5GAA Automotive Association

Predictive Edge Analytics and Network Slicing Enabling Mobility-as-a-Service in Global MEC Scenarios

5GAA Automotive Association
Technical Report

CONTACT INFORMATION:
Lead Coordinator – Thomas Linget
Email: liaison@5gaa.org

MAILING ADDRESS:
5GAA c/o MCI Munich
Neumarkter Str. 21
81673 München, Germany
www.5gaa.org

VERSION:	1.0
DATE OF PUBLICATION:	
DOCUMENT TYPE:	Technical Report
EXTERNAL PUBLICATION:	Yes
DATE OF APPROVAL BY 5GAA BOARD:	

Contents

Foreword

This Technical Report has been produced by 5GAA.

The contents of the present document are subject to continuing work within the Working Groups (WG) and may change following formal WG approval. Should the WG modify the contents of the present document, it will be re-released by the WG with an identifying change of the consistent numbering that all WG meeting documents and files should follow (according to 5GAA Rules of Procedure):

x-nnzzzz

(1) This numbering system has six logical elements:

(a) x: a single letter corresponding to the working group:
where x =
T (Use cases and Technical Requirements)
A (System Architecture and Solution Development)
P (Evaluation, Testbed and Pilots)
S (Standards and Spectrum)
B (Business Models and Go-To-Market Strategies)

(b) nn: two digits to indicate the year. i.e. ,17,18 19, etc

(c) zzz: unique number of the document

(2) No provision is made for the use of revision numbers. Documents which are a revision of a previous version should indicate the document number of that previous version

(3) The file name of documents shall be the document number. For example, document S-160357

Contents

Introduction

This Technical Report (TR) provides an analysis of how **predictive edge analytics** can be generated and delivered in distributed mobile edge computing (MEC) deployments. In particular, when such deployments involve multiple mobile network operators (MNO), original equipment manufacturers (OEM) and additional third parties such as service providers, application developers, internet provider (IP) interconnect operators and road transport authorities (RTA), the generation and delivery of such predictive edge analytics is not a trivial task and may require interaction across different stakeholders.
The use of predictive edge analytics and situation awareness can warn the application about potentially undesirable effects, including poor user experience, limited support of selected features or when the service could be no longer available. This early notification enables ‘closed-loop adaptation’ to be implemented. Before the issue arises. In this way the application may continue its operation minimizing the undesired effects.
Furthermore, the introduction of **network slices**, based on sensed context awareness, is an additional tool that can be used in heterogeneous MEC systems for enhanced and more predictable network performance. This TR analyses the state of the art of network slicing in the context of MEC systems and tries to identify potential gaps and additional requirements for future study.

1. Scope

Edge computing is an important topic in the vehicle-to-everything (V2X) field, as many use cases [3] require ultimate latency and reliability as well as significant data exchange from and between vehicles and infrastructure. The support of specific performance requirements is key for the realisation of those use cases. When such requirements are not fulfilled by the underlying network and infrastructure, the application may require closed-loop adaptation in order to cope with the potential undesired effects, which may also include poor user experience, limited support of selected features or indeed service unavailability.

The usage of predictive edge analytics and situation awareness helps provide early notifications to the application about potentially undesirable events or situations. It enables closed-loop adaptation which may limit or avert the effects of a degradation in performance affecting the network and/or MEC infrastructure. Previous 5GAA Work Items [1] have studied how automotive applications may perform such 'service adaptation' based on early notifications. Other 5GAA Work Items [10], [11] have studied how early notifications may be generated in the 5GS and delivered to the application. However, no previous work has studied how predictive edge analytics can be generated and delivered in distributed MEC deployments. In particular, when such deployments involve multiple MNOs, OEMs and additional third parties, such as service providers, application developers, IP interconnect operators and RTAs, the generation and delivery of such predictive edge analytics is not a trivial task and may require interaction across different stakeholders.

The delivery of predictable performance requirements is associated with the concept of Service Level Specifications (SLS) and Service Level Agreements (SLA): it depends also on how the network resources are allocated and utilised in the network and in the MEC environment. For this reason, network slicing is an additional tool that can be used to implement logical networks that can provide specific capabilities and characteristics [19]. GSMA has defined the customisable network capabilities that can be associated with the service provided by a network slice. These capabilities are provided based on an SLA between the mobile operator and the business customer [20]. Moreover, usage and adaptation of network slices based on sensed context awareness is an additional tool for use in heterogeneous MEC systems for enhanced and more predictable network performance. This study analyses the state of the art of network slicing in the context of MEC systems and tries to identify potential gaps and additional requirements for future study.

2. References

[1]	5GAA T-220058 PRESA Technical Report "Predictive QoS and V2X Service Adaptation". Available at: https://5gaa.org/news/5gaa-technical-report-on-predictive-qos-and-v2x-service-adaptation/
[2]	5GAA A-200150 MEC4AUTO Task2 Technical Report "MEC-for-Automotive-in-Multi-Operator-Scenarios". Available at: https://5gaa.org/wp-content/uploads/2021/03/5GAA_A-200150_MEC4AUTO_Task2_TR_MEC-for-Automotive-in-Multi-Operator-Scenarios.pdf
[3]	5GAA A-200146 MEC4AUTO Technical Report "Use cases and initial test specifications review". Available at: https://5gaa.org/wp-content/uploads/2021/07/5GAA_MEC4AUTO.pdf
[4]	5GAA T-190032 "Use case Description Cooperative Lane Merge", v1.1.
[5]	3GPP Technical Specification 22.886 "Study on enhancement of 3GPP Support for 5G V2X Services (Release 16)". Available at: https://www.3gpp.org/DynaReport/22886.htm
[6]	3GPP Technical Specification 23.288 "Architecture enhancements for 5G System (5GS) to support network data analytics services (Release 17)". Available at: https://www.3gpp.org/DynaReport/23288.htm
[7]	ETSI GS MEC 030 "MEC V2X Information Service API" v2.1.1. Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/030/02.01.01_60/gs_MEC030v020101p.pdf
[8]	ETSI White paper #36, "Harmonising standards for edge computing – A synergised architecture leveraging ETSI ISG MEC and 3GPP specifications", 1st edition, July 2020. Available at: https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_wp36_Harmonising-standards-for-edge-computing.pdf
[9]	5GAA A-220928 gMEC4AUTO Task 1 Technical Report "Moving toward federated MEC demos/trials (global MEC)", 5gaa.org/moving-toward-federated-mec-demos-trials
[10]	5GAA A-191007 NESQO Technical Report "Architectural_Enhancements for QoS_ Predictability".
[11]	5GAA A-200055 eNESQO Technical Report "5GS Enhancements for Providing Predictive QoS in C-V2X". Available at: https://5gaa.org/wp-content/uploads/2020/05/5GAA_A-200055_eNESQO_TR_final.pdf
[12]	ETSI ISG MEC MEC(21)000622r1 "Reply LS to 5GAA WG4 on Enhancements on ETSI MEC 030 to Support end to end Predictive Edge Analytics", December 10, 2021.
[13]	3GPP Technical Specification 23.287 "Architecture enhancements for 5G System (5GS) to support Vehicle-to-Everything (V2X) services" (Release 17). Available at: https://www.3gpp.org/DynaReport/23287.htm
[14]	5GAA A-200094 Technical Report "V2X Application Layer Reference Architecture (V2XSRA)" available at https://5gaa.org/wp-content/uploads/2020/06/5GAA_A-200094_V2XSRA-Application-Layer-Reference-Architecture-final.pdf
[15]	5GAA 211007 "PRESA Questions to WI ToD on PQoS with responses".
[16]	ETSI Group Specification MEC 003, "Multi-access Edge Computing (MEC); Framework and Reference Architecture". Final draft GS MEC 030 v3.1.1 (2022-03). Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/03.01.01_60/gs_MEC003v030101p.pdf
[17]	ETSI Group Specification MEC 040, "Multi-access Edge Computing (MEC); Federation enablement APIs". Draft GS MEC 030 v3.1.1 (2023-02). Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/040/03.01.01_60/gs_MEC040v030101p.pdf
[18]	GSMA PRD, Permanent Reference Document, OPG.02, "GSMA Operator Platform Telco Edge Requirements – October 22",Version 3.0, 03 October 2022 Available at: https://www.gsma.com/futurenetworks/wp-content/uploads/2022/10/Operator-Platform-Telco-Edge-Requirements.-v.3-October22.pdf
[19]	3GPP Technical Specification 23.501 "System architecture for the 5G System (5GS); Stage 2, (Release 17)" V17.2.0 (2021-09). Available at: https://www.3gpp.org/DynaReport/23501.htm
[20]	GSMA NG.116 "Generic Network Slice Template Version 5.0", 01 June 2021. Available at: https://www.gsma.com/newsroom/wp-content/uploads/NG.116-v5.0-8.pdf

[21]	3GPP Technical Specification 28.541 "5G Network Resource Model (NRM); Stage 2 and stage 3 (Release 16)", V16.4.1 (2020-03). Available at: https://www.3gpp.org/DynaReport/28541.htm
[22]	3GPP Technical Report 23.794 "Study on enhanced IP Multimedia Subsystem (IMS) to 5GC integration (Release 17)", V17.0.0 (2019-12). Available at: https://www.3gpp.org/DynaReport/23794.htm
[23]	5GAA A-180295 "NESQO Update of Analysis of Tele-Operated Driving Use case", Ericsson, Oct 2018.
[24]	3GPP Technical Specification 23.548 "5G System Enhancements for Edge Computing; Stage 2, (Release 17)", V17.0.0 (2021-09). Available at: https://www.3gpp.org/DynaReport/23548.htm
[25]	5G-ACIA whitepaper on "5G for Connected Industries and Automation", Second Edition, Feb. 2019. Available at: https://5g-acia.org/wp-content/uploads/5G-ACIA_WP_5G-for-Connected-Industries-and-Automation-Second-Edition_SinglePages.pdf
[26]	GSMA, "Network slicing Use case Requirements", April 2018.Available at: https://www.gsma.com/futurenetworks/wp-content/uploads/2018/07/Network-Slicing-Use-Case-Requirements-fixed.pdf
[27]	3GPP Technical Report 23.700-41 "Study on enhancement of network slicing; Phase 3". Available at: https://www.3gpp.org/DynaReport/23700-41.htm
[28]	OGC Standards, available at https://www.ogc.org/docs/
[29]	3GPP Technical Report 23.558 "Architecture for enabling Edge Applications (Release 17)". Available at: https://www.3gpp.org/DynaReport/23558.htm
[30]	3GPP Technical Report 23.748 "Study on enhancement of support for Edge Computing". Available at: https://www.3gpp.org/DynaReport/23748.htm
[31]	3GPP Technical Report 23.758 "Study on application architecture for enabling Edge Applications". Available at: https://www.3gpp.org/DynaReport/23758.htm
[32]	ETSI Group Specification MEC 012 "Multi-access Edge Computing (MEC);Radio Network Information API" V2.2.1 (2022-02). Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/012/02.02.01_60/gs_MEC012v020201p.pdf
[33]	3GPP Technical Specification 29.520 " 5G System; Network Data Analytics Services; Stage 3". Available at: https://www.3gpp.org/DynaReport/29520.htm
[34]	GSMA NG.127- "E2E Network slicing Architecture", Version 1.0-2, 03 June 2021. Available at: https://www.gsma.com/newsroom/wp-content/uploads/NG.127-v1.0-2.pdf
[35]	ETSI Technical Specification 102.894-2 "Intelligent Transport Systems (ITS);Users and applications requirements; Part 2: Applications and facilities layer common data dictionary". Available at: https://www.etsi.org/deliver/etsi_ts/102800_102899/10289402/02.01.01_60/ts_10289402v020101p.pdf
[36]	5GAA Technical Report STiCAD "Safety Treatment in Connected and Automated Driving Functions". Available at: https://5gaa.org/news/sticad-safety-treatment-in-connected-and-automated-driving-functions/
[37]	5GAA Technical Report ToD "Tele-Operated Driving (ToD): System Requirements Analysis and Architecture". Available at: https://5gaa.org/wp-content/uploads/2021/09/5GAA_ToD_System_Requirements_Architecture_TR.pdf
[38]	5GAA White paper "A visionary roadmap for advanced driving use cases, connectivity technologies, and radio spectrum needs". Available at: https://5gaa.org/news/5gaa-publishes-updated-2030-roadmap-for-advanced-driving-use-cases-connectivity-technologies-and-radio-spectrum-needs/
[39]	3GPP Technical Report 23.700-91 "Study on enablers for network automation for the 5G System (5GS); Phase 2". Available at: https://www.3gpp.org/DynaReport/23700-91.htm
[40]	3GPP Technical Report 23.700-81 "Study on Enablers for Network Automation for 5G; Phase 3". Available at: https://www.3gpp.org/DynaReport/23700-81.htm
[41]	ETSI GS MEC 030 "MEC V2X Information Services API" v3.1.1. Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/030/03.01.01_60/gs_mec030v030101p.pdf

3. Abbreviations

For the purposes of the present document, the following abbreviations apply:

G3GPP	3rd Generation Partnership Project
5GAA	5G Automotive Association
5GC	5G Core
5GS	5G System
5QI	5G QoS Identifier
AC	Application Client
ACR	Application Context Relocation
AECC	Automotive Edge Computing Consortium
AF	Application Function
AMF	Access and Mobility Management function
API	Application Programming Interface
ARP	Allocation and Retention Priorities
B2B	Business-to-business
CAM	Cooperative Awareness Message
CAPIF	Common API Framework
CLC	Cooperative Lane Change
CN	Core network
DENM	Decentralised Environmental Notification Message
DN	Data Network
DNAI	Data Network Address Identifier
DNN	Data Network Name
DNS	Domain Name System
E2E	End-to-End
EAS	Edge Application Servers
EASDF	Edge Application Server Discovery Function
EC	Edge Computing
ECS	Edge Configuration Server
ECSP	Edge Computing Service Provider
EDN	Edge Data Network
EEC	Edge Enabler Client
EES	Edge Enabler Servers
EHE	Edge Hosting Environment
eMBB	Enhanced Mobile Broadband
eNB	evolved Node B
eNESQO	enhanced NESQO
ETSI ISG	ETSI Industry Specification Group
ETSI	European Telecommunications Standards Institute
EWBI	East/Westbound Service Interface
GBR	Guaranteed Bit Rate
GFBR	Guaranteed Flow Bit Rate
gNB	Next Generation Node B
GNSS	Global Navigation Satellite Systems
GSMA OPG	GSM Association Operator Platform Group
GST	Generic Network Slice Template

HMTC	High-Performance Machine-Type Communications
HPLMN	Home Public Land Mobile Network
HTTP	Hypertext Transfer Protocol
HV	Host Vehicle
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IQN	In-advance Quality of Service Notification
ISG	Industry Specification Group
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LBO	Local Breakout
LS	Liaison Statement
MANO	Management and Orchestration
ME	Mobile Edge
MEAO	Mobile Edge Application Orchestrator
MEC	Multi-access Edge Computing
MEC4AUTO	MEC for Automotive
MEO	Multi-access Edge Orchestrator
MEP	MEC Platform
MEPM	MEC Platform Manager
MIoT	Massive IoT
MNO	Mobile Network Operator
MOS	Mean Opinion Score
MSP	Mobility Service Provider
NAT GW	Network Address Translation GW
NBI	Northbound Service Interface
NEF	Network Exposure Function
NESQO	Predictable QoS and E2E Network slicing for Automotive Use cases
NEST	NEtwork Slice Type
NF	Network Function
NFV	Network Function Virtualisation
NFVI	Network Function Virtualisation Infrastructure
NG-RAN	Next Generation RAN
NMS	Network Management System
NSaaS	Network slice as a service
NSAC	Network Slice Admission Control
NSC	Network Slice Customer
NSP	Network Slice Provider
NWDAF	Network Data Analytics Function
OAM	Operation and Maintenance
OEM	Original Equipment Manufacturer
OP	Operator Platform
OSS	Operations Support Systems
OTT	Over-the-Top
P-QoS	Predictive QoS
PCF	Policy Control Function
PDB	Packet Delay Budget
PDU	Protocol Data Unit

PER	Packet Error Rate
PF	Prediction Function
PGW	PDN Gateway
PNF	Physical Network Function
POI	Point of Interconnection
PoP	Point-of-Presence
PRESA	Predictive QoS and V2X Service Adaptation
PSA	PDN Session Anchor
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RNIS	Radio Network Information Service
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSU	Road Side Units
RTA	Road Transport Authority
RV	Remote Vehicle
SBI	Southbound Service Interface
SBMA	Service Based Management architecture
SDO	Standard Developing Organisation
SIM	Subscriber Identity Module
SLA	Service Level Agreement
SLS	Service Level Specification
SMF	Session Management Function
S-NSSAI	Single Network Slice Selection Assistance Information
SP	Service Provider
SST	Slice Service Type
STICAD	Safety Treatment in Connected and Automated Driving Functions
TA	Tracking Area
TM Forum	TeleManagement Forum
TN	Transport Network
ToD	Tele-operated Driving
TR	Technical Report
TS	Technical Specification
UE	User Equipment
UL CL/BP	Uplink Classifier Branching Point
UL	Uplink
UNI	User to Network Interface
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communications
USRP	UE Route Selection Policy
V2X	Vehicle-to-Everything
VAF	Virtualised Application Functions
VIS	V2X Information Services
VNF	Virtual Network Function
VPLMN	Visited Public Land Mobile Network
VxFs	Virtualisation Everything Functions
WI	Work Item
ZSM	Zero touch network & Service Management

4 The state of the art

4.1 State of the art on predictive analytics in MEC systems

This section summarises the main conclusions achieved by previous 5GAA Work Items, studies and papers available in the topic of analytics, QoS prediction and network and adaptation that can be relevant in the context of MEC systems. The following associations and SDOs have produced relevant work in the context of predictive analytics and MEC technology:

- 5GAA
- 3GPP
- ETSI MEC
- GSMA

The list is non-exhaustive and may be updated in the future. The next sections summarise the state of the art produced by these entities because it is relevant to later topics discussed in this Technical Report.

4.1.1. 5GAA state of the art

5GAA has run several Work Items in the context of predictive analytics, with special focus on introducing QoS prediction and in-advance QoS notifications into the 5GS and V2X applications. Relevant WIs include:

- NESQO [10] was the first Work Item to focus on the architectural enhancements required within 5GS to introduce predictive QoS (P-QoS) and in-advance QoS notifications (IQNs). While in the baseline 5GS (Release 15 or earlier), an application could only react to QoS changes, which happen without prior notice to the application, the introduction of predictive QoS enables a new 'proactive' behaviour to deal with situations where QoS is predicted to change from what was pre-agreed, thanks to the introduction of IQNs. The proposals introduced by NESQO in this scope included:
  - A set of requirements for the 3GPP system and the 5GS described in Section 5.3 of [10], which are the result of a use case analysis of six priorities in Section 5.2 of [10]. Those requirements are divided into functional requirements, non-functional requirements, and requirements for the 5GS.
  - A high-level procedure for QoS prediction, described in Section 6.1 of [10], where the application and the network can cooperate and exchange information in order to implement P-QoS. Such a procedure also describes the high-level information flow and data exchanges by different parties.
  - A proposed content and structure for the IQN is described in Section 6.4 of [10].

- A set of proposed metrics described in Section 5.2 of [10] are relevant for P-QoS specified for each of the six prioritised use cases including – among others – the KPIs to be predicted and included in the IQN, based on information availability, relevant time horizons, and the goal of such prediction.
- A number of liaison statements (LS) between 5GAA and 3GPP SA1 and SA2. 5GAA issued the requirements to 3GPP SA1 and SA2, and later several information exchanges were needed to achieve a common understanding of those requirements and P-QoS features. Following this work 3GPP later introduced a solution for V2X application adaptation based on QoS sustainability analytics in Release 16.
- On edge computing, NESQO investigated edge deployment of user plane functions (UPF) for C-V2X applications, in combination with network slicing and an evolved QoS framework.

3 eNESQO [11] continued the work of NESQO on how to make QoS predictions and on the use of QoS prediction in automotive applications. eNESQO concluded reviewing current 3GPP solution for QoS sustainability analytics and providing a gap analysis or areas of improvements, summarised in section 5.6 of [11].

3 PRESA [1] is the last 5GAA WI that continued on P-QoS, with the main focus on how the V2X applications can implement service adaptations. PRESA developed a methodology to study how adaptations can be implemented in the context of any 5GAA-defined use case. Some use cases were selected for detailed analysis in order to determine the service adaptation and relevant time requirements for the prediction time horizon. PRESA also analysed 3GPP Release 18's ongoing work on predictive analytics in light of 5GAA requirements and relevant gaps identified for future study.

3 STiCAD [36] tried to determine, propose and evaluate possibilities for telecommunication operators, vendors, and any further identified stakeholders to provide what is necessary in order to enable OEMs to better treat safety for systems that exist beyond a single vehicle. To achieve this, it was decided to find representative safety requirements for two selected use cases that cover the V2X scenarios of direct communication and network-based information delivery. Predictive QoS has been identified as a network requirement as part of the strategies for fault avoidance.

