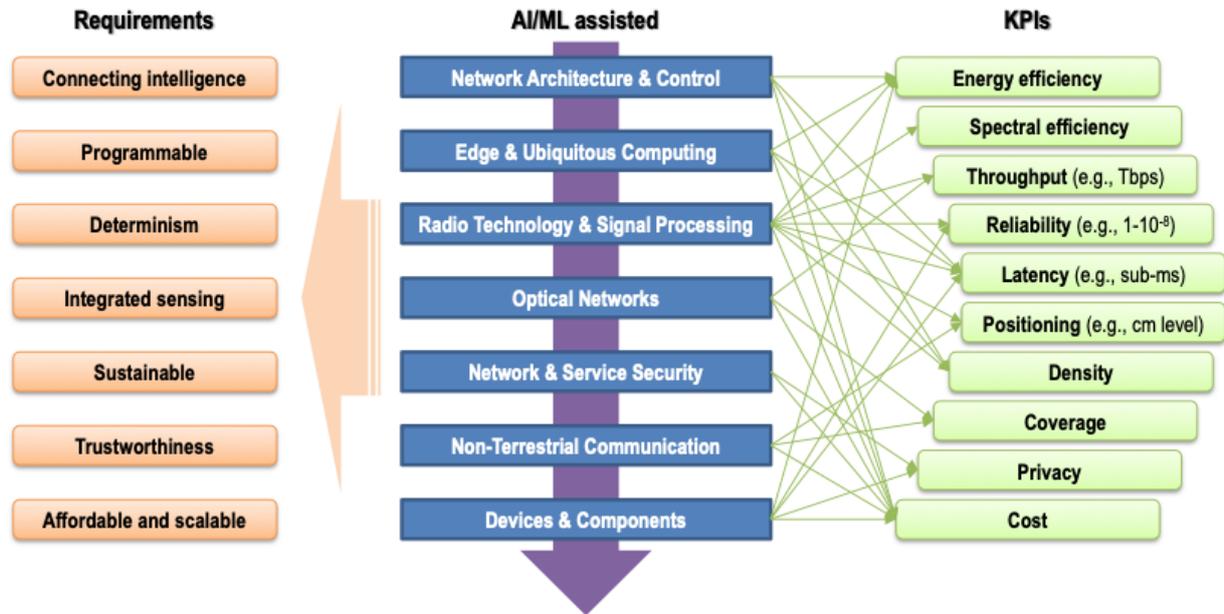


Figure 8-2 Technologies for 6G and relationship with 6G goals (requirements) and KPIs

Table 8-2 Key Technologies for 6G

Key Technology	Description
Network Architecture and Control	Allowing for a very versatile, pervasive and autonomic resource control, with the capability to maintain the controllability of all available resource components in multi-tenant environments. This controllability shall be preserved regardless of (i) nature, density and address of resources, and their topology distribution; (ii) usage loads and fluctuations; (iii) own operational decisions of the operator; and (iv) operational decision of 3 rd parties.
Edge & Ubiquitous Computing	Allowing for in-network computing, with computing and networking resources merged into a single computing continuum. This continuum stretches from user/IoT devices to centralized cloud, meaning that edge computing as an independent concept would fade away. The main objective is to go for a much leaner resource management of distributed microservices, while reducing delays and increase responsiveness (e.g., zero-perceived latency).
Radio Technology & Signal Processing	Allowing for an improved air interface design, with innovation on different facets: spectrum (spectrum reutilization, optical wireless communication, mmWave and THz communications), radio building blocks (waveform, multiple access and full-duplex designs, coding and modulation) and add-ones (integrated positioning, sensing and communication, wireless edge caching, etc.)
Optical Networks	Approaching optical infrastructure featuring higher capacity, lower latency, increased programmability, enhanced reconfigurability, increased environmental hardening and cost reduction. This will require looking for disruptive solutions at the physical layer (wavelength bands, space division multiplexing), data layer (hybrid use of electronic and optical switching) and management layer (AI-assisted, data-driven reconfiguration of devices), all backed with appropriate security and fault protection solutions.
Network and Service Security	Allowing for an end-to-end trustworthy system, building on a security management framework that (i) is green, ensuring that is not detrimental to the level of protection; (ii) incorporates novel security strategies such as deception moving target defence and DLT; and (iii) protect AI assets; and (iv) can be offered as a service
Non-Terrestrial Networks	Satellites can act as a main single backhaul segment for rural areas, aircraft, vessels and trains; as additional backhaul means to opportunistically provide additional connectivity

	resources; or as a pure transport subnetwork. This line of working is about current satellite architecture towards a flexible, and yet scalable and cost-efficient, hierarchical architecture able to support space-borne and air-borne nodes, and seamlessly integrated with terrestrial networks (PLMNs and NPNs).
Devices & Components	Keeping up with the progress on the different communication system assets, such as processors, memories, RF, DAC/ADC, antennas, packaging, and optical components. In all these components, disruptive approaches are needed to approach Shannon's limits and push Moore's limits out further.

8.2.2 5G-CLARITY approach to 6G

From the key technologies listed in Section 7.2.1, we identify the ones that 5G-CLARITY should focus on towards the support 'network of networks' scenarios in 6G. For those which are in-scope, we provide guidelines on how they may allow 5G-CLARITY system to evolve appropriately.

