H2020 5G-TRANSFORMER Project

Grant No. 761536

Definition of the Mobile Transport and Computing Platform

Abstract

This deliverable provides the first version of the Mobile Transport and Computing Platform (5GT-MTP) design. The deliverable addresses the following aspects of the 5GT-MTP, namely: the internal architecture of the 5GT-MTP; the 5GT-MTP northbound interface abstraction towards the service orchestrator (5GT-SO); the workflows between the 5GT-SO and the 5GT-MTP as well as workflows among the various components of the 5GT-MTP; and the mapping of the 5G-TRANSFORMER use cases to the 5GT-MTP.

Document properties

Document number	D2.1
Document title	Definition of the Mobile Transport and Computing Platform
Document responsible	Charles Turyagyenda (IDCC)
Document editor	Charles Turyagyenda (IDCC)
Editorial team	Paola Iovanna (TEI), Giuseppe Imbarlina, Luca Valcarenghi (SSSA)
Target dissemination level	Public
Status of the document	Final
Version	1.0

Production properties

Reviewers	Andres Garcia-Saavedra, Thomas Deiß, Xi Li, Giada Landi, Dmitriy Andrushko, Pantelis Frangoudis, Philippe Bertin, Cao-Thanh Phan.

Disclaimer

1 Table of Contents

List of Contributors

Partner Short Name	Contributors
TEI	Paola Iovanna, Giuseppe Imbarlina, Erin Elizabeth Seder, Teresa Pepe
UC3M	Antonio de la Oliva, Antonio Pastor, Carlos J. Bernardos
NEC	Andres Garcia Saavedra, Fabio Giust, Josep Xavier Salvat
ATOS	Juan Brenes Baranzano
NOK-N	Thomas Deiß
IDCC	Charles Turyagyenda
TID	Luis Miguel Contreras Murillo
ORANGE	Thouraya Toukabri, Frederic Klamm, Philippe Bertin
BCOM	Farouk Messaoudi, Cao-Thanh Phan
NXW	Giada Landi
MIRANTIS	Dmitriy Andrushko
CTTC	Josep Mangues
POLITO	Christian Vitale, Carla Chiasserini
EURECOM	Adlen Ksentini, Pantelis Frangoudis
SSSA	Luca Valcarenghi, Nicola Sambo
ITRI	Chia-Lin Lai

List of Figures

List of Tables

List of Acronyms

Acronym	Description
3G	Third Generation
3GPP	3rd Generation Partnership Project
5G	Fifth Generation
5GPPP	5G Infrastructure Public Private Partnership
5GT-MTP	Mobile Transport and Computing Platform
5GTP	5G-TRANSFORMER Provider
5GT-SO	Service Orchestrator
5GT-VS	Vertical Slicer
AN	Access Network
API	Application Programming Interface
AS	Application Server
BSS	Business Support System
BTS	Base Transceiver Station
CAM	Cooperative Awareness Messages
CAS	Collision Avoidance Server
CPU	Central Processing Unit
CR	Cloud Robotics
CSMF	Communication Service Management Function
DC	Data Centre
DENM	Decentralised Environmental Notification Message
DNS	Domain Name System
E2E	End to end
eNodeB	Evolved Node B
EPC	Evolved Packet Core
EPCaaS	EPC as a Service
ETSI	European Telecommunication Standardization Institute
IaaS	Infrastructure as a Service
ICA	Intersection Collision Avoidance
IoT	Internet of Things
ISG	Industry Steering Group
KPI	Key Performance Indicator
LC	Lifecycle
LCM	Lifecycle Management
LTE	Long Term Evolution
M&E	Media and Entertainment
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MEO	Multi-access Edge Orchestrator
MEP	Multi-access Edge Platform
MEPM	MEP Manager
MIMO	Multiple Input Multiple Output
MLPOC	Multiple Logical Point of Contact
MME	Mobility Management Entity
MNO	Mobile Network Operator
MVNE	Mobile Virtual Network Enabler
MVNO	Mobile Virtual Network Operator
NaaS	Network as a Service
NBI	Northbound Interface
NETCONF	Network Configuration Protocol