3 Tele-operated driving [37] focused on system requirements analysis and architecture for ToD services. The technology assists, complements and accelerates semi- and fully automated driving in various scenarios. The study covered the vehicle sub-system, ToD operator sub-system, infrastructure sub-system, and C-V2X networks for end-to-end deployment of ToD services. System application layer architectures and the underlying communication network architectures for different ToD use cases and scenarios are presented in this study with the focus on interfaces among different stakeholders and with considerations on service interoperability in multi-OEM, multi-service provider, multi-RTA and multi-MNO environments. This study also envisages that market deployment of ToD services will follow a multi-stage roadmap,

starting from confined areas, then evolving into dedicated public roads and areas, and finally covering the cross-regional (long-haul) mobility of automated vehicles. The study also investigates how QoS influences the settings of the driving behaviour, e.g. level of automation, trajectory, speed, inter-vehicle distance, and how predictive QoS can be an effective tool for ToD applications to adapt behaviours, improve service availability and the ToD user experience.

3 Finally, the 5GAA White Paper “A visionary roadmap for advanced driving use cases, connectivity technologies, and radio spectrum needs” [38] synthesised the 5GAA vision of the future and developed the Association’s forward-looking C-V2X roadmap in its latest version, until 2030. It focuses on advanced driving use cases which pave the way to automated driving, teleoperation, automated valet parking (AVP), and sensor sharing – all of which contribute to global safety, mobility, environmental stewardship and transportation equity goals. The Paper identifies MEC as an enabler for automated driving functions, such as AVP, especially to reduce efforts and complexity in the operation of such a service. In the Paper predictive QoS is listed as a 5GS network enhancement which can benefit certain automotive use cases, while network slicing is considered an important tool for separating network resources in order to provide a more consistent service and QoS.

4.1.2. 3GPP state of the art

3GPP has introduced in Release 16 the network data analytics function (NWDAF) which provides network analytics for several purposes, to be used in the network and in the application. Later, in the edge computing specific studies, 3GPP has started to investigate how analytics can be used in the functionalities that are typical of edge deployments.

The relevant 3GPP specification for use of NWDAF to predict QoS for V2X services is the following 3GPP Technical Specifications:

3 3GPP TS 23.287 “Architecture enhancements for 5G System (5GS) to support Vehicle-to-Everything (V2X) services” [13].
3 3GPP TS 23.288 “Architecture enhancements for 5G System (5GS) to support network data analytics services” [6].

The relevant 3GPP specifications for the use of predictive analytics in MEC environment are the following 3GPP Technical Specifications:

3 3GPP TS 23.548 “5G System Enhancements for Edge Computing” [24].
3 3GPP TS 23.558 “Architecture for enabling Edge Applications” [29].

The following 3GPP studies are also relevant as some of the solutions propose the use of predictive analytics:

3 3GPP TR 23.748 “Study on enhancement of support for Edge Computing” [30].
3 3GPP TR 23.758 “Study on application architecture for enabling Edge Applications” [31].

The following specification provides Stage 3 details on the network analytics data services:

- 3GPP TS 29.520 "5G System; Network Data Analytics Services; Stage 3" [33]

Finally, the following studies have brought several proposed enhancements to NWDAF in the 3GPP System, some of those have been selected for normative respectively in Release 17 and Release 18:

- 3GPP TR 23.700-91 "Study on enablers for network automation for the 5G System (5GS); Phase 2" [39].
- 3GPP TR 23.700-81 "Study on Enablers for Network Automation for 5G; Phase 3" [40].

4.1.2.1. Network analytics used in 5GS to support V2X services

3GPP TS 23.287[13] describes how to use the analytics provided by NWDAF to support QoS sustainability analytics for V2X services including autonomous driving.

As described in Clause 5.4.5.2 about notification on QoS sustainability analytics to the V2X application server (AS), it has been specified that the AS server may request notifications 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 adjusting the application (how the AS makes use of such data is outside of 3GPP scope).

From a functional perspective, the potential QoS change to assist application adjustment is based on the notification of QoS sustainability analytics defined in clause 6.9 of TS 23.288 [6]. The V2X AS acting as an 'application function' communicates with the NEF, which corresponds to the NF consumer in clause 6.9.4 of TS 23.288 [6].

4.1.2.2. Relevant network analytics in 5GS

3GPP TS 23.288 [6] describes the following analytics that can be relevant in MEC systems:

- **Observed service experience analytics**, described in clause 6.4, can provide observed service experience (i.e. average observed service or mean opinion score (MOS) and/or variance of observed service MOS indicating the distribution of services such as audio-visual (AV) streaming as well as non-AV streaming (i.e. V2X and web browsing services) analytics, in the form of statistics or predictions, to a service consumer. These analytics can provide the following sub-types:
  - Service experience for a network slice: service experience for a user equipment (UE) or a group of UE or any UE in a network slice;
  - Service experience for an application: service experience for a UE or a group of UE or any UE in an application or a set of applications;
  - Service experience for an edge application over a User Plane (UP) path: service experience for a UE or a group UEs or any UE in an application or

a set of applications over a specific UP path (UPF, DNAI and EC server);

- Service experience for an application over a radio access technology (RAT) type or frequency: service experience for a UE or a group of UEs in an application or a set of applications over a specific RAT type and/or frequency.

3 **Network performance analytics**, described in clause 6.6, provides either statistics or predictions on next-generation node B (gNB) status information, gNB resource usage, communication performance and mobility performance in an Area of Interest; in addition, NWDAF it may provide statistics or predictions on the number of UEs that are located in that area of interest.

3 **QoS sustainability analytics** described in clause 6.9, provides analytics information regarding the QoS change statistics for an analytics target period in the past in a certain area or the likelihood of a QoS change for an analytics target period in the future in a certain area.

3 **DN performance analytics** described in clause 6.14, provides analytics for user plane performance (i.e. average/maximum traffic rate, average/maximum packet delay, average packet loss rate) in the form of statistics or predictions to a service consumer.

3 **UE related analytics** described in clause 6.7, such as UE mobility analytics, UE communication analytics, expected UE behavioural parameters related network data analytics, abnormal behaviour related network data analytics.

With the exception of QoS sustainability analytics [13], 3GPP does not specify how an application can make use of analytics, e.g. adapting application behaviour. It is interesting to investigate how 5GS can help in the generation of such analytics in the context of MEC deployments, especially those involving multiple domains.

4.1.2.3. Network analytics used in 5GS supporting edge computing specifications

At least two 3GPP technical specifications can be mentioned in this context:

3 3GPP TS 23.548 [24] defines the Stage 2 specifications for enhancing the 5GS to support edge computing, limited to non-roaming and local breakout (LBO) roaming scenarios. The specification assumes an edge hosting environment (EHE) is deployed in the data network (DN) beyond the PDN session anchor for the user plane function (PSA UPF). In this local breakout connectivity model for edge computing, a protocol data unit (PDU) session has a PSA UPF in a central site (C-PSA UPF) and one or more PSA UPFs in the local site (L-PSA UPF). The C-PSA UPF provides the IP anchor point when an uplink (UL) classifier is used. The edge computing application traffic is selectively diverted to the L-PSA UPF using the UL classifier or multi-homing branching point mechanisms. The L-PSA UPF may be changed due to, for example, UE mobility. In such a configuration, 3GPP has defined analytics in the procedure for edge application server (EAS) (re-)discovery over the session breakout connectivity model. Clause 6.2.3

describes the EAS discovery procedure with EASDF, i.e. when the UE performs a DNS query for the EAS which is handled by the EASDF, the session management function (SMF) may perform UPF selection based on service experience or DN performance analytics for an edge application, as described in TS 23.288 [6]. The SMF may use such information to determine the DNAI and the associated N6 traffic routing information for the DNAI. In fact in step 16 of cl. 6.2.3.2.2, SMF may perform UL CL/BP and Local PSA selection and insert UL CL/BP and Local PSA based on EAS information received from EASDF, other UPF selection criteria and the selected analytics for an edge application.

3 3GPP TS 23.558 [29] defines several scenarios for utilising analytics. One relates to the edge enabler server, which provides supporting functions needed for EASs and edge enabler clients (EEC). Its functionality includes providing configuration information to EECs, interacting with the 3GPP core network, registering the EEC and EAS, etc. The EES may subscribe to the "UE expected behaviour analytics" (UE mobility), which can be used when serving UE location subscribe request coming from EAS or for application context relocation (ACR) management events.

4.1.2.4. Network analytics usage related to 3GPP studies (non-normative)

The following studies describe other analytics usage in the 3GPP system:

3 3GPP TR 23.748 [30] studies and performs evaluations of potential architecture enhancements to support edge computing in the 5G core network (5GC). As part of this study, "Key Issue #3: Network information provisioning to local applications with low latency" addresses exposure of information to application functions deployed in the edge (e.g. EAS). In "Solution #44 (KI#3): Network information exposure to local AF with low latency", the network exposure function deployed in the edge (NEF-Edge) may report analytics to the edge application, together with other exposed information.

3 3GPP TS 23.758 [31] is the 3GPP study on application architecture for enabling edge applications. In this study, the following solutions reference the use of analytics in different contexts:

  - "Solution #11 (KI#13): QoS management for 5G edge applications" refers to the use of the QoS sustainability analytics procedure in clause 5.4.5.2.2 of TS 23.287 between the NEF and an AF (in this case EAS). Such analytics can be used to notify the EAS of potential QoS changes for an indicated geographic area and time interval in order to adjust the application behaviour in advance of potential QoS change.
  - "Solution #27 (KI#9): Relocation of application context considering analytics" considers how to exploit analytics to help the source EES select the optimal target edge data network.

4.1.3. ETSI MEC state of the art

ETSI has standardised a MEC V2X information service (VIS) [7] that, among other functionalities, can provide QoS predictions to service consumers. VIS can be implemented by a MEC platform or a MEC application. Examples of a service consumer can be a MEC application, a MEC platform or a V2X AS deployed or integrated with the MEC infrastructure. In a different configuration, a service consumer can also be a vehicle application (e.g. V2X application) or an OEM application that can access the service via a MEC application.

This service is quite interesting for the automotive industry, not only because it is has also been specified by ETSI ISG MEC, but also because it has been conceived as a service for dispatching journey-specific QoS predictions relating to a particular route or set of routes, thus matching the scenario of vehicle UEs moving across one or multiple MNO networks. In fact, the VIS service consumer – the entity requesting the QoS prediction – can indicate a set of potential routes for a vehicle UE: the routes can be described, for example with an “origin” and “destination” but also optionally through a set of intermediate “waypoints”. Moreover, the VIS service consumer can request a QoS prediction according to a point or time when it estimates the vehicle UE to be at a specific location (times can be associated with every location along the route (origin, destination or intermediate waypoint). Another reason this API is interesting is because ETSI ISG MEC has confirmed that the VIS can evolve to provide QoS prediction for the end-to-end user plane link also in multi-domain (e.g. multi-MNO or multi-OEM) environments [12].

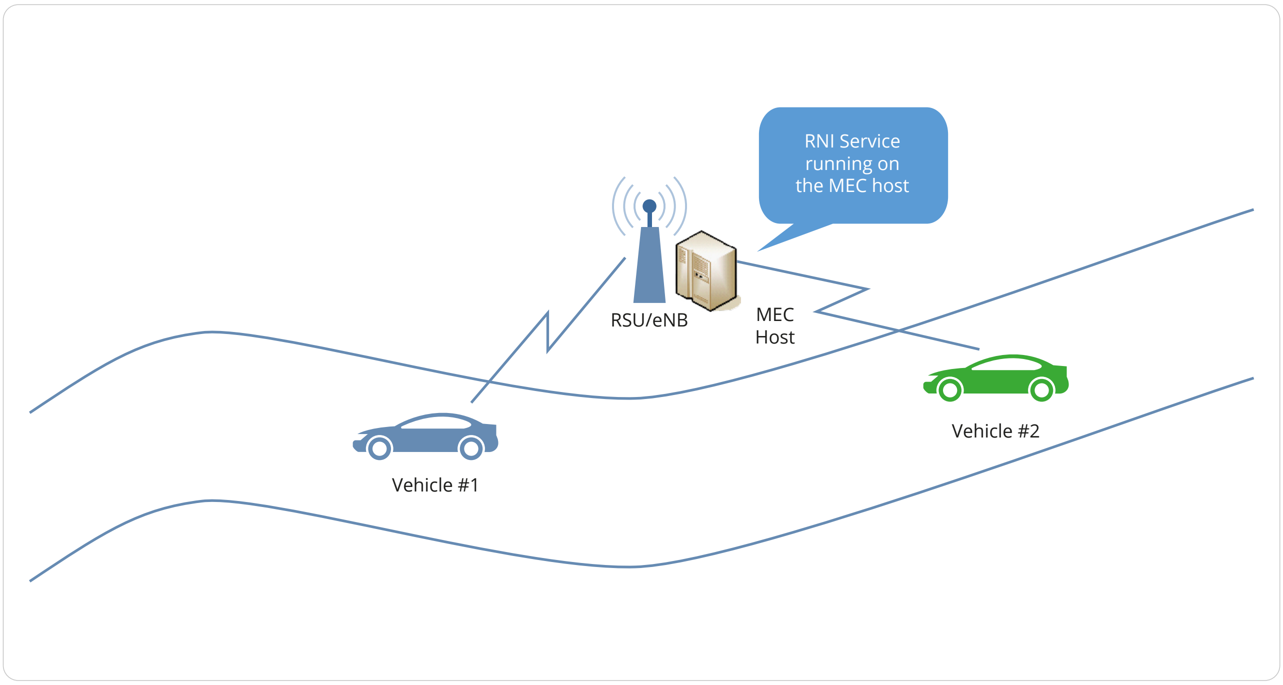

Figure 1: Exemplary V2X system scenario where the MEC host is deployed in collocation with a roadside unit (RSU)/eNB providing coverage (V2X communication) (Source [7])

Proposed 5GAA enhancements to the VIS API have been described in Section 5.4 of this Technical Report.

4.1.4. GSMA state of the art

GSMA described in details the requirements and provided data abstraction and attributes for a generic slice template (GST) [20]. The quality of service of the network slice is one key feature that is represented by GST. The QoS attributes below are typical information provided by GST:

- Service availability;
- Area of service;
- Delay tolerance;
- Downlink throughput per network slice;
- Downlink maximum throughput per UE;
- Uplink throughput per network slice;
- Uplink maximum throughput per UE;
- Latency from (last) UPF to application server.

Performance prediction is another key QoS attribute of the GST. This attribute defines the capability allowing the mobile system to predict the network and service status. Predictive QoS can be done for various key quality indicators (KQIs) and key performance indicators (KPIs). KQIs reflect the end-to-end service performance and quality, while KPIs reflect the performance of the network. The prediction is done for a specific point of time in the future and for a specific geolocation. Only the KQIs of communication services offered by the network slice provider (NSP) can be predicted. For over-the-top (OTT) services, the NSP is not able to access the KQIs. The first parameter to be considered for the performance P-QoS is the so-called "prediction availability", which contains a list of KQIs and KPIs available for prediction. If the list is empty, the parameter is not available in the network slice and the other parameters might be ignored. The availability is applied to the throughput, latency and service success rate. The prediction frequency is another important attribute of the GST's performance prediction.

The prediction frequency describes how often KQI and KPI values are provided. Typical timescales considered for the predicted values are per second, per minute or per hour. Additional information should be also considered such as prediction nature, i.e. active prediction where the network actively informs the NSC and/or the UE proactively about the predicted values. Alternatively, the NSC and/or UE are only informed if the predicted KPI or KQI value crossed a defined threshold or passive prediction, where the NSC and/or UE requests predictions from the network via its APIs.

A prediction (request as well as reply) is always associated with a point of time in the future and a geolocation. A prediction provided by the network slice to the UE and/ or customer (prediction reply) should always be associated with a confidence interval. The reliability of the prediction depends on many parameters, e.g. which KPI to predict, look ahead of time, etc.

The GSMA operator platform group (OPG) describes in [18] the overall architecture and requirements for building a federation of operators to optimise service deployment in multi-operator environments. The overall high-level architecture and interfaces of a typical operator platform (OP) is given below:

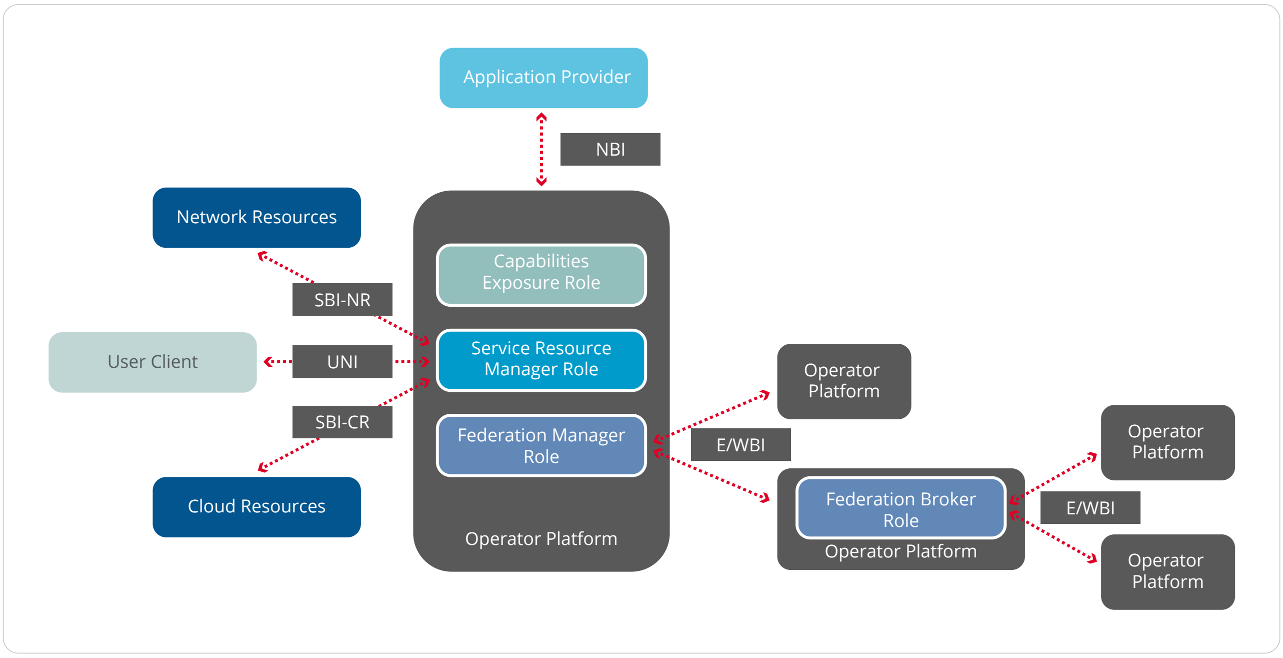

Figure 2: OP Roles and Interfaces Reference Architecture [18]

The different interfaces provided by the OP are briefly recalled in the following:

- Northbound interface (NBI): interface between the application providers. The OP "exposes" its capabilities to the application provider, such as the geographical footprint reachable via the OP, edge cloud resource catalogue exposure, application resource consumption monitoring, etc.
- Southbound interface (SBI): interface between the OP and underlying network, including edge resources and charging engine. The SBI-NR linking the network resources with the OP is of particular interest since it is responsible for receiving statistics/analytics, e.g. to influence application placement or mobility decisions, steer traffic in the mobile network towards applications orchestrated in edge clouds, and capabilities such as QoS information in the network.
- East-west bound interface (EWBI): interface that connects two OPs that are members of the same federation. The current specification of the GSMA OPG [18] defines basic features for the EWBI, for example edge cloud resource exposure and monitoring towards partner Ops, service availability in visited networks, etc. More details are found in [18].
- User to network interface (UNI): interface connecting the operator platform with the end user. The primary function of this interface is to help a user client interact with the OP, to enable the matching of an application client with an application instance on a "cloudlet" (a point of presence for the edge cloud, where edge applications are deployed)). UNI is able also to authenticate the end user and contribute to providing E2E security credentials for the system.

As described above, the network analytics and capabilities are exposed to the OP by the network through SBI-NR. This could be done, for example, through a 3GPP network exposure function (NEF) or network data analytics function (NWDAF). The EWBI that is connecting two OPs from the same federation is transmitting service availability information not information related to QoS predictions. The provision of such information over the EWBI could be one feature of the evolution of GSMA-OPG.

The above descriptions present the state of the art of predictive QoS in GSMA where the following considerations were summarised: generic slice templates with a specific focus on performance attribute of the network slice. This performance attribute is applied either to network-related QoS metrics (KPIs) or end-to-end QoS metrics (KQIs). Another important aspect is the architecture of GSMA-OPG where the different interfaces of the OPs were presented. Regarding, the specific focus of gMEC4AUTO study, i.e. multi-domain deployment and service continuity of V2X services, further exposing predictive QoS capabilities between operator platforms through East-West-bound interface may be considered in future versions of GSMA-OPG [18].

4.2. State of the art on network slicing in MEC systems

This section includes a summary of the state of the art on network slicing functionality for MEC applications according to the documents published by 3GPP, 5GAA and GSMA.

4.2.1. Introduction

One of the key technological ingredients that make up a 5G system is network slicing. A network slice is a logical network that provides specific capabilities and characteristics [19]. It may be designed to support specific functionality, specific network or customer requirements, or a specific traffic class. While there are several definitions of a network slice in literature, for the purpose of this paper the 3GPP TS 23.501 [19] definition is applied. Also it is considered that for a MEC application a network slice is identified by an S-NSSAI [19]. Network slices can be used for separating network resources and to dedicate those resources to a specific service or customer (e.g. an OEM) in order to provide a more consistent service, according to a predefined SLA. Network slices are associated with one or more network instances and network slice bundles, or business slices. These are introduced later in this report.