- Network architecture and control.** Driven by the capability to meet ultra-reliable and low latency requirements, NPN capabilities were defined in Rel-16, and we are beginning to see their use in vertical industries for industrial automation or in campus networks, both in private and hybrid modes (SNPN and PNI-NPN, respectively). 5G-CLARITY system already supports these capabilities, as reported in 5G-CLARITY D2.2 [1] and in this deliverable. However, for 6G, these capabilities will be pushed out to current thought limits. The trend will likely further expand resulting in increasing demand for industry customers with application to special purpose networks (e.g., network of swarm or drones) or to meet even smaller range sub-networks (e.g., in-body, in-robot, in-car sub-networks) that can generally operate in a standalone fashion but may benefit from connectivity to the wide area network. To keep up with these new scenarios, 5G-CLARITY need to evolve into a more flexible and scalable system, with modifications on three strata: network and application function stratum, management and orchestration stratum, and intelligence stratum. At the first stratum, there is a need to *expand SBA* not only to RAN control plane, but also beyond operator's trusted domain, to allow for an easy yet controlled plug-and-play of functions from different networks into a single bus. At the second stratum, *information models governing OSS shall evolve* from today's parent-child relationships to more name-containment relationships, to allow for recursive definition of managed resources, and thus composability patterns across them. Finally, at the intelligence stratum, there is a need to work on the *formalization of intent language*, shifting from today's 5G-CLARITY proprietary intent semantics towards open, global, and user-friendly intent semantics. This is key to facilitate easy composition and integration of networks from different stakeholders at operation time.
- Edge & Ubiquitous Computing.** 5G-CLARITY will need to leverage the cloud continuum with a twofold purpose. On the one hand, allowing an effortless of portability of workloads (telco/IT/application functions) across all the nodes building up this continuum, from in-device edge to hyperscaler cloud, including the compute tiers in between (i.e., on-prem/far/telco edge). On the other hand, the harmonization of all cloudification technologies (e.g., VM, unikernels, containers) and associated service delivery models (e.g., IaaS, CaaS, and FaaS/serverless solutions) under a single unified resource management framework that facilitates the natural evolution of telco systems towards distributed (micro)service-based architecture.
- Radio Access Technology & Signal Processing.** With a clear focus on multi-access connectivity from the very beginning, 5G-CLARITY need to keep working on the *evolution of air interfaces*, and their combination/aggregation, in such a way 6G KPIs can be met. These KPIs, illustrated in Figure 8-1, impose far more stringent requirements, such as Tbps throughput, sub-ms latency to the networking layer, 12-nine's reliability, 10e-8 packet error rate, cm-level accuracy localization and extreme

energy consumption (and their combination/aggregation). In addition to this work, lines of work revolving around *spectrum reutilization* and *integrated positioning, sensing and communication* are need when having multiple sub-networks, of different nature and with different scales, connected to each other following the ‘network of networks’ paradigm. On the one hand, it is crucial to make a joint use of licensed and unlicensed, across all bands yet intelligently, using cognitive radio-based solutions. On the other hand, integrated positioning, sensing and communication will provide context awareness, which has been identify as a key enabler in many ‘network of networks’ scenarios, including autonomous cars, factories of the future, smart cities and public safety (e.g., V2X vulnerable road user discovery). To that end, 5G-CLARITY will take as a basis the positioning and localization capabilities presented in WP2 and developed in WP3, using them as a starting point to work out challenges ahead, including interference management, time sharing between positioning and communications, and information/estimation theory limitations, to name a few.

- **Network and Service Security.** According to the expectations captured in Table 8-2, work in 5G-CLARITY system shall be articulated into two main directions: trust distillation and secure data-driven management. The aim of *trust distillation* is to identify, evaluate, certify and monitoring the level of security in complex and fragmented environments, with the use of AI models. These models can help toto detect and deter miscreants, with a dynamic identification of devices and workloads and issuing adaptable, context-aware responses, thus giving trust to providers and hosted services. On the other hand, the *secure data-driven management* represents the capability of protecting the new oil, which is data, based on telemetry trustworthiness and AI resilience (AI/ML functions and models need to be hardened to avoid malfunctioning and cyberattacks).

Though not listed in [47] and [48] (and therefore not captured in Table 8-2 and Figure 8-2), we identify integrated sensing and communications (ISAC) as another 6G key technology, and also in roadmap for 5G-CLARITY evolution. ISAC represents a paradigm change where the previously competing sensing and communication operations can be now optimized together (allowing considerable gains in terms of spectral, energy and cost efficiency) via the shared use of a single hardware platform and a joint signal processing framework. The lines of work in this technology cover multiple dimensions, from physical layer design (e.g. waveform, sequence, coding, modulation, beamforming) to network architectures and transmission protocols, with the security and privacy as a transversal yet critical feature to be support.

8.3 6G use cases

8.3.1 State-of-the-art

Different initiatives have started to outline the use cases for the upcoming generation. Among these initiatives, there is Hexa-X [49], which is the European 6G flagship project. HEXA-X project list a wide variety of use cases, and cluster them into five use case families, according to the types of usages, as well as the research challenges and values addressed. Table 8-3 captures these use case families and provides a brief description of their intended scope.

