Fifth Generation Cross-Border Control

Deliverable D3.2
Intermediate E2E, MEC & Positioning Architecture

Version: v1.0
2021-01-31

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825050. Any 5GCroCo results reflects only the authors’ view and the Commission is thereby not responsible for any use that may be made of the information it contains.

http://www.5g-ppp.eu
# Deliverable D3.2

## Intermediate E2E, MEC & Positioning Architecture

<table>
<thead>
<tr>
<th>Grant Agreement Number:</th>
<th>825050</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Name:</td>
<td>Fifth Generation Cross-Border Control</td>
</tr>
<tr>
<td>Project Acronym:</td>
<td>5GCroCo</td>
</tr>
<tr>
<td>Document Number:</td>
<td>5GCroCo/D3.2</td>
</tr>
<tr>
<td>Document Title:</td>
<td>Intermediate E2E, MEC &amp; Positioning Architecture</td>
</tr>
<tr>
<td>Version:</td>
<td>v1.0</td>
</tr>
<tr>
<td>Delivery Date:</td>
<td>2021-01-31</td>
</tr>
<tr>
<td>Editor(s):</td>
<td>Francesca Moscatelli (NXW)</td>
</tr>
<tr>
<td>Authors:</td>
<td>Jens Schwardmann (BOSCH), Jesus Alonso-Zarate, Juan Luis de la Cruz, Ramon Casellas, Raul Muñoz, Ricard Vilalta, Roshan Sedar, Francisco Vázquez-Gallego (CTTC), Oliver Goldich (DTAG), Maciej Muehleisen, Massimo Condoluci (ERI), Apostolos Kousaridas (HWDU), Miguel Catalán (I2CAT), Sokratis Barmpounakis (NKUA), Wolfgang Scheidl (NOKIA), Francesca Moscatelli, Leonardo Agueci (NXW), Laurent Gallo, Mathieu Lefebvre Jean-Marc Odinot, Stefan Wendt, Laurent Mussot (Orange), Rodrigo Silva, Antonio Eduardo Fernandez Barciela (PSA)</td>
</tr>
<tr>
<td>External Reviewers:</td>
<td>Giada Landi (NXW), Marisa Catalan (I2CAT), Sebastian Euler, Mazen Abdel Latif (ERI), Andreas Heider-Aviet (DTAG), Chan Zhou (HWDU)</td>
</tr>
<tr>
<td>Keywords:</td>
<td>5G, Standalone 5G, Non-Standalone 5G, New Radio, Cross-Border / -MNO Handover, End-to-End QoS, QoS Prediction, Network Slicing, Network Slice Templates, Positioning, Mobile Edge Computing, Mobile Edge Cloud, MEC, Management and Orchestration, MANO, Software Defined Networking, SDN</td>
</tr>
<tr>
<td>Status:</td>
<td>Final</td>
</tr>
<tr>
<td>Dissemination level:</td>
<td>Public</td>
</tr>
</tbody>
</table>
## Version History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date of release</th>
<th>Description of changes</th>
<th>Last editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>31/01/2021</td>
<td>First version</td>
<td>Francesca Moscatelli</td>
</tr>
</tbody>
</table>
Abstract

This deliverable takes the user story descriptions from Deliverable D2.2 [1], the use case requirements from Deliverable D2.1 v3 [2] and the application architecture from Deliverable D3.1 [3], to be published end of January 2021, to provide an overview of the 5GCroCo end-to-end network architecture including solution concepts like predictive and end-to-end Quality of Service (QoS), Mobile Edge Computing/Cloud (MEC)¹ and precise positioning. Project partners provided descriptions of the individual solutions they evaluated in the project and jointly worked on a common architecture allowing to realize this. A key goal for such architecture has been to assure cross-telco-vendor, cross Mobile Network Operator (MNO), cross-border, and cross-car Original Equipment Manufacturer (OEM) interoperability from the start by at least identifying potential interoperability issues and ideally, already providing solution options.

After introducing the scope and objective of the document, a first analysis is provided to describe how the Fifth Generation (5G) solution components in this project extend beyond today’s baseline. Existing 4G infrastructures, already rolled out, are mostly used for Mobile Broadband (MBB) services and initial 5G rollouts are mostly tackling the same MBB and similar eMBB services for the broad public. The extensions proposed in 5GCroCo have the objective to go beyond this initial set of MBB and eMBB services, targeting the requirements of Cooperative, Connected and Automated Mobility (CCAM).

Following the extended abstract, the MBB baseline is described in more detail. Besides just focusing on Third Generation Partnership Project (3GPP) specifications, also virtualization and Software Defined Networking (SDN) associated aspects, related specifications and open-source best practice are presented since they are all used in today’s commercially deployed networks for MBB services. After that, use cases requirements and further challenges of the telco industry are discussed as prerequisite to evaluate how to evolve the baseline to support cross-border CCAM. Some of the solutions are already standardized and often well evaluated, but not in cross-telco-vendor, cross-MNO, cross-border, and cross-car-OEM context. This is particularly true for MEC and end-to-end QoS aspects. In other cases, especially in the context of QoS prediction, interfaces are not standardized yet and only first ideas about how to use these interfaces and on performance of different prediction algorithms exist. However, further research will be necessary and conducted before and during the trial phases of the project. A clear distinction is made between what is possible with initial 5G deployments, which still use a 4G Evolved Packet Core (EPC), and future ones that also include a 5G Core.

The document describes and discusses solutions in the field of enabling cross-border / -MNO handover, cross-MNO virtualization and SDN techniques, 3GPP QoS framework, including its 5G evolution, adaptive connectivity to most suitable MEC hosts, QoS prediction with and without networks support and three precise positioning solutions, one of them within the scope of 3GPP

¹ The Mobile Edge Cloud enables Mobile Edge Computing, so the expansion of the acronym depends on the sentence. The “M” should not stand for “Multi-access” since this would point at just one subset of MEC-related specifications, namely the European Telecommunications Standards Institute (ETSI) ones.
specifications. Based on this, WP4 will describe the deployment architectures to be used in the 5GCroCo trials.

The final goal for the architecture designed in WP3 is to support its scalable and realistic applicability to continent-wide deployments, as enabler for solid business models, to be defined and evaluated in WP5.
Contents

1 Introduction ........................................................................................................................................ 13
  1.1 Objective of the Document ........................................................................................................ 13
  1.2 Structure of the Document ......................................................................................................... 14
  1.3 Key 5G Solutions ....................................................................................................................... 15
2 Baseline Network Architecture ........................................................................................................ 21
  2.1 General Connectivity ................................................................................................................ 21
    2.1.1 Interfaces Toward Application and Backend ........................................................................ 24
    2.1.2 Network Service Continuity at Country Borders ................................................................. 31
    2.1.3 Network Orchestration and Control .................................................................................... 32
  2.2 Predictive and End-to-End QoS ............................................................................................... 36
    2.2.1 QoS Classes ........................................................................................................................ 37
    2.2.2 Session Management .......................................................................................................... 41
  2.3 MEC ........................................................................................................................................... 42
  2.4 Precise Positioning .................................................................................................................... 43
    2.4.1 GNSS .................................................................................................................................... 43
    2.4.2 Inertial Systems .................................................................................................................... 43
3 Challenges and Use Case Requirements ......................................................................................... 44
  3.1 General Connectivity ................................................................................................................ 44
    3.1.1 Network Service Continuity at Country Borders ................................................................. 44
    3.1.2 Network Orchestration and Control .................................................................................... 45
  3.2 Predictive and End-to-End QoS ............................................................................................... 46
    3.2.1 Tele-operated Driving .......................................................................................................... 46
    3.2.2 HD Mapping ....................................................................................................................... 46
    3.2.3 ACCA .................................................................................................................................. 47
    3.2.4 All Use Cases ...................................................................................................................... 47
  3.3 MEC ........................................................................................................................................... 47
    3.3.1 Tele-operated Driving .......................................................................................................... 48
    3.3.2 HD Mapping ....................................................................................................................... 48
    3.3.3 ACCA .................................................................................................................................. 48
3.4 Precise Positioning............................................................................................................. 48
  3.4.1 Tele-operated Driving ................................................................................................. 49
  3.4.2 HD Mapping.............................................................................................................. 49
  3.4.3 ACCA ......................................................................................................................... 49

4 State of the Art and Solutions ................................................................................................. 50
  4.1 General Connectivity .................................................................................................... 51
    4.1.1 Network Service Continuity at Country Borders: Cross-MNO Handover Optimization .......................................................................................................................... 51
    4.1.2 Network Orchestration and Control ........................................................................ 57
  4.2 Predictive and End-to-End QoS .................................................................................... 62
    4.2.1 QoS Classes and Session Management ................................................................. 62
    4.2.2 QoS Prediction ........................................................................................................ 67
    4.2.3 Prediction for Reduced Cost .................................................................................. 76
    4.2.4 QoS Prediction Algorithmic Models ...................................................................... 77
  4.3 MEC ............................................................................................................................... 87
    4.3.1 Related Technical Specifications and Studies and their Applicability ..................... 87
    4.3.2 Solutions for Selected Challenges .......................................................................... 92
    4.3.3 Relation to other Key 5G Solutions ....................................................................... 114
    4.3.4 Use Case Specific Solutions ................................................................................ 116
  4.4 Precise Positioning ......................................................................................................... 125
    4.4.1 Enhanced Vehicle Positioning Based on UWB ....................................................... 125
    4.4.2 Vehicle Positioning Based on Sidelink Measurements ......................................... 129
    4.4.3 3GPP GNSS Real-Time Kinematic (GNSS-RTK) ................................................... 133
  4.5 Security ........................................................................................................................ 137
    4.5.1 Used Methods and Results of the Security Task Force ......................................... 137
    4.5.2 Security Guideline for MQTT Protocol .................................................................. 138
    4.5.3 Security Guideline for Docker .............................................................................. 140
    4.5.4 Privacy/General Data Protection Regulation (GDPR) ........................................... 142

5 Interoperability Issues ........................................................................................................ 144
  5.1 Across Telco Vendors .................................................................................................. 144
    5.1.1 General Connectivity ............................................................................................. 144
    5.1.2 Predictive and End-to-End QoS .......................................................................... 145
5.1.3 MEC ................................................................. 145
5.1.4 Precise Positioning ............................................. 145
5.2 Across MNOs and Borders ......................................... 145
  5.2.1 General Connectivity ........................................... 145
  5.2.2 Predictive and End-to-End QoS ............................... 146
  5.2.3 MEC .................................................................. 146
  5.2.4 Precise Positioning .............................................. 147
5.3 Across Car OEMs ..................................................... 147
  5.3.1 General Connectivity ............................................ 147
  5.3.2 Predictive and End-to-End QoS ............................... 147
  5.3.3 MEC .................................................................. 148
  5.3.4 Precise Positioning .............................................. 148
6 Selected Open Issues and Refinements for Next Version .............. 149
  6.1 Deployment Aspects ................................................ 149
  6.2 MEC .................................................................. 149
  6.3 QoS Prediction Algorithms and Interfaces ....................... 150
  6.4 Security Aspects .................................................... 150
7 References .................................................................. 151
  A. Used Approach and Methods of the Security Task Force .......... 159
    A.1 Interviews with Partners ........................................ 159
    A.2 Security Analysis (Penetration Testing) of Applications ........ 161
  B. Classification of Data Centres and their Interconnection .......... 164
  C. Common Legend for Architecture Figures .......................... 166
# List of Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>Third Generation Partnership Project</td>
</tr>
<tr>
<td>5G</td>
<td>Fifth Generation</td>
</tr>
<tr>
<td>5GCroCo</td>
<td>5G Cross-Border Control</td>
</tr>
<tr>
<td>5QI</td>
<td>5G QoS Identifiers</td>
</tr>
<tr>
<td>ACCA</td>
<td>Anticipated Cooperative Collision Avoidance</td>
</tr>
<tr>
<td>AECC</td>
<td>Automotive Edge Computing Consortium</td>
</tr>
<tr>
<td>AF</td>
<td>Application Function</td>
</tr>
<tr>
<td>ALG</td>
<td>Application Layer Gateway</td>
</tr>
<tr>
<td>AMF</td>
<td>Access Management Function</td>
</tr>
<tr>
<td>AOA</td>
<td>Angle-of-Arrival</td>
</tr>
<tr>
<td>AOD</td>
<td>Angle-of-Departure</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>APN</td>
<td>Access Point Name</td>
</tr>
<tr>
<td>AppD</td>
<td>Application Descriptor</td>
</tr>
<tr>
<td>AUSF</td>
<td>Authentication Server Function</td>
</tr>
<tr>
<td>BBU</td>
<td>Base Band Unit</td>
</tr>
<tr>
<td>BDT</td>
<td>Background Data Transfer</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>BSS</td>
<td>Business Support System</td>
</tr>
<tr>
<td>C2C-CC</td>
<td>CAR 2 CAR Communication Consortium</td>
</tr>
<tr>
<td>CAM</td>
<td>Cooperative Awareness Messages</td>
</tr>
<tr>
<td>CCAM</td>
<td>Cooperative, Connected and Automated Mobility</td>
</tr>
<tr>
<td>CCU</td>
<td>Communication Control Unit</td>
</tr>
<tr>
<td>CDN</td>
<td>Content Delivery Network</td>
</tr>
<tr>
<td>CEE</td>
<td>Cloud Execution Environment</td>
</tr>
<tr>
<td>CEP</td>
<td>Connection End Point</td>
</tr>
<tr>
<td>CLI</td>
<td>Command Line Interface</td>
</tr>
<tr>
<td>CNCF</td>
<td>Cloud Native Computing Foundation</td>
</tr>
<tr>
<td>CPM</td>
<td>Cooperative Perception Message</td>
</tr>
<tr>
<td>CSC</td>
<td>Cloud SDN Controller</td>
</tr>
<tr>
<td>CSS</td>
<td>Cloud SDN Switch</td>
</tr>
<tr>
<td>CUPS</td>
<td>Control and User Plane Separation</td>
</tr>
<tr>
<td>DENM</td>
<td>Decentralized Environmental Notification Message</td>
</tr>
<tr>
<td>DL</td>
<td>Down Link</td>
</tr>
<tr>
<td>DNN</td>
<td>Data Network Name</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name System</td>
</tr>
<tr>
<td>E2E</td>
<td>end-to-end</td>
</tr>
<tr>
<td>ECCD</td>
<td>Ericsson Cloud Container Distribution</td>
</tr>
<tr>
<td>eMBB</td>
<td>enhanced MBB</td>
</tr>
<tr>
<td>eNB</td>
<td>Evolved Node B</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>EPS</td>
<td>Evolved Packet System</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>E-SMLC</td>
<td>Evolved Serving Mobile Location Center</td>
</tr>
<tr>
<td>E-UTRAN</td>
<td>Evolved Universal Terrestrial Radio Access Network</td>
</tr>
<tr>
<td>FR2</td>
<td>Frequency Range 2</td>
</tr>
<tr>
<td>GBR</td>
<td>Guaranteed Bit Rate</td>
</tr>
<tr>
<td>gNB</td>
<td>Next Generation Node B</td>
</tr>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>GPDR</td>
<td>Evolved Serving Mobile Location Center</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>GSMA</td>
<td>GSM Association</td>
</tr>
<tr>
<td>GST</td>
<td>Generic Network Slice Template</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>GW</td>
<td>Gateway</td>
</tr>
<tr>
<td>HD</td>
<td>High-definition</td>
</tr>
<tr>
<td>H2020</td>
<td>Horizon 2020</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HMAC</td>
<td>Hashed Message Authentication Code</td>
</tr>
<tr>
<td>HOT</td>
<td>Heat Orchestration Template</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IFA</td>
<td>Interfaces and Architecture</td>
</tr>
<tr>
<td>IMS</td>
<td>Internet Multimedia Subsystem</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPv4</td>
<td>IP Version 4</td>
</tr>
<tr>
<td>IPv6</td>
<td>IP Version 6</td>
</tr>
<tr>
<td>IPX</td>
<td>IP Exchange</td>
</tr>
<tr>
<td>ISG</td>
<td>Industry Specification Group</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>KVM</td>
<td>Kernel-based Virtual Machine</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LCM</td>
<td>Lifecycle Management</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-of-Sight</td>
</tr>
<tr>
<td>LPP</td>
<td>LTE Positioning Protocol</td>
</tr>
<tr>
<td>LRM</td>
<td>Linear Regression Model</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long Short-Term Memory</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MAC</td>
<td>Message Authentication Code</td>
</tr>
<tr>
<td>MANO</td>
<td>Management and Orchestration</td>
</tr>
<tr>
<td>MBB</td>
<td>Mobile Broadband</td>
</tr>
<tr>
<td>MBMS</td>
<td>Multimedia Broadcast Multicast Service</td>
</tr>
<tr>
<td>MCPTT</td>
<td>Mission Critical Push to Talk</td>
</tr>
<tr>
<td>MDO</td>
<td>Multi-Domain Orchestrator</td>
</tr>
<tr>
<td>MEC</td>
<td>Mobile Edge Computing/Cloud</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MME</td>
<td>Mobility Management Entity</td>
</tr>
<tr>
<td>MNO</td>
<td>Mobile Network Operator</td>
</tr>
<tr>
<td>MQTT</td>
<td>Message Queuing Telemetry Transport</td>
</tr>
<tr>
<td>NAS</td>
<td>Non-access Stratum</td>
</tr>
<tr>
<td>NAT</td>
<td>Network Address Translation</td>
</tr>
<tr>
<td>NBI</td>
<td>Northbound Interface</td>
</tr>
<tr>
<td>NE</td>
<td>Network Element</td>
</tr>
<tr>
<td>NEF</td>
<td>Network Exposure Function</td>
</tr>
<tr>
<td>NEP</td>
<td>Node Edge Point</td>
</tr>
<tr>
<td>NEST</td>
<td>NETwork Slice Type</td>
</tr>
<tr>
<td>NF</td>
<td>Network Function</td>
</tr>
<tr>
<td>NFV</td>
<td>NF Virtualization</td>
</tr>
<tr>
<td>NFVI</td>
<td>Network Function Virtualization Infrastructure</td>
</tr>
<tr>
<td>NFVO</td>
<td>NFV Orchestrators</td>
</tr>
<tr>
<td>NIS</td>
<td>Network Information Security</td>
</tr>
<tr>
<td>NLOS</td>
<td>Non-LOS</td>
</tr>
<tr>
<td>NRM</td>
<td>Network Resource Model</td>
</tr>
<tr>
<td>NS</td>
<td>Network Service</td>
</tr>
<tr>
<td>NSD</td>
<td>Network Service Descriptor</td>
</tr>
<tr>
<td>NSMF</td>
<td>Network Slice Management Function</td>
</tr>
<tr>
<td>NSS</td>
<td>Network Slice Subnet</td>
</tr>
<tr>
<td>NSSF</td>
<td>Network Slice Selection Function</td>
</tr>
<tr>
<td>NST</td>
<td>Network Slice Template</td>
</tr>
<tr>
<td>NWDAF</td>
<td>Network Data Analytics Function</td>
</tr>
<tr>
<td>ODL</td>
<td>Open Daylight</td>
</tr>
<tr>
<td>OEM</td>
<td>Original Equipment Manufacturer</td>
</tr>
<tr>
<td>ONAP</td>
<td>Open Networking Automation Platform</td>
</tr>
<tr>
<td>ONF</td>
<td>Open Networking Foundation</td>
</tr>
<tr>
<td>ONOS</td>
<td>Open Network Operating System</td>
</tr>
<tr>
<td>OPAF</td>
<td>Open Platform for NFV</td>
</tr>
<tr>
<td>OSI</td>
<td>Open Systems Interconnection</td>
</tr>
<tr>
<td>OSM</td>
<td>Open Source MANO</td>
</tr>
<tr>
<td>OSR</td>
<td>Observation-Space Representation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>OSS</td>
<td>Operation Support System</td>
</tr>
<tr>
<td>OTN</td>
<td>Optical Transport Network</td>
</tr>
<tr>
<td>OVF</td>
<td>Open Virtualization Format</td>
</tr>
<tr>
<td>OVS</td>
<td>Open vSwitch</td>
</tr>
<tr>
<td>PCF</td>
<td>Policy Control Function</td>
</tr>
<tr>
<td>PCRF</td>
<td>Policy and Charging Rules Function</td>
</tr>
<tr>
<td>PDN</td>
<td>Packet Data Network</td>
</tr>
<tr>
<td>PDU</td>
<td>Protocol Data Unit</td>
</tr>
<tr>
<td>PF</td>
<td>Prediction Function</td>
</tr>
<tr>
<td>P-GW</td>
<td>PDN GW</td>
</tr>
<tr>
<td>PNF</td>
<td>Physical NF</td>
</tr>
<tr>
<td>PoP</td>
<td>Point-of-Possession</td>
</tr>
<tr>
<td>QCI</td>
<td>QoS Class ID</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RNN</td>
<td>Recurrent Neural Network</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RRU</td>
<td>Remote Radio Unit</td>
</tr>
<tr>
<td>RSRP</td>
<td>Reference Signal Received Power</td>
</tr>
<tr>
<td>RSRQ</td>
<td>Reference Signal Received Quality</td>
</tr>
<tr>
<td>RSSI</td>
<td>Received Signal Strength Indication</td>
</tr>
<tr>
<td>RSU</td>
<td>Roadside Unit</td>
</tr>
<tr>
<td>RTA</td>
<td>Road Traffic Authority</td>
</tr>
<tr>
<td>RTK</td>
<td>Real-Time Kinematic</td>
</tr>
<tr>
<td>Rx</td>
<td>Receiving / Receiver</td>
</tr>
<tr>
<td>SA</td>
<td>Service Architecture</td>
</tr>
<tr>
<td>SBI</td>
<td>Southbound Interface</td>
</tr>
<tr>
<td>SCEF</td>
<td>Service Capability Exposure Function</td>
</tr>
<tr>
<td>SD</td>
<td>Slice Differentiator</td>
</tr>
<tr>
<td>SDK</td>
<td>Software Development Kit</td>
</tr>
<tr>
<td>SDN</td>
<td>Software Defined Networking</td>
</tr>
<tr>
<td>SEPP</td>
<td>Security Edge Protection Proxy</td>
</tr>
<tr>
<td>S-GW</td>
<td>Serving GW</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identity Module</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference Plus Noise Ratio</td>
</tr>
<tr>
<td>SIP</td>
<td>Service Interface Point</td>
</tr>
<tr>
<td>SIPTO</td>
<td>Selected IP Traffic Offload</td>
</tr>
<tr>
<td>SLA</td>
<td>Service Level Agreement</td>
</tr>
<tr>
<td>SMF</td>
<td>Session Management Function</td>
</tr>
<tr>
<td>SO</td>
<td>Service Orchestrator</td>
</tr>
<tr>
<td>SP</td>
<td>Service Provider</td>
</tr>
<tr>
<td>SPI</td>
<td>Security Protocol Index</td>
</tr>
<tr>
<td>SSC</td>
<td>Session and Service Continuity</td>
</tr>
<tr>
<td>SSR</td>
<td>State-Space Representation</td>
</tr>
<tr>
<td>SST</td>
<td>Slice/Service Template</td>
</tr>
<tr>
<td>TAPI</td>
<td>Transport API</td>
</tr>
<tr>
<td>TAU</td>
<td>Tracking Area Update</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TDOA</td>
<td>Time-Difference-of-Arrival</td>
</tr>
<tr>
<td>TFT</td>
<td>Traffic Flow Template</td>
</tr>
<tr>
<td>TOA</td>
<td>Time-of-Arrival</td>
</tr>
<tr>
<td>TOD</td>
<td>Time-of-Departure</td>
</tr>
<tr>
<td>ToD</td>
<td>Tele-operated Driving</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of Flight</td>
</tr>
<tr>
<td>TOS</td>
<td>Type of Service</td>
</tr>
<tr>
<td>TS</td>
<td>Technical Specification</td>
</tr>
<tr>
<td>Tx</td>
<td>Transmitting / Transmitter</td>
</tr>
<tr>
<td>TWR</td>
<td>Two-way-Ranging</td>
</tr>
<tr>
<td>UDM</td>
<td>Unified Data Management</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UL</td>
<td>Up Link</td>
</tr>
<tr>
<td>UPF</td>
<td>User Plane Function</td>
</tr>
<tr>
<td>UUID</td>
<td>Universally Unique ID</td>
</tr>
<tr>
<td>UWB</td>
<td>Ultra-Wideband</td>
</tr>
<tr>
<td>V2I</td>
<td>vehicle-to-infrastructure</td>
</tr>
<tr>
<td>V2N</td>
<td>vehicle-to-network</td>
</tr>
<tr>
<td>V2P</td>
<td>vehicle-to-pedestrian</td>
</tr>
<tr>
<td>V2V</td>
<td>vehicle-to-vehicle</td>
</tr>
<tr>
<td>V2X</td>
<td>Vehicle-to-Everything</td>
</tr>
<tr>
<td>VCoC</td>
<td>Vehicle Control Center</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>VIM</td>
<td>Virtualized Infrastructure Manager</td>
</tr>
<tr>
<td>VLAN</td>
<td>Virtual LAN</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>VNF</td>
<td>Virtual NF</td>
</tr>
<tr>
<td>VNFC</td>
<td>VNF Component</td>
</tr>
<tr>
<td>VNFD</td>
<td>VNF Descriptors</td>
</tr>
<tr>
<td>VNFM</td>
<td>VNF Manager</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>VoLTE</td>
<td>Voice over LTE</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual Private Network</td>
</tr>
<tr>
<td>VSB</td>
<td>Vertical Service Blueprints</td>
</tr>
<tr>
<td>VSD</td>
<td>Vertical Service Descriptor</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
<tr>
<td>WG</td>
<td>Working Group</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
1 Introduction

Deliverable 3.2 provides the initial version of the network architecture supporting cross-border CCAM, as designed in the 5GCroCo project. It is part of a complete end-to-end architecture when combined with the application architectures described in Deliverable D3.1 [3], which will be published end of January 2021.

1.1 Objective of the Document

This document describes the cross-border CCAM architecture envisioned by the 5GCroCo project. It provides the required level of detail for Work Package (WP) 4 to create trial deployment architectures for the small- and large-scale test sites. Its current and final version go beyond that and provide an architecture applicable for real-world commercial cellular networks. These networks should support CCAM in all its aspects and not just the three 5GCroCo use cases, described in Deliverable D2.2 [1], which were selected as representatives of a wide spectrum of services. First trials have been conducted and the corresponding result Deliverable D4.2 [4] provides information what instantiations of the architecture described in this document were used for that. Furthermore, Deliverable D4.1 [5] provides insights on instantiations of this architecture to be used in further trials and demonstrations planned.

The relation of this document to the other ones in WP3 and to selected WP2 and WP4 documents is shown in Figure 1-1.
Figure 1-1: Role of Deliverable 3.2 within WP3 and towards WP2 and WP4.

Beyond that, the document supports WP5 to discover business related topics and WP6 to discover standardization gaps.

1.2 Structure of the Document

In the next section an extended abstract is provided summarizing what needs to be added on top of a 4G MBB or Fifth Generation (5G) eMBB network to enable cross-border CCAM. After that, this document follows a different structure compared to other similar architecture documents. Rather than presenting the state of the art in research and technology, it presents a baseline network architecture in Section 2 according to what is currently commercially deployed in 4G Long Term Evolution (LTE) and currently rolled out 5G New Radio networks. Section 3 then provides use case requirements and further challenges the telco industry is facing. In Section 4, the state of the art for predictive and end-to-end QoS, MEC, and precise positioning is provided, analysing what needs to be added to the baseline architecture reported in Section 2 to converge towards the requirements provided in Section 3 and solve the challenges. In many cases, solutions beyond the state of the art are needed. Those are also described in Section 4. A clear distinction between non-standalone 5G New Radio, as it is currently being deployed, and standalone 5G New Radio, as it will be deployed in a few years, is made. The currently deployed non-standalone 5G New Radio networks still rely on the 4G EPC and, therefore, the full set of 5G features will not be available immediately.
Section 5 identifies cross-telco-vendor, cross-MNO, cross-border, and cross-car-Original Equipment Manufacturer (OEM) issues. In Section 6 solutions for some of these issues are provided. Further ones will be described in later iterations of the document, especially integrating the feedback and the experience that will be gathered from designing, executing and evaluating the trials in WP4. In particular, one more iteration of this document will follow with Deliverable D3.3 planned for October 2021. Section 7 keeps track of the open issues and refinements to be iteratively solved in coming document versions.

This deliverable is self-contained but closely related to the application architecture provided in Deliverable D3.1 [3]. Some topics, especially security (see Section 4.5) and QoS prediction (see Section 4.20) are covered in both deliverables. Most of the content for these topics is in this deliverable.

Figure 1-2: Split Between Deliverable D3.1 (Application Architecture) and D3.2 (Network Architecture)

Figure 1-2 illustrates the split between the two deliverables. Section 2.1.1 further explains the different types of components in the backend and in which case they are covered within this deliverable.

1.3 Key 5G Solutions

This project considers input from different standardization bodies to scope its solution architecture for CCAM in cross-border environments. The 3GPP 5G New Radio specifications are considered the main set of input documents. Some key aspects are widely considered being a part of 5G but are not within the scope of 3GPP specifications. For these, other standardization bodies, as well as best practices from open-source communities, are considered instead.

The baseline for this project is, therefore, the state of the art of technologies implemented and integrated today in 5G networks that are currently being deployed. It includes Physical and Virtual Network Functions (PNFs and VNFs) and SDN to interconnect them. For the baseline, 3GPP 5G New Radio Access Network (RAN) with a 4G LTE EPC are considered. This is referred to as non-standalone 5G New Radio. This deployment is being rolled out in most parts of the world today, including Europe. Experience from previous network generation rollouts shows that this deployment could remain in place for several years. Some, but by far not all, 5G Core features are also available with 4G EPC since the LTE specifications are also being evolved. According
to this baseline, all use cases will benefit from the increased capacity, reduced latency and improved reliability offered by design by 5G.

On top of this baseline, components are added, and/or their configuration is optimized to serve the challenges of cross-border CCAM. The three use cases allow to more systematically discover the different facets of those challenges in order to design a network capable of supporting a wide range of advanced use cases.

QoS prediction has been identified as one key solution. Its baseline is the 3GPP QoS framework currently mostly used for voice calls. For this rather simple application, the current way of describing QoS requirements by delay budgets, packet loss rates and throughputs is enough. More advanced use cases have more complex requirements. A first version of Generic Network Slice Templates (GSTs) was defined by Global System for Mobile Communications Association (GSMA) [6] and we consider them a solution to describe service requirements across MNOs and country borders. Their generic principle is equally applicable to 4G and 5G core networks, but the respective document [6] explicitly references the 5G Core specifications [7]. Furthermore, the standalone 5G New Radio with its defined Slice/Service Type (SST) for Vehicle-to-Everything (V2X) will allow better integration with the vehicle and backend, especially for identifying the right slice.

Besides information on instantaneous QoS, looking ahead in time and allowing selection of alternative QoS is being studied in 3GPP (QoS Sustainability Analytics [8]). The required new interfaces and/or changes to existing ones might only become available for the 5G Core. For the project trials and intermediate deployments and tests, proprietary interfaces can be used but those always come at the risk of cross-telco-vendor issues towards the backend and cross-car-OEM ones on the vehicle and modem side. Interfaces proven to be beneficial can be proposed for standardization. Appropriate inputs for different prediction algorithms will be evaluated in the context of the trials conducted in WP4.