Objective 4 of gMEC4AUTO “Setup of network slicing for C-V2X services” has identified the following areas as relevant to be studied in the context of MEC deployment:

- **Area of study 1** – Study the topics of dynamic slice management, optimisation and adaptation across multi-operator MEC networks.
- **Area of study 2** – Identify needs and requirements to define the interface between application layer and modem layer of 5G devices to enable URSP rules and traffic descriptor policies to be realised in C-V2X services.
- **Area of study 3** – Identify potential requirements to be considered in the definition of the network slice SLS/SLA in relation to edge deployments for automotive applications. Specifically, study if potential additional requirements for a network slice (on top of those already described in GST/NEST) are needed in order to support edge computing services in the resource sharing scenarios defined in MEC4AUTO (e.g. scenarios 2 and 3).

The following sections describe the state of the art in the relevant standards development organisations (SDO) in relation to the above three areas of study. This analysis aims to create a common understanding in order to identify potential new requirements and solutions.

4.2.2. 5GAA state of the art

NESQO and eNESQO TRs can be considered a first attempt to lay out the foundations for using network slicing in automotive applications, which can be considered a baseline for all of the three identified areas of study.

NESQO TR [10] in Section 4.5 describes an overview of network slicing and the benefits it can bring to automotive applications. It also provides state-of-the-art concepts for a network slice template and GSMA generic slice template (GST) and network slice type (NEST) and network slice roaming in Section 4.6. It also discusses different methods for traffic separation, alternatives to network slicing in Section 4.7 and 4.8. In Section 7 it performs a system analysis for network slicing and edge UPF deployment in C-V2X.

eNESQO TR [11] continued from the previous study, putting its main focus on further detailing aspects and mechanisms for making QoS predictions, and on application and network reactions to such predictions. It concluded that prediction and network slicing could be combined to provide further advantages for QoS management, when the prediction function is deployed in the MNO network.

4.2.3. 3GPP state of the art

In the context of the **first area of study** listed in Section 4.2.1, 3GPP has introduced a number of enhancements during Rel-17. With the introduction of a new network function – the network slice admission control (NSAC) – the 5G system has introduced a level of dynamic control into the core network control plane on the allocated network slice resources. This happens with the help of two new counters which can be limited by the 3GPP system for each network slice (each network slice is from the UE perspective identified by so-called single network slice selection assistance information or S-NSSAI) – this is specified in clause 5.15.11 of TS 23.501 [19]:

- the maximum number of UEs that can be registered with a network slice;
- the maximum number of PDU sessions that can be created in a network slice.

At the moment, resource sharing of a common resource pool by multiple slices is not within scope of SA2 Rel-17 specification. In the case of roaming, depending on the operator's policy, a roaming agreement or an SLA between the visited public land mobile network (VPLMN) and the home PLMN (i.e., HPLMN), NSAC for roaming UEs can be performed by the VPLMN.

At present, the deployment of a consistent SLA across network slices provided by different PLMNs remains quite difficult and this poses challenges to V2X service interoperability when two V2X applications make use of different PLMNs. The same issues apply when a V2X application is roaming in a different PLMN. While different PLMNs may implement the same S-NSSAI with a standard SST (e.g. V2X slice), there is still no standard definition of the service that should be provided by the S-NSSAI with a standard SST. Ongoing work in GSMA aiming to define a standard network slice template for V2X looks promising to resolve some of those interoperability issues. For example, the definition of the QoS supported by the V2X NEST is a first step to make sure that network slices provided by different PLMNs will offer consistent QoS. It is also very important that 5GAA defines what are the QoS parameters and characteristics required for each use case in order to make sure that V2X NEST is aligned with those specifications. In this way the GSMA-defined V2X NEST could be associated with a set of 5GAA use cases supported by the V2X NEST, according to the required QoS.

Rel-18 Study in 3GPP SA2 [27] may allow more functionality in terms of dynamic slice management, optimisation and adaptation across and especially within a single PLMN. Among the aspects that could be improved, the following are noteworthy:

- At the moment the 5GS is missing support for service continuity between network slices, in the event an existing slice cannot serve a PDU session because of specific limitations. How to allow service continuity if an existing slice cannot serve the PDU session in the current cell (due to OAM reasons) or target cell (due to mobility), or a slice was not allowed by a specific UE due to NSAC is one of the problems that could be studied.
- Right now, the 5G system control plane does not know what slice is available for each tracking area (TA), or even finer granularity (e.g. cell). How to allow the network to gather S-NSSAI support per TA or location is another problem which could be studied.
- In roaming cases, a UE activating a service/application requiring a network slice, which is not offered by the serving network but is available in the same area from other network(s) (e.g. the one provided by another mobile operator), is also another issue related to network slicing which could be studied.

In the context of the **second area of study** listed in Section 4.2.1, the issues related to network slice selection and identification are particularly important in the 5GS and relevant to MEC deployments. The 5G system allows for one or more network slices, each of them identified by S-NSSAI to be used by UE applications to connect to edge application servers or EAS, based on input from application developers. An EAS is assumed to be deployed within a local data network that can be identified by a data network name (DNN). A PDU session provides a required IP/Ethernet pipe for an application client running on a UE to interact with an appropriate (edge) application server. PDU sessions may be shared by several application clients using the same UE as long as they need connectivity to the same DNN and use the same network slice. When a UE needs to connect to multiple DNNs or network slices, it has to request more PDU sessions.

In order to establish a PDU session with appropriate S-NSSAI and DNN to route given application traffic to an appropriate local EAS, a UE has to be configured with a set of UE route selection policy (URSP) rules. These rules can be seen – with some simplifications – as instructions for the UE on how to use S-NSSAI to reach which DNN and for which traffic. USRP could be either preconfigured on the UE or dynamically generated by the policy control function (PCF) of the 5GS based on input from application developers who can use a control application function (AF) to pass application-specific information on to the 5G core. The URSP rules include DNN, S-NSSAI and other relevant network parameters that can be used by the UE for matching edge application traffic, e.g. traffic from edge application clients installed on the UE to EASs.

Network slice selection is also related to network node selection, as different nodes may be linked to different network slices. In a homogenous type of network that provides the same service to all the UEs, it is irrelevant which node for a specific network function is selected for the traffic of a UE. However, in a heterogeneous type 5GS, where different network slices can provide differentiated services, only appropriate network nodes need to be chosen to serve a specific UE traffic. In fact, mobile network operators may deploy differentiated network nodes which are optimised for specific traffic classes. Node differentiation may apply within the 3GPP system and in the data network (e.g. EAS).

This differentiation of network node may be achieved with the help of S-NSSAI and DNN. Accordingly, based on appropriate policy information as derived by the PCF of the 5GS for a given application traffic, during a PDU session establishment procedure, access and mobility management function (AMF) of the 5GS will be configured to choose appropriate session management function (SMF) based on S-NSSAI to which a given PDU session is associated. Another set of session management-related policies will help a chosen SMF identify appropriate UPFs that can direct given application traffic to an EAS identified by DN Access Identifier (DNAI). In this way, the PDU session is established in order to route any application traffic from a client running in a UE to an appropriate (edge) AS in local DN. The procedure is defined in TS 23.548 [24].

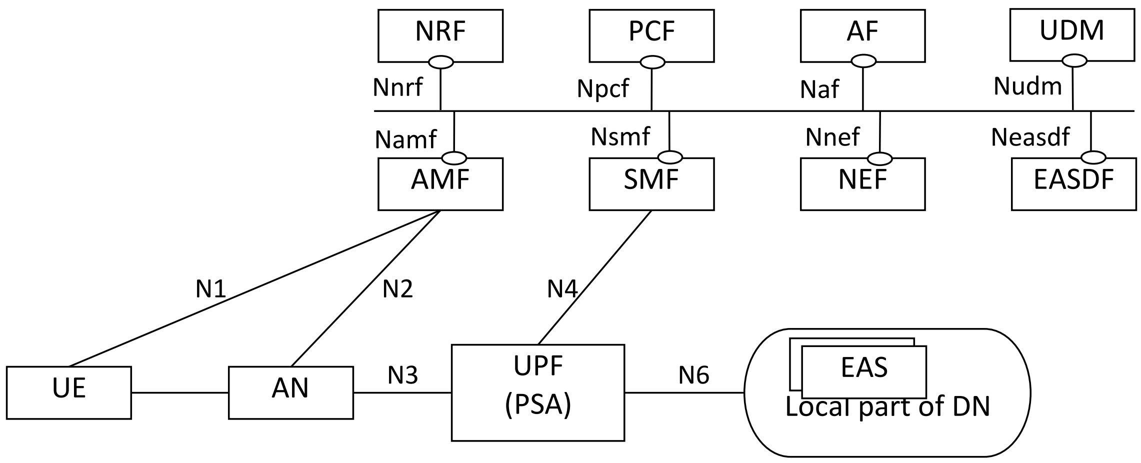

Figure 3: Addition of EASDF in 3GPP core network architecture

Once an appropriate PDU session is established for the edge hosting environment, the next step is for the application running in a UE to identify the appropriate EAS. This process involves DNS query and response. In order to facilitate the network-controlled DNS query-response mechanism, an EAS discovery function (EASDF) was introduced in Rel-17 as depicted in Figure. 3. One of the criteria to be used by the SMF of the 5GS to discover the appropriate EASDF is S-NSSAI, which needs to be input by the SMF as specified in clause 6.3.23 of [19].

As the network nodes, network slices may also be optimised for a specific traffic class. This means that two different network slices cannot necessarily be similar to each other in terms of services they support because parts of the service demand special hardware features, software processing capabilities and different Layer 1 wave-forming, numerologies and techniques, and other technology. As part of the NEST work [20], GSMA clearly indicated in clause 3.4.26 that different slice/service types (SST) may support different QoS performance requirements – especially different 5QIs [19]. This fundamental fact is later acknowledged by 3GPP SA5 in TS 28.541 [21]. In fact Chapter 6 of TS 28.541[21] introduces an attribute called *perfReq* as part of *sliceProfile* for indicating different performance requirements as supported by different SSTs.

This means that different SSTs – such as URLLC, eMBB and MIoT – have their own *perfReq* as indicated respectively by *eMBBPerfReq*, *uRLLCPerfReq* and *mIoTPerfReq*. In conclusion, supported QoS is one way in which network slicing may be able to support differentiated services.

3GPP has defined new slice types for V2X in Stage-2 specification [19]. While Stage-3 specifications support V2X slice type, but how they can support V2X slice type in CT and RAN has not been fully specified. Thus in the gMECAUTO Work Item, solutions are assumed to be developed upon these slice-related features except new V2X slice type. Using of new slice type can be reconsidered in the future if both Stage-2 and Stage-3 specifications have been completed.

Table 1 (from Table 5.15.2.2-1 of [19]): Standardised SST values

Slice/service type	SST value	Characteristics
eMBB	1	Slice suitable for the handling of 5G enhanced mobile broadband.
URLLC	2	Slice suitable for the handling of ultra- reliable low latency communications.
MIoT	3	Slice suitable for the handling of massive IoT.
V2X	4	Slice suitable for the handling of V2X services.
HMTC	5	Slice suitable for the handling of high-performance machine-type communications.

In the context of the **third area of study**, the definition of what kind of support of edge services is provided by a PLMN is currently not part of GSMA NEST [20].

4.2.4. GSMA state of the art

GSMA described detailed requirements and attributes for its generic slice template (GST) [20]. In this context, network slice is defined as a logical network that provides specific network capabilities tailored to the specific requirements agreed between network slice customer (NSC) and network slice provider (NSP). The network slice could span multiple network domains used by a NSP (e.g. access network, core network and transport network). A network slice is comprised of dedicated and/or shared resources, e.g. in terms of functionality, processing power, storage, and bandwidth.

The GST is a set of general attributes characterising a type of network slice/service. It is thus not tied to any specific network deployment in the sense that the attributes could be used for different deployments of network slices. The network slice type (NEST) is a GST with associated attributes, i.e. specific values for throughput, latency, service availability, etc. that correspond to a given set of requirements supporting a NSC use case and deployment. The NEST is an input to the network slice preparation performed by the NSP, as shown in the figure below:

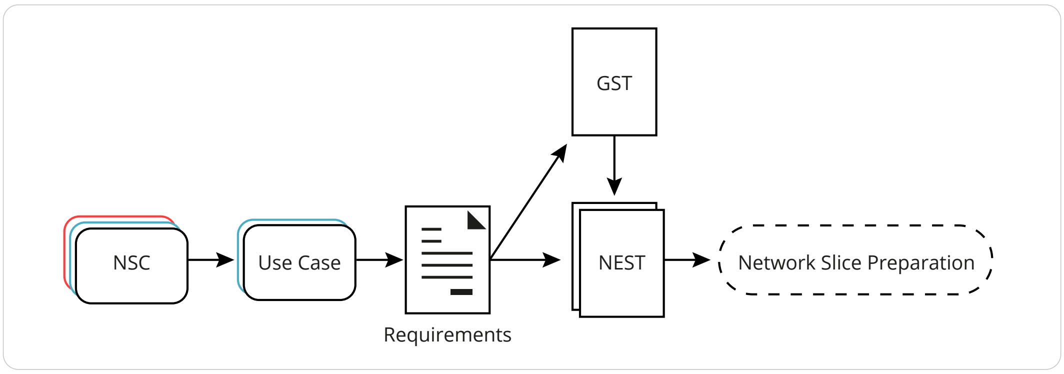

Figure 4: GST and NEST in the context of the network slice lifecycle [20]

In [34], components and high-level requirements were presented for the definition of a *blueprint* for end-to-end (E2E) network slicing. These requirements can be summarised as: requirements for SLA and SLS associated with the network slice; slice provisioning and operation requirements; slice isolation requirements with a specific focus on slice performance; slice management and slice security isolation; and slice capabilities exposed to applications through northbound APIs guaranteeing E2E slice service level requirements and ensuring E2E slice management and orchestration. The overall network slice architecture, including the management plane, as described in [34], is shown in the figure below.

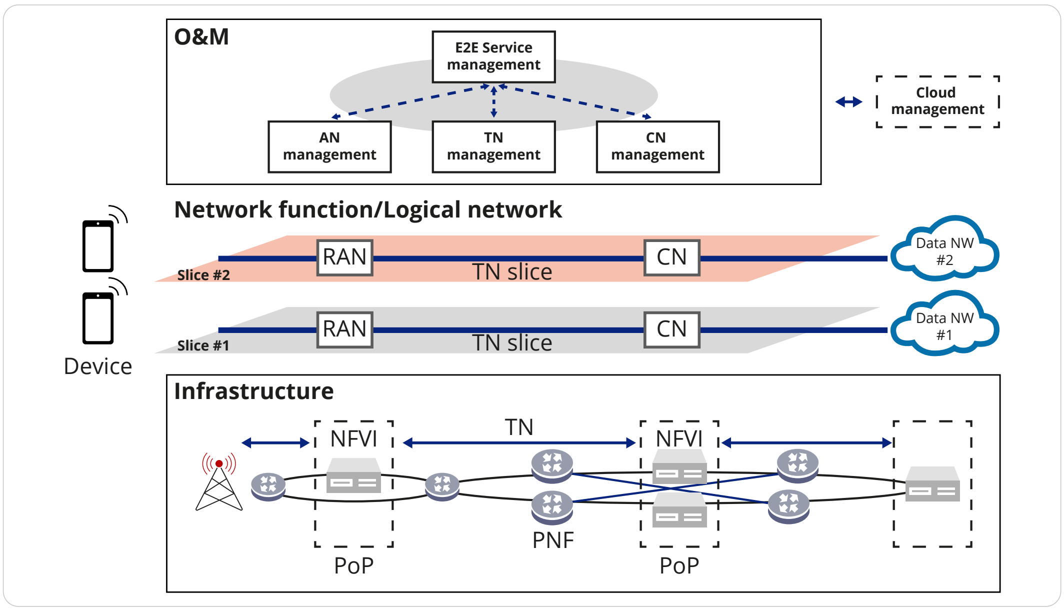

Figure 5: Overall architecture for network slicing for mobile network [34]

The architecture above consists of three parts: infrastructure stratum, network stratum, and management stratum.

The compute, storage and networking fabric of the infrastructure stratum can be used to implement physical network nodes and/or to define a distributed cloud environment, i.e. a NFV infrastructure (NFVI). While the former delivers bespoke physical network functions (PNFs), the latter allows a multi-site virtualised execution environment for VxF hosting, including virtualised network functions (VNFs) and virtualised application functions (VAFs).

The network infrastructure shown in the figure above consist of a RAN with micro- and macro cells attached via dedicated fibres to a multi-tier transport network (TN). This capillarity in TN design, based on distributing computing capacity across points of presence (PoPs) physically deployed at different aggregation levels, paves the way for the rapid decentralisation of the slice deployment.

In this example, we consider a three-tier TN, with operator's managed points of presence classified into regional and central PoPs. Unlike central PoPs, which represent typical core cloud sites, regional PoPs take the role of edge computing nodes. These nodes provide virtualisation capabilities closer to service delivery endpoints in order to reduce the delay budget, making them ideal to host functions and applications for URLLC services. The network and application function stratum is formed of a collection of PNFs and VxFs. It provides the user control and application plane functionality across the different network segments, including RAN, TN and CN. A network slice is defined as an E2E composition of PNFs and VxFs. The resource provisioning and allocation of the individual VxFs depend on the needs of the associated network slices.

Network slicing is built and designed on the basis of requirements described in the GST [20]. To fulfil these requirements a specific design has to be achieved in RAN, TN and CN. As a result, the network slice is the interlocking of the RAN sub-slice, TN sub-slice and CN sub-slice. Even though TN slicing is not well defined by SDOs, the TN is fully part of the network slice as it can be dedicated to a slice or shared between slices. The design of the TN sub-slice is essential to fulfil E2E slice requirements, such as E2E latency, E2E availability, isolation level, and throughput. Such shared network nodes can also be potentially possible in the CN and requires special attention to avoid unauthorised access.

Regarding the slice operation and management stratum, it was proposed by GSMA to use service-based management architecture (SBMA) in order to solve integration and scalability challenges arising when managing E2E network slicing. Finally, the operation and management (OAM) stratum conveys the OSS functionality that allows for the deployment and operation of individual network slices. This stratum aims at handling the operational complexity introduced by network slicing. This architecture style means migrating from functional blocks exposing telecom-style point-to-point interfaces (e.g. network managers/element managers providing 3GPP Itf-N interfaces) to management services exposing APIs based on web-based technology. Different SDOs have already captured the benefits of having SBMA in their specifications. For example, 3GPP SA5 and ETSI ISG ZSM defined their architectural framework based on SBMA. Even ETSI ISG NFV, which originally chose an interface-centric approach for the design of the management and orchestration (MANO) framework, has now decided to migrate towards a SBMA from NFV.

From the device perspective, GSMA [in [34] also described 3GPP work on the specification of UE route selection policy (URSP) that enables the steering of device-aided traffic to a PDU session on a network slice. The device uses URSP to determine which PDU session is chosen for particular traffic based on a URSP rule, which consists of rule precedence, traffic descriptor, and route selection descriptor (plus route selection validation criteria in 3GPP Rel-16). The traffic descriptor has many possible entries, such as DNN, IP address, OS ID plus APP ID, etc., in order to determine whether a URSP rule matches. This traffic on a PDU session can be from a single application or from a group of applications with a limit of eight network slices that can be used by a device.

GSMA proposed an extension to the static network slice deployment described previously. The dynamic network slice, i.e. network slice as service (NSaaS), represents a service delivery model that allows the operators to customise network slices to individual customers, and eventually enable these customers to gain access to some network slice management capabilities. It is up to the operator to decide which specific management capabilities are made available to each customer, typically exposed through customer-facing APIs (e.g. TM Forum APIs).

A baseline NSaaS will require at least the standardisation of dynamic slicing mechanisms, so individual network slices can be deployed and operated in an E2E manner; and the *definition of mechanisms allowing for network slice capability exposure* to individual B2B customers, and the integration of corresponding customer-facing network APIs into the operator's service platform will be considered in the next phase of Operator Platform Group (OPG) work [18]. An advanced NSaaS will require full automation in the entire orchestration pipeline, focusing on the assurance phase and coordinated work among multiple SDOs.

The operator platform's (OP) slicing capability is calculated from its slice parameters and performance metrics. These capabilities range from slice identifiers, such as an S-NSSAI, network slice type, SST, etc., to slice load or performance information.

The OP architecture is defined by the GSMA OPG as an architecture for enabling and coordinating network and application functions between members of operator federations. The detailed requirements corresponding to OPG architecture are described in detail in [18]. One example of an OPG architecture and deployment scenario is shown in the edge sharing illustration below.