Table 8-3 Hexa-X Families of Use Cases

Use Case Family	Description	Examples
Sustainable Development	Illustrates how 6G can contribute to the society transformation, targeting UN sustainable development goals and the EU Green Deal, providing global access to digital services and energy-optimized infra facilities.	E-Health for all; institutional coverage; earth monitor; autonomous supply chains
Massive twinning	Involves the massive use of digital twins to represent, interact and control actions in the	Digital Twins for manufacturing; immersive smart city; digital twins for

	physical world	sustainable food production
Telepresence	Covers immersive telepresence for enhanced interactions, involving mixed reality or merged reality, providing extreme and fully immersive experience	Fully merged cyber-physical worlds; mixed reality co-design; immersive sport event; merged reality game/work.
Robots-cobots	Includes application scenarios involving interacting robots, at home (to facilitate everyday life) as well as professional/industrial environments (to improve the efficiency of processes)	Consumer robots; AI partners; interacting and cooperative mobile robots; flexible manufacturing
Local trust zones	Targeting both humans and machines, it encompasses networks of different types and scale, from in-body networks to wide area deployment of sensors networks.	Precision healthcare; sensor infrastructure web; 6G IoT micro-networks for smart cities; infrastructure-less network extensions and embedded networks; local coverage for temporary usage; small coverage, low power network of networks for production & manufacturing; automatic public security.

Building on these use case families, HEXA-X has additionally:

- Identified seven enabling services harness new capabilities: (i) compute-as-a-service, for resource-constrained devices; (ii) AI-as-a-service; (iii) AI-assisted V2X; (iv) flexible device type change; (v) energy-optimized services; (vi) internet-of-tags; and (vii) security as a service.
- Defined novel qualitative and quantitative indicators to assess use case behaviour, clustering them into four dimensions: (i) the evolution of Key Performance Indicators (KPIs) to meet 6G technical targets; (ii) revolution of new E2E measures; (iii) the need to capture new network capabilities; and (iv) the definition of meaningful and measurable Key Value Indicators (KVI), featuring the trustworthiness, inclusiveness and sustainability abilities on value creation for products, services and society.

A representative subset of HEXA-X use cases is pictured in Figure 8-3, to discuss deployment scenarios (left side) and associated KPIs/KVIs (right side). One use case has been chosen for each use case family, based on the representation of the use case family. The selected use cases are E-health for all, immersive smart city, fully merged cyber-physical worlds and interacting, cooperating mobile robots, and dynamic and trusted local connectivity.

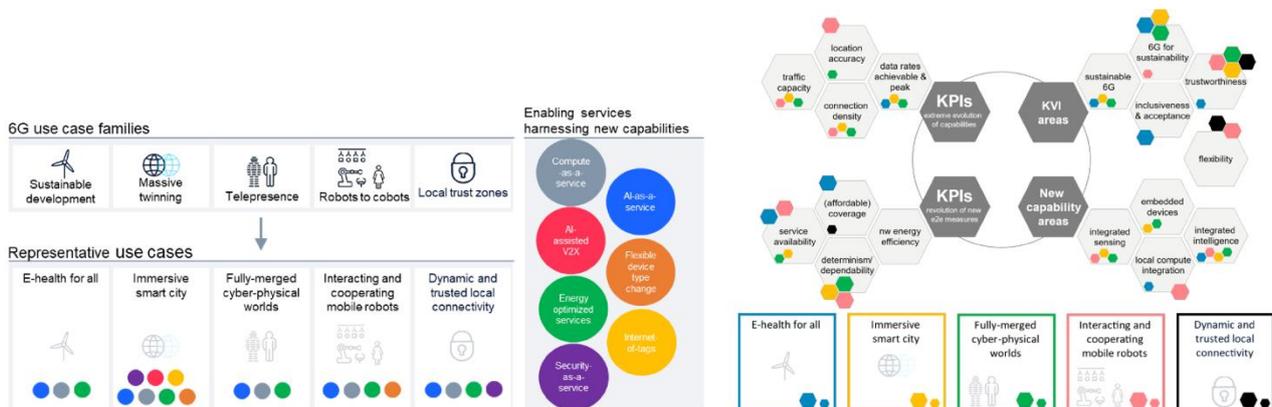


Figure 8-3 Representative HEXA-X use cases, enabling services and mapping to KVIs/KPIs [50]

Other representative initiative is Oulu 6G Flagship initiative [51], which has issued several white papers on 6G, covering a wide range of topics. The complete list of white papers can be found at [52]

8.3.2 5G-CLARITY approach to 6G

From all the use cases collected from HEXA-X and Oulu 6G-Flagship initiatives, in this section we clarify the ones that we foresee are within the scope of evolved 5G-CLARITY system. Figure 8-4 and Figure 8-5 represents selected use cases and Table 8-4 shows their relationships with the technology levers listed and elaborated in Section 7.2.2.