For a special subtopic of prediction allowing to deliver large data volumes at reasonable monetary price, the Background Data Transfer (BDT) functionality was identified as a candidate to support a selected user story of the High-definition (HD) Mapping use case. The currently specified interface might need to be improved or replaced because it is intended for communication (e.g., software update) with usually geographically static Internet of Things (IoT) devices. Also, here QoS prediction is needed to identify the best times and/or places to download HD map updates.

3GPP specifications only allow QoS management within the RAN and core domain. End-to-end performance guarantees are difficult or even impossible to fulfil with one end of the communication in the public Internet. MEC enables operators to deploy backend applications within their domain. This often requires the so-called Local Breakout through additional
gateways$^2$. The capabilities of the 4G EPC to dynamically select and especially switch the gateway to reach the closest or otherwise best suitable MEC host are very limited. **The 5G Core adds capabilities for more dynamic and seamless gateway selection.** EPC and 5G Core are currently not capable to provide this in a cross-MNO environment as experienced across country borders, but also within the same country where usually multiple MNOs are present. Within 5GCroCo, solutions will be studied for addressing these aspects.

Handover from one MNO to another one across country borders is technically feasible but rarely enabled and the required links for the interfaces across MNOs are not present. **The project will demonstrate the benefit of such interfaces for enabling and improving handover between different MNOs.** Today, roaming is usually realized with Home Routed Roaming using the packet gateway in the home network. Particularly in context of MEC, it is preferable to use gateways in the visited network. As of today, no solution allowing service continuity across country borders (cross-MNO handover) and using a gateway in the visited network is specified. **We will evaluate if and how the capabilities of the 5G Core for more dynamic and seamless gateway selection can be applied across MNOs.** Technology baselines and planned improvements in 5GCroCo are summarized in Table 1-1.

<table>
<thead>
<tr>
<th>Technology</th>
<th>5G eMBB Baseline</th>
<th>5GCroCo Improvement(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5G RAN</td>
<td>RAN is essential for end-to-end QoS but extensive experience on parametrization exists only for (e)MBB, especially for Voice over LTE (VoLTE)</td>
<td>Find suitable RAN parameter$^3$ ranges providing the required QoS for the use cases within a network used for many different services</td>
</tr>
<tr>
<td>Network Slicing</td>
<td>GSMA defines this in a broader sense [9] while 3GPP standardized it for the 5G Core to provide virtual subnetworks [7] and, independent of this, also in context of network management and orchestration [10]. It is commonly used as synonym for QoS mechanisms. Those</td>
<td>See &quot;end-to-end QoS&quot; and &quot;MANO and SDN - cross-border / -MNO&quot;</td>
</tr>
</tbody>
</table>

$^2$ Before introducing the correct technical terms used in different 3GPP generations, and also later in the document where precise technical terms are not needed, the term “gateways” will be used to describe the nodes where the mobile radio network is interconnected to the public Internet or other networks.

$^3$ For example, Acknowledged vs. Unacknowledged Radio Link Control mode, Hybrid Automatic Repeat Request retransmission limit, link adaptation algorithm parameters, etc.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-to-end QoS</td>
<td>Dedicated bearers are supported but not widely commercially used except for VoLTE. Lack of experience with parametrization and requirement description. No “true end-to-end” since only possible between gateway and User Equipment (UE), not to servers on the public Internet.</td>
<td>GSTs provide common templates to describe use case requirements. MNOs and vendors obtain the necessary experience to map them to configurations that might be network deployment/topology dependent, but common best practice likely exists. 3GPP has defined a default SST value for V2X Network Slices. It can serve as common ground for MNOs to define slices offering same or otherwise acceptable QoS for certain services in different networks, even when roaming. Evaluate the potential of MEC to enable true end-to-end QoS beyond what is possible with servers on the public Internet where the MNO has no control over the full end-to-end path (see “Local Breakout for MEC” below).</td>
</tr>
<tr>
<td>Predictive QoS</td>
<td>Not present</td>
<td>Evaluate, and if needed, improve interfaces from related 3GPP study item [11] allowing to exploit QoS prediction, also across different MNOs. Identify further improvements on the network and the device side that will be implemented with proprietary interfaces in 5GCroCo but will later be considered for standardization.</td>
</tr>
<tr>
<td>Background Data Transfer</td>
<td>Standardized but very static solution. The application proposes the time window it provides the network with the</td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>Status</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>can tolerate for communication and the</td>
<td>information it needs to</td>
<td>propose the best possible place or time for a data transmission, e.g., an underutilized cell.</td>
</tr>
<tr>
<td>network selects a slot within this window.</td>
<td></td>
<td>This is intended for stationary IoT devices.</td>
</tr>
<tr>
<td>This is intended for stationary IoT devices.</td>
<td></td>
<td>Evaluate solutions within the currently standardized interface and potentially propose extensions to the existing interface or new ones where beneficial.</td>
</tr>
<tr>
<td>Local Breakout for MEC</td>
<td>Not commercially deployed.</td>
<td>No seamless dynamic gateway changes with 4G EPC.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluate 5G Core Session and Service Continuity (SSC) and other edge cloud support mechanisms to design solutions, potentially service specific, for service continuity when changing the gateway used to access MEC hosts (see next row).</td>
</tr>
<tr>
<td>Cross-border / -MNO handover</td>
<td>Technically feasible but</td>
<td>required links for interfaces across MNOs not in place. Service continuity only possible when home network gateway is used (see previous row)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluate the benefit of the solution as decision basis to deploy the links for cross-MNO interfaces. Determine the QoS requirements for these links.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describe further handover improvements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describe solutions for service continuity including using gateways in visited network.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design and implement an architecture and software solution for enabling the cross-border / -MNO orchestration of virtualized automotive services within tailored Network Slices.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluate the proposed solution in a trial, collecting relevant insights towards commercialization.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identify gaps with respect to current mainstream.</td>
</tr>
<tr>
<td>operations to support the provision of network services across multiple administrative domains.</td>
<td>standardization bodies (e.g., ETSI NFV)</td>
<td></td>
</tr>
</tbody>
</table>
2 Baseline Network Architecture

Deliverable D3.1 [3] defines the application architectures in the vehicles and backend while considering the network as a black box. Interfaces between ‘vehicles and network’ and ‘backend and network’ were not further specified. This section provides a first refinement on the network architecture considered to connect the vehicles with the backend. This section is limited to providing a baseline architecture according to what current 3GPP 4G networks deploy for MBB services and what 3GPP 5G networks, that are currently being rolled out, provide for eMBB. Section 4 will then describe which features beyond (e)MBB already exist, either already specified or at least researched, and which further solutions are needed to evolve the baseline architecture described in this section to support cross-border CCAM.

This, and all following sections, are divided into the subtopics “general connectivity”, “predictive and end-to-end QoS”, “MEC”, and “precise positioning”.

The architecture figures in this and other sections follow a common set of drawing elements defined in Annex C.

2.1 General Connectivity

In the following, first the 3GPP 4G and 5G user plane architectures are described. After that, the baseline control plane architecture is explained, followed by the interfaces that can be used towards the application in the vehicle and the backend. Furthermore, the baseline about today’s service experience when crossing a country border is provided due to its importance for the project.
Figure 2-1 shows the user plane of a 4G mobile radio network including the cardinalities among nodes. The evolved Node B (eNB) is nowadays typically split into a Remote Radio Unit (RRU) located at the base station site with the antennas and a Base Band Unit (BBU) that can be up to multiple kilometres away and drive multiple RRUs. RRUs and BBUs are currently realized with dedicated physical hardware and, therefore, considered to be PNFs. The eNB is connected to a Serving Gateway (S-GW). Each S-GW can be responsible for multiple eNBs. The S-GW is connected to a Packet Data Network (PDN) Gateway (P-GW) connecting it to a PDN, usually the

---

4 This can change in the future
Internet. This is done over the SGi-interface. There can be multiple P-GWs in a network. Typically, there is more than just one mobile radio network in an area. Communication between the networks is only possible through the PDN over the SGi-interfaces of the networks.

![Figure 2-2: Simplified Depiction of 4G Mobile Radio Network User Plane without Cardinalities](image)

Figure 2-2 provides a simplified depiction of a 4G mobile radio network where the previously described cardinalities are not shown. This representation depicts the path the data of one user served by the shown eNB takes through the network. In the following, if not stated otherwise, this figure will serve as baseline; however, the statements about cardinalities of Network Functions (NFs) remain true.

The same architecture is used for non-standalone 5G New Radio where RRUs and BBUs form next generation Node Bs (gNBs). Besides gNBs, eNBs still must be present to provide the control plane.

![Figure 2-3: 5G Standalone Mobile Radio Network User Plane Architecture](image)

Figure 2-3 shows the user plane of a standalone 5G New Radio mobile radio network with 5G RAN and Core. P-GW and S-GW are replaced by one or more User Plane Functions (UPFs). The UPFs can be chained and the one providing access to a PDN is called Protocol Data Unit (PDU) Session Anchor. The equivalent of the 4G SGi-interface is the N6-interface.
2.1.1 Interfaces Toward Application and Backend

Figure 2-4 shows the use case independent high-level architecture defined in Deliverable D3.1 [3] that is considered in 5GCroCo. It consists of the vehicle, the backend, and the network between them.

![Vehicle - Network - Backend](image)

**Figure 2-4: 5GCroCo High-level Architecture from Deliverable 3.1 [3]**

In its turn, Figure 2-5 shows how the high-level architecture can be mapped into the 3GPP 4G Evolved Packet System (EPS) architecture with Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and the EPC. In the following, the commonly used shorter terms RAN and Core (Network)\(^5\) will be used. Initially, user plane aspects will be described, followed\(^6\) by the control plane NFs in the Core. The purple and yellow components are for network orchestration and control and will be explained in Section 2.1.3.

The RAN includes the eNB and User Equipment (UE) in the vehicle. The UE is part of the Communication Control Unit (CCU) defined in Deliverable D3.1 [3]. The whole CCU, all other components in the vehicle, and the networks and other links between them, form the vehicle domain. The network domain consists of the eNB and the Core Network together with Operation Support System (OSS) and Business Support System (BSS). The Uu-interface between eNB and UE is, therefore, the only interface between the vehicle and network domains. It is used for user- and control plane communication.

---

\(^5\) In some sentences readability is improved by including the word “network”, while in others it will be omitted.

\(^6\) Also, the acronyms in the figure will be expanded and explained then
Figure 2-5: Mapping of High-Level Architecture to 3GPP 4G Evolved Packet System Architecture [14]

Figure 2-6 shows the evolution of the 4G EPS architecture towards non-standalone 5G New Radio [15]. The eNB is still present for control plane communication but the gNB is used for user plane communication. The BBUs\(^7\) of the eNB and gNB exchange control plane information over the X2-interface. User plane data is also exchanged over the X2-interface as long as the path from the gNB towards the S-GW is not established.

\(^7\) It is expected that both logical BBUs are physically in the same component making X2 a purely internal interface.
While making a difference in achieved performance, it usually makes no difference regarding architecture and interfaces whether 4G LTE RAN or 5G New Radio RAN is used. Figure 2-7 is thus introduced to explain the backend and its interfaces towards the network, since they are the same for 4G EPC and non-standalone 5G New Radio.

Figure 2-7: Mapping of High-Level Architecture to 3GPP Architecture. The RAN generation is not mentioned, so the figure is equally applicable to 4G and 5G RAN.
User plane communication between the backend and the network is realized over the SGi-interface. On the network domain, the interface terminates at the P-GW, while the part at the backend side is called application server (AS).

A clear distinction is made between ASs hosted on the public Internet and the ones that are deployed within\(^8\) the MNO domain, defined as being deployed on a MEC host. This is further explained in Section 4.3.

Besides SGi for the user plane, the Receiver (Rx)-interface is also shown allowing control plane communication between an Application Function (AF) and the Policy and Charging Rules Function (PCRF). It is, among others, used to request dedicated bearers for QoS, as further explained in Section 4.2.1. Specifications do not cover deployment aspects, so the AF can be in the public Internet or in the MNO domain.

The EPC specification was extended introducing further specialized interfaces for certain domains and functionalities like IoT or Multimedia Broadcast Multicast Service (MBMS). The 4G LTE-V2X Architecture [16] does not introduce new interfaces between network and backend and solely uses the SGi-interface. The Service Capability Exposure Function (SCEF) [17] is an extension to the EPC. It adds the T8-interface shown in Figure 2-8 and further described in Section 4.2.3 for BDT. The T8-interface also provides QoS related services and, therefore, serves as alternative to the Rx-interface.

\(^8\) It is especially important that the MNO controls the servers hosting the application and the transport network providing the connectivity. Many different models can exist where other entities manage or even own parts of the servers or transport network. For the solutions scoped in 5GCroCo, it is enough to assume that the MNO itself or through other parties is capable to manage and configure the servers and transport network between gateways (P-GW) and MEC instances.
In the case of standalone 5G New Radio, the EPC is replaced by the 5G Core, as shown in Figure 2-9. The role of the SGi interface is provided by the N6-interface between UPF (see Section 2.1) and the application server. Backend entities using control plane interfaces of the network are now consistently always called AF. They can use the N5-interface for QoS Flow related services provided by the Policy Control Function (PCF) or the N33-interface also for QoS Flow and all other services provided by the Network Exposure Function (NEF) [18], for example BDT [19] as described in Section 4.2.3. Means to support QoS prediction over N5- and/or N33-interface are being studied [11] and are further described in Section 4.2.2.1.
In the EPS, the Uu-interface was also used for control plane communication between UE and Mobility Management Entity (MME) for the Non-Access Stratum (NAS) set of protocols. With standalone 5G New Radio, a clearer definition has been done by introducing the N1 interface for this.

NEF and PCF were already introduced above as functions taking over the roles of EPC functions NEF and PCRF, respectively. Further functions are present in the 5G Core [7]:

- **Session Management Function (SMF):** Manages the path between Session Anchor UPF(s) and UE. Equal to the control plane parts of S-GW and P-GW in EPC.
- **Access Management Function (AMF):** Manages network access and mobility similar to the MME in EPC.
- **Unified Data Management (UDM):** Provides similar functionality to the Home Subscriber Server (HSS) of the EPC but also many other services require persistent or temporary data storage provided by the UDM.
- **Authentication Server Function (AUSF):** Handles security, as mostly done by the HSS in the EPC.
- **Network Slice Selection Function (NSSF):** Supports selecting one or more Network Slices a UE should attach to. It has no equivalent in the EPC.
Figure 2-10 and Figure 2-11 summarize the interface between vehicles and the network and backend and the network for non-standalone 5G New Radio (with 4G EPS) and standalone 5G New Radio, respectively.

**Figure 2-10: Interfaces between Vehicles and Network / Backend and Network in 4G EPS and Non-standalone 5G New Radio**

Besides those, further proprietary\(^9\) interfaces between backend and network can exist. The AFs using them on the network side would usually have to be deployed in the MNO domain for security reasons. As of today, such interfaces do not exist between UEs and the network, and therefore, not between vehicle and network domain. Communication between vehicle and backend must be done through the user plane of the network in case involvement from vehicle side is required, as shown in Figure 2-12.

---

\(^9\) In the sense that there is no 3GPP specification defining them. They could be defined elsewhere but remain proprietary from 3GPP point of view.
2.1.2 Network Service Continuity at Country Borders

Today’s networks usually do not allow cross-border / -MNO handovers. Experience shows that when leaving a country, a UE will stay connected to the network of the previous country (home\textsuperscript{10} network) until it is so far away that it loses synchronization to the last serving cell in the home network. For many seconds, or even minutes, radio link quality can be very low making even simple MBB services and voice calls infeasible. After loss of synchronization, the UE will perform a scan and attach to a new network in the new country (visited network). It will then establish a new PDN connection usually resulting in a new Internet Protocol (IP) address. Being served by a different network than the home one is called “roaming” [14].

Analyses performed indicate that this procedure is time consuming and introduces delays in the range of seconds or even larger [20]. The delay considerations in [20], also take timers for error handling procedures into account that in previous generations networks may even reach the magnitude of seconds. The analysis in [20] shows that the delay to attach to a visiting operator is up to 100 seconds because of the sequential process and the context transfer procedure, whereas towards the home network it can be up to 9 seconds.

\textsuperscript{10} The special case of handover between two visited networks can of course also occur but is, where not stated differently, equally covered.
Figure 2-13: Roaming Architecture with Home Routing

Figure 2-13 shows the roaming architecture typically used today. When the UE re-establishes a connection to the visited network (bottom), it will usually still use a P-GW in its home network. This could be the same P-GW as before the roaming process, but the figure shows the case where it is a different P-GW in the same home MNO domain. The S-GW in the visited network and the P-GW in the home network communicate over the S8-interface. The MME in the visited network and the HSS in the home network communicate over the S6a interface. Interfaces S8 and S6a are realized over an IP Exchange (IPX) network. This could be either the public Internet or any other wide area network, usually realized on the same infrastructure as the public Internet. What kind of IPX network is used is up to the choice of the MNOs.

2.1.3 Network Orchestration and Control

As stated in Section 2.1, today’s mobile radio core networks are usually virtualized, and this kind of deployment is therefore considered as baseline. For this, first Management and Orchestration (MANO) of NFV Network Services (NSs) as the general concept will be explained and available solutions deployed in today’s networks will be presented. Furthermore, SDN as concept to dynamically interconnect the PNFs and VNFs will be discussed together with available solution deployed and used in mobile radio networks.

2.1.3.1 MANO

ETSI Industry Specification Group (ISG) for NFV has released a set of specifications for virtualized networks. In specification [21], ETSI specifies a MANO framework targeting the
provisioning of NFV NSs and VNFs as well as related Lifecycle Management (LCM) operations, including the configuration of the VNFs and the hosting Network Function Virtualization Infrastructure (NFVI). Figure 2-14 depicts the NFV MANO framework as defined by ETSI, in the following of this section we provide also a brief description of the most relevant represented functional entities.

**Figure 2-14: ETSI NFV MANO Framework Functional Architecture**

**NFV Orchestrator (NFVO):** this component is responsible for the orchestration of NFVI resources across multiple VIMs. It’s also responsible for the LCM of NFV NSs.

**VNF Manager (VNFM):** this component handles the LCM of VNF instances, it can be assigned to a single VNF instance or multiple VNF instances that can be of different types, indeed VNFM are assumed to be generic common functions applicable to any type of VNF.

**Virtualized Infrastructure Manager (VIM):** this component is responsible for controlling and managing the NFVI compute, storage and network resources, usually within one operator's infrastructure domain. A VIM can be specialized with respect to the type of offered resources or may be capable of managing multiple types of NFVI resources.

**NFVI:** it includes all the hardware (e.g., compute, storage, and networking) and software (e.g., hypervisors) components that together enables the deployment of VNFs[1].

Furthermore, 3GPP Technical Specification (TS) 28.500 [22] “[…] clarifies the relationship between 3GPP management architecture and ETSI ISG NFV Management and Orchestration architecture”, stating that VNF instances can be considered as virtualized Network Elements (NEs) as defined by 3GPP. Then, a VNF instances may comprise several NFs in the form of Virtual Network Function Components (VNFCs). Regarding Network Slicing, 3GPP TS 28.541 [23] specifies the Network Slice Network Resource Model (NRM), where a Network Slice is composed of one or several Network Slice Subnets (NSSs) that, as depicted in Figure 2-15, may correspond to NFV NSs as defined in [24]. Finally, according to 3GPP TS 28.531 [25], the Network

[1] Will be called “hardware” or “infrastructure” or just “host(s)” in this document
Slice Management Function (NSMF) plays the role of Network Slice management Service Provider (SP), providing the management functionalities for the provisioning of Network Slices that may include also NFV NSs.

![Diagram](image-url)

**Figure 2-15: 3GPP Network Slice NRM Class Diagram**

In the following, available MANO solutions are described, along with their relation to the ETSI architecture and major open-source solutions.

2.1.3.1.1 OSM for Network Function Virtualization
Open Source MANO (OSM) is an ETSI-hosted initiative to develop an open-source NFV MANO software stack aligned with ETSI NFV [26], [27]. It has support from multiple MNOs and telco-vendors. Latest releases (starting from release FIVE up to release EIGHT) provide novel features that include Network Slicing, support for PNFs, policy and monitoring, support for VNF configuration, and support for network edge functions.

2.1.3.1.2 SONATA NFV Service Platform
SONATA NFV Service Platform is a MANO framework created by Horizon 2020 (H2020) SONATA project that allows the creation of a versatile and modular ecosystem that serves to service developers and testers, telecom operators or vertical industries, managing the full lifecycle of network services [28]. Its latest release 5.1 is from January 2020.

2.1.3.1.3 ERI’s Network Functions Virtualization Infrastructure Solutions
ERI’s virtualization environment [29] is based on OpenStack. It follows the architecture defined in ETSI NFV and OSM. The “Cloud Manager” serves as NFVO. It supports Open Virtualization Format (OVF) and Heat Orchestration Templates (HOTs) to define VNFs, Virtual Machines (VMs) and virtual networks through SDN (see Section 2.1.3.2.2.2).

The “Cloud Execution Environment (CEE)” serves as Virtualized Infrastructure Manager (VIM). The host operating system of CEE is Ubuntu Server providing QEMU and Kernel-based Virtual Machine (KVM) capability to run virtual machines hosting VNFs or applications.

The OpenStack based solution was originally used for VNFs and VMs hosting end user applications. As an alternative for application VMs, Ericsson Cloud Container Distribution (ECCD)
based on Kubernetes is available to manage and orchestrate containerized VMs. ECCD can also be used to manage and orchestrate VNFs [30] under the constraint that the respective VNF is available in a containerized deployment version.

2.1.3.2 Software Defined Networking (SDN)
The key essence of SDN is separating control plane functions of a data network from the user plane ones. The user plane function of a network is defined as packet forwarding being the decision of how an incoming packet at a network node should be forwarded. All other tasks, especially provisioning the information required to make this forwarding decision, is considered control plane. The concept particularly gained interest in context of the OpenFlow [31] set of specifications promoted by the Open Network Foundation (ONF) but should not be limited to that within this document and the 5GCroCo project. For this, the different SDN solutions provided by different project partners, are presented in the following sections. Moreover, SDN is used together with NFV techniques to manage and orchestrate virtual networks on top of the physical network infrastructure, e.g., for establishing the network connectivity between the different VNFs and PNFs composing an NFV NS.

2.1.3.2.1 SDN-enabled Switches
Network elements such as L2 switches have been enabled by SDN in order to program their data flows. A commonly used open-source software switch is Open vSwitch (OVS) [32]. The main purpose of OVS is to provide a switching stack for hardware virtualization environments, while supporting multiple protocols and SDN standards.

2.1.3.2.2 SDN Controllers
Considering the networking aspects, data connectivity between data plane functional entities will be enabled using a control plane based on SDN. In this context\(^\text{12}\), a control plane is a system and a set of functions especially dedicated to provisioning of connectivity services, with higher abstractions, refined functional architectures, better addressing the (evolving) requirements. A control plane is specially designed for and focusing on the configuration and specifics of the forwarding and switching operations of the network elements, ideally with standard interfaces operating across domains ensuring vendor inter-operability. In particular, an SDN controller handles the configuration of switches/routers in the data plane via southbound Application Programming Interfaces (APIs) (e.g., compliant with OpenFlow or NETCONF protocols), while the northbound APIs offers applications-related procedures to deploy intelligent networks.

\(^{12}\) There can be relations to the 3GPP control plane (see Section 2.1) but in context of SDN the role of the control plane is clearly defined and has a narrower scope than for 3GPP or other International Organization for Standardization (ISO) / Open Systems Interconnection (OSI) reference model [109] related specifications.
SDN will be used in selected 5GCroCo scenarios to configure the data plane forwarding functions across multiple layers (Layer0/1\textsuperscript{13}, Layer 2, and Layer 3) and technologies.

2.1.3.2.2.1 Open-Source SDN Controllers
As stated, the SDN Controller is a key element in the SDN architecture. Amongst the several projects, two main ones will be further described:

Open Network Operating System (ONOS) [33] is an open-source project backed by an expanding community of developers and users. It provides the control plane for an SDN-based network, managing network components, such as switches, and running software applications. ONOS provides functionality of operating systems, including Application Programming Interfaces (APIs) and abstractions, resource allocation, and permissions management, as well as user-facing software such as a Command Line Interface (CLI) and a Graphical User Interface (GUI). It also provides system applications managing the entire network rather than a single device. This can dramatically simplify management, configuration, and deployment of new software, hardware and services. The ONOS platform and applications act as an extensible, modular, distributed SDN controller, providing a useful and usable platform for software programs designed for a particular application or use case. It can run as a distributed system across multiple servers, providing fault tolerance in the face of server failure and potentially supporting live/rolling upgrades.

Open Daylight (ODL) [34] is an open-source software SDN controller and framework that allows network control and automation. The ODL Project is hosted by the Linux Foundation and includes contributions of more than 35 vendor solutions and applications. It has been used as the foundation for commercial solutions. It is also included, contributed and distributed in other open-source frameworks, including Open Networking Automation Platform (ONAP), OpenStack, and Open Platform for NFV (OPNFV).

2.1.3.2.2.2 ERI's SDN Solutions
ERI provides SDN through the “Cloud SDN Switch (CSS)” and “Cloud SDN Controller (CSC)” each combining a set of different solutions. Interaction with Ericsson Network Functions Virtualization Infrastructure [29] (see Section 2.1.3.1.1) is done through integration of the OpenStack component “Neutron” allowing to configure virtual networks within a host but also ones within physical switching (Layer 2) and routing (Layer 3) hardware. In case of Layer 2, Virtual Local Area Network (VLAN) tagging [35] is used. CSC supports different protocols including OpenFlow [31] to be able to control not only ERI’s software-based CSS but also other software and hardware switches and routers.

2.2 Predictive and End-to-End QoS
QoS differentiation for packet-switched services was available in mobile radio network generations before 4G. To the best of our knowledge, it was not commercially used because the

\textsuperscript{13} Layer 0/1 refers to physical layer Technology (e.g., Dense Wavelength Division Multiplexing (DWDM), Optical Transport Network (OTN)).
prioritization of voice call services was achieved by handling them through the circuit-switched domain providing exclusively reserved resources per call. 4G networks do not provide a circuit-switched domain and, therefore, gave momentum to QoS differentiation within the All-IP network. Using dedicated bearers to give Voice over IP (VoIP) packets priority over other ones is widely deployed and, therefore, considered as baseline. The concept of QoS prediction is not available in today’s LTE networks and can only be realized by crowdsourcing information available on the UE side.

In the following, the concept of QoS classes and methods to use them is described.

### 2.2.1 QoS Classes

QoS classes and respective bearers providing them are established across RAN and Core. Therefore, substantial parts of the QoS extensions becoming available with 5G can be considered as part of non-standalone 5G New Radio, which, according to Section 2, is the baseline architecture for 5GCroCo. This section, therefore, jointly describes 4G and 5G QoS classes. A first solution to support QoS prediction for automotive has been introduced for 5G by 3GPP, during the work for Rel. 16 [36].

3GPP specifications [7] and [37] define standardized QoS classes grouped in guaranteed- (GBR) and non-guaranteed (non-GBR) bit rate ones. The bit rate for the GBR ones can be set upon bearer setup but each class has a defined packet delay budget, packet loss rate and priority level. Each QoS class (GBR or non-GBR) is identified by a QoS Class Identifier (QCI). 5G New Radio extends this concept to 5G QoS Identifiers (5QI). It adds delay critical GBR as third kind of QoS classes. Furthermore, the priority level is considered as “default” and can be changed to allow prioritization within the same QoS class. Maximum burst data volumes are defined for the delay critical GBR classes and all GBR classes define a default averaging window over which the data rate is calculated.

Besides the standardized QCI and 5QI values, customised configurations are possible. However, end users cannot determine the QoS they obtain when requesting a bearer with non-standard QCI or 5QI.