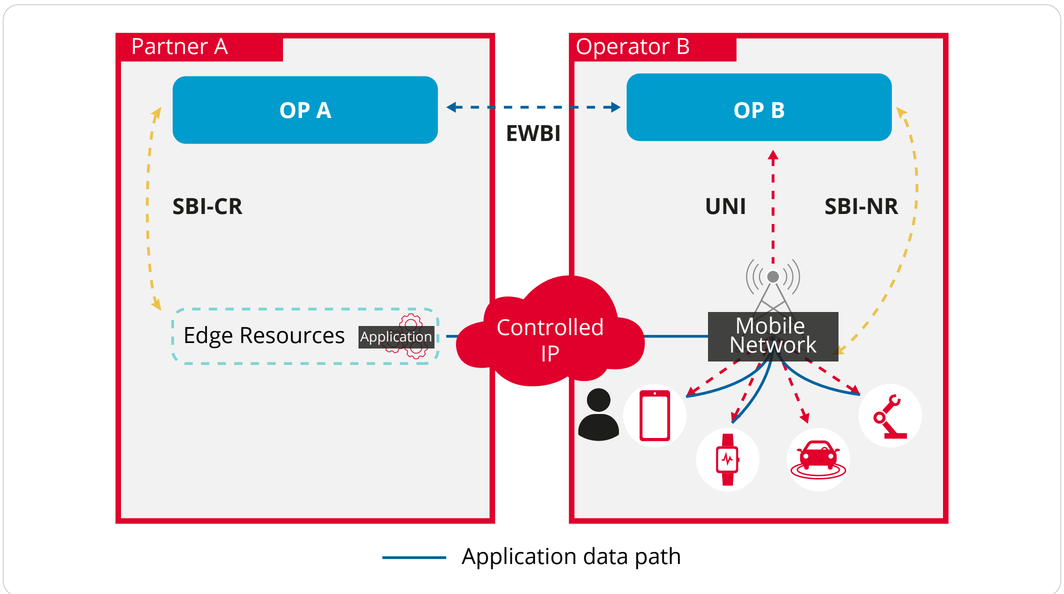

Figure 6: Edge node sharing scenario [18]

In this scenario, the operator B is using edge resources deployed in the network of the partner A. Both operator B and partner A are deploying operator platforms OP A and OP B. OP A and OP B are interconnected by the East/Westbound interface, which enables operator B's OP to retrieve the application instance context and access edge resources, not otherwise available, and provide these resources to their users through the user-to-network interface (UNI).

This approach allows service discovery and delivery to be performed in the same way as when the application was delivered from a cloudlet in operator B's own network. The network resources managed by operator B, who is providing the actual mobile network connection to the user and IP connectivity between partner A's edge node and operator B, is managed to ensure end-to-end QoS delivery for the subscriber.

This E2E QoS is ensured by a controlled IP network connecting the partner A's edge resources with the network of operator B. Responsibility for the management of the edge cloud resources depends on the agreement between the operator/partner. Most likely, operator B has a long-term allocation of resources in partner A's cloudlets accessed as part of overall subscriber edge services.

The southbound interface NR is able, within the current specification of OPG, to configure network slices for the OP's specific application needs and/or to expose slice information from network B to the applications. However, the exposure of slice information to multiple OPs that are members of the federation should be considered in future evolutions of the OPG specification [18].

In summary, the text above presented a status of E2E network slicing in GSMA where the following aspects are presented: a GST used to describe and characterise the network slices, and end-to-end slice architecture with a special focus on SBMA. As to the gMEC4AUTO study (i.e. multi-domain deployment and service continuity of V2X services), slice capability exposure between East-Westbound Ops and capabilities for dynamic NSaaS management and orchestration should be considered in future versions of GSMA-OPG [18] standardisation.

5 High-level architectural considerations on predictive analytics in MEC systems

This chapter studies specific scenarios where predictive edge analytics can be generated and distributed with the help of the MEC platform. In this context, it identifies the requirements and proposes a set of enhancements to be addressed via services provided by the MEC platform.

5.1. Introduction

The chapter is composed of four main parts:

- Section 5.2 introduces the concepts of end-to-end predictive edge analytics, end-to-end user plane link, and the concept of analytics network domain.
- Section 5.3 describes the scenarios selected for the study, specifically the scenarios where a single MNO is involved or when there are multiple MNOs.
- Section 5.4 focuses on the V2X information services (VIS) API provided by the MEC platform and identifies a number of potential enhancements providing more extensive support of predictive edge analytics for the identified scenarios.
- Section 5.5 summarises the conclusions and the issues that are still open.

5.2. Definitions

Some key definitions are introduced in this section, since they will be used throughout the remaining sections.

5.2.1. End-to-end user plane link

The end-to-end user plane link can be defined as the network path between two vehicle application instances that have to exchange V2X messages through the Uu interface in the realisation of a V2X use case.

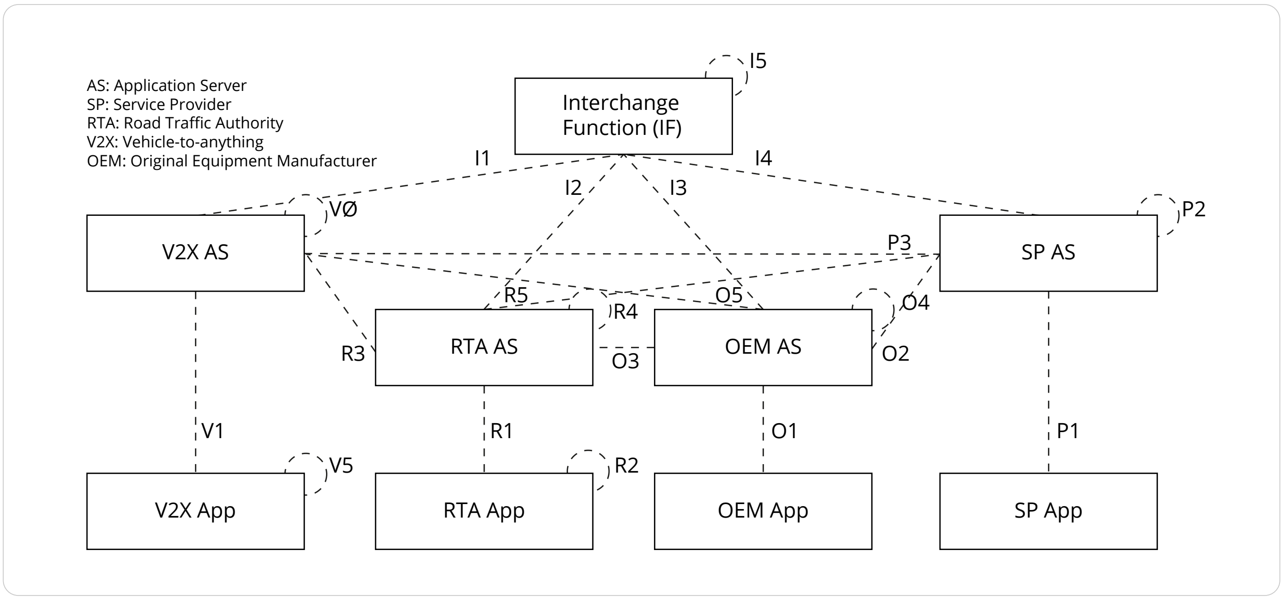

Figure 7: 5GAA V2X application layer reference architecture

Based on the 5GAA V2X application layer reference architecture [14], the vehicle application instance can include one of the following:

- Vehicle-to-anything application (V2X App)
- Original equipment manufacturer application (OEM App)
- Road traffic authority application (RTA App)
- Service provider application (SP App)

Those application instances are assumed to be deployed in the vehicle. Each of these application instances may establish a user plane link to exchange data with a counterpart application instance that is assumed to be deployed in the edge data network, which are respectively:

- Vehicle-to-anything application server (V2X AS)
- Original equipment manufacturer application server (OEM AS)
- Road traffic authority application server (RTA AS)
- Service provider application server (SP AS)

5.2.2. Different domains for predictive analytics

In the context of this Technical Report, an analytics network domain (or simply a network domain) is an administrative grouping of multiple private computer networks or local hosts within the same infrastructure and for which it is possible to identify a provider of predictive analytics (e.g. a prediction function, an NWDAF enabler, etc.) that provides analytics with scope in that domain.

According to that definition, the RAN of a specific MNO A is not a network domain, unless the RAN has a provider of predictive analytics that pertains specifically to the RAN domain. Nor can the core network be considered a network domain because the NWDAF, which is an enabler in the CN, can provide analytics for the whole 3GPP network domain (segment UE-UPF).

In a specific multi-MNO MEC deployment, examples of network domains are described in the table below:

Table 2: Example of analytics network domains and relevant analytics provider

Network domain	Example of analytics provider
MNO A 3GPP network (segment UE-UPF)	MNO A NWDAF instance
MNO B 3GPP network (segment UE-UPF)	MNO B NWDAF instance
MEC platform deployed in DN of MNO A	VIS deployed in MEC platform of MNO A
IP interconnect domain between MNO A and MNO B	OTT PF providing link level prediction service

The analytics are considered end-to-end when they refer to the user plane link between two application instances deployed; for instance in two vehicles (in the V2N2V case), or in the vehicle and in the application server (V2N case), or in the vehicle and in the RTA (V2N2I case). One way to provide analytics for the E2E link is by combining the analytics that relate to different analytics domains, in order to cover all related network segments.

5.3. Analysis of the scenarios identified in the Work Item

This section analyses the scenarios identified in reference [9] and within the Work Item's scope. The objective of the analysis includes the following:

- Explain the benefits of predictive edge analytics in the context of V2X applications deployed in MEC infrastructure.
- Understand which solutions are available to support the generation of predictive edge analytics.
- Identify requirements and potential gaps in the available solutions for future studies.

MEC4AUTO Technical Report "Use cases and initial test specifications review" [3] describes a number of situations that can benefit from the deployment of MEC nodes. Cooperative lane change (CLC) is listed as one of the scenarios actively discussed in literature. 5GAA has specified the use case in reference [4] and 3GPP has studied it in section 5.23 of reference [5], producing very specific QoS requirements in terms of latency and reliability. 5GAA has also specified for this use case predictive QoS on the end-end link to be a requirement for the use case, including support for functionality in multi-OEM, multi-MNO, and cross-border environments. This use case can be characterised by the following:

- Involves vehicles exchanging data (e.g. their intended trajectories to coordinate their lateral (steering) and longitudinal controls (acceleration/deceleration), to ensure a smooth manoeuvre [5].
- Several messages need to be exchanged over a certain period of time among the involved vehicles, and SLRs should be supported during the whole lane merge/change operation.

- Predicted QoS notifications of the E2E user plane link (as defined below) are required from the communication network. Receiving a predicted QoS notification about a potential QoS change (e.g. deterioration) during the time of the manoeuvre may determine different actions by the vehicles involved (e.g. abort manoeuvre or switch to a different communication mode).

It has to be noted that the CLS example has been chosen as representative for the scenario of two vehicles needing to exchange information with a specific set of SLRs over a certain period of time. Other use cases may also fit such a scenario.

As explained later, when this use case is implemented using the Uu interface with MEC infrastructure, the latter may facilitate the production and delivery of QoS predictions. Predictive QoS is not the only analytic that can be produced and delivered thanks to MEC nodes. Indeed, P-QoS is just one example of generic predictive edge analytics; others include observed service experience, network or network slice load, etc.

5.3.1. Single MNO scenarios

In the context of this document, single MNO scenarios are the cases in which vehicles connect via the Uu interface and the same MNO provides the connectivity to all of the vehicles towards the edge infrastructure, which is generally connected via a data network (DN) accessible via the UPF node of the MNO. The MNO network provides connectivity between the UE deployed in the vehicle and the UPF. Several existing solutions can provide predictive edge analytics and they usually differ on the data that is available for the generation of the analytics and on the network segment they can cover. According to the state of the art, the solutions for predictive edge analytics in single-MNO scenarios include the following:

- OTT solutions deployed inside or outside the MNO network, usually within the DN. These solutions are in general based on measurements taken by the application in the UE and in the DNs, and do not have access to internal MNO data, unless an agreement for data exposure is in place between the OTT service provider and the MNO. Limits and capabilities of these solutions have been discussed in reference [10]. While these solutions may have issues in terms of prediction accuracy, they can potentially cover the end-to-end user plane link.
- The 3GPP solution for QoS sustainability analytics described in [6] and [13]. The solution is based on OAM data, and can be complemented by additional data available in the MNO or sourced from the application or a third party. This solution can be more accurate, but there are limitations which have been discussed in reference [10]. While these solutions may have better prediction accuracy, they do not cover the end-to-end link user plane; rather they are limited to the 3GPP scope of the network between the UPF and UE.
- ETSI MEC V2X information services API described in [7]. The accuracy of this solution depends on the data available for the generation of analytics and on information returned to the API service consumer. This type of API-based solution can cover the end-to-end user plane link.

5.3.2. Multi-MNO scenarios

In the context of MEC4AUTO [2], multi-MNO scenarios involve two or more vehicles connected to edge infrastructure using an Uu interface provided by different MNOs. It is assumed that the vehicles need to exchange data in order to realise one or more of the automotive use cases using the facilities provided by MEC nodes.

In such multi-MNO cases, one or more of the application instances deployed on one vehicle need to exchange data with a counterpart application instance connected via a different MNO. MEC4AUTO has studied the challenges of how a vehicle, which has radio access to MNO B, can use a MEC application operated by MNO A without missing the MEC-KPIs (i.e. low latency). The following three multi-MNO scenarios have been studied within the 5GAA MEC4AUTO scope:

1. Both MNO A and MNO B have MEC platform and MEC application X.
2. Both MNO A and MNO B have MEC platform, but MEC application X is available only in MNO A.
3. Only MNO A has MEC platform and MEC application X is available only in MNO A.

Note that inter-MNO connectivity in the **Scenario 3** can be realised through two different options, depending if the inter-MNO link is delivered using the N9 interface connecting the UPF nodes or via IP interconnect (please see Sections 6.3.3 and Section 6.4.3 of [1]for more details).

The E2E user plane link may assume very different configurations depending on the specific scenario that is considered. It may also assume a different configuration depending on the MEC resource sharing scenario defined in [2].

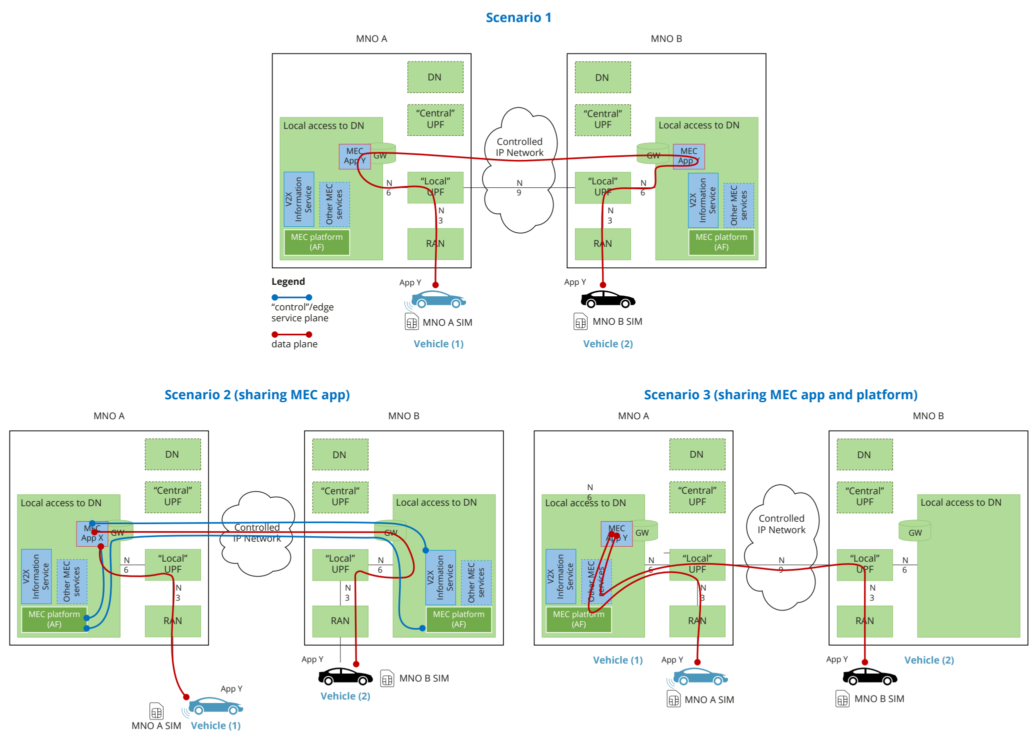

Figure 8: Example MEC scenarios (1, 2, and 3) studied in MEC4AUTO

Figure 8 describes some of the MEC resource-sharing scenarios studied in a previous 5GAA Work Item, MEC4AUTO [2]. The end-to-end user plane link between two vehicle applications is displayed as a red line connecting both vehicles. As it can be observed the E2E user plane link between two applications running in the vehicle and making use of the services in the MEC infrastructure may assume very different configurations depending on the specific scenario.

In the following sections those scenarios will be analysed to obtain analytics that can cover the end-to-end user plan link. The following observations should be noted:

3 MEC4AUTO assumed vehicles 1 and 2 have a global SIM, however here it is assumed that each vehicle has a SIM card provided by the MNO that the vehicle is connecting to. This means that vehicle A has a SIM card provided by MNO A and vehicle B has a SIM card provided by MNO B.

In the scenarios studied in MEC4AUTO TR [2], the end-to-end communication link includes network segments in multiple MNO networks and potentially also outside MNO networks. Each of these network segments may require a specific forecast to be provided by a relevant prediction function. For this, the P-QoS of potentially each of the traversed network segments may be needed to determine the aggregated prediction.

5.3.2.1. Scenario 1: Both MNO A and MNO B have MEC platform and MEC application X

According to MEC4AUTO [2] and the newer methodology defined in the subsequent Work Item gMEC4AUTO [9], this scenario can be categorised by the following dimensions:

Presence of MEC application	vehicle (1): MNO A	vehicle (2): MNO B
Presence of MEC platform	vehicle (1): MNO A	vehicle (2): MNO B
Vehicle subscriptions	vehicle (1): MNO A	vehicle (2): MNO B
Available interconnection between MNOs	N9	
Roaming	No	

NOTE – similar considerations could be made if the available interconnection between MNOs is based on *IP interconnect*.

As SLRs need to be supported for the duration of the use case execution, it is required that both vehicles have subscribed to QoS prediction notifications via the relevant QoS attributes for the E2E user plane link (shown in red in Figure 8).

NOTE – the QoS attributes required for the QoS prediction depend on the specific use case. Reports [10] and [11] provide more information on the QoS attributes requested by specific use cases.

The realisation of such use cases requires the deployment of one or more network functions capable of supporting the collection of relevant input data [11], generation and delivery of the QoS prediction notifications to both vehicles.

As it can be seen in Figure 8, it is possible to identify at least five network domains traversed by the end-to-end user plane link for which QoS prediction could be advisable:

1. MNO A RAN, core network and central DN domains
2. MNO A edge DN domain
3. MNO B RAN, core network and central DN domains
4. MNO B edge DN domain
5. IP interconnect domain between MNO A and MNO B

At the time of writing the present document, **domains 1 and 3** may support the generation of QoS predictions and deliver such information to the respective MEC applications in MNO A and MNO B using a standardised solution provided by 3GPP (more information is available in references [6] and [13]). The prediction function (PF) for such domains is the 5G network data analytics function (NWDAF) which can be accessible to the application either directly or through network exposure (e.g. NEF). As inter-PLMN analytics are presently not supported by NWDAF, the realisation of the targeted scenarios requires that the MEC applications in MNO A and MNO B can first obtain QoS predictions from their respective operator's NWDAF, and share them with each other so both vehicles (associated with MNO A and B) are aware of the prediction related to domains 1 and 3 (i.e. multi-MNO communication of predictions). It is of course required that MNO A and MNO B both deploy NWDAF in their core network and provide predictions with sufficient accuracy in the time and location requested by the application. In general, it is assumed that the application instance that can obtain QoS prediction from the NWDAF is the MEC application deployed in the data network. The application instances running in vehicle (1) and vehicle (2) may obtain P-QoS information from their MEC applications. Another limitation of such a solution is the lack of support for roaming (e.g. when one of the vehicle roams into or visits a network which is different from the home operator MNO A or MNO B). Other limitations have also been identified by 5GAA in [10], and it is expected that some of those may be addressed during the Rel-18 timeframe or later.

In the present document, however, it is assumed that the delivery of QoS prediction to the vehicle happens via the MEC application, independent of the entity that generates the QoS prediction (e.g. core network, external or OTT prediction function, MEC platform, etc.). Direct delivery of QoS predictions to the UE from the core network is also possible and has been studied in [10], however this is not considered in this document.

As already stated, the 3GPP solution only covers RAN and core network domain – the network segment between the UE and the UPF node working as a PSA for the PDU session, or the connectivity service provided by the MNO to the UE deployed in the vehicle. Currently, 3GPP has no plans to extend this prediction to other domains, however nothing prevents vendor-specific implementations to support domains outside of the 3GPP scope. These solutions can still provide QoS predictions within the MNO (RAN and core network) domains.

When it comes to the other domains, related to MNO edge DNs and interconnect between MNOs (**domains 2, 4 and 5**), the following can be observed:

- ETSI GS MEC 030, on "V2X information services (VIS) API" [7], provides a service for journey-specific QoS prediction. However, the current specifications do not include a prediction service that refers to the network segment-related edge data network domains, or to the IP interconnect domain between the two MNOs in multi-MNO scenarios. Yet it could still be possible to support a prediction for this domain if sufficient information is provided to ETSI MEC VIS either to retrieve or to calculate such information. Recently, ETSI MEC has confirmed that such an API can evolve to support multi-MNO scenarios [12]. In fact, this solution typically falls into the use cases of MEC Federation (following GSMA requirements [18]), which is currently being standardised [17][18] in order to support the API exposure in such heterogeneous environments.

