Project funded by the Horizon 2020 Framework Programme of the European Union,

Grant agreement Nº: ***951947.***

Start date: **01/09/2020**

Duration: **36 months**

D3.1 Analysis of 5GMED Infrastructure Requirements and 5G Handover between Networks and Cross-Border

WP	WP3
WP Leader	RETE
Responsible Author	Jad Nasreddine (I2CAT)
Contributors	I2CAT, RETE, VLO, AAE, CMS, ATOS, ATC, HSP, IRT, CTTC, NBC, VEDE, VDF
Dissemination Level	PU
Nature	RE

Dissemination Level:	
PU	Public

Nature	
RE	Report

Synopsis	Deliverable D3.1 is the first report of WP3. It provides the network requirements in terms of network KPIs, which are derived from the service KPIs of the use cases presented in D2.1. It then provides an analysis of the cross-border corridor and identifies the technological challenges to deploy a network that can satisfy the requirements of the use cases in cross-border scenarios. Finally, it presents and analyses the existing solutions to overcome the technological challenges.
List of Keywords	Architecture, 5G, V2X, Cellular, Network KPI, HetNet, Network Slicing, Multi-connectivity, MEC, virtualisation, seamless communication, Roaming

Funded by the Horizon 2020 Framework Programme of the European Union

DOCUMENT HISTORY

Version	Status[1]	Date	Comments	Author
0.1	Draft	30/03/2021	ToC drafted.	I2CAT
0.2	Draft	15/04/2021	First Version with some contributions	I2CAT
0.3	Draft	23/04/2021	Second Version with some contributions	All partners
0.4	Draft	28/04/2021	Third Version with some contributions	All partners
0.5	Draft	30/04/2021	Fourth Version with some contributions	All partners
0.6	Draft	7/4/2021	Internal version	I2CAT
0.7	Draft	14/4/2021	Internal version	I2CAT
0.8	Draft	19/4/2021	Internal version	I2CAT
0.9	Draft	21/5/2021	Final draft for internal review	I2CAT
1.1	Draft	23/02/2022	First version for re-submission	I2CAT
1.2	Draft	16/03/2022	Partners' contributions integrated	All partners
1.3	Draft	18/03/2022	First version ready to internal review	I2CAT
1.4	Under review	23/03/2022	Internal review	I2CAT, RETE
1.5	Under review	27/03/2022	Integration of partners feedback and editor review	All partners
1.6	Under review	29/03/2022	External review	IRT, CXFR, AXBY
1.7	Under review	30/03/2022	Integration of comments	I2CAT
1.8	Under review	01/03/2022	Review by TM	I2CAT
2	Issued	02/03/2022	Final Document	I2CAT
2.1	Draft	23/09/2022	First version for second re-submission	I2CAT
2.2	Draft	16/10/2022	Partners' contributions integrated	All partners
2.3	Under review	24/10/2022	Review sections 3 and 4	I2CAT
2.4	Under review	31/10/2022	Review from I2CAT and NBC	I2CAT, NBC
2.5	Under review	21/11/2022	Review from I2CAT	I2CAT
2.6	Under review	27/11/2022	Final contribution from partners	All partners
2.7	Under review	01/12/2022	WP3 internal review	CTTC, I2CAT, SNCF
2.8	Under review	03/12/2022	Review from I2CAT	I2CAT
2.9	Under review	05/12/2022	TM review	I2CAT
2.91	Under review	07/12/2022	Integration of TM review	I2CAT
2.92	Under review	22/12/2022	External review	IRT, RETE, AXBY
2.93	Under review	05/01/2023	Integration of External reviews	I2CAT
2.94	Under review	06/01/2023	Final review by Editor	I2CAT
2.95	Under review	11/01/2023	Integrating partners and TM corrections	All partners
3	Final document	13/01/2023	Final document	I2CAT

[1]
Status (a status is associated to each step of the document life cycle)
Draft. This version is under development by one or several partner(s)
Under review. This version has been sent for review
Issued. This version of the document has been submitted to EC

Funded by the Horizon 2020 Framework Programme of the European Union

PEER-REVIEWED BY

Partner	Reviewer
IRT	Philippe Veyssiere
RETE	Jose Antonio Aranda
AXBY	Luca Petrucci

Funded by the Horizon 2020 Framework Programme of the European Union

TABLE OF CONTENTS

LIST OF TABLES 6
LIST OF FIGURES 7
LIST OF ACRONYMS 8
EXECUTIVE SUMMARY 11
1. Introduction 12
2. Use Case Analysis for Network Design 13
2.1. Overview of the Use Cases 13
2.1.1. UC1: Remote driving 14
2.1.2. UC2: Road infrastructure digitalisation 16
2.1.3. UC3: FRMCS applications and business service continuity 18
2.1.4. UC4: Follow-Me infotainment 20
2.2. Network KPIs 22
2.2.1. Definition of Network KPIs 23
2.2.2. UC1: Remote driving 25
2.2.3. UC2: Road infrastructure digitalisation 27
2.2.4. UC3: FRMCS applications and business service continuity 30
2.2.5. UC4: Follow-ME Infotainment 36
2.3. Network Requirements 38
3. 5GMED Cross-border Corridor Analysis 40
3.1. Geographical features of the corridor 40
3.2. Need for a heterogeneous radio access network 41
3.3. Availability of computing resources in the cross-border section 44
3.4. Required network exchange points 45
4. Technical Challenges and State-of-the-Art Solutions 47
4.1. 5GMED Cross-border technical challenges 47
4.1.1. Challenge #1: Multi-Connectivity and integration of heterogeneous networks 47
4.1.2. Challenge #2: Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core 48
4.1.3. Challenge #3: 5G roaming at cross-border with low latency and interruption time 49
4.1.4. Challenge #4: MEC deployment and Inter-MEC connectivity 49
4.2. Solutions to Challenge #1: Multi-Connectivity and integration of heterogeneous access networks 49

4.2.1. Non-3GPP Inter-Working Function (N3IWF) 50
4.2.2. Access Traffic Steering/Switching/Splitting (ATSSS) 51
4.2.3. Adaptive Communication System-Gateway (ACS-GW) 53
4.3. Solutions to Challenge #2: Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core 53
4.3.1. Virtual Machines and Containerization 53
4.3.2. Service Orchestration and Automation Platforms 54
4.3.3. Raemis Core overview 55
4.3.4. Network slicing in the context of 5G and cross-border scenarios 57
4.3.5. Network Automation and observability 58
4.4. Solutions to Challenge #3: 5G roaming at cross-border with low latency and interruption time 59
4.4.1. 5G roaming architecture 59
4.4.2. Techniques for low latency roaming 61
4.5. Solutions to Challenge #4: MEC deployment and inter-MEC connectivity 63
4.5.1. ETSI MEC architecture 64
4.5.2. Operator platform concept 67
5. Conclusions 71
References 72

Funded by the Horizon 2020 Framework Programme of the European Union

LIST OF TABLES

Table 1. Overview of service KPIs for UC1: Remote driving. 16
Table 2. Overview of service KPIs for UC2: Road infrastructure digitalisation. 18
Table 3. Overview of service KPIs for UC3: FRMCS applications and business service continuity. 20
Table 4. Overview of service KPIs for UC4: Follow-Me Infotainment. 22
Table 5. UC1 network KPI calculation method. 26
Table 6. UC1 network KPI values. 27
Table 7. UC2 network KPI calculation method. 29
Table 8. UC2 network KPI values. 30
Table 9. UC3 Network KPI calculation method. 33
Table 10. UC3 Network KPI values. 35
Table 11. UC4 Network KPI calculation method. 37
Table 12. Network KPI values for the EMT service in UC4 37
Table 13. Network KPI values for TP service in UC4 38
Table 14. Most stringent network requirements for the highway scenario. 38
Table 15. Most stringent network requirements for the railways scenario. 39
Table 16. Identifiers and geographical locations of Vodafone gNBs in the Spanish side. 42
Table 17. List of N3IWF features. 50

LIST OF FIGURES

Figure 1. High-level cross-border 5GMED network architecture (extracted from [1])..... 14
Figure 2. High-level functional architecture of UC1 (extracted from [1])..... 15
Figure 3. High-level functional architecture of UC2 (extracted from [1])..... 17
Figure 4. High-level functional architecture of UC3 (extracted from [1])..... 19
Figure 5. High-level functional architecture of UC4 - automotive scenario (extracted from [1]). 21
Figure 6. Simplified Representation of the 5GMED network segments. 23
Figure 7. End-to-end latencies break-down in UC1. 26
Figure 8. End-to-end latencies break-down in the REM service of UC2. 27
Figure 9. End-to-end latencies break-down in the AID service of UC2. 28
Figure 10. Traffic regulation End-to-end latencies break-down in the three service of UC2. 29
Figure 11. End-to-end latencies break-down in FRMCS P1 service and FRMCS P2 service of UC3 for uplink (left) and downlink (right). 31
Figure 12. End-to-end latencies break-down in FRMCS P3 service of UC3 for uplink (left) and downlink (right). 32
Figure 13. End-to-end latencies break-down in B1 and B2 services of UC3 for uplink (left) and downlink (right). 33
Figure 14. End-to-end latencies break-down in UC4 36
Figure 15. Map showing the different areas of the cross-border corridor between Figueres and Perpignan. 40
Figure 16. Coverage simulation for 5G New Radio (NR) at 3.5 GHz (band N78) in the Spanish side (using four gNBs that will be installed in existing Vodafone sites). The blue colour denotes where the estimated received power is higher or equal to a given threshold (-115 dBm). 42
Figure 17. Coverage simulation for 5G New Radio (NR) at 3.5 GHz (band N78) in the French side (using six gNBs). The green colour denotes where the estimated received power is higher or equal to a given threshold (-93dBm). 44
Figure 18. MEC sites distribution over the cross-border corridor. 45
Figure 19. Interconnection networks. 46
Figure 20. Integrating non-3GPP technologies in the 5G core network via N3IWF. 50
Figure 21. ATSSS simplified architecture for user plane. 51
Figure 22. ATSSS simplified control and data plane between UE and UPF. 52
Figure 23. Elements of the Raemis Core [8]. 56
Figure 24. Network Slicing within 5GMED Architecture. 58
Figure 25. 5G System Roaming architecture – Service Based Interface Representation (HR) [18]. 60
Figure 26. 5G System Roaming architecture – Service Based Interface Representation (LBO) [18]. 60
Figure 27. Roaming architecture with N14 interface between AMFs [20] 62
Figure 28. 5G Core components with Edge Data plane offload. 66
Figure 29. Federation among multiple OP instances [24]. 69
Figure 30. OP high level architecture for Edge Computing [24]. 69

LIST OF ACRONYMS

5GC	5G Core
ACS-GW	Adaptive Communication System-Gateway
AI	Artificial Intelligence
AID	Automatic Incident Detection
ALS	Application Layer Security
AMF	Access and Mobility Management Function
AP	Access Providers
API	Application Programming Interface
ATSSS	Access Traffic Steering, Switching and Splitting
CA	Certification Authority
CCAM	Cooperative and Connected Automated Mobility
CDN	Content Delivery Network
CI/CD	Continuous Integration and Continuous Delivery
CUPS	Control and User Plane Separation
C-V2X	Cellular V2X
DAM	Data Analytics Module
DAS	Distributed Antenna System
DN	Data Network
DNN	Data Network Name
E	Edge
E/AO	Edge/Access Operator
EMT	Enjoy Media Together
eUICC	Embedded Universal Integrated Circuit Card
EWBI	East-Westbound Interface
FRMCS	Future Railway Mobile Communication System
gNB	gNodeB
GBR	Guaranteed Bit Rate
GKE	Google Container Engine
GPS	Global Positioning System
GTP-U	GPRS Tunnelling Protocol User plane
GUI	Graphical User Interface
H2020	Horizon 2020
HD	High Definition
HetNet	Heterogeneous Network
HMI	Human Machine Interface
hPLMN	home PLMN
HR	Home Routed
IKE	Internet Key Exchange
IPUPS	Inter PLMN User Plane Security
IoT	Internet of Things
IPSec	Internet Protocol Security
IPX	Internet Protocol packet eXchange
KPI	Key Performance Indicator
LAGW	Local Access Gateway
LAN	Local Area Network
LBO	Local Breakout

MEC	Mobile Edge Computing
MME	Mobility Management Entity
MNO	Mobile Network Operator
MPTCP	Multipath TCP
MRM	Minimum Risk Manoeuvre
MTU	Maximum Transmission Unit
N3IWF	Non-3GPP Inter-Working Function
NAS	Non-Access Stratum
NBI	Northbound interface
NGAP	Next Generation Application Protocol
NF	Network Function
NFV	Network Function Virtualization
NFVI	Network Functions Virtualization Infrastructures
NR	New Radio
NRI	Network Reselection Improvement
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NWDAF	Network Data Analytics Function
ODD	Operational Design Domain
OP	Operator Platform
OPG	Operator Platform Group
OS	Operating System
OSS	Operations Support System
OTA	Over-The-Air
OTT	Over-The-Top
PCN	Private Core Network
PDU	Protocol Data Unit
PKI	Public Key Infrastructure
PLMN	Public Land Mobile Network
PMF	Performance Measurement Function
PMFP	PMF protocol
POI	Point of Interest
QoS	Quality of Service
PCC	Policy and Charging Configuration
PCF	Policy Control Function
PCN	Private Core Network
PDR	Packet Data Rule
PDU	Protocol Data Unit
POI	Points Of Interest
PRINS	Protocol for N32 Interconnect Security
RAN	Radio Access Network
RAT	Radio Access Technology
REM	Relay of Emergency Messages
REST	Representational state transfer
RRA	Request for Remote Assistance
RS	Remote Station
RSU	Road-Side Unit
RTT	Round Trip Time

RV	Remote Vehicle
SA	Stand-Alone
SBA	Service-Based Architecture
SBI	Service-Based Interface
SCTP	Stream Control Transmission Protocol
SDF	Service Data Flow
SDN	Software Defined Network
SEPP	Security Edge Protection Proxy
SIM	Subscriber Identity Module
SLA	Service-Level Agreement
SM	Session Management
SMF	Session Management Function
S-NSSAI	Single Network Slice Selection Assistance Information
SoR	Steering of Roaming
SP	Service Provider
SUPI	Subscription Permanent Identifier
TAN	Train Access Network
TCP	Transmission Control Protocol
TCU	Telematic Control Unit
TEC	Telco Edge Cloud
TEI	Tunnel Endpoint Identity
TFR	Traffic Flow Regulation
TLS	Transport Layer Security
TM	Teleoperation Manoeuvre
TMC	Traffic Management Centre
TNGF	Trusted Non-3GPP Gateway Function
TP	Tour Planning
UDP	User Datagram Protocol
UE	User Equipment
UPF	User Plane Function
UDM	Unified Data Management
UNI	User-Network Interface
USIM	Universal Subscriber Identity Module
V2X	Vehicle to Everything
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
vPLMN	Visited PLMN
VR	Virtual Reality
VTC	Valeo Teleoperation Cloud
W-AGF	Wireline Access Gateway Function
WLAN	Wireless LAN
WPA2	Wi-Fi Protected Access 2

Funded by the Horizon 2020 Framework Programme of the European Union

EXECUTIVE SUMMARY

The objective of this deliverable is to analyse the requirements of the use cases defined in D2.1 [1] and the characteristics of the cross-border corridor. This analysis will be used to infer the requirements of the 5GMED network and identify the possible challenges and solutions to be investigated in the project. This document first provides a brief description of the four use cases and their related services: remote driving use case, road infrastructure digitalisation use case, Future Railway Mobile and Communication System (FRMCS) applications and business service continuity use case, and Follow-ME infotainment use case. Then, it uses the service performance requirements defined in D2.1 to derive the network Key Performance Indicators (KPIs) that will be considered when designing the network architecture and test cases. From the analysis performed in this document, the most challenging network KPIs are the very low latency and the very high throughput required in certain services, in addition to the short mobility interruption time when crossing the border.

Thereafter, the cross-border corridor is analysed in terms of geographical features (i.e., challenging orography, presence of a tunnel), the availability of sites for the 5G gNBs and their potential coverage, and the proximity between the railway track and highway. Based on this analysis, and to achieve the 5GMED objectives, it was decided to complement the 5G network with other radio access technologies, i.e., 5.9 GHz C-V2X and IEEE 802.11ad at 70 GHz. In addition, Satellite links will provide connectivity in the region where there is no ground connectivity. In the tunnel, a Distributed Antenna System (DAS) will be used to provide 5G coverage all along the tunnel. Furthermore, the required Mobile Edge Computing (MEC) sites were identified.

Finally, four cross-border technical challenges related to the use cases and the technological enablers provided by the 5G network were identified to be solved by 5GMED. These challenges are as follows: 1) multi-connectivity and integration of heterogeneous networks, 2) virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core, 3) cross-border 5G roaming with low latency and low interruption time, and 4) MEC deployment and inter-MEC connectivity. The challenges are described, together with possible solutions from standards and telecommunication alliances. In addition, the solutions and platforms selected by 5GMED are highlighted. These challenges will be considered in the detailed design of the 5GMED network architecture within WP3.

1. Introduction

The 5GMED project will build a scalable multi-stakeholder network and compute infrastructure to support Cooperative, Connected and Automated Mobility (CCAM) and Future Railways Mobile Communication System (FRMCS) use cases along the Barcelona-Perpignan cross-border corridor using 5G system compliant with 3GPP Release 16. The 5GMED infrastructure will be designed to meet the performance requirements of the CCAM and FRMCS use cases. To highlight the numerous challenges faced in the deployment of CCAM and FRMCS use cases, especially at country borders, the 5GMED consortium has considered four use cases, namely: the remote driving use case, the road infrastructure digitalisation use case, the FRMCS applications and business service continuity use case, and the follow-me infotainment use case. The description of these use cases is presented in D2.1 [1].

The 5GMED use cases require seamless and high-quality communications in heterogenous, high speed, and challenging cross-border scenarios. This will be particularly challenging in the geographical area chosen to deploy the 5GMED infrastructure due to the presence of a railway tunnel with specific propagation conditions and a border separating two countries with one Mobile Network Operator (MNO) in each country and different service providers. Therefore, we need to carefully design a network architecture that can support the defined use cases in both railway and highway cross-border scenarios.

In summary, the aim of the project is to show that the selected CCAM and FRMCS services can be supported in different countries using mainly 5G system. In addition, 5GMED must show how to integrate non-5G technologies in a transparent way to improve system performance. To meet the requirements of the use cases, 5GMED will develop a 5G-based network architecture that facilitates:

- Cross-MNO and inter-radio access technology (inter-RAT) seamless connection to minimize mobility interruption times at the cross-border and where different access technologies are mandated by the nature of the environment. In addition to the 5G system, Cellular V2X (C-V2X) at 5.9 GHz, IEEE 802.11ad at 70 GHz, and satellite links will be also used in 5GMED.
- Flexible and optimized service (use case) rollout on the 5GMED infrastructure with customized quality of service (QoS) profile for each service using the 5G network slicing technology.
- Artificial Intelligence (AI)-powered distributed edge computing to mainly manage network functions, and in some cases to predict QoS, and process data from sensors at the edge.
- Network and service orchestration in cross-border scenarios to allow seamless data flow and to guarantee end-to-end quality of service and continuity among different administrative domains.
- Self-sustainable network infrastructure sites to be deployed in isolated regions of the corridor where the power grid is not available.
- Connectivity to remote sites either via fibre or advanced radio link technologies to be able to connect the user to the 5GMED core network.

The document is organized as follows. In Section 2, we define the network KPIs, provide an analysis of the network KPIs required in each use case, and derive their values based on the target service KPIs defined in D2.1 [1]. In Section 3, we analyse the topological characteristics and the challenges of the highway and railway segments of the cross-border corridor. In Section 4, we identify and analyse the technological challenges that must be accounted in the design and implementation of the 5GMED network architecture and present existing solutions to overcome those challenges. Finally, this document is concluded in Section 5.

2. Use Case Analysis for Network Design

The aim of this section is to derive the network KPIs from the service KPIs presented in D2.1 for each use case. These network KPIs will be used to identify the network requirements in terms of data rate, latency, mobility interruption time, reliability, etc. Therefore, the main functional components of each use case are mapped on a high-level 5GMED network architecture, which includes all the technologies considered in 5GMED (i.e., 5G, C-V2X, IEEE 802.11ad, and satellite).

In Section 2.1, we briefly describe the 5GMED use cases and their services. In Section 2.2, we present the mapping between the service KPIs of each use case and the target values of the network KPIs. Finally, the derived network KPIs are consolidated in Section 2.3.

2.1. Overview of the Use Cases

The following four use cases have been defined in 5GMED to show the capabilities of the 5GMED network infrastructure to provide seamless, reliable, high data rate, and low latency communications in a challenging cross-border and multi-RAT scenario:

- Use case 1 (UC1): Remote driving.
- Use case 2 (UC2): Road infrastructure digitalisation.
- Use case 3 (UC3): FRMCS applications and business service continuity.
- Use case 4 (UC4): Follow-Me infotainment.

In the rest of this section, we provide an overview of the 5GMED use cases to remind the reader of the main functionalities and the high-level functional architecture of each use case. A more detailed description of the use cases and their services is available in D2.1 [1].

It should be noted that all use cases definitions were developed using the high-level 5GMED network architecture presented in D2.1 (depicted in Figure 1). This architecture is designed for an ***AI-enabled*** network that will be implemented in ***cross-border*** scenarios, and it includes the following layers and interfaces. The final detailed design of the 5GMED network architecture and the corresponding interfaces will be presented in deliverable D3.3 [2].

- The **Cloud Layer** that hosts all backend applications that should run in a cloud.
- The **Slice Management Layer** that is responsible of creating and managing network slices. A slice manager will be available in each country.
- The **Orchestration Layer** that is responsible of network and service orchestration. An orchestrator will be available in each country, and the different orchestrators will be connected through a cross-border interface developed in 5GMED.
- The **MEC layer** that hosts AI-powered services and distributed applications requiring low latency. A MEC layer will be available in each country, and the different MEC layers will be connected through a cross-border interface developed in 5GMED.
- The **Network Infrastructure Layer** that provides wireless connectivity in both countries. It integrates a set of radio access network technologies as well as backhauling networks. A different network infrastructure will be present in each country, and the different infrastructures will be connected through standardized interfaces.
- The **Data Analytics layer** that contains a set of AI modules managed by the network operator to optimize network configuration. A data analytics layer will be available in each country and

the different data analytics layers will be connected through an interface developed in 5GMED.

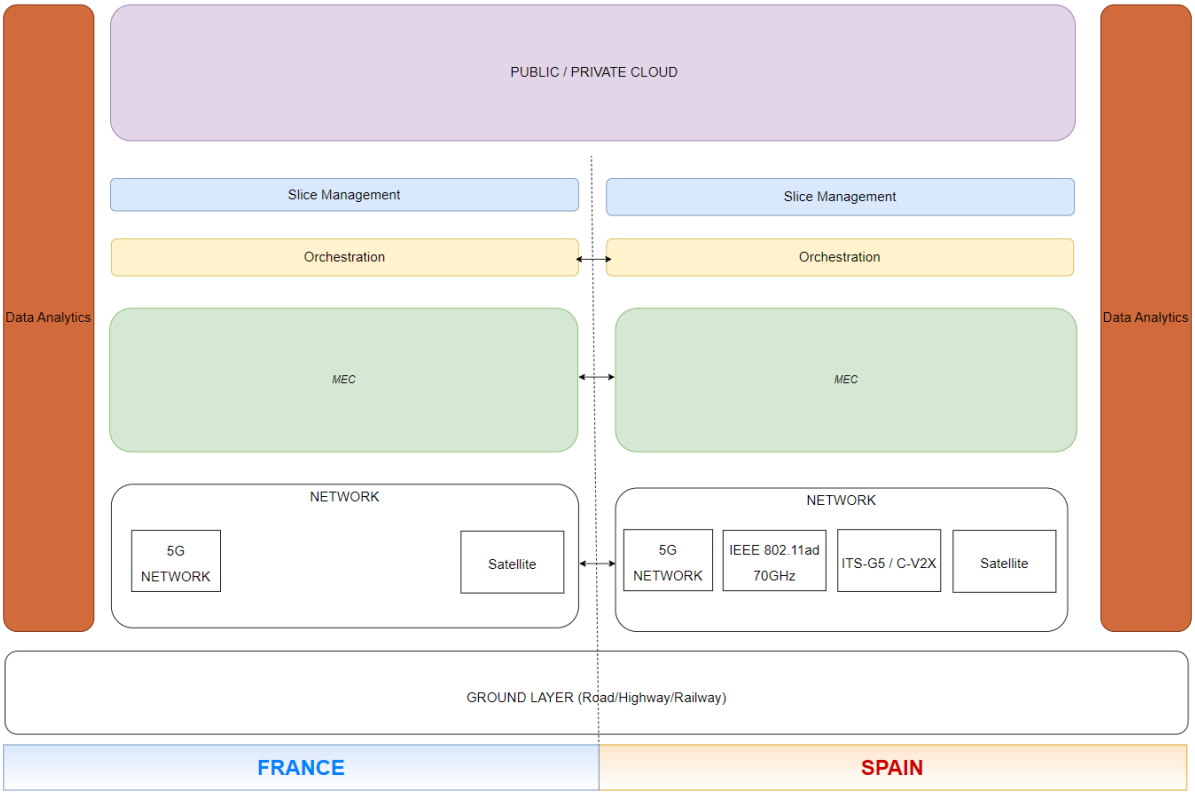

Figure 1. High-level cross-border 5GMED network architecture (extracted from [1]).