For each QCI and 5QI, example services are provided. V2X is mentioned as example service for QCI/5QI value 3 (GBR), 75 (GBR, MBMS only) and 79 (non-GBR).
Table 2-1,
Table 2-2, and Table 2-3 list the standardized QCI and 5QI values for GBR, non-GBR, and delay critical GBR QoS classes, respectively. Fields only present in 5G New Radio are highlighted in grey. QCI value 75, marked in green, is only standardized for 4G LTE but until now not for 5G New Radio due to lack of a 5G New Radio MBMS specifications. The default priority value for the 5QI classes is ten times the one of the respective QCI classes.
<table>
<thead>
<tr>
<th>QCI/5QI Value</th>
<th>Resource Type</th>
<th>(Default)/Priority Level</th>
<th>Packet Delay Budget</th>
<th>Packet Error Rate</th>
<th>Default Maximum Data Burst Volume</th>
<th>Default Averaging Window</th>
<th>Example Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBR</td>
<td>2/20</td>
<td>100 ms</td>
<td>$10^{-2}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Conversational Voice</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4/40</td>
<td>150 ms</td>
<td>$10^{-3}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Conversational Video (Live Streaming)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3/30</td>
<td>50 ms</td>
<td>$10^{-3}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Real Time Gaming, V2X messages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Electricity distribution – medium voltage, Process automation - monitoring</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5/50</td>
<td>300 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Non-Conversational Video (Buffered Streaming)</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td>0.7/7</td>
<td>75 ms</td>
<td>$10^{-2}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Mission Critical user plane Push-To-Talk voice (e.g., Mission Critical Push to Talk (MCPTT))</td>
</tr>
<tr>
<td>66</td>
<td></td>
<td>2/20</td>
<td>100 ms</td>
<td>$10^{-2}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Non-Mission-Critical user plane Push-To-Talk voice</td>
</tr>
<tr>
<td>67</td>
<td></td>
<td>-/15</td>
<td>100 ms</td>
<td>$10^{-3}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>Mission Critical Video user plane</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>2.5/-</td>
<td>50 ms</td>
<td>$10^{-2}$</td>
<td>N/A</td>
<td>2000 ms</td>
<td>V2X messages</td>
</tr>
<tr>
<td>71</td>
<td></td>
<td>-/56</td>
<td>150 ms</td>
<td>10-6</td>
<td>N/A</td>
<td>2000 ms</td>
<td>&quot;Live&quot; Uplink Streaming</td>
</tr>
<tr>
<td>72</td>
<td></td>
<td>-/56</td>
<td>300 ms</td>
<td>10-4</td>
<td>N/A</td>
<td>2000 ms</td>
<td>&quot;Live&quot; Uplink Streaming</td>
</tr>
<tr>
<td>73</td>
<td></td>
<td>-/56</td>
<td>300 ms</td>
<td>10-8</td>
<td>N/A</td>
<td>2000 ms</td>
<td>&quot;Live&quot; Uplink Streaming</td>
</tr>
<tr>
<td>74</td>
<td></td>
<td>-/56</td>
<td>500 ms</td>
<td>10-8</td>
<td>N/A</td>
<td>2000 ms</td>
<td>&quot;Live&quot; Uplink Streaming</td>
</tr>
<tr>
<td>76</td>
<td></td>
<td>-/56</td>
<td>500 ms</td>
<td>10-4</td>
<td>N/A</td>
<td>2000 ms</td>
<td>&quot;Live&quot; Uplink Streaming</td>
</tr>
</tbody>
</table>
### Table 2-2: Standardized QCI/5QI Values for non-GBR QoS Classes [7], [37]

<table>
<thead>
<tr>
<th>QCI/5QI Value</th>
<th>Resource Type</th>
<th>(Default) Priority Level</th>
<th>Packet Delay Budget</th>
<th>Packet Error Rate</th>
<th>Default Maximum Data Burst Volume</th>
<th>Default Averaging Window</th>
<th>Example Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Non-GBR</td>
<td>1/10</td>
<td>100 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Internet Multimedia Subsystem (IMS) Signaling</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>6/60</td>
<td>300 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>7/70</td>
<td>100 ms</td>
<td>$10^{-3}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Voice, Video (Live Streaming) Interactive Gaming</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>8/80</td>
<td>300 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>9/90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Video, etc.</td>
</tr>
<tr>
<td>69</td>
<td></td>
<td>0.5/5</td>
<td>60 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Mission Critical delay sensitive signaling (e.g., MCPTT signaling)</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>5.5/55</td>
<td>200 ms</td>
<td>$10^{-6}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Mission Critical Data (e.g., example services are the same as QCI 6/8/9)</td>
</tr>
<tr>
<td>79</td>
<td></td>
<td>6.5/65</td>
<td>50 ms</td>
<td>$10^{-2}$</td>
<td>N/A</td>
<td>N/A</td>
<td>V2X messages</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>-6/68</td>
<td>10 ms</td>
<td>$10^{-8}$</td>
<td>N/A</td>
<td>N/A</td>
<td>Low Latency eMBB applications Augmented Reality</td>
</tr>
</tbody>
</table>

### Table 2-3: Standardized 5QI Values for Delay Critical GBR QoS Classes [7]

<table>
<thead>
<tr>
<th>5QI Value</th>
<th>Resource Type</th>
<th>Default Priority Level</th>
<th>Packet Delay Budget</th>
<th>Packet Error Rate</th>
<th>Default Maximum Data Burst Volume</th>
<th>Default Averaging Window</th>
<th>Example Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>Delay Critical GBR</td>
<td>19</td>
<td>10 ms</td>
<td>$10^{-4}$</td>
<td>255 bytes</td>
<td>2000 ms</td>
<td>Discrete Automation</td>
</tr>
<tr>
<td>83</td>
<td></td>
<td>22</td>
<td>10 ms</td>
<td>$10^{-4}$</td>
<td>1354 bytes</td>
<td>2000 ms</td>
<td>Discrete Automation</td>
</tr>
<tr>
<td>84</td>
<td></td>
<td>24</td>
<td>30 ms</td>
<td>$10^{-4}$</td>
<td>1354 bytes</td>
<td>2000 ms</td>
<td>Intelligent transport systems</td>
</tr>
<tr>
<td>85</td>
<td></td>
<td>21</td>
<td>5 ms</td>
<td>$10^{-5}$</td>
<td>255 bytes</td>
<td>2000 ms</td>
<td>Electricity Distribution- high voltage</td>
</tr>
</tbody>
</table>

#### 2.2.2 Session Management

Session management is used to setup, change, and release dedicated bearers providing a certain QoS class as described above in Section 2.2.1. The Rx-interface [38] (see Section 2.1.1) allows an AF in the backend to influence this and, more in general, to interact with the core network via the PCRF. A summary of some of the procedures carried over Rx is provided in the following.
The AF can provide the minimum supported, minimum desired, maximum supported, and/or maximum desired bandwidth for uplink and downlink [39]. These parameters originate from voice, audio and video applications where often source rate adaptation can be used.

When a bearer is established, the AF will receive an indication of successful resource allocation. Upon normal termination, a release indication is provided. Abnormal termination, usually caused by insufficient radio resources, is indicated by a loss indication. A recovery indication is provided to the AF when the terminated bearer is activated again.

The sessions for which the requests and notifications apply are usually\textsuperscript{14} identified by a filter consisting of local (UE side) IP address, remote (AS) IP address, source port, destination port, and transport layer protocol (User Datagram Protocol (UDP) or Transmission Control Protocol (TCP)) as provided in the Protocol Number (IP Version 4 (IPv4)) or Next Header (IP Version 6 (IPv6)) field of the used IP protocol version. Wildcards and ranges for each entry are possible [40]. Below, the allowed combinations are described. For each combination a minimum of one field must contain information and the rest can be omitted and is then considered as wildcard.

**IPv4 and IPv6:** Remote Address and Subnet Mask, Protocol Number (IPv4) / Next Header (IPv6), Local Address and Subnet Mask, Local Port Range, Remote Port Range, Type of Service (TOS) (IPv4) / Traffic Class (IPv6).

**IPv6:** Remote Address and Subnet Mask, Local Address and Subnet Mask, TOS (IPv4) / Traffic Class (IPv6), Flow Label (IPv6).

On the network side, the P-GW is responsible for mapping incoming downlink traffic to the right bearer according to the filter information stored in a downlink Traffic Flow Template (TFT). In the uplink, the same is done in the UE based on an uplink TFT.

The Rx-interface can be then used to establish, modify and release sessions on demand. Furthermore, the AF gets notified when a session is dropped due to insufficient resources to provide the requested QoS. This can happen when higher prioritized sessions compete for the same resource, especially in RAN. These higher prioritized sessions can be newly established or handed over from other cells. In addition, the own session could be intended for handover to another cell since the vehicle moves, but the target cell could have insufficient resources to accept it.

The Rx-interface is currently used to establish dedicated bearers for VoLTE when a call is started and releasing them after hanging up.

A detailed list of procedures carried between AF and PCRF can be found in [38].

### 2.3 MEC

Within the scope of 5GCroCo, the term MEC describes application server hosting capabilities within the domain controlled by the MNO (see Section 2.1.1). Some MNOs already today host

\textsuperscript{14} Other parameters, e.g., Security Protocol Index (SPI) for IPSec traffic, are also possible in some context.
application servers to provide services to their customers. Furthermore, Content Delivery Networks (CDNs) exist bringing services close to the peering points of MNOs but remain within public Internet domain.

In 2020 first MNOs across the world started MEC service deployment or are planning the deployment. A study conducted in Q1 2020 [41] revealed that 9 out of 30 interviewed MNOs started deployment (first movers) and 17 are planning it (followers).

2.4 Precise Positioning

Today's vehicles usually combine information from Global Navigation Satellite Systems (GNSSs) and their inertial system for global positioning. Different sensors like radars, cameras, and lidars are then used for relative positioning of objects. GNSSs and inertial systems are described in the following.

2.4.1 GNSS

Vehicles are usually equipped with a GNSS receiver for tracking and positioning purposes. The usual accuracy achieved with GNSS systems is of few meters and depends on the features of the receiver. Furthermore, as accuracy is linked to the number of satellites that can be tracked by the receiver, positioning might be specially affected in urban areas, e.g., due to high buildings, narrow streets, glass buildings, or low visibility of the sky in general. This could lead to errors of several meters or even positioning service interruption.

2.4.2 Inertial Systems

The information provided by gyroscopes, accelerometers, and magnetometers can be used to infer the direction of a vehicle or to estimate its speed. This data can be employed to correct the error from GNSS systems or to estimate the absolute position of a vehicle during GNSS service interruption. To do this, the system needs to extrapolate starting from earlier reliable position information.
3 Challenges and Use Case Requirements

This section describes the requirements of the three use cases for the different technical areas and requirements common for all 5GCroCo use cases and further ones. Where possible, user stories from Deliverable 2.1 [2] are referenced.

3.1 General Connectivity

Use case agnostic NFV MANO and SDN requirements based on some challenges that telecommunication companies experience with these technologies are discussed in this section. Furthermore, requirements for service continuity across country borders are described per use case.

3.1.1 Network Service Continuity at Country Borders

The 5GCroCo use cases have different requirements for service continuity resulting from different consequences in case the network is not capable to transmit data within the expected latency bounds or even not at all. Providing this continuity when crossing a country border is a challenge. It depends on the use case requirements if service continuity is a prerequisite or not. This particularly applies to increased delay and packet loss when entering an area covered by a foreign visited network instead of being handed over to the next cell of the home network, as it would happen in a non-border-crossing situation.

3.1.1.1 Tele-operated Driving

This use case requires continuous data streaming in uplink and downlink. Still, certain service degradation, as potentially happening when crossing country borders, could be tolerated for this use case if the duration of violation of QoS requirements, especially packet loss and delay, remains short. User Story 3 will test if MNO handover is possible while keeping the teleoperation operational, as described in Deliverable 2.1 [2]. Furthermore, the QoS prediction could indicate such upcoming QoS degradation resulting in appropriate countermeasures on application side. User Story 4 is intended to evaluate operation under reduced network performance conditions. The reduced uplink QoS requirements stated there might tolerate more severe and/or longer service degradation during handover than the other user stories.

3.1.1.2 HD Mapping

User Story 1 of HD Mapping considers the time required to download a HD map update. Service interruption due to crossing borders can negatively influence the time or even cause a download to fail forcing the vehicle to retry the download. This similarly applies for User Story 2 where the uplink is considered but smaller data volumes are expected. User Story 3 is a combination of both. In the worst case both involved vehicles are negatively affected when the first vehicle needs more time for uploading changes due to the border crossing and then the following vehicle needs more time for downloading the update for the same reason.
In the delay tolerant User Story 4 of the use case the network would schedule the HD map update download to either finish before crossing the border or start the download after network related procedures like handover or reattachment are completed. The later requires cross-MNO interaction since the network in the first country would have to estimate the optimal time and/or place for the download in the network of the visited country.

3.1.1.3 ACCA
The Anticipated Cooperative Collision Avoidance (ACCA) user stories consider the time required to collect information, identify and fuse already known hazards, and distribute warnings to all relevant vehicles to increase road safety. Service interruption due to border crossings can negatively influence the total reaction time and increase the overall detection failure rate.

The delay requirements and service availability requirements of the use case must also be met when crossing a border. It cannot be accepted having areas where vehicles cannot send or receive messages despite available network coverage in principle, especially when it comes to critical information.

MEC application servers are an essential part of the use case and it must therefore be assured that service continuity is also preserved in this context. This includes today’s case of Home Routed Roaming, meaning that all traffic breaks out in the home network and the MEC host can actually be far away but would still need to exchange data with MEC hosts in the visited network in the proximity of the vehicle. It should also work when Local Breakout Roaming is enabled and MEC hosts of the visited network are used.

3.1.2 Network Orchestration and Control
In the following, challenges and requirements for NFV MANO and SDN are described. A relation to 5GCroCo use cases is not discussed since the topics are considered valid for all three use cases, and many other use cases from automotive and other domains. Networks are managed and orchestrated, including the underlying transport networks, for a mix of simultaneously provided services.

Managing the hardware, VNFs running on it, and the configuration of the underlying transport network on a wide area, usually spanning a whole country, is a resource-intensive task for MNOs. Networks are permanently evolving; a clear example is the ongoing evolution from 4G to 5G. This process is expected to happen in life commercial networks without service interruption. Regular software updates are also common, e.g., to add or fix features. As for every complex system, further sources of unexpected failure exist. Therefore, MNOs expect from their MANO and SDN systems, usually part of their OSS, to seamlessly support network extension, software upgrades and recovery in case of failure. This should be possible with minimum effort, and preferably in an automated manner.

Furthermore, the rollout, operation, and maintenance of new kinds of domains or customer group-/segment-specific Network Slices with specialized requirements should be as simple and automated as possible, including application server deployment on MEC hosts, where required.
Especially in automotive context, MNOs should be enabled to offer same, similar, or at least well defined QoS. This requires standardized means to describe and process such requirements.

3.2 Predictive and End-to-End QoS

This section discusses in general and per use case the requirements for QoS prediction and end-to-end QoS.

3.2.1 Tele-operated Driving

Deliverable D2.1 v3 [2] provides communication related application-level requirements for each user story of the use case. They directly or indirectly correspond to network QoS requirements, as further described and summarized in Section 4.2.1.1.

If the requested QoS for the uplink video transmission and potentially other telemetry information is not met, the remote driver cannot perceive the situation and is unable to provide commands to the vehicle. In this case it is beneficial to command the vehicle into a safe state or reduce the QoS demand by certain countermeasures assuring requested QoS can be met by the network. Examples for such counter measures are change of video codec or compression, reduction of vehicle speed, safe stop, or re-routing (more information will be provided in D3.1). This all similarly applies for the downlink where insufficient QoS can lead to steering commands not being received or being received too late.

Vehicle and remote driver need time to completely execute these counter measures and therefore QoS prediction is required for the imminent future. Having information about instantaneous or past QoS is not enough. Furthermore, network and application should be able to interact about deriving QoS prediction estimations. For instance, the network might require information from the vehicle to provide the prediction, e.g., the route, while information about predicted QoS change can be provided to the vehicle, or the Vehicle Control Center (VCoC) or both. In addition, the application can provide alternative QoS requirements that the network can potentially support. Both features allow the application to maintain its session regardless of QoS degradations and/or to adapt its behaviour according to the supported QoS.

3.2.2 HD Mapping

The following only applies to User Story 4 where HD map updates are not considered delay critical. MNOs might be willing\textsuperscript{15} to charge less for the data volume produced during HD map update download if the application in the vehicle prepones or postpones the download to a time and/or place proposed by the network. The update must be completed before the vehicle enters the area for which it requests the update.

The network must determine the best time or place to start the download. Prediction algorithms can be used for that but additional information from the vehicle might be required. The application

\textsuperscript{15} This should be discussed as a business topic under WP5
must at least provide the maximum time the vehicle can wait for the information and the data volume to the network.

Besides reducing cost, predicting the throughput performance is another role prediction can play in HD Mapping User Story 4. The vehicle obtains the required information to perform HD map downloads in area of high throughput and/or to prevent them in areas of low throughput. It can also be used to adapt the size of the requested HD map tile if versions with different degrees of detail are available.

### 3.2.3 ACCA

The ACCA Use Case description in Deliverable D2.2 [1] states that two different messages priorities are required. Getting prioritization usually requires dedicated bearers and establishing them can take time.

The ACCA use case therefore requires a solution where prioritized sessions and related dedicated bearers are available when the vehicle enables the service or can be established much quicker than in today’s networks. The filters used to identify the respective messages and their priorities must reflect the backend application server deployment and configuration and be flexible since the deployment can change.

### 3.2.4 All Use Cases

Most previous sections discuss QoS requirements. This includes assuring that QoS requirements are met and possible reactions when this is not the case. A prerequisite for all this is a unified and machine readable and processible way to define requirements for use cases. Today’s ways of defining Service Level Agreements (SLAs) between customers and MNOs do not fulfil this requirement since they are limited to few commitments, e.g., regarding overall network availability.

When roaming, the visited network should also be aware and capable of understanding and fulfilling the QoS requirements or at least in the same way describe what it can fulfil.

Implementation, configuration and update effort in the CCUs to use dedicated bearers for QoS should be minimized.

### 3.3 MEC

This section provides the requirements of the use cases for MEC. The assumption is that each MNO runs its own MEC, so all requirements stated in Section 3.1.1 about service continuity on border crossings apply. Furthermore, this also applies when MEC hosts from different MNOs within the same country are used.

For all use cases applying MEC, vehicles should be served by the most suitable\(^{16}\) MEC host at a given time.

---

\(^{16}\) It e.g., depends on the use case, individual vehicle state and overall network state considering all traffic flows to and from the MEC hosts what is “most suitable”.
As already shortly discussed in Section 3.1.2, MNOs need an easy-to-use system for network orchestration and control, potentially also including MEC application server deployment. For some use cases or realizations of those, other parties than the MNO may be the ones deploying the MEC application servers. MNOs must therefore provide appropriate and secure access to respective parts of their OSS managing third-party MEC application server deployments.

### 3.3.1 Tele-operated Driving

It is assumed that the VCoC is located outside of the domain controlled by the MNO. MEC is therefore not directly required for this use case and will not be explicitly evaluated for it in this document. MEC e.g., plays a role when discussing possibilities for deployment and interfaces for QoS prediction for Tele-operated Driving (ToD), as discussed in Section 4.2.2.1.

### 3.3.2 HD Mapping

MEC application servers can provide caching to minimize the HD map update duration in downlink for User Story 1. Furthermore, they can help to faster close the loop of locally updating the HD map and downloading the updated information, as done in User Story 3. In this case there is a clear relation between geographic area and serving MEC host. Methods described for ACCA in Section 3.3.3 are applicable. Inter-MEC host data exchange could be omitted but, in this case, updates uploaded to a MEC application server on a MEC host serving a certain area will not be available to vehicles served by other MEC hosts until the update has propagated to a central server or servers and from there to further MEC hosts.

In User Story 4 the case is evaluated where HD map updates are considered not delay critical and the network can actually exploit this to decide on an optimal place or time to perform the HD map update download. In this case it plays a minor role if MEC or public Internet hosting is used.

### 3.3.3 ACCA

For ACCA there should be a clear relation between geographical areas and MEC hosts within one MNO. Vehicles must be capable to connect to the most suitable MEC host as they move and it must be assured, they are served by the MEC application server it hosts. Implementation, update, and configuration effort on the CCU to achieve this must be minimized.

Regions of Interest (ROIs) for a message might be served by more than one MEC host, so communication between MEC hosts must generally be possible, even if the MEC hosts are controlled by different MNOs in the same or different countries regardless of whether Home Routed or Local Breakout Roaming are used. Nevertheless, the situation of inter-MEC communication not being possible or no MEC facilities at all must also be considered and fallback solutions must be designed.

### 3.4 Precise Positioning

For the 5Gcroco use cases, the following requirements regarding precise positioning can be considered according to Deliverable D2.1 v3 [2].
3.4.1 Tele-operated Driving
The minimum manoeuvring range is 100 m for speeds of 15 km/h. In trajectory-based user stories, absolute positioning should have an accuracy of 50 cm.

3.4.2 HD Mapping
Updates sent to the HD map provider must be georeferenced with lane accuracy. TomTom provides HD maps at 1 m accuracy or less [42]. More accurate referencing of detected changes uploaded to the backend could help fusing information from different vehicles for a better common understanding of what has changed.

3.4.3 ACCA
Obstacles sent to the ACCA service must be georeferenced at least with lane accuracy (5 m). More accurate referencing of detected changes uploaded to the backend could help fusing information from different vehicles for a quicker common understanding what has changed.
4 State of the Art and Solutions

Section 2 provided a baseline network architecture according to what to expect from a non-standalone 5G New Radio (e)MBB service. The state of the art in standardization and research goes substantially beyond that. This section focuses on what to add to the baseline or how to customize it to fulfil the use case requirements. Most subtopics start with a general description of the state of the art before explaining how it is used to serve the use cases. Still, several research challenges exist, especially when it comes to algorithms in context of prediction and how to customize the different standalone 5G New Radio enhancements of the 5G Core for MEC support for the use cases.

For better understanding of the implications of deploying standalone 5G New Radio with a 5G Core, separate parts cover this in many sections.

Throughout this deliverable it was already required to address different stakeholders and their roles. For the sake of a clearer definition, the V2X Application Layer Reference Architecture [43] is applied in a slightly modified version. For this, parts with limited relevance for 5GCroCo are not included and the term “backend” is used instead of the term “cloud”.

Figure 4-1: 5GAA Application Layer Reference Architecture with 5GCroCo Terms and Colouring; Based on [43]

Figure 4-1 is in line with the 5GAA Application Layer Reference Architecture but applies the common colouring scheme used in 5GCroCo. It shows the three stakeholders Road Traffic Authority (RTA), OEM and SP. Deliverable D3.1 [3] provides real-world architectures for the three
5GCroCo use cases where this reference architecture is applied. It is not precluded that on the vehicle side or backend side application software is combined within one node. OEM is clearly defined as being the entity that manufactured the corresponding vehicle. RTA is used as generic term summarizing all public entities or private entities acting on behalf of public entities to deliver services in context of road infrastructure. In contrast to this, SPs summarize all other entities that are neither OEM nor RTA. The MNO is normally responsible to operate the network. It can also deliver services, but would in this case also be considered as SP. For this deliverable, especially MEC Section 4.3, only the distinction between these three stakeholders is important. Deliverable D3.1 provides more information of different combinations and compositions that are possible between them.

4.1 General Connectivity
This section describes the state of the art beyond the MBB-baseline (see Section 2.1) for service continuity at country borders, NFV MANO, SDN and solutions for the fulfilling the requirements and tackling the challenges described in Section 3.1.

4.1.1 Network Service Continuity at Country Borders: Cross-MNO Handover Optimization
The reason that connectivity is sometimes lost for minutes when changing country borders and thus changing the serving mobile network, is that the UE must connect to a new network and this procedure takes some time (see Section 2.1.2).

In an internal engagement\(^{17}\) outside of 5GCroCo scope a fictional country border was created by configuring the mobile network to emulate two different MNOs, see Figure 4-2.

Two virtual mobile core networks were deployed. The radio network was configured to support these two mobile core networks with two radio cells (Cell 2 and Cell 3 in Figure 4-2) belonging to a virtual network operator and Cell 1 to the one. These two networks were then configured for different levels of roaming to measure the connection interruption.

A baseline is defined according to Section 2.1.2 where only the HSS in the home network is consulted to obtain user subscription data when a UE attempts to connect to a visiting mobile network.

\(^{17}\) https://www.ericsson.com/en/blog/2019/5/connected-vehicle-cross-border-service-coverage
Figure 4-3 shows an improvement to the baseline roaming architecture (see Figure 2-13 in Section 2.1.2). Besides the S6a-interface between home HSS and visited MME, the MMEs are also connected over the S10-interface. As a result, same handover procedures as within the same network with MME change apply. In this case the gateway (P-GW) could remain unchanged in order to provide session continuity; however, the details of such operation have to be defined.
This is a 3GPP specification conforming solution, but in real deployments several issues can exist as further discussed in Section 5.1.1.1 and Section 5.2.1.1.

4.1.1.1 Evolution with Standalone 5G New Radio

Figure 4-4 shows the home-routed roaming architecture with 5G Core. There, Security Edge Protection Proxies (SEPPs) are used instead of directly connecting to the HSS or respective 5G Core entities of the home network.

The same principle as with non-standalone 5G New Radio (see Section 4.1.1 above) can be applied. In this case the N14 interface between AMFs in the different networks must be present.
Same as for the non-standalone 5G New Radio case with 4G EPC (see Figure 2-13 in Section 2.1.2), the home-routed case is shown. For standalone 5G New Radio it is expected that Local Breakout routing with PDU Session Anchor UPF in the visited network receives more momentum.

The current 5G Core procedure specifications [44] do not support seamless handover with N14 interface between two networks and Local Breakout. The PDU session needs to be re-established potentially causing service interruption.

To the best of our knowledge solutions for that are currently not discussed in standardization.

4.1.1.1 Possible Future Improvements

Handover for standalone 5G New Radio is described in [44] and involves the AMF, UDM and PCF as shown in Figure 4-5. In roaming case [45], interaction between the AMF in the visited network, UDM in the visited network, UDM in the home network and PCF in the home network is needed.

According to [46] the time required to perform a single attachment procedure (without considering roaming delays) is in the range of hundreds of milliseconds (~330 ms), which is unsuitable for
many V2X use cases, e.g., ToD (see Section 3.2.1). When roaming (or re-attachment), other criteria should be considered as well, like load in the gNB and the core networks, communication performance between the different operators, or roaming agreements that may prioritize some users, or some operators in favour of others.

Cross-MNO handover is discussed in the previous Section 4.1.1, aiming at keeping session continuity and minimizing interruption time. This is supported by the 3GPP specifications but to the best of our knowledge not deployed in current 4G networks since it requires interfaces across networks managed by different MNOs. According to [47] in cross-MNO handover the home and visited network have to interact in order to exchange information:

- Static information, for example, neighbour cell lists, interconnecting traffic and signalling links, etc.;
- Dynamic information, for example real-time signalling information related to target cell selection, etc.

Involved VNFs include, among others, AMFs, UDMs, PCFs, and gNBs.

In order to be able to reduce this time, particular actions can be performed in advance, as shown in Figure 4-5. These interactions relate to the in advance UE context transfer to the visited network. When the UE gets close to the border\(^\text{18}\), the UDM of the home network communicates with the remote AMF in order to perform the respective attach, according to such approach. Then the UE context is transferred to the AMF of the visited network making it available when the UE performs the respective attach when it crosses the border.

\(^{18}\) This may be determined by geographical criteria, Tracking Area Update (TAU) related approaches, etc.
Figure 4-5: Registration Procedure
4.1.2 Network Orchestration and Control

In the scope of 5GCroCo, several commercially less mature aspects of SDN are key to the use cases to be demonstrated, including the need to provision connectivity services across multiple domains\(^{19}\) and the need to control several layers. In the following, we present basic solutions and building blocks that are be adopted within 5GCroCo. Multiple solutions will be considered, either extending existing projects and open-source software or building ad-hoc solutions for the specific purposes of the 5GCroCo project objectives.

The need to orchestrate multiple technologies and domains is a key 5GCroCo requirement. Most SDN Controllers offer proprietary\(^{20}\) interfaces to applications\(^{21}\), and such SDN controllers are arranged following an approach commonly referred to as “vendor domains” or “islands”. This heterogeneity, due to having different controller interfaces in a multi-domain context, forces the use of “plugins” and it is difficult and expensive to extend e.g., with the so-called umbrella management systems, used by the operators to deploy services spanning multiple domains. As a driving motivation and clear problem statement, there is a need for a standard interface, with common models, to act as a controller Northbound Interface (NBI).

![TAPI Use Case on the Orchestration of Multiple Network Domains](image)

Figure 4-6: TAPI Use Case on the Orchestration of Multiple Network Domains

---

\(^{19}\) Technology domains such as packet and optical transport or administrative domains

\(^{20}\) Or, at best, open yet no standardized interfaces

\(^{21}\) Or, more generically, high level controllers or other functional layers referred to as Network Orchestrators
The Transport API (TAPI) [48] published by the ONF meets the main requirements to be a protocol and interface used between an orchestrator and multiple domain controllers (see Figure 4-6). A TAPI based interface offers multiple services, as shown in Figure 4-7. The key ones are the topology and connectivity service. The services are modelled in the Yang modelling language [49].

**Common Context:** The TAPI context is the shared information between a TAPI client (user) and the TAPI server (SDN controller). The model defines a TAPI domain as being able to provide services between Service Interface Points (SIPs) mainly characterized by their universally unique identifiers (UUIDs). A basic operation for a client is to retrieve the context in order to obtain the list of SIPs, so connectivity services are requested between two or more exported SIPs.

**Topology context and models:** If a given TAPI server supports topology model, it augments the TAPI shared context with a list of topologies. Each topology is composed of a list of nodes, which, in turn, have Node Edge Points (NEPs). Links connect two NEPs. The model is flexible enough to support recursive topologies and different levels of abstraction. The level of exported details is configurable by policies. A client is thus able to obtain an abstracted view of the topology and map TAPI SIPs to external NEPs.

**Connectivity context and models:** The third model augments the shared context in order to support connectivity services. The instantiation of a connectivity service relies on the instantiation of several connections (e.g., one end-to-end and internal at each TAPI node). For this, Connection End Points (CEPs) are instantiated over NEPs. Each connection involves two or more CEPs.

![Figure 4-7: TAPI High Level Functional Architecture and Interfaces.](image-url)
TAPI will be used as a key interface between network controllers associated to different domains, allowing end-to-end transport connectivity services.

4.1.2.1 Network Slice and Service Life Cycle Management
MANO and SDN together allow to (semi) automatically deploy, adapt, and tear down virtual network services. These services include the ASs and the VNFs assuring connectivity. VNFs can and typically will be reused by different services. The ASs are considered to be deployed in the MEC and/or cloud allowing end-to-end control.

The 5GCroCo orchestration framework aims at providing functionalities and mechanisms for managing end-to-end Network Slices in support of automotive services deployed across different geographical, administrative and technological domains. In such a context, different challenges have to be taken into consideration for orchestrating end-to-end 5GCroCo services and instantiating the associated slices. In particular, the end-to-end service must be decomposed into multiple service components to be instantiated in the underlying single administrative domains, where the different VNFs and ASs composing the end-to-end service chain can be either placed in a centralized public Cloud or the MEC. For instance, depending on the specific service, the service decomposition can result in a set of centralized management functions plus several distributed functions running in MEC hosts for data processing in proximity to the vehicles.

An end-to-end Service Orchestrator (SO) is proposed to address these functionalities. The SO, starting from the high-level requirements of the service, determines its decomposition in functional elements, identifying their nature (i.e., VNFs or MEC ASs) and virtual infrastructure requirements (e.g., network connectivity and required virtual computing resources). This decomposition results in the generation of an end-to-end Network Slice that is built through a Network Slice Template (NST), which refers a Network Service Descriptor (NSD) referring to a set of different VNF Descriptors (VNFDs) and MEC Application Descriptors (MEC AppDs) [50]. The SO may decide to share sub-components among different Network Slice instances and/or to deploy the service across different administrative domains using the concept of NSSs. In this case, the end-to-end NST will also include a number of nested NSSs, each of them mapped to an NSD. Finally, the SO will identify the interconnection among the functional elements and the geographical area where to deploy each component for guaranteeing a proper service coverage.

22 When strictly following 3GPP specifications, a VNF belongs to a specific Network Slice except for ones explicitly serving multiple/all slices. When looking at software implementations certain components, e.g., databases, can be reused.
The SO resides on top of a component in charge of coordinating the multi-domain service deployment of the NSSs in the target domains, with explicit indications about the geographical deployment of MEC ASs, as shown in Figure 4-8. The Multi-domain Orchestrator (MDO) coordinates the lifecycle management between the end-to-end Network Slice and its slice subnets, those provided by vMNOs’ NSMFs. In turn, the lifecycle management of the NSSs will be also reflected on the corresponding NFV Network Services instantiated at the underlaying NFVOs. Therefore, with respect to existing baseline orchestration and management tools and/or interfaces, the MDO and the underlying NSMFs/NFVOs should support new advanced functionalities such as:

- Integrated MEC AS LCM.
• Advertisement of domain specific capabilities, e.g., geographical information.

• On-boarding of heterogeneous application descriptors, e.g., NSDs, VNF packages and MEC AppD across multiple domains.

• End-to-end slice and service (including MEC ASs) instantiation and orchestration with geographical constraints.

• Automated duplication of MEC ASs to guarantee service coverage.

Following this vision, the 5GCroCo solution foresees the implementation of a hierarchical, centralized multi-domain orchestration layer composed of two main functional components: The SO and the MDO. Summarizing, the SO is responsible for the service decomposition based on the service logic and the high-level service requirements, while the MDO handles the on-boarding, advertisement and discovery of functions across the catalogues of the different domains, the coordination of the LCM of the end-to-end Network Slice and service components across multiple NSMFs and the translation of descriptors and interface messages supported by the NSMFs in each domain. These adaptation functionalities are critical to overcome the fragmentation of interfaces and information models of multi-vendor orchestrators and solve the interoperability issues mentioned in Section 5.1.1.2. APIs and Software Development Kits (SDKs) for multi-vendor integration exist. In particular, ETSI NFV SOL standards represent an attempt to specify standard, cross-vendor, packages [51] and descriptors [52] [53] formats. In the same way, ETSI defines also standard interfaces for NFV MANO components (e.g., the NFVO northbound interface specified in [54]) with the purpose of enabling interoperability among multi-vendor solutions.