23 use cases, clustered in to 5 families

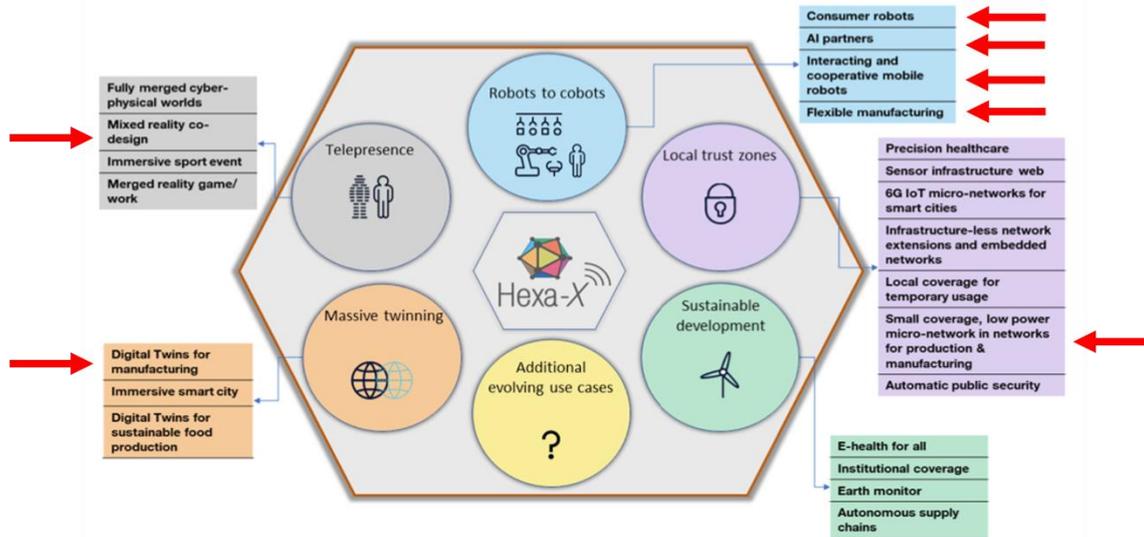


Figure 8-4 In-scope HEXA-X use cases (indicated by red arrows)

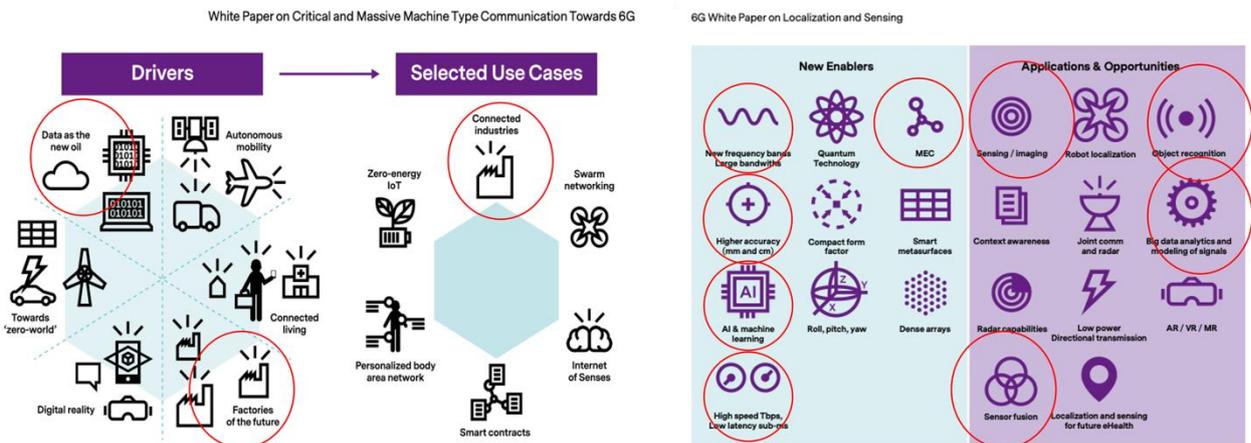


Figure 8-5 In-scope Oulu 6G Flagship use cases (red circled)

Table 8-4 On the Relationship with In-Scope Technologies with Relevant 6G Use Cases

Initiative	Use Case	Network Architectures and Control	Edge & Ubiquitous Computing	Radio Access Technology & Signal Processing	Network & Service Security	ISAC
HEXA-X	Mixed reality co-design	X	X		X	
	Digital Twins for manufacturing	X	X	X	X	
	Small coverage, low power network of networks for	X	X		X	X

	production & manufacturing					
	Consumer robots	X	X	X	X	
	AI partners	X	X		X	
	Interacting and cooperative mobile robots	X	X	X	X	X
	Flexible manufacturing		X	X	X	X
6GFLAGSHIP	Connected industries	X	X	X	X	X
	Sensing/imaging		X		X	X
	Sensor fusion		X	X	X	X
	Object recognition		X	X	X	
	Big data analytics	X	X		X	

8.4 6G Architecture Vision

8.4.1 State-of-the-art

The first steps towards a future-proof architecture for 6G has been reported in [47]. The expectation is for 6G system to be a **programmable, smart service execution platform** available to provide (and consume) network, cloud, and IT capabilities to (from) third-party consumers through open, global and user-friendly APIs, thereby enabling full app-to-network integration. This platform shall build on the principles of i) *full cloud-nativeness*, spanning E2E and cross-plane; ii) *software and hardware disaggregation*, entirely agnostic to the underlying technology; iii) *AI everywhere*, natively integrated and offered for consumption using ‘AI-as-a-service’ models; iv) *reliability, availability, and resilience*, beyond 5G, both in terms of service and infrastructure provisioning; and v) *digital inclusion and global service coverage*.