2.1.1. UC1: Remote driving

The objective of the remote driving use case is to provide tele-operated driving to an autonomous vehicle that encounters a complex situation (e.g., accident, on-board component failure, etc.) and fails to accomplish its intended autonomous driving task. When this happens, the vehicle autonomously stops in the emergency lane and requests remote assistance from a teleoperation centre. Then, a remote driver tele-operates the vehicle to a safe harbour to continue driving. During teleoperation, video images and data from sensors on-board the vehicle (e.g., 360º camera, LIDAR) must be perceived by the remote driver with sufficient quality and short delay, and the actuators of the vehicle must execute commands from the tele-operator reliably and with low latency. Thus, all the data traffic exchanged between the vehicle and the teleoperation centre is handled by the 5G network, which must meet the strict target values of the remote driving's service KPIs.

In this use case, the three main services that have been considered are:

- **Minimum Risk Manoeuvre** (MRM), in which the autonomous vehicle broadcasts an alert about its hazardous status to its environment while executing the safety MRM manoeuvre itself, i.e., it autonomously stops in a safe place on the emergency lane.
- **Request for Remote Assistance** (RRA), in which the vehicle contacts the Valeo Teleoperation Cloud (VTC) for assistance. This is a mandatory step prior to tele-operated driving.
- **Teleoperation Manoeuvre** (TM), in which a remote driver operates the vehicle from his/her Remote Station (RS).

The high-level functional architecture of UC1, presented in D2.1, is illustrated in Figure 2. Apart from the UE (i.e., TCU) on-board the remote vehicle, it can be observed that:

- The **public cloud** includes the cloud services, i.e., Valeo Teleoperation Cloud (VTC), Traffic Management Centre (TMC), and the environmental analysis module. In addition, it contains the remote station required for teleoperation. It should be noted that one set of these cloud services will be used in any country as the service will be managed by the same operator in all countries.
- The **data analytic layer** includes the QoS prediction module.
- The **network infrastructure layer** includes only the 5G network.

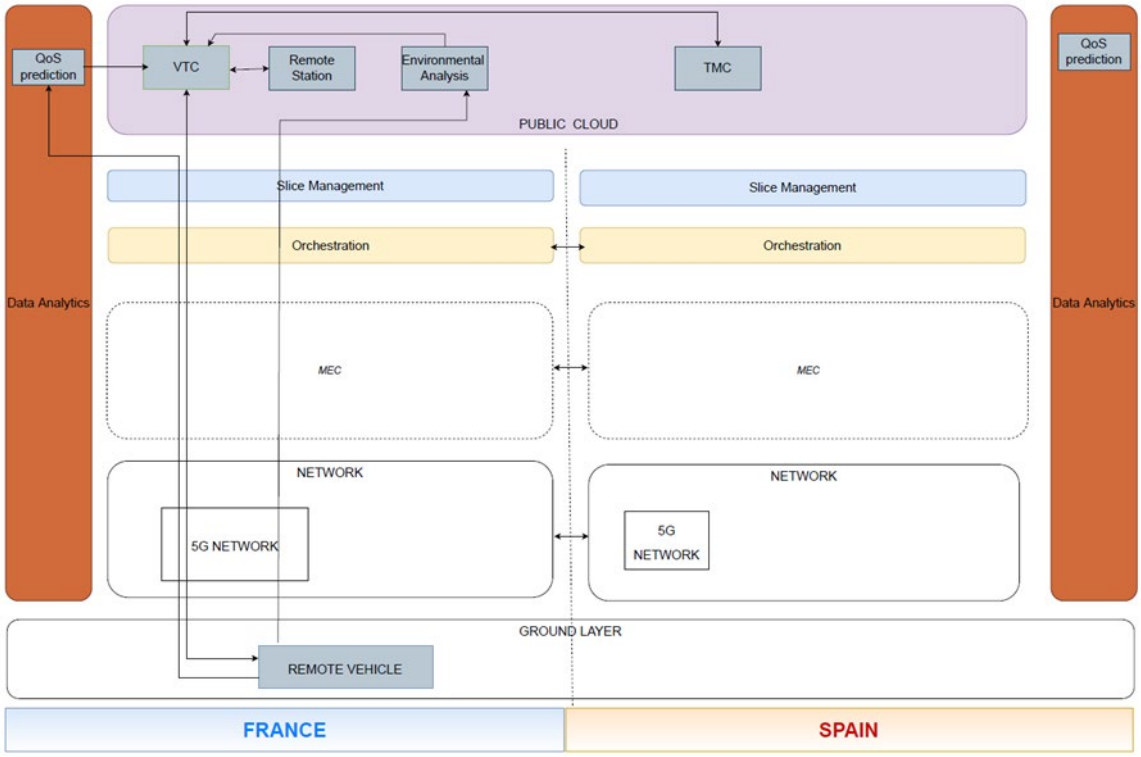

Figure 2. High-level functional architecture of UC1 (extracted from [1]).

As it was described in D2.1, the TM service is the most critical service of the remote driving use case in terms of latency, data-rate, reliability, and mobility interruption time. It requires a high uplink data rate for the transmission of high-quality videos, very low latency for real-time remote control, very high reliability to avoid teleoperation errors, and very low mobility interruption time when crossing the border. Therefore, the service KPIs of the TM service shall be the one considered in the design of the 5GMED network infrastructure for the remote driving use case. Table 1 provides an overview of the **most stringent** service KPIs for UC1 [1].

Table 1. Overview of service KPIs for UC1: Remote driving.

KPI	Requirement
Data Rate	High data rate for the transmission of high-quality videos
Latency	Very Low latency for real-time remote control
Reliability	High reliability to avoid teleoperation errors
Mobility interruption time	Low mobility interruption time when crossing the border

2.1.2. UC2: Road infrastructure digitalisation

The objective of the road infrastructure digitalisation use case is to ensure safe and efficient mobility on highways where connected vehicles coexist with non-connected vehicles. To this end, a Traffic Management Center (TMC) shall generate intelligent traffic management strategies (i.e., velocity, lane change) by processing the information received from vehicles and from roadside sensors (LiDAR, HD cameras, etc.). These strategies will be forwarded to and executed by connected cars in the affected area. The TMC is divided into two elements: the TMC Edge, which is located at the edge and controls different segments of the highway under the coverage of a MEC; and the TMC Global, which is located in the cloud and controls the whole highway by coordinating the set of TMC Edges.

Two types of traffic management strategies have been considered: (i) warning traffic strategies, and (ii) global traffic strategies. The warning traffic strategies focus on the detection of hazardous events (e.g., stopped vehicle, obstacles, etc.) and the generation of real-time warning notifications and strategies (change of lane, lower speed, etc) that will be delivered to vehicles approaching the risk area affected by the hazardous event. These events are detected by vehicles' on-board sensors or by cameras on the infrastructure and evaluated by the TMC. In global traffic strategies, the TMC analyses the traffic situation to detect abnormal behaviours (e.g., traffic jam, congestion, etc), devises a traffic strategy, and finally sends regulation commands to groups of vehicles, e.g., change lane or adjust speed.

The data traffic exchanged between the vehicles and the TMC in this use case is mainly handled by the 5G network and by a set of C-V2X roadside units that could in the future fill some gaps, if any, in 5G coverage along a cross-border corridor.

In this use case, the three services that have been considered are:

- **Relay of Emergency Messages (REM),** in which hazards are detected by connected and autonomous vehicles that send warning messages to the TMC edge through the V2X gateway. The V2X gateway handles the warning messages received from both 5G and C-V2X networks and forwards the hazard information to the TMC edge. The TMC edge evaluates the hazard information sent by different vehicles and generates a traffic strategy that will be transmitted to those vehicles that are in the corresponding segment controlled by the TMC edge. In addition, it sends all the information to the TMC global.
- **Automatic Incident Detection (AID),** in which the cameras deployed along the infrastructure are responsible for hazard detection. The only difference from the REM is that it has different hazard detection source application; instead of in-vehicle sensors in REM, road video sensors deployed along the road are used for AID. The cameras are directly connected to the TMC Edge through the 5G network. Once the hazard is detected, the TMC Edge informs the V2X gateway, which sends the warning messages to the vehicles near the hazard.

- **Traffic Flow Regulation (TFR),** in which the TMC global regulates the traffic flow by detecting abnormal behaviour (e.g., low speed, accordion phenomena, etc.) and sends regulation commands to a group of vehicles in circulation.

The high-level functional architecture of this use case, presented in D2.1, is represented in Figure 3. It can be observed that:

- The **Public Cloud** includes two TMC Global modules, one for each country. Different TMC Global are needed because each country will have a different road operator.
- The **MEC layer** in each country is responsible for hosting V2X Gateway application, which interacts with a local TMC, i.e., TMC Edge. The cross-border inter-MEC interface will be used to reduce latency in hazard notification between different countries.
- The **Network Infrastructure layer** contains 5G and C-V2X. It should be noted that C-V2X will be deployed only in the Spanish side.

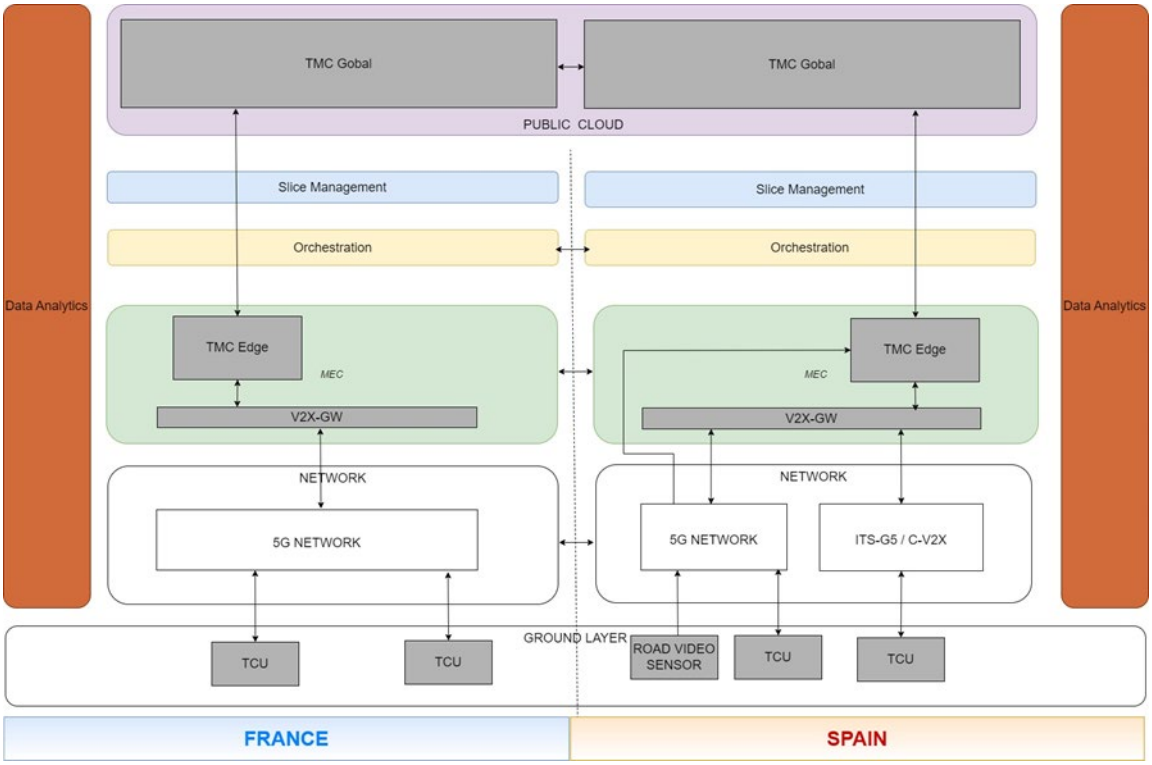

Figure 3. High-level functional architecture of UC2 (extracted from [1]).

As described in D2.1, this use case requires low latency and low mobility interruption time at the borders because the warning messages and traffic strategies should reach the concerned incoming cars in time to facilitate safe and efficient reaction to avoid accidents. Table 2 provides an overview of the **most stringent** service KPIs for UC2 [1].

Table 2. Overview of service KPIs for UC2: Road infrastructure digitalisation.

KPI	Requirement
Latency	Low latency for real-time warning and traffic strategy messages
Mobility interruption time	Low mobility interruption time at the borders because the warning messages and traffic strategies should not be lost and reach the concerned incoming vehicles
Data rate	High data rate is required especially in dense traffic situation

2.1.3. UC3: FRMCS applications and business service continuity

The objective of the FRMCS applications and business service continuity use case is to showcase how a Train Access Network (TAN) can be implemented to satisfy the requirements of all the stakeholders involved around the railway communications: infrastructure operators, train operators, and MNOs. The TAN architecture must be designed in a way that it enables the provision of five FRMCS performance and business services without disturbing the operation of the FRMCS critical services.

In this use case, the data traffic exchanged between the train and the ground (cloud or edge) application can go through one of the following access networks: 5G, 70 GHz IEEE 802.11ad, or satellite. To manage all these technologies, the data traffic exchanged between the train and the ground passes through two Adaptive Communication System-Gateway (ACS-GW) units: one ACS-GW unit is installed on-board the train, and another ACS-GW unit is installed on the ground. The ACS-GW unit on the ground is the one located closest to the train's geographical position of the train. The ACS-GW units choose the access network that is most suitable for train-to-ground/ground-to-train communication in each service depending on the train's position and network connection quality.

In this use case, the five services that have been considered are:

- **Advanced Sensor Monitoring on Board (FRMCS P1)**, in which massive on-board sensors are used to monitor the status of non-critical systems of the train and transmit their readings directly to the train control centre on ground to take decisions, which will be sent directly to the train staff.
- **Railway Track Safety – Obstacle Detection (FRMCS P2)**, in which hazards will be detected on the rail tracks using an on-board LIDAR that sends its data stream to the closest edge server, where an AI module processes the sensors' data. The warnings resulting from the AI processing are sent directly to the train control centre, which takes the necessary decision and forwards it to the trains moving towards the hazard.
- **Passenger safety and comfort (FRMCS P3)**, in which a dangerous situation on-board (e.g., fire, fights) will be detected using cameras on-board and AI processing in the MEC. The flow of information is similar to FRMCS P2.
- **High Quality Wi-Fi to passengers (B1)**, in which high-quality Wi-Fi access is provided to the passengers through the TAN and the nearest Break-out Internet Point available in the edge.
- **Multi-tenant Mobile Service (B2)**, in which passengers with 5G UEs on board the train will have high-bandwidth and low-latency 5G access through a neutral 5G small cell deployed on board the train. The small cell, which is a part of what we call train neutral MNO, will be connected to a neutral 5G core network on the ground that provides services to MNOs. Providing that the passengers' MNOs have a roaming agreement with the train neutral MNO, the passengers will transparently benefit from a 5G service on board the train.

The high-level functional architecture of UC3, presented in D2.1, is illustrated in Figure 4. It can be observed that:

- The **Private cloud** contains the Train Control Centre that manages the services and the 5G-Core of the neutral MNO.
- The **MEC layer** in each country hosts the AI modules needed by the FRMCS P2 and P3 services and the B1 edge break-out points. The latter allows the Wi-Fi access point on the train to access to internet. In addition, one ACS-GW will be deployed in each country.
- The **Network Infrastructure layer** contains 5G, satellite, and IEEE 802.11ad networks. It should be noted that the IEEE 802.11ad network will be deployed only on the Spanish side.

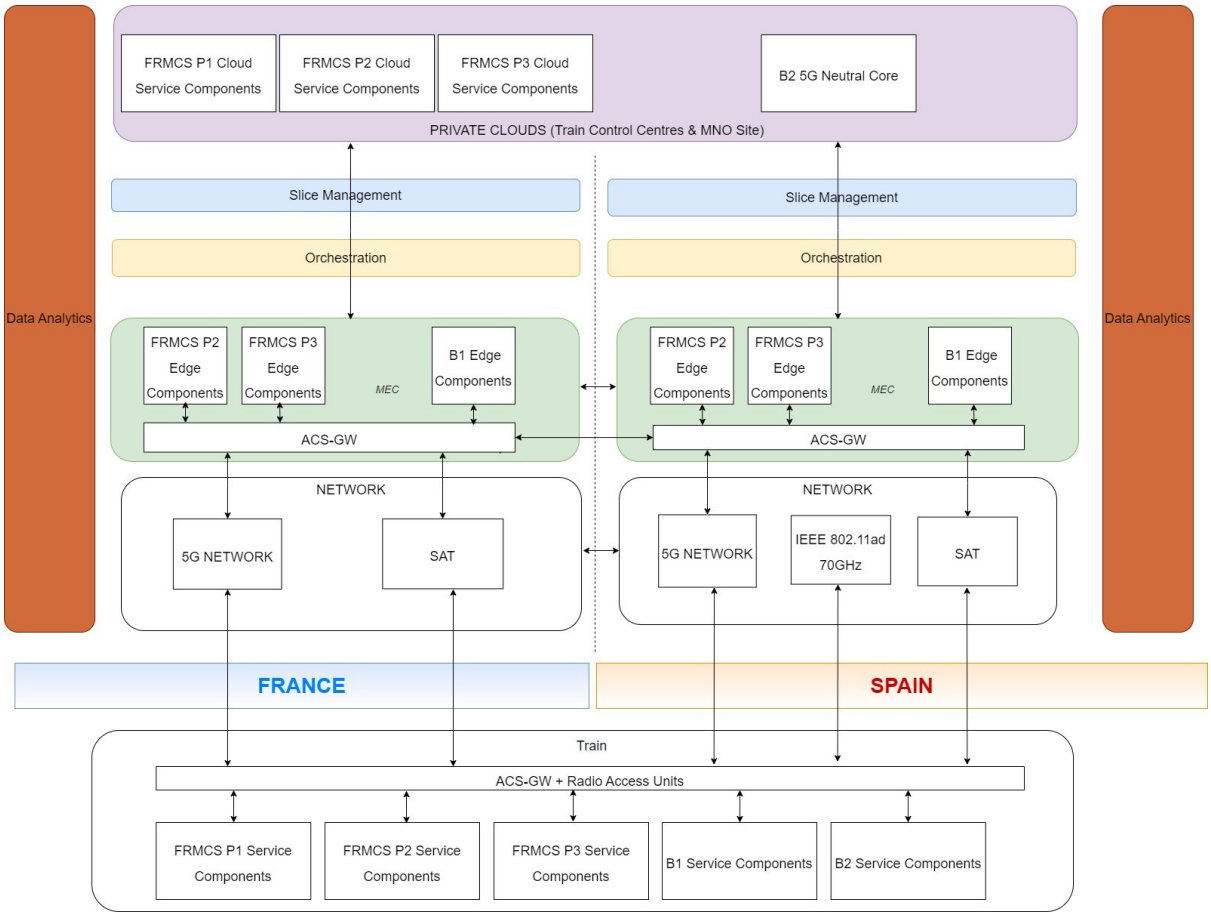

Figure 4. High-level functional architecture of UC3 (extracted from [1]).

As it was described in D2.1, this use case requires very high data rate especially in service B1 and high reliability in service FRMCS P1, P2, and P3, in addition to low latency and low mobility interruption time in all service. Table 3 provides an overview of the **most stringent** service KPIs for UC3 [1].

Table 3. Overview of service KPIs for UC3: FRMCS applications and business service continuity.

KPI	Requirement
Data Rate	Very high data rate to enable all users in the train having high quality Wi-Fi connection
Latency	Low latency for real-time services such as video conferencing in B2
Reliability	High reliability specially to avoid errors in sensor data
Mobility interruption time	Low mobility interruption time is expected especially when crossing the borders or changing RAT

2.1.4. UC4: Follow-Me infotainment

The aim of the follow-me infotainment use case is to distribute several types of high-quality media contents, such as live-streaming of 360º video, video-conferencing, and virtual reality video, synchronously to passengers travelling at high speed by car or train. It is based on two main concepts:

- The follow-me concept [28], which allows the migration of edge applications along the corridor following the user's movements.
- The Infotainment concept, which is a set of advanced media applications that are used to showcase the Follow-me concept.

In this use case, the two services that have been considered are:

- **Enjoy Media Together (EMT),** which is composed of two main functionalities: the EMT video streaming allowing users to create a virtual "video room" to watch high-definition media contents together, and the EMT Video Conferencing, allowing users to interact among them through chat-like applications and video calls.
- **Tour Planning (TP)**, which is composed of three main functionalities that give the users the opportunity to customize the rest of their tour if they decide to make changes before arriving at their destination: the TP high-resolution media allows users to enjoy on-demand high-quality media content (HD, FHD, or UHD/4K resolution videos and photos) related to Points Of Interest (POIs) along their route; the 360-degree high quality videos aim at giving the user a more comprehensive view of some POIs, and the immersive media functionality provides Virtual Reality (VR) video streaming of specific POIs.

Figure 5 shows the mapping between the UC4 main components and the high-level 5GMED network architecture presented in D2.1. As it can be observed, the UEs of the infotainment clients (in the Ground Layer) will be connected to different EMT/TP Edge server instances in the MEC nodes, i.e., providing access close to the service consumers. The EMT/TP Edge servers and the UEs need to reach the EMT/TP Cloud servers (where all media contents are stored), which are deployed at the Cloud (top). The direct access from the UEs to that cloud server will be for those operations that do not require low latency (e.g., the service registration process or media catalogues checking), while the media contents consuming will always be done through one of the instantiated EMT/TP Edge servers to provide lower latency connections. The follow-me concept will be implemented through the media services migration among the different EMT/TP Edge servers, following the end-users' movements. This migration will be triggered from the Data Analytics Module (DAM) through the orchestration function based on UE and network metrics (e.g., informing about the end-user's position and speed). This is intended to avoid service disruptions while the end user is moving, and at the same time, assuring better latency than if the connection where directly to the cloud.

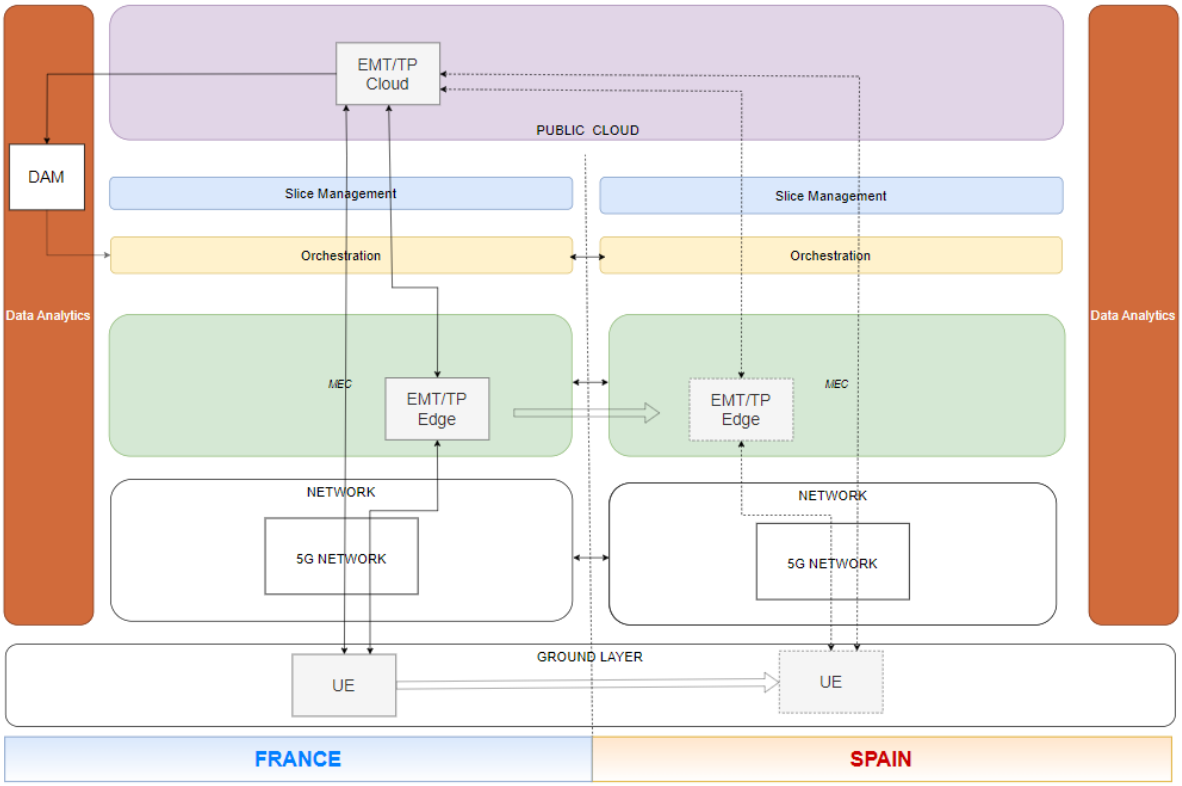

Figure 5. High-level functional architecture of UC4 - automotive scenario (extracted from [1]).