- For the IP interconnect domain between MNO A and MNO B, there is no standard solution available. However, this could be covered by an OTT-type solution, as described in [10].

There are two potential general approaches for the MEC application to cover QoS predictions also for domains 2, 4 and 5:

- Each of the relevant domains traversed by the user plane link implements a prediction function (PF)[10], [11]. The application on the vehicle is connected to the MEC application which requests or subscribes to QoS prediction information from the PF running in each of the domains, collects the prediction notifications, and aggregates the final result as an aggregated prediction covering the end-to-end link.
- A prediction function is available to calculate the aggregated end-to-end predictions, either directly supporting prediction generation for the E2E link across the multiple domains or by collecting the domain specific information on behalf of the application from external sources, and later performing the aggregation and delivery to the MEC application.

The second approach requires less effort on the application side and could benefit from a standard API. In fact, as also stated in [8], for matters related to QoS a standard solution is advisable.

In summary, the solutions available for QoS prediction described so far are limited in general to a specific domain and do not provide the support for multi-domain QoS prediction. At present, there is no standard API available to support multi-domain QoS prediction in the edge environment. MEC V2X API is a potential candidate to fulfil such a requirement [12].

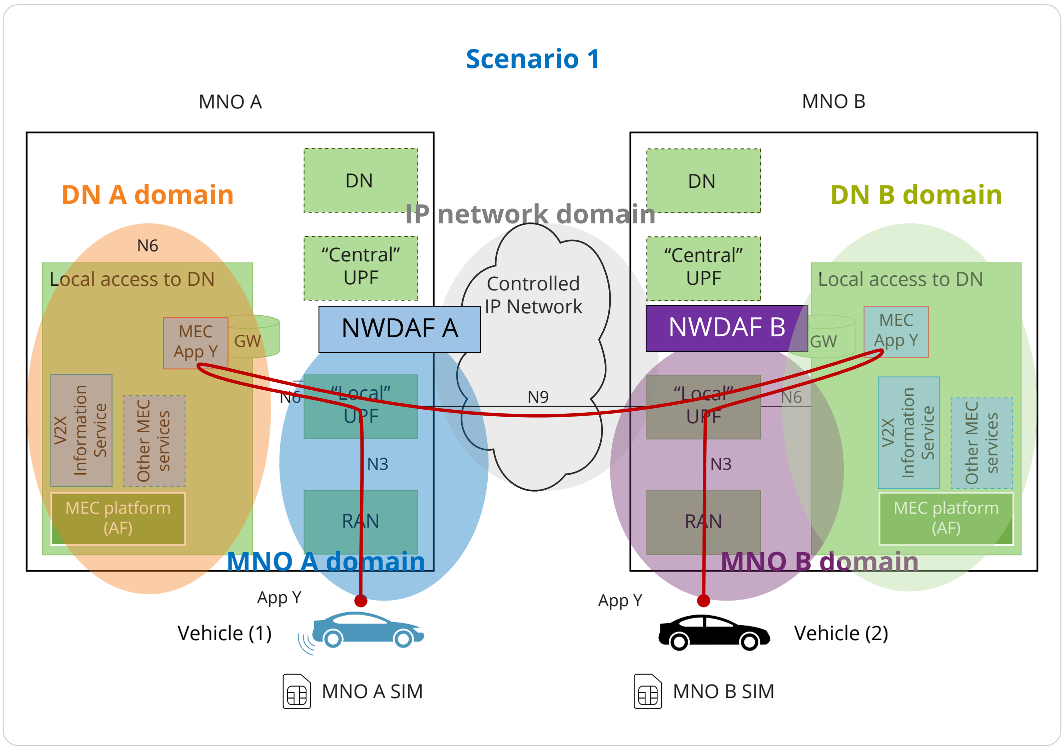

Figure 9: Scenario 1 of 5GAA MEC4AUTO TR and indication of the five different prediction domains for the end-to-end link

5.3.2.2. Scenario 2: Both MNO A and MNO B have MEC platforms, but MEC application X is available only in MNO A

According to [2] and the newer methodology defined in [9], this scenario can be categorised by the following dimensions:

Presence of MEC application	vehicle (1): MNO A	vehicle (2): MNO B
Presence of MEC platform	vehicle (1): MNO A	vehicle (2): MNO A
Vehicle subscriptions	vehicle (1): MNO A	vehicle (2): MNO B
Available interconnection between MNOs	N9	
Roaming	No	

NOTE – similar considerations apply if the available interconnection between MNOs is based on IP interconnect.

As it can be observed in Figure 10, from the point of view of supporting the E2E user plane link QoS prediction this scenario presents the same issues as in Scenario 1 – the network domains are the same:

1. MNO A RAN, core network and central DN domains
2. MNO A Edge DN domain
3. MNO B RAN, core network and central DN domains
4. MNO B Edge DN domain
5. IP interconnect domain between MNO A and MNO B

However, an aspect to be considered for Scenario 2 is that MEC applications serving vehicle (1) and (2) are deployed in the MEC platform of MNO A. Since both instances of the MEC application are deployed in the MNO A MEC host, Scenario 2 has one additional problem to be solved compared to Scenario 1. In fact, MNO A may not be able to offer P-QoS information for RAN, core network and the edge data network domains of MNO B, unless there is some kind of agreement between both operators to exchange such information. In this case, the MEC platform of MNO A may request such information from the MEC platform of MNO B. This solution typically falls into the use cases of MEC Federation (following GSMA requirements [18]), which is currently being standardised [16][17] in order to support the API exposure in such heterogeneous environments.

Another possibility for the MEC platform of MNO A to offer P-QoS information for RAN, core network and other MNO B domains, is if MNO B offers up its own NWDAF interface via network exposure phase. In that event, the MEC platform or another node can request such information from the NWDAF of MNO B via a NEF.

In conclusion, concerning domains 1 and 3 the same considerations of Scenario 1 apply. Concerning domain 4, as stated above, this part of the analytics may not be available for the MEC application deployed in MNO A, unless the MEC platform of MNO B makes such information available at the application instance deployed in MNO A data network for vehicle (2). Concerning domain 5, similar considerations to Scenario 1 also apply.

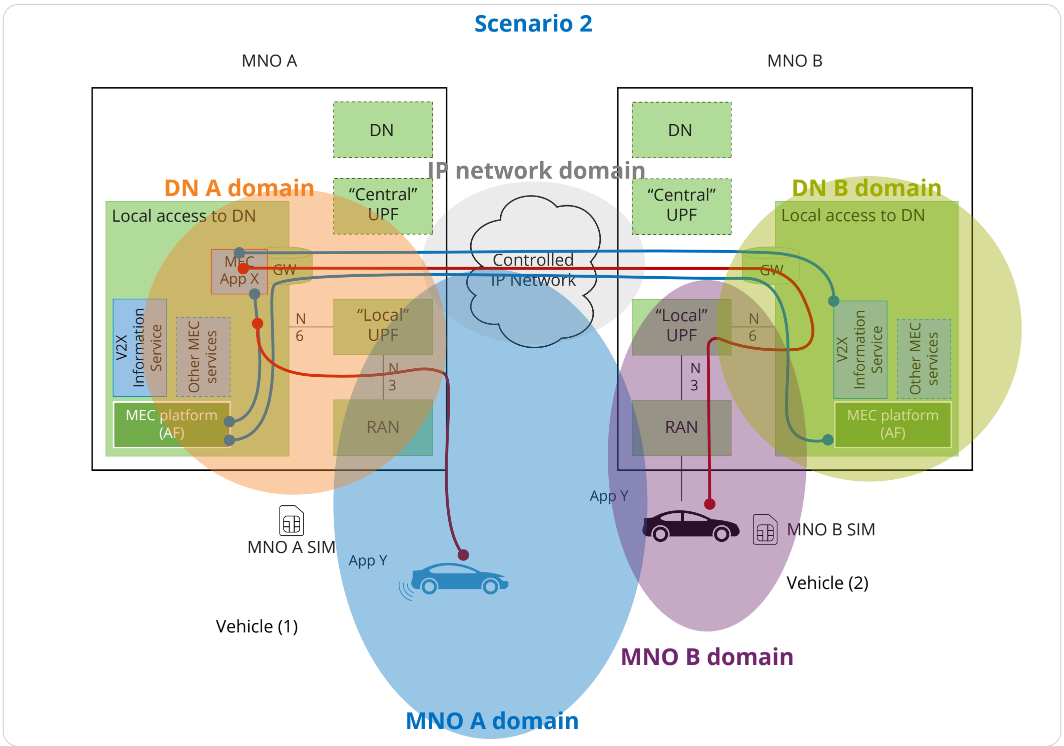

Figure 10: Scenario 2 of 5GAA MEC4AUTO TR and indication of the five different prediction domains for the end-to-end link

5.3.2.3. Scenario 3: Only MNO A has a MEC platform and MEC application X is available only in MNO A

According to [2] and the newer methodology defined in [9], this scenario can be categorised by the following dimensions:

Presence of MEC application	vehicle (1): MNO A	vehicle (2): MNO A
Presence of MEC platform	vehicle (1): MNO A	vehicle (2): MNO A
Vehicle subscriptions	vehicle (1): MNO A	vehicle (2): MNO B
Available interconnection between MNOs	N9	
Roaming	No	

NOTE – similar considerations could be applied here if the available interconnection between MNOs is based on IP interconnect.

As it can be observed in Figure 11, from the point of view of supporting end-to-end user plane link QoS prediction this scenario presents the same issues as Scenario 1, however there are only four domains relevant for achieving the E2E user plane link QoS prediction:

1. MNO A RAN, core network and central DN domain
2. MNO A Edge DN domain
3. MNO B RAN, core network and central DN domain
4. N/A
5. IP interconnect domain between MNO A and MNO B

As in Scenario, 2 the MEC applications serving vehicle (1) and (2) are deployed in the MEC platform of MNO A. Since both instances of the MEC application are deployed on the MNO A MEC host, MNO A may not be able to offer QoS prediction information for RAN and core network domains of MNO B, unless there is some kind of agreement between both operators to exchange such information. In this case, MNO A's MEC platform may request such information from MNO B. Alternatively, if MNO B offers its NWDAF interface via network exposure, the MEC platform or another node can request such information from the NWDAF of MNO B via NEF.

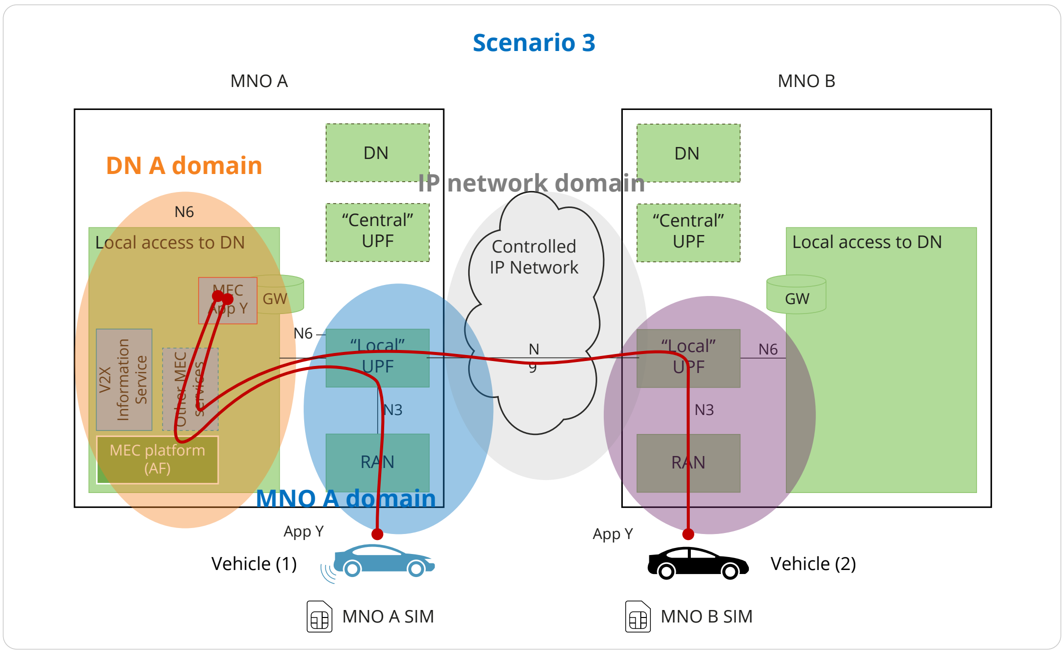

Figure 11: Scenario 3A of 5GAA MEC4AUTO TR and indication of the four different prediction domains for the end-to-end link

In summary, domains 1, 2, 3 apply the same considerations as those observed in Scenario 1 (as a side note, domain 4 is not present in this scenario because the MEC infrastructure is not available at MNO B).

5.3.2.4. Scenario 4: Analytics on user plane link with a third-party

Figure 12 represents a deployment that is similar to Scenario 1 except the two vehicles also exchange information with a third party. This third party could be an external service provider or an RTA. In the latter case, 5GAA has recognised that in specific V2X use cases, such as tele-operated driving (ToD), it can be useful for the application to receive QoS predictions for the link between the RTA and the vehicle, and the link between the RTA and the RTA AS [15].

For the purpose of generating analytics in Scenario 4, a sixth (6) domain is added to the list for **third parties.** In the event of MEC deployments, the MEC platform and application can also support the generation of analytics for this domain, either directly or in cooperation with a PF located in the third-party domain.

In this scenario, the MEC application or platform may provide this information to the vehicle that is served or potentially also share it with other vehicles through the remote MEC platform.

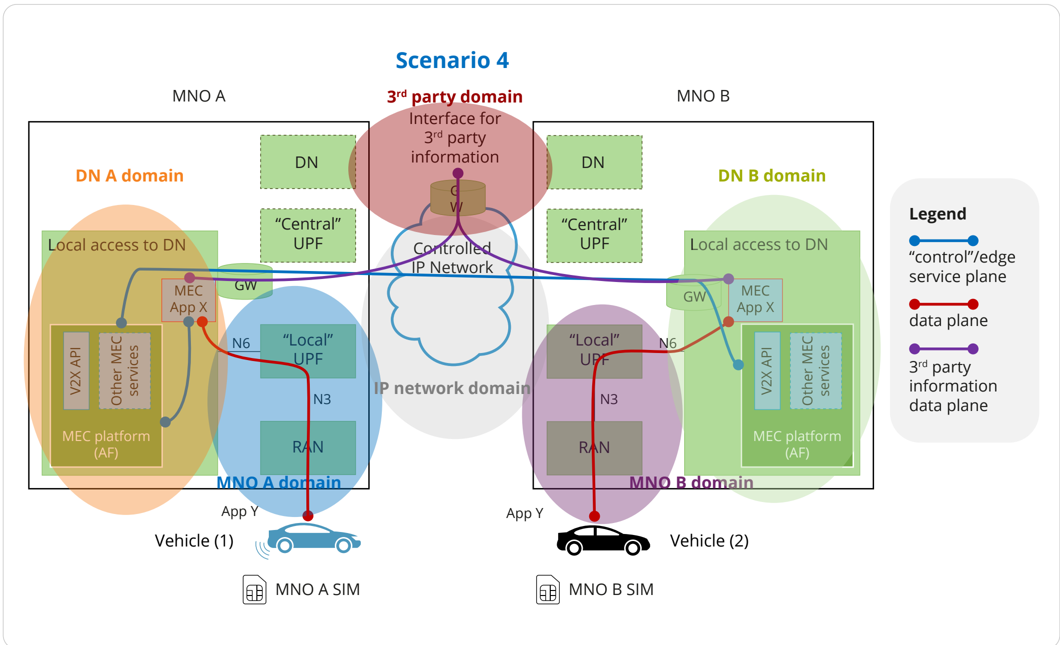

Figure 12: Scenario 4: Analytics on user plane link with a third-party and indications of the 5 different prediction domains for the end-to-end link. It is equivalent to Scenario 1 of 5GAA MEC4AUTO TR + interface with third parties.

In summary, domain 6 QoS prediction can be covered via a PF deployed in the third-party domain or by a PF deployed in the MEC platform.

5.4. Potential enhancements to MEC VIS API

This chapter lists a number of potential enhancements that could be added to MEC VIS API for journey-specific QoS predictions, in order to provide better support for predictive edge analytics in multi-domain environments.

5.4.1. Cooperation with external prediction functions

This section introduces to the VIS the possibility to cooperate with external prediction functions in order to enable E2E QoS predictions involving multi-domain deployments.

According to [7], the VIS service for journey-specific QoS predictions *"may assist in implementing a framework for cooperative acquisition, partitioning and distribution of information for efficient, journey-specific QoS prediction. That is, the VIS service may be utilised to identify space/time correlations between radio quality data collected by different vehicles in a V2X system and a specific vehicle's planned journey for better prediction of the quality of the communication network along the designated route. As a consequence, the VIS may expose relevant (i.e. journey-specific) information about the QoS prediction to authorised UEs."*

In addition to the data collected by the different vehicles, VIS may also cooperate with external PFs located in different network domains, such as the ones identified in Section 5.3.2, in order to provide more accurate QoS predictions to the service consumers. For example, in 3GPP system domains the VIS may cooperate with the analytics service provided by the NWDAF described in cl. 6.9 of [6]. NWDAF can provide predictions for the 3GPP network domain (e.g. RAN and core network segment between the UE and the UPF). Other network domains, such as IP interconnect or DNs in the service provider domain (including the MEC host) may also deploy a domain-specific PF that can provide analytics information to a calling service consumer via the VIS. Some operators may also deploy a NWDAF service that can also provide predictions for DN domains, such as the DN performance analytics described in cl. 6.14 of [6] for user plane performance (i.e. average/maximum traffic rate, average/maximum packet delay, average packet loss rate) in the form of statistics or predictions to a service consumer. However other DNs, especially those not in the operator domain, may also provide a prediction service to the DN which is different from the standardised 3GPP one. VIS may aggregate such information to provide an end-to-end QoS prediction, especially in multi-domain scenarios. VIS may also internally generate a prediction for one or more domains where the PF is not available based on collected information. This information can be either obtained from the NEF of the domain or through other MEC services such as a radio network information service, or RNIS [32].

In a simple scenario, VIS may also estimate QoS predictions for network domains which do not have a PF available based on past measured information or through RNIS. This may be a workable solution, especially for short-term predictions. In general, a QoS prediction for a link composed of several network domains can be a combination of actual predictions and estimations based on past measurements.

As an example of the possible cooperation of VIS with external prediction functions it is possible to consider the architecture of Scenario 1 described in Section 5.3.2.1. An alternative representation of this scenario for the use case "cooperative lane merge" is represented in the figure below.

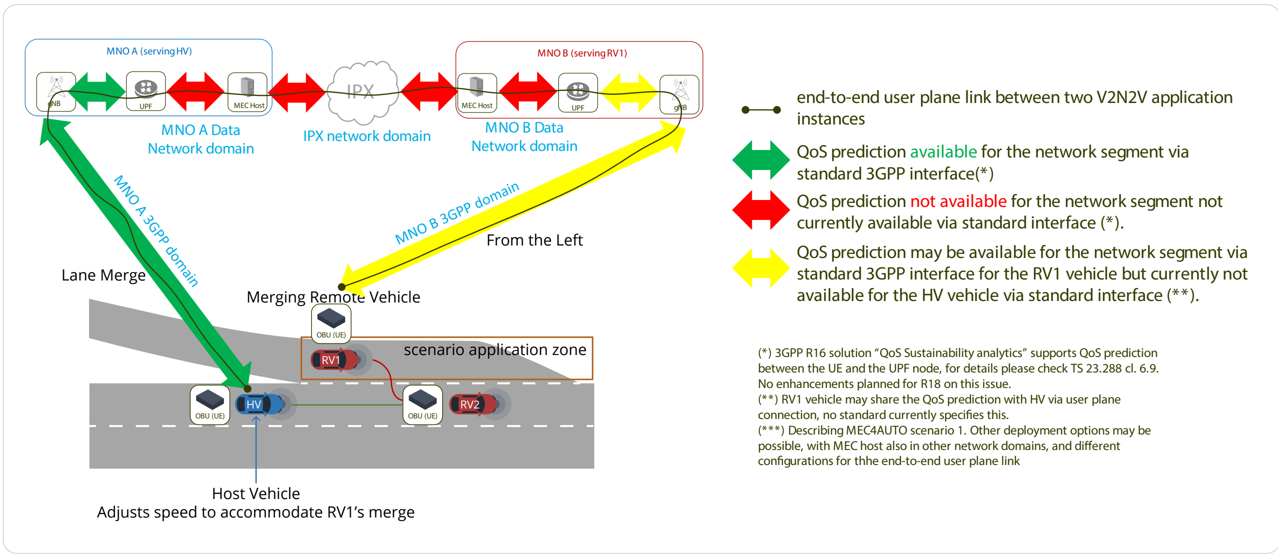

Figure 13: Using QoS prediction for the end-to-end user plane link between two application instances (case V2N2V) in cooperative lane merge, assuming the deployment is according to MEC4AUTO Scenario 1

In this example the host vehicle needs to make decisions based on the QoS prediction of the end-to-end link with RV1 and RV2. For simplicity, we will consider that the use case requires QoS prediction only of the latency KPI. Therefore, it is assumed that – in order to provide the response to the service consumer – the VIS needs to compute the end-to-end latency of the user plane link between the two vehicles (e.g. HV and RV1 or HV and RV2, hereon called vehicle A and vehicle B) and for this reason it needs to retrieve the predictions for the individual latency contributions of the different network segments from the prediction functions available for the relevant domains. Once the VIS has retrieved the latency prediction of each network segment, it can compute the E2E latency by simply summing each component, according to the number of times the network segment has to be traversed in the end-to-end user plane path.