The Vertical Slicer\(^{23}\) prototype [55] developed in the context of the 5G-TRANSFORMER\(^{24}\) H2020 Phase 2 project has been selected as baseline for the SO and MDO components. The Vertical Slicer already provided functionalities for translating vertical services, defined via Vertical Service Blueprints (VSB) and Descriptors (VSD), into Network Slices and NSs that are then instantiated through an underlying NFVO. In particular, the Vertical Slicer prototype is characterized by proprietary data models for VSD, VSD and NSTs, while VNFDs and NSDs are based respectively on ETSI NFV specifications [56], [57]. In the same way, the NBI, handling requests for the instantiation and management of end-to-end vertical services, is designed from scratch, while the Southbound Interface (SBI) between the Vertical Slicer and the underlying NFVO is based on ETSI NFV Interfaces and Architecture (IFA) 013 specification [58]. Within 5GCroCo the Vertical Slicer framework has been reengineered in order to fulfil the SO and MDO requirements in terms of new supported functionalities. In particular, with respect to the SO functional component, the provided extensions include the implementation of the Network Slice data-model according to 3GPP TS 28.541 v16.6.0 specification [23] as well as the implementation and integration of a Domain repository for maintaining domains’ relevant information. Moreover, enhancements to the

\(^{23}\) [https://github.com/nextworks-it/slicer](https://github.com/nextworks-it/slicer)

\(^{24}\) [http://5g-transformer.eu/](http://5g-transformer.eu/)
Translator and Arbitrator internal modules have been provided for supporting (i) the mapping between automotive vertical services into the associated Network Slices and (ii) the arbitration of their slice subnets across different domains in cross-border scenarios. Concerning the MDO functional component, the Vertical Slicer prototype have been extended with a new coordination logic for handling the lifecycle management of Network Slices in cross-domain deployments. Finally, modular and driver-based mechanism have been implemented to interact with different kinds of NSMFs and NFVOs deployed at the per-domain level, based on a unified and abstract interface exposed to the internal components of the Vertical Slicer. In particular, two software plugins have been developed and integrated for controlling underlying NSMFs and NFVOs based on OSM and SONATA.

4.2 Predictive and End-to-End QoS

The following section first discusses solutions for end-to-end QoS within the respective existing 3GPP QoS framework. After that, QoS prediction is discussed.

4.2.1 QoS Classes and Session Management

Dedicated bearers and related QoS classes allow to provide service differentiation for different kinds of information within a use case and between different services using the same network. Besides standardized QoS classes with defined delay budget and loss, further ones can be realized. Such unspecified QoS classes are not further considered since the end user cannot know what QoS to expect from a 5QI/QCI value associated with such a class, as explained in Section 2.2.1. It depends on the use case which standardized QoS classes should be used and how the respective dedicated bearers should be established and handled.

A dedicated section on how this evolves with standalone 5G New Radio is omitted since this was explained in Section 2.2.1.

4.2.1.1 Tele-operated Driving

This section describes the dedicated bearer configurations suitable for ToD. Apart from that, Section 4.2.2 evaluates the role of QoS prediction and for that it is not necessarily assumed that dedicated bearers are used. Using the default, best-effort, bearer can be enough given that QoS prediction is also available. This is to be finally answered after corresponding trials documented in Deliverable D4.3 planned for October 2021.

According to Deliverable D2.1 v3 [2], ToD is enabled on demand. Therefore, dedicated bearers should be established when needed. This should be done through the newer T8-interface provided by SCEF in case of non-standalone 5G New Radio and through the N6-interface of the NEF for standalone 5G New Radio. Using the Rx-interface with non-standalone 5G New Radio will not be pursued because it is not following the service-oriented architecture approach of 5G where webservice-like APIs are used.

In the uplink, video data is transmitted with different minimal and maximal bitrate and further requirements. Those are summarized in Table 4-1.
Table 4-1: Network Related Uplink QoS Requirements for ToD

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>direct control, low velocity</td>
<td>10</td>
<td>50</td>
<td>40</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>indirect control, low velocity</td>
<td>8</td>
<td>30</td>
<td>80</td>
<td>99</td>
</tr>
<tr>
<td>3</td>
<td>high velocity</td>
<td>10</td>
<td>30</td>
<td>40</td>
<td>99.9</td>
</tr>
<tr>
<td>4</td>
<td>Slim Uplink, indirect control, low velocity</td>
<td>10</td>
<td>50 (less when “Slim uplink” is active)</td>
<td>40</td>
<td>99</td>
</tr>
</tbody>
</table>

The difference between minimum and optimal required bitrate does not directly result from a variable source bitrate by the cameras but is intended to leave room for application adaptation in conjunction with QoS prediction. This will be discussed later in Section 4.2.2. The Application Level Latency is directly provided but the Application Level Reliability must be mapped to a network QoS parameter. Worst case is assumed where every lost packet is considered a failure to deliver the service. The Application Level Reliability therefore directly maps to packet error rate.

As long as adaptation and QoS prediction is not considered, the optimal bitrate should be requested for the dedicated bearer and it should be realized over a GBR one. Using a non-GBR one should not be precluded but will not be evaluated in detail in this report.

None of the GBR QoS classes meets the 40 ms requirement and only CQI/5QI class 3 would support the 80 ms requirement for User Story 2 but the 50 ms delay budget is close to the 40 ms required for the other user stories. From the Delay Critical GBR classes introduced with 5G, 5QI class 83 and 84 meet the requirement with 10 ms and 30 ms delay budget, respectively. The other two delay critical classes are precluded since they are limited to packet (burst) sizes of 255 bytes, which is usually not the case for video streaming. All mentioned classes fulfil the packet error rate requirement.

Table 4-2 summarizes the requirements for the downlink.

Table 4-2: Network Related Downlink QoS Requirements for ToD

<table>
<thead>
<tr>
<th>User Story</th>
<th>Name (see [59] and [1] for details)</th>
<th>Maximum required bitrate [Mbit/s]</th>
<th>Application Level Latency [ms]</th>
<th>Application Level Reliability [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>direct control, low velocity</td>
<td>0.5</td>
<td>40</td>
<td>99.9</td>
</tr>
<tr>
<td>2</td>
<td>indirect control, low velocity</td>
<td>0.3</td>
<td>80</td>
<td>99.9</td>
</tr>
</tbody>
</table>
Same as for uplink, the Application Level Reliability is directly mapped to packet error rate for the downlink.

The most demanding requirement for defining suitable QoS classes is the maximum latency. For this much lower bitrate than in uplink, it is assumed that non-GBR classes can also be considered. 5QI value 65 was ruled out for downlink due to too high packet error rate of 99%. Different than for video in uplink, a packet (burst) size of 255 bytes can also be considered large enough making all Delay Critical GBR classes suitable.

Table 4-3 summarizes the suitable CQI/5QI values for downlink and uplink. The ones almost meeting the requirements are in braces.

### Table 4-3: Suitable CQI/5QI Values for ToD

<table>
<thead>
<tr>
<th>User Story</th>
<th>Name (see [59] for details)</th>
<th>5QI/CQI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uplink GBR</td>
<td>Uplink Delay Critical GBR</td>
</tr>
<tr>
<td>1</td>
<td>direct control, low velocity</td>
<td>(3) 83, 84</td>
</tr>
<tr>
<td>2</td>
<td>indirect control, low velocity</td>
<td>3 83, 84</td>
</tr>
<tr>
<td>3</td>
<td>high velocity</td>
<td>(3) 83, 84</td>
</tr>
<tr>
<td>4</td>
<td>Slim Uplink, indirect control, low velocity</td>
<td>(3) 83, 84</td>
</tr>
</tbody>
</table>

4.2.1.2 HD Mapping

Downloads and map difference uploads are realized over bulk data transfer, where the downloads and uploads will usually be segmented into many network packets. With a normal subscription, this data transfer is done with default best effort priority meaning CQI/5QI value 9. CQI/5QI value 8 can be used to get priority over best effort services using class 9 but this would usually come at extra subscription cost.
In case of User Story 4 where minimizing transmission costs is a goal, also lower priority than CQI/5QI value 9 could be accepted. This can be achieved in a 5G system by increasing the priority level beyond the default value of 90 for CQI/5QI value 9.

4.2.1.3 ACCA

The ACCA use case applies asynchronous communication triggered e.g., according to conditions specified by the CAR 2 CAR Communication Consortium (C2C-CC) [60]. This on-demand type of communication is usually not suitable for GBR QoS classes leaving the non-GBR ones as option. GBR ones should not be completely precluded but will not be further discussed in this report. The allowed maximum Service Level Latency is 1000 ms [2] including all downlink and uplink delays and potentially also inter-MEC communication. Furthermore, processing in the MEC application server must also fit within this delay budget. QCI/5QI values 70 (200 ms), 6, and 8 (300 ms) are considered to be not appropriate as they would not leave much time for processing in the application server. 5QI value 80 (10 ms) is substantially overfulfilling the requirement. Table 4-4 summarizes the QCI/5QI values potentially appropriate for the ACCA use case.

Table 4-4: QCI/5QI Values Potentially Applicable for ACCA Use Case

<table>
<thead>
<tr>
<th>QCI/5QI Value</th>
<th>Resource Type</th>
<th>Default Priority Level</th>
<th>Packet Delay Budget</th>
<th>Packet Error Rate</th>
<th>Default Maximum Data Burst Volume</th>
<th>Default Averaging Window</th>
<th>Example Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Non-GBR</td>
<td>10</td>
<td>100 ms</td>
<td>10^{-5}</td>
<td>N/A</td>
<td>N/A</td>
<td>IMS Signalling</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>100 ms</td>
<td>10^{-3}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Voice, Video (Live Streaming)</td>
</tr>
<tr>
<td>69</td>
<td>5</td>
<td>60 ms</td>
<td>10^{-5}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)</td>
</tr>
<tr>
<td>79</td>
<td>65</td>
<td>50 ms</td>
<td>10^{-2}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>V2X messages</td>
</tr>
</tbody>
</table>

For the larger, not prioritized, ROI no dedicated bearer is required. Messages can be transmitted using the default bearer providing best effort service. Critical Information for the smaller ROIs needs to be prioritized. For selecting an appropriate QCI/5QI value it also plays a role if the service is realized over TCP or UDP to prevent bad TCP performance due to too frequent packet losses interpreted as congestion. So, for TCP, lower packet error rates are preferred. Therefore, QCI/5QI values 5 or 69 are preferred for TCP and 7 or 79 for UDP. The proposed solution is shown in Table 4-5.

Table 4-5: Possible Mappings of ROIs to QCI/5QI Values

<table>
<thead>
<tr>
<th>Transport Layer Protocol</th>
<th>QCI/5QI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>5, 69</td>
</tr>
<tr>
<td>UDP</td>
<td>7, 79</td>
</tr>
</tbody>
</table>

25 Higher values mean lower priority.
As for ToD (see Section 4.2.1.1), dedicated bearers for the respective QCI/5QI classes can be established through the backend to network interfaces T8 (non-standalone 5G New Radio) or N6/N33 (standalone 5G New Radio). This could either be done from the Traffic Management System on the public Internet or a MEC-hosted Geoservice application server. In both cases this is done only once when the vehicle starts the service.

An alternative solution is using preconfigured dedicated bearers. Those are set up when communication to or from a preconfigured server IP address, port number, and transport layer protocol (TCP or UDP) is detected through a P-GW. Each of the entries, summarized in uplink and downlink TFTs, can include wildcards and ranges (see Section 2.2.2). Permission to use this dedicated bearer is obtained through the Subscriber Identity Module (SIM) card and its respective entry in the HSS.

Radio resource schedulers are the key components of QoS. They are not subject to standardization, so different algorithms are used, and MNOs can configure them differently. MNOs try to provide defined performance often by assuring sufficient capacity for critical services preventing congestion and keeping the risk of not fulfilling the requested QoS low. Performance perceived by the subscriber could differ among MNOs even if the same QCI/5QI class is provided, as described in Section 5.2.2.

Network Slicing is available with standalone 5G New Radio. A dedicated SST for V2X was introduced and one way of using it could be to agree on common and well tested performance provided by certain 5QI classes. Furthermore, the Slice Differentiator (SD) can be used for a more fine-grained differentiation.

This requires coordination across MNOs as usually done in the GSMA. A prerequisite is a clear understanding of what is required from the network so the QoS provided in the home network can either be also provided in a visited one or a clear agreement exists about what minimum performance and other features to expect when roaming.

4.2.1.4 Cross-MNO Alignment on Service Level QoS Requirement Description

GSMA has published a Non-binding Permanent Reference Document describing GSTs [6] intended to provide the means to better describe network requirements for standalone 5G New Radio networks. It currently defines 39 attributes, but not all are mandatory. Some examples are:

- Availability
- Coverage
- Delay tolerance
- Deterministic communication
- Downlink throughput per Network Slice
- Downlink throughput per UE
- Maximum supported packet size
- Mission critical support
- Number of connections
- Number of terminals
- Performance monitoring
- Performance prediction
- Positioning support
- Reliability
- Session and Service Continuity support
- Supported device velocity
- Uplink throughput per Network Slice
- Uplink throughput per UE

A filled-out GST is called NEtwork Slice Type (NEST) and can be used by the MNO to either instantiate appropriate Network Slices or identify ones providing the requirements. It is expected that MNOs will agree on certain NESTs for certain sets of services and each MNO will then have to configure its network in a way fulfilling the requirements under the conditions that enough resources are available. Partial outage due to congestion at certain times or places cannot be precluded.

In Section 4.1.2.1, an example is provided about how Network Slices can be orchestrated according to machine readable requirements. It goes beyond the above since it also includes a description of the application server. This remains a highly complex task since every network is different and therefore no simple mapping of requirements to network (slice) instances exists.

### 4.2.2 QoS Prediction

Substantial parts of this section are applicable for many different use cases. Within this project the applicability of the solutions below should mainly be evaluated for ToD and the performance prediction version of HD Mapping User Story 4 (see Deliverable D2.2 [1]). Synergies to prediction methods for cost reduction, as described in Section 4.2.3, is expected so at a later stage parts might be reused and individually obtained insights during trials will complement each other.

The solution below considers support from the network side, e.g., by using information about the cell load. Without information from the network side, a crowd-sourcing QoS prediction approach is still possible. An indication for the impact of network information on the accuracy of QoS predictions can be found in [61].

The need for QoS prediction in general is based on missing information of the expected network quality and in advance for a location and/or time even if no PDU session is established. In 4G and 5G mobile radio networks various methods are available to ensure QoS for applications, but they are only enforced by the scheduler based on the available radio-resources that can change over location and time.

A vehicle is moving and depending on the geographical environment, coverage gaps (e.g., tunnels, mountains, forests) can occur. Traffic prioritization will not solve the problem that the QoS required by an application cannot be met if radio resources are not available at all or their quality is too bad. Prediction enables in-advance information of the expected network quality for a given
location and time to a specific subscriber of the network. It can include direct or indirect\textsuperscript{26} coverage information of the radio network and available\textsuperscript{27} network capacity.

\textbf{Figure 4-9: General Principle of Network QoS Prediction}

Figure 4-9 shows the general principle how such prediction can be realized. The goal is to predict the QoS in the future at a different time and location. The prediction algorithm, or set of algorithms, is in the center of this task. It takes observed parameters as input. Those do not necessarily exactly resemble the current and past network QoS due to many constraints such as sampling, quantization, delayed availability, noise, etc. Besides network parameters, further environmental ones can be collected, e.g., GNSS positions. This can be done in the network, the vehicle or any other available source given that the necessary interfaces exist (see Section 4.2.2.1).

The prediction algorithm can be stateful. The exact meaning of that depends on the algorithm but this can typically include a previous training phase and/or feedback on the accuracy of predictions. The output of the algorithm is the predicted QoS. Once the time and place for which the prediction was done is reached, it can be evaluated against reality. This evaluation is again subject to the limitations of observing the parameters. This evaluation and its result allow to determine the correctness of the previous estimation and can be potentially used to improve future predictions.

\textbf{4.2.2.1 Deployment and Interfaces}

This section describes different deployment options for the Prediction Function (PF). The following entities and definitions are used in the next sections:

\textsuperscript{26} “Direct” is what can be measured while “indirect” means what is actually perceived by higher layers, and/or the application

\textsuperscript{27} Per cell or subscriber
• Prediction Client: triggering request of network QoS prediction to application server, receiving QoS prediction from AS, providing QoS prediction to other applications in charge for adaptation at vehicle side.

• Prediction Function (PF): generator of prediction.
  o Prediction can be a result of composition from several PFs
  o In the 3GPP architecture, the Network Data Analytics Function (NWDAF) is the PF

• Application Function (AF): component interacting with the network to request and/or receive QoS prediction / network information.

• Application server (AS): component handling QoS prediction request / response from / to client, interacting with the AF to request / receive network QoS prediction. AF and AS could be jointly implemented.

4.2.2.1.1 3GPP Study on QoS Prediction for V2X Communication
One approach for realizing the PF involves the usage of the interfaces and services provided by the network over the NEF and the user plane. 3GPP has worked in this option and therefore Figure 4-10 depicts the high-level interfaces. Tin this case the PF is implemented at the NWDAF and the AF or the AS can ask for QoS prediction information via the NEF. Section 4.2.2.1.2 provides an overview of different deployment options that could be considered, taking into account different architectural assumptions as well as the location of the AF and the AS.

Figure 4-10: Realization of QoS Prediction at NWDAF

The status of this study is described in the following.
3GPP, has already specified some requirements to support QoS monitoring and notifications for V2X applications [62]. For instance:

- “The 3GPP system shall be able to authenticate and authorize V2X application for requesting information on the quality of service of the ongoing connection or on the estimated quality of service.”
- “The 3GPP system shall be able to notify V2X applications with updated estimation of unfulfilment (or re-fulfilment) of quality of service for a certain geographic area, at least a certain amount of time before when the actual change occurs.”
- “The 3GPP system shall be able to provide quality of service information to the V2X application in a resource efficient way.”
- “The 3GPP system shall be able to support an efficient and secure mechanism to gather information (e.g., location information, reliability information, timing information, latency information, velocity information), in order to generate information about quality of service in a resource efficient way.”

In 3GPP Service Architecture (SA) Working Group (WG), architectural solutions are under discussion about the notifications on potential QoS change [11], [36]. The goal is to enable 5G communication systems to provide analytics information regarding potential QoS change upon request from a V2X Application Server. To this aim, 3GPP in Rel. 16 introduced two different solutions for 5GS [36]. One solution is formally known as “QoS Change based on Extended NG-RAN Notification to support Alternative Service Requirements”, which has been introduced to support V2X applications that can operate with different configurations (e.g., different bitrates or delay requirements). In this solution, the V2X AS, acting as the AF, can provide, in addition to the requested level of service requirements, Alternative Service Requirements to the 5GS which are then mapped to Alternative QoS Profiles. This enables the 5GS to act on the Alternative QoS Profiles and apply them for the extended NG-RAN notification [7], i.e., when NG-RAN is not able to fulfil the main QoS Profile, it triggers a notification towards the core network indicating that the main QoS Profile can no longer be guaranteed and includes also an indication of which of the Alternative QoS Profile NG-RAN is expected to currently fulfil. Finally, the UE may be notified via NAS signalling about the QoS Profile that NG-RAN is expected to currently fulfil, thus allowing the application to understand which configuration to use that is better suited for the currently fulfilled QoS Profile. Another solution is formally known as “Notification on QoS Sustainability Analytics”, enabling a V2X AS to receive from NWDAF analytics information regarding the QoS change statistics for a target period in the past in a certain area or the likelihood of a QoS change for a target period in the future in a certain area (i.e., QoS prediction) [8].

The procedure to provide “QoS Sustainability Analytics” is presented in Figure 4-11, The NF consumer can be an AF or an AS. In this solution, the V2X AS subscribes / requests QoS Sustainability Analytics and provides NWDAF with Analytics Filter Information containing:

- QoS requirements, either in the form of standardized / pre-configured 5QI, QoS Characteristics attributes including Resource Type, PDB, PER and their values;
- Location of interest for notifications (in the form of a path or geographical area);
- Analytics target period: relative time interval, either in the past or in the future, that indicates the time period for which the QoS Sustainability analytics is requested;
- Reporting Threshold(s), which apply only for subscriptions and indicate conditions on the level to be reached for the reporting of the analytics, i.e., to discretize the output analytics and to trigger the notification when the threshold(s) provided in the analytics subscription are crossed by the expected QoS KPIs. The level(s) relate to:
  - for a 5QI of GBR resource type, the Reporting Threshold(s) refer to the QoS flow Retainability KPI\(^{28}\);
  - for a 5QI of non-GBR resource type, the Reporting Threshold(s) refer to the RAN UE Throughput KPI\(^{29}\).

If the Analytics target period is in the future, which is the focus of QoS prediction, the NWDAF detects the need for notification about a potential QoS change based on comparing the expected values for the KPI of the target 5QI against the Reporting Threshold(s) provided by the V2X AS in any cell in the requested area for the requested Analytics target period. The expected KPI values are derived from the statistics for the 5QI obtained from OAM. The analytics feedback contains the information on the location and the time when a potential QoS change may occur and what Reporting Threshold(s) may be crossed.

The utilization of this procedure for V2X applications is discussed in TS 23.287 [36]. It is considered that a V2X AS may request notifications on QoS Sustainability Analytics for an indicated geographic area and time interval in order to adjust the application behaviour in advance with potential QoS change. The V2X AS may also request past statistical information for the purposes of adjustment of the application, how V2X AS makes use of such data is outside of 3GPP scope. The V2X AS acts as an AF that communicates with the NEF, which corresponds to the NF consumer for the Procedure for "QoS Sustainability" analytics defined in TS 23.288. The V2X AS can either subscribe to notifications from the NEF (i.e., a Subscribe-Notify model) or request a single notification from the NEF (i.e., a Request-Response model).

\(^{28}\) Number of abnormally released QoS flows during the time the QoS Flows were used per timeslot, per cell, per 5QI and per S-NSSAI.

\(^{29}\) Average UE bitrate in the cell (Payload data volume on RLC level per elapsed time unit on the air interface, for transfers restricted by the air interface), per timeslot, per cell, per 5QI and per S-NSSAI.
It should be noted that according to the current specification in 3GPP, the OAM provides the input data to the NWDAF in order to derive the predictions. For the sake of completeness, we should mention that, according to implementation, NWDAF could collect monitoring data by the RAN and/or other core network entities (i.e., this is not part of the current standard). The latter may provide the benefit of a more precise prediction, but signalling complexity, required computational resources, and multi-vendor interaction are some of the issues that should be considered.

4.2.2.1.2 Deployment Options
The PF client is the device that triggers the transmission of prediction request message as well as the final recipient of prediction response message. Different deployment options could be considered, taking into account the above architectural assumptions as well as the location of the AS and AF (e.g., located inside or out of an MNO domain).

Figure 4-12 describes the different deployment options, where NWDAF is always involved (in a 3GPP-based architecture) as the entity that derives analytics and predictions from the network side:

- Option 1: AS / AF are located out of an MNO domain.
- Option 2: AS / AF are located inside the MNO domain.
- Option 3: AS is located out of an MNO domain and the AF is located inside an MNO domain, having also PF functionality (e.g., in collaboration with the NWDAF).
- Option 4: AS and AF are located out of an MNO domain, while the AF can also have some PF functionality (e.g., by using the NWDAF response).
Options 3 and 4 provide the benefit of scalability, since multiple application servers can be served by using one AF. If the PF is located inside the domain of an MNO then it is easier to access the MNO data (e.g., RAN, core network entities) that could impact the QoS prediction result.

**Figure 4-12: QoS Prediction – Deployment Options**

### 4.2.2.2 QoS Prediction Message Specification

#### 4.2.2.2.1 Message Sequence Charts (MSC)

Figure 4-13 and Figure 4-14 present the flow charts how the PF client located at the vehicle side requests the network to deliver QoS prediction service. At both cases the PF Client sends the prediction request message to the PF (via the backend AS) providing context and configuration information that will help the network to deliver the QoS prediction. The prediction response message can be sent: a) from the AF to PF client, via the AS (Figure 4-13), or b) from the PF AF directly to the PF Client, in case the PF AF is hosted at a MEC (Figure 4-14). The intra MNO signalling and the entities involved in the generation of QoS prediction (Step 3) could be based on the procedure described in Figure 4-11. The content of the prediction request and prediction response messages are presented in Section 4.2.2.2.2. The figures do not consider the whole vehicle architecture, only the PF Client is shown, as it represents the end point of the vehicle-backend communication regarding QoS prediction. QoS prediction-based application adaptations take place in further components running on the car PC. More information about the later will be provided in deliverable D3.1.
4.2.2.2 Message Content
As presented in section 4.2.2.1.1, initial messages for the specification of QoS prediction request and response messages have been defined in 3GPP, also with contributions of 5GCroCo partners as presented in Deliverable D6.1 [63]. The content of the prediction request and prediction response messages that 5GCroCo has defined is given in the following in a tabular format, since it shall not prescribe a specific message format. The QoS prediction request message in the Table
4-6 provides additional UE capabilities, as additional contextual information e.g., UE antenna features that could help the derivation of more precise QoS prediction. The QoS prediction response message in Table 4-7 provides as output also deterministic values or classes of values, while in 3GPP the response message is correlated with the reporting threshold of the received request message.

### Table 4-6 QoS Prediction Request

<table>
<thead>
<tr>
<th>Category</th>
<th>Content</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>UE capabilities</td>
<td>CAT Category</td>
<td>To tailor the prediction to the UE, all capabilities relevant for the prediction have to be communicated.</td>
</tr>
<tr>
<td></td>
<td>Supported bands</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of antennas on vehicle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UE ID</td>
<td></td>
</tr>
<tr>
<td>QoS Prediction configuration</td>
<td>Requested confidence for the uplink predictions</td>
<td>Some parameters to configure the prediction. The thresholds are only needed if updates shall be received. In case of data rates, updates are only supplied if the old or new prediction is below this value. In case of latencies, updates are only supplied if the old or new prediction is above this value.</td>
</tr>
<tr>
<td></td>
<td>Requested confidence for the downlink predictions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>QoS Requirements for a specific V2X Service (optional)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information whether updates shall be received for changed predictions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uplink data rate threshold [Mbit/s]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Downlink data rate threshold [Mbit/s]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uplink latency threshold [ms]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Downlink latency threshold [ms]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uplink packet error rate threshold [%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Downlink packet error rate threshold [%]</td>
<td></td>
</tr>
<tr>
<td>Location Information and Observation period</td>
<td>List, each entry contains: Longitude</td>
<td>The route for which the prediction shall be obtained. Timestamps per location define when the vehicle is expected to be there.</td>
</tr>
<tr>
<td></td>
<td>Latitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Altitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coordinated Universal Time (UTC) Timestamp</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4-7: QoS Prediction Response

<table>
<thead>
<tr>
<th>Category</th>
<th>Content</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction information</td>
<td>Type of the prediction (full/update)</td>
<td></td>
</tr>
</tbody>
</table>
Upcoming route with prediction

<table>
<thead>
<tr>
<th>List, each entry contains:</th>
<th>Applicable Area</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Latitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altitude</td>
</tr>
<tr>
<td>Thresholds</td>
<td>Min. uplink data rate [Mbit/s]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min. downlink data rate [Mbit/s]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max. uplink latency [ms]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max. downlink latency [ms]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uplink packet error rate [%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Downlink packet error rate [%]</td>
<td></td>
</tr>
</tbody>
</table>

Contains the prediction of QoS for each position.
Each prediction is supplied with the requested confidence.
In case an update is supplied, only those positions are included for which an entry has changed.
The “validity duration” defines a period around the “timestamp” of the request, during which the prediction is valid.
It is to be determined, whether the predictions will be deterministic values or classes of values.

4.2.3 Prediction for Reduced Cost

The SCEF [19] provides BDT configuration and notification services to application servers over the T8-interface. With standalone 5G New Radio the same is provided through the NEF and the N33-interface. It can be used to support HD map update download at reduced cost if the MNO enables the services with the respective subscription. This can be used within the cost reduction focused version of HD Mapping User Story 4 (see Deliverable D2.2 [1]), where within certain bounds the best place or time for data transfer should be suggested by the network. Obeying the suggestion provided by the network would be rewarded by reduced cost, e.g. by reducing only a fraction\textsuperscript{30} from the data plan for each transferred byte.

The application server provides the data volume (downlink, uplink or both), the desired time window within which the transmission should finish and optionally a location area where the BDT request is valid. Further components are required in case this information is available in the vehicle, not the backend, since the T8-interface is on the backend side. In this case the application in the vehicle must provide the required information to the application server in the network.

\textsuperscript{30} E.g. transferring 100 MB reduces the data plan by just 70 MB if 30 % reduced cost is agreed
The SCEF will reply with the time window to be used for communication. The application server may send and/or receive the agreed data volume within this time window and individual charging will be applied to it.

This API was originally intended for use cases like software update or data collection from static IoT sensor nodes where the network would reply with an off-peak hour for the transmission. For HD mapping the current API might not be sufficient. The following solutions are possible:

- (Machine) learning based approach where the network determines the best time window to reply for certain requests received at certain locations. From experience

- the network would predict where the requesting vehicle is going and estimate the time window for the reply when the vehicle will likely be in a suitable spot to conduct the data transfer.

- Extending the API or defining a new one to e.g., allow the vehicle to also provide route information including waypoints and estimated time of arrival or by an event-driven approach where the SCEF notifies the AF when a suitable location is reached.

### 4.2.4 QoS Prediction Algorithmic Models

The described mechanism is based on a Fusion Machine Learning scheme, which combines different data sources, in order to predict with high accuracy, the expected QoS, which is then provided to the vehicles. The described Predictive QoS algorithm is capable of processing diverse types of context information depending on their availability in different deployments and network architectures.

On the one hand, in order to adequately assess the available QoS for the respective V2X service, the mechanism needs to identify the different parts that comprise the end-to-end communication path. As also described in the respective 3GPP’s study on application layer support for V2X services, this end-to-end communication path depends on the type of the V2X application/service, i.e., vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-network (V2N) or vehicle-to-pedestrian (V2P). On the other hand, besides the communication aspect, additional delay-introducing components contribute to the end-to-end network performance, such as device buffers, computing VNFs/PNFs, backhauls, as well as core network components.

In order to optimize the processing of the different context parameters, they are grouped in five different categories, which are illustrated in Table 4-8 that follows.