In the railway scenario, the scope of the UC4 services is different and not all the test cases defined for the automotive UC4 will be checked, including all those regarding the follow-me functionality. The approach in the railway scenario will be reduced to an Over-The-Top (OTT) deployment with just some functionalities in the cloud[2]. For this, only the EMT/TP Cloud server will be deployed, leaving aside all the other components (Figure 5). The UEs of UC4 will be eventually connected to the Wi-Fi access point inside the train. As there will be no MEC connection, the testing will include the Mobility interruption time only when roaming.

As described in D2.1, this use case requires very low latency, very high reliability, and low mobility interruption time at the borders and at inter-RAT handover. Otherwise, the services' performance will

[2] UC4 is the only UC in the project with application to both: automotive and railway scenarios, addressed in WP4 and WP5 respectively. However, it has been agreed in the consortium that, following the project reviewers' recommendations, the UC4 deployment should be simplified in WP5 (railway scenario). The main reason for this is to minimize duplication of efforts, and since UC4 can be extensively showcased in WP4 (automotive), the approach in WP5 would be only to do a reduced deployment, with the overall objective of demonstrating that the UC4 infotainment services could also be accessed by train users.

not be acceptable by the users. In addition, the services of this use case are very demanding in terms of data rate. Table 4 provides an overview of the most stringent service KPIs for UC4 [1].

Table 4. Overview of service KPIs for UC4: Follow-Me Infotainment.

KPI	Requirement
Data Rate	Very high data rate is required especially for virtual reality services
Latency	Very low latency is required especially for video conferencing and interactive services
Reliability	Very high reliability is required to provide users with high quality experience
Mobility interruption time	Low mobility interruption time is required especially for real time interactive services

2.2. Network KPIs

This section presents the general network KPIs typically used to characterize the 5G network performance and the effects of roaming due to cross-border mobility. The target values of the network KPIs will be derived from the service KPIs of each use case.

The network KPIs are defined in Section 2.2.1 and their target values are calculated later in Section 2.2.2, Section 2.2.3, Section 2.2.4, and Section 2.2.5, for UC1, UC2, UC3, and UC4, respectively. For each use case, we first describe the breakdown of the end-to-end latencies considered in each service. Following that, the calculation method used to compute each network KPI is presented along with the target values.

Figure 6 depicts a high-level simplified representation of the network segments. This representation will be used to show how we extract network KPIs from service KPIs and includes only the layers involved in the user plane data flow. It should be noted that in the following, the 5GMED network infrastructure layer will refer to all networks connecting vehicles/trains to MEC/clouds, i.e., 5G network, C-V2X, IEEE 802.11ad, and satellite. The elements inside boxes with dashed contours are optional and will be used only for specific services or use cases. More specifically:

- In the automotive scenario, the connected elements (e.g., vehicle, roadside camera) are directly connected to the 5GMED network infrastructure via 5G modems.
- In the railway scenario, the connected elements (e.g., small cell, WiFi Access point, or FRMCS application) are connected to the 5GMED network infrastructure through the ACS-GW.
- In some specific use cases and services, the data traffic from the application client (which is located inside the connected element) can be processed by a backend application allocated at the edge (e.g., in UC2, in the B1 service of UC3, etc.), but in general, the data traffic generated by the application client will bypass the edge.
- Depending on the characteristics of the use case and the scenario (highway or railway), the backend application may be hosted (i) in a public cloud, as in the Valeo Teleoperation Cloud for UC1 or the high-quality Wi-Fi for of UC3; (ii) in a private cloud, as it is the case for the TMC Global of UC2; or (iii) at the edge, as in the case of the V2X Gateway of UC2 and the Infotainment Edge server of UC4.
- In the railway scenario (UC3 and UC4), the path between the connected element in the train and the 5GMED network infrastructure is enabled by the ACS-GW installed on-board the train.

On the other side, the path between the 5GMED network infrastructure and the edge or the transport network is enabled by the ACS-GW on-ground.

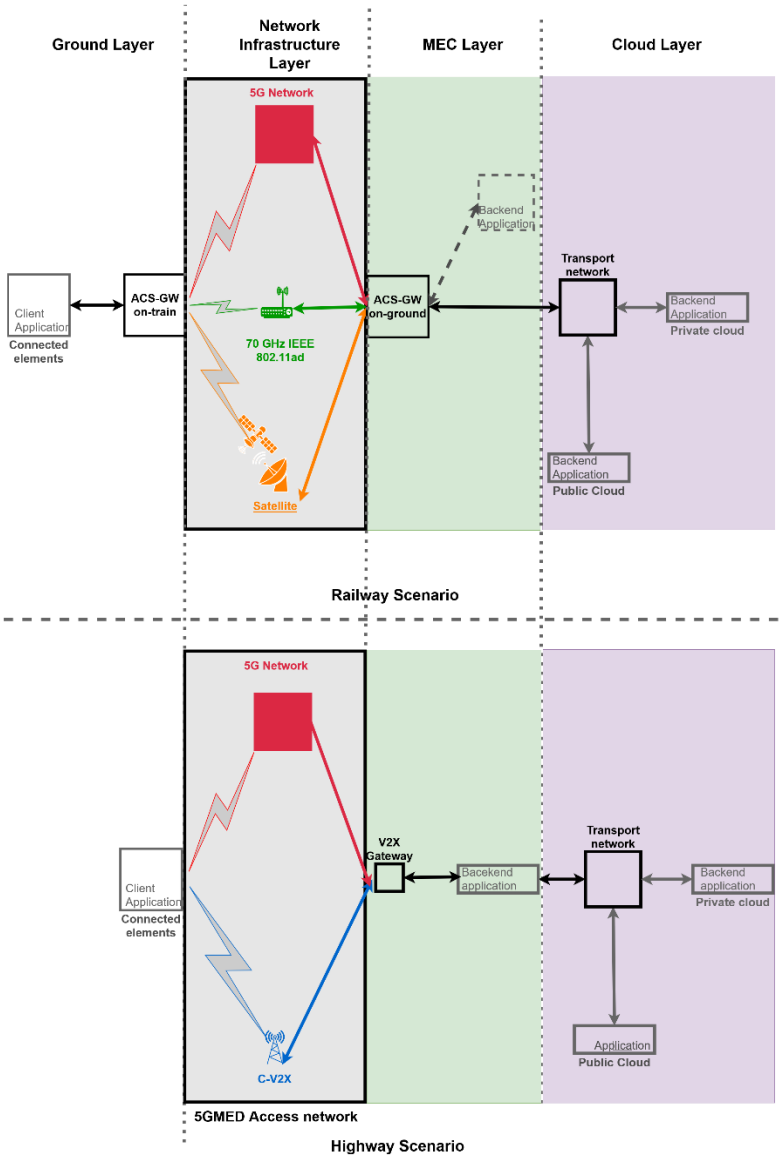

Figure 6. Simplified Representation of the 5GMED network segments.

2.2.1. Definition of Network KPIs

The network KPIs are defined as follows:

- **Downlink Network Latency**: It is defined as the downlink latency that is introduced by the *5GMED network infrastructure, only including 5G, IEEE 802.11ad, satellite, or C-V2X latencies.* In other terms, it is the latency introduced by the 5GMED network infrastructure when packets are transmitted from the backend application to the connected element, as shown in Figure 6. Since we focus on the latency introduced by the 5GMED network infrastructure, if the backend application is hosted in the cloud, then the downlink network latency does not

include the latency introduced by the transport network. Similarly, in the railway scenario, the downlink network latency does not include the latency introduced by the operations of the ACS-WS on-train and ACS-GW on-ground. In UC2, this latency does not include the latency introduced by the V2X Gateway. Later, for each use case, we will explain in detail how the downlink network latency is calculated.

- **Uplink Network Latency**: It is defined as the uplink latency that is introduced by *the 5GMED network infrastructure, only including 5G, IEEE 802.11ad, Satellite, or C-V2X latencies.* In other terms, it is the latency introduced by the 5GMED network infrastructure when packets are transmitted from the connected element to the backend application. Since we focus on the latency introduced by the 5GMED network infrastructure, if the backend application is hosted at the cloud, then the uplink network latency does not include the latency introduced by the transport network. Similarly, in the railway scenario, the uplink network latency does not include the latency introduced by the operations of the ACS-WS on-train and ACS-GW on-ground. In UC2, this latency does not include the latency introduced by the V2X Gateway. Later, for each use case, we will explain in detail how the uplink network latency is calculated.
- **Downlink Network Reliability:** It is defined as the ratio between the number of packets successfully received without errors by the connected element (or ACS-GW on-train in the railway scenario) divided by the total number of packets transmitted by the backend application.
- **Uplink Network Reliability:** It is defined as the ratio between the number of packets successfully received without errors by the backend application (or ACS-GW on-ground in the railway scenario) divided by the total number of packets transmitted by the connected element.
- **Uplink Jitter**: It is defined as the mean difference in uplink network latency between the fastest and slowest packets received by the backend application (or ACS-GW on-ground in the railway scenario).
- **Downlink Jitter**: It is defined as the mean difference in downlink network latency between the fastest and slowest packets received by the connected element (or ACS-GW on-train in the railway scenario).
- **Uplink data rate**: It is defined as the amount of data bits transmitted by all simultaneously connected elements (or ACS-GW on-train in the railway scenario) within a certain time window. In another way, this is the application-level data rate seen by the 5GMED network infrastructure. The uplink data rate is calculated from the service data rate, denoted as *d*, by considering the length of the headers added by the transmitting elements and the headers added by the ACS-GW in the railway scenario. The headers added by the transmitting elements are composed of an IPv4 header (i.e., 20 bytes), denoted as *I*, and a transport header, denoted as *T*. The latter can be a TCP header (i.e., 20 bytes) or an UDP header (i.e., 8 bytes). The length of the header added by the ACS-GW, denoted as *A,* is 28 bytes. It should be noted that we assume the maximum transmission unit (MTU) as our packet size *P* inside the 5GMED network infrastructure, and it is considered to be 1500 bytes. The calculation also considers the number of individual transmitters *n*. The uplink packet data rate can be expressed as:

$$D = d \times n \times \frac{P}{P - I - T - A} \qquad \text{(Equation 1)}$$

- **Downlink data rate:** It is defined as the amount of data bits transmitted by the backend application (or ACS-GW on-ground in the railway scenario) within a certain time window. It is calculated as the uplink data rate.

- **Mobility interruption time:** It is defined as the maximum time interval in which there is no connection between the connected element and the backend application, or between the two ACS-GWs in the railway scenario.
- **Service Migration time:** It is defined as the time elapsed since the migration of a service (available from a specific edge server) starts until the service is available and running in the target edge server.

As an exception, the values of the following network KPIs are considered equal to the values of the corresponding service KPIs: downlink/uplink network reliability, uplink/downlink jitter, mobility interruption time, and service migration time. Therefore, the derived network KPIs are considered in the worst scenario where all the delay and packet losses will occur only in the 5GMED network infrastructure and not in any other segment.

In the rest of this section, we compute the target values of the network KPIs from the service KPIs considered in each use case. It should be noted that all network KPI values provided in this section are a first estimation and might change slightly during the project due to the optimisation and development that will be performed. They are given here as reference values for the design the 5GMED network infrastructure. In addition, the parameters used in the calculations depend on the ongoing optimization of the network architecture and configuration. Therefore, they might also change during the project lifetime.

2.2.2. UC1: Remote driving

In this section, we first describe the breakdown of the end-to-end latencies of UC1 (defined in D2.1), present the calculation method used to compute each network KPI, and finally provide the target values.

In Figure 7, we depict the different segments that introduce latency in the data flows of UC1 services. It should be noted that in this representation, only the layers related to the flow of user plane traffic in the use case for one country (see Figure 2) is depicted to clearly show how the network latency is calculated. As it can be observed, the service KPI *data end-to-end latency* (Data_L), defined in D2.1, is composed of the *uplink network latency* (UL_L) and the *uplink cloud latency* (UL_Cloud_L), which is the latency between the 5GMED network infrastructure and the remote station in the Valeo Teleoperation Cloud.

The service KPI *command E2E latency* (Command_L) is composed of the *downlink network latency* (DL_L) and the *downlink cloud latency* (DL_Cloud_L), which is the latency between the remote station and the 5GMED network infrastructure.

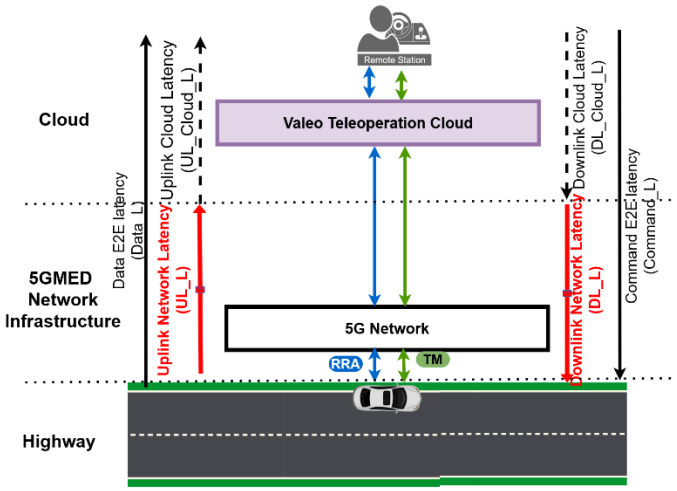

Figure 7. End-to-end latencies break-down in UC1.

In Table 5, we explain how the network KPIs are calculated based on the service KPIs of UC1 defined in D2.1 [1].

Table 5. UC1 network KPI calculation method.

KPI Name	Unit	Calculation method
Downlink Network Latency	[ms]	It is calculated from the command end-to-end latency (Command_L), defined in D2.1 for UC1, using the following equation (see Figure 7): • DL_L = Command_L – DL_Cloud_L
Uplink Network Latency	[ms]	It is calculated from the data end-to-end latency (Data_L), defined in D2.1 for UC1, using the following equation (see Figure 7): • UL_L = Data_L – UL_Cloud_L
Downlink Network Reliability	[%]	It is equal to the command reliability KPI defined in D2.1 for UC1.
Uplink Network Reliability	[%]	It is equal to the sensing reliability KPI defined in D2.1 for UC1.
Uplink Data-Rate	[Mbps]	Calculated by applying (Equation 1 on the uplink service data rate KPI, defined in D2.1 for UC1, and considering TCP in the transport layer.
Downlink Data-Rate	[Mbps]	Calculated by applying (Equation 1 on the downlink service data rate KPI, defined in D2.1 for UC1, and using TCP in the transport layer.
Mobility interruption time	[s]	It is the same as in the service KPI with the same name defined in D2.1 for UC1.

During preliminary tests, it was found that the average value of the uplink and downlink cloud latencies (UL_Cloud_L and DL_Cloud_L) is 15 ms.

In Table 6, we provide the target values of the network KPIs for UC1. As detailed above, they were computed using the target values of the service KPIs of UC1 defined in D2.1.

Table 6. UC1 network KPI values.

Network KPI Name	Service 1: Minimum Risk Maneuver	Service 2: Request for Remote Assistance	Service 3: Teleoperation Maneuver
Downlink Network Latency	-	-	Between 5 and 35 ms
Uplink Network Latency	-	90 ms	90 ms
Downlink Network reliability	-	-	99 %
Uplink Network reliability	-	-	95 %
Uplink packet Data-Rate	-	5.15 Mbps(1)	10.3 Mbps(1)
Downlink packet Data-Rate	-	0.5 Mbps	1 Mbps
Mobility interruption time	-	1 s	100 ms

(1) One connected autonomous vehicle as defined in D2.1

2.2.3. UC2: Road infrastructure digitalisation

In this section, we first describe the breakdown of the end-to-end latencies of UC2 services (defined in D2.1), then we present the calculation method used to compute each network KPI, and we finally provide the target values.

In Figure 8, we depict the different segments of latencies that are experienced by the data flows in the REM service. It should be noted that in this representation, only the layers related to the flow of user plane traffic in the use case for one country (see Figure 3) is depicted to clearly show how the network latency is calculated. As it can be seen, the service KPI *Hazard End-to-End latency* (Hazard_E2E_L) is composed of the *uplink network latency* (UL_L), the *V2X Gateway processing time* (GW_P) in the uplink direction, the *TMC edge processing time* (Edge_P), the *V2X Gateway processing time* (GW_P) in the downlink direction, and the *downlink network latency* (DL_L). We assume that the values of the *uplink network latency* and the *downlink network latency* are equal to be able to compute their target value.

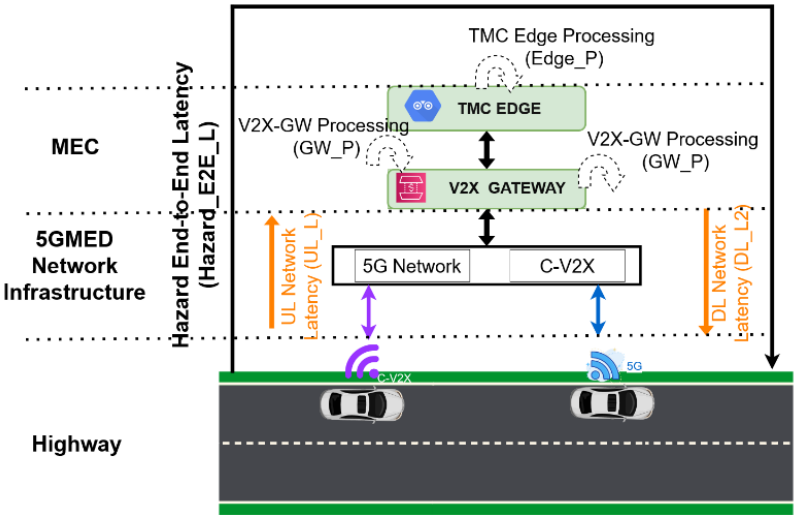

Figure 8. End-to-end latencies break-down in the REM service of UC2.

In the AID service, the roadside cameras are always sending their video streams to the TMC edge, which is responsible for detecting the hazard. Therefore, the *hazard end-to-end latency* (Hazard_E2E_L), defined in D2.1 for UC2, is composed of the following segments of latencies that are experienced by data flows in this service as depicted in Figure 9: the *TMC edge processing time* (Edge_P), the *V2X-GW processing time (GW_P)*, and the *downlink network latency* (DL_L).

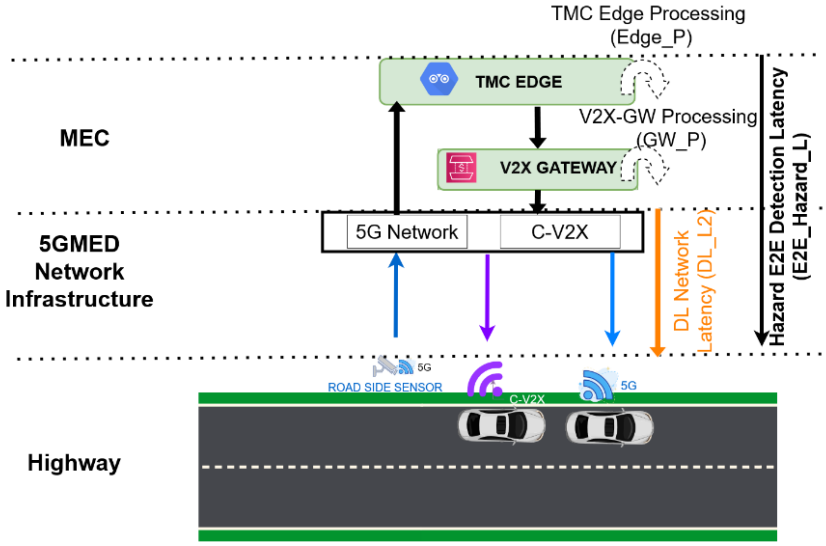

Figure 9. End-to-end latencies break-down in the AID service of UC2.

In the three services, an abnormal situation (e.g., a vehicle driving at very low speed) will be detected by either the TMC Edge, using the video streams sent by roadside cameras, or by connected vehicles. Therefore, the *Traffic Regulation End-to-End latency* (Regulation_L) can be calculated as depicted in Figure 10. The Regulation_L is composed of the *TMC edge processing time* (Edge_P), the *Edge to Global Latency* (E2G_L), the *TMC Global Latency* (Global_L), the *Global to Edge Latency* (G2E_L), the *V2X Gateway processing time* (GW_P), and the *downlink network latency* (DL_L).

For the REM and AID services, we have two service KPIs that includes downlink network latencies (i.e., Hazard End-to-End detection latency and Traffic Regulation End-to-End latency). Therefore, the downlink network latency for these two services is computed for each service KPI as described above and the minimum between them is considered as the network latency (i.e., the most stringent one).

We assume that the TMC Edge processing time in the downlink is negligible as it only forwards the messages received from the TMC Global.

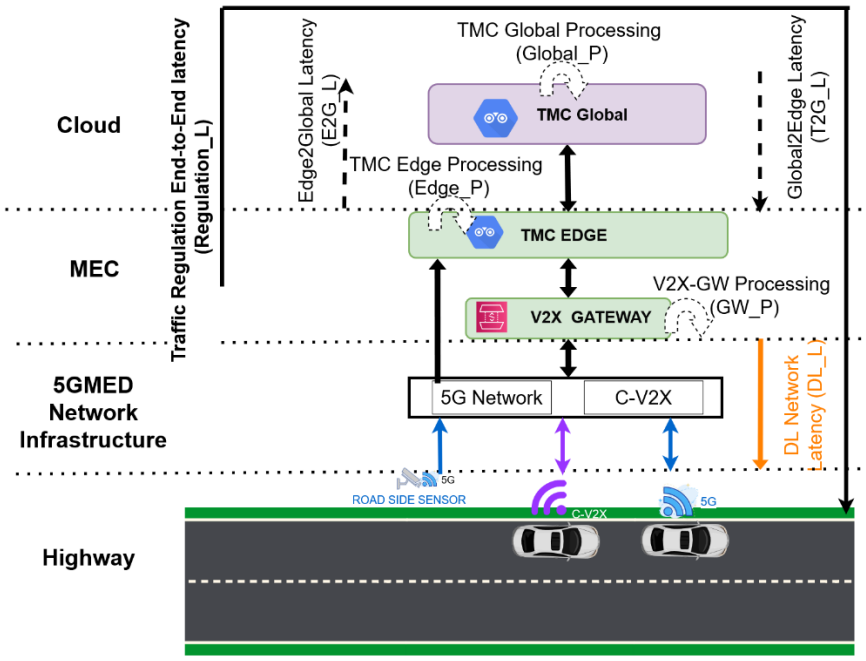

Figure 10. Traffic regulation End-to-end latencies break-down in the three service of UC2.

In Table 7, we explain how the network KPIs used for UC2 are calculated based on the service KPIs of UC2 defined in D2.1 [1].

Table 7. UC2 network KPI calculation method.