In the deployment considered here the external prediction functions are two NWDAF entities of both operators and a local link level prediction service available for the IP interconnect. Again for the sake of simplicity, when calculating the E2E predicted latency the MNO A DN and MNO B DN are not considered. When such latency predictions are deemed relevant, they could be retrieved according to cl. 6.14 of [6]. As an example, the NWDAF addresses could be either pre-configured in VIS or dynamically discovered. (Exactly how VIS discovers the external prediction functions, e.g. NWDAFs, is out of scope in this TR.

The above example considers the case in which VIS receives a consumer's service request for a journey specific prediction for the end-to-end user plane link between two vehicle UEs, vehicle A and vehicle B. According to Scenario 1, the connectivity to those vehicle UEs is provided by two different MNOs, respectively MNO A and MNO B, and both MNOs have deployed a MEC platform and application. It is assumed that both MNOs support QoS prediction and for such reason have deployed the NWDAF enabler, which is "exposed" for external applications via the NEF. Both MNO A and MNO B can provide analytics of type *QoS sustainability* and *DN Performance* [6] which relate respectively to the 3GPP domain (network segment between the UE and the UPF) and the data network (segment between the UPF and MEC host) domain of each MNO network. It is also assumed that the controlled IP network also provides link-level QoS prediction information with a proprietary service based on HTTP/REST. Such a scenario can be realised according to the implementation shown in the Figure 14.

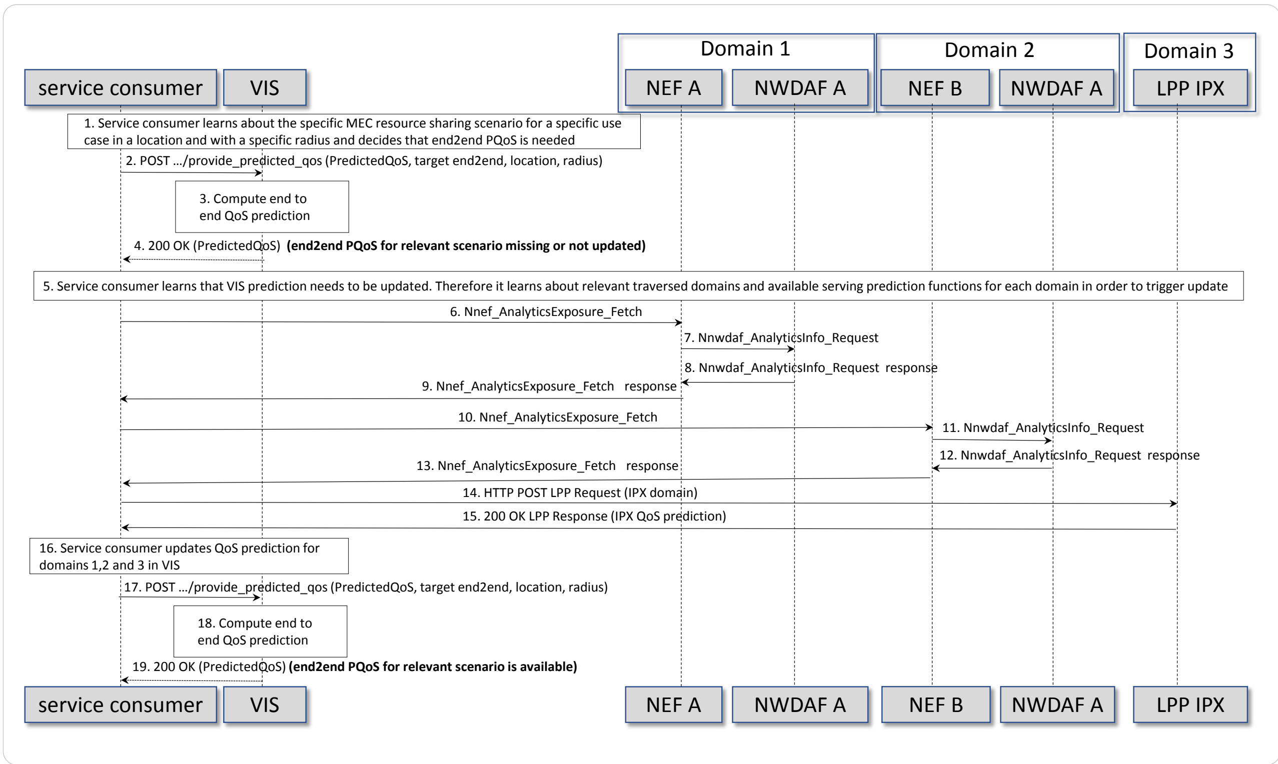

Figure 14: VIS cooperating with other prediction functions

In step 1, a service consumer learns about the specific MEC resource-sharing scenario (e.g. MEC4AUTO Scenario 1) for a specific use case (e.g. cooperative lane merge), and that predictive QoS for end-to-end user plane link is needed for the application entities participating in this scenario. The area in which the use case will take place can be identified by a location (e.g. in GNSS coordinates) and a radius.

In step 2, the service consumer sends a request to VIS for journey-specific QoS predictions for the user plane link between vehicle A and vehicle B, or for the area where the specific use case is supposed to take place. As specified in cl 6.2.5 of [7], the VIS may expect from the service consumer the following information in the *PredictedQoS* attributes: information related to the potential routes of the vehicular UEs, containing the location of the vehicular UEs and the estimated time at the location. An alternative way could be to provide VIS the information about the region for which the end-to-end P-QoS is required, for example with a location centred on the vehicle A (e.g. in GNSS coordinates) and a radius to include all vehicles in a geographical area where the use case is supposed to take place.

In step 3, VIS tries to serve the request received in step 2, however it does not have sufficient information to serve the request, therefore in step 4 VIS provides a response to the service consumer with an error code indicating that insufficient information is available to serve the request received in step 2.

In step 5, the service consumer has learned that VIS prediction needs to be updated. Therefore, it knows that it has to collect all the information necessary to perform an update of the VIS. In this example, the analytics need to be collected from three separate domains (MNO A or domain 1, IPX domain or domain 2, MNO B or domain 3). In this step, the service consumer learns about the specific MEC resource-sharing, either by accessing the configuration setup or by performing dynamic queries to other nodes (e.g. DNS or other server). The consumer also learns the addresses of the prediction functions available for the domains traversed by the user plane link in the specific scenario. It is therefore assumed that there is an NWDAF available in each of the MNO networks capable of providing P-QoS for each MNO network. The NWDAF is accessible via NEF to servers located outside the MNO network. In this scenario, there is also a (proprietary) network prediction function available in the IP interconnect network between MNO A and MNO B. If one or more of the domains do not provide prediction functions, the VIS has to estimate the relevant QoS prediction using other means (e.g. by collecting sufficient data or by replacing the prediction with a measurement of the relevant quantity which can be assumed to be valid for a short time in the future).

In step 6, the service consumer requests analytics from MNO A by issuing a *Nnef_AnalyticsExposure_Fetch* request addressed to the NEF of MNO A. This request contains the relevant information required by the NWDAF of MNO A to provide the required analytics, as described in [6]. The MNO A in this case comprises two separate domains, which are addressed each by a different analytics request:

- The **3GPP domain** is comprised of the network segment between the UE A and the UPF of the MNO A. For this segment the service consumer may request the QoS sustainability analytics. In the request, it has to indicate an analytics target period in the future that can be provided according to the estimated time (at the location) of the request received in point 1. The service consumer should also include other analytics filter information, such as the relevant 5G QoS 5QI required by the application in question, the S-NSSAI, etc. For more information on all of the information to be included in the service request please check clause 6.9.1 of [6].
- The **data network domain**, where the MEC host is located, is comprised of the network segment between the UPF of MNO A and the terminating server or MEC host in the DN of MNO A. For this network segment, the service consumer may request the DN performance analytics. For more information on the service request please check clause 6.14.1 of [6].

In step 7, NEF in MNO A, after performing authentication and authorisation, invokes the *Nnwdaf_AnalyticsInfo_Request* service operation of NWDAF, in order to call up the relevant analytics.

In step 8, NWDAF in MNO A responds by providing the required analytics output/prediction to NEF. This may include information on the packet delay budget prediction for the network segment between the UE A and the UPF of the MNO A, and the latency of the network segment between the UPF of MNO A as well as the terminating server in the DN of MNO A.

In step 9, the NEF invokes the *Nnef_AnalyticsExposure_Fetch* response service operation to provide the analytics response to service consumer. NEF may issue restrictions in the service operation (e.g. removing/hiding part of the information received in step 5 from NWDAF). At this point VIS has received the analytics (prediction) for the domains related to MNO A.

The content of the QoS prediction for the 3GPP network domain is described in the table [6] below for convenience.

Table 3 (based on Table 6.9.3-1 of [6]): Content of the QoS prediction in the 3GPP network

Information	Description
List of QoS sustainability analytics (1..max)	
>Applicable area	A list of TAIs or cell IDs within the location information that the analytics applies to.
>Applicable time period	The time period within the Analytics target period that the analytics applies to.
>Crossed reporting threshold(s)	The reporting threshold(s) that are met or exceeded or crossed by the statistics value or the expected value of the QoS KPI.
>Confidence	Confidence of the prediction.

Since in this example it is assumed that the service consumer is interested in the latency prediction, the relevant information that can be provided by the MNO is highlighted in grey in the table. For the QoS sustainability analytics, the information is reported in terms of which threshold(s) are met or exceeded by the expected value of the QoS KPI. In order to get a latency prediction for a GBR QoS flow, the relevant QoS KPI that the analytics should refer is the retainability KPI for the GBR flow. In this case the information on whether the threshold is met/not met on the retainability KPI means that the corresponding packet delay budget associated to the 5QI of the guaranteed bit rate (GBR) QoS flow is fulfilled/not fulfilled. For more information please check reference [6] and [33].

The content of the QoS prediction for the DN domain is described by Table 4 (6.14.3-2 of [6]) which is also included below for convenience.

Table 4 (based on Table 6.14.3-2 of [6]): Content of the QoS prediction for the DN domain in the 3GPP network

Information	Description
Application ID	Identifies the application for which analytics information is provided.
S-NSSAI	Identifies the network slice for which analytics information is provided. See Note 1.
DNN	Identifies the data network name (e.g. internet) for which analytics information is provided. See Note 1.
DN performance (0-x)	List of DN performance for the application.
> Application server instance address	Identifies the application server instance (IP address/FQDN of the application server).
> Serving anchor UPF	The involved anchor UPF. See Note 2.
> DNAI	Identifier of a user plane access to one or more DN(s) where applications are deployed as defined in TS 23.501 [19].
> Performance	Performance indicators
>> Average traffic rate	Average traffic rate predicted for UEs communicating with the application. See Note 3.
>> Maximum traffic rate	Maximum traffic rate predicted for UEs communicating with the application. See Note 3.
>> Average Packet Delay	Average packet delay predicted for UEs communicating with the application. See Note 3.
>> Maximum Packet Delay	Maximum packet delay for predicted for UEs communicating with the application. See Note 3.
>> Average Packet Loss Rate	Average packet loss predicted for UEs communicating with the application. See Note 3.
> Spatial Validity Condition	Area where the DN performance analytics applies.
> Temporal Validity Condition	Validity period for the DN performance analytics.
> Confidence	Confidence of this prediction.
Note 1: The item "DNN" and "S-NSSAI" shall not be included if the consumer is an untrusted AF. Note 2: The item "serving anchor UPF" shall not be included if the consumer is an AF. Note 3: Analytics subset that can be used in "list of analytics subsets that are requested", "preferred level of accuracy per analytics subset" and "reporting thresholds".	

Since in the example it is assumed that the service consumer is interested in the latency prediction, the relevant information that can be provided by the MNO is highlighted in grey in the table. For more information please check reference [6] and [33].

In step 10, the service consumer repeats similar operations as in step 6 by issuing a *Nnef_AnalyticsExposure_Fetch* request addressed to the NEF of MNO B.

In step 11, NEF in MNO B, after performing authentication and authorisation, invokes the *Nnwdaf_AnalyticsInfo_Request* service operation of NWDAF, in order to request the relevant analytics.

In step 12, the NWDAF in MNO B responds by providing the required analytics output/ prediction to NEF of MNO B.

In step 13, the NEF invokes the *Nnef_AnalyticsExposure_Fetch* response service operation to provide the analytics response to the service consumer, similarly as in step 9

In step 14, the service consumer may also engage a PF providing link level latency prediction for the network segment between the MNO A and MNO B (IP interconnect). If such a function is not available, service consumer may estimate the predicted latency for this network segment, for example by replacing the prediction with a measurement of the current delay for that network segment.

In step 15, the service consumer receives from the link-level latency PF the requested latency prediction.

In step 16 the service consumer updates the QoS prediction in VIS according to the information received in steps 9, 14 and 15, respectively from domains 1, 2 and 3.

After step 16, VIS is supposed to have updated P-QoS prediction for all relevant domains of the use case considered in the example, therefore can effectively serve incoming P-QoS requests.

In step 17, the service consumer (it can be a different consumer from the one that has performed the update) requests VIS QoS prediction for the end-to-end user plane link for a specific use case, for example by providing a location and radius.

In step 18, VIS computes the predicted end-to-end latency for the user plane link between MNO A and MNO B. For example the predicted latency can be obtained by performing the sum of the generated latencies in each of the link segments in the networks, considering also the number of times those segments are traversed in the end-to-end link.

In step 19, VIS provides the journey-specific prediction – computed in the previous step – to the service consumer. The information is provided inside a *PredictedQoS* data structure as defined in [7] or with the extensions proposed in Section 5.4.3.

Observation 5.4.1-1 – in addition to the cooperative acquisition of analytics, in multi-domain (e.g. multi-MNO, multi-OEM, etc.) MEC scenarios [2], VIS may aggregate analytics sourced from external PFs in order to provide (predictions/estimations) of the QoS that cover the end-to-end user plane link between two V2X application instances.

5.4.2. Support for subscription/notification

This section proposes introducing support for subscription/notification for continuous issuance of en route predictive QoS notifications in VIS.

According to [7], the service consumer (e.g. a V2X application or a MEC application) sends a HTTP POST request to VIS to receive the P-QoS corresponding to potential routes of a vehicular UE. The response contains the required information. Figure 195 describes such interaction.

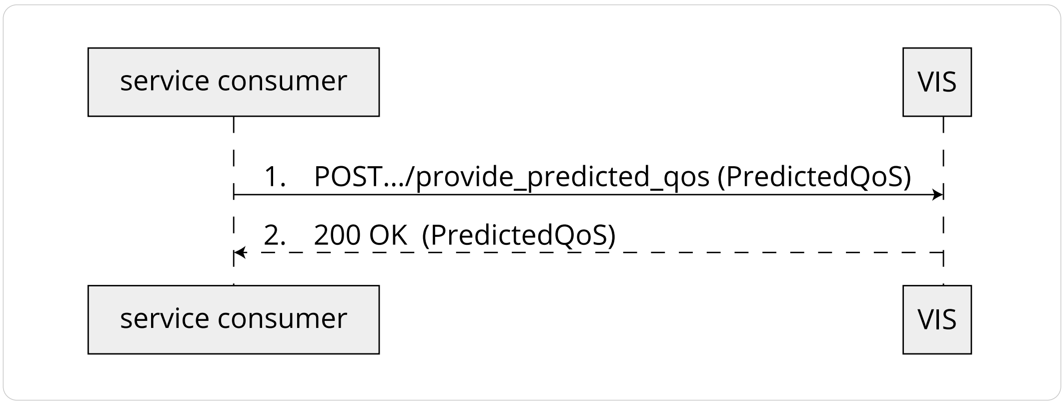

Figure 15: Flow of a VIS service consumer (e.g. V2X application) requesting the predicted QoS of a UE with potential routes (Source [7])

One limitation of the currently specified MEC VIS is that it is based on a request/response type of interaction for QoS predictions. Request/response interactions can be used when the service consumer needs to make a (one-time) instant decision based on the current value of the prediction. However, subscription/notification models have, in general, more flexibility. For example, the application may subscribe to a specific route when the vehicle starts to drive towards a destination, and VIS may continuously provide in-advance predictions on relevant events along the route, as such information becomes available. This model of interaction is more suitable compared to continuously polling VIS for predictions. Another advantage of using subscriptions is that VIS may relay the information contained in the subscription to the providers of the prediction (e.g. PF when available). For example, when the provider is NWDAF [6], VIS may relay the information contained in the subscription received by the service consumer in the subscription issued to the NWDAF. For more information on which information could be received in the subscription for P-QoS, see Section 5.4.3. Another advantage of using subscriptions to specific route information about the vehicular UE is that VIS may relay the route prediction to providers (again, PF when available) of multiple network domains along the route. The prediction providers may use the analytics from VIS to update their own predictions. Current signalling for updating the subscription for V2X information event notification may be used to handle vehicular UE route updates. In a multi-domain environment, VIS and the application may implement back-to-back subscriptions with the PFs in the specific network domains, such as the one described in cl. 6.9 of [6]. For example, in a scenario with multiple MNOs involved, VIS may relay the *QoSSustainabilityAnalytics*, described in cl. 5.1.6.2.19 of [33], and use such information to compute the QoS prediction in the multi-MNO scenario. In turn, the PFs may use the subscription information to fine tune the algorithm used to calculate the prediction. Also, the notifications generated by VIS can be used by the service consumer to implement a call-back routine to trigger specific actions when a relevant prediction is available.

Observation – in addition to the request/response model, VIS may also support subscription/notification interaction with the service consumer. In multi-domain (e.g. multi-MNO, multi-OEM, etc.) MEC scenarios [2], when aggregating analytics from external PFs, VIS and the application may implement back-to-back subscriptions with external source of analytics, such as those described in cl. 6.9 of [6].

5.4.3. Changes to the data model

This section proposes changes to the data model aiming to abstract predicted QoS information to facilitate the evaluation of V2X service support levels. The data model used by the VIS journey-specific QoS prediction service is based on the type PredictedQoS, which is described in the table below (source [7] cl. 6.2.6).

Table 5: PredictedQoS data type (source: [7])

Name	Data type	Cardinality	Remarks
timeGranularity	TimeStamp	0..1	Time granularity of visiting a location.
locationGranularity	String	1	Granularity of visited location. Measured in metres.
Routes	Structure (inlined)	1..N	Information relating to the potential routes of a vehicular UE.
>routeInfo	Structure (inlined)	2..N	Information relating to a specific route. The first structure shall relate to the route origin and the last to the route destination. Intermediate waypoint locations may also be provided.
>>location	LocationInfo	1	Vehicular UE location.
>>time	TimeStamp	0..1	Estimated time at the location.
>>rsrp	Uint8	0..1	Reference Signal Received Power as defined in ETSI TS 136 214 Shall only be included in the response.
>>rsrq	Uint8	0..1	Reference signal received quality as defined in ETSI TS 136 214 Shall only be included in the response.
Note: The data type of locationGranularity is a string which indicates the granularity of a visited location by means of latitudinal and longitudinal margins.			

Since the service is intended to provide journey-specific QoS predictions, it is expected that the service consumer describes the journey for which it is requesting such predictions by populating the routes information (*routeinfo*) element in the request or subscription. That element may incorporate multiple potential routes of the vehicular UE *routeInfo*. Focusing on a given potential route, information is completed with one or more route points (origin, destination and intermediate waypoints), and the expected time at each route point. Such information can be easily provided by an application running in the vehicle acting as an interface with the navigation system.

As described in [7] cl. 6.2.6, the VIS replies with the predicted QoS for the queried routes, based on the available information, by populating the "reference signal received power" *(rsrp)* and "reference signal received quality" *(rsrq)* information elements in the response. Some considerations on the format of the QoS prediction returned in the response can be made according to previous 5GAA studies, such as [10] and [11]. One of the key aspects of the QoS prediction in the context of an automotive application is that it needs to be actionable: this means that the information is used in the application to trigger a specific application reaction. The application reaction is use-case specific and depends on the predicted information (value, value range or other statistical information) of specific QoS KPIs that depend on the specific use case. For every 5GAA use case that can benefit from QoS prediction, 5GAA has defined a list of QoS KPIs that need to be predicted.

5GAA has listed in Section 6.4 of [10] a number of measurements that can be returned in the QoS prediction in terms of KPIs and associated features. Later reference [11] in Section 4 has analysed a number of use cases in terms of potential reaction and related predicted QoS KPIs. The 5GAA Work Item PRESA has provided a more extensive analysis of QoS KPIs needed by V2X applications and related usage in terms of triggering service specific actions or adaptations [1]. For this reason, *rsrp* and *rsrq* are measurements that could be considered too generic in order for the application to determine a specific application reaction.