---

31 Depending if the request was for uplink, downlink or both directions.

32 E.g. because there is only one or few major roads in the request area.
Table 4-8: QoS Prediction Input Context Parameters

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Input metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility information</td>
<td>Required Information related to the mobility of the vehicle/UE, in order to request the prediction. The optional mobility information is only required in the case of Trajectory prediction of the UE.</td>
<td>Required: UE Route including Latitude, longitude and timestamp for each location Optional: Velocity, acceleration, heading, predicted path, trajectory constraints (e.g., road limits), etc.</td>
</tr>
<tr>
<td>Radio parameters</td>
<td>Radio-related, passive measurements provided by the UE. These metrics are complementary, in order to enhance the prediction.</td>
<td>Received Signal Strength Indication (RSSI), Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), CQI, Signal to Interference Plus Noise Ratio (SINR), client (GPS, velocity, heading), geolocation map</td>
</tr>
<tr>
<td>RAN/facilities layer</td>
<td>Latency-introducing network components related to queues, computing resources’ availability, etc.</td>
<td>Roadside Unit (RSU) -related network load, RSU-queues load, Number of UEs associated to Base Station (BS), availability of MEC/cloud computing resources</td>
</tr>
<tr>
<td>Backhaul / end-to-end / Network performance perspective</td>
<td>Active network measurements related to the end-to-end service.</td>
<td>packet loss, jitter, latency, throughput.</td>
</tr>
<tr>
<td>Application-specific information</td>
<td>Specific, application-oriented information that influences the performance of a V2X service. For example, some V2X scenarios are group-based communications, such as</td>
<td>Service priorities, group-based communication type, etc.</td>
</tr>
</tbody>
</table>
It is important to note that the above table illustrates the information sets that can be potentially processed during the offline training process. The proposed algorithm does not need all the afore-presented information in order to perform the real-time prediction process. Depending on the respective deployment and the available real-time information, the algorithm is capable of generating a prediction based only a sub-set of these metrics, such as the location information, RSRP/RSRQ, velocity, etc.

4.2.4.1 System Model

We consider a prediction communication system, in which $M$ mobile devices (i.e., vehicles) are consuming $N$ different V2X services and are notified by the network about the predicted downlink (DL) / uplink (UP) data rate, packet error rate and end-to-end latency network Key Performance Indicators (KPIs), in two possible ways: a) in a periodic manner, b) upon change of the predicted QoS values/value classes.

The PF can be deployed either at a MEC server or at an NWDAF, as part of the Core NFs. In the case of a MEC-based deployment, the delays and packet losses in backhauls and core networks are considered obviously equal to zero, according to the input parameters modelling presented in the previous subsection. The described algorithm generally considers a 5G wireless communication system, where mobile devices consume different V2X services e.g. corresponding to the 5GCroCo use cases HD mapping and ToD, which make use of prediction. The PF can either provide V2X service-agnostic QoS Prediction Responses, to which different services may subscribe in order to receive the respective messages; alternatively, the PF may be set up to provide V2X service-specific predictions, tailored to the specific services’ requirements and respective QoS Prediction Request messages.

4.2.4.1.1 Geographical Space as a Grid

The considered geographical space (map) is modelled based on grid-tile approach, comprising small rectangle cells towards applying the prediction in a discrete manner (Figure 4-15 (a)). Based on a recursive QoS metric assessment, mapped to the discrete map grid cells, the second step...
is the clustering of grid cells, in a way that the clusters demonstrate similar behaviour in terms of QoS metrics (Figure 4-15 (b)).

![Figure 4-15: (a) Initial Grid Modeling, (b) Grid after the QoS-based Correlation Clustering](image)

The major advantage of this approach is the minimization of the computational overhead, as the prediction is applied per cluster. The geolocation information (i.e., latitude, longitude) that is received as one of the algorithm's inputs, is directly translated to one specific grid cell, or clustered cell, respectively. The QoS within each geographical cell is considered to be homogeneous and hence, within each cluster.

### 4.2.4.1.2 Time-based Modelling

The time of the prediction model is initially discretized into $T$ slots of a pre-defined duration, namely $\text{QoS}_{\text{window}}$, for which the QoS metrics of a specific grid cell are considered static and is considered as the prediction horizon for each single prediction. An example $\text{QoS}_{\text{window}}$ for a single cell/cluster cell could be defined at 1 minute; this translates to 60 min $\times$ 24 hours $\times$ 7 days = 10,080 slots for a weekly-based prediction model. The second step is to normalize the time dimension, depending on the periodicity of the model to be generated (e.g., time is normalized from 0 to 1 for one-week duration using min/max normalization or standardization).

Based on the seasonality of the data we want to capture, the interval could be daily, weekly, monthly or yearly. In order to choose the appropriate interval, during data analysis 1) the interval that suggests a cyclic pattern should be chosen and 2) the training data must be sufficient for each timestamp.

According to the validity duration requested by each V2X application, the predicted data rate/latency/packet error rate values are extracted based on a one-to-one mapping, depending on the cluster cells that are traversed during this time duration. The traversed cells are computed based on the vehicle's position and mobility characteristics (i.e., velocity and heading).

Figure 4-16 illustrates in high-level the afore-described concept. For the requested validity duration $d$, the vehicle path comprises 3 adjacent cluster cells. Let us denote time duration $d$, where $d = d1 + d2 + d3$.  

---

80
4.2.4.2 Machine Learning Algorithms

The proposed mechanism exploits two different Machine Learning (ML) approaches in order to generate the QoS prediction models. On the one hand, a Static Multiple Linear Regression Model is used; besides the regression model approach, a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) architecture operates in a complementary manner.

4.2.4.2.1 Static Multiple Linear Regression Model

This approach relies on the intuition that the service traffic-related data follow a seasonal pattern on a weekly interval and therefore during training the model is able to capture that seasonality. Moreover, the model can be extended to perceive cyclic patterns in the data, meaning that larger intervals (monthly / yearly) can be considered (e.g., more traffic during holidays etc.). In the case that during data analysis, the data lacks any seasonality, the static model demonstrates a lower prediction performance due to the unpredictable nature of the data, thus, in such cases -where smaller datasets of limited time periods are available - a more dynamic prediction methodology is more suitable and will be presented later, as part of the LSTM-based forecasting (Section 4.2.4.2.2). The overview of the training phase of the Linear Regression Model (LRM) is illustrated in Figure 4-17.
For each QoS KPI that is relevant to the specific V2X service configuration, the respective regression models are generated (e.g., for DL data rate, UL data rate, DL/UL latency, DL/UL packet error rate, etc.). Those regression models are generated per grid cell, or grid cluster respectively, after the clustering step has taken place. Each regression model is accompanied by the respective confidence intervals (Figure 4-18).
One significant advantage of the Linear Regression Model approach is that as in order to augment the historical measurements-based model, we correlate the initial regression models that are built, with near-real time information retrieved from the network, such as the road traffic, i.e., the number and density of the vehicles/UEs in a specific area/cluster cell. By using Pearson Correlation, the initial regression model for the specific cell is adapted and an enhanced regression model is generated.

Finally, performance- and memory-wise, this approach is extremely efficient and scalable, which is crucial in a dense environment, where numerous QoS prediction requests must be handled and addressed in a timely manner, especially for challenging services such as URLLC-related traffic type.

4.2.4.2.2 Long Short-Term Memory-based (LSTM) Forecasting

The prediction task under study can also be categorized into the class of time-series forecasting problems. This state-of-the-art approach introduces a single model for the whole grid, predicting the QoS values on each cell, in the same manner as in predicting the pixel values on a video stream based on the previous values. A simplified architecture of the LSTM approach is illustrated in Figure 4-19.
The example illustrates 1 feature and 1 timestep. This is however fully scalable and it is possible to use multiple timesteps and features. The first two layers of the network are implemented as LSTM, while the last layer is Fully Connected.

Towards offline training, the available dataset is split into training/testing (e.g., 80/20 split). The network predicts the DL data rate for the following time slot. For each predicted time slot the result is compared with the actual.

The loss metric (Mean Squared Error) is employed for the evaluation of the LSTM model performance.

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \]

The prediction provided by the LSTM model, compared to the previous Linear Regression Model, is more dynamic in forecasting, by learning underlying patterns in the time-series data, that a simple Regression may miss.

LSTMs are not great at learning long term dependencies and therefore, it may be unable to capture the seasonality of the data (if present). Moreover, the computational and memory demands for training, as well as, real-time operation are high, which may cause a bottleneck, given the available infrastructure.

4.2.4.2.3 Combination of the two Approaches

A hybrid approach will also be evaluated during the 5GCroCo trials execution, which exploits both algorithms, in order to mitigate performance issues in the LSTM approach (Figure 4-20). The steps followed are:

- Grid creation
- Creation of Multiple Linear Regression for each cell
- Map Grid Cell Clustering (as already described)
- LSTM applied on a cluster basis, instead of grid tiles.

![Simplified Example LSTM Architecture for QoS Prediction](image)
4.2.4.2.4 Non Map-Grid Based Regressors: Random Forest Regressor
This implementation is more straightforward, consisting of one Random Forest model for the entire geographical region, without partitioning the map in a grid, and therefore, the latitude and longitude data are not converted into Cell Id. The input data of the model are, as in the previous models, the basic three [timestamp, latitude, longitude]. There is no need for normalization or standardization of the input parameters, while also having the ability to specify a confidence interval.

4.2.4.2.5 Model Evaluation
The Loss Function that was used for the training of the model and for evaluation is the **Quantile Regression Loss function**. The loss in Quantile Regression for an individual data point is defined as:

\[
L(\xi_i|a) = \begin{cases} 
\alpha \xi_i, & \xi_i \geq 0 \\
(a - 1)\xi_i, & \xi_i < 0 
\end{cases}
\]

where alpha (\(\alpha\)) is the required quantile (a value between 0 and 1 -> 0-100% quantile) and \(\xi_i = y_i - f(x_i)\), where \(f(x)\) is the predicted (quantile) model and \(y\) is the observed value for the corresponding input \(x\). The average loss over the entire dataset is shown below:

\[
L(y, f|a) = \frac{1}{N} \sum_{i=1}^{N} L(y_i - f(x_i)|a)
\]
This Loss Function was used, due to the fact that a confidence interval was needed in order to provide an upper or lower \textit{threshold} for each predicted QoS metric (UL/DL data rate, UL/DL latency and UL/DL packet loss) prediction, based on the requested interval. For example, if the requested confidence interval is 90\%, then the Quantile Loss function of the ML Model will have as input the $\alpha$ (alpha) with value of 0.1, in order to estimate the 10th and 90th quantile. An illustration of different quantiles is shown below. (note that a quantile of 0.5 is equal to the median, the same as using the Mean Absolute Error (MAE) loss function).

![Quantile Regression for Different $\alpha$ Values](image)

\textbf{Figure 4-21}: Quantile Regression for Different $\alpha$ Values

In order to test the model’s performance in terms of accuracy, precision and recall, 10-fold cross validation was used. Ten-fold cross validation iterates 10 times, where each time 10\% of the dataset is used for testing (validation data) and the other 90\% for training (different part of the dataset each time). After calculating the overall loss of the predicted values against the real values, a mean loss is calculated, stating the model’s performance. Since the QoS prediction is based on interpolation, meaning that there will not be a request for an input $x$ (i.e., daily interval 00.00 - 23.59) outside the value range in the training (hence no extrapolation), the data are shuffled before each cross validation.
4.3 MEC

MEC describes the capability of application server hosting within the domain controlled by the MNO, e.g., in its data centers or cabinets (see Section 2.1.1, especially Footnote 8). This includes the case of using the same gateway (4G EPC P-GW, 5G PDU Session Anchor UPF) as for public Internet access, but also additional gateways are being deployed.

Modified versions of many text parts and figures of this section were also contributed to 5GAA, but the corresponding document is not yet public and therefore cannot be referenced and to a joined white paper of the Horizon 2020 ICT-18 projects [64].

4.3.1 Related Technical Specifications and Studies and their Applicability

4.3.1.1 3GPP

The sections below provide an overview of MEC-related features in 3GPP specifications. 3GPP uses the term “Edge Computing” but for consistency the term “MEC” will be also used in 3GPP context in this document.

4.3.1.1.1 RAN

No 3GPP RAN specification was added or evolved for MEC. According to the definition from Section 4.3 there is a relation between MEC and RAN due to deployment of cell sites and their association to gateways. This is important as MEC-hosts will likely be close to or even collocated with the gateways. A Next Generation Node B (gNB) can usually reach different gateways, potentially even all gateways in the network. When establishing a packet data network session, the Core network can use different criteria to select an appropriate gateway, as further described.
in Section 4.3.2.2. In context of MEC proximity, but also balanced load\textsuperscript{33}, are appropriate criteria. Due to mobility and related radio handovers a gateway that was selected during packet data network session initiation might become far away and others might be closer. Switching the gateway is further discussed in Section 4.3.2.1.2 but a prerequisite for this is a notification from the RAN to the Core network about radio handovers. Theoretically, this can be done for every handover but practically a more coarse-grained approach is used and Tracking Area Updates (TAUs) are used as potential trigger for gateway changes.

4.3.1.1.2 Service and System Aspects Working Group 2 (SA2) - Architecture

5G Core network specifications \cite{7} introduce two MEC-specific features to support gateway mobility. One is session and service continuity, which is further described in Section 4.3.2.2, and the other is the AF influence on traffic steering API, further described in Section 4.3.2.1.2.

4.3.1.1.3 Service and System Aspects Working Group 5 (SA5) - Telecom Management

3GPP SA5 is specifying MANO-related topics which also includes MEC. This is most obvious with the recently launched study on enhancements of edge computing management \cite{65}. Even before that an implicit relation existed as deploying application servers to MEC hosts can rely on solutions used to deploy VNFs to their physical hosts, as further described in Section 4.3.3.2.

\textsuperscript{33} E.g. by taking into account the number of sessions served by a gateway
Figure 4-23 is based on a publication [10] created by key people involved in 3GPP SA5 5G standardization. It defines different roles of actors providing and consuming services from each other in context of virtualized networks. The 5GCroCo project sees the need for unified terms for different stakeholders, but the ones provided in 3GPP SA5 context were not considered applicable without adaptation and refinement. This is described in Section 4.3.2.

4.3.1.2 Automotive Edge Computing Consortium (AECC)
The Automotive Edge Computing Consortium (AECC) was founded in 2017 to bring together industry stakeholders to drive the evolution of edge network architectures and computing infrastructures for connected vehicles. A key goal is to more efficiently support high volume data services.

The AECC Technical Report [66] summarizes all findings of the Technical Solution Working Group and is being continuously updated. Its cellular network architecture consists of the 3GPP 4G architecture including QoS features and the 5G architecture, where QoS is included by default. The core network can either be standalone or non-standalone with a 5G Core or a 4G EPC, respectively. In both cases the RAN can be a mix of 5G New Radio and 4G LTE cells.

AECC, so far, considers six key issues and proposes solutions for them. Only a selected subset of the six key issues and solutions is presented below:
**Edge Data Offloading:** This key issue combines the challenges of routing different kinds of data traffic to either MEC- or public Internet hosted application servers. According solutions correspond to what is described in Section 4.3.2.1.1 mentioning that different Access Point Names (APNs), or Data Network Names (DNNs) in case of standalone 5G New Radio, can be used for such separation. It furthermore discusses how MEC-hosts can be connected to 3GPP networks and conclude that this should be done through the SGi and N6 interface for non-standalone and standalone 5G New Radio, respectively, as further described in Section 4.3.2.1.1.

**Server Selection:** This key issue discusses how an appropriate server can be found depending on the location of the vehicle. Besides one IP anycast based solutions, three other solutions are based on Domain Name System (DNS). They differ according to how the DNS server obtains the information on where the vehicle is located to provide an appropriate MEC host address. This is further described in Section 4.3.2.3.1.

**Opportunistic Data Transfer:** This relates to use cases where some flexibility exists with regard when and/or where large data volumes should be transmitted. The variant of HD Mapping User Story 4 focusing on cost reduction (see Deliverable D2.2 [1]) is such an example for an application where immediate data transmission is not needed but there is a certain deadline, e.g., a map tile must be downloaded before entering the area it maps. Among other solutions, BDT, as described in Section 4.2.3, is described.

### 4.3.1.3 Cloud Native Computing Foundation (CNCF)

Common practice and solution for cloud computing is aligned by CNCF. Many aspects of MEC (see also Section 4.3.3.2) can, but do not have to, be realized based on the principles of CNCF, commonly referred to as “cloud native”. It can be summarized as:

> “Cloud native technologies empower organizations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify this approach.”

[67]

In the following, the statement and major terms used are further described:

**Public, private, and hybrid clouds:** A public cloud provides computation, connectivity and storage resources to many businesses and other kinds organizations using it, according to [68]. Public clouds are operated by cloud providers (see Figure 4-24). The largest ones are typically referred to as “hyperscale cloud providers” (e.g., AWS, Microsoft, Google). Private clouds are operated by the same business or other kind of organization that also uses it to run its applications on it. “Hybrid” refers to a mix of both where some applications or parts of them are on public and other

---

34 In this deliverable and throughout the 5GCroCo project the term “Public Internet” is used especially to point at the location of a backend application server, in contrast to MEC hosting. The term public cloud here stresses the MANO aspect of cloud computing that includes the entire lifecycle of the backend service, as described in Section 4.3.3.2 and Section 4.1.2.
on private clouds. According to this definition, MNOs providing MEC services would be considered as public clouds. The attribute “MEC” therefore used to further distinguish MNO from hyperscale hosting, where it is not precluded that MNOs cooperate with hyperscale cloud providers to offer MEC services, as further described in Section 4.3.3.2.

**Containers:** VMs represent a way of virtualization requiring most resources to host a machine, compared to other methods. VMs include the whole operating system (OS) with applications running on top of it. KVMs are an improvement where the Kernel of the operating system is shared among the VMs running on top of it. Containers are an even more lightweight virtualization solution and especially used in hyperscale clouds. Docker is the most commonly used open source software to create containerized applications. A container includes the application itself and all helper applications and libraries it requires to run. They can therefore be deployed and executed without having to resolve further software dependencies. Kubernetes is commonly used to manage tasks related to container deployment and further operations and maintenance tasks. The role of Containers and Kubernetes as their MANO system is further described in Section 4.3.3.2.

**Microservices:** An overall backend application is commonly referred to as “the service” while the subparts constructing it are called “microservices”. Containers allow to create and deploy self-contained applications (microservices) that can cooperate with each other to provide an overall service. Microservices allow distributed software development and evolution where only interfaces between the microservices must be defined. They can include the logic to discover each other, connect to each other and provide fail-over mechanisms.

**Service meshes:** Discovery, communication and fail-over can be part of the microservice implementation. It can also be decoupled from it, meaning application developers do not need to implement such logic, and handled by the service mesh.

**Immutable Infrastructure:** Libraries, or even the OS hosting the container might require updates or configuration changes. The paradigm of immutable infrastructure states that such changes should not be made during runtime. Instead, new (micro)service instances are deployed where these updates and/or changes have been done. The service mesh assures they are seamlessly integrated into the running overall service. Newly deployed microservices can be disabled and another configuration attempt can be done in case of undesired behaviour. Old containers can be gradually replaced by new ones.

**Declarative APIs:** In order to explain declarative APIs, their opposite, being imperative APIs, are explained first. Such APIs are invoked to get a task done step by step. A set of API calls is used to obtain intended outputs and/or state changes. Follow-up calls then built upon those e.g., by using output from previous calls as input for follow-up calls. Declarative APIs [69], [70] are used

---

35 [https://www.docker.com/resources/what-container](https://www.docker.com/resources/what-container)

36 [https://www.redhat.com/en/topics/microservices/what-is-a-service-mesh](https://www.redhat.com/en/topics/microservices/what-is-a-service-mesh)
to get an overall task done without exposing the underlying steps. An example would be “deploy two MEC-enabled mobile radio networks with interconnected MEC hosts” instead of “deploy RAN”, “deploy Core”, “interconnect RAN with Core”, “interconnect Core with MEC hosts”, “interconnect MEC hosts”, and so on.

4.3.2 Solutions for Selected Challenges
Before listing solutions for different challenges related to MEC in 5GCroCo context, we first evolve the high-level application reference architecture from Figure 2-4: 5GCroCo High-level Architecture from Deliverable 3.1 Figure 2-4 in Section 4 to also consider the stakeholders of the MEC ecosystem, of which parts are based on the general Cloud Computing ecosystem.

![Figure 4-24: Roles Related to Virtualized Networks in Context of 5GCroCo](image)

**Figure 4-24** combines the roles from 3GPP SA5, explained in Section 4.3.1.1.3, with the high-level application architecture that introduces the roles MNO, RTA, OEM, and SP. Those entities can host their backends through hosting services provided by Cloud Providers (see also Appendix B). This includes MEC and public Internet hosting. The backend software is provided by Backend Software Providers and for this and any other role it is not precluded that the same entity has multiple roles. An OEM could e.g., develop backend software by itself. The MNO also requires software and hardware to form the mobile radio network. This is provided by Network Equipment Vendors and the VNFs, being software, can be hosted by Cloud Providers without precluding that MNOs are at the same time also Cloud Providers. The required MANO software for virtualization is provided by Virtualization Software Providers. The hardware and its local interconnection are hosted/provided through Data Centers of different types, as further described in Appendix B. They are interconnected through links and services provided by Wide Area Network (WAN) Providers.
All this is set up so vehicles, shown on top of the figure, can be served through the network operated by the MNO. Required software in the vehicle comes from Client Software Providers.

4.3.2.1 Interfaces Between MEC and 3GPP Network

The following section describes how MEC and 3GPP networks can interface with each other and which implications for 3GPP architecture choices might result from it. Subsections for user plane and control plane are provided.

4.3.2.1.1 User Plane

The only 3GPP specification compliant option of connecting ASs, including those hosted on MEC, to the 3GPP network is through the SGi interface for non-standalone 5G New Radio with 4G EPC (see Figure 4-25) and N6 interface for standalone 5G New Radio with 5G Core (see Figure 4-26). In case of standalone 5G New Radio many options exist as there can be an arbitrary number of UPFs along the path and UPFs can serve as branch points to separate MEC and public Internet traffic.

Figure 4-25: MEC Deployment for Non-standalone 5G New Radio
Figure 4-26: MEC Deployment for Standalone 5G New Radio; with two UPFs Connected to BBU (Top) and with Branching Point UPF (Bottom)

As the 3GPP architecture builds on top of TCP/IP and common Local Area Network (LAN)/WAN technologies underneath it (e.g., Ethernet) further options emerge. The ETSI-MEC white paper on MEC Deployment in 4G Networks [71] and AECC Technical Report [66] report and also list the option to connect through the S1 interface (4G and non-standalone 5G New Radio) or N3 interface (standalone 5G New Radio). This deployment option is also called "bump-in-the-wire". The 5GCroCo project will not pursue this option for the reasons explained below.

MEC usually requires increasing the number of gateways\(^{37}\) where the SGi/N6 interface is exposed. The term "Local Breakout" is commonly\(^ {38}\) used for that and it is not precluded that existing gateways that e.g., provide access to the public Internet can also be used to enable MEC access. For standalone 5G New Radio Local Breakout is achieved by adding UPFs that can act as PSA UPFs.

From 3GPP Release 14 on an equivalent solution to adding UPFs for standalone 5G New Radio exists, as shown in Figure 4-27. Control and User Plane Separation (CUPS) can be used in the 4G EPC to achieve Local Breakout by deploying Serving Gateway (S-GW) User Plane (S-GW-U) and P-GW User Plane (P-GW-U) Network Functions (NFs). These NFs have less functionality than full S-/P-GWs that also include the control plane. The control plane functionality is centrally deployed in S-/P-GW Control Plane (S-/P-GW-C) NFs, e.g., collocated with gateways for public Internet access.

\(^{37}\) The term "gateway" equally refers to 5G Core Protocol Data Unit (PDU) Session Anchor (PSA) User Plane Functions (UPFs) (PSA UPFs) and 4G EPC Packet Data Network Gateways (P-GWs) and is used when it is not required to distinguish between the two core network generations.

\(^{38}\) Strictly speaking, according to 3GPP specifications, this term refers to using a gateway in the visited network when roaming, rather than one in the home network (called “Home Routing”). We use the term “Local Breakout” also in non-roaming context. In context of roaming, we use the term “Local Breakout Roaming” to be more precise.
CUPS is an optional feature and might not be available. In this case full S-/P-GWs can be deployed for Local Breakout. They require more resources in terms of computation, storage and memory than S-/P-GW-Us. The ETSI-MEC white paper on MEC Deployment in 4G Networks [72] also lists the option of deploying S-GWs without P-GWs and calls it “Local Breakout”. This option does not comply to 3GPP specifications as the S-GW is omitted. It can be interpreted that the P-GW is integrated in the MEC-host and/or the MEC application server it runs. In this case it corresponds to the option above. It furthermore lists the option to deploy a complete 4G EPC also including the pure control plane nodes MME and HSS. While this can make sense in other domains it has no benefit in the context of CCAM and only results in even higher computational, memory and storage resource demand.

The S1 interface bump-in-the-wire approach, corresponding to N3 interface for standalone 5G New Radio, is not suggested and will not be used or further evaluated in the 5GCroCo project. It is realized by rerouting traffic with certain characteristics to MEC hosts. This is problematic or even impossible if encryption is used. Furthermore, 3GPP security, lawful interception and charging functions were not designed for this [73].

In the following, different options are discussed enabling the vehicle to access MEC application servers and the public Internet. While MEC access is optional, public Internet access should always be possible. There are different ways to decide which P-GW(s) should be used by a UE. Figure 4-28 shows the option where an APN different from the default one providing public Internet access is used. In this example access to the public Internet is provided through the edge P-GW. Other MEC hosts belonging to the same MNO but connected through different Edge P-

---

39 The ETSI-MEC white paper is not focussing on automotive use cases

40 This is called Data Network Name (DNN) in the 5G New Radio specifications. Consequently, for non-standalone 5G New Radio, the term DNN would apply in the UE and APN in the core network. The term APN will be used for non-standalone 5G New Radio related parts and DNN for standalone 5G New Radio ones.
GWs can be accesses through the public Internet. Optionally, the MNO can configure its transport network to provide connectivity between these MEC hosts.

![Diagram of MEC with one APN](image)

**Figure 4-28: MEC with one APN**

Alternatively, a UE can use two different APNs, one towards the edge P-GW and one towards the P-GW providing public Internet access, as shown in Figure 4-29. This requires according support in the UE including an interface towards the application and/or operating system on the CCU allowing to establish two PDU connections with two different APNs. The vehicle would then have one IP address per APN. It is not precluded that also the edge P-GW provides Internet access, but this is optional, and it depends on the routing table of the vehicle if it is used.
Corresponding options exist for standalone 5G New Radio where APNs are called Data Network Names (DNNs). Besides that, branchpoint UPFs, as shown in Figure 4-30, can be used to filter traffic according to defined rules, e.g., based on IP addresses and port numbers. This is only needed in the uplink as for the downlink the traffic enters the network either from the top or the right N6 interface and will all be routed towards the BBU. In this case only one DNN is needed but it corresponds to two different N6 interfaces.
4.3.2.1.2 Control Plane

Besides application servers, also AFs can be MEC-hosted and use the T8, Rx, N33, and N5 interface to interact with the Core network (see Section 2.1.1). Corresponding implications are described in Section 4.3.3.3 around QoS prediction, but mentioned findings are also applicable for other control plane interfaces. These interfaces are specified for non-standalone and standalone 5G New Radio, except for the following one, which is most relevant in context of MEC.

The 5G Core introduces the “Application Function influence on traffic routing” interface. It is not available with non-standalone 5G New Radio as it is mostly intended to be used in conjunction with SSC mode 3 (see Section 4.3.2.2) only available with 5G Core. The interface can be used in two ways, that can also be combined. The first way allows the AF to influence the selection of the gateway to be used when a packet data network session is established. For the moment we have no reason to believe this has a role for CCAM. According to Section 4.3.2.1.1 geographical proximity, optionally combined with load balancing, are typically used to select a gateway. Further information from the AF does not have benefit given that all gateways provide access to a MEC-host running the required application server and public Internet hosting is used as backup where no MEC-hosted application server, or no MEC hosting at all, are available.

The second feature provided by this interface is information from the network about events corresponding to gateway changes. This allows to trigger application server changes based on events related to gateway changes (see Section 4.3.2.2). Section 4.3.2.3 further discusses how gateway and application server switching relate to each other and what triggers can be used to decide when to switch. This is further discussed per use case in Section 4.4.2.3. The main challenge is the granularity at which gateway and application server switching happen. An application server is typically only responsible to provide backend functionality for one use case. A session, which is affected by gateway switching, usually serves several use cases. Given this, it is currently unclear if and how CCAM services can benefit from this interface. This will be further investigated and published in Deliverable D3.3 planned for October 2021.

4.3.2.2 Session Continuity when Switching Gateways

According to Section 4.3.1.1.1 a packet data network connection to a gateway is triggered from the currently serving radio cell and then a gateway selection algorithm is invoked which in the simplest case corresponds to a fixed mapping between cells and gateways e.g., according to geographical or network-topological proximity (e.g., hop count). Further criteria like number of already served connections by the candidate gateways, for load balancing reasons, are also possible. A MEC host close to or collocated with that gateway would then be used to deliver the backend service.
Figure 4-31: A Vehicle Communicating with an MEC-hosted Application Server through a Gateway (P-GW-U) Close to it in a Non-standalone 5G New Radio Network with CUPS

Figure 4-31 shows the situation in a non-standalone 5G New Radio network with CUPS (see Section 4.3.2.1.1). The packet data network connection was established through the upper P-GW-U as gateway. The application client communicates with the application server on the MEC host closed to this gateway. Section 4.3.2.3.1 discusses how this application server can be discovered. Selecting one or more application servers and selecting the gateway to be used are independent tasks. Switching the gateway can be done at the same time as switching the server, but it does not have to. The figure shows one example of relations between S- and P-GWUs. Other combinations are possible as all nodes are normally within the same IP network and can communicate with each other. For example, the Edge P-GW-U could be using the same S-GW-U as the one used to access the public Internet. The control plane is not shown except for the MME which plays a role in signalling when gateway switching is being done. In the following, different variants of gateway switching are presented, first for non-standalone and then for standalone 5G New Radio.
Switching the gateway was originally not envisioned in the 4G EPC used for non-standalone 5G New Radio. Figure 4-32 shows how it can be achieved by triggering the gateway selection process again. The first subfigure corresponds to the situation shown in Figure 4-31 where the vehicle is served by the upper radio cell and communicates with the upper MEC-hosted application server through the upper gateway. The second subfigure then shows that the vehicle moved down and is now served by the lower radio cell but still uses the upper gateway and MEC-hosted application server. The vehicle then triggers a request to disconnect to the MME. This is triggered by the connection management software on the CCU and could for example exploit information about cell IDs or Tracking Area Codes. This information does not directly indicate if there is a closer gateway for the current serving cell than the one used at the moment. The connection management software would require further information about the network topology and this information would have to be kept up to date. The third step shows that after disconnecting, there is no MEC packet data network connection available. There is therefore an interruption of the communication unless the vehicle used a separate APN (see Section 4.3.2.1.1) to access the public Internet. The vehicle issues a connect request to the MME to establish a new packet data network connection. The gateway selection algorithm in this case selects the lower gateway, as shown in the fourth step. The application client then starts to communicate with the lower MEC-hosted application server.
Figure 4-33: Gateway Switching in Non-standalone 5G New Radio Network with Access to the Lower MEC-hosted Application Server before Gateway Switching

Figure 4-33 stresses the fact that server and gateway selection are independent from each other and can use different triggers. In this example the application client in the vehicle starts to communicate with the lower MEC-hosted application server while still using the upper gateway. This could be e.g., triggered by an application server algorithm based on the geographic position rather than being triggered through radio handovers and/or status of the packet data network connection.
Figure 4-34: Gateway Switching in Non-standalone 5G New Radio Network with SIPTO above RAN

Figure 4-34 shows gateway switching with a 4G EPC supporting Selected IP Traffic Offload (SIPTO) above RAN [14]. The first step is identical as before but now the trigger for gateway switching comes from the network, namely the MME. Upon radio handover with TAU the MME can decide to also trigger a gateway reselection. It can exploit network topology information for that, e.g., through the gateway selection algorithm it would execute and/or further topology information that could be configured and kept up to date from the OSS. The MME sends a disconnect request to the vehicle indicating that it should immediately reconnect. From step 2b on the same procedure as with vehicle-side triggered gateway reselection is executed. Correspondingly, there can be a period of no connectivity after disconnecting the packet data network connection and before establishing it again.
Figure 4-35: A Vehicle Communicating with an MEC-hosted Application Server through a Gateway (UPF) Close to it in a Standalone 5G New Radio Network

Figure 4-35 depicts the starting point in a standalone 5G New Radio network. As for the non-standalone case, it is only one example and many options exist how many UPFs are deployed and how they are interconnected. The example builds upon Section 4.3.2.1.1 that focuses on how traffic could be separated between MEC and public Internet, which is not the focus here. The SMF takes the role the MME had for standalone 5G New Radio.