Network KPI Name	Unit	Calculation method
Uplink Network Latency	[ms]	It is calculated from Hazard_E2E_L and Regulation_L, defined in D2.1 for UC2, using the following equations (see Figure 8): • REM service: UL_L = (Hazard_E2E_L – 2xGW_P – Edge_P)/2
Downlink Network Latency	[ms]	It is calculated from Hazard_E2E_L and Regulation_L, defined in D2.1 for UC2, using below equations (see Figure 8, Figure 9, and Figure 10): • REM service: DL_L is the minimum between (Hazard_E2E_L – 2xGW_P – Edge_P)/2 and (Regulation_L – Edge_P – E2G_L – Global_P – G2E_L – GW_P) • AID service: DL_L is the minimum between Hazard_E2E_L – GW_P – Edge_P and (Regulation_L – Edge_P – E2G_L – Global_P – G2E_L - GW_P) • TFR service: DL_L = Regulation_L – Edge_P – E2G_L – Global_P – G2E_L – GW_P
Network Reliability (in uplink and downlink)	[%]	It is the minimum value between the Hazard Notification Reliability and Traffic Regulation Reliability defined in D2.1 for UC2.
Uplink Data-Rate	[kbps]	Calculated by applying (Equation 1 on the uplink service data rate, defined in D2.1 for UC2, and using UDP.
Downlink Data-Rate	[Kbps]	Calculated by applying (Equation 1 on the downlink service data rate, defined in D2.1 for UC2, and using UDP. The number of users will be considered as 1 because the same message is forwarded to all users (broadcast message).
Mobility interruption time	[s]	It is the same as in the service KPI with the same name defined in D2.1 for UC2.

The following values of processing times were measured during preliminary tests:

- TMC edge processing time (Edge_P): 100 ms taking as a reference Lenovo ThinkSystem SE350 with an NVIDIA Tesla T4 GPU.
- V2X Gateway processing time (GW_P) based on previous experience in other projects: 30 ms.
- Round-trip latency between TMC Edge and TMC global (E2G_L + G2E_L): 40 ms.
- TMC global processing time (Global_P): 150 ms based on the experience of the Lenovo ThinkSystem SE350 in Castellolí.

In Table 8, we provide the target values of the network KPIs for UC2. They have been computed using that target values of the service KPIs of UC2 defined in D2.1.

Table 8. UC2 network KPI values.

Network KPI Name	**Target Value**		
	Service 1: Relay of emergency messages	**Service 2: Automatic Incident Detection (AID)**	**Service 3: Traffic Flow Regulation (TFR)**
Uplink Network Latency	20 ms	-	-
Downlink Network Latency	20 ms	70 ms	280 ms
Network Reliability (in uplink and downlink)	99.9 %	99.9 %	-
Uplink Data-Rate	51 kbps(1)	82 Mbps(1)(2)	82 Mbps(1)(2)
Downlink Data-Rate	7.1 Kbps	7.1 kpbs	3 kpbs
Mobility interruption time	80 ms	100 ms	100 ms

(1) One connected autonomous vehicle and four connected vehicles as defined in D2.1
(2) Eleven roadside cameras as defined in D2.1

2.2.4. UC3: FRMCS applications and business service continuity

In this section, we first describe the breakdown of the end-to-end latencies of UC3 services (defined in D2.1), then we present the calculation method used to compute each network KPI, and we finally provide the target values.

In Figure 11, Figure 12, and Figure 13, we depict the different segments that introduce latency in the data flows of UC3 services. It should be noted that in this representation, only the layers related to the flow of user plane traffic in the use case for one country (see Figure 4) is depicted to clearly show how the network latency is calculated.

FRMCS P1 and FRMCS P2

As it can be observed in Figure 11 for the FRMCS P1 service, the backend application is allocated in the cloud and the *uplink cloud end-to-end latency* (UL_Cloud_E2E_L) is composed of the ACS-GW on-train processing time (ACS_T_P), the *uplink network latency* (UL_LP1), the *ACS-GW on-ground processing time* (ACS_G_P), and the *edge to cloud latency (E2C_L)*. For the *downlink network latency* (DL_LP1), we can have the same reasoning on *downlink cloud end-to-end latency* (DL_Cloud_E2E_L).

As can be observed in Figure 11 for the FRMCS P2 service, the backend application is in the edge. Therefore, the *uplink edge end-to-end latency* (UL_Edge_E2E_L) is composed of the ACS-GW on-train

processing time (ACS_T_P), the *uplink network latency* (UL_LP2), and the *ACS-GW on-ground processing time* (ACS_G_P). For the *downlink network latency* (DL_LP2), we can have the same reasoning on *downlink edge end-to-end latency* (DL_Edge_E2E_L).

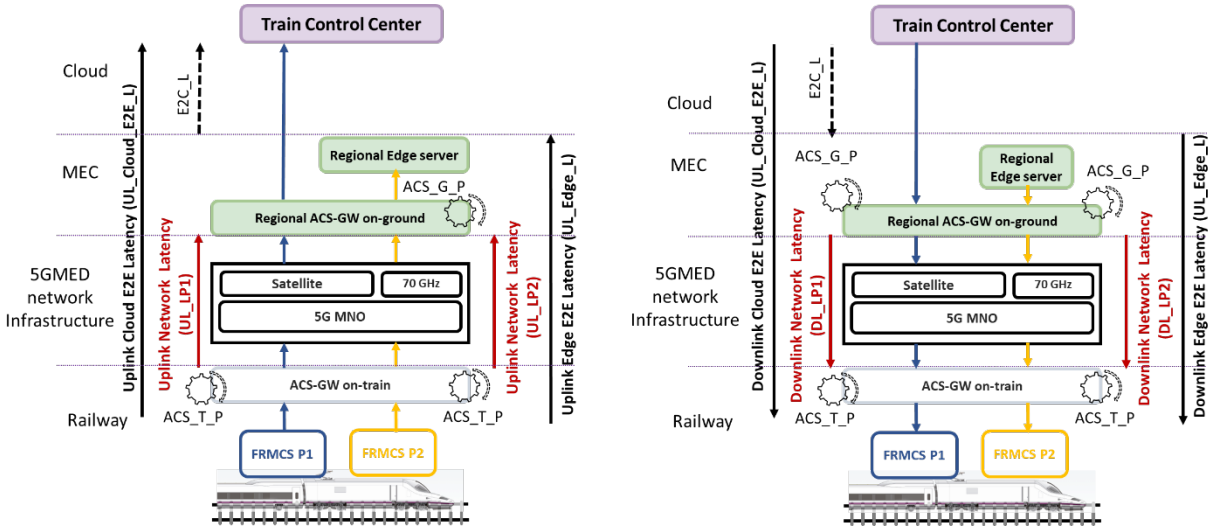

Figure 11. End-to-end latencies break-down in FRMCS P1 service and FRMCS P2 service of UC3 for uplink (left) and downlink (right).

FRMCS P3

As it can be observed in Figure 12 for the FRMCS P3 service, there is one data flow between the application client in the train and a backend application in the edge, and a second data flow between the edge and a second backend application in the cloud. For the first data flow, the *uplink cloud end-to-end latency* (UL_Cloud_E2E_L) is composed of the ACS-GW on-train processing time (ACS_T_P), the *uplink network latency* (UL_LP3_1), the *edge to cloud latency (E2C_L)*, and the *ACS-GW on-ground processing time* (ACS_G_P). For the second flow, the *uplink edge end-to-end latency* (UL_Edge_E2E_L) is composed of the ACS-GW on-train processing time (ACS_T_P), the *uplink network latency* (UL_LP3_2), and the *ACS-GW on-ground processing time* (ACS_G_P). For the *downlink network latency* (DL_LP3), we can have the same reasoning on *downlink cloud end-to-end latency* (DL_Cloud_E2E_L). and *downlink edge end-to-end latency* (DL_Edge_E2E_L).

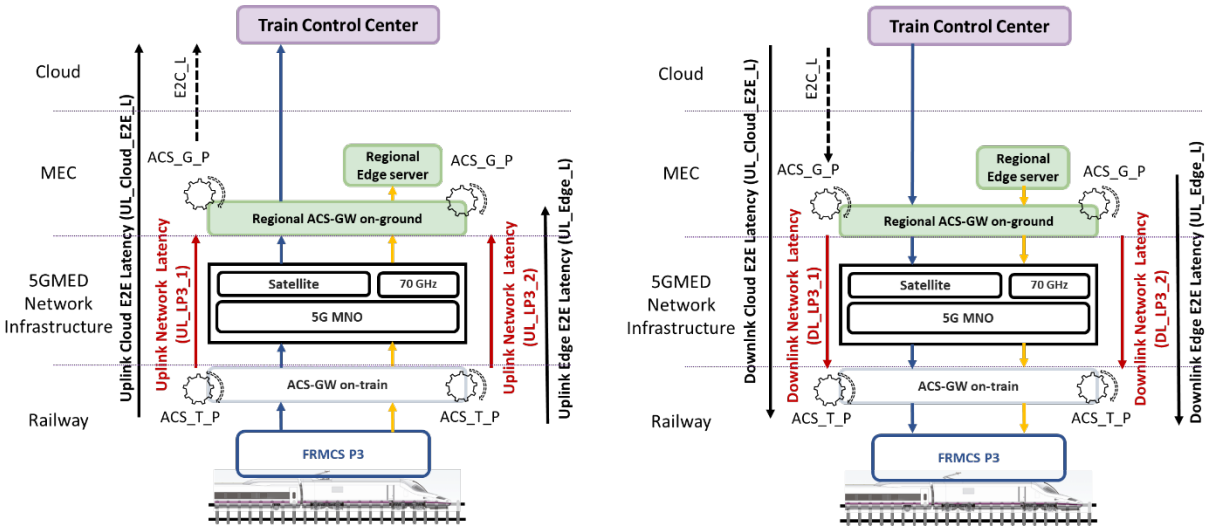

Figure 12. End-to-end latencies break-down in FRMCS P3 service of UC3 for uplink (left) and downlink (right).

Train Business Services B1 and B2

As can be observed in Figure 13 for the high-quality Wi-Fi for passengers, the backend application is in the edge. Therefore, the *uplink edge end-to-end latency* (UL_Edge_E2E_L) is composed of the *ACS-GW on-train processing time* (ACS_T_P), *uplink network latency* (UL_LB1), and the *ACS-GW on-ground processing time* (ACS_G_P). For the *downlink network latency* (DL_LB1), we can have the same reasoning on *downlink edge end-to-end latency* (DL_Edge_E2E_L).

As it can be observed in Figure 13 for the multi-tenant mobile service, the destination will be another network (e.g., internet or another PLMN) through the core network of the neutral operator. Therefore, in addition to the delay in the 5GMED network infrastructure, the data will have delays due to home routed roaming between the neutral core network and the home core network, in addition to the delay inside the latter. In Figure 13, we can see that End-to-End latency between UE's using the 5G small-cell o-board (E2E_5G_L) is composed of the *uplink network latency* (UL_LB2), the *edge to cloud latency (E2C_L)*, the *ACS-GW on-ground processing time* (ACS_G_P), the ACS-GW on-train processing time (ACS_T_P), and the *other_latencies* (O_L) due to the connection to the other network. For the *downlink network latency* (DL_LB2), we can have the same reasoning on End-to-End latency between UE's using the 5G small-cell o-board (E2E_5G_L).

It should be noted that the network latency calculated here does not account for roaming interruption time, as these two KPIs are represented in a separate network KPI.

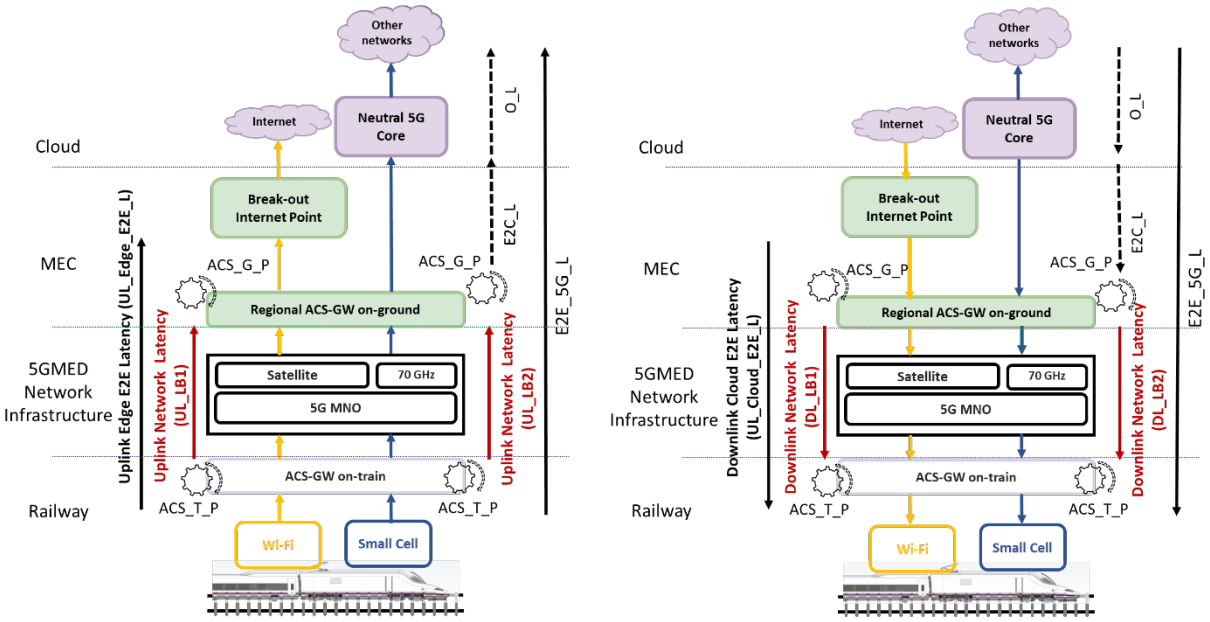

Figure 13. End-to-end latencies break-down in B1 and B2 services of UC3 for uplink (left) and downlink (right).

In Table 9, we explain how the network KPIs used for UC3 are calculated based on service KPIs of UC3 defined in D2.1.

Table 9. UC3 Network KPI calculation method.

Network KPI Name	Unit	Calculation method
Uplink Network Latency	[ms]	For the FRMCS P1 service, it is calculated from the uplink cloud latency (UL_Cloud_L) defined in D2.1 using the following equation (see Figure 11): • UL_LP1 = UL_Cloud_L – ACS_T_P – ACS_G_P – E2C_L For the FRMCS P2 service, it is calculated from the uplink edge latency (UL_Edge_L) defined in D2.1 using the following equation (see Figure 11): • UL_LP2 = UL_Edge_L – ACS_T_P – ACS_G_P For the FRMCS P3 service, it is calculated from the uplink cloud latency (UL_Cloud_L) and uplink edge latency (UL_Edge_L) defined in D2.1 using the following equation (see Figure 12): • UL_LP3 = min (UL_LP3_1, UL_LP3_2) • UL_LP3_1 = UL_Cloud_L – ACS_T_P – ACS_G_P – E2C_L • UL_LP3_2 = UL_Edge_L – ACS_T_P – ACS_G_P For the B1 service, it is calculated from the uplink edge latency (UL_Edge_L) defined in D2.1 using the following equation (see Figure 13): • UL_B1 = UL_Edge_L – ACS_T_P – ACS_G_P For the B2 service, it is calculated from the End-to-End latency between UE's using the 5G small-cell o-board (E2E_5G_L) defined in D2.1 using the following equation (see Figure 13): • UL_B2 = E2E_5G_L – ACS_T_P – ACS_G_P – E2C_L – O_L

Network KPI Name	Unit	Calculation method
Downlink Network Latency	[ms]	For the FRMCS P1 service, it is calculated from the downlink cloud latency (DL_Cloud_L) defined in D2.1 using the following equation (see Figure 11): • DL_LP1 = DL_Cloud_L – ACS_T_P – ACS_G_P – E2C_L For the FRMCS P2 service, it is calculated from the downlink edge latency (DL_Edge_L) defined in D2.1 using the following equation (see Figure 11): • DL_LP2 = DL_Edge_L – ACS_T_P – ACS_G_P For the FRMCS P3 service, it is calculated from the downlink cloud latency (DL_Cloud_L) and downlink edge latency (DL_Edge_L) defined in D2.1 using the following equation (see Figure 12): • DL_LP3 = min (DL_LP3_1, DL_LP3_2) • DL_LP3_1 = DL_Cloud_L – ACS_T_P – ACS_G_P – E2C_L • DL_LP3_2 = DL_Edge_L – ACS_T_P – ACS_G_P For the B1 service, it is calculated from the downlink edge latency (DL_Edge_L) defined in D2.1 using the following equation (see Figure 13): • DL_B1 = DL_Edge_L – ACS_T_P – ACS_G_P For B2, it is calculated from the End-to-End latency between UE's using the 5G small-cell o-board (E2E_5G_L) defined in D2.1 using the following equation (see Figure 13): • DL_B2 = E2E_5G_L – ACS_T_P – ACS_G_P – E2C_L – O_L
Uplink Network Reliability	[%]	It is equal to the service KPI uplink reliability defined in D2.1 for UC3.
Downlink Network Reliability	[%]	It is equal to the service KPI downlink reliability defined in D2.1 for UC3.
UL Jitter	[ms]	It is equal to the same as service KPI defined in D2.1 for UC3.
DL Jitter	[ms]	It is equal to the same as service KPI defined in D2.1 for UC3.
Uplink Data-Rate	[Mbps]	It is calculated by applying (Equation 1 on the service KPI uplink service data rate defined in D2.1 for UC3. All services use TCP, except the P1 service that uses UDP. Then, we have to sum over all services.
Downlink Data-Rate	[Mbps]	It is calculated by applying (Equation 1 on the service KPI downlink service data rate defined in D2.1 for UC3. All services use TCP, except the P1 service that uses UDP. Then, we have to sum over all services. Here, we multiply by the number of users only in B1 service.
Mobility interruption time	[s]	It is the same as in the service KPI defined in D2.1 for UC3.

The following values were measured during preliminary tests:

- Processing time in the ACS-GW on-ground and ACS-GW in-train (ACS_G_P, ACS_T_P): 1 ms in average. To give an estimate of the latency introduced by the ACS-GW processing, we measured the latency of a comparable application performing similar processing tasks as we still did not test the solution with the three technologies. A similar application is the XDP Katran Load Balancer [4], which performs similar functions: (i) parsing; (ii) flow tracking; (iii) encapsulation; and (iv) forwarding. This use case was measured on an environment similar to

the one that will be used to execute the ACS-GW; a Linux server with Intel Xeon Silver 4110 CPU clocked at 2.1GHz and an Intel X710 10G NIC.
Such a use case introduces an end-to-end delay of about 0.1 ms. We took a conservative approach by indicating a large 0.1 – 1 ms latency range for the following reasons: (i) we take into account a worst case scenario in which the ACS-GW processing requires more operations (e.g. classification); (ii) the variation of latency in XDP can be considerable because of the nature of software processing in the kernel, which can occasionally give rise to latency spikes; (iii) the test was single core. When going to multicore, we might have some extra latency introduced by possible locking of shared data; (iv) we are not sure if we need extra primitives to support the issues that are still under development.

- Latency from the edge to the cloud (E2C_L): 45 ms in average. The tests were done in early stage of the project and may be reduced as the project advances.
- Other latencies in the small cell scenario: In [2], the latency when roaming with home routed roaming in Europe is on average 100 ms, whereas the latency average would be 30 ms if there was no roaming. This means that roaming introduces around 70 ms additional latency. Therefore, one can estimate the average latency inside one operator to be around 30/2 = 15 ms when there is no roaming. Therefore, the additional latency will be 70 + 15 = 85 ms. It should be noted that the values presented in [3] are not for 5G networks, as there have been no mass deployments of the 5G SA networks, in addition to the fact that this is done using home routing roaming without optimization. Therefore, this can be considered as the most stringent scenario for determining 5G network requirements.

In Table 10, we provide the target values of the network KPIs for UC3. They have been computed using the target values of the service KPIs of UC3 defined in D2.1.

Table 10. UC3 Network KPI values.

Network KPI Name	**Target Value**				
	FRMCS P1	**FRMCS P2**	**FRMCS P3**	**B1**	**B2**
Uplink Network Latency	953 ms	198 ms	198 ms	98 ms	68 ms
Downlink Network Latency	953 ms	198 ms	198 ms	98 ms	68 ms
Uplink Network Reliability	99.9 %	99 %	99 %	98 %	97 %
Downlink Network Reliability	99.9 %	99 %	99 %	98 %	97 %
UL Jitter	-	-	-	5 ms	40 ms
DL Jitter	-	-	-	5 ms	40 ms
Uplink Data-Rate	5.2-6.2 Mbps[(1)]	10.5-41.9 Mbps[(2)]	4.3-17.3 Mbps[(3)]	52.4 – 65.5 Mbps[(4)]	0.62 – 4 Mbps
	Total between 5.2 + 10.5+ 4.3 + 52.4 + 0.62 = 73 Mbps and 6.2 + 41.9 + 17.3 + 65.5 + 4 = 134.9 Mbps				
Downlink Data-Rate	-	-	-	52.4 – 65.5 Mbps	0.62 – 3.1 Mbps
	Total between 5.2 + 52.4 + 0.62 = 58.22 Mbps and 5.2 + 65.5 + 3.1 = 73.8 Mbps				
Mobility interruption time	1 s	1 s	10 s	10 s	1 s

(1) Equivalent to the traffic generated by 2500 sensors' data parameters that are periodically transmitted to ground every 100 ms.
(2) Two Lidar sensors as described in D2.1.
(3) Two cameras as described in D2.1.
(4) The target value provided in D2.1 refers to the maximum value defined for a complete train (12 coaches), whereas we are dimensioning our network for one coach.

2.2.5. UC4: Follow-ME Infotainment

In this section, we first describe the breakdown of the end-to-end latencies of UC4 services (defined in D2.1), then we present the calculation method used to compute each network KPI, and we finally provide the target values.

In Figure 14, we depict the different segments that introduce latency in the data flows of UC4 services. It should be noted that in this representation, only the layers related to the flow of user plane traffic in the use case for one country (see Figure 5) is depicted to clearly show how the network latency is calculated. The figure represents the end-to-end latency break-down for both: highway and railway scenarios. However, the relevant numbers here should be those in the highway scenario, since (as introduced in Section 2.1.4) the deployment in the railway scenario has been limited to just showcasing the basic functionality (i.e., in the railway scenario, the UC4 service deployment will be only in the cloud, and the connection of UEs will be performed through the train Wi-Fi), so that higher latency values than in the highway scenario should be expected. As can be seen in the highway scenario, the service *end-to-end latency* (E2E_L) is composed of the *application processing time* (App_P) and the *network latency* (N_L2). On the other hand, in the railway scenario the service *end-to-end latency* (E2E_L) is composed of the *application processing time* (App_P), the *ACS-GW on-train processing time* (ACS_T_P), the *network latency* (N_L1), the *ACS-GW on-ground processing time* (ACS_G_P), and the *edge to cloud latency (E2C_L)*.

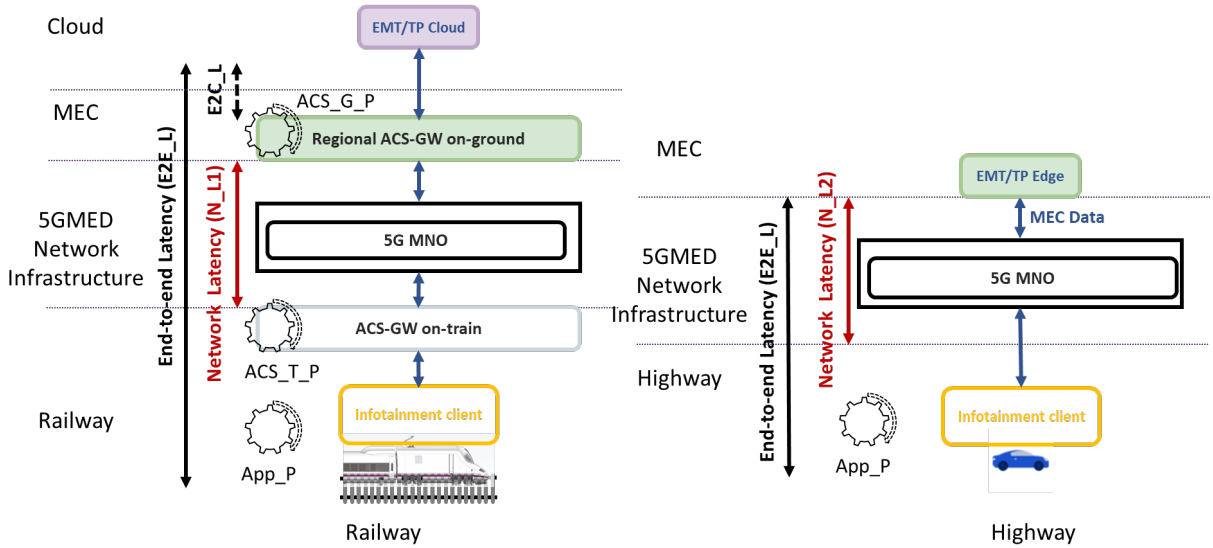

Figure 14. End-to-end latencies break-down in UC4

In Table 11, we explain how the network KPIs used for UC4 are calculated based on service KPIs of UC4 defined in D2.1.