Therefore, it is suggested to replace (or complement) the prediction on *rsrp* and *rsrq* with more high-level QoS KPI-based predictions that can be representative of the application requirements, such as those listed in Section 5.6.3 of [11], or more specifically the following:

- **Bitrate** or **throughput**: This can be mapped – in the case of interaction with a 3GPP domain where the prediction function is NWDAF and according to cl. 6.9 of [6] – to a specific value of the GFBR [19] or of the RAN UE throughput.
- **Latency**: This can be mapped – in the case of interaction with a 3GPP domain where the PF is NWDAF and according to cl. 6.9 of [6] – to a specific value of the Packet Delay Budget (PDB) [19].
- **Error rate**. This can be mapped – in the case of interaction with a 3GPP domain where the PF is NWDAF and according to cl. 6.9 of [6] – to a specific value of the Packet Error Rate (PER) [19].
- **Coverage**: VIS may be used to provide information on the availability of radio coverage at a specific location. Despite the fact that 3GPP has not included this analytic as part of [6], this can be provided, for example, via interfacing with a coverage map server, such as the Web Coverage Service (WCS), a standard issued by the Open Geospatial Consortium [28].
- **Capability**: VIS may be used to provide information on the availability of a specific functionality or service at a predetermined time and location. One capability may be the support of a specific service which can be defined as a combination of 5QI [19] and a range or values of QoS KPIs or network features, such as multi-connectivity or QoS monitoring. The semantics and use of this capability can be handled during implementation.

Such a new model of reporting QoS could be further enhanced to provide data stream (re: 5QI for 3GPP data streams) granularity, instead of radio link granularity, as in the current reference [7]. This enhancement has the advantage of supporting QoS prediction also for treating differentiated traffic, as supported by the 5G QoS model. Such an enhancement suits the multi-MNO scenario if the provider of the prediction in each MNO network is the NWDAF service “QoS sustainability analytics” described by cl. 6.9 of reference [6].

It should be noted, however, that services like “QoS sustainability analytics” can only return the predicted value of the relevant QoS KPIs that refers to specific 3GPP measurement points (as in [11]). What is expected instead from VIS is to provide a prediction of the QoS KPIs related to the end-to-end user plane link between the two V2X application level instances. This means that the measurements are related to the application layer (or layer-7 according to OSI standard) between the two V2X applications in Scenarios 1, 2 and 3 and between the vehicle and the application entity in the third-party domain offered in Scenario 4. Such information has much higher value for the application, since it encompasses all data transport and may more effectively be used to trigger application-specific reactions.

Lastly, by comparing the currently supported features of VIS in [7] with requirements listed in [11], other potential enhancements to be considered for VIS API arise, including the following:

- Support of **notice period/time horizon**, which is defined as the minimum time interval required for the consumer to receive the notification before the event happens (time of the prediction). This is better described in [10] and the areas of improvement 6 of [11].
- Support of the **threshold model**, as defined in cl. 6.9 of [6], which is requested in subscription/notification in order to define the triggering conditions and allow the consumer to be notified any time one or more predicted KPI falls within or outside of a predefined range.
- Support for **confidence,** or the confidence of a specific prediction. The semantics and use of *confidence* can be fleshed out during implementation.

With the proposed modifications, the data model for the service can look like in the following table:

Table 6: PredictedQoS data model with proposed enhancements

Name	Data type	Cardinality	Remarks
Target	String	0..1	Specifies whether the target of the predicted QoS is either: a QoS prediction request tailored to a single vehicular UE if the information element is not populated (as of current ETSI MEC 030 API), or E2E involving another vehicular UE, in which case it includes an identifier of the relevant vehicular UE.
timeGranularity	TimeStamp	0..1	Time granularity of visiting a location.
locationGranularity	String	1	Granularity of visited location. Measured in metres.
noticePeriod	TimeStamp	0..1	Information on when the predicted QoS is needed at the service consumer interface. The value of the notice period depends on the application reaction that has to be triggered by the service consumer.
Routes	Structure (inlined)	1..N	Information relating to the potential routes of a vehicular UE.
>routeInfo	Structure (inlined)	2..N	Information relating to a specific route. The first structure shall relate to the route origin and the last to the route destination. Intermediate waypoint locations may also be provided.
>>location	LocationInfo	1	Vehicular UE location.
>>time	TimeStamp	0..1	Estimated time at the location.
~~>>rsrp~~	~~Uint8~~	~~0..1~~	~~Reference Signal Received Power as defined in ETSI TS 136 214 [i.13].~~ ~~Shall only be included in the response.~~
~~>>rsrq~~	~~Uint8~~	~~0..1~~	~~Reference Signal Received Quality as defined in ETSI TS 136 214 [i.13].~~ ~~Shall only be included in the response.~~
>>qos	Structure (inlined)	1	Predicted QoS at the related time and vehicular UE location
>>>stream	Structure (inlined)	1..N	Predicted QoS at the related time and vehicular UE location for the specific data stream. In the 3GPP network case, this is mapped to a QoS flow. Stream needs to also contain the stream ID which, in case of the 3GPP network, can be mapped on to the 5QI or QCI.
>>>>qosKpi	Structure (inlined)	1..N	This structure contains the prediction for a specific QoS KPI related to a given data stream.
>>>>>kpiName	String	1	The name of the KPI (e.g. latency, UL bitrate, etc.).
>>>>>kpiValue	String	1	Information on the predicted value for the specific QoS KPI. It can be in different forms, such as upper bound and lower bound, CDF, actual value, etc.
>>>>>confidence	String	0..1	Confidence of the prediction, as returned by the relevant domain PF. The value and the measurement of the confidence depends on the SLA.
Note: The data type of locationGranularity is a string which indicates the granularity of a visited location by means of latitudinal and longitudinal margins.			

Observation – A number of changes can be applied to the existing data model for journey-specific QoS predictions that can be summarised as follows:

In the request and/or in the subscription:

- Support for the notice period/time horizon, as described in [9] and in area of improvement 6 of [10].
- Support for the threshold model, as defined in cl. 6.9 of [6] .

In the response and/or in the notification:

- Replacing (or complementing) the prediction on *rsrp* and *rsrq* with more high-level QoS KPI-based predictions that can be more representative of the application requirements, such as the ones listed in Section 5.6.3 of [10], with measurements of those QoS KPIs related to the application layer between the two vehicle applications in Scenarios 1, 2 and 3 and between the vehicle and the application entity in the third-party domain in Scenario 4.
- Support reporting of QoS prediction per data stream (e.g. QoS flow) rather than per radio link, since the same radio link may support differentiated traffic treatment according to the 5G QoS model.
- Support for the prediction confidence.

5.4.4. V2X message interoperability and QoS prediction

MEC facilitates V2X interoperability in a multi-vendor, multi-network and multi-access environment. In this context, VIS facilitates the interaction between the MEC applications and the V2X message distribution server (e.g. message broker) which is needed to distribute V2X messages related to non-session-based V2X services. For this purpose, ETSI MEC 030 specification [7] includes a publish/subscribe API for V2X message interoperability, described in Section 5.5.9 of the above specification. The service allows a consumer (a MEC application or a MEC platform):

- to subscribe the V2X messages which come from different vehicle OEMs or operators,
- to publish V2X messages towards different vehicle OEMs or operators, and
- to receive notifications when an another vehicle has posted a message according to this service, via a call-back mechanism.

In this way, the service consumer/application can use the standard ETSI VIS API to send and receive messages to the V2X message distribution service without having to interact with it directly.

The following tables describe the data structures that can be used in this service [7].

Table 7: Attributes of the V2xMsgPublication data type

Attribute name	Data type	Cardinality	Description
stdOrganization	Enum	1	Standardization organization which defines the published V2X message type: • ETSI: European Telecommunications Standards Institute. See note 1.
msgType	Enum	1	Published V2X message type. Its value is defined by the standardization organization indicated by the attribute stdOrganization. See note 2.
msgEncodeFormat	String	1	The encode format of the V2X message, for example base64.
msgContent	String	1	Published V2X message content. Its format is defined by the standardization organization indicated by the attribute stdOrganization.
NOTE 1: Other standardization organizations could be added as needed. NOTE 2: The V2X message types of ETSI shall be used as specified in ETSI TS 102 894-2 [6], clause A.114.			

Table 8: Attributes of the V2XMsgNotification data type

Attribute name	Data type	Cardinality	Description
notificationType	String	1	Shall be set to "V2xMsgNotification".
timeStamp	TimeStamp	1	Date and time of the notification generation.
stdOrganization	Enum	1	Standardization organization which defines the published V2X message type: • ETSI: European Telecommunications Standards Institute. See note 1.
msgType	Enum	1	Published V2X message type. Its value is defined by the standardization organization indicated by the attribute stdOrganization. See note 2.
msgEncodeFormat	String	1	The encode format of the V2X message, for example base64.
msgContent	String	1	Published V2X message content. The format of the string is defined by the standardization organization indicated by the attribute stdOrganization.
_links	Structure (inlined)	1	Links to resources related to this notification.
>subscription	LinkType	1	A link to the related subscription.
NOTE 1: Other standardization organizations could be added as needed. NOTE 2: The V2X message types of ETSI shall be used as specified in ETSI TS 102 894-2 [6], clause A.114.			

Table 9: Attributes of the V2xMsgSubscription data type

Attribute name	Data type	Cardinality	Description
subscriptionType	String	1	Shall be set to "V2xMsgSubscription".
callbackReference	URI	1	URI selected by the service consumer to receive notifications on the subscribed V2X message. This shall be included both in the request and in response.
_links	Structure (inlined)	0..1	Hyperlink related to the resource. This shall be only included in the HTTP responses and in HTTP PUT requests.
>self	Link Type	1	Self-referring URI. The URI shall be unique within the VIS API as it acts as an ID for the subscription.
filterCriteria	Structure (inlined)	1	List of filtering criteria for the subscription. Any filtering criteria from below, which is included in the request, shall also be included in the response
>stdOrganization	Enum	1	Standardization organization which defines the subscribed V2X message type: • ETSI: European Telecommunications Standards Institute. See note 1.
>msgType	Enum	0..N	Subscribed V2X message type. Its value is defined by the standardization organization indicated by the attribute stdOrganization. See note 2.
expiryDeadline	TimeStamp	0..1	Time stamp
NOTE 1: Other standardization organizations could be added as needed. NOTE 2: The V2X message types of ETSI shall be used as specified in ETSI TS 102 894-2 [6], clause A.114.			

When it comes to the delivery of V2X messages in the V2N2V and V2N2I scenarios, the latency at which those messages are delivered to the destination may be significant for the application endpoint. When an application needs to exchange V2X messages, knowing the latency at which those messages will be received can be important information to use in the service implementation. For example, if the latency is too high the application may decide not to wait/not to send a specific V2X message and determine related countermeasures (e.g. abort service or implement a decision according to data that is locally available instead of data cooperatively acquired form other vehicles).

For these reasons, one potential enhancement could be to add information/analytics about the short-term QoS prediction when communicating using the V2X message interoperability API and V2X message distribution server. This information can be used by the application as it complements the contextual information needed to implement service-related decisions. Specifically, latency information (as part of QoS) can become useful for non-session-based V2X services. As an example, the information on the latency prediction could be returned by VIS to the service consumer when performing a V2X message service subscription. In a typical scenario in which there could be multiple senders sending messages to a vehicle UE, VIS may retrieve the predicted latency (e.g. by cooperating with external PFs) and provide this information when the consumer subscribes to the service. For example, if a car is approaching an intersection, all nearby cars may want to send messages to signal their presence, intentions, next manoeuvres, etc. Each sender is connected to the receiving car with a different end-to-end user plane link, for which the VIS may retrieve the latency prediction. Once this information is known to the VIS, it may decide to embed it in the response to the V2X message service subscription. Furthermore, every time this prediction changes (because of changes in the configuration, in the network, or in the vehicle taking part in the use cases, or simply because a different QoS prediction is received from a PF) VIS may need a mechanism to forewarn the service consumer. For this reason, the resulting latency prediction for the end-to-end user plane links may be included as optional information also in other VIS responses, such as those provided to the service consumer when replying to a V2X message publication (as in cl. 5.5.9.2) or a V2X message notification (as in cl. 5.5.9.3).

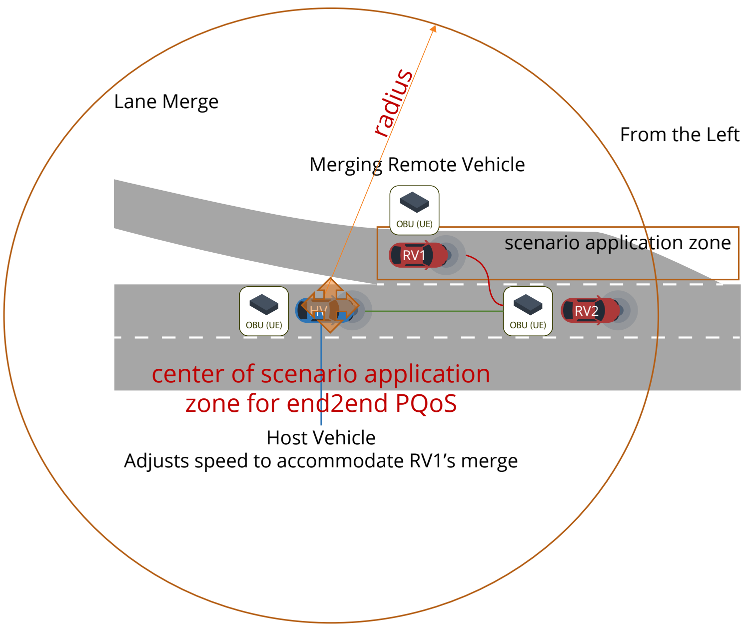

Figure 16: Road traffic scenario of lane merge with three vehicles; the end-to-end user plane links for this scenario are HV-RV1, RV1-RV2 and HV-RV2

As an example, consider the following road scenario in which vehicle RV1 needs to perform a lane merge operation and communicate this to RV2 and HV. V2X messages are exchanged according to the V2X message interoperability functions provided by ETSI GS MEC 030 [7]. VIS is provided end-to-end QoS prediction information for the network domains related to the end-to-end user plane links between each vehicle. In the example, the applicable links are HV-RV1, HV-RV2 and RV1-RV2. For every service request received in the scenario described above, the VIS may reply including the QoS prediction of the latency in the HTTP response message provided to the service consumer. The latency QoS prediction thus received by the service consumer can be regarded as the worst case among those generated for E2E links HV-RV1, HV-RV2 and RV1-RV2.

Observation – it is proposed to enhance the V2X message interoperability service described in cl. 5.5.9 of ETSI GS MEC 030, so when the VIS receives requests *V2xMsgSubscription*, *V2XMsgPublication*, or *V2XMsgNotification* it can optionally include (if such information is available) the latency P-QoS of the E2E user plane link related to the V2X message interoperability service.

Another possibility for the service consumer to retrieve the QoS prediction (e.g. latency prediction) for the V2X message interoperability service is to use the *identifier* for an active V2X message subscription in the request or subscription for QoS prediction, described in the previous section (as in cl. 5.4.2 and enhancements proposed in this document). In such cases, VIS may use the identifier for an active V2X message subscription to determine the group of the vehicles to be associated with the notification and therefore the individual links for which the QoS prediction would be requested. For each link, the VIS can determine the domains traversed when contacting the relevant V2X message distribution server and finally determine the relevant QoS prediction for that network segment. If the QoS prediction is available (VIS may retrieve such information as described in the previous chapters), VIS may return the resulting QoS prediction to the service consumer (e.g. latency). In the example of Figure 16 this refers to the worst case latency scenario among the links HV-RV1, RV1-RV2 and HV-RV2, considering that the group in the *V2XMsgSubscription* contains three vehicles, HV, RV1 and RV2.

The proposed enhancements related to the request for QoS prediction sent to VIS by the service consumer include the possibility of adding an identifier for the *V2XMsgSubscription*.

Figure 17: Example implementation for a consumer subscribed to V2XMsgSubscription service and using the ID to request QoS prediction for the group in the V2XMsgSubscription

As shown in Figure 17, in step 1 the service consumer acting for a vehicle (e.g. in a lane merge use case) subscribes to *V2XMsgSubscription* as in cl. 5.5.9.1 of [7].

In step 2, VIS responds including among other information the attributes *self*, which is an ID for the subscription (unique in VIS) and the attribute *msgType* which, according to cl. A.114 of [35], can contain the *Type* of the ITS message, e.g. DENM, CAM, POI, etc. These attributes can thus be used as subscription *identifiers* for the V2X Message interoperability service.

In step 3, as specified in cl. 5.5.9.3 of [7], VIS may provide a notification to the service consumer.

In step 4, the message to the service consumer results in a *204 No Content* reply, simply to recognise that the message was received.

In step 5, according to the enhancements proposed in this document, the service consumer can use the subscription identifier of the V2X message interoperability service to request the QoS prediction related to the group affected by the service. This can be done by including in the request the following attributes: *target end2end* to specify that this is the end-to-end QoS prediction between two application endpoints, the *id* of the message subscription and the *msgType* received in step 2.

In step 6, VIS computes the end-to-end prediction and in step 7 it responds with the QoS prediction. This information is the P-QoS aggregated for the E2E user plane links related to the group identified by the *V2XMsgSubscription*.

It is important to note that the same vehicle may have multiple subscriptions for *V2XMsgSubscription* interoperability services, e.g. mapped to different use cases (e.g. one for a lane merge operation, one for a platooning use case, etc.). Each of these subscriptions may identify different groups and therefore different values for the QoS prediction. This can be easier to understand by the following picture:

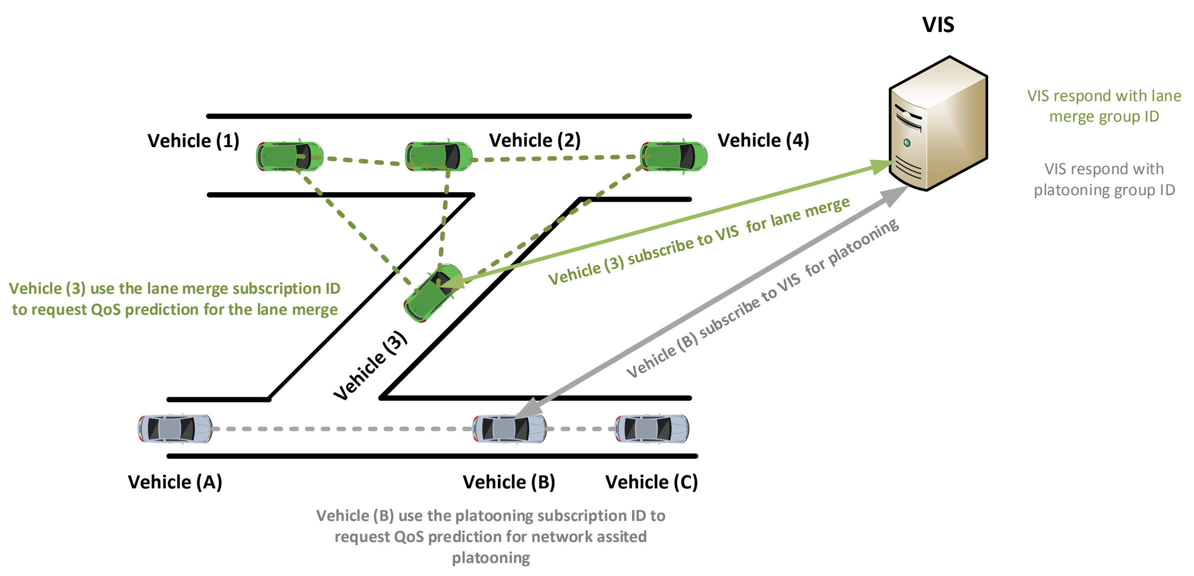

Figure 18: Example of the usage of subscription ID for the request of QoS predictions for lane merge and platooning use cases

Observation – it is proposed to enhance the service for journey-specific QoS predictions (in the case of end-to-end user plane QoS predictions) by including an additional input in the request or subscription: the identifier of a V2X message subscription. The VIS may respond to such requests by providing the QoS prediction of the worst-case latency calculated for all the user plane links related to the specific V2X message subscription.

5.5. Open issues and summary

Section 5 has analysed the scenarios relevant to the generation and delivery of predictive edge analytics and the solutions available in the industry. In general, each of the scenarios requires generation and delivery of analytics in up to six different network domains. While a standard solution that may cover all or part of those domains is advisable, at the time of writing none of the solutions can provide a standard interface for the generation of analytics in multi-domain environments.

While PF entities (for the actual P-QoS and related local data analytics generation) can leverage standardised solutions or vendor-specific implementations, the support of MEC platforms, and the existing ETSI MEC VIS (V2X information service API) may help in the distribution and delivery of those analytics, providing a single point of contact for the application instances and the relevant MEC applications. However, this solution has to be complemented by proper standardisation enabling MEC Federation among various domains (ongoing standardisation work is currently available in [16] and [17]). Some of the suggested 5GAA enhancements have already been implemented in the recently released specification [41]. Other enhancements have been suggested for future study.

6 High-level architectural considerations on network slicing in MEC systems

In the context of the areas of study n. 1 and 3 listed in Section 4.1.1 one the topics that could be further studied is the prediction of the availability of network slices and business slices/business bundles. This is further described in Section 6.1.