The 5G Core used with standalone 5G New Radio supports the same principles like described above for the 4G EPC used with non-standalone 5G New Radio. Vehicle-side triggered gateway switching is called Session and Service Continuity (SSC) mode 1. Network-side triggered switching through SIPTO above RAN is called SSC mode 2.
Figure 4-36: Gateway Switching in Standalone 5G New Radio Network; Each Small Figure Corresponds to Figure 4-35 Except for Changing Data Paths According to the Black Arrow

Figure 4-36 shows SSC mode 3 that is only available with standalone 5G New Radio [7]. It is network triggered but instead of instructing the UE in the vehicle to disconnect and connect again it is instructed to issue another packet data network connection\textsuperscript{41}. This is then established to the lower gateway, as shown in the third subfigure. The connection to the upper gateway is also still open. Section 4.3.2.4 briefly discusses how the path for different traffic flows is decided. At some point the network can decide to instruct the UE in the vehicle to disconnect from the upper UPF. In reality it is not easy to decide when to do that, as further discussed in Section 4.3.2.3. Like for non-standalone 5G New Radio, switching the gateway and deciding when to use which MEC-hosted application server(s) and when to change application server are independent processes that can have independent triggers, as further discussed in the next section.

4.3.2.3 End-to-end Session Continuity when Switching MEC-hosts and/or Gateways

While session continuity (see Section 4.3.2.2) falls within the scope of 3GPP standardization, further challenges must be solved for end-to-end service continuity of applications and those are typically not subject to 3GPP specifications. In the following it is assumed that gateways towards MEC-hosts are deployed in the same way as those allowing public Internet access in today's mobile radio networks. This usually means carrier-grade Network Address Translation (NAT) and

\textsuperscript{41} Terminologies are slightly different in 5G Core specifications than for 4G EPC but for the sake of better understanding both generations in direct comparison we use the 4G EPC terms.
different IP address ranges per gateway assigned to UEs when the packet data network connection is established, as shown in Figure 4-37 with example\textsuperscript{42} IP addresses and ranges.

\textbf{Figure 4-37: Example MEC-enabled Network with two Gateways and Corresponding IP Address Ranges and NAT}

When switching the gateway, the UE will obtain a new network-internal IP address and will be visible towards application servers, incl. MEC-hosted ones, through the public IP address of the NAT router associated with that gateway. If NAT is not\textsuperscript{43} used with gateways towards MEC-hosts, the UEs would communicate with MEC-hosted application servers with their public IP addresses, as shown in Figure 4-38.

\textsuperscript{42} The used netmasks and corresponding IP network sizes (number of hosts within the same broadcast domain) are also just examples that likely do no correspond to real network deployments

\textsuperscript{43} Which would make it very unlikely that these gateways can also be used to access the public Internet. Vehicles would then need another packet data network connection for public Internet access, as described in Section 4.3.2.1.1.
In any case, with or without NAT, the IP address visible towards the MEC-hosted application server changes if the gateway is switched. In case of SSC mode 3 (see Section 4.3.2.2) the UE obtains another IP address when establishing a packet data network connection through the second gateway. It depends on the routing policies in the CCU (see Section 4.3.2.4) through which gateway the MEC-hosted application server is reached and what IP address it therefore sees as source IP address of the packets it receives from the vehicle.

Many applications are capable to deal with changing source IP addresses of clients. This is therefore discussed in more detail under the use cases in Section 4.3.4. A common solution is to keep required state information on the client side and avoid using the source IP address to identify a client. One common example are authentication tokens provided by the client with every request instead of associating authentication information to a specific IP communication flow between client and server. This can lead to increased processing delays as authentication algorithms need to be executed more often. The cloud native “Service mesh” paradigm described in Section 4.3.1.3 helps to relax the problem. It is intended to enable many different microservices, each with different IP address, to respond to client request\(^{44}\). In this case IP addresses change on the

\(^{44}\) Which corresponds to a request/reply communication pattern. It is an open challenge how this can also be applied to publish/subscribe patterns as e.g., used with MQTT
server-side, but it equally promotes the paradigm the avoid stateful (also called “context”) information at the different servers so they can each handle a request.

In case TCP is used, a changed source IP address would result in corresponding TCP connections to abort, as the source IP address is part of the unique identifier of a TCP connection. According to Section 4.3.2.2 the vehicle could reach a target MEC-hosted application server before switching the gateway to e.g., already establish a TCP connection, but this connection would abort after switching the gateway and obtaining a new client IP address. Variant of TCP like Multipath TCP [74] and alternative transport protocols like QUIC\footnote{https://quicwg.org/} do not use source IP addresses to identify connections but instead use unique IDs. This way two things can be achieved: an application client can connect to a MEC-hosted application server before switching the gateway and continue using that connection after switching it. Furthermore, the application client in the vehicle could still be able to receive from the previous MEC-hosted application server through its new IP address. QUIC and Multipath TCP have protocol-specific means to make the application-server-side connection endpoint aware of the IP address change. Simply speaking the application server would realize that it is receiving data with an already known unique ID but from a different source IP address. It will therefore conclude that it is the same connection but the IP address of one connection end point has changed and/or, in case of Multipath TCP, another sub-flow with different IP address was added to an already existing TCP connection.

Above solutions, together with the session continuity features described in Section 4.3.2.2, can be triggered by different events. Table 4-9 lists different steps that are executed when switching MEC-hosted application servers and gateways. Examples for triggers are provided, but not for all steps as for some steps it is still being researched (See Section 6.2).

Table 4-9: Session and Service Continuity Related Events and Examples for Triggers

<table>
<thead>
<tr>
<th>Step</th>
<th>Event</th>
<th>Example Triggers</th>
<th>Application</th>
<th>Gateway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Server</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Old</td>
<td>New</td>
</tr>
<tr>
<td>0</td>
<td>Precondition</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Contact new server before gateway/server switching (e.g., to establish a transport)</td>
<td>To be determined if possible</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
<td>Trigger Conditions</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Disconnect from old gateway (SSC mode 1 and 2 only)</td>
<td>CCU trigger (SSC mode 1); MME/SMF trigger (SSC mode 2); see Section 4.3.2.2 for more detailed examples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Connect to new gateway</td>
<td>CCU trigger (SSC mode 1); MME/SMF trigger (SSC mode 2 and 3); see Section 4.3.2.2 for more detailed examples</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Use new application server</td>
<td>Detecting that a new gateway is being used (as prerequisite, further conditions can also be used)</td>
<td>X&lt;sup&gt;46&lt;/sup&gt; X&lt;sup&gt;47&lt;/sup&gt; X</td>
<td></td>
</tr>
<tr>
<td>5,6</td>
<td>Stop using old application server</td>
<td>To be determined; if Step “6,5” is done first, it can serve as (part of the) trigger</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6,5</td>
<td>Disconnect from old gateway (SSC mode 3 only)</td>
<td>To be determined</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

### 4.3.2.3.1 Application Server Discovery

Switching the application server requires a method to identify an appropriate application server. Three methods have been identified for that. The first one is maintaining a mapping of geographical areas to MEC-hosted application servers. This would require information from the network, as also discussed for gateway switching with SSC mode 1 (see Section 4.3.2.2). Geographical coordinates make limited sense, as the applications server decision is related to the serving cell and closest gateway allowing access to MEC-hosts associated with that cell. Cell IDs therefore make more sense and even just considering Tracking Area Codes should be sufficient. For the particular case of cross-border / -MNO scenarios it would even work to start by considering the network ID. Besides obtaining the required information from MNOs and keeping it up to date, a further challenge comes from IDs and ID changes only being visible in the CCU.

<sup>46</sup> Only if Step 1 was executed

<sup>47</sup> Only with SSC mode 3

<sup>48</sup> If old server can be accessed through new gateway or SSC mode 3 is used; with SSC mode 3 the old or new IP address can be used, depending on CCU routing policies. With SSC mode 1 and 2 only the new source IP address can be used.

<sup>49</sup> Already disconnected in case of SSC mode 1 and 2. Step “6,5” can be done before “5,6” for SSC mode 3 and in this case the old gateway is already disconnected at this point.
while the logic to select an application server is associated with the application clients that might run on other units, as further described in Deliverable D3.1 [3].

The second solution is IP Anycast [75] where the same IP address is always used to address the backend application server but depending on e.g., the used gateway and further involved IP routers different application server instances, all with this anycast IP address, are being reached. The anycast IP address for a given service can be preconfigured, obtained by one of the mechanisms described above, especially if different anycast IP addresses are used in different MNO networks or provided through DNS, as described further below. This solution can only be used if the IP routers along the path support it. It is transparent for the client application and in case no MEC-hosted application server replies the request can automatically propagate to an application server hosted on the public Internet.

The third solution is a set of solutions all based on DNS. Three prerequisites are required for this to fork. First, it must be assured that DNS queries are actually executed at the times when application servers should be changed. This depends on the application client implementation and the operating system providing DNS resolution services. Second, the reply for the request must come from a DNS server capable to participate in the application server discovery task. Third, the DNS server must have further information to decide what IP address to provide for a request coming from a certain vehicle in a certain location served by a certain radio cell and using a certain gateway. The following text will mainly focus on options for the second and third challenge. The first challenge, to assure that there actually is a DNS query when the application server should be changed, is very use case dependent and therefore discussed in Section 4.3.4.

The AECC Technical Report [66] (see also Section 4.3.1.2) distinguishes three concepts to assure location dependent DNS replies pointing at the most suitable MEC-hosted application server for a given vehicle location and/or serving cell. The DNS system normally assures that DNS requests are forwarded to a server capable of providing a reply. This can cause extra delays, so it is favourable to use a DNS server capable of resolving FQDNs of MEC-hosted application servers directly. This can be achieved by providing the DNS server IP through DHCP or other means of IP address configuration used in the network. This does not work if the operating system is configured to use other DNS servers than what is provided through DHCP. Alternatively, DNS requests can be rerouted, e.g. in the NAT routers. With increasing adaptation of DNS security features this method will likely become less effective. A recently concluded 3GPP study [76] provides more insights on how the mobile network can provide DNS IP addresses to clients.

Once it is assured that the right DNS server is used it must be enabled to provide different replies depending on vehicle location and/or serving cell. DNS requests do not directly contain information about location or the serving cell. Such information needs to be determined otherwise. There could be different DNS servers configured for different gateways serving the packet data network connection intended to reach MEC hosts. This is possible, as establishing a packet data network connection to a gateway also contains IP configuration, e.g. through DHCP, and a

---

50 Or more coarse-grained information like Tracking Area or Mobile Network Code
corresponding DNS server address could then also be provided or the gateway or corresponding NAT router could forward all DNS requests to an appropriate DNS server. The drawback is that this would require a DNS server for each MEC location, which usually corresponds to edge gateway locations. Alternatively, a more central DNS server could exploit the source IP address of the DNS request to determine the gateway serving the vehicle. This requires that the DNS request is sent through the gateway and, if deployed, corresponding NAT router, that is also used to access MEC hosts. The DNS server would require current information on which MEC-hosted application server IP address should be returned depending on which source IP was used for the DNS request. Such configuration can be conducted by interaction between network OSS/MANO and MEC MANO, as described in Section 4.3.3.2. The DNS server can be within control of the MNO and in this case requires information from the MEC MANO system. The DNS server can also be within control of a service provider operating the corresponding application servers and/or the entire MEC-hosting MANO. In this case information from the network OSS/MANO system would be needed to map source IP addresses of DNS requests to appropriate responses for the closes MEC-hosted application server.

Alternatively, the DNS server can exploit further information provided by the network through the UE location interface available for non-standalone 5G New Radio through SCEF [77] and standalone 5G New Radio through NEF [78] or through Core network internal interfaces towards the MME (4G EPC) or SMF (5G Core) if the DNS server is within MNO control. The DNS server would then receive information about the serving cell ID and/or Tracking Area of a vehicle and can exploit it for an appropriate DNS response. In case of standalone 5G New Radio also information about changes in the gateway as available through the “AF influence on traffic routing” interface (see Section 4.3.2.1.2) can be used.

Currently, exploiting the source IP address rather than the UE location interface appears to be the solution with less complexity.

4.3.2.4 CCU-side Network Interface and Route Selection
Several sections above discuss the challenge of CCU-side decisions on which network interface to use, if several are available. This can result from methods applied to distinguish between MEC and public Internet related data traffic (see Section 4.3.2.1.1) and/or from applying SSC mode 3 for gateway switching (see Section 4.3.2.2 and 4.3.2.3). It also applies for the special case of using DNS for application server discovery where some solutions require DNS requests to be sent through a certain network interface with corresponding source IP address and route (see Section 4.3.2.3.1). In case of non-standalone 5G New Radio the IP routing table, managed by the OS, plays a key role for network interface selection and routing. Besides that, applications can also influence which network interface and corresponding source IP address should be used. The solution space is further extended as different modems apply different means to expose themselves to the operating system if several logical channels, e.g. corresponding to different

51 Which could be controlled by the same MNO, but does not have to
sessions through different gateways, exist. So, in case of non-standalone 5G New Radio the application and CCU developer might require further insights what is enabled on the network side.

For standalone 5G New Radio UE Route Selection Policy (URSP) was specified [7] as common solution for the network to steer traffic routing on the client device through defined filters and procedures to deploy and update those filter rules on the device. By the time of writing this document no information on how this is used in modems and OSs was publicly available. URSP must be consistently propagated from the modem to the OS, e.g. to the IP routing tables. Without these insights it was not yet possible to describe the potential role of URSP in this architecture. This is planned for Deliverable D3.3 scheduled for October 2021 as more information from modem and OS vendors should be available by then. It must especially be evaluated within the common application architecture described in Deliverable D3.1 [3] where it cannot be assumed that application clients run on the CCU but are separated through an in-vehicle network.

4.3.2.5 Cross-MNO Inter-MEC Communication
MEC hosts often have non-public IP addresses and therefore cannot interact when belonging to different networks as is the case when trying to interact between MEC hosts from different MNOs. This can be solved through firewalls and similar security gateways that are usually anyway already present. But the result would be loss of end-to-end control of the path as communication between MEC hosts would be realized through the public Internet only providing best effort service, as shown in Figure 4-39.

![Figure 4-39: MEC Hosts Interacting through the Public Internet](image)

Three solutions are therefore proposed.

The first solution is showed in Figure 4-40 where the MEC hosts were moved out of the MNO data centers, also referred to as Points-of-Presence (PoPs), and moved to a colocation data centre. Appendix B provides more information on what “colocation” means and what other types
of data centres exist. The colocation data centre provider can assure controlled connectivity between the MEC hosts within the data centre through “cross connect” (see Appendix B). End-to-end QoS control also requires controlled connectivity between the gateways (depicted as “GW” in the figure) and the colocation data centre. This can be realized by WAN providers as inter-site connectivity or through Blended IP (see Appendix B). The figure shows one WAN provider, but different ones can be used. It is also not precluded that WAN and colocation data centre provider are the same entity. Furthermore, involved MNOs can also have the role of colocation data centre or WAN provider or both.

Figure 4-40: MEC Host Deployment in Shared Data Centres

Figure 4-41 presents a variant of the previous solution where not only the MEC platforms but also the gateways used to reach them are located in a colocation data centre. The precise realizations and involved transport network and WAN providers required to realize that for each MNO can be different. But it is a generally valid assumption that MNOs can deploy parts of their core network, in this case gateways, in any data centre, also a colocation one and that these gateways are connected through controlled IP networks.

52 “Controlled” means that measures to achieve certain QoS, e.g., latency or throughput, can be applied. This is not considered possible in the public Internet. It can be possible on the same physical infrastructure realizing the WAN enabling the public Internet (see Figure 4-42) but it is assumed that such measures are not active and the public Internet remains a best-effort service.
Different than the previous solutions, Figure 4-42 presents an approach not requiring a colocation data center. In this example the MNOs obtain inter-site connectivity or Blended IP services from one or more appropriate WAN providers to obtain a better-than-best-effort connection between their PoPs. Firewall and security gateways are present in all described solutions but here they are particularly highlighted to stress the fact that each MEC platform usually resides within its own domain, e.g. an MNO-specific LAN. Communication between the MEC applications requires more network engineering than just mutual knowledge of IP addresses, as those addresses are usually not accessible outside of the respective MNO domains.

Three solution were presented how end-to-end QoS can be maintained when more than one MNO and its respective MEC platform is involved for a service. The findings are equally applicable if only one MEC platform is involved but more than one MNO. There is no preference among the solutions. Their applicability depends on individual MNO properties like existing deployments and further services they offer (e.g., if they are also in the data center and/or WAN business) and what the services from other stakeholders (shared data center and WAN provider) cost.
4.3.3 Relation to other Key 5G Solutions

MEC has architecture implications on many of the key 5G solutions described in other sections. Those are described in the sections below.

4.3.3.1 Network Service Continuity at Country Borders

According to Section 2.1.2 there is no uninterrupted cross-border / -MNO service continuity currently deployed and the CCU of a vehicle would at some point let go off the network of the previous country and then, after scanning, select a new serving network in the country just reached, when crossing a border. At this point the packet data network connection can be established again. Today, Home Routed Roaming is typically applied but the network can be configured to use Local Breakout Roaming and by this also provide access to MEC-hosts of the serving MNO even in roaming conditions. When enabling cross-border / -MNO handover, as described in Section 4.1.1, there is no initial registration and packet data network establishment when entering the new network but just a handover that, from vehicle point of view, is not\(^{53}\) different from a handover within the same MNO. As with every radio handover, the gateway used to reach the MEC-host remains unchanged. This then corresponds to Home Routed Roaming. Achieving Local Breakout Roaming would require switching the gateway. This can be done with the same methods as available for gateway switching within the same MNO network. Those are described in Section 4.3.2.1.2 where it is also stated that only SSC mode 3 would provide an uninterrupted switching of gateways.

---

\(^{53}\) Different kinds of handovers exist within the same MNO network. S1-handover with change of MME within the same MNO is what is closest to cross-border / -MNO handover. This equally applies for standalone 5G New Radio with corresponding interface and 5G Core nodes. It is an open question, to be determined through trials, how long the service is interrupted during handover and if it is different between cross-border / -MNO handover and S1-handover within the same MNO.
### 4.3.3.2 Network Orchestration and Control

Section 4.1.2.1 describes life cycle management of software components including VNFs and MEC-hosted application servers across multiple MNOs, potentially serving different countries. This section focusses on the aspect of jointly managing VNFs and MEC-hosted application servers. Section 2.1.3 describes today's baseline of mobile radio networks being deployed as interconnected VNFs typically running in VMs and managed by corresponding MANO systems like OSM or OpenStack. This leads to the solution of using the same MANO system for MEC-hosted application servers to be deployed in VMs. According to Section 4.3.1.3 the Cloud Native paradigm uses Containers instead of full VMs to deploy application servers. This leads to a solution where two different MANO systems are used, one for VMs to deploy the VNFs for the mobile radio network and one for Containers for MEC-hosted application servers. Kubernetes is usually used as MANO system for Containers. In yet another evolution, VNF deployments are also done according to the Cloud Native paradigm which again leads to the situation of VNFs and MEC-hosted application servers using the same kind of MANO system, namely Kubernetes. It does not necessarily have to be the same MANO system, as MNOs might decide to use one system for VNFs while partnering with a Hyperscale cloud provider for MEC-hosting. In this case, the MANO system of that cloud provider could be reused but at least for connectivity (see Section 4.3.2.1.1) a coordination between the MANO systems would be needed.

**Table 4-10: Different Options for VNF and MEC-hosted Application Server MANO**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>VNF MANO</th>
<th>MEC-hosted Application Server MANO</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM-based for VNFs and application servers</td>
<td>VM-based (e.g. OpenStack-based)</td>
<td>Same as for VNFs (VM-based)</td>
</tr>
<tr>
<td>VM-based for VNFs, Container-based for application servers</td>
<td>VM-based (e.g. OpenStack-based)</td>
<td>Container-based (usually Kubernetes-based)</td>
</tr>
<tr>
<td>Container-based for VNFs and application servers, separate MANO systems of same kind</td>
<td>Container-based (usually Kubernetes-based)</td>
<td>Container-based (usually Kubernetes-based)</td>
</tr>
<tr>
<td>Container-based for VNFs and application servers, common MANO system</td>
<td>Container-based (usually Kubernetes-based)</td>
<td>Same as for VNFs (Container-based)</td>
</tr>
</tbody>
</table>

Table 4-10 summarizes the different options for VNF and MEC-hosted application server MANO. Not every possible combination is included, as it would be e.g., very uncommon to have separate VM-based MANO systems for the same option. This option is motivated by reusing the already deployed VM-based MANO system also for MEC-hosted application servers. While having the
same (kind of) MANO system might be an advantage as it reduces system complexity, it can have the drawback of using a solution that is not specialized for the task. Kubernetes is very mature and feature rich for Cloud Native application server deployments while VM-based solutions like OSM or OpenStack evolved over years to meet specific requirements of telco industry, especially with regard to supporting different SDN control interfaces. The fourth solution is, to the best of our knowledge, currently not deployed. It is currently an open question how the two MANO systems of the third solutions would cooperate. There could either be another component on top of the two MANO systems coordinating them or one MANO system would instruct the other MANO system how to deploy and interconnect components. It would then be open which system is the instructing one and which one receives the instructions.

It is currently not possible to further narrow down the solution space and identify most likely scenarios as this is not so much a technical as it is a question on migration from legacy systems, partnership between MNOs and cloud providers and evolution of the MANO systems and their interplay.

4.3.3.3 QoS Prediction

Section 4.2.2.1.2 discusses deployment options to realize QoS prediction. 3GPP Core specification for non-standalone [14] and standalone [7] 5G New Radio distinguish between trusted AFs considered to be “within MNO domain” and untrusted ones outside of that domain. This is not related to MEC-hosting and it cannot be automatically precluded that MEC-hosting automatically means being trusted. It is still easier to create such a trust relation towards a MEC-hosted AF as it is typically collocated with Core network components and a trusted end-to-end network connection can be established. AFs used as interface towards the network for QoS prediction hosted on the public Internet would require enhanced security measures and the MNO might generally decide that no AF hosted on the public Internet can be considered to be trusted.

When trusted, the AF could also directly interact with the NWDAF which contains the PF. According to Section 4.2.2.1.2 this has no impact on the performance of QoS prediction and is rather a design choice of the MNO if a NEF, as required for untrusted AFs, must be present between NWDAF and AF. The NEF would need to implement and expose all interfaces available at the NWDAF to assure all QoS prediction features are available to the AF and from there to the AS.

4.3.4 Use Case Specific Solutions

Previous generally applicable assessment of the state of the art will be discussed per use case in the following sections.

---

54 Not precluding that an MNO could at the same time also be Cloud Provider
4.3.4.1 Tele-operated Driving
MEC for ToD is not explicitly evaluated, according to Section 3.3.1, but for example discussed in terms of possibilities to expose QoS prediction services (see Section 4.3.3.3).

4.3.4.2 HD Mapping
User Story 1 and 4 follows a request/reply communication pattern to download map tiles. For User Story 3 a permanent TCP connection towards the HD map content server is used to inform vehicles about map updates that were just received from other vehicles (according to User Story 2) and trigger their download through a request/reply communication pattern. So, User Story 3 is a backend-triggered request/reply communication pattern and the solution must assure triggers are not lost. In the following, MEC-hosted application server and gateway switching are described for the HD Mapping Use Case. It is not intended to trial this within WP4, according to Deliverable D4.1 [5] and D4.2 [4]. For trials, the focus lies on comparing the performance gain from hosting HD map content closer to the vehicles.

The application client in the vehicle can easily determine when a request for a map tile according to User Story 1 or 4 is about to happen and knows when a download is ongoing. It cannot anticipate when it will get a trigger to download a recently updated map tile according to User Story 3. In the following, we assume that for User Story 3 uploads of HD map deviations are received by one MEC-hosted application server and propagated to neighbouring ones. The vehicle is normally served by a single MEC-hosted application server and only during transition from one server to another two MEC-hosted application servers could be used for a short period of time.

Figure 4-43 shows an example scenario in which applied MEC solutions for the HD Mapping Use Case will be described. It includes three MEC-hosted HD map application servers within two MNOs. In reality, there would be more than two MEC-hosted application servers with the upper MNO and the lower one would have more than one, but those are not shown. Any intra-MNO procedures described for the upper MNO are equally valid for the lower one. It is assumed that only one packet data network session through one edge gateway is used and that this connection also allows to access the public Internet.

The UE established a packet data network connection to the upper gateway and the application client is connected to the corresponding MEC-hosted HD map application server. The vehicle is driving downwards and will be handed over to the lower cell of MNO 1 and then to MNO 2. It is assumed that both MEC-hosts can be reached through any gateway of MNO 1 but the MEC-host at MNO 2 can only be reached through MNO 2 gateways and it is not possible to reach MNO 1 MEC-hosts when served by MNO 2. This is a conservative assumption. In case MEC-hosts are also reachable from other MNOs through the public Internet, solutions would become less complex as sudden changes of reachability are avoided, just like for the intra-MNO case.
Figure 4-43: Example Scenario for HD Mapping Use Case with Gateway and MEC-hosted Application Server Switching within the Same MNO and Cross-border / -MNO

Table 4-11 lists the steps required to switch between the three gateways and corresponding MEC-hosted application servers as the vehicle moves from top to bottom. It should not be precluded
that it could be done differently. In this example, no pending transmissions are aborted, which requires an interface between CCU and application client to negotiate the right point of time to switch the gateway. Alternatively, switching can be done independent of application state and aborted transmissions get restarted with the new MEC-hosted application server once switching is completed. The application client would at least need a mechanisms to detect when the connection is available again, e.g. by regularly sending ping messages. It must start with a DNS request and execute required initial connection steps toward the HD map application server if needed.

Table 4-11: Events and Triggers for HD Mapping Use Case Gateway and MEC-hosted Application Server Switching Intra-MNO and Cross-border / -MNO

<table>
<thead>
<tr>
<th>Step</th>
<th>Event</th>
<th>Trigger</th>
<th>Application Server</th>
<th>Gateway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MNO 1</td>
<td>MNO 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Upper</td>
<td>Lower</td>
</tr>
<tr>
<td>0</td>
<td>Precondition</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>Disconnect from MNO 1 upper gateway</td>
<td>Change of Tracking Area (vehicle triggered / SSC mode 1)(^\text{55}) \footnote{SIPTO above RAN / SSC mode 2 (network triggered) not suggested as it would not allow to wait for pending transmissions to finish before disconnecting}</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>Connect to MNO 1 lower gateway</td>
<td>Disconnect completed (see row above)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Connect to HD map application server (IP address discovered through DNS request)</td>
<td>Connect completed (see row above)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

\(^\text{55}\) SIPTO above RAN / SSC mode 2 (network triggered) not suggested as it would not allow to wait for pending transmissions to finish before disconnecting
The transition between the MNOs happens through a cross-border / -MNO handover as described in Section 4.1.1. On the vehicle side it is just indicated through a change of Tracking Area Code and/or Mobile Network Code. The connection manager in the CCU would then align with the application client to determine when ongoing transmissions are over and prevent new ones from starting and then trigger disconnect and connect\(^{56}\) of the packet data network connection. The gateway selection algorithm used in the Core network of MNO 2 will assure a local gateway is assigned, corresponding to Local Breakout Roaming. The CCU will inform the application client that the connection is up again, or the application client will determine this through a mechanism that polls if connectivity is available. A DNS request will be executed to discover a MEC-hosted application server. The IP address returned belongs to the IP range used for MEC-hosts. IP routing after the gateway (from vehicle point of view) will assure that data is routed to the MEC-host, not the public Internet.

The procedure is very similar with standalone 5G New Radio and SSC mode 3 but can be completely transparent to the application client. Connecting to another gateway is triggered from the network and in the meantime the application client can continue to communicate with the upper application server through the upper gateway. The DNS cache must be purged after establishing the connection to the second gateway. Routing on the CCU must be adjusted to assure all DNS requests have the source IP address corresponding to the second gateway and it must be assured that the application client always uses DNS before communicating with the server. As the DNS cache was purged, the next request, e.g. for a tile download, will first trigger a DNS request. The DNS reply will contain the IP address of the lower MEC-hosted application

\(^{56}\) During transition, when the vehicle had no connectivity, tiles might have been updated but the trigger to download them was not received. Instead, the new HD map server is requested for information about such updates in the very recent past during the time of no connectivity

\(^{57}\) It is assumed that MNOs use the same APN. If not, the CCU would need a mapping of Mobile Network Codes to APNs. So, aligning on common APNs, e.g. within GSMA, is very beneficial to prevent the need for such a mapping.
server because this is the configured answer when receiving a DNS request from a source IP address associated with the lower gateway. Additional application client logic might be required to determine that now a different IP address than before is used and to execute required initial steps with the new server, e.g. for authentication. From this point on the lower MEC-hosted application server is used to request tile downloads (User Story 1 and 4), to upload detected map tile deviations (User Story 2) and to be informed about pending HD map tile updates that should be downloaded immediately (User Story 3). Ongoing transmissions can continue through the upper MEC-hosted application server, but new ones will go through the lower one.

It is an open question how to decide when to release the packet data network connection to the upper gateway. The HD map application might stop to transfer data at some point but there might still be communication from other applications and/or TCP keep-alive transmissions ongoing.

The above equally applies for gateway and MEC-hosted application server switching across MNOs under the assumption that SSC mode 3 can be used across MNOs.

4.3.4.3 ACCA
For the ACCA use case a vehicle is usually served by one MEC-hosted Geoservice. Alternatively, especially if no MEC-hosted Geoservice is available because the MNO provides no MEC-hosting or no Geoservice backend is deployed, a Geoservice collocated with the TMS on the public Internet can be used. It can provide all services the MEC-hosted Geoservice provides, but higher latencies may be experienced. Changing the Geoservice is required when entering the area served by a different MEC-host and when changing the serving MNO, e.g. when crossing a border. The Geoservice is used to send reports about detected hazards and to receive notifications about hazards. More details are provided in Deliverable D3.1 [3].