Table 11. UC4 Network KPI calculation method.

Network KPI Name	Unit	Calculation method
Network Latency (uplink and downlink)	[ms]	It is calculated from the service KPI end-to-end latency (E2E_L), defined in D2.1 for UC4, using the following equation (see Figure 14): • For the railway scenario: N_L1 = E2E_L – App_P – ACS_T_P – ACS_T_P – E2C_L • For the highway scenario: N_L2 = E2E_L
Data rate (uplink and downlink)	[Mbps]	Calculated by applying (Equation 1 on the KPI uplink service data rate, defined in D2.1 for UC4, and using UDP. All services use TCP except the videoconferencing that uses UDP. As a user can use only one server, the total data rate will be the maximum between the data rates of all services. In the train scenario, the ACS-GW header should be taken into account. The maximum between the two scenarios will be considered.
Jitter (uplink and downlink)	[ms]	The same as the service KPI jitter defined in D2.1 for UC4.
Mobility Interruption time	[ms]	The same as the service KPI mobility interruption time defined in D2.1 for UC4.
Reliability	[%]	It is equal to the service KPI reliability defined in D2.1 for UC4.
Service Migration Time	[s]	The same as the KPI service migration time defined in D2.1 for UC4.

The value of the application processing time (App_P) is based on the values in [5][6], where it is considered to be 15 ms. This value might change as the application is still under development. The values of E2C_L, ACS_T_P, and ACS_G_P are the same as in UC3. It should be noted that in the highway we use the upper bound of E2E_L because the application server in the cloud and the application cannot have the best performance.

Table 12 and Table 13 provide the target values of the network KPIs for the two services in UC4 (EMT and TP). They result from the target values of the service KPIs defined in D2.1. The intention is to test those values with up to 5 simultaneous users accessing the service, as defined in D2.1.

Table 12. Network KPI values for the EMT service in UC4

Network KPI Name	Target Value	
	Functionality 1: Video Streaming	Functionality 2: Video conferencing
Network Latency (uplink and downlink)	< 4 s	Railway: 23 ms Highway: 5 - 85 ms
Data rate (uplink and downlink)	> 105 Mbps	> 5.2 Mbps
	Total: Max(105, 5.2) = 105 Mbps	
Jitter	< 400 ms	< 2 ms
Mobility Interruption time	< 1000 ms	< 100 ms
Reliability	99.9 %	99.9 %
Service Migration Time	20 – 35 s	20 – 35 s

Table 13. Network KPI values for TP service in UC4

Network KPI Name	Target Value		
	Functionality 1: High resolution media	**Functionality 2: 360 high resolution media**	**Functionality 3: Immersive media**
Network Latency (uplink and downlink)	< 1 s	Railway: 23 ms Highway: 5 - 85 ms	Railway: 3 ms Highway: 5 - 65 ms
Data rate (uplink and downlink)	105 Mbps	523 Mbps	523 Mbps
	Total: max(106, 523, 523) = 523 Mbps		
Jitter	10-50 ms	< 10 ms	< 10 ms
Mobility Interruption time	< 1000 ms	< 30 ms	< 30 ms
Reliability	99.9 %	99.9 %	99.9 %
Service Migration Time	20 – 35 s	20 – 35 s	20 – 35 s

2.3. Network Requirements

The network architecture of 5GMED must satisfy ***the most stringent*** requirements of all use cases and their services as the different use cases will be demonstrated separately. We present in Table 14 and Table 15, respectively, the most stringent target values of the network KPIs for the highway and railway scenarios. These target values will be considered in the design of the 5GMED network infrastructure. For each network KPI, the table details the target value selected from the most demanding use case, i.e., minimum latency, maximum reliability, maximum data-rate, minimum mobility interruption time, minimum jitter, and minimum service migration time. It should be noted that for the design of the network slices, the target KPIs shall be extracted from Section 2.2 for each use case.

Table 14. Most stringent network requirements for the highway scenario.

Network KPI Name	Target Value (Use case)
Downlink Network Latency	5 – 35 ms (UC1 and UC4)
Uplink Network Latency	5 – 65 ms (UC4)
Downlink Network Reliability	99.9% (UC2, UC4)
Uplink Network Reliability	99.9% (UC2, UC4)
Uplink Data-Rate	523 Mbps (UC4)
Downlink Data-Rate	523 Mbps (UC4)
Mobility interruption time	30 ms (UC4)
Uplink Jitter	2 ms (UC4)
Downlink Jitter	2 ms (UC4)
Service Migration time	20 – 35 s (UC4)

Table 15. Most stringent network requirements for the railways scenario.

Network KPI Name	Target Value (Use case)
Downlink Network Latency	3 ms (UC4)
Uplink Network Latency	3 ms (UC4)
Downlink Network Reliability	99.9% (UC3, UC4)
Uplink Network Reliability	99.9% (UC3, UC4)
Uplink packet Data-Rate	523 Mbps (UC4)
Downlink packet Data-Rate	523 Mbps (UC4)
Mobility interruption time	30 ms (UC4)
Uplink Jitter	2 ms (UC4)
Downlink Jitter	2 ms (UC4)
Service Migration time	20 – 35 s (UC4)

As it can be observed in the table, the most stringent requirements for the network design are:

- Latencies in UC1 and UC4 that are between 3 and 35 ms, which should be achievable by 5G networks that have the target of latency in the order of milliseconds.
- Data rate in UC4 that is higher than 523 Mbps. It should be noted that this is done in order to test the capability of 5G networks. This is the required data rate for the service to work using 5 users. In case this is not achievable, the number of users can be reduced.
- Minimum mobility interruption time in UC4, which is 30 ms. This is one of the challenges that 5GMED is working to overcome by testing and proposing new architectures and procedures to reduce the interruption time when roaming across the border.

3.5GMED Cross-border Corridor Analysis

This section presents an analysis of the 5GMED cross-border corridor and highlights the associated challenges for the deployment of the network and compute infrastructure.

3.1. Geographical features of the corridor

The 5GMED project will showcase the deployment of the above explained use cases in the "Figueres-Perpignan" cross-border corridor. We identify two main sections of the cross-border corridor: France and Spain. Each section presents different areas with very specific challenges for each use case. Figure 15 depicts the geographic areas within the 5GMED cross-border corridor.

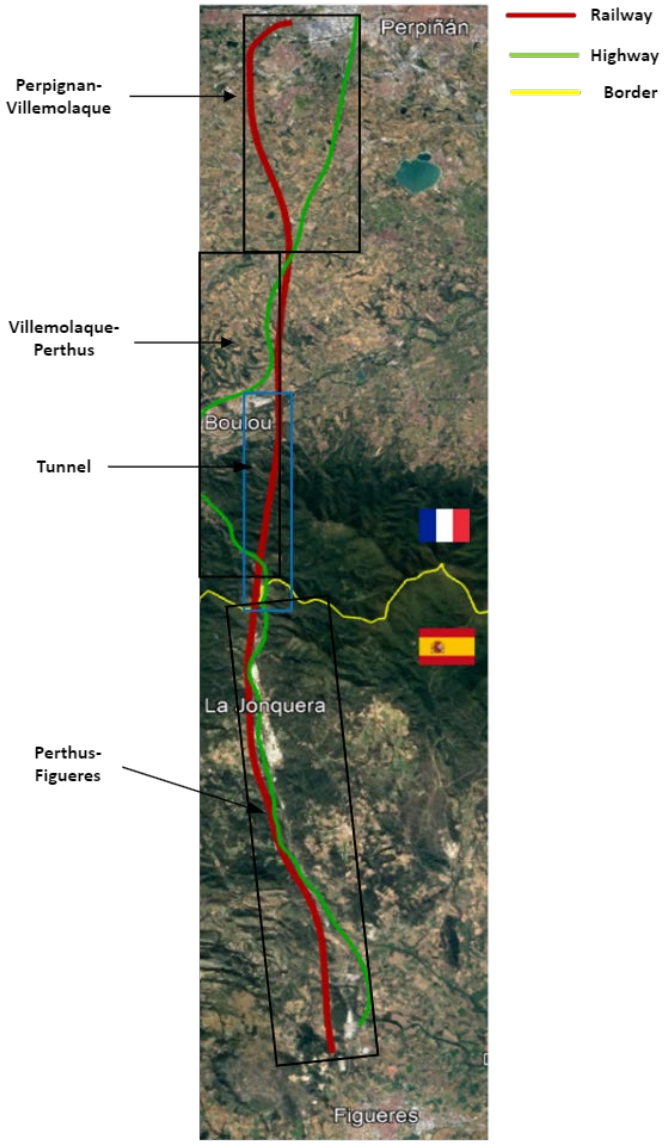

Figure 15. Map showing the different areas of the cross-border corridor between Figueres and Perpignan.

In the following subsections, each area is described starting from the southernmost point in Figueres up to northernmost point in Perpignan.

From Figueres to Perthus

The first segment of the corridor has the advantage of having the high-speed train infrastructure physically close to the highway. This will help to provide high-quality coverage for all use cases using the same infrastructure. Nevertheless, the orography of this stretch shows that the propagation of the 5G radio signal may be severely impaired by hills surrounding highway and railway. In the next section, more details on the 5G radio planning will be provided. All the UCs will be validated in this first segment.

Tunnel

The railway cross-border segment corresponds to the Perthus tunnel that spreads over 8 kilometres between the two countries. Currently, there is no indoor connectivity inside the tunnel and only mission-critical services, such as GSM-R and TETRA, are guaranteed inside the tunnel. A dedicated connectivity solution will be deployed by 5GMED based on 5G technology. Only UC3 and UC4 will be validated in this area because the other two use cases are automotive use cases.

From Perthus to Villemolaque

In this segment, the highway and the railway draw away from each other because the railways are mostly in the tunnel. Therefore, a single infrastructure will not be sufficient to provide a reliable connectivity service for all the UCs. Specifically, automotive users may connect either to the Spanish 5G RAN or the French 5G RAN depending on their position and the coverage in each country. The automotive use cases will be evaluated in this segment, namely UC1, UC2 and UC4.

From Villemolaque to Perpignan

In the 9-Km French stretch of the corridor between Villemolaque and Perpignan, the railway and highway branch off, which will most probably require specific coverage for train track and for the highway. Further details on the potential 5G cells site selection will be given in the following section. All use cases can be validated in this segment.

3.2. Need for a heterogeneous radio access network

After describing the segments of the 5GMED cross-border corridor in the previous section, this section describes the need for incorporating multiple radio access network solutions in 5GMED.

In Spain, between the city of Figueres and the border, four Vodafone gNBs will be deployed to offer 5G coverage along a stretch of 20 Km. Due to the need for reusing infrastructure already in place, these four gNBs will be installed in existing Vodafone sites close to the AP-7 highway. The identifier and geographical location of these gNBs are listed in the table below.

Table 16. Identifiers and geographical locations of Vodafone gNBs in the Spanish side.

gNodeB ID	Name	WGS84 Latitude	WGS84 Longitude
1	GI_PONT_DE_MOLINS	42.31481695	2.928325103
2	GI_CAPMANY	42.35995598	2.902749906
3	GI_JUNQUERA	42.41741102	2.869317108
4	GI_PERTHUS	42.46411114	2.866625101

The chosen band will be the N78 portion assigned to Vodafone by the Spanish spectrum regulator, i.e., 3710-3800 MHz. A coverage simulation performed with a planning software tool (Atoll [7]) shows that there will be two coverage gaps between the deployed gNBs, depicted in Figure 16, because of the area orography and signal propagation characteristics in the frequency band. The simulation in Figure 16 is based on the re-use of four existing 4G Vodafone sites on which gNBs will be installed.

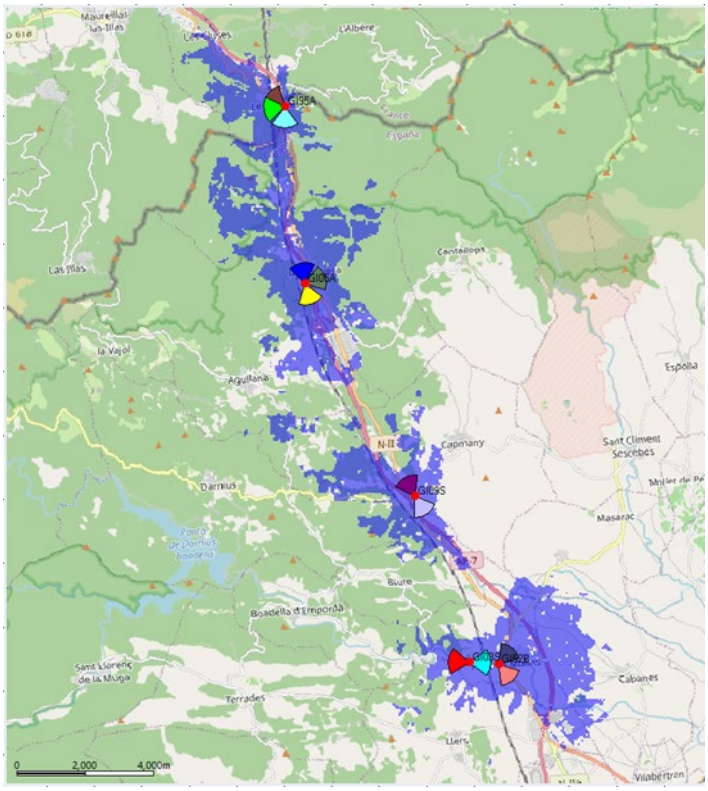

Figure 16. Coverage simulation for 5G New Radio (NR) at 3.5 GHz (band N78) in the Spanish side (using four gNBs that will be installed in existing Vodafone sites). The blue colour denotes where the estimated received power is higher or equal to a given threshold (-115 dBm).

To fill the two 5G coverage gaps in the Spanish side depicted in Figure 16, 5GMED will explore the following alternatives:

1. **Deploying additional cells connected to the same 5G Core (5GC) as the Vodafone cells**: The additional cells, which will be provided by a vendor that could be different from the Vodafone gNB's vendor, can operate at the following bands (We cannot use the N78 band as it is for commercial use):
    a. N77-upper (3.8GHz – 4.2GHz), which relies on a good support in terms of available Customer Premise Equipment (CPE)s. However, it requires requesting spectrum for R&D purposes from the Spanish regulator. Given the experience of 5GMED consortium partners on similar requests, we anticipate that obtaining a license for a 40 MHz carrier should be possible, which should be enough for the 5GMED use cases.
    b. N257 (26.5GHz – 29.5GHz) or N258 (24.25–Hz - 27.5GHz) bands. These bands would provide much higher capacity, but the limited availability of commercial CPEs is a risk for the project. Moreover, the lower RAN performance in these bands at high-mobility scenarios might represent another potential risk.
2. **Deploying additional cells connected to a dedicated 5G core**: If interoperability problems arise when the Vodafone gNBs and the N77/N257/N258 cells are connected to the same 5G Core, a separate 5G core network could be devoted to the N77/N257/N258 cells. In this case though, the two 5G networks would be independent IP networks from a mobility perspective and we would lose the ability to perform smooth handover between the N78 and the N77/N258 cells leading to a roaming situation.
3. **Deploying 5.9 GHz C-V2X technology for UC2**: In the case of UC2, where communication is based on ETSI ITS-G5 messaging [8], both the Uu interface (vehicle to network) or the PC5 interface (vehicle to infrastructure) can be used. Hence, only for UC2, the coverage gaps can be filled by deploying dedicated PC5 cells. In order to allow the vehicles to have seamless connection to the two networks, they will be equipped with a Telematics Control Unit (TCU) with both Uu and PC5 interfaces.

On the French side, another 5G RAN will be deployed to provide continuous 5G coverage across the corridor. The 5GMED consortium will be supported by a French MNO that will provide the 5G spectrum necessary to meet the requirements of the use cases as well as part of the infrastructure, e.g., cell sites, transport networks, etc. Figure 17 shows a potential radio coverage generated with the Atoll tool for the French stretch of the highway (green line) and the railway (red line), except for the Perthus tunnel. Six sites have been initially considered to ensure 5G coverage, although further assessment will be required to verify site availability and the necessary logistics, including the power grid, transport network, and related permissions.

Aside from 5G connectivity, the 5GMED project will rely on other ad-hoc solutions for railway and highway communications. In particular, the following additional radio access technologies are required to ensure the right degree of coverage for the railway use cases. The idea is to experiment other technologies that could be complementary to 5G in the scope of a large-scale corridor coverage over Europe.

- **IEEE 802.11ad at 70 GHz train-to-ground**: A dedicated high-capacity train-to-ground radio communication infrastructure will be set up between the LFP campus in Llers and the south entrance of the Perthus tunnel, where the 5G RAN infrastructure cannot provide simultaneous coverage to the highway and the railway.
- **In-tunnel connectivity**: The best option to cover this section is to deploy additional N78 cells and use a Distributed Antenna System (DAS).

- **Satellite connectivity**: This is required to provide ubiquitous coverage to the train along the corridor except for the tunnel segment, and to serve those services of UC3 which are delay tolerant.

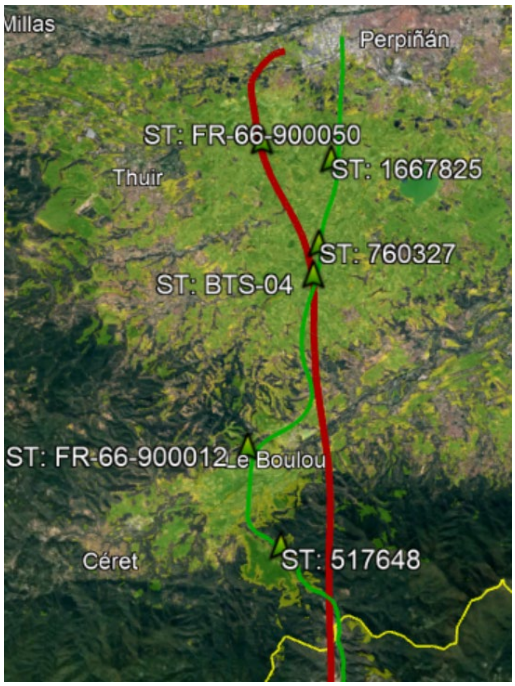

Figure 17. Coverage simulation for 5G New Radio (NR) at 3.5 GHz (band N78) in the French side (using six gNBs). The green colour denotes where the estimated received power is higher or equal to a given threshold (-93dBm).

3.3. Availability of computing resources in the cross-border section

In addition to the wireless connectivity options described in the previous section, there are few locations available along the cross-border corridor where 5GMED can deploy computing resources. The potential locations of the MEC sites are highlighted in the figure below.

All the MEC sites are located within the LFP premises due to the large availability of resources, including fiber connectivity, energy sources, and space for equipment. Moreover, the Perthus tunnel features four service rooms that can potentially host MEC nodes.

- At the French side, two sites have been identified, namely the CTE1 (Caseta Tecnica Este 1) in the outskirts of Perpignan, and CTN (Caseta Tecnica Norte) right outside the north entrance of the Perthus tunnel.
- At the Spanish side, a potential location is the CTS (Caseta Tecnica Sur) at the south entrance of the Perthus tunnel, and the PCC (Puesto de Control Central) that corresponds to a building of the LFP campus in Llers. The PCC will host an interconnection point with the radio connectivity solutions, i.e., 70 GHz ground-to-train, 5G and satellite technologies.

As mentioned earlier, all these MEC sites can rely on a continuous supply of electricity and are connected through a large amount of dark fiber that is already deployed. This is a key benefit since the time-to-deployment is lower than designing and building an optical backbone network from scratch. Furthermore, optical fiber ensures higher reliability and lower latency as compared to a microwave-based backhaul system.

Figure 18. MEC sites distribution over the cross-border corridor.

3.4. Required network exchange points

As 5GMED infrastructure is composed of multiple radio access networks from different stakeholders, it is necessary to define several exchange points. The following exchange points depicted in Figure 19 have been identified:

1. **VDF-CLNX_ES**: Required to connect the four Vodafone gNBs to the Cellnex transport network that will provide access to the 5G Core on the Spanish side. The VDF 5G traffic will be delivered in a single aggregation point due to existing architecture limitation and security reasons. MEC computing resources will not be collocated with the gNBs due to the small size of the sites but will rather be placed in the traffic aggregation point.
2. **CLNX_ES-LFP**: Required to connect the 70 GHz railways infrastructure and the MEC nodes deployed at the LFP premises with the Cellnex network on the Spanish side, so that railways services can reach the application functions hosted in the cloud.
3. **HSP-CLNX_ES**: Required to deliver satellite traffic from the Hispasat network to the Spanish 5G Core network deployed by Cellnex.
4. **CLNX_FR-LFP**: Required to connect the French 5G RAN with the LFP network where MEC resources are located.
5. **LFP_Cross-Border** Required to enable roaming between Spanish and French 5GC as well as to provide connectivity between the different MEC resources.

The detailed design of this interconnection networks will be reported in D3.2.

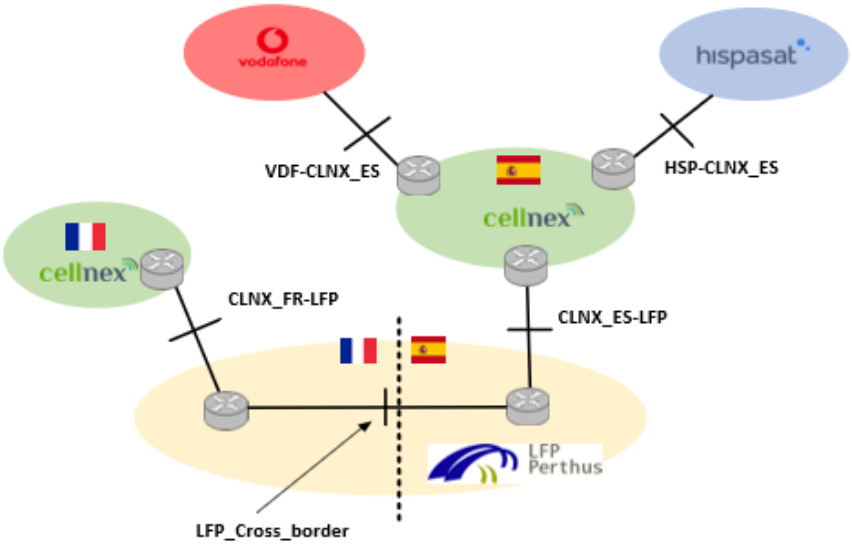

Figure 19. Interconnection networks.

4. Technical Challenges and State-of-the-Art Solutions

The aim of 5GMED is to develop a flexible architecture that meets the strict requirements of CCAM and FRMCS services, especially in cross-border scenarios. In previous Section 2, we have analysed the requirements of the use cases chosen by 5GMED from the different CCAM and FRMCS services. This analysis led to the conclusion that providing continuous and seamless service experience will require an architecture that can support challenging levels of end-to-end latency, data rate, reliability, and interruption time. These service requirements are especially challenging due to the high mobility of CCAM and FRMCS services. They become even more challenging in cross-border scenarios where two 5G networks will be involved. Therefore, this section is dedicated to the technical challenges related to the implementation of the 5GMED four use cases and their requirements. In addition, we provide some guidelines to the possible solutions that will be investigated by the 5GMED project based on the technology enablers identified in D2.1 [1].

Firstly, we identify and explain these technical challenges and then we provide an analysis of possible solutions for each challenge in the rest of the subsections. The final architectural choices that 5GMED will make to address these challenges will be reported in D3.2 [9].

4.1. 5GMED Cross-border technical challenges

Based on the previous discussion in Section 2 and Section 3, we identify the following specific cross-border technical challenges that are further described in next subsections:

- Challenge #1: Multi-Connectivity and integration of heterogeneous networks
- Challenge #2: Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core
- Challenge #3: 5G roaming at cross-border with low latency and low interruption time
- Challenge #4: MEC deployment and Inter-MEC connectivity

4.1.1. Challenge #1: Multi-Connectivity and integration of heterogeneous networks

As discussed in the previous section, the environment where the 5GMED network will be deployed includes areas with irregular orography, rural areas with dense vegetation, and even a long tunnel. Such diverse environment is very challenging for radio communications and requires the usage of a heterogenous network where different technologies can be appropriate for different environmental challenges. This problem is not particular for 5GMED project and would appear in many cross-border networks. Therefore, the performance analysis of the different technologies will help in upscaling 5GMED solution to whole Europe.