Figure 8-6 pictures a high-level diagram of envisioned 6G system architecture, with no bindings to any reference SDO. The take aways of this architecture are listed below:

- The infrastructure will be per se a flexible on-demand provision, with great volatility. Represented as “**Resource Pool**”, such an infrastructure will be allocated across the administrative boundaries of independently operating, separate owners (e.g., Owners A, B and C). of used resources. This liquid infrastructure, where clear and stable boundaries no longer exist, calls for a pervasive, extremely capable resource control, implemented by means of “Unified Controllability” layer.

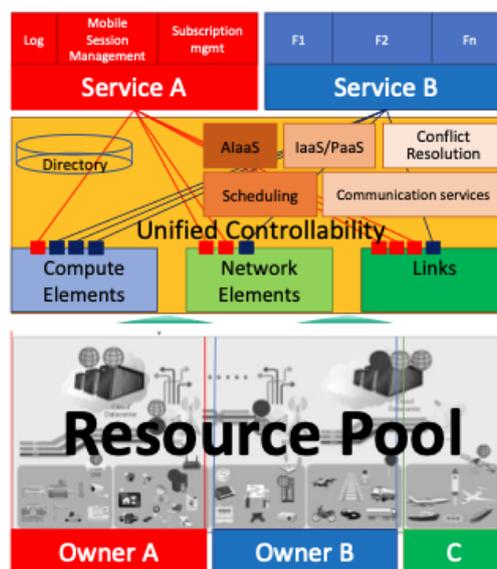


Figure 8-6 6G Architecture Vision [47]

- The “**Unified Controllability**” layer oversees interconnect (glue) and reach all the components building the “Resource Pool”, regardless of their nature, geographical and topological positions, to robustly deliver data or commands within their semantic constraints between the latter and to dynamically represent these resources as simple and stable abstractions, e.g., “Network/ Compute/ Links”. This layer will create an impression of an elastic e2e infrastructure with stable API-like interfaces for transport, compute, AI instrumentations to programmability of different abstractions, qualities, and automation levels to the service layer (both in development and runtime environments).
- The **hosted services and applications** will be built as a flexible composition (*à la carte*) of modular functionalities, built out of “Network/Compute/links” whose operation is governed with built-in capabilities of the “unified controllability layer”, i.e., AlaaS, IaaS/PaaS, conflict resolution, scheduling, and communication services. The result is that the system increasingly looks like a computer program executed within the programmable infrastructure. Using high level programming languages, compilers, and interpreters, it is expected for intent-based expressions of needs and constraints of services to be translated to microservice-based and data-driven allocations.

8.4.2 5G-CLARITY approach to 6G

5G-CLARITY system is designed according to much of the principles mentioned in Section 7.4.1, making this system a good starting point to keep evolving 5G-CLARITY towards target 6 architecture goals. This forward compatibility was demonstrated in 5G-CLARITY D2.2 [1], through the table below. This table shows how to extend current 5G-CLARITY system features and capabilities to materialize extensions mentioned in Section 7.2.2.

Table 8-5 5G-CLARITY Road Towards 6G: Architecture Evolution

Innovation Aspect	Innovation Description	Means to Meet the Innovation
New/Enriched WAT	Even though 5G-CLARITY system is currently designed to support (and make a combined use of) 5G NR, Wi-Fi and LiFi, new WATs (e.g. mmWave, 6G radio) may need to be accommodated in the medium and long run, if upcoming private network scenarios require that. Likewise, these scenarios might require updates on existing WATs (e.g., new features in Wi-Fi 7, new features in Rel-18 5G NR, etc.) and ICAS. 5G-CLARITY shall be able to cope with both situations.	Way forward: develop WAT-specific models and integrate them in the Management and Orchestration stratum, using NETCONF as protocol. Supporting principles: extensibility, model-driven operation.
multi-WAT aggregation	This innovation represents the ability to introduce new algorithms for multi-WAT aggregation, to cope with the pace of evolution of access technologies. Scenarios that might motivate these algorithms is the integration of a new WAT, or a new optimization dimension (e.g., increase reliability without decreasing spectrum efficiency) derived from the combined use of existing WATs.	Way forward: algorithms can be incorporated as a separate VNF (proxy) or updates to the CPE. The architecture is not impacted by these two solutions. Supporting principles (stratum): multi-WAT protocol stack (network and application function stratum)
New xApps	Service innovation in RAN operation can be done with the onboarding of new xApps. These xApps can complement CU-CP hosted RRM functionality, by bridging new functionality applicable to real-time use cases, including QoS/QoE optimization, per-UE controlled load balancing, seamless handover control and advanced per-slice SLA	Way forward: develop a new feature as a new xApp. Once validated and tested (out of the scope of 5G-CLARITY), it can be easily onboarded to 5G-CLARITY dRAN (RIC) and injected to gNB-CU/DU using E2