Figure 4-44 shows almost the same example scenario as for the HD Mapping Use Case in Section 4.4.2.3.2. An essential difference is the proposal that for public Internet access a separate gateway should be used, which is achieved through another packet data network connection with a different APN. This corresponds to the default APN typically called “default” or “internet” or an empty APN string. The routing table in the CCU is set to use this as default gateway. The communication characteristics of the HD Mapping Use Case allow for a temporary loss of connectivity when switching gateways without SSC mode 3. For the ACCA Use Case this cannot be accepted as there is no way to know when the vehicle might need to send a hazard report and especially not when the backend sends a hazard notification to the vehicle. Using a separate gateway for public Internet access assures that there is always one Geoservice available even during temporary loss of connectivity when disconnected from one gateway but not yet connected to the other one.
Figure 4-44: Example Scenario for ACCA Use Case with Gateway and MEC-hosted Application Server Switching within the Same MNO and Cross-border / -MNO
Table 4-12 presents an example how gateway and MEC-hosted application server switching for the ACCA use case can be realized with non-standalone 5G New Radio, which does not provide SSC mode 3 or with standalone 5G New Radio if SSC mode 1 or 2 is used.

### Table 4-12: Events and Triggers for ACCA Use Case Gateway and MEC-hosted Application Server Switching Intra-MNO and Cross-border / -MNO

<table>
<thead>
<tr>
<th>Step</th>
<th>Event</th>
<th>Trigger</th>
<th>Application Server</th>
<th>Gateway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MNO 1</td>
<td>MNO 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Upper</td>
<td>Lower</td>
</tr>
<tr>
<td>0</td>
<td>Precondition</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>Disconnect from MNO 1 upper gateway</td>
<td>Change of Tracking Area detected in CCU (vehicle triggered / SSC mode 1) or trigger from network (SIPTO above RAN / SSC mode 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Connect to MNO 1 lower gateway</td>
<td>Disconnect completed (see row above)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>Connect to lower Geoservice (IP address discovered through DNS request)</td>
<td>Connect completed (see row above), to be further evaluated what triggers are possible with SIPTO above RAN / SSC mode 2</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Intra-MNO gateway and application server switching complete**

| 4-6 | Same as for intra-MNO; instead of Tracking Area change, also Mobile Network Code change can be used as | X | X |
The ACCA client application would have to assure that reports that have been detected while switching gateways and Geoservices are not sent before the connection to the MEC-hosted Geoservice has been established. This could be negotiated with the CCU in case of vehicle triggered / SSC mode 1 gateway switching. Else, especially for the RSA-flavour, where UDP is used, hazard reports could be lost but this would be concealed as the same report is sent again, periodically. For the PSA-flavour using TCP it could be eventually detected that the report cannot be sent, but this timeout can take very long. Any pending hazard notifications should be immediately received when connected to the new Geoservice according to the same procedures that are used when a vehicle enters a new area (updates its ROI) to receive hazard notifications it did not get before because it was not subscribed to that area.

Application server discovery is challenging in this scenario, because it must be assured that DNS requests are not routed through the network interface a gateway providing public Internet access. Routing tables must be set in a way that all MEC-related data traffic is routed through the network interface and with source IP address associated with edge gateways. This can be done if MEC-related data traffic always relates to a certain IP address range that would then be configured accordingly in the routing table. The challenge is how to obtain this information from the MNO to set routing tables accordingly. The DNS server must be within this IP range to assure DNS requests are sent with the right source IP address so that the replies correspond to the correct Geoservice depending on which gateway is used. It will be evaluated if and how URSP (see Section 4.3.2.4) can be used to simplify the solution.

With standalone 5G New Radio and SSC mode 3 there is no need to have two packet data network connections in order to have public Internet access as backup solution. Furthermore, no special information exchange between CCU and application client is needed to prevent sending hazard reports to MEC-hosted Geoservices while gateway switching is ongoing.

It is challenging to trigger the application client to establish a connection to the new Geoservice after switching gateways. In case of vehicle triggered / SSC mode 1 switching the trigger is in the vehicle and the CCU could inform the application client when switching is done. In the other cases the trigger comes from the network and it is up to modem and OS implementation and configuration how it is exposed to the CCU.

---

58 At least not to detect loss of connectivity due to gateway switching.
It is assumed that the same mechanisms can be applied for intra-MNO and cross-border / -MNO switching under the condition that SSC mode 3 also works across MNOs.

4.4 Precise Positioning

According to Section 3.4, a combination of GNSS and vehicle sensors can satisfy most of the minimum requirements of the different use cases. However, in some locations where satellites cannot be easily tracked, like urban or rugged areas, it could lead to inaccuracies of several meters and significant latencies. In addition, more precise accuracies will enhance the general performance of all the use cases. Thus, in this section, two precise positioning solutions are described. The first one uses Ultra-Wideband (UWB) [79] radio technology to enhance global positioning, while the second one is based on 5G New Radio sidelink to provide relative position information between vehicles.

Two main requirements and/or limitations should be considered when implementing the UWB positioning solution. First, UWB anchors have to be deployed close to the road. Thus, regarding the test sites, the most adequate locations for this technology will be cities, as is the case of the Barcelona testbed, or specific areas of the highway (e.g., entrances/exits, tolls, accident blackspots, cross-borders, etc.). On the other hand, as common solutions based on Two-way-Ranging (TWR), it needs to exchange at least three messages between the tags and the anchors, the periodicity of the measurements is limited by the bandwidth of the technology; i.e., by the number of vehicles using the service in an area. Thus, this type of solution seems more suitable for use cases where a precise location is needed occasionally and can be triggered by an event (e.g., obstacle detection or ToD activation). Additional details on the application of the UWB solution to the different use cases can be found in Section 4.4.1. Regarding the relative positioning solution based on 5G New Radio sidelink, it is particularly interesting for the ToD use cases, where it can facilitate coordinated manoeuvres and track the movement of neighbour vehicles. Also, it can be combined with a precise global positioning mechanism to obtain the global location of a group of vehicles (e.g., if not all vehicles in the area are able to obtain an accurate location from GNSS or other positioning system). Section 4.4.2 provides more details on the usage of this solution in the different use cases.

4.4.1 Enhanced Vehicle Positioning Based on UWB

The proposed enhanced positioning system is based on the hybridization of different technologies to provide a more accurate location of the vehicles. The objective to this system is to determine the position of vehicles in the street with lane accuracy. With this aim, three different technologies are proposed, which can complement each other. The first two technologies, GNSS and inertial systems, are described in Section 2.4. This solution combines them with UWB relative positioning.

Using UWB technologies makes it possible to measure distances with a precision of several centimetres based on the Time of Flight (TOF) of messages sent between two endpoints. Thus, knowing the distance from the device to be located, referred to as tag, to three or more reference devices, referred to as anchors, it would be possible to calculate the relative position of the tag through trilateration. Anchors are considered static references, which, thus, can be accurately
geopositioned by professional topographers to minimize errors. Hence, UWB systems would require the deployment of dedicated anchor devices on the road but this would serve to improve the precision of GNSS technologies in use cases which demand lane accuracy. This supports the 5GCroCo use cases (see Section 3.4) and further ones like e.g., prioritizing emergency vehicles such as ambulances and other blue light services.

This enhanced positioning system is envisioned to be used in conjunction with vehicular communication to provide the position of the vehicles whenever necessary. Figure 4-45 shows an example on how the enhanced positioning system could be employed to facilitate the priority access of an ambulance in a crowded road. In this case, UWB anchors would be installed as an element of the road infrastructure; for example, they could be deployed in street lamps (depicted as yellow circles in the figure) and traffic lights at street crossings. First, the ambulance will accurately locate itself using the enhanced positioning system to determine its position with lane accuracy. This information will be sent by the CCU. Secondly, vehicles in the path will be notified and will also accurately locate themselves and for this benefit from the UWB architecture. Finally, vehicles in the lane of the ambulance (depicted in red) could proactively move to a different lane before the ambulance arrives. Also, vehicles in the crossing would stop to let the ambulance pass through. Vehicles not interfering with the path of the ambulance (depicted in blue) will not need to change their lane though they might be warned to enable the incorporation of other vehicles in their path.

Figure 4-45: Enhanced Positioning System Based on UWB to Facilitate the Driving of Priority Vehicles
4.4.1.1 Architecture
The architecture of the enhanced positioning system relies on two elements (shown in Figure 4-46):

- **Locating unit**: This is the equipment that would be installed in each vehicle to allow its enhanced positioning. This unit is composed of the following modules:
  - Positioning technologies: As mentioned previously, three types of positioning technologies will be considered: a GNSS receiver supporting an update rate of 1 Hz; a nine-axis inertial measurement unit and a UWB tag. The details of the UWB tag will be described in the next section.
  - Main processor: Runs the enhanced positioning algorithms and controls the different positioning technologies.
  - Communication module: Transmits the position information to the vehicle CCU or to an external platform. The CCU will insert this position information into Cooperative Awareness Messages (CAMs), Decentralized Environmental Notification Messages (DEMNs) or Cooperative Perception Messages (CPMs).

- **UWB infrastructure**: As mentioned previously, the infrastructure devices are supposed to be installed in lampposts or traffic lights and their positions must be well-known and accurate. The devices consist on the following modules:
  - The UWB anchor
  - Processor: This module is optional and could be included to provide higher computing resources to the device.
  - Communication interface: Optionally, a communication interface could be included to transmit information to an external platform.
The following section summarizes the main details of the UWB implementation proposed for the 5GCroCo project.

### 4.4.1.2 UWB Ranging System

The UWB ranging system is based on the radino32 module that integrates a DW1000 Integrated Circuit chip from DecaWave. It enables decimetre-level accuracy for positioning systems and can complement other system approaches.

The key features of this technology are: It allows high localization precision, does not interfere with other radio communication systems, is resistant to multipath signal propagation and noise, and uses low power transceivers.

Unlike traditional radio technologies (like Bluetooth or Wi-Fi), UWB operates with the so-called TOF of the radio signal rather than RSSI, which makes the technology much more precise and enables it to obtain very exact ranging measurements. Distance is obtained as:

\[
\text{Distance} = \text{TOF} \times \text{speed of light}
\]

The TOF is calculated using the DecaWave Two Way Ranging Protocol [80] between two modules based on UWB Institute of Electrical and Electronics Engineers (IEEE) 802.15.4a standard. There are three messages exchanged between the UWB tag and the UWB anchor in order to get a precise measure of the TOF: Poll, Response and Final. TOF and distance are calculated on the UWB anchor, based on the timestamps calculated in both devices. Finally, a report message is optionally employed in order to transfer distance measurements from the UWB anchor back to the UWB tag as shown in Figure 4-47. The TOF is then calculated as: 

\[
\text{TOF} = \frac{(T_{RR} - T_{SP}) - (T_{SR} - T_{RP}) + (T_{RF} - T_{SR}) - (T_{SF} - T_{RR})}{4}
\]
4.4.1.3 Application to 5GCroCo Use Cases

UWB-based positioning can be applied to the 5GCroCo use cases as follows:

4.4.1.3.1 Tele-operated Driving
When ToD is required, the ego vehicle can use UWB to provide its precise position and/or the precise position of obstacles detected by its sensors. This can provide the required accuracy to define a trajectory avoiding the obstacles. The solution can make use of inertial sensors to update the position under high speeds.

4.4.1.3.2 HD Mapping
UWB precise positioning can be used in order to obtain the precise position of the vehicle and, relative to it, of the changes in the HD map uploaded to the HD map server in the backend. Precise location will significantly enhance the quality of the uploaded data, increasing accuracy and avoiding uncertainties and inconsistencies. It could also improve the fusion of changes detected by different vehicles if the provided information about vehicle position and position of the detected change is more precise.

4.4.1.3.3 ACCA
The vehicle detecting the obstacle can use UWB to provide its precise location and/or the precise location of the obstacle to be sent to the MEC-hosted application server. UWB ranging can be triggered by the obstacle detection event. It could also improve the fusion of hazards detected by different vehicles or generated from the MEC-hosted application server if provided information about vehicle position and position of the detected change is more precise.

4.4.2 Vehicle Positioning Based on Sidelink Measurements
This positioning system is based on using receiver measurements on the 5G New Radio sidelink to provide relative position information between a transmitting vehicle and a receiving vehicle. In particular, the measurements will be beam based, focusing on the use of Frequency Range 2 (FR2) (above 7 GHz) potentially included in future 3GPP specification releases of 5G New Radio V2X sidelink. The measurements and the resulting relative positioning estimate could be
furthermore combined or fused with other on-board sensors on the vehicle (i.e., lidar, radar, camera, etc.) to provide higher reliability and accuracy.

The main goal is to meet the 5G New Radio V2X relative position requirements [62] which state that the relative lateral positioning accuracy should be 0.1 m between UEs and the relative longitudinal positioning accuracy should be less than 0.5 m between UEs. These relative position requirements were derived to support coordinated manoeuvres between vehicles i.e., overtaking, lane merging and platooning. Examples of overtaking and platooning manoeuvres are depicted in Figure 4-48. We will focus on the performance for different types of position measurements and the performance differences for different locations of beam forming antenna arrays on the vehicle.

4.4.2.1 Geometry and Signalling Scheme

We consider vehicles that are equipped with multiple antenna panels/arrays mounted on different parts of the vehicles. An example of the considered setup is depicted in Figure 4-49, where the antenna arrays are mounted on the bumpers of the vehicle. The relative position of the transmitter arrays with respect to the reference point of the vehicle may or may not be known to the receiver.

Disjoint subsets of the available resource elements are assigned to the transmission arrays of the vehicles, so as to enable orthogonal access of the arrays to the channel. Each array transmits a beamformed reference signal on its allocated resources. The beamforming strategy, as well as the resource allocation, can be tailored to better accommodate specific manoeuvres.
4.4.2.2 Measurements and Position and Orientation Estimation

With $K_T$ transmit panels and $K_R$ receive panels, up to $K_TK_R$ links can be established between the transmitting (Tx) and receiving (Rx) vehicles, with the quality of the links depending on their relative position and orientation. The following set of measurements can then be obtained at the receiver:

- **Angle-of-Arrival (AOA) and Angle-of-Departure (AOD):** For each Tx-Rx panel pair, an AOA measurement is attainable if the Rx panel is equipped with two or more antenna elements. The quality of the AOA measurements will depend and the size of the Rx array and geometry, as well as the AOAs themselves. The quality of the AOD measurement for each Tx-Rx pair is in turn influenced by the size and geometry of the corresponding Tx panel, with the additional requirement that the beamforming strategy is properly selected to enable the identifiability of the AOD. Trade-offs such as the accuracy of the measurement versus robustness to time-variation due to mobility have to be taken into account during the design of the beamforming strategy.

- **Time-of-Arrival (TOA):** Each of the Rx panels can measure the TOA of the signal from each of the Tx panels. The accuracy of each TOA measurement is influenced by the bandwidth of the signal transmitted by the corresponding Tx array, which is in turn determined by the implemented resource allocation strategy. The TOA measurements can be further translated to the following:
  - TOF: If the clocks at the transmitted and the receiver are synchronized and the Time-of-Departure (TOD) is known to the receiver through some protocol or a timestamp, the TOA measurement can be translated to a $\text{TOF} = \text{TOA} - \text{TOD}$ measurement, which is effectively a range measurement. Nevertheless, the requirement for Tx-Rx clock synchronization imposes harsh requirements on the system, which may be difficult to realize in practice.
Time-Difference-of-Arrival (TDOA): If the Tx-Rx clocks are not synchronized and/or no time-stamps are available, the TDOA of signals corresponding to different Tx-Rx panel pairs can be calculated: the TOA of a Tx-Rx panel pair is used as reference and (up to) \( K_T K_R \) - 1 TDOA measurements are obtained as the difference of the TOA of a given pair from the reference TOA.

Accurate estimates of the relative position and orientation of the Tx vehicle to the Rx vehicle can be obtained by using the above-mentioned measurements. They can potentially also be fused with measurements from other on-board sensors. Furthermore, the geometric relation of the positions of the arrays with the reference point of the corresponding vehicle can be exploited. In the mapping of angle and time measurements to position estimates, the reliability of each of the measurements, as expressed by its corresponding Fisher information, is taken into account by appropriate weighting. The relative importance of angle vs. time measurements is also evaluated based on position and orientation Fisher information and Cramér-Rao bound analysis. This is done separately for the lateral and the longitudinal position component, as different accuracy requirements might be present.

### 4.4.2.3 Application to 5GCroCo Use Cases

Sidelink measurement-based positioning can be applied to the use cases as follows:

#### 4.4.2.3.1 Tele-operated Driving

It enables the tele-operated vehicle to be aware of the positions and trajectories of the neighbouring vehicles. The better the awareness of the environment including moving objects, the better the safety and the more reliable the operation is. Furthermore, the sidelink-based relative positioning can enhance the accuracy and availability of the localization of the vehicle. First, by fusing the sidelink-based relative positioning results with the traditional Line-of-Sight (LOS)-based positioning result (from BS), higher vehicle positioning accuracy can be obtained. Second, when the vehicle is under non-LOS (NLOS) condition to the BS, it is difficult to obtain its accurate position based on traditional LOS-based positioning method. By using sidelink-based relative positioning and available position information of the neighbouring UE (vehicle, RSU, etc.), the tele-operated vehicle can obtain its own position accurately. In this way, the availability of the localization of the vehicle is enhanced.

#### 4.4.2.3.2 HD Mapping

The relative position between vehicles could be used to increase the accuracy of detected HD map changes and support fusing the information obtained from different vehicles by including relative positioning and orientation information of surrounding vehicles to the uploaded changes of each vehicle.

#### 4.4.2.3.3 ACCA

It allows vehicles to locate each other with very low latency, including their orientation. Sidelink-based positioning can be especially useful in cases where the detected hazard is caused (e.g., an accident or a traffic jam) by another vehicle equipped with this technology. Also, the relative position between vehicles could be used to increase the accuracy of detected hazards, fusing the
information obtained from different vehicles. It also permits to compute potential lane change recommendations in front of an accident and a jam.

### 4.4.3 3GPP GNSS Real-Time Kinematic (GNSS-RTK)

Originally, the main driver for precise positioning in 3GPP standardization was locating people who dialled the emergency call number. Being an US-based requirement\(^\text{59}\), it was called Enhanced 911 (E911) according to the US emergency phone number and requires accuracy within few meters. With 3GPP Rel. 15 GNSS-RTK was introduced to the LTE standard\(^\text{81}\) to provide correction information to GNSS signals, allowing to reach centimetre-level accuracy\(^\text{82}\),\(^\text{83}\). With Rel. 16 the same solution was formally also standardized for 5G New Radio\(^\text{84}\) but the name “LTE Positioning Protocol (LPP)” was kept. Before being standardized in 3GPP, GNSS-RTK was already possible as the required correction information can be provided through any IP-based data connection. The main advantage of the 3GPP standardized GNSS-RTK solution is the option to broadcast the correction information within a System Information Block (SIB) transmitted by eNBs\(^\text{60}\) or gNBs. This allows to substantially reduce the data volume for correction information as this information is the same for all receivers within a given area. The 3GPP specifications also allow unicast transmission of the information. In this case there is no\(^\text{61}\) advantage from reduced data volume, but it can be used without requiring special features in the RAN on g/eNBs or UEs. The solution can still benefit e.g., from authentication through the 3GPP core network and from location information of the cells to deliver the right correction information according to cell location.

Figure 4-50 depicts the 3GPP GNSS-RTK architecture according to\(^\text{85}\). It is part of the 3GPP Location Service architecture which covers all generations starting from 2G. The newly added NF Evolved Serving Mobile Location Center (E-SMLC) is therefore also commonly called Location Server.

---

\(^{59}\) [https://www.fcc.gov/general/9-1-1-and-e9-1-1-services](https://www.fcc.gov/general/9-1-1-and-e9-1-1-services)

\(^{60}\) Incl. eNBs serving as anchor cells for non-standalone 5G New Radio, making GNSS-RTK therefore already available for NR with Rel. 15

\(^{61}\) At least not in RAN and the part of the Core Network traversed to reach the server providing the GNSS-RTK information. The traffic load through the Internet towards the MNO network is reduced as the correction information is only provided once for each E-SMLC.
Figure 4-50: 3GPP GNSS-RTK Architecture

Figure 4-50 depicts two variants: unicast and broadcast transmission. For the unicast case, the E-SMLC is considered to be an application server delivering its information over the SGi interface to each client, as shown in Figure 4-51. The required correction information for all areas where the service is provided does only need to be provided to the E-SMLC from a Network RTK Server, not to each GNSS-RTK Client in the vehicles. All 3GPP-specified Core and RAN components, incl. UEs, are not show in the figure as they are transparent for the communication. Core network information from the MME could still be used to determine the location of vehicles through corresponding cell or Tracking Area Identifiers (IDs) and/or for authentication. This is not used in the example and the client applications need to inform the E-SMLC about cell ID changes to assure they get the correction information for their area.
Figure 4-51: Message Flow of Unicast GNSS-RTK Network Assistance Data Delivery through E-SMLC

Figure 4-52 shows the broadcast variant of 3GPP GNSS-RTK. In this case the E-SMLC is considered as NF of the Core Network connected to the MME through the SLs interface. For the sake of limiting complexity, the figure does not show multiple MMEs and g/eNBs but only multiple UEs and GNSS-RTK Clients all served by the same g/eNB. In reality, the E-SMLC only provide network assistance data relevant for the area served by a specific MME and the MME further separates the information and provides to each g/eNB only the information that is relevant for the area it serves. This way there is no need for vehicles to provide position updates.
The broadcasted information can be encrypted and UEs can request the decryption key from their MME when attaching and/or performing a TAU. The MME will only provide the key to UEs subscribed to the service according to information stored in the HSS.

Rel. 15 specifications enabled to use Observation-Space Representation (OSR) GNSS-RTK information. In Rel. 16 this was extended to State-Space Representation (SSR) allowing to further reduce the required data volume by submitting different kinds of correction information according to required update frequency to assure centimetre accuracy. OSR relies on reference stations with known positions. The offset between position obtained through GNSS and precisely known position of the reference station is used to determine the correction information. As the position error might be different in different locations, a rough device location must be determined. It is enough to use the ID of the serving cell as cell areas are usually much smaller than areas served by a reference station. Non-physical reference stations are realized by interpolating the error from several physical reference stations.

SSR separates the GNSS error into different factors that contribute to it. Some, e.g., satellite orbit and clock errors and satellite signal biases, are area independent and/or slowly changing and therefore rarely need to be updated. Others like atmospheric satellite signal delays in the ionosphere and troposphere are area dependent and/or change faster and according correction...
is provided more frequently. By using different update frequencies for different correction signal components, the data volume is further reduced.

Many 4G and 5G modems also include a GNSS receiver and could therefore use the 3GPP GNSS-RTK solution in a positioning engine on the modem to improve position accuracy and provide precise GNSS positions to other functions of the unit hosting the modem. Until such integrated solutions are available, the modem can provide the RTK-GNSS assistance information to a separate positioning engine, e.g., a software application running on the unit hosting the modem, to provide precise GNSS positions.

4.5 Security
This section refers to the used approach/methods and describes the outcomes of the audits from the Security Task Force to enhance the level of cyber security in the project. It provides a general overview of security measures considered most relevant for the project. How they are actually applied within the different use cases is covered in Deliverable D3.1 [3].

As the software in the project is in development and constantly changing, security tests have not yet been concluded. Still the conducted Security Audits (Penetration Tests) showed that there is a need for security guidelines especially for the topics of Message Queuing Telemetry Transport (MQTT) and Docker security. These two technologies are widely used in the project, as further described per use case in Deliverable D3.1 [3], and therefore the creation of security guidelines is of great benefit for the overall cyber security level of the project.

4.5.1 Used Methods and Results of the Security Task Force
To generate evidence-based improvements in terms of cyber security a practical “hands on” approach was followed. This way it was possible to find out first-hand how far the development has progressed, and in which state the respective components are in terms of cyber security.

A detailed description of the approach and methods used for the interviews and Penetration Tests is provided in Annex A.

4.5.1.1 Results of the Interviews with Partners
In a first step a series of interviews was conducted with project partners to identify the scope for security audits. Based on the outcomes of the interviews several objects for Penetration Testing could be identified and necessary steps to conduct the tests were defined.

4.5.1.2 Results of Security Analysis (Penetration Testing) of Applications
Security analysis (Penetration Testing) of several client and backend applications was conducted, and the findings and suggested solutions were reported to the respective project partners in a standardized report format.

---

62 Further topics not yet covered have been identified. They will be covered in the next version of this deliverable and are listed in Section 6.4.
The conducted security audits (Penetration Tests) showed that there is a need for security guidelines especially for the topics of MQTT and Docker security. These two technologies are widely used in the project and therefore the creation of security guidelines is a great benefit for the overall cyber security level of the project.

Due to confidentiality no in-depth information about the tested objects, the respective findings and involved partners can be shared in this deliverable.

### 4.5.1.3 IoT Firmware Devices Analysis

5G modems could also be identified as a crucial component in the project. Therefore, we conducted tests on the firmware of two 5G modems. Static and dynamic analysis techniques were used to find vulnerabilities. The compliance of the firmware to various standards and recommendations published by renowned organizations in the industry was also tested. Findings got classified according to severity as *High, Medium, Low or Information*. This reflects the likely impact of each issue. The respective findings were reported to the corresponding manufacturer. Due to confidentiality no in-depth information about the tested objects and the respective findings can be shared in this report.

### 4.5.2 Security Guideline for MQTT Protocol

The ACCA and ToD use cases apply MQTT for different message types, as further described in Deliverable D3.1 [3]. The MQTT-protocol in version v3.1.1 does not include state-of-the-art security features like password encryption or other features. This section explains options how to harden the communication between the MQTT clients and the MQTT broker to increase the security of the MQTT-protocol v3.1.1.

If the MQTT-protocol v3.1.1 [86] traverses an unsecure zone like the Internet, where hackers are suspected, the network traffic between the MQTT clients and the MQTT broker must be encrypted, for example using TLS encryption. If all MQTT brokers are situated in a secure zone behind a Firewall or an Application Layer Gateway (ALG), then the network traffic must be encrypted at least between the MQTT client and the Firewall / ALG. If the network traffic between MQTT clients and MQTT brokers is not encrypted, all username / password credentials or client identifiers are transmitted in clear text, because the MQTT-protocol in version v3.1.1 does not take care for encrypting username / password credentials or client identifiers.

#### 4.5.2.1 Authentication

If authentication is an issue, MQTT-protocol v3.1.1 provides authentication with client identifier or identification with username / password credentials on application level. If no encryption of the network traffic is provided, then the client identifier and the username / password credentials are transmitted in clear text.

#### 4.5.2.2 TLS Encryption

Using TLS Encryption [87] as encryption mechanism will put the responsibility of encryption to the TCP/IP protocol. That mechanism of encryption will protect all parts of the MQTT message. For
security reasons, only the current version of TLS (TLS 1.3) should be used because the older versions of TLS or SSL are not treated as secure anymore.

4.5.2.3 Payload Encryption
Using MQTT payload encryption as encryption mechanism provides security only for the application layer. This type of encryption mechanism is not defined in the MQTT specification and is completely application specific. Only the payload of the message is encrypted, all “MQTT PUBLISH” metadata remains unencrypted. The security issue of that encryption mechanism is that attacker can replay the message or modify parts of the message.

4.5.2.4 End-to-end Payload Encryption
Using MQTT Payload Encryption End-to-end as encryption mechanism provides security only for the MQTT payload on the application layer. This type of encryption mechanism works without having to configure the MQTT brokers, because the MQTT broker can use the unencrypted packet metadata for routing, QoS-handling, et cetera. Only trusted clients have access to the decryption key and will be able to read the data. The security issue of that encryption mechanism is that it does not protect client identifier and username / password credentials on the connection itself.

4.5.2.5 Payload Encryption Client to Broker
Using MQTT Payload Encryption Client to Broker as encryption mechanism will encrypt the MQTT payload in the communication between MQTT publisher and MQTT broker. The MQTT broker has to decrypt the message received from the MQTT publisher before distributing it to the MQTT clients. For this encryption mechanism the environment between MQTT broker and MQTT subscribers must be inside a secured area. The security issue of that encryption mechanism is that it does not protect client identifier and username / password credentials on the connection itself.

4.5.2.6 Data Integrity with Checksums
If data integrity is an issue, checksums can be used for MQTT messages if also TLS encryption is in use. It is recommended to do not use checksums without TLS encryption because checksums can get altered within the MQTT packet. The security issue of using checksums for MQTT messages is that it does not protect client identifier and username / password credentials on the connection itself. Still this method is not as secure as the use of digital signatures.

4.5.2.7 Data Integrity with Message Authentication Code
If data integrity is an issue, Message Authentication Code (MAC) for MQTT messages can be used for data integrity and for authentication check. MAC algorithms like Hashed MAC (HMAC) are very fast compared to digital signatures. The shared secret key must be exchanged securely prior to the MQTT communication. MACs can be used securely even if TLS is not deployed. The security issue of using MACs for MQTT messages is that all clients who are aware of the secret key can sign and verify because the same key is involved in both processes.
4.5.2.8 Data Integrity with Digital Signatures
If data integrity is an issue, digital signatures for MQTT messages can be used for data integrity, authentication check and for non-repudiation. Digital signatures use public / private key cryptography. Provisioning and revocation of public / private keys adds complexity to the system. Only if there is a guarantee that only a specific client can publish to a specific topic then digital signatures can be a good fit for MQTT messages.

4.5.2.9 Data Integrity with Digital Signatures and Encryption without TLS
If data integrity is an issue, data integrity check and encryption for MQTT messages can be used. If MQTT is used without TLS encryption, a data integrity check will enhance the security. Even if the message is decrypted and, after modification, encrypted again, the integrity check still fails if the message was altered.

4.5.2.10 Data Integrity with Digital Signatures and Encryption with TLS
If data integrity is an issue, data integrity check and TLS encryption for MQTT messages can be used. If TLS is deployed and proper authentication and authorization mechanisms are in place, the data integrity check will not add much additional security. This makes only sense if sensitive data is transmitted with the command PUBLISH packets.

4.5.2.11 MQTT without Security Features
Using MQTT-protocol v3.1.1 without security features like for example encryption lets hackers easily intercept the communication between MQTT clients and MQTT broker. An automated Man-in-the-Middle-Attack is able to change the values sent to or from MQTT clients to the MQTT broker(s) because it is possible to intercept unencrypted “MQTT Publish messages”, make changes in real time and sending these rogue packets to the MQTT broker. This can be done by interacting with the fields of the MQTT layer, make changes, adapt the lengths field if the length of the message was changed and recalculate the control fields. The MQTT client will not be able to see that the packets received are rogue packets if the security features in the description above are not in use.

4.5.3 Security Guideline for Docker
The basic security guideline for Docker gives developers of containerized workloads an overview of security relevant aspects for building and running docker containers. It recaps container basics, covers the threat models and relevant attack vectors, and focusses on common docker vulnerability patterns. For a comprehensive understanding of each vulnerability pattern, commonly seen insecure methods and their problems would have to be shown. In addition, secure alternatives would have to be discussed and shown with examples.

---

63 [https://docs.docker.com/engine/security/](https://docs.docker.com/engine/security/)
Containers are mostly regular processes that share certain resources with their host and other processes but are also isolated in other areas. Additionally, differences between containers and VMs are covered, which is also an important aspect with regards to security.

When dealing with container technologies, it is important to know the difference between these two terms:

- **Image**: A read-only template with instructions for creating a container
- **Container**: A runnable instance of an image

In an environment running containers, the following threat vectors must be paid attention to and protected against:

**Container escapes**: Malicious actors trying to escape the isolated environment to gain access to (sensitive) resources outside the intended environment. Usually, the goal is to directly access sensitive resources on the host system and to achieve persistence. Malicious actors can be malicious software or even an attacker abusing a vulnerability in the containerized software.