The first challenge in deploying such heterogeneous network is to guarantee that the used technologies should consider the requirement of the use cases described in Section 2.2 and Section 2.3 especially in terms of data rate, latency, and reliability. In addition, an efficient inter-RAT handover solution to move from one access technology to another should be develop and respect the requirement on mobility interruption time of each use case.

The problem of integrating heterogeneous networks can be split into two technical challenges:

1. Provide a transparent access to the application functions, regardless of the network being used. For example, if the V2X application server of UC2 is deployed on a dedicated Data Network Name (DNN), this V2X application server should be reachable both if the vehicles use a Uu interface or a PC5 interface.
2. Provide a solution to provide the service inside the train regardless of the actual access network(s) being used to provide the train-to-ground connectivity. A multi-connectivity gateway will be used allowing to:
    - Define service-level policies defining how access networks should be used for each service.
    - Aggregate capacity available in multiple access networks enables simultaneous access through the different access networks.
    - Divert traffic to a backup access network when the main one is not available.

It should be noted that mobility interruption time (i.e., the service interruption duration due to switching from one technology to another) should be the key factor when choosing the best solution for the heterogeneous network design. In addition, all involved technologies should be able to support the requirements of the services in terms of throughput, reliability, and latency.

4.1.2. Challenge #2: Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core

The diversified requirements of the use cases presented in Section 2, which are potential services to be provided in vehicular and railway networks, can be better supported and more scalable if isolation mechanisms such as slicing are implemented in the network. This will provide a more dedicated and isolated set of resources that can be optimized for each use case. In addition, the existence of different and isolated sub-networks, and the high dynamics in the resources required by each network require a very flexible architecture that allows resource allocation at every level of the network. This flexibility is provided by virtualization and network automation, which are building blocks of the 5G Standalone Core (5G SA). Therefore, the solution to be implemented by the 5GMED project should be based on a 5G SA core implementation supporting the following capabilities:

- **Virtualization**: The 5GMED elements shall be deployable as network functions, either in cloud native (container) or Virtual Machine (VM)-based approach. As these resources need to be dynamically managed in a coherent way across the network, there is a need for an orchestrator that can coordinate the lifecycle of these resources. A key point in 5GMED is to enable cross-border orchestration that provides a seamless service experience for all use cases.
- **Network slicing support**: The 5GMED network shall support network slicing capabilities to support simultaneously the diversified use cases in terms of QoS requirements. A key point in 5GMED is to study how slice management should be implemented in cross-border scenarios.
- **Network automation and observability**: The 5GMED architecture should support Application Programming Interface (API)-driven configuration to automate the provisioning and configuration of the complex infrastructure used in 5GMED. In addition, it should provide observability, e.g., with metrics at the DNN and UE levels, which are required to optimize the network support for the use cases. In a similar way to orchestration, the network automation and observability should take into account the cross-border and cross-network scenarios.

4.1.3. Challenge #3: 5G roaming at cross-border with low latency and interruption time

All 5GMED use cases require a low mobility interruption time when roaming between the Spanish and the French 5G networks, as was described in Section 2. For mobile networks up to 5G NSA, roaming interconnection can span tens of seconds because UEs stick to their home network until the radio link fails completely, and after the failure it needs to start scanning until a cell to camp on is found. In addition, there might be additional delays due to connection establishment in the core network. This interruption duration can be even larger in railway scenarios where trains are moving at very high speed and carrying a big number of connected equipment that shall perform roaming quasi simultaneously when crossing the borders. Therefore, 5GMED needs to investigate mechanisms to reduce these roaming interconnection times.

In addition, most of today's implementations of roaming are based on the so-called Home Routed (HR) roaming, where user traffic is always forwarded to its home network even when it is connected to another network, which will induce high latency. This problem will also be considered in 5GMED to reduce this latency.

4.1.4. Challenge #4: MEC deployment and Inter-MEC connectivity

Many services in the cross-border scenario require very low latency as it was described in Section 2 (e.g., UC1 and UC4). Fortunately, 5G networks were designed with many features and tools to support such requirements. One of these features is edge computing and in particular Mobile Edge Computing (MEC). The deployment of MEC for services such as CAM and FRMCS is accompanied by challenges, such as the migration of the resources from one MEC to another to follow the user; this might lead to a change in the IP address of the application server and the UE and may require a restart of user session.

Another key challenge in 5GMED MECs is to provide a consistent Application/MEC-Network interface across different MEC domains. Note that 5GMED application functions in several use cases need to be instantiated both in the Spanish and French MEC sites. It is not realistic to assume in this scenario that vertical service providers will customize their services to the MEC capabilities of the operators in each country. Instead, a standardized MEC interface should be offered in all MEC sites. Therefore, the 5GMED network architecture shall account inter-MEC connectivity even between different PLMNs when roaming.

4.2. Solutions to Challenge #1: Multi-Connectivity and integration of heterogeneous access networks

To overcome the first 5GMED challenge explained in Section 4.1.1, we present some standardized solutions. In particular, we present in Section 4.2.1 the Non-3GPP Inter-Working Function (N3IWF) function mechanisms and in section 4.2.2 the Access Traffic Steering/Switching/Splitting (ATSSS) function defined by 3GPP to connect heterogeneous access network and enable multi-connectivity. In Section 4.2.3, we discuss the usability of these two standards, and we explain the reason behind the selection of the ACS-GW as a solution for the multi-connectivity problem in the 5GMED project.

4.2.1. Non-3GPP Inter-Working Function (N3IWF)

3GPP introduced the N3IWF function in release 15 to support non-3GPP access to the 3GPP core networks. The N3IWF is responsible for adapting access and authentication protocols of the non-3GPP untrusted access towards the 5GC. The term "untrusted" refers to non-3GPP access networks not controlled by the 5G operator, e.g., an airport Wi-Fi hotspot, domestic network, 70 GHz IEEE802.11ad. As a result, UE and N3IWF shall set up an end-to-end security association, regardless of the security measures established at layer 2, i.e., Wi-Fi Protected Access 2 (WPA2). The resulting architecture is shown in Figure 20. As can be seen from the figure, a UE can simultaneously have a connection using NR-Uu interface or utilise Wi-Fi as a unique access method through NWu interface. In both cases the UE should provide the 3GPP credentials to grant access, i.e., Universal Subscriber Identity Module (USIM) / Embedded Universal Integrated Circuit Card (eUICC).

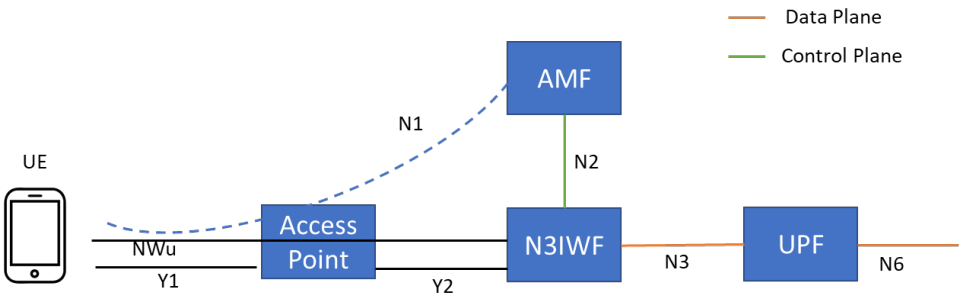

Figure 20. Integrating non-3GPP technologies in the 5G core network via N3IWF.

The N3IWF is connected to the AMF through the N2 interface. It also enables the N1 interface allowing the UE to communicate with AMF over a NAS interface. In addition, the N3 interface is used by the N3IWF to interact with User Plane Function (UPF). The N3IWF is responsible for establishing a secure connection over NWU interface for both user plane and control plane traffic. Internet Protocol Security (IPSec) and Internet Key Exchange (IKE) are to establish a secure tunnel to carry NAS messages and packets for the DN.

Table 17 summarizes the main features of the N3IWF for both control and user planes.

Table 17. List of N3IWF features.

Control Plane	Support to establish IPsec tunnels with the UE over NWu interface and employing IKEv2/IPsec protocols
	Set-up of signalling IPsec security association for protecting NAS messages
	Set-up of IPsec security association for protecting PDU session traffic
	N2 interface termination using NG Application Protocol (NGAP) and Stream Control Transmission Protocol (SCTP) protocols towards AMF
	Forwarding uplink and downlink control plane NAS (N1) signalling messages between the UE and AMF
	Forwarding NAS messages to authenticate, register and grant UE access to the 5GCN
	Forwarding NAS messages to set up PDU sessions

	Handling N2 signalization from SMF (Session Management Function) forwarded by AMF, associated with PDU sessions and QoS
	Selection of the AMF
User plane	N3 interface termination using GPRS Tunnelling Protocol User plane (GTPU) protocol towards UPF
	Forwarding uplink and downlink user plane packets between the UE and UPF
	Decapsulation/encapsulation of packets for IPsec and GTPU tunnelling
	N3 user plane packet marking in the uplink
	Enforcing QoS associated with N3 packet marking

4.2.2. Access Traffic Steering/Switching/Splitting (ATSSS)

The ATSSS function was introduced in 3GPP release 16 to allow the joint use of the 3GPP and non-3GPP access. It has three functionalities: steering, switching, and splitting. Steering is used to route user-plane traffic according to the service and select the best link to use. Switching enables the possibility of initiating an Inter-RAT handover without service interruption. Splitting enables the joint use of more than one link.

ATSSS can be enabled in the untrusted access over the N3IWF, and some of the trusted configurations of the non-3GPP access, e.g., Trusted Non-3GPP Gateway Function (TNGF) and Wireline Access Gateway Function (W-AGF). Figure 21 shows a high-level user-plane architecture of a UE equipped with Wi-Fi and 5G NR, accessing the same 5G core network by means of the ATSSS.

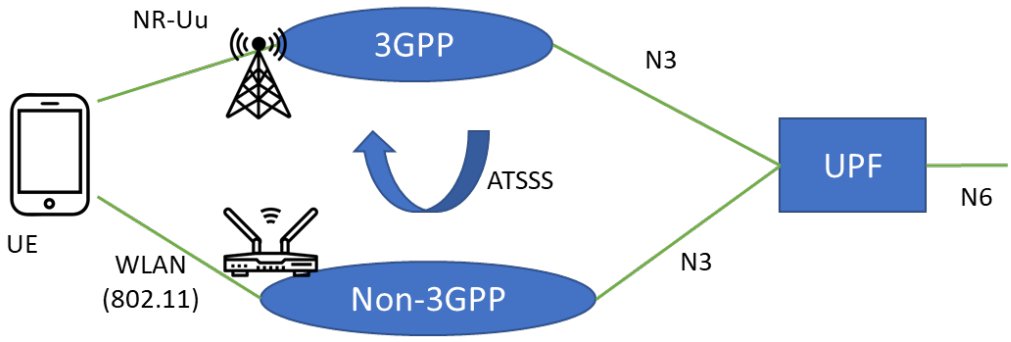

Figure 21. ATSSS simplified architecture for user plane.

The ATSSS support can be enabled by ad-hoc user-plane functions running in the UE and the UPF, in addition to new control-plane policy rules for enforcing traffic management on the two links by building a multi-access PDU layer (N3), as illustrated in Figure 22.

For each Service Data Flow (SDF) type, the ATSSS-enabled PCF creates Policy and Charging Configuration rules that are converted into N4 rules for the UPF by the ATSSS-enabled SMF. Similarly, the ATSSS-enabled AMF translates them into N1 rules for the UE. Note that the ATSSS functions of the UPF manage the downlink traffic through the inter-working between Multipath TCP (MPTCP) at higher layer and/or lower layer multi-link protocols (called ATSSS-LL).

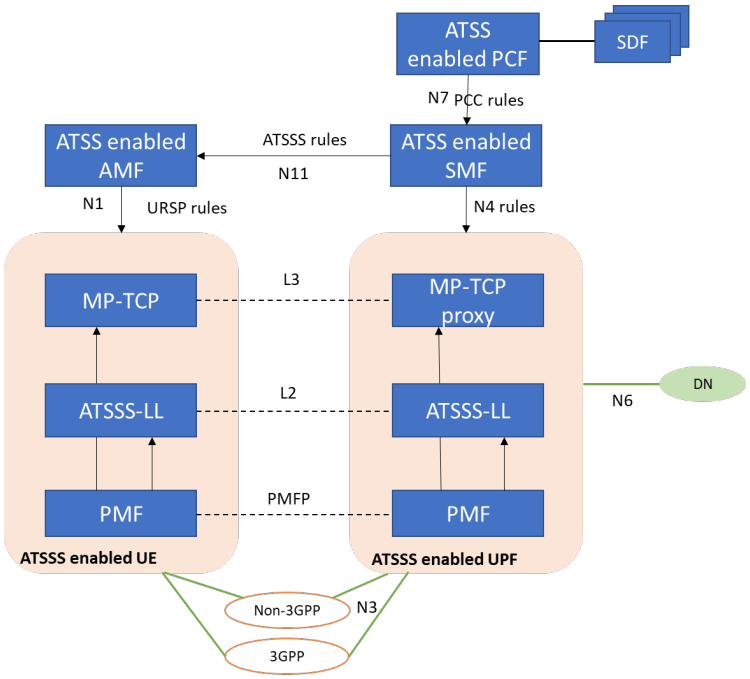

Figure 22. ATSSS simplified control and data plane between UE and UPF.

The core network is informed through a multi-access PDU request over the uplink NAS layer that a new multi-access PDU session is requested and that ATSSS-LL functionality and/or MPTCP are required to set up the MA PDU session, which will be regulated as follows:

- Indicating if MP-TCP or ATSSS-LL will be used for each matching SDF
- Indicating one of the following steering modes:
  - **Active-Standby**, in which a link will be selected as master and the second as a slave. The former is the one that will handle the traffic until it becomes unavailable. In the case of Guaranteed Bit Rate (GBR) traffic sessions, it is the only mode available.
  - **Smallest-delay**, in which the link with the lowest Round Trip Time (RTT) will be selected during the session establishment. If the link disappears, its traffic will be handled by the backup link.
  - **Load-balancing**, in which the traffic will be distributed between the links according to a certain load ratio. If one of the links fails, all the traffic will be handled by the other link.
  - **Priority-based**, in which a link associated with higher priority can handle the traffic until congestion appears.

For the correct operation of MPTCP, a session layer between both endpoints is needed, otherwise, a regular TCP connection on a single link will be set up by the sender. An application talks to the transport layer over socket APIs, while MPTCP interacts as a middle layer, establishing multiple TCP connections with the endpoint.

Furthermore, two strategies can be applied to configure MPTCP: a) channel bonding, employing multiple links, thus increasing the data rate, b) link failover, ensuring service continuity in case of failure. A scheduler then regulates the outgoing traffic on the links according to a policy, a congestion control mechanism or an ad-hoc multipath scheme. Moreover, traffic coming from multiple TCP flows is reordered to maintain the TCP byte stream order required by the application.

ATSSS-LL defines a layer-2 multi-access mechanism at the user plane of the 5GC, between UE(s) and UPF(s). In the case of Ethernet PDU, this is the only mechanism available in ATSSS, whereas if PDU is IPv4/v6, ATSSS-LL complements MPTCP by managing the traffic, which is not making use of MPTCP. Note that the UEs and UPFs will need to include the Performance Measurement Function (PMF) component for the ATSSS. This entity measures values of interest by using the UDP-based PMF Protocol (PMFP), whose packets are carried over the same PDU layer used by the user-plane data. As a result, PMF can infer RTT and load values, which will capture the same traffic conditions of the whole data path between UE and UPF.

4.2.3. Adaptive Communication System-Gateway (ACS-GW)

The above standard solutions require that the non-trusted systems and the equipment connected to them support the standardized interfaces, which is not the case in what exists in the market, especially in terms of existing UEs. Even if during the project life such equipment will become available, the objective of 5GMED consortium is to provide a solution that is transparent to the user; therefore, any user that enters the train should be able to get access whatever the UE he/she uses. Therefore, it was decided to develop an in-house solution that provides seamless multi-connectivity while hiding the complexity of the network from the user, and we call it Adaptive Communication System-Gateway (ACS-GW). As it is a solution fully developed by the 5GMED consortium, it will provide high flexibility to integrate any function needed in the project and fulfil the requirement of the different use cases, especially in terms of mobility interruption time.

4.3. Solutions to Challenge #2: Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core

In order to overcome the second technical challenge explained in Section 4.1.2, a relevant 5G Core implementation that will be considered in 5GMED is the Raemis product from Druid [17]. The Raemis platform is a 3GPP compliant 5G Core implementation that supports virtualisation, network slicing, and automation/observability capabilities required by 5GMED. This section introduces the concepts of virtualization and containers and presents the main features of the Raemis 5G core, whereas the intended use of these features in the context of 5GMED will be described in D3.2 [9]. In addition, we explain the possible paths investigated by the 5GMED consortium to bring automation and network slicing to the cross-border 5GMED architectural design.

4.3.1. Virtual Machines and Containerization

Virtualization describes the separation of the resources of a service from its underlying hardware infrastructure. Virtualization techniques may be applied to different layers, such as the network, storage, operating systems, and applications.

The virtual infrastructure provides a layer in which a virtual abstraction between the compute, storage, and network is created, making a single hardware virtually act as multiple machines running independently from each other. Through virtualization, the administrators can allocate dynamically pooled resources across different applications, utilizing the infrastructure efficiently.

Before the birth of virtualization, the software was tightly coupled to corresponding hardware, so having several applications running on the same infrastructure may introduce issues and problems [10]. Another disadvantage of this setup was the inefficient way of using the resources, in a way that additional hardware deployment can be necessary to support requirements during usage spikes. Thereby, such legacy setup would lead to costly deployment and a larger footprint that dynamically cannot allocate resources based on user demand and business needs.

With the emergence of both virtualization and containerization, some of the limitations mentioned above are addressed, allowing several applications, independent from each other, to be deployed on the same hardware. In virtualization, this is accomplished by using a hypervisor that is responsible for logically assigning and separating physical resources. The hypervisor resides between the physical infrastructure and the Operating System (OS). It allows the deployment of a guest operating system that is running on a virtual machine (VM). The guest OS is independent and unaware of the other guest OS running on the same physical hardware; in fact, all these OS are protected and not in any way affected by any issues of the others [11].

Recently, application containers, leveraging containerization, have started gaining the attention of software and application developers, as part of the application modernization. In fact, this technique has somehow become an alternative to virtual machines. In a nutshell, containers are software packages that are necessary to define a specific application. These containers are more lightweight than VMs. Containers only include the dependencies to package their applications, whereas VMs contains the complete set of OS, the hypervisor, apart from the actual applications. Lastly, the containers incur less overhead on the applications as they do not have a virtualization layer or hypervisor. Having said this, the containers are proven to provide more flexibility and versatility in improving the utilization of resources [12].

Although people normally coin containers as “docker containers,” Docker is only one of the platforms that enable developers to package their software in the form of containers. Some of the container runtimes available apart from Docker are containerd, CRI-O, and Mirantis Container Runtime, all of which Kubernetes supports.

As mentioned above, we will use in 5GMED docker containers that will enable us to isolate the different use case applications and provide the needed resources to fulfil the requirements especially in terms of latency and reliability.

4.3.2. Service Orchestration and Automation Platforms

To fully take advantage of the power of edge computing and network flexibility, different frameworks such as virtualization and orchestration must be utilized and well-adapted to the nature of the Edge environment. Edge Computing may be seen as a technology that combines the capabilities of networks, infrastructure, and applications to provide an enhanced experience to end users. It addresses the different requirements of new use cases, such as extremely low latency together with high throughput.

To complement the virtualization technologies mentioned above, orchestration tools are leveraged to provide resources, networking capabilities, and applications to end-users. Container orchestration is defined as the process of managing the lifecycle of the containers, while service orchestration is the process of delivering services end-to-end by stitching all the components in an automated manner to create a service.

Docker-swarm and Kubernetes [13] are the most popular choices for container orchestration. Kubernetes, commonly known as K8S, is an open-source project widely known and used for running and managing containers. This container orchestration platform allows all the pieces to operate in a synchronized way to maintain the defined and desired state of the containerized workloads in an automated way. Most of the container management platforms (e.g., Google Container Engine (GKE), OpenShift, Tanzu, and Rancher) have Kubernetes distributions that can also manage K8S clusters.

However, it is not only the infrastructure that needs to be managed but also the network functions and applications running on top of it. According to Gartner [15], there are 6 capabilities of Orchestration and Automation.

1. Workflow Orchestration – creation and management of workflows across applications running on-premises and in the cloud.
2. Event-driven Automation – simplified processes that normally involve manual intervention and scripting.
3. Self-service Automation – orchestration enablement to users and developers.
4. Scheduling, Monitoring, Visibility and Alerting – real-time monitoring that improves operational processes and Service-Level Agreements (SLA)s.
5. Resource Provisioning – allocation and assignment of resources both on-premises and in the cloud.
6. Managing Data Pipelines – automation of ingestion and processing of data.

In the context of 5GMED, the overall service orchestration tool that will be used is the NearbyOne solution by Nearby Computing. NearbyOne is a cloud-native solution, completely implemented as a set of micro-services, packaged as Docker containers that can be deployed either using Kubernetes manifests or pushed to the service provider premises using a cloud-based tool integrated into its Continuous Integration and Continuous Delivery (CI/CD) platform. The description of this solution will be presented in details in D3.2 [14].

The orchestrator will manage the entire edge ecosystem, i.e., MEC Infrastructure, some Network elements, and Applications. It will offer an end-to-end solution for the 5GMED project, not only the usual Kubernetes-based and VM-based products (e.g., RedHat OpenShift and GKE, OpenStack, VMware), that orchestrate and manage only the containers. The orchestrator will not only manage the container resources, but also the provisioning of the MEC nodes, the upgrades, taking into account that the corresponding applications are also managed by the orchestrator. This is achieved through the solution's deep application and network functions integration, to ensure QoS at the Edge.

Some of the applications and the network functions, such as ATOS and ATC applications, as well as Raemis Core, will be onboarded by the orchestrator. This onboarding procedure allows the 5G core network functions and the applications to be orchestrated and managed through NearbyOne. Orchestration will allow dynamic service migration from one MEC node to another in real time.

4.3.3. Raemis Core overview

Figure 23 depicts the elements of the Raemis 5G core. It follows the Control and User Plane Separation (CUPS) architectural model, and hence the Raemis data plane function is a common network element shared across all generations of cellular technology. It provides the User Plane Function (UPF) of the 5G SA core. It must be noticed that the 5G core based on the Druid Raemis 4G/5G solution is 3GPP

compliant. This aspect makes Druid Raemis platform a suitable solution for the Core network of 5GMED 5G system.

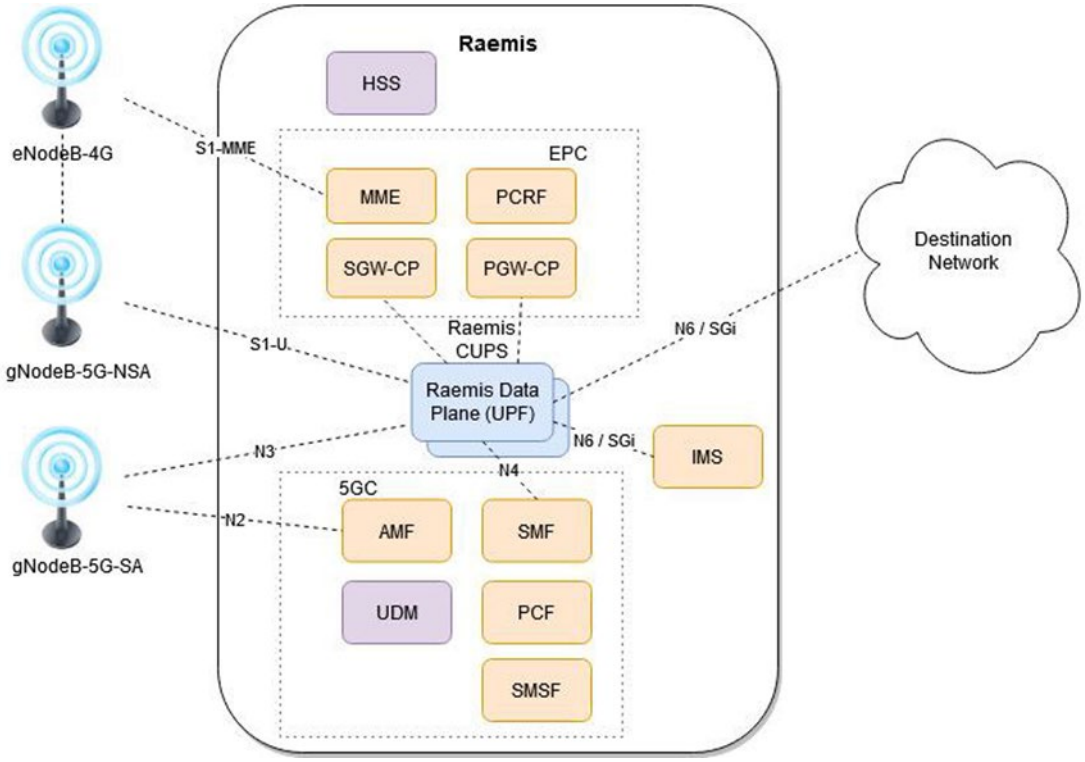

Figure 23. Elements of the Raemis Core [8].