In the context of the area of **study 2** listed in Section 4.1.1 one of the topics that could be further studied is how a MEC application can use and select multiple network slices. This is further described in Section 6.2.

Section 6.3 summarises the suggested improvements and the issues that are still open.

6.1 Prediction of the availability of network slices and business slices/business bundles

When a V2X application demands simultaneous use of multiple network slices it can bring additional complexity to be solved by network planners and mobile operator's end customers such as OEMs. This is because network slicing can be viewed differently by different stakeholders – for instance, the perception of "slice" by a corporate customer or sales representative of an MNO may be different from that of a network planning engineer of an MNO. Hence, it is recommended that a different term is used. In order to facilitate this, the concept of "business slice" or "business bundle" [26] can be used.

It is anticipated that MNOs could deploy a single network slice type that satisfies the needs of multiple verticals, as well as multiple network slices of different types packaged as a single product targeted towards business customers (a business bundle or business slice) having multiple and diverse requirements (for example, a vehicle may need simultaneously a high bandwidth slice for infotainment and an ultra-reliable slice for telemetry, assisted driving, etc.) [26].

In simple terms, MNO customers are interested in using a network infrastructure to meet their requirements. Hence, for the customer the network slice is a customised (virtual) network with certain characteristics and guarantees. That is what is defined as "business slice" or "business bundle" [25].

A business slice is what is offered by the MNOs to the customer and always associated with an SLA. The SLA represents a contract between the network provider and its customer which stipulates a specified level of service, support options, and guaranteed level of system performance in terms of, for instance, downtime or uptime.

On the other hand, in order to support services with diverse requirements, MNOs have to streamline their network resources to provide different services efficiently and economically because non-optimised solutions can increase their overhead. Hence, parts of the resources are configured in order to support, for instance, low latencies and high reliability while other resources are configured to provide high data throughput. This is called an end-to-end (requirement driven) network slice, and is depicted in the figure below.

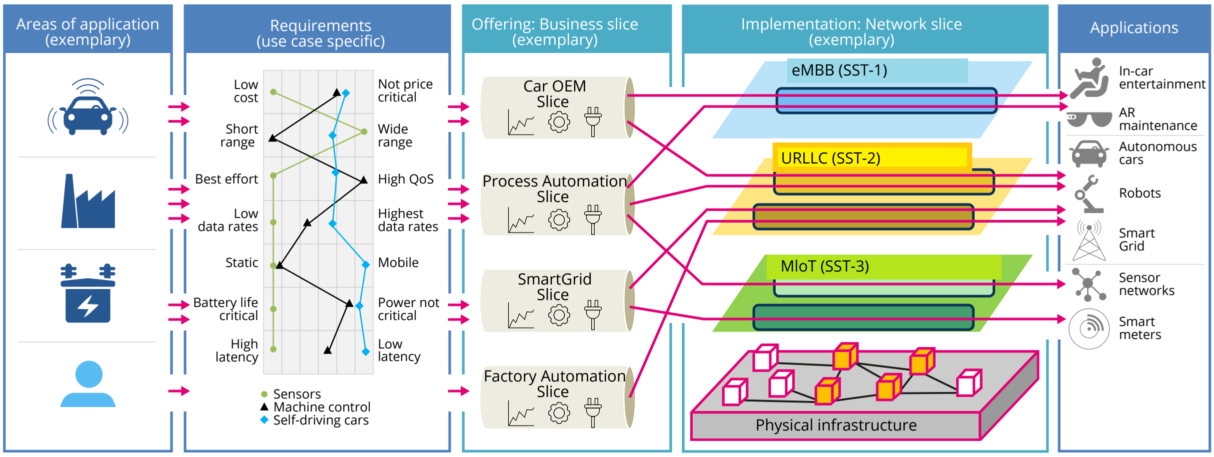

Figure 19: Business slice vs. network slice (Source: T-Mobile)

Figure 19 shows how three network slices available in a network are each mapped with a different SST – i.e., SST1-3. In the example, SST1 is optimised for large bandwidth (e.g. eMBB type of service) without supporting demanding latency requirements on fast-moving devices. SST2 is optimised for an URLLC slice aimed at services with low latency and high reliability requirements. SST3 is optimised for MIoT and is able to provide connectivity to multiple sensor devices and suitable for small data packets and maximum energy efficiency.

For instance, a car OEM-1 may require network slice types URLLC and MIoT for its fleet operation. The network operator may study its SLA and create a business slice called OEM-1 that in fact consists of network slice types URLLC and MIoT. Similarly, another car OEM-2 requires eMBB and URLLC for its fleet operation. In response, a network operator may create a business slice called OEM-2 consisting of network slice types URLLC and eMBB. This way it may appear that a corporate customer gets a slice (business slice) that can support any 5QI. In fact, a business slice may consist of one or more network slices.

Accordingly, depending on individual SLAs, a new business slice can be created out of one or many (pre-existing) network slices – a network slice can be identified by an S-NSSAI. Hence, for network planning and dimensioning, what is important to MNO engineers is the ability to identify network slice instances through the S-NSSAI. This is because network slices and related network slice instances need to be dimensioned individually in order to meet the specific customer demands. In turn, the overall business slice can meet customer requirements thanks to the proper dimensioning of the implementing network slices and slice instances.

While a vehicle traverses different areas (e.g. cells, tracking areas) of a mobile network, which slice is supported may change depending on the current position. It is reasonable to assume that MNOs will not deploy unlimited resources ubiquitously. It would be of great advantage for a MEC application to know in advance which slice is available in a predetermined area or along a predetermined route. For this purpose, it is possible to envisage a new function that can provide such predictions based on existing and inferred information.

For this purpose we may distinguish network-level (e.g. performed by NWDAF or other network functions) and application-level (performed by the application) prediction types. Network-level prediction functionality has to be associated with the availability of network slices (i.e. S-NSSAI) and an application-level predictor may bundle network level predictions of network slices to provide a business slice (or business bundle). Network-level prediction deals with what the 5GS provides in terms of network-level resources, PDU sessions, QoS flows, and QoS model guarantees. While application-level prediction operates at the highest layer (i.e., Layer 7 according to OSI-layer model) and it may derive prediction on guarantees from application-layer perspectives. Given an application may require the support of multiple network slices, the application-level predictor can have a **collective view**, although that is not required for a predictor that operates at the network-level.

Observation 1 – in MEC deployments covering wide area networks and supporting network slices with differentiated services it is suggested to consider the introduction of new prediction functionality:

- A network-level prediction functionality – which could be provided by NWDAF or other prediction functions within the 3GPP system – in order to support the prediction of which network slice (e.g. S-NSSAI) is available at a specific location.
- Such functionality can be complemented by higher level prediction functionality – which could reside in the application layer – with the purpose of aggregating network level predictions of network slices in order to provide predictions related to the availability of business slices or business bundles [25].

It has also to be noted that - although dynamic slice management when a single application requires multiple simultaneous slice support is an issue to be solved in the context of a single mobile network operator - the same issue will still pose even bigger challenges when multiple MNOs are involved.

6.2 A MEC application using multiple slices

For the reason mentioned in Section 4.1.4, if an application requires traffic with diverse and extreme performance requirements, a single SST may not be able to support all required traffic from that application as each SST may be optimised to support only a few 5QIs and related QoS parameters efficiently and economically, although it is technically possible to make a single SST support all possible 5QIs.

Although service differentiation and some form of performance guarantee is possible through the 5G QoS model (e.g. GBR and delay critical GBR type QoS flows), ensuring more strict performance guarantees especially for critical services can be achieved relatively easily with the use of network slicing [25]. For instance, although the 5GS strives to guarantee requirements of GBR traffic streams, it is possible that occasionally those guarantees cannot be fulfilled by the 5GS. There is a possibility for the network to downgrade GBR traffic or even drop the QoS flow if a serving base station struggles with available resources. Looking at Table 5.7.4-1 of TS 23.501[19], it is possible to note that there are some GBR traffic QoS flows that have higher priority (i.e. lower priority level) than delay-critical GBR traffic. For instance, 5QI=85 corresponds to delay-critical GBR with PL=21. Suppose a given cell serves GBR of 5QI=1 and delay-critical GBR of 5QI=85. Unless allocation and retention priorities (ARP) are carefully configured, there may be a possibility for GBR to receive prioritised treatment over delay-critical GBR in overloaded situations. Under such circumstances, if there are different network slices to carry different types of traffic streams, it is possible to minimise one type of traffic, pre-empting other more important traffic during (extreme) peaks. This is possible assuming that each network slice has proper resources allocated and provisioned. For instance, all conversational voice type services can be carried by an eMBB SST while controller-to-control communication can be carried by HMTC SST. In this way it is possible to define clear boundaries and avoid different traffic types adversely impacting each other. Network slicing boundaries can be dimensioned/arranged in order to provide different types of services, and make it possible that traffic from one slice (e.g. non-critical) will not impact traffic from a different (more critical) slice, as they are served by segregated network resources. As a reminder, different SSTs such as URLLC, eMBB, MIoT, and HMTC (refer to TS 23.501 clause 5.15.2.2 [19]) could be designed to support different service types. In other words, a network slice or SST suggests a type of service and coarse granular QoS – which can be used to provide a further level of traffic differentiation in addition to the functionality already provided by the 5G QoS model.

Observation 2 – traffic segregation across different slices may be used to minimise impact across different traffic classes and facilitate QoS guarantees.

It may be possible, especially in V2X scenarios, that a V2X application has extreme requirements (e.g. traffic with very high bitrate *combined* with traffic demanding very low latency) [5], such that it can be supported efficiently and economically through traffic separation and simultaneous access to different and distinct SSTs. The IP multimedia subsystem (IMS) is an example of this [22], as described in clause 5.3 of [5], tele-operated driving (ToD) and [23] advanced driving. This is illustrated in Table 10: and Figure 20 for the ToD case.

Table 10: Composition of ToD in terms of different traffic classes (Source: Ericsson [23])

Application class	Recommended 5QI value	5QI features
Video and audio	2	Resource type: GBR PDB: 150ms PER: 10^{-3} SST: e.g., eMBB
Vehicle command	83	Resource type: Delay Critical GBR PDB: 10ms PER: 10^{-4} SST: e.g., URLLC
Status information	4	Resource type: GBR PDB: 300ms PER: 10^{-6} SST: e.g., URLLC
Conversational voice	1	Resource type: GBR PDB: 100ms PER: 10^{-2} SST: e.g., eMBB

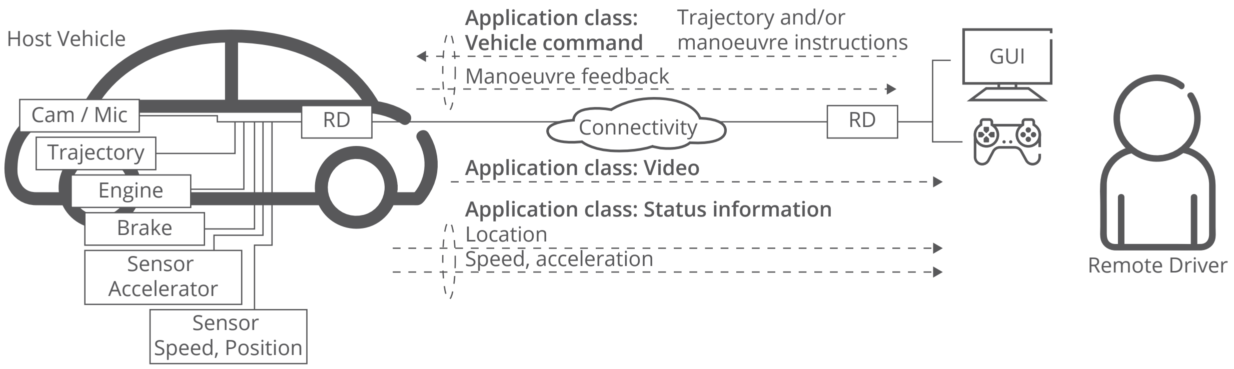

Figure 20: Example realisation of tele-operated driving (Source: [23])

The current assumption within 3GPP is that traffic from a single application can be satisfied by a single PDU session, thus a single S-NSSAI can be used, since S-NSSAI is PDU session-specific. If, on the other hand, a single application generates traffic with diverse and extreme performance requirements, measures have to be taken for a 5GS to enable the creation of multiple PDU sessions per application to be mapped onto different slices – this is depicted in Figure 21. It is not realistic to assume that ToD requirements will be supported economically and efficiently by a single S-NSSAI, because some of those requirements can be considered diverse and extreme. Instead, for deployment reasons, it is easier to achieve those requirements using at least two different S-NSSAIs (with SSTs associated with these S-NSSAIs, respectively uRLLC and eMBB). Use of a set of different and extreme QoS requirements can also be observed for other use cases such as advanced driving or lane-merge.

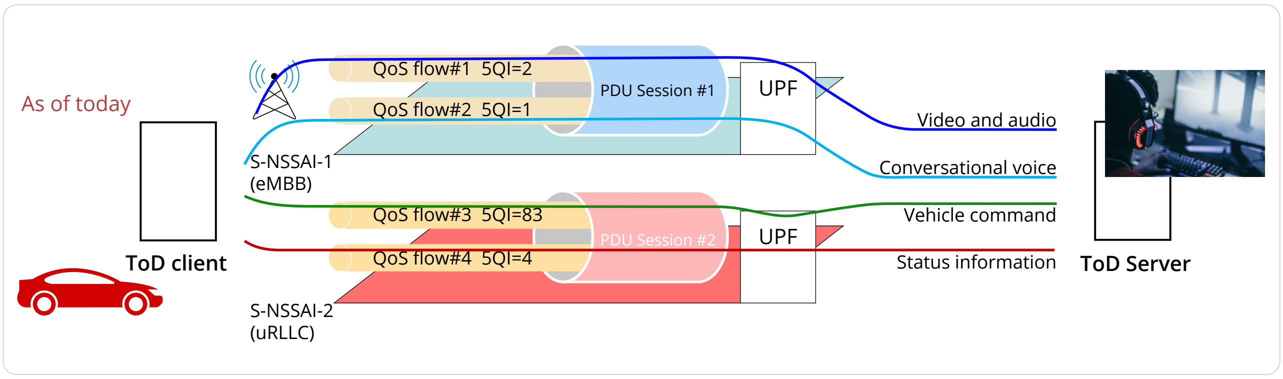

Figure 21: ToD demanding at least two SSTs of eMBB and URLLC type

Observation 3 – it is possible that connectivity for a V2X application in the MEC environment may require the use of one or more S-NSSAIs, each of them mapped onto different SSTs.

In the same way, a single V2X application may require the use of one or many edge application servers depending on traffic types or slices they handle. It may be possible that a single EAS can handle traffic belonging to different slices. However, in case different EASs are required, measures have to be taken to find those that are close to each other so the application can run smoothly thanks to similar propagation delays.

Observation 4 – measures need to be taken to find an EAS that can serve diverse traffic types supported by different slices.

Support of different S-NSSAIs (on required SSTs) and different EASs may not always be possible along the entire route of a vehicle moving over a wide geographical area and across country borders. It is quite useful if a prediction functionality checks availability of slices and EASs along the route and send prompts if gaps (slices or EASs) are identified/anticipated. This can help a vehicle plan potential detours based on input from prediction functionality and depending on the services that may be needed along the planned path, as already discussed in the previous chapter.

6.3 Open issues and summary

Network slicing is a powerful tool that can be used in MEC deployments to provide connectivity and manage related network resources. This Technical Report has suggested a number of potential enhancements that can be provided in the context of 5G networks to better support MEC applications in the automotive domain.

One potential enhancement includes the possibility for the 5GS to predict the availability of network slices (or S-NSSAIs) in a specific area or location. Slices supporting specific services are not expected to be available everywhere. The application may benefit from knowing such information in advance (as opposed to when the application instance in the vehicle is already in the specific area or location) in order to better adapt to the conditions supported with the available network slices. Such network functionality can be complemented by higher level prediction functionality – which could reside in the application layer – with the purpose of aggregating network level predictions of network slices to provide predictions about the availability of business slices or business bundles of network slices.

Another potential enhancement that has been suggested concerns the possibility for the 5GS to enable the traffic directed towards and originated from a MEC application across multiple slices. The traffic segregation across different slices may be used to minimise impact across different traffic classes and facilitate QoS guarantees. This can be particularly important for applications with extremely diverse traffic classes, such as traffic with very low latency together with traffic requiring extremely high throughput. The 5GS may be able to support such requirements by segregating the different traffic categories into different network slices, which can be engineered to support different categories of 5QIs. In these cases, it is requested that the 5GS can help the MEC application find an EAS serving diverse traffic types supported by different network slices. Further work may be required to map the automotive use cases to specific 5GS QoS parameters and characteristics, and also to define standard network slice templates aligned with those 5GS parameters and characteristics. Standardised templates associated with a set of use cases and required MEC functionality may help in the provisioning of standardised QoS and network characteristics across multiple mobile network operator deployments.

6. Conclusions

Edge computing is an important topic in V2X, as many use cases [3] demand lower latency and higher reliability and need to handle large amounts of data being exchanged between vehicles and with road infrastructure. The support of specific performance requirements is key for the realisation of those use cases. When such requirements are not fulfilled by the underlying network and infrastructure, the application may require closed-loop adaptation to cope with the potential undesired effects, which may also include poor user experience, limited support of selected features, or the service no longer being available to the user. Other 5GAA Work Items [10],[11] have studied how such early notifications may be generated in the 5GS and delivered to the application. However, no previous work has studied how predictive edge analytics can be generated and delivered in distributed MEC deployments. In particular, when such deployments involve multiple MNOs, OEMs and additional third parties – such as service providers, application developers, IP interconnect operators, and road transport authorities – the generation and delivery of such predictive edge analytics is not a trivial task and may require interaction across different stakeholders. In this heterogeneous context, the MEC platform with its standard APIs may facilitate such interaction and help in achieving better contextual awareness and predictive edge analytics.

This TR provided a technical analysis of how **predictive edge analytics** can be generated and delivered in distributed MEC deployments. In particular, it described how to provide predictive edge analytics in applications deployed in distributed MEC environments, with particular focus on multi-domain, inter-MNO and inter-OEM aspects. It analysed the current available solutions; it proposed a reference architecture for predictive edge analytics aligned with available 3GPP and ETSI MEC standards, and tried to identify potential gaps and suggested improvements. Some of those improvements proposed by 5GAA in [12] have contributed to the normative work in ETSI ISG MEC, and have been included in the latest release of the VIS specification [41].

More in detail:

- Section 5 analysed the scenarios relevant to the generation and delivery of predictive edge analytics and the solutions available in the industry. In general, each of the scenarios requires generation and delivery of analytics in up to six different network domains. While a standard solution that may cover all or part of those domains is advisable, at the time of writing none of the solutions can provide a standard interface for generating such analytics in a multi-domain environment.
- While PF entities (for the actual P-QoS and related local data analytics generation) can leverage standardised solutions or vendor-specific implementations, the support of MEC platforms and the existing ETSI MEC VIS (V2X Information Service API) may help in the distribution and delivery of those analytics, providing a single point of contact for the application instances and the relevant MEC applications. However, this solution has to be complemented by proper standardisation enabling the MEC Federation among various domains (ongoing standardisation work is currently available in [16] and [17]). Some of the suggested 5GAA enhancements have been implemented in the latest release of the VIS specification [41]. Other enhancements have been suggested for future study.

Furthermore, the usage and adaptation of **network slices** based on sensed context awareness is an additional tool that can be used in heterogeneous MEC systems for enhanced and more predictable network performance. In this context, this TR analysed the state of the art of network slicing in the context of MEC systems and tried to identify potential gaps and additional requirements for future study.

More in detail:

- One potential enhancement includes the possibility for the 5GS to predict the availability of network slices (or S-NSSAIs) in a specific area or location. Slices supporting specific services are not expected to be available everywhere. The application may benefit from knowing such information in advance (as opposed to when the application instance in the vehicle is in the specific area or location) in order to better adapt to the conditions supported in the available network slices. Such network functionality can be complemented by higher level prediction functionality – which could reside in the application layer – with the purpose of aggregating network level predictions of network slices in order to provide predictions on the availability of business slices or business bundles of network slices.
- Another potential enhancement that has been suggested concerns the possibility for the 5GS to enable the traffic directed towards and originated from a MEC application across multiple slices. The traffic segregation across different slices may be used to minimise impact across different traffic classes and facilitate QoS guarantees. This can be particularly important for applications with extremely diverse traffic classes, such as traffic with very low latency together with traffic requiring very high throughput. The 5GS may be able to support such requirements via the segregation of different traffic categories in different network slices, which can be engineered to support different categories of 5QIs. In such cases, it is requested that the 5GS can help the MEC application find an EAS that can serve diverse traffic types supported by different network slices. Further work may also be required to map the automotive use cases to specific 5GS QoS parameters and characteristics, and also to define standard network slice templates aligned with those 5GS parameters and characteristics. Standardised templates associated with a set of use cases and required MEC functionality may help in providing standardised QoS and network characteristics across multiple mobile network operator deployments.

5GAA is a multi-industry association to develop, test and promote communications solutions, initiate their standardisation and accelerate their commercial availability and global market penetration to address societal need. For more information such as a complete mission statement and a list of members please see https://5gaa.org

5GAA Automotive Association