**Network attacks**: Malicious actors trying to bypass network restrictions to attack network services outside the container. This can be services running in other containers, on the host itself or even on other hosts in the same network.

**Resource starvation**: Malicious actors using up limited physical resources, such as CPU and memory, to reduce system stability and impact other applications on the same host.

**Compromised host**: When an attacker gains low privileged access to a host with an insecurely configured Docker environment, this may be used to escalate privileges to full system access.

**Image integrity**: Insecure coding practices might allow an attacker to inject malicious software into images during the build phase.

The following, commonly seen, Docker vulnerability patterns are covered in the security guideline. The list is based on findings from security assessments and on common security best practices.

**Secure user mapping**: By default, containers run as root user which gives the application a wide range of permissions, which are often not necessary. This also makes container escapes easier and provides an attacker higher chances to access sensitive host resources. The solution is to use lower privileged users within the container and to map root users inside the container to low privileged users on the host.

**Network segmentation**: By default, all containers running on a host can reach each other over the network. By using a separate virtual Docker network for each (set of) container(s), this can be limited.
Secure defaults and hardening: This covers the host setup and configuration, especially with regards to the Docker daemon, as well as methods to ensure a secure software environment within the container. The key aspects are to install as few software components as possible, only install them from trusted sources, to avoid giving containers access to sensitive host resources or container features, and to limit container privileges (process capabilities) as much as possible.

Security contexts: Even though containerized workloads easily allow one to run multiple applications on the same physical host, applications with different security requirements must still be physically separated. For example, the following properties should be considered: Environment (test, production), user base (internal only, Internet-facing), data sensitivity (sensitive internal data, sensitive customer data, no sensitive data).

Secrets management: Most applications need access to secrets, such as API keys or database passwords. It is easy to accidentally embed a secret into an image. As a result, it might get leaked and becomes hard to change if the need arises. Using Docker secrets is a secure way of dealing with such sensitive data.

Resource protection: Define resource limits for containers to prevent rogue applications to negatively impact other workloads or the host system itself.

Immutable paradigm: The environment of a container should change as little as possible during its runtime. Software should not be installed or updated. Instead, new image versions should be built and used.

Patch management: Patching all relevant components (applications and dependencies within the container, container technology, host operating system and installed software) regularly is as important as in any other IT environment.

It can be concluded that there are many opportunities for a developer to accidentally create an insecure image or container. Many aspects are not secure by default, while others can be easily accidentally made insecure. However, there are many controls that allow for highly secured environments, if used correctly. A sufficient level of security can only be achieved by combining protection mechanisms at different levels. Security should always follow a "defence in depth" approach.

4.5.4 Privacy/General Data Protection Regulation (GDPR)
At this point only a general and generic statement about privacy can be made as there is not yet a final picture on where, how and which amount of GDPR relevant data (personal data) will be processed/ transferred in the respective systems.

Facilitating a rollout that entails a considerable amount of data generated on an international level certainly poses substantial regulatory and compliance challenges, including the consideration of the impact on a person's reasonable expectation of privacy.
Correlating location, speed, route and destination data (just to name a few examples) with additional information regarding the passenger including the date and time of the trip allow getting a concrete picture of when, where and how a person travels, especially if specific data categories are stored or logged over a longer period of time. As a result, the data storage has a potential to generate a precise and comprehensive picture of a person’s movements that can further reflect on, for instance, familial and professional associations. Currently navigation features in modern cars often contain saving specific locations in memory and use locations and planned routes to identify additional information relevant to the trip, including points of interest. Collecting data from the surrounding of the vehicle relating to the operation of the vehicle automation or assistance system can cause a high number of additional privacy concerns that need to be addressed, including the safe use of driver-assistance options.

In that context, the privacy by design requirement in accordance with Art. 25 GDPR [88] shall be enforced by building privacy into devices at the outset, rather than as an afterthought and to consider, by conducting a privacy risk assessment, testing the privacy measures before launching the products and minimizing collected and retained data.

Although certain privacy rights are granted to passengers, pedestrians, and drivers as data subjects (Art. 4(1) GDPR [88]) the data processor still ultimately controls the data as data owner (Art. 4(7) GDPR [88]). Accessibility, exchange and reuse of status road data, dynamic road data and traffic data shall be granted in accordance with the Commission Delegated Regulation (EU) 2015/962 [89] with road safety on an international and national level as paramount objective.


To assure compliance of the involved systems to privacy related legal requirements it can be recommended to follow relevant International Organization for Standardization (ISO) Standards such as the following:

5 Interoperability Issues

The following section identifies interoperability issues for the different solutions described in Section 4. It divides the issues into the categories cross-telco-vendors, cross-MNOs and borders, and cross-car-OEMs.

Potential solutions for many, but not all, of the issues are discussed in previous section and according references are provided. Further issues will be resolved in Deliverable D3.3 planned to be published in October 2021. Those are listed in Section 6.

5.1 Across Telco Vendors

This section discusses challenges when dealing with components provided by different vendors.

It is common that mobile radio networks consist of software and hardware components from different vendors. With the introduction of All-IP networks any network company became a potential vendor. Furthermore, the Information Technology (IT) industry was added to the set of potential vendors with introduction of virtualization in 5G. In broader sense open-source communities also qualify as “vendors” in this context.

5.1.1 General Connectivity

Cross-border, MANO, and SDN interoperability issues between telco vendors are discussed in this section.

5.1.1.1 Network Service Continuity at Country Borders

In Section 4.1.1 it is proposed to deploy the S10 interface across different MNOs to enable cross-MNO radio handover. ERI has successfully tested this with its MME products but this does not guarantee it is also possible with MMEs from other vendors. Every VNF is extensively tested before being deployed. This also includes cross-vendor interoperability tests, that sometimes reveal issues that were not detected during single-vendor testing. To the best of our knowledge cross-vendor tests of the S10 interface where the MMEs serve different networks are not common and therefore undetected issues might exist.

5.1.1.2 Network Orchestration and Control

Although often based on the same open-source components, MANO and SDN solutions are usually not compatible across vendors due to e.g., customization and packaging that tightly couples the different components forming the overall MANO and SDN solution. For example, the carrier-grade MANO and SDN solutions introduced in Section 2.1.3.1.3 and Section 2.1.3.2.2.2 come with a set of installation templates tailor made for certain high-performance hardware configurations that were tested and certified. Manually selecting a subset of the components and using them together with other ones from other sources or directly from their open-source repositories is not foreseen.
5.1.2 Predictive and End-to-End QoS
The radio schedulers in the gNBs are the key components to deliver the requested QoS according to CQI/5G class (see Section 2.2.1). Their algorithms and implementations are vendor specific. This is a minor problem if the respective QoS is fulfilled but products from different vendors might show different behaviour and therefore different behaviours and / or performance in case of congestion and insufficient resources to fulfil the QoS requirements of all subscribers.

5.1.3 MEC
With different cloud management systems in place, as discussed in Section 5.1.1.2, migrating a MEC application from a cloud managed by a system from one vendor to another one managed by a different vendor might become complex.

5.1.4 Precise Positioning
No cross-telco-vendor issues for the currently described precise positioning solutions were identified.

5.2 Across MNOs and Borders
MNOs are typically active within one country and even ones active in several countries typically act separately per country. One of the reasons is the spectrum licensing per country. Crossing a border and driving a certain distance into a visited country will therefore eventually result in a change of serving MNO.

Besides country borders, further cross-MNO topics exist. Within one country it cannot be expected that all vehicles are served by the same MNO. Vehicles might also be allowed to roam into different networks within one country.

5.2.1 General Connectivity
Cross-border, MANO and SDN interoperability issues between MNOs and across country borders are discussed in this section.

5.2.1.1 Network Service Continuity at Country Borders
In Section 4.1.1. it is described how deploying the S10 interface between MMEs in different mobile networks operated by different MNOs can enable cross-border / -MNO radio handover. In reality, IP connectivity between different data centers across borders must be established and each MME might have to be connected to several MMEs in the other network. Furthermore, complexity is added since each country has several MNOs that might need to be interconnected.

Using IPX networks to reach the HSS in the home network is a well-established enabler for roaming. For the S10-interface, performance is currently not known when deployed over an IPX network and the influence of S10-interface performance on cross-border / -MNO handover performance is also not known. Besides that, security topics like authorization and authentication must be evaluated for such setting, including their potential influence on performance, e.g., when Virtual Private Network (VPN)-tunnels are applied on the path.
Also, Section 4.1.1 presents how this service continuity concept could evolve with standalone 5G New Radio and beyond. These enhancements relate to interconnecting AMFs of different MNOs, similarly to the previous inter-MME connections. Furthermore, it includes proactive registration to the operators of the neighbouring countries. Such solutions however need additional study.

5.2.1.2 Network Orchestration and Control
Network orchestration and control usually happens within a single network and country, so no issues were identified.

5.2.2 Predictive and End-to-End QoS
An SLA with an MNO usually only applies when connected to the home network, not when roaming. Even when extended to visited networks one would have to guarantee that the agreement is identically interpreted by the visited network providing same performance or at least the one agreed for roaming case. The QCI/5GI classes described in Section 2.2.1 only give an indication of what QoS to expect. Besides the cross-vendor issues described in Section 5.1.2, diverging behaviour can also result from different parametrization of the schedulers. SLAs might require more information than provided in the QCI/5QI tables in Section 2.2.1. Network Slice templates, as described in Section 4.2.1.4, support a common, machine readable understanding of QoS requirements across MNOs.

A challenge arises if a prediction for a different than the currently serving network is needed. This can be particularly true when crossing a country border. Furthermore, the reliability of the prediction or other parameters defining its confidence might be different for different MNOs. In the worst case involved MNOs do not offer a QoS prediction service. Furthermore, a common method to identify and address the PF offered by an MNO is required to continue using a PF after being handed over to a different MNO.

5.2.3 MEC
The issue of having different, potentially not compatible, cloud management systems was discussed in Section 5.1.3. If MNOs have customized management systems, third party software MEC application server providers face the challenge of having to customize their software images for the different systems.

A P-GW in the home network is always used in case of home-routing. Local Breakout for MEC is therefore prohibited. Network supported APN configuration might not be possible in a visited network or the APN configuration from the home network might not be accepted by the visited one. MNO-dependent server discovery mechanisms might also differ in different networks.

The above examples mostly refer to non-standalone 5G New Radio where certain customization on the client side might be required for certain use cases. MEC-related mechanisms in standalone 5G New Radio are, as far as we evaluated them by now, designed in a way to be transparent for the client application in the vehicle. MNOs can therefore apply different solutions and still achieve the same goal, e.g. seamless gateway switching and end-to-end service continuity when
switching the MEC-hosted application server. Some of the mechanisms need to be supported by the OS and for within architectures where the client application is not running on the CCU.

Many discussed use cases require information exchange between neighbouring MEC hosts. Within one MNO this is possible through the transport network outside of the 3GPP domain but doing this across MNOs might be more challenging. Solutions for that were described in Section 4.3.2.4.

5.2.4 Precise Positioning
As described in Section 4.4.1, the UWB-based positioning solution needs the deployment of anchors close to the road. Thus, operation in different countries will depend on the presence of these anchors and on the interoperability between the anchors and the tags installed in the cars. The GNSS-RTK solution described in Section 4.4.3 is also specified for roaming but it is an open question, to be especially determined through trials, if the user plane variant works in roaming conditions and if it is seamlessly handed over without unacceptable service interruption.

5.3 Across Car OEMs
Vehicles from different brands are considered in the project and in reality, there are even more brands and each brand has different models with different series. Even within the same series the software and hardware can change, so a huge set of different components must be considered for interworking. This space is drastically reduced if not all CCUs need to be evaluated for mutual interworking but only towards the backend that could potentially be customized, to some extent, for certain car brands and the software and hardware they use. The effort for this customization must be also considered and cross-car-OEM interoperability must be then assured within the backend. This is further described in the real-world architecture for the ACCA Use Case in Deliverable D3.1 [3].

5.3.1 General Connectivity
Different than for the previous sections, this section is focused on issues resulting from the modem and antenna as the interface of the vehicles towards the network.

Some aspects of service quality are beyond network influence and depend on the modem, antenna type and antenna placement. These can be different for different car brands and even within the same brand, since different suppliers can provide the CCUs. The network relies on support from the UE for channel estimation in order to perform link adaptation and handover. Procedures are mostly specified and unified, but per-modem performance differences can exist. End-to-end performance can be different even if everything is identical except for modem and/or antennas.

5.3.2 Predictive and End-to-End QoS
Although standardized and best practice open-source interfaces for interaction with the modem exist (see Section 4.2.2.1), behaviour can differ across different modems. This is particularly true when querying channel quality parameters. Experience shows that general availability, sampling
rate, range, and quantization differ across modems. It cannot be precluded that two different modems report different values for the same channel quality.

### 5.3.3 MEC

For the ACCA use case the realization of the vehicle application is tightly coupled with the MEC realization. This includes e.g., server discovery, description of geographical regions and realization of geocasting. Some vehicle OEMs already have systems in place for other or similar services with certain solutions that might be intended for reuse. This can result in the desire to not have identical, and therefore naturally interoperable vehicle applications, but allow a certain degree of customization as long as interoperability through the backends is assured.

### 5.3.4 Precise Positioning

In the case of the UWB-based precise positioning solution, as mentioned in Section 5.2.4, interoperability issues can arise between the anchors deployed close to the road and the tags installed on the cars. Besides, the operation of the sidelink-based solution (see Section 4.4.2) could be influenced by the existence of different radio equipment, which might not be interoperable, and their antenna type and placement, what might impact the accuracy of the measurements. Finally, as mentioned in in Section 4.4.3, the specific implementation of the GNSS-RTK solution can be either in the modem or on the unit hosting the modem. When done on the modem no interoperability issues are expected, as the modem would just provide more precise GNSS information through the same interface as used without GNSS-RTK. If done on the unit, interoperability issues could exist depending on how the positioning engine exposes the corrected GNSS-RTK information.
6 Selected Open Issues and Refinements for Next Version

This section lists topics that were not at all or not significantly enough discussed in this deliverable. It does not claim to be complete so the scope of follow up versions of this document (Deliverable D3.3 planned for October 2021) can be extended.

6.1 Deployment Aspects

Deployment aspects were abstractly discussed in Section 4.3.2.1.2 in context of cross-MNO end-to-end QoS. The performance of many functions discussed in the deliverable depends on the deployment of software components on real server hardware in data centers or cabinets. For this and other deployment challenges, generally applicable reference deployment architectures will be defined, reflecting an abstract view of mobile radio network deployments in different countries. Where possible, MNOs will provide topology information about their networks or parts of them.

6.2 MEC

Section 4.3 describes different challenges and solutions for MEC including how they relate to other key 5G solutions discussed in this document. For non-standalone 5G New Radio no uninterrupted gateway switching is possible, so only use case specific mitigation techniques can be applied to limit the effect. This requires further evaluation of the severity of the effect, in term of duration of the period where connectivity is lost during switching of gateways. It shell furthermore be discussed what implication it would have to not conduct any gateway switching at all with non-standalone 5G New Radio, as the negative effect of service outage might be larger than from not using the closest MEC-hosted application server through an appropriate gateway. For standalone 5G New Radio further details of SSC mode 3 regarding options for the transition period, when two gateways are used simultaneously, must be evaluated in more detail. This is closely related to CCU-side network interface selection and routing described below.

It was hardly discussed yet how the OS and connection management system in the CCU can deal with multiple IP addresses and/or virtual network interfaces that result from some of the solutions discussed in Section 4.3.2.1.1, 4.3.2.2, and 4.3.2.3. The issue mostly arises when multiple gateways are available simultaneously either during gateway and/or MEC-hosted application server switching or to distinguish between MEC and public Internet data network access. URSP was identified as potential solution but could, so far, not be further evaluated without an understanding how device and OS vendors will realize it, as it is not purely realized in the UE and modem. Furthermore, the architecture so far focused on IPv4 and does not exploit further possibilities that might become available when using IPv6.

Almost all described solutions must be practically evaluated within the second round of trials planned for mid-2021. In this context further issues might be detected, and the architecture will then be adjusted. Although no standalone 5G New Radio network is used for trials, methods to...
emulate its behaviour (e.g. [97]) will be used to also discover potential issues for this deployment and provide solutions. Furthermore, findings from other projects using standalone 5G New Radio will be considered.

The interplay of MEC and VNF MANO systems will likely evolve throughout 2021. This will be monitored and corresponding speculative assumptions in Section 4.3.3.2 will be updated where needed.

### 6.3 QoS Prediction Algorithms and Interfaces

Section 4.2.4 describes QoS prediction algorithms and input data they might require and how performance might change depending on input data availability and quality. Corresponding tests and trials, as documented in Deliverable D4.2 [4], have just started. Different algorithms must be further evaluated to find a compromise between access to input data and prediction performance, since access to input data might include certain effort or other issues might exist, e.g., privacy. In addition, the realisation of QoS prediction in cross-border environments (e.g., roaming and non-roaming cases) may require further investigation to make sure that existing interfaces and signalling are adequate.

### 6.4 Security Aspects

Further Penetration Testing of the 5GCroCo architecture will be done for selected components of the trial systems deployed in WP4. This mostly applies to the topics covered so far, namely MQTT and Docker. The privacy assessment, which is currently generic, will be refined to be more specific towards selected part of the trial network and application (use case) architecture.

As new topic, the ETSI-ITS security architecture [98] and PKI will be evaluated for its applicability for CCAM services over cellular networks.
7 References


[18] 3GPP, “TS 29.522 v16.0.0 5G System; Network Exposure Function Northbound APIs; Stage 3,” 2019.


[22] 3GPP, “TS 28.500 v15.0.0 Telecommunication management; Management concept, architecture and requirements for mobile networks that include virtualized network functions,” 2018.


[38] 3GPP, “TS 29.214 v15.6.0 Policy and charging control over Rx reference point (Release 15),” 2019.


[40] 3GPP, “TS 23.060 v16.0.0 General Packet Radio Service (GPRS); Service description; Stage 2 (Release 16),” 2019.


[47] 3GPP, “TS 22.129 v15.0.0 Service aspects; Handover requirements between UTRAN and GERAN or other radio systems (Release 15),” 2018.


[65] 3GPP, “TR 28.814 v0.2.0 Study on enhancements of edge computing management,” 2020.


[75] IETF, “RFC 1546 Host Anycasting Service”.

[76] 3GPP, “TS 23.748 v17.0.0 Study on enhancement of support for Edge Computing in 5G Core network (5GC)”.

[77] 3GPP, “TS 29.122 v15.4.0 T8 reference point for Northbound APIs,” 2019.

[78] 3GPP, “TS 29.522 v16.0.0 5G System; Network Exposure Function Northbound APIs; Stage 3,” 2019.


[85] 3GPP, “TS 23.271 v15.2.0 Functional stage 2 description of Location Services (LCS),” 2019.


A. Used Approach and Methods of the Security Task Force

This Annex describes the approach/methods used in the Security Task Force.

A.1 Interviews with Partners

In a first step a series of interviews was conducted with project partners to identify scope for security audits. For these interviews, a structured guideline, presented ¡Error! No se encuentra el origen de la referencia., was prepared. This guideline was separated into different sections. The first section helped to identify potential scope for technical security analysis. The second section was needed to identify all the requirements needed to be able to conduct the security audits. In the third section possible impact on connected systems was checked. The fourth section helped to clarify organizational topics to be able to conduct the tests smoothly. In the fifth section the confidentiality, format, and date to deliver the results was defined. In the sixth section the respective contact persons were defined and the last section was used in case of remarks.

Table A-1: Interview Guideline

<table>
<thead>
<tr>
<th>Interview Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 What is your preferred scope?</strong></td>
</tr>
<tr>
<td><strong>Examples</strong></td>
</tr>
<tr>
<td>• Web-application?</td>
</tr>
<tr>
<td>• Mobile applications?</td>
</tr>
<tr>
<td>• Infrastructure? Hosted by participant?</td>
</tr>
<tr>
<td>• Sensors and onboard systems (IoT)?</td>
</tr>
<tr>
<td>• Middleware (services, middleware etc.)?</td>
</tr>
<tr>
<td>• Communication channels between all components?</td>
</tr>
<tr>
<td><strong>2 Requirements</strong></td>
</tr>
<tr>
<td>• Scheduled period for pen test execution (start-/ end-date), proposed effort</td>
</tr>
<tr>
<td>• Accessibility of application</td>
</tr>
<tr>
<td>• Whitelisting of IP-addresses necessary?</td>
</tr>
<tr>
<td>• Test data</td>
</tr>
</tbody>
</table>
3 Impact on Connected Systems
- Does the test object have connected systems that are impacted by the pen test?
- Are these systems in scope or should they be disconnected for the pen test?
- Pre risk evaluation?

4 Organisational Topics
- Start/stop mail each day as part of the pen test?
- Handling of critical and high vulnerabilities, reported immediately?
- Permission to attack (PTA)
- Schedule date for presentation/read out session of report
- Timetable
- Milestones

5 Results
- Confidentiality: Public, non-public, internal, confidential, etc.?
- Format
- Report: Delivery-date of draft/final version

6 Communication
- Contact participant
- Contact tester

7 General Remarks
A.2 Security Analysis (Penetration Testing) of Applications

Security Analysis (Penetration Testing) of several Client and Server Applications was conducted, and the findings and suggested solutions were reported to the respective project partners in a standardized report format.

Table A-2: Table gives an overview of the content of the reports delivered.

<table>
<thead>
<tr>
<th>Table of Contents for Security Analysis (Penetration Testing) Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Management Summary</strong></td>
</tr>
<tr>
<td>1.1 Scope</td>
</tr>
<tr>
<td>1.2 Goal</td>
</tr>
<tr>
<td>1.3 Results</td>
</tr>
<tr>
<td>1.4 Suggested Measures</td>
</tr>
<tr>
<td><strong>2 Approach</strong></td>
</tr>
<tr>
<td>2.1 Testing Method</td>
</tr>
<tr>
<td>2.2 Scope and Timetable</td>
</tr>
<tr>
<td>2.3 Test Classes Performed</td>
</tr>
<tr>
<td>2.4 Actions to be Taken After the Conducted Test</td>
</tr>
<tr>
<td>2.5 Disclaimer</td>
</tr>
<tr>
<td><strong>3 Vulnerability Summary</strong></td>
</tr>
<tr>
<td>3.1 Total Risk Per System</td>
</tr>
<tr>
<td>3.2 Risk of Each Vulnerability</td>
</tr>
<tr>
<td><strong>4 Detailed Analysis</strong></td>
</tr>
<tr>
<td><strong>5 Version History</strong></td>
</tr>
<tr>
<td><strong>Appendix A Risk Calculation</strong></td>
</tr>
<tr>
<td>A.1 Definition of the Term Likelihood</td>
</tr>
<tr>
<td>A.2 Definition of the Term Severity</td>
</tr>
<tr>
<td>A.3 Total Risk</td>
</tr>
</tbody>
</table>
In the management summary an overall overview was provided about the scope, goal and results of the tests as well as suggested measures. Suggested measures can be further distinguished between actions with immediate need for action and further measures. Actions with immediate need for action typically consist of the following actions:

- Correction of the discovered vulnerabilities
- Recheck of the assessed applications
- Execution of a full security test

Correction of discovered vulnerabilities and recheck are always strongly recommended. Execution of a full security test with for example an in-depth code review provides an even better overview but is associated with more effort.

In the mid- and long-term it is recommended to take the following further measures to mitigate / solve identified problems:

- **Security acceptance tests**: A security assessment should be done for every system before its use in production in an extent that reflects the system's criticality. By performing such a test before rollout to production, risks can be drastically reduced and potential downtimes avoided.

- **Source code assessment of business-critical applications**: Applications which process business-critical information can be tested thoroughly through a source code assessment. This will reveal vulnerabilities that are difficult to find in blackbox penetration tests.

- **Improving the security awareness of developers through trainings**: A very important measure for improving the quality and security of applications is the awareness of developers about common security vulnerabilities, their root causes and possible counter measures. Educational measures can improve the security awareness of employees.

- **Improving the security awareness of employees against social engineering through training**: A very important countermeasure against social engineering and technical attacks is to maintain a high level of security awareness. Educational measures can improve the security awareness of employees and increase the security level.

The second section describes the approach used in the respective test. Typically, security assessments are conducted to check the security of a complete system or single system components. The tools, methods and techniques used fall into three categories:

- Well known throughout both the computer security and “hacker” communities.
- In-house tools developed to extend the boundaries beyond the usual hacker's toolkit.
- Expert knowledge. Security consultants look for vulnerabilities that may not be discovered by using automated tools.
The section “Scope and Timetable” as well as “Test Classes Performed” are very individual to each test and thus cannot be explained in this deliverable in detail. Generally, test classes are distinguished according to best practice into the following classes:

- **Proprietary Applications**: In this class security checks towards proprietary developed applications are performed.
- **Patch Status**: In this class security checks towards outdated software are performed.
- **Standard Software**: In this class security checks towards deployed standard software are performed.
- **Configuration**: In this class security checks towards the configuration of test items are performed.
- **Infrastructure**: In this class security checks towards infrastructure are performed.

Actions to be taken after the test are individual to the scope of the test, but mainly consist of “clean-up activities” before systems are continued to be used after the audit.

The section “Vulnerability Summary” consists of a total risk per system overview as well as risk assessment of each vulnerability.

In the section “Detailed Analysis” each vulnerability found is explained in detail, showing the respective proof of concept and recommended solutions. In addition, a risk estimation and classification according to industries best practices was made for each vulnerability.

In the appendix the used risk calculation, definitions of the terms likelihood and severity are defined and as well as the used risk calculation explained.

All security risks discovered were evaluated with a risk score. The risk score is calculated from a risk matrix, which consists of likelihood and severity. The likelihood describes the probability that an attacker discovers the vulnerability and can exploit it. The severity refers to the severity of the vulnerability as well as its impact. By multiplying likelihood and severity, the risk score is determined, which allows an assessment of the risks posed by a vulnerability.
B. Classification of Data Centres and their Interconnection

One of the best known standards for data centres is the Telecommunications Industry Association (TIA) specification TIA-942 [99]. That one mainly covers physical aspects like cabling, floor plan and especially redundancy. TIA also published a position paper [100] on how Edge data centres are different from traditional ones. The key statement is that traditional data centres are well defined, especially through TIA-942, while very different types of Edge data centres can exist. They can be as small as one half-sizes server rack or even just a single (small) server. But even a traditional data centre can qualify as “Edge” from the perspective of end users close to it.

A further refinement of the classification of traditional data centres is provided in [101] and [102]. Three types are listed: Onsite (aka “Enterprise”) data centres, colocation facilities, and hyperscale data centres.

Onsite data centres serve a single company and are usually located on this company’s premises. The company is responsible for all related task including utilities (power, cooling), facility management and IT tasks.

Colocation facilities rent spaces to multiple tenants allowing them to host their equipment there and provide power, cooling, monitoring, and Wide Area Network (WAN) connectivity. Further services like e.g. installation or reboot can also be provided by the colocation facility provider. Retail colocation facilities typically provide space in racks while wholesale ones provide more flexible floor space where for example own racks can be placed. Services provided by colocation data centres are evolving [103] and converging towards what it offered by hyperscale data centres, described below.

Hyperscale data centres are the ones used by hyperscale cloud providers like e.g. Microsoft and Amazon Web Services (AWS) to physically host the Information and Communication Technology (ICT) equipment for their services. These data centres are not intended to host any tenant-specific equipment. The equipment is selected and deployed by the hyperscale cloud provider in a very large volume and usually with the capability to be extended when needed. Customers of the service obtain shared access to CPUs, memory, network connectivity, and storage to host their virtualized applications on virtual machines or, today more commonly, containers.

---

64 This is a broader field than MEC, as MEC is limited to access through mobile radio networks while „Edge“ in general also covers fixed broadband, e.g. through Digital Subscriber Line (DSL)

65 Here and in the following the term „responsible“ always includes the option to outsource the task to a subcontractor

66 Containers are a lightweight alternative to deploying whole virtual machines. They run on a host operating system managed by the hyperscale cloud provider but include the actual application and all supporting
In [104] four different types of data centre interconnection are discussed, namely peering exchange, cross connect, inter-site connectivity, and Blended IP.

Peering exchange means that two networks, e.g. the WANs serving the respective data centres, are directly interconnected without a third party involved. This is only possible if at least one pair of border routers, each belonging to one of the networks, can be physically connected through a physical or virtual link.

Cross connect is available in case of colocation data centres allowing to directly connect servers from different tenants e.g. by butting them in the same Local Area Network (LAN).

Inter-site connectivity is the generic term describing the case where multiple data centres need to be interconnected and peering exchange is not possible. A variety of such services exist though WAN providers.

Blended IP is an extension of inter-site connectivity, where multiple inter-site connectivity providers are combined to interconnect sites while still providing better than best effort QoS guaranteed through SLAs.

Many hyperscale cloud providers like provide “on-ramp” services where the tenant can manage its interconnections. Those are often realized through Blended IP interconnections.
# C. Common Legend for Architecture Figures

<table>
<thead>
<tr>
<th>Icon / Graphical Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>_______</td>
<td>User plane interface or transport link(s) within a local or wide area network</td>
</tr>
<tr>
<td>_______</td>
<td>Control plane interface specified by 3GPP</td>
</tr>
<tr>
<td>_______</td>
<td>Proprietary interface not defined by normative 3GPP specifications. It can be defined by other bodies or commonly used software implementations</td>
</tr>
<tr>
<td>_______</td>
<td>MANO control plane interface</td>
</tr>
<tr>
<td>_______</td>
<td>Optional user plane transport link(s)</td>
</tr>
<tr>
<td>_______</td>
<td>Crossing links are connected</td>
</tr>
<tr>
<td>_______</td>
<td>Crossing links are not connected but there was no other feasible option to draw it without this overlap</td>
</tr>
<tr>
<td><strong>PNF</strong></td>
<td>Physical Network Function (PNF) according to [105]. “PNF” is replaced by the name of the PNF (spelled out or defined acronym)</td>
</tr>
<tr>
<td><strong>VNF</strong></td>
<td>Virtual Network Function (VNF) according to [105]. “VNF” is replaced by the name of the VNF (spelled out or defined acronym). End user applications are not considered VNFs (see next row) according to [105] but contrary to [106] calling them “Mobile Edge App VNF”.</td>
</tr>
</tbody>
</table>
| Application hosted in virtualized environment (Cloud) that can, but does not have to, be managed by the same MANO system as the VNFs (see row above). “APP” can be replaced by the name of the application, preferably according to the Application Architecture in Deliverable D3.1 [3]. “Place” can be optionally used to provide additional deployment[67] information, e.g. “MEC” (see Section 4.3).

| Virtual Network Function (VNF) related to MANO and SDN. “MANO VNF” is replaced by the name of the function (spelled out or defined acronym).

| The blue dot represents the APN endpoints. One is on the UE side, the other at the P-GW (or UPF PDU Session Anchor for 5G Core). Different colors are used in case of multiple APNs.

---

[67] Terminologies for different deployment options (server locations) will be defined after more information on possible locations in real mobile radio networks is obtained and related terms describing them.