The UPF Network Function (NF) handles all the data plane traffic, combining the functionality of both the SGW and PGW in LTE. It is responsible for performing packet routing to external Data Networks (DNs) via the N6 interface. The Raemis UPF implements the standard N3 interface with the gNB element and the standard N6 interface towards the DNs. The UPF data plane functions are controlled by the Session Management Function (SMF) element, which as a similar function in 5G to the PGW and SGW control plane in terms of 4G CUPS architecture. It is the NF responsible for the management of data sessions.

The Raemis Access and Mobility Management Function (AMF) implements the standard N1 and N2 interfaces, allowing the connection of any standards based 5G Standalone RAN and UE. Note that N1 is the standardised interface between the UE and the AMF but it is not depicted in the figure for brevity. The AMF in 5G systems is the NF responsible for the registration, authentication, and mobility services of the UE. It contacts the Unified Data Management (UDM) to provide the registration and authentication information and with the SMF to enable the set up and management of data sessions.

The Policy Control Function (PCF) element manages subscriber policy control, controlling the service level and network access available to a subscriber. It performs subscriber policy control, controlling the service level and network access available to a subscriber. The Raemis core platform can be delivered as virtual machines or as containers. The choice for 5GMED is the containerized version. It has orchestration capabilities allowing scalable and resilient virtualised infrastructure.

The Raemis 5G core network elements are designed with a built-in API. They implement the Representational state transfer (REST) API model. The API schema for the associated objects is available in the REST API specification. This will facilitate exposing the internal state of the elements to external functions in an easy way.

In addition, the REST API model will allow the integration of applications using the REST API to collect, expose, and analyse data collected from the elements. For instance, the Raemis Private Core Network (PCN) Graphical User Interface (GUI) is an easy-to-use Web-based interface enabling to monitor and control the elements of the system.

The Raemis core will be deployed as the core for the two mobile operators across the border together with interfaces, including N14 interface required for optimizing the roaming. The possibility of deploying this interface was one of the main reasons to choose the Raemis core as it will allow us to fulfil the requirement of very low mobility interruption time when crossing the borders.

4.3.4. Network slicing in the context of 5G and cross-border scenarios

Network Slicing is a key concept of 5G SA networks. Its purpose is to provide multiple logical networks, known as network slices, which can operate over a common virtualised infrastructure. These logical units are composed of several chunks of resources, such as computing, networking, transport, and radio [16]. This will allow us to isolate the different use cases and provide different QoS for each use case based on the requirement that was presented in Section 2. A fully activated slice enables end-to-end connectivity among network elements including UEs. In addition, network slicing allows the deployment of independent services with performance guarantees. This is critical in 5GMED to separate the different use cases that have different requirements in terms of resources and QoS.

In the 5GMED architecture, network slicing is applied at the different 5G subnetworks, i.e., core network, transport network, and radio access network. In addition, the resources in edge computing elements can also be shared by different slices and require slice management. Figure 24 represents the 5GMED components that provide such network slicing capabilities. 5GMED's slice management layer will provide the necessary mechanisms for the management of end-to-end Network Slice Instances (NSI) to support the various use cases. These NSI can be of different granularity, ranging from use case-based slices to service-based slices. In order to deploy these slices, the 5GMED consortium will deploy the NearbyOne orchestrator provided by Nearby Computing. This orchestrator incorporates a Network Slice Management Function (NSMF) that can communicate through Southbound Interfaces (SBIs) with (i) i2CAT's Slice Manager and RAN Controller, for the setting-up the RAN elements of the network slices, (ii) CTTC's Transport Slice Manager, for the setting-up the transport network components of the network slices, and (iii) Nearby's core slice manager, for setting-up of the core network and edge elements of the core and edge slices. Through these interfaces, the NearbyOne orchestrator will communicate QoS requirements or any other requirements to each slice manager. In particular, the Raemis platform supports this concept by design and goes a step further in bringing the slicing concept directly to Edge Private Networks. With Raemis Edge Slicing, one can define dedicated network slices that ensure a specific level of QoS and security for individual groups of users.

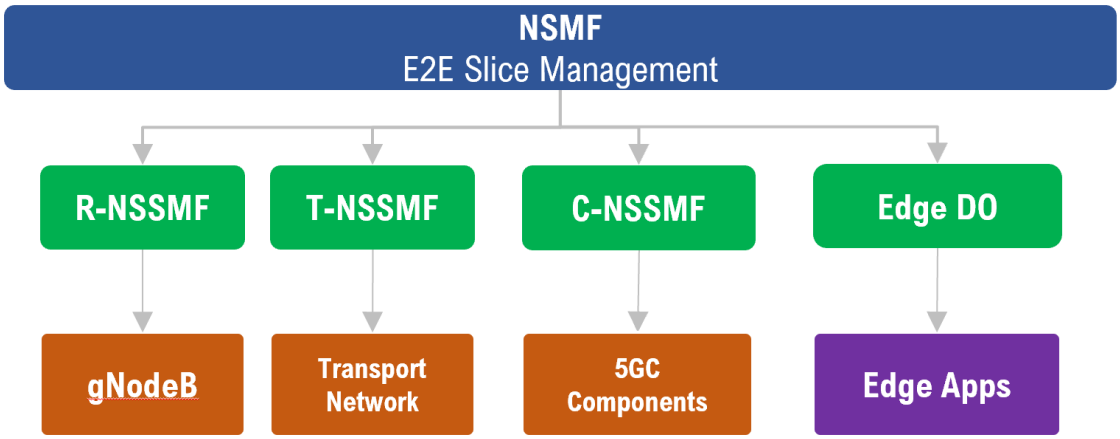

Figure 24. Network Slicing within 5GMED Architecture.

In the cross-border scenario, different operators may have different slicing policies, configurations, and availability. So, moving a slice from one PLMN to another is a challenging problem. As network slicing is an already standardised feature of 5G networks, standard slices are the same in all PLMNs and have the same Single Network Slice Selection Assistance Information (S-NSSAI). Therefore, a UE moving from one PLMN to another can attach to the same standard slice. In case a service provider wishes to have a customized slice, it should use the same S-NSSAI in all PLMNs where its service will be deployed, and this should be done by agreements between the different operators that will create these slices in a manual way.

What could be dynamically adapted in cross-border scenarios is the resources that can be assigned to a slice when UEs are crossing the border, which can be done through the orchestrators and their cross-border interfaces.

4.3.5. Network Automation and observability

Artificial intelligence will be one of the main building blocks of 5G networks. It will be used to enhance user experience and minimize network costs. One approach could be that an operator provides some AI functions as services that can be used by the applications, allowing these applications to use information collected by network elements to optimize their performance. Furthermore, by providing these AI services through standardized interfaces, some critical information can be collected by third party applications and can be used by any application using this service. For instance, information collected by a road operator (e.g., number of cars and their speed, accidents) can be used by the network to optimize the resources allocated to teleoperation application slice.

In a cross-border scenario, the AI service and its decision should be communicated to the UE even when changing the PLMN. This will pose a problem as different PLMNs might not have the same AI modules implemented and normally they do not exchange information about their networks.

In this context, the 5GMED consortium will investigate the feasibility of including an AI or data analytic layer in the 5GMED architecture, including the possibility of standardized interfaces between the AI

layers of different operators. This will convert the 5GMED architecture into an AI-enabled architecture. This layer should contain a data lake and a set of generic AI modules. The former will be responsible for collecting, storing, and exposing a set of network and service KPIs and measurements, in addition to possible third-party data. The AI modules will be generic AI modules using the information of the data lake or directly collected information from the network to optimize the resources allocated to each slice or service. In addition, they can be used to minimize the mobility interruption time and the service migration time by predicting signal quality changes on arrival to borders between countries.

4.4. Solutions to Challenge #3: 5G roaming at cross-border with low latency and interruption time

In order to overcome the third technical challenge described in Section 4.1.3, we present in this section how roaming is implemented in 3GPP 5G systems. These solutions will be used as a starting point for the 5GMED project to design an architecture that reduces mobility interruption time during international roaming. In particular, we focus on the general architecture enabling roaming from core network point of view in Section 4.4.1, and describe radio access techniques to achieve low latency in Section 4.4.2. As most of the advanced standardized techniques on the radio side explained hereafter are not mature yet in terms of equipment, the 5GMED consortium will be focusing on the optimization of the roaming from within the core network, in addition of course to the legacy optimization techniques in the RAN side. In case RAN features will be available during the execution of the project, they might be also implemented in the final solution.

4.4.1. 5G roaming architecture

The deployment of SA 5G networks is still ongoing, and the majority of commercial 5G networks are providing their services over non-Standalone networks, which are built using 4G core network. Our objective in this context is to deploy and evaluate the performance of three roaming solutions that will be explained in this section: home routed roaming, local breakout roaming, and UE roaming with AMF relocation and RAN assistance. Therefore, the 5GMED project will provide one of the first results on roaming using 5G SA networks especially in terms of mobility interruption time that was one of the main KPI for all use cases.

4.4.1.1. 3GPP architectures

Two legacy roaming architecture were defined in 4G core networks and are still in use in 5G networks: Home Routed (HR) roaming and Local Breakout (LBO) roaming. The reference architectures of 5G HR and LBO roaming are depicted in Figure 25 and Figure 26, respectively.

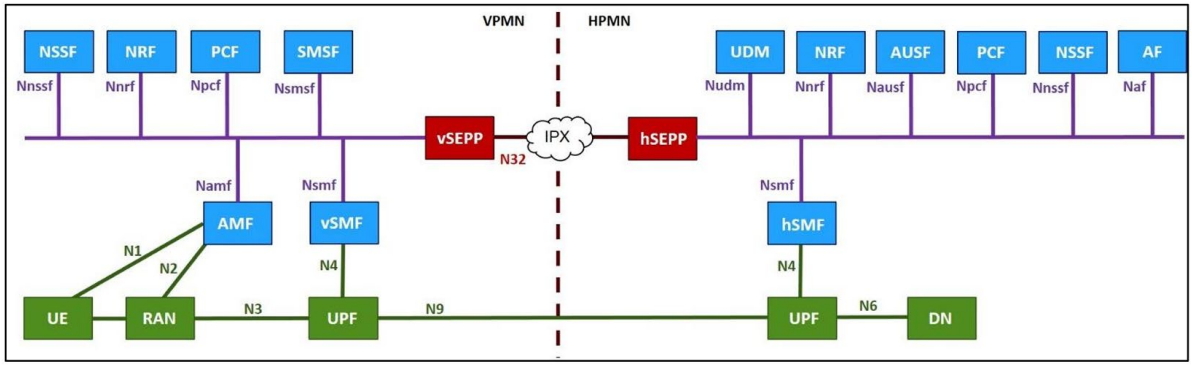

Figure 25. 5G System Roaming architecture – Service Based Interface Representation (HR) [18]

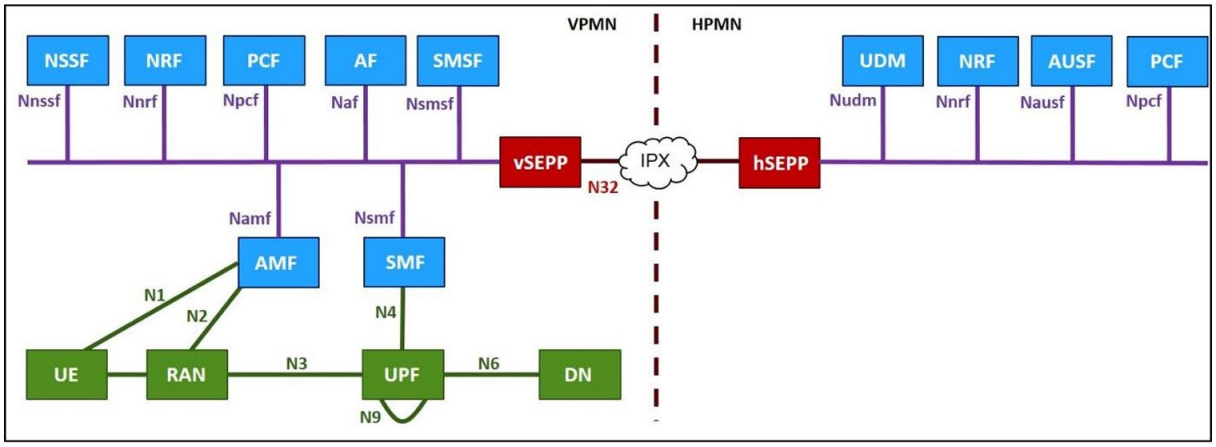

Figure 26. 5G System Roaming architecture – Service Based Interface Representation (LBO) [18]

It should be noted that in the two architectures, the visited and home core networks are connected through the N32 interface created between two Security Edge Protection Proxy (SEPP) to allow secure communication.

The main difference between LBO and HR roaming is the involvement of the UPF and SMF of the hPLMN in the HR roaming, whereas these two NFs will not be connected in the LBO roaming. This difference impacts the quality of the connection in different ways; in the HR roaming, an additional delay will be added in the user plane as the data flow should go through the UPF in the hPLMN after it is forwarded by the UPF in the vPLMN. In the LBO roaming, the UE must establish a new Protocol Data Unit (PDU) context, which might lead to significant interruption time that is needed to establish this new PDU context. Some techniques were proposed to reduce the delay and interruption time as it will be explained in Section 4.4.2. It should be noted that in the case of HR roaming the visited UPF and home UPF are connected through the N9 interface.

4.4.1.2. Interface procedure implementation

During the establishment of a PDU session in HR roaming, the following actions are performed [19]:

- The AMF selects the SMFs in the vPLMN and the hPLMN.
- The AMF provides the identifier of the selected SMF in the home network to the selected SMF in the visited network.

- The Non-Access Stratum (NAS) Session Management (SM) procedure terminates in the SMF in the vPLMN, and the related information will be forwarded to the SMF in the hPLMN.
- The SMF in the vPLMN forwards the Subscription Permanent Identifier (SUPI) of the UE to the SMF in the hPLMN.
- Based on UE subscription and limitation, its requests can be accepted or rejected by the SMF in the hPLMN.
- UE subscription information is obtained from the UDM in the hPLMN by the SMF in the hPLMN.
- QoS requirements associated with a PDU session can be sent to the SMF in the vPLMN by the SMF in the hPLMN during or after the establishment of the PDU session.
- The N9 interface between the two SMFs can carry user plane forwarding information.
- The SMF in the vPLMN may check QoS requests from the SMF in the hPLMN with respect to roaming agreements.

When a PDU session is to be established in the case of LBO, the AMF selects an SMF in the vPLMN. In this case, the SMF in the vPLMN may reject the N11 message from the AMF related with the PDU session establishment request message with a proper N11 cause. This triggers the AMF to select both a new SMF in the VPLMN and an SMF in the hPLMN to handle the PDU Session using home routed roaming.

4.4.2. Techniques for low latency roaming

This section discusses two mechanisms to reduce roaming delays. First, Network Reselection Improvements (NRI) defined by 5GAA are discussed in Section 4.4.2.1. Second, steering of roaming techniques are presented in Section 4.4.2.2.

4.4.2.1. Network reselection improvements

The 5GAA Cross-Working group work item [20] investigates NRI to reduce roaming delays in automotive use cases identifying three main roaming architectures:

- **UE roaming with new registration**:

  This architecture is based on the concept of HR roaming and relies on the implementation of the N16 interface between the SMF in the home network and its counterpart in the visited network. In this approach the SMF in the home network is responsible of choosing the UPF in the home network, whereas the SMF in the visited network is responsible of choosing the UPF in its network. The roaming process starts when the signal from the home network of a user starts to become very low. The connection of the UE to its home network will persist until the signal is totally lost. At this point, the UE will start scanning the spectrum to find a carrier of a PLMN that appears in the roaming list. The UE will try to attach to this PLMN by completing registration and authentication, in a similar way as if the UE was turned on. If the attachment is successful, the user plane connection should be re-established by creating a new PDU context. Therefore, this legacy roaming may result in large interruption time (in the order of hundreds of seconds), which is unacceptable for most of the use cases' services.

- **UE roaming with AMF relocation and RAN assistance**:

  In this architecture the AMF function of the home and visited networks are interconnected through the N14 interface. The architecture is depicted in Figure 27 [20]. This roaming architecture can achieve roaming times as low as 1 second, if the following features are implemented:

- **Early home network release**: The radio nodes in the home network should be configured to release the connection of the UE once there is still a good signal level from the home network, and an attachment to the visited network can be completed. Finding this threshold requires drive test optimization around the radiofrequency border.
- **Scanning assistance**: When releasing the UE, the home network can include redirect information in the NAS release message saving scanning time for the UE to find the visited network, using for instance Steering of Roaming (SoR) described in the next section.
- **Reducing failed attachments**: If the two PLMNs, home and visited, are configured as equivalent PLMNs in the UE, then the attachment request of the UE in the visited network is guaranteed to succeed.
- **Reducing attachment time in the visited network**: Thanks to the N14 interface between AMFs the visited network can retrieve the UE context from the home network's AMF, and so a full attachment and registration is not required.
- **Reducing user plane establishment time**: Through the N14 interface the visited network is made aware of the UPF and IP address of the home network. User plane is res-established as part of the tracking area updated in the visited network.

- **UE roaming with AMF relocation and handover**:

This architecture builds on additional optimizations to the ones described in the UE roaming with AMF relocation and RAN assistance and can deliver roaming times as low as 100 ms. The architecture builds on the N14 architecture described in Figure 27 but also configures the radio nodes (gNBs) so that cells across the border on the visited network are configured as neighbouring cells. This means that the home network instructs the UEs to scan for the quality of the visited network cells across the border, and a network handover is triggered when the handover threshold is crossed. Notice that this architecture requires further coordination between the home and visited networks, as low-level parameters such as cell IDs need to be communicated across networks.

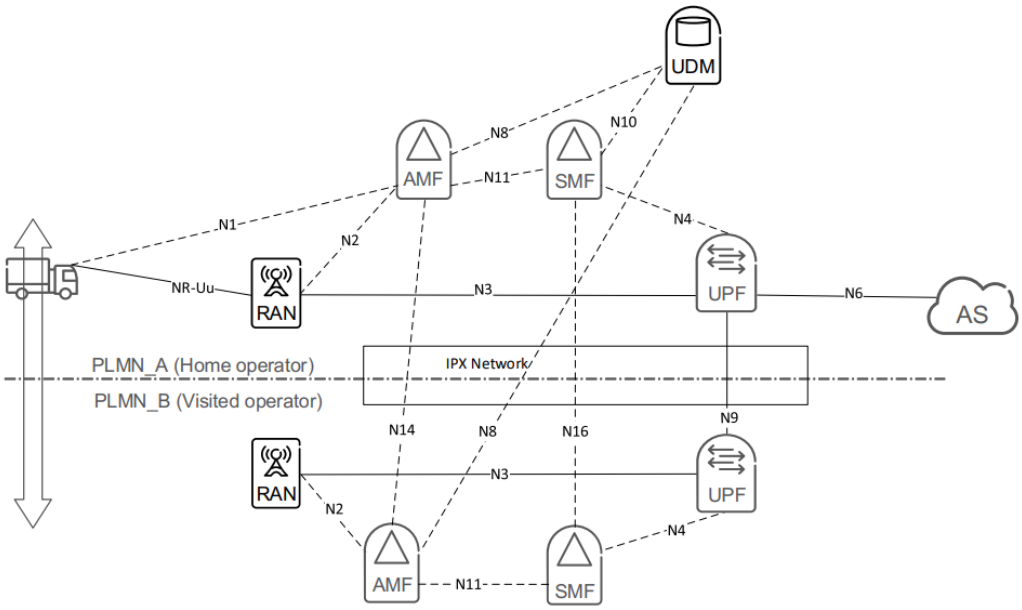

Figure 27. Roaming architecture with N14 interface between AMFs [20]

4.4.2.2. Steering of roaming in 5GS

The Steering of Roaming (SoR) solution in 5GS [21][22] is a control-plane solution introduce in 3GPP release 16 specifications. It allows the hPLMN to update UEs with the list of preferred PLMN/access-technology combinations. It was developed to fulfil the following aspects that are missing from the currently deployed SoR mechanisms. The control plane SoR solution allows the hPLMN to direct the UE during or after registration on the vPLMN [21]:

- Provide a standardized directive of roaming mechanisms when the UE is trying to register with a vPLMN. Specifically, the objective is to allow the hPLMN to request a UE, which is in automatic mode, to try to attach to a different vPLMN from the one it is currently using or trying to attach to. The recommended vPLMN should be available and not be in a forbidden list.
- Enable the UE to detect and stop using a misbehaved vPLMN, if it is modifying or not correctly transmitting the control plane SoR information provided by hPLMN.

The control plane SoR in 5G can be used during registration or after registration to a vPLMN by allowing the hPLMN to securely update the "Operator Controlled PLMN Selector with Access Technology" list in the UE based on operator policies (e.g., vPLMN preferences, UE location). This is done by providing the hPLMN protected list of preferred PLMN/access technology combinations via NAS signalling.

To allow UEs to use the SoR, the hPLMN must configure their UE's USIM to support the reception of SoR information at the initial registration in a vPLMN. In addition, the hPLMN should update the UDM by the fact that this USIM expect to receive SoR during the registration phase.

The 5G SoR approach has the following benefits:

- It is a Subscriber Identity Module (SIM)-based steering like the Over-The-Air (OTA) current steering.
- The preferred PLMN list is sent through signalling and not through an independent channel like SMS, as it is currently done. This will make the delivery of the preferred list much more effective compared to SMS, where the list might not reach the SIM due to different reasons.

Due to the on-going development of the 5G features, the 5GMED consortium will be focusing on the optimization of the roaming from within the core network only. In case RAN features will be available during the execution of the project, they might be also implemented in the final solution.

4.5. Solutions to Challenge #4: MEC deployment and inter-MEC connectivity

To overcome the fourth technical challenge described in Section 4.1.4, we present in Section 4.5.1 the ETSI MEC architecture and its support in Raemis and Nearby One platform. This is the architecture that is currently selected by the 5GMED project to deploy MECs. The operator platform proposed by GSMA as a federated architecture with unified interfaces in Section 4.5.2 as a possible solution for inter-MEC connectivity that will be 5GMED project.

4.5.1. ETSI MEC architecture

The 5G architecture was defined as a Service Based Architecture (SBA) and it is highly relying on Network Function Virtualisation (NFV) and Software Defined Network (SDN). These two technologies are known to be key enablers for edge computing.

Multi-Access Edge Computing (MEC) is the concept of implementing communications handling capabilities at the Edge of the network, closer to the cellular radio access points. This contrasts with previous centralised core network architectures, which brought user traffic to a central point in the network before processing could take place. The MEC system is composed of MEC hosts and the MEC management.

The MEC management comprises the MEC system level management and the MEC host level management. The latter manages a MEC host functionality and the applications running on it. It comprises the MEC platform manager and the virtualization infrastructure manager. The MEC system level management has an overview of the complete MEC system and is composed mainly of the MEC orchestrator [26].

The **MEC orchestrator** has the overall view of the entire MEC system in terms of the deployed MEC hosts, the available resources, as well as the MEC services. It is also responsible for onboarding the applications to the appropriate MEC Hosts based on the defined constraints, such as latency, resources, and available services among others [27].