	<p>monitoring. In addition to providing add-on features, industry is exploring the possibility of implementing complex, hard-to-evolve RRM procedures through xApps, in order to reduce lifecycles of innovations, decouple them from the lifecycles (e.g., 1/2-year release cycle) from traditional CU-CP hosted RRM functionalities.</p>	<p>interface. Supporting principles (stratum): RAN functional splitting and O-RAN (network and application function stratum).</p>
<p>New VxFs</p>	<p>Given the ever-evolving nature of B5G/6G ecosystem, new network functions will be integrated into the Rel-17+ 5GC, producing/consuming services to/from internal and external application functions. The evolution towards fully cloud-native environments will require accommodating all these workloads as containerized VxFs, coexisting with 4G/5G workloads, most of them deployed in VM-based environments.</p>	<p>Way forward: The 5G-CLARITY NFVI is built out of different clusters (resource zones), all orchestrated from a single NFVO. This approach is ideal to flexibly accommodate VxFs as needed, taking into account their semantics and the virtualization technology. For example, VM-based and containerized workloads can be segregated into different clusters. In the same way, RAN, 5GC and application functions can be hosted by different clusters. This clustered-based approach while having a single NFVO is ideal for the seamless accommodation of new VxFs, without the impact of system architecture. Supporting principles (stratum): clustered NFVI (infrastructure stratum) and SBMA (reproducibility)</p>
<p>AI marketplace</p>	<p>This innovation represents the ability of introducing (updating) new (existing) AI capabilities, by uploading new (versions) of ML models and making them available to subscribed consumers, e.g. new applications, new management functions, etc.</p>	<p>Way forward: ML models can be seamlessly loaded into the AI engine (Section 8.1) which is originally designed to do this. Supporting principles (stratum): cloud native and service-oriented (intelligence stratum)</p>
<p>Enlarge intent scope</p>	<p>The applicability of intent concept in 5G-CLARITY is limited to AI models and slice management, which act as intent providers. However, in the long run, with much more technology heterogeneity in private networks, and more (non-telco) industry actors willing to take on the network operator role, it is important to abstract underlying details to streamline the operation of these private networks. This will be even more critical with the upcoming ‘network of networks’ paradigm, based on the aggregation of nanoscale ad-hoc subnetworks within the same private network, and where the complexity is exponentially increased as the number of these subnetworks scale. To solve this, the “intentization” is a must, and shall be spanned beyond short-term/mid-term practices, at both fulfilment (provisioning and configuration management) and assurance (performance and fault management) time.</p>	<p>Way forward: New system modules can be integrated with the Intent Engine (Section 8.2) as new providers. Updates to the Intent engine northbound interfaces do not affect the rest of the system Supporting principles (stratum): ease of use, language/ML framework independence (intelligence stratum).</p>

9 Conclusions

The outcomes of these deliverable are two. On the one hand, 5G-CLARITY D2.4 has provided the final version of 5G-CLARITY system architecture, implementing necessary extensions over the initial original architecture design detailed in 5G-CLARITY D2.2 [1], based on the lessons learnt from the activities done in WP3 (implementation of Network and Application Function stratum), WP4 (Management and Orchestration stratum and Intelligence stratum) and WP5 (set up and execution of in-project pilots). On the other hand, 5G-CLARITY D2.4 has reported on the architecture evaluation, building on the end-to-end modelling tools worked out in 5G-CLARITY D2.3, and further extend them, according to the features updated in the final release of 5G-CLARITY system. To proceed with this evaluation, several relevant use cases have been defined and reported. The evaluation results produced for these use cases using the modelling tools developed indicate that the 5G-CLARITY solution can offer clear benefits with respect to the relevant state-of-the-art.

Sections 2 and 3 have covered activities in scope of 5G-CLARITY T2.2. In more detail, Section 2 has reported on the final architecture of 5G-CLARITY system, highlighting the progress in relation to the original design captured in deliverable 5G-CLARITY D2.2 [1]. Section 3 has given a deep dive on the system capabilities when coming to advanced NPN scenarios, illustrating solutions to cope with multi-site deployments, mobility context and data aggregation in public-private network environments. Section 4 provides a list of the scenarios under evaluation and highlights extensions / refinements over the studies performed in 5G-CLARITY D2.3 [2].

Section 5, 6 and 7 have covered activities in scope of Task 2.3 presenting in detail the evaluations that have been carried out per scenario considered in the project. Attention is given on the extensions of the end-to-end modelling tools originally introduced in 5G-CLARITY D2.3 [2], to perform more advance evaluation of 5G-CLARITY architecture. These include (a) a multi-stage optimization framework to enable optimal placement of virtualized functions to appropriate resources taking into consideration end-user mobility, (b) LiFi communication channel model, (c) Modelling of user mobility in the SDN-enabled LiFi/Wi-Fi integrated network and (d) WAT positioning models. In the same sections, the final evaluation results indicating the benefits of the proposed 5G-CLARITY architecture and technologies and the associated trade-offs are reported.

To conclude this report, Section 8 has detailed pathways to evolve 5G-CLARITY system towards a 6G-ready service platform for private networks, with focus on the use cases and priorities outlined in on-going 6G initiatives, and with the capabilities to fully realize the 'network of network's paradigm.

10 Bibliography

- [1] 5G-CLARITY, “5G-CLARITY D2.2 Primary System Architecture,” 2020. Available at: https://www.5gclarity.com/wp-content/uploads/2020/12/5G-CLARITY_D22.pdf.
- [2] 5G-CLARITY, “5G-CLARITY D2.3 Evaluation of the Primary System Architecture,” 2021. Available at: https://www.5gclarity.com/wp-content/uploads/2021/10/5G-CLARITY_D23.pdf.
- [3] 5G-CLARITY, “5G-CLARITY D2.1 Use cases and requirements,” 2020. Available: https://www.5gclarity.com/wp-content/uploads/2020/06/5G-CLARITY_D2.1.pdf.
- [4] 5G-CLARITY, “5G-CLARITY D3.1 State-of-the-Art Review and Initial Design of the Integrated 5G NR/Wi-Fi/LiFi Network Frameworks on Coexistence, Multi-Connectivity, Resource Management and Positioning,” 2020. Available at: https://www.5gclarity.com/wp-content/uploads/2020/09/5G-CLARITY_D3.1.pdf.
- [5] 5G-CLARITY, “5G-CLARITY D3.2 Design Refinements and Initial Evaluation of the Coexistence, Multi-Connectivity, Resource Management and Positioning Frameworks,” 2021. Available at: https://www.5gclarity.com/wp-content/uploads/2021/06/5G-CLARITY_D32.pdf.
- [6] 5G-CLARITY, “5G-CLARITY D4.2 Validation of 5G-CLARITY SDN/NFV Platform, Interface Design with 5G Service Platform, and Initial Evaluation of ML Algorithms,” 2021. Available at: https://www.5gclarity.com/wp-content/uploads/2021/11/5G-CLARITY_D42.pdf.
- [7] O-RAN Alliance, “O-RAN Technical Report, O-RAN Working Group 2 Non-RT RIC: Functional Architecture v1.0, 2021.03.1”.
- [8] O-RAN Alliance, “O-RAN, O-RAN Working Group 2 (Non-RT RIC and A1 interface WG) A1 interface: Type Definitions, 2021.07.16”.
- [9] “Amazon EC2,” [Online]. Available: <https://aws.amazon.com/ec2/pricing/on-demand/>.
- [10] “Keycloak,” [Online]. Available: <https://www.keycloak.org/>.
- [11] “Prometheus,” [Online]. Available: <https://prometheus.io/>.
- [12] 5G-CLARITY, “5G-CLARITY D4.3: Evaluation of E2E 5G infrastructure and service slices, and of the developed self-learning ML algorithms,” (In Preparation), 2022.
- [13] “Amazon pricing,” [Online]. Available: <https://aws.amazon.com/s3/pricing/>.
- [14] ZSM 002 ETSI GS, “Zero-touch network and Service Management (ZSM): Reference Architecture”.
- [15] Network Management Research Group , “charter-irtf-nmrg-02,” [Online]. Available: <https://datatracker.ietf.org/doc/charter-irtf-nmrg/>. [Accessed 30 March 2022].
- [16] 3GPP, “5G System Architecture for the 5G System (5GS); State 2 (Release 16), document 3GPP TS 23.501, Version 16.5.0, Jul. 2020.,” 2020.
- [17] S. Rommer , 5G Core Networks: Powering Digitalization, Academic Press., 2019.
- [18] 5G-CLARITY, “5G-CLARITY D5.1 Specification of Use Cases and Demonstration Plan,” 2021. Available at:

https://www.5gclarity.com/wp-content/uploads/2021/02/5G-CLARITY_D51.pdf.

- [19] J. Prados-Garzon and e. al., “5G Non-Public Networks: Standardization, Architectures and Challenges,” *IEEE Access*, pp. 153893-153908, 2021.
- [20] M. Corici, P. Chakraborty, T. Magedanz, A. S. Gomes, L. Cordeiro and K. Mahmood, “5G Non-Public Networks (NPN) Roaming Architecture,” 2021.
- [21] I. Leyva-Pupo, C. Cervelló-Pastor, C. Anagnostopoulos and D. P. Pezaros, “Dynamic scheduling and optimal reconfiguration of UPF placement in 5G networks,” in *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2020.
- [22] “Open Air Interface,” [Online]. Available: <https://gitlab.eurecom.fr/oai/openairinterface5g/-/wikis/OpenAirFeatures>.
- [23] A. Tzanakaki, M. Anastasopoulos, A. Manolopoulos and D. Simeonidou, “Mobility aware Dynamic Resource management in 5G Systems and Beyond,” in *International Conference on Optical Network Design and Modeling*, 2021.
- [24] H. D. and R. S., “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized,” *IEEE Trans. Veh. Technol*, p. 77–92, 1986.
- [25] H. Haas, “Introduction to indoor networking concepts and challenges in LiFi,” *Journal of Optical Communications and Networking*, pp. A190-A203, 2020.
- [26] X. Wu and H. Haas, “Handover Skipping for LiFi,” *IEEE Access*, pp. pp. 38369-38378, 2019.
- [27] H. Alshaer, H. Haas and O. Y. Kolawole, “An Optimal Networked LiFi Access Point Slicing Scheme for Internet-of-Things,” in *IEEE International Conference on Communications Workshops*, 2021.
- [28] P. Fahamuel, J. Thompson and H. Haas, “Study, analysis and application of optical OFDM, single carrier (SC) and MIMO in intensity modulation direct detection (IM/DD),” in *IET Intelligent Signal Processing Conference*, 2013.
- [29] S. B. G. S. H. Merlin, “TGax Simulation Scenarios”. 2015 [Online]. Available:,” pp. <https://mentor.ieee.org/802.11/dcn/14/11-14-0980-16-00ax-simulation-scenarios.docx>, 2015 .
- [30] E. Perahia and R. Stacey, *Next Generation Wireless LAN: 802.11n and 802.11ac*, Cambridge University Press, 2013.
- [31] H. Alshaer, H. H. and O. Kolawole, “An Optimal Networked LiFi Access Point Slicing Scheme for Internet-of-Things,” 2021.
- [32] R. G. Gallager, *Stochastic Processes: Theory for Applications*, 1st ed Cambridge University Press, 2014.
- [33] H. Haas, L. Yin, C. Chen, S. Videv, D. Parol, E. Poves, H. Alshaer and M. Islim, “Introduction to indoor networking concepts and challenges in LiFi,” *Journal of Optical Communications and Networking*, pp. A190-A202, February 2020.
- [34] L. Lin, Z. Ghassemlooy, C. Lin, X. Tang, Y. Li and S. Zhang, “An Indoor Visible Light Positioning System Based on Optical Camera Communications,” *IEEE Photonics Technology Letters*, pp. 579-582, 2017.
- [35] R. Shilkrot and D. M. Escriva, “Mastering OpenCV 4 - Third Edition,” [Online]. Available: <https://www.oreilly.com/library/view/mastering-opencv-4/9781789533576/848e4e77-32ec-499d->