The **MEC platform manager** is responsible for the lifecycle management of the applications, which includes their instantiation, maintenance, and termination. Moreover, it is responsible for handling traffic rules and DNS configuration. Lastly, it serves as the Element Manager focused on the FCAPS management of the MEC platform and is responsible for the following functions [27]:

The **MEC host** is an entity that contains a MEC platform and virtualization infrastructure, providing the necessary resources for computation, storing, and networking functionalities necessary to run a MEC application. Well-known Network Functions Virtualization Infrastructures (NFVIs) are OpenStack and Kubernetes. OpenStack relies mainly on using VMs as the virtualization technology, whereas Kubernetes relies on virtual containers. Kubernetes containers are considered the next evolution of virtual machines. Indeed, in the 5GMED project, Kubernetes is a key component of the MEC platform, with the aim of hosting virtualized applications. The MEC host is composed of:

- **MEC platform**, which contains a set of functionalities required to run MEC applications and enable them to consume MEC services.
- **MEC applications**, which are instantiated on the virtualisation infrastructure of the MEC host. Basically, these are actual applications like Augmented Reality/Virtual Reality, Edge Cloud Gaming, Video Analytics etc.
- **Virtualisation infrastructure**, which contains compute, storage, and network resources, virtual machines.

The ETSI MEC specification states that the functionality of the MEC platform is provided through three well-defined interfaces [27]:

- **Mp1**: used for communication among the MEC platform and the MEC applications with the aim of providing both service registration and discovery. Through this interface, it is possible to obtain information regarding the MEC applications such as their availability, as well as to

configure the MEC applications with network policies such as traffic rules or DNS rule activation. Moreover, the MP1 interface enables access to persistent storage, which is typically in the form of volumes, shares, or containers. It must be noted that persistent storage is independent of any running MEC instance. This storage is used for any data that needs to be reused, hence it exists beyond the life span of a specific MEC instance.

- **Mp2**: Reference point between the data plane of the virtualization infrastructure and the MEC platform. It is used to instruct the data plane on how to route the traffic among services, networks, and applications. This reference point is not further specified by ETSI.
- **Mp3**: Reference point between MEC platforms. It is used to control communication between MEC platforms.

To support MEC applications, the 5G SA core network combines the necessary NFs for the user, i.e., UPF, as well as the control plane NFs, i.e., AMF, SMF, PCF, UDM, NSSF. On one hand, the SMF can establish multiple PDU sessions (IPv4, IPv6, or Ethernet) that enable the UE to communicate with multiple UPFs and access services in distinct Data Networks (DNs), e.g., local DN hosting MEC applications. Moreover, the 5G Core also includes the Network Data Analytics Function (NWDAF) that enables network analysis information, which can be leveraged by MEC applications.

4.5.1.1. Support of ETSI REM by Raemis

Two models of Edge Computing are supported by the Raemis platform. The basic model is pure MEC data plane offload. This model has been promoted by the ETSI MEC working group and an overview of the Raemis support for this model is provided next.

The MEC data plane offload model is not an effective approach, however, in providing resilience and survivability to the network Edge. To address this requirement, which is key for critical communications use cases, the Raemis platform also provides the distributed Edge Core model, which provides fully resilient service at the network edge in addition to the data plane offload capabilities.

Figure 28 shows the 5G core components arranged for Edge data plane offload. The Raemis Data plane module is common across 4G and 5G and it consists of:

- **Data Plane Controller**: Orchestrates the packet switching operation and logically sits close to the SMF and its functionality.
- **Packet Switch**: Implements the packet switching functionality of the user plane including Edge local breakout, referred to as Uplink Classifier UL-CL in 5G, based on the instructions sent to it by the Controller.

Figure 28 shows subscribers Bob, Susan, and Fred configured for MEC local offload to their enterprise LANs 1, 2, and 3, respectively, while present in the appropriate tracking areas. The SMF element interacts with the AMF to maintain knowledge of the current area of the user and supplies this to the Controller. The Controller is supplied with additional user level information by the MEC orchestrator which defines MEC behaviour at a subscriber level.

The Raemis Local Access Gateway (LAGW) elements are used as a gateway to the various local enterprise LANs. The MEC orchestration function orchestrates the Edge and core elements into an overall MEC service for specific users.

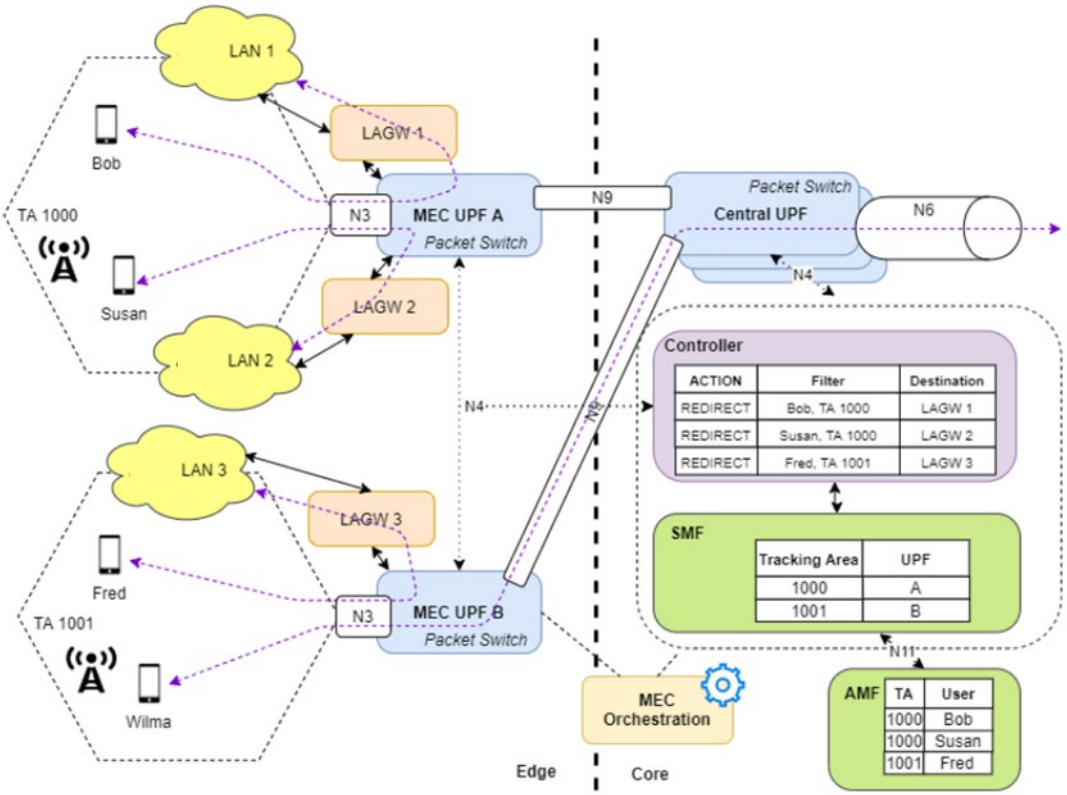

Figure 28. 5G Core components with Edge Data plane offload.

4.5.1.2. Support of ETSI MEC by Nearby One

Nearby One Solution fulfils all the requirements defined by ETSI MEC with respect to the MEC Orchestrator and MEC Platform Manager. It also covers the functionality of the MEC platform component, referred to as Nearby Blocks. Besides, the solution natively integrates with third-party Virtual Infrastructure Managers (VIM). Moreover, the solution extensively supports the integration of Docker containers into both VIMs and bare metal servers.

Nearby One has also the capability of provisioning and monitoring the infrastructure when needed. As the solution natively implements the functionalities of the MEC Orchestrator and the MEC platform manager, it implements the functionalities of the interfaces from Mm1-Mm9 through the Orchestration Platform where Mm1 and Mm2 are externally exposed to operator's Operations Support System (OSS).

The Mm1 serves as the interface between the MEC Orchestrator and the OSS, which is used for triggering the instantiation and termination of MEC Applications in the MEC System. This integration is done by leveraging standardized APIs, which can be accessed via REST interfaces or using a GraphQL interface. On the other hand, the Mm2 is the interface between the OSS and the MEC Platform Manager, which is used to configure the MEC platform and perform fault and performance management. The rest of the interfaces are internal to the solution but may be exposed when needed.

4.5.2. Operator platform concept

Growth in service innovation provides several opportunities and challenges to Service Providers (SPs). Opportunities come in form of providing better service than the competitors, becoming more relevant to society, and solving complex problems in human civilization through advanced capabilities of communication services. Challenges come in many forms and some of them are related to management of the complex ecosystem of service platforms, seamless integration with non-telecom capabilities, managing the balance between cost and benefits, and remaining innovative from new service feature perspective [23].

Collaboration between SPs and Edge/Access Operators (E/AOs) under some kind of federation is one of the most promising options of SPs to reduce risk and cost. In the last decade, infrastructure sharing SPs and E/AOs has been flourishing as it provides sustainability to telecom industry through shared cost and risk [24].

The future of communication services is evolving around platforms and freedom away from very heavy core network systems that historically provided a monolithic structure to SPs and E/AOs. SPs and E/AOs are pushing more capability towards edge of the network and MEC is proving exceptionally beneficial, since most of the analytical functions can be logically hosted at MEC platform and service configuration can be routed through MEC-based capability.

Therefore, adopting federation and collaboration at service platform, especially MEC layer capabilities may prove beneficial for SPs and E/AOs, who are consistently chasing the high paced innovation in service domain. MEC federation and collaboration provides a unique proposition in multi-industry services, e.g., Industry vertical solutions or industrial automation. Similarly, 5GMED project with its different use cases requires a MEC infrastructure supporting multi-tenant services and implementing unified or standardised interfaces.

There are multiple ways through which SPs and E/AOs can take advantage of MEC federation and collaboration. In one of the most recent and thorough efforts, GSMA has initiated "Operator Platform Concept" and recently presented its main idea [25].

The Operator Platform (OP) concept is defined in [25] as "a common platform exposing operator services/capabilities to customers/developers in the 5G-era in a connect once, connect to many models". The main objective of the OP is the standardisation of a common federation interface that allows operators to expand their coverage and collaborate with hyperscalers and other service providers **Error! Reference source not found.**.

The operator platform group brings together operators, platform developers, edge cloud providers, Open-Source Projects, etc. This generic platform will be highly flexible to be able to serve the plethora of emerging applications from healthcare to industrial Internet of Things (IoT), which will be very advantageous for any operator. Having a unified platform will be also beneficial to the operators, especially with the trend of having their assets and capabilities consistently available across networks and national boundaries.

The first phase of Operator Platform will be In Phase 1, the Operator Platform will federate the edge computing infrastructure of different operators, which is an interesting approach for 5GMED. This will create a common platform where application providers can deploy innovative, resource consuming,

distributed, and low-latency application using a set of unified APIs in close proximity to their clients and at cross-borders.

4.5.2.1. OP architecture and APIs

In OP architecture, each operator can hold an instance independent of the deployment in other operators. It consists of a common exposure and capability framework, including federation interface towards other operators and the platforms that provide the capabilities. This architecture is based on the following APIs [25]:

- **Northbound interface (NBI) APIs**: responsible of service management. The NBI is designed to fulfil enterprise and application provider use case requirements. It follows existing cloud APIs principles to facilitate using existing cloud platforms. Any application provider, another enterprise or service provider can be a consumer of network capability and services.
- **East-Westbound Interface (EWBI) APIs**: responsible of exchanging information between operators.
- **Southbound Interface (SBI) APIs**: connecting the operator platform with the specific operator infrastructure that will deliver the network services and capabilities to the user.
- **User-Network Interface (UNI) APIs**: responsible of allowing final equipment to set communications towards OP and opening new capabilities at user level e.g., dynamic service requests or location data.

From the operator's point of view, network and service functions shall be offered by application providers in the form of NBIs. This can be done through the OP by enabling capabilities to be tied to these functions.

4.5.2.2. Main players

Figure 29 shows the different players, the APIs, and the relationships among them. The main players in the OP are [25]:

- **Application provider**: The owner of the application or service. It offers services or applications to the end user through network capabilities, such as cloud or edge, that can be reached via the OP.
- **Network Resources:** Uses the SBI-NR to allow interaction between mobile network and the Service Resource Manager Role of the OP.
- **Cloud Resources:** Uses the SBI-CR to enable connectivity to the underlying infrastructure to perform the orchestration of workloads to the Edge Cloud Infrastructure.
- **Operator Platform:** It facilitates access to the Edge Cloud Capability of an operator or even other operators that are part of the federation. It enables connectivity to the end users to access the applications. It consists of the following roles:
    - Capabilities Exposure Role
    - Service Resource Manager Role
    - Federation Manager Role
    - Federation Broker Role
- **End user**: The customer of the operator that uses the UNI to communicate with the OP to access the application provider's services through operator's network resources.

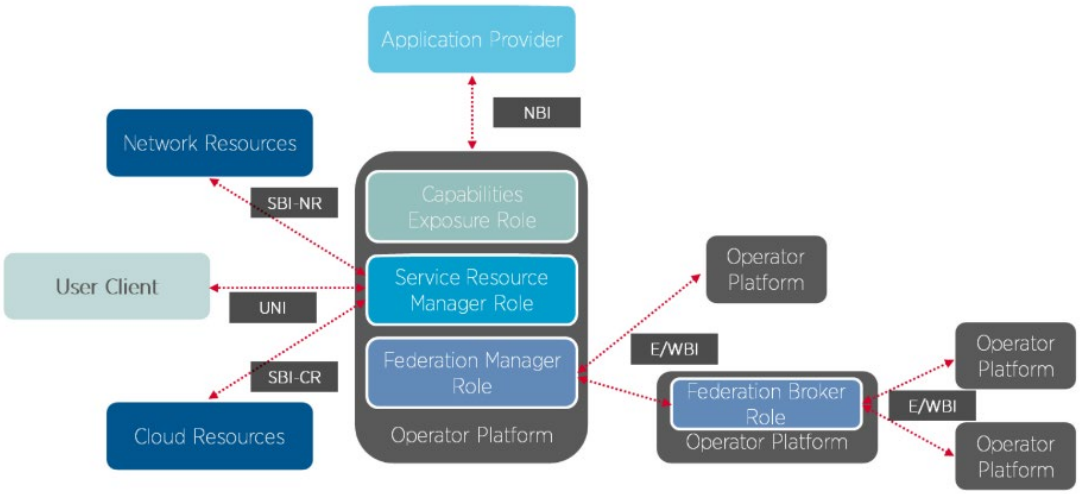

Figure 29. Federation among multiple OP instances [25].

4.5.2.3. OP architecture for edge computing

Figure 30 shows the high-level architecture of the OP for edge computing. It consists of the following elements:

- Operator platform: Single point of entrance for application providers to deliver edge computing capabilities in close vicinity of the users. Edge computing capabilities across different footprints are connected through OP.
- Application provider: Uses NBIs to request edge computing capabilities through OP. This element makes use of the NBI to enable service management and fulfilment.
- Edge computing platform: It is connected to the OP through SBI for resource management to provide the infrastructure that will deliver the network services to the users.
- End user: the customer who requests application usage through UNI and uses resources allocated on a proper edge instance. The selection of the resources is based on user's location on network requirements/status.

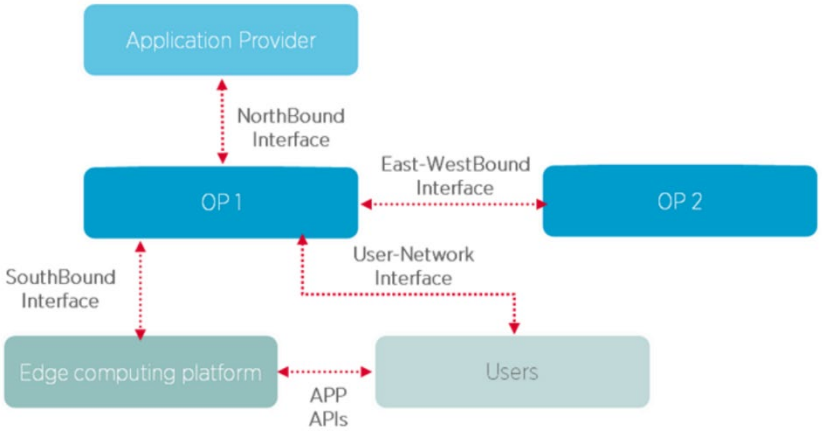

Figure 30. OP high level architecture for Edge Computing [25].

In line with the growth in service innovation, new use cases have also emerged that have stringent requirements in terms of compute power, low latency, and throughput intensive, among others. With these requirements, the Mobile Network Operators (MNO) intend to capitalize on their edge and network capabilities.

The intention to capitalize on the edge and network capabilities led to the creation of Telco Edge Cloud, also known as TEC. It aims to expose and provide a platform that federates the operators' telco edge resources providing a single interface to TEC customers that hides the network complexities. In this manner, TEC customers can deploy different services across different operators, delivering better end-user experiences [29].

To achieve this, the Operator Platform (OP) created by GSMA/OPG, which is a set of modules, is leveraged by the operators, allowing them to put the applications of different application providers closer to the customers. The association between the OP and one or more operators allows application providers to reach various networks, enlarging their reach in terms of providing services to end users [30].

4.5.2.4. CAMARA project and Inter-MEC connectivity

An open-source project called CAMARA within the Linux Foundation aims to define, develop, and test APIs. CAMARA works closely with the GSMA OPG to ensure that API requirements, as well as API definitions, are aligned [31].

The scope of the project is limited to the following:

- Collect API requirements from GSMA Operator Platform Group (OPG) and other sources
- Define Service APIs
- Create test plan / cases / tools from an API consumer perspective
- Develop and test Service APIs
- Create developer friendly documentation for Service APIs

Concerning the 5GMED project, the NearbyOne Orchestrator, also a member of GSMA and the CAMARA project, closely follows the developments and adheres to the requirements of GSMA OPG to support the federation and cross-border interfaces between orchestrators, and cross-border interfaces between MECs. Moreover, NearbyOne natively supports some of the requirements, and in effect, fewer developments are necessary to make the solution fully compliant with GSMA OPG's requirements in terms of the federation.

The deployment of MEC layer will allow us to reduce the latency to reach the stringent requirement of very low latency in UC4. In addition, the availability of inter-MEC interface will reduce the mobility interruption time and service migration time. In addition, it will reduce the packet loss, and therefore increase reliability, by forwarding user traffic received in the original MEC to the following one.

5. Conclusions

In this document, we have provided an analysis of the use cases defined in 5GMED and the technical challenges that might face the consortium when deploying such use cases in a challenging environment such as the "Figueres – Perpignan" cross-border corridor.

Each use case was analysed separately, where network KPIs were calculated from service KPIs based on specific mapping rules. The obtained network KPIs were then consolidated to pinpoint the most challenging ones. It was found that data rates, latency, and mobility interruption time values will be among the key factors to be considered when designing the network architecture. This document has provided the networking requirements from each of the 5GMED use cases, together with the high-level architecture of the network.

Furthermore, the cross-border corridor was analysed in terms of possible technologies to be used depending on the geographical characteristics of the corridor segments (i.e., tunnel, spectrum licensing, etc.), availability of computing resources, coverage gaps, and technical challenges. Regarding technical challenges, we have identified the following:

- Multi-Connectivity and integration of heterogeneous networks to allow using different technologies, which can be required in certain zones, in a transparent manner to the services.
- Virtualization, network automation, and network slicing support in 5G Stand-Alone (SA) Core to enable service isolation and resource optimisation, and therefore enable the calculated KPIs of the use cases.
- Low latency 5G roaming at cross-border to reduce mobility interruption time, which is one of the main challenging KPI.
- Inter-MEC connectivity to enable services using edge computation to be served without service interruption when moving from one MEC to another

Existing solutions in the state of the art to overcome these challenges have been analysed, especially standardisation and off-the-shelf solutions, and they will be the starting point of the solutions to be adopted by the final network architecture of 5GMED.

Furthermore, the roaming scenarios between the Spanish and French networks have been analysed to implement and optimize the roaming strategies and minimize the impact on the users. Finally, network orchestration has been exposed, addressing the cross-border situation and challenges.

The work in WP3 is currently under development and it will be extended in D3.3 and D3.4, where it shall present the adopted 5GMED network architecture, cross-border interfaces, and their implementation in the small-scale and large-scale testbeds.

References

[1] 5GMED Project, D2.1. Initial definition of 5GMED use cases, March 2022.
[2] 5GMED Project, D3.3. First release of 5GMED ICT infrastructure.
[3] GSMA, "How roaming affects mobile speeds in Europe," Feb. 2020, (available at https://www.gsma.com/membership/resources/how-roaming-affects-mobile-speeds-in-europe/).
[4] https://github.com/facebookincubator/katran.
[5] https://www.optofidelity.com/blog/comparing-vr-headsets-tracking-performance#:~:text=Today%20latencies%20below%205ms%20are,the%20pose%20to%20the%20content.
[6] https://www.qualcomm.com/documents/whitepaper-making-immersive-virtual-reality-possible-mobile.
[7] https://www.forsk.com/atoll-overview.
[8] ETSI EN 302 663 V1.2.0, Intelligent Transport Systems (ITS); Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band, Nov. 2012.
[9] 5GMED Project, D3.2. 5GMED ICT architecture and initial design, March 2022.
[10] VMWare, "Virtualization Overview", White Paper, xx.
[11] J. Kremer "Cloud Computing and Virtualization" White Paper, October 2022, (available at https://cloud.report/Resources/Whitepapers/b7de8ebf-b2c9-4c77-b1d3-1f4bafd8ad02_143.pdf). .
[12] Q. Zhang, L. Liu, C. Pu, Q. Dou, L. Wu and W. Zhou, "A Comparative Study of Containers and Virtual Machines in Big Data Environment," 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), 2018, pp. 178-185, doi: 10.1109/CLOUD.2018.00030.
[13] M. Moravcik, M. Kontsek, "Overview of Docker container orchestration tools," 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), 2020, doi: 10.1109/ICETA51985.2020.9379236.
[14] 5GMED project, D3.2. 5GMED ICT architecture and initial design, May 2022.
[15] C. Saunderson, M. Bhat, D. Betts, H. Ennaciri "Market Guide for Service Orchestration and Automation Platforms, Gartner Research, August 10, 2021 (available at https://www.gartner.com/en/documents/4004556).
[16] NGMN Alliance, "Description of Network Slicing Concept," 2016 (available at https://www.ngmn.org/wp-content/uploads/ 160113_NGMN_Network_Slicing_v1_0.pdf).
[17] Raemis™ – Cellular Network Technology, (available at https://www.druidsoftware.com/raemis-cellular-network-technology/).
[18] GSMA, Official Document NG.113, "5GS Roaming Guidelines", version 5.0, December 2021.
[19] 3GPP TS 23.501 - Technical Specification Group Services and System Aspects, "System Architecture for the 5G System (5GS)", Stage 2, Release 17, v17.2.0, September 2021.

[20] 5GAA Automotive Association Technical Report, "Cross-Working Group Work Item Network Reselection Improvements (NRI)," v1.0, 2020.
[21] GSMA, "Steering of Roaming Implementation Guidelines," May 2020, Version 5.0 (available at https://www.gsma.com/newsroom/wp-content/uploads/IR.73-v5.0-1.pdf).
[22] 3GPP TS 29.550 - Technical Specification Group Core Network and Terminals, "5G System; Steering Of Roaming Application Function Services," Stage 3, Release 18, v17.4.0, June 2022.
[23] ITU-T technical specification, FG-NET2030 – Focus Group on Technologies for Network 2030. "Network 2030 Architecture Framework," June 2020.
[24] A. Antonopoulos et al., "Energy-efficient infrastructure sharing in multi-operator mobile networks," in *IEEE Communications Magazine*, vol. 53, no. 5, pp. 242-249, May 2015.

[25]GSMA, "Operator Platform Telco Edge Proposal," January 2020.
[26]ETSI White Paper No. 28, "MEC in 5G networks," June 2018.
[27]ETSI GS MEC 003 V2.2.1, "Multi-access Edge Computing (MEC); Framework and Reference Architecture," Dec. 2020
[28]T. Taleb , A. Ksentini, and P.A. Frangoudis, "Follow-Me Cloud: When Cloud Services Follow Mobile Users," IEEE Transactions on Cloud Computing, Vol. 7, No. 2, April-June 2019.
[29]GSMA white paper, "Telco Edge Cloud - Value & Achievements," March 2022.
[30]GSMA white paper, "Operator platform concept - phase 1: Edge cloud computing," January 2020.
[31]https://camaraproject.github.io/Scope.html