9945-cb0352e28236.xhtml.

- [36] Matlab, “Camera Calibration,” [Online]. Available: https://de.mathworks.com/help/vision/ug/camera-calibration.html?s_tid=srchtitle_pinhole%2520camera_7.
- [37] I. Std, “IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks-Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE”.
- [38] A. S. Stordal, H. A. Karlsen, G. Naevdal, H. J. Skaug and B. Valles, “Bridging the ensemble kalman filter and particle filters: the adaptive gaussian mixture filter,” *Computational Geosciences*, pp. 293-305, 2011.
- [39] D. Alspach and H. Sorenson, “Nonlinear Bayesian estimation using Gaussian sum approximations,” *IEEE transactions on automatic control*, vol. 17, pp. 439-448, 1972.
- [40] F. Gustafsson, “Particle filter theory and practice with positioning applications,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, pp. 53-82, 2010.
- [41] M. Goodarzi, V. Sark, N. Maletic, J. Gutiérrez, G. Caire and E. Grass, “DNN-assisted Particle-based Bayesian Joint Synchronization and Localization,” *arXiv preprint arXiv:2110.02771*, 2021.
- [42] J. Chen, S. Guan, Y. Tong and L. Yan, “Two-dimensional direction of arrival estimation for improved archimedean spiral array with MUSIC algorithm,” *IEEE Access*, vol. 6, pp. 49740-49745, 2018.
- [43] J. Heaton, Introduction to neural networks with Java, Heaton Research, Inc, 2008.
- [44] S. Boyd, S. P. Boyd and L. Vandenberghe, Convex optimization, Cambridge university press, 2004.
- [45] 5G-PICTURE, “D4.3 "Integration of developed functions with 5G-PICTURE orchestrator,” 2020.
- [46] A. Andrade and H. de-Souza, “Proposal for the application of Technological Prospecting approach to a Scientific and Technological Institution,” *International Journal of Humanities and Social Science Invention*, pp. 61-67, 2017.
- [47] 5G IA, “European Vision for the 6G Network Ecosystem,” White Paper, [Online]. Available: <https://5g-ppp.eu/wp-content/uploads/2021/06/WhitePaper-6G-Europe.pdf>.
- [48] “Strategic Research and Innovation Agenda”.
- [49] “HEXA-X project [Online]. Available at: <https://hexa-x.eu>,” [Online].
- [50] M. Hoffmann, M. Uusitalo, M.-H. Hamon and et al., ““6G vision, use cases and key societal values,” Hexa-X Deliverable D1.1, February, 2021. Available at: https://hexa-x.eu/wp-content/uploads/2021/02/Hexa-X_D1.1.pdf”.
- [51] 6G Flagship , [Online]. Available: <https://www.oulu.fi/6gflagship/>.
- [52] 6G Flagship, [Online]. Available: <https://www.oulu.fi/6gflagship/6g-white-papers>.
- [53] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker and J. Turner, “OpenFlow: Enabling Innovation in Campus Networks,” *ACM Computer Communication Review*, pp. 69-74, 2008.

- [54] M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppanen, V. Koivunen and M. Valkama, "Joint Device Positioning and Clock Synchronization in 5G Ultra-Dense Networks," *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol. 16, 2017.
- [55] M. Goodarzi, D. Cvetkovski, N. Maletic, J. Gutiérrez and E. Grass, "Synchronization in 5G Networks: a Hybrid Bayesian Approach towards Clock Offset/Skew Estimation and Its Impact on Localization," *EURASIP Journal on Wireless Communications and Networking*, 2021.
- [56] "Free5GC," [Online]. Available: <https://www.free5gc.org/>.
- [57] H. Alshaer, H. Haas and O. Y. Kolawole, "An Optimal Networked LiFi Access Point Slicing Scheme for Internet-of-Things," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021.
- [58] "O-RAN Alliance," [Online]. Available: <https://www.o-ran.org>.
- [59] T. Taleb, I. Afolabi and M. Bagaa, "Orchestrating 5G network slices to support industrial internet and to shape next-generation smart factories," *IEEE Netw.*, vol. 33, no. 4,, pp. pp. 146–154,, Jul./Aug. 2019..
- [60] T. Werthmann, H. Grob-Lipski and M. Proebster, "Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks," in *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, 2013.
- [61] "Amarisoft," [Online]. Available: <https://www.amarisoft.com/>.