NF	Network Function
NFP	Network Forwarding Path
NFV	Network Function Virtualization
NFVI	Network functions virtualization infrastructure
NFVIaaS	NFVI as a Service
NFV-NS	NFV Network Service
NFV-NSO	Network Service Orchestrator
NFVO	NFV Orchestrator
NFVO-RO	Resource Orchestrator
NSD	Network Service Descriptor
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NSSMF	Network Slice Subnet Management Function
OFDM	Orthogonal Frequency Division Multiplexing
OLE	On site Live Event Experience
OSS	Operating Support System
PM	Performance Management
PNF	Physical Network Function
PNFM	PNF Manager
PoA	Point of Access
PoP	Point of Presence
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAM	Random Access Memory
RAN	Radio Access Network
REST	Representational State Transfer
RNIS	Radio Network Information Service
ROADM	Reconfigurable Optical Add-Drop Multiplexer
S/P-GW	Serving/Packet Gateway
SBI	Southbound Interface
SDN	Software Defined Networking
SLA	Service Level Agreement
SLPOC	Single Logical Point of Contact
SMS	Short Messaging Service
TD	Technology Domain
TETRA	Terrestrial Trunked Radio
TOSCA	Topology and Orchestration Specification for Cloud Applications
UC	Use Case
V2I	Vehicle to Infrastructure
V2X	Vehicle to Everything
VA	Virtual Application
vCPU	Virtual CPU
VDC	Virtual Data Centre
VDU	Virtual Deployment Unit
vEPC	virtual Evolved Packet Core
vEPCaaS	vEPC as a Service
VIM	Virtual Infrastructure Manager
LL	Logical Link
LLD	Logical Link Descriptor
VM	Virtual Machine
VNF	Virtual Network Function

VNFC	VNF Component
VNFD	VNF Descriptor
VNFFG	VNF Forwarding Graph
VNFM	VNF Manager
VSD	Vertical Service Descriptor
VSI	Vertical Service Instance
WIM	Wide area network Infrastructure Manager
XCI	Crosshaul Control Infrastructure
XFE	Crosshaul Forwarding Element
XML	Extensible Mark-up Language
XPU	Crosshaul Processing Unit

Executive Summary and Key Contributions

The 5G-TRANSFORMER project aims to transform today's rigid mobile transport networks into an SDN/NFV-based **Mobile Transport and Computing Platform** (5GT-MTP), that brings the "Network Slicing" paradigm into mobile transport networks by provisioning and managing slices tailored to the specific needs of vertical industries. The primary responsibilities of the 5GT-MTP are two-fold. First is the coupling of radio, transport, storage and computational resources required by the vertical services; and second is providing an abstracted view of the resources, to the service orchestrator (5GT-SO) thus hiding the complexity of the specific underlying technologies. Hence the current architecture of NFV MANO with the related interfaces must be reviewed to deal with the challenges of the 5GT-MTP.

This deliverable provides the first version of the 5GT-MTP design. The deliverable addresses the following aspects of the 5GT-MTP, namely: the internal architecture of the 5G-MTP; the 5G-MTP Northbound Interface (NBI) abstraction towards the Service Orchestrator (5GT-SO); the workflows between the 5GT-SO and the 5GT-MTP as well as workflows among the various components of the 5GT-MTP; and the mapping of the 5G-TRANSFORMER use cases to the 5GT-MTP. The following highlights the key achievements in this deliverable:

- A comprehensive description of the 5GT-MTP system architecture including Physical Network Functions (PNFs), Virtual Network Functions (VNFs), Virtual Infrastructure Manager (VIM), Wide area network Infrastructure Manager (WIM), and 5GT-MTP Single Logical Point of Contact for resource orchestration (5GT-MTP NFVO-RO SLPOC).
- A detailed description of the 5GT-MTP workflows associated with the following lifecycle events: instantiating a non-nested network service, modifying a non-nested network service, terminating a non-nested network service, VNF instantiation, VNF termination and monitoring of virtual resources.
- An exhaustive characterization of the 5GT-MTP technical requirements at different stages of the service lifecycle, i.e. service discovery, assurance, fulfilment and decommissioning.
- An in-depth description of the 5GT-MTP NBI towards the 5GT-SO specifying the abstraction of resources exposed by 5GT-MTP NFVO-RO SLPOC.
- A detailed mapping of the vertical use cases (i.e. automotive, entertainment, e-Health, e-Industry and MNO/MVNO) to the 5GT-MTP. Particularly, each use case describes the resource abstraction exposed by the 5GT-MTP to the 5GT-SO.
- An analysis of the 5GT-MTP innovations beyond the state-of-the-art, namely the integration of a MEC platform, the ability to compose a connectivity service and expose it to the 5GT-SO, and the decoupling of the VIM from the NFVO and VNFM.
- Baseline examples of the YANG information modelling for the following components: logical link (i.e., a physical path connecting two physical node interfaces), computational resources, and storage resources.

Future work is anticipated to expand and refine these results by filling gaps identified, such as: the definition of data models for the abstracted resources, the specification of resource orchestration algorithms and the extension of standard interfaces to support the 5GT-MTP and 5GT-SO interaction.

1 Introduction

In contrast to previous mobile communication technologies, 5G promises to support a variety of vertical use cases' requirements rather than merely providing a general-purpose albeit high-capacity pipe. 5G-TRANSFORMER replaces the rigid "one-size-fits-all" deployments with a customizable SDN/NFV-based 5G Mobile Transport and Computing Platform (5GT-MTP) capable of simultaneously supporting a diverse range of networking and computing requirements specific to the vertical industries, i.e. vertical slicing. The 5GT-MTP is an integral part of the 5G-TRANSFORMER architecture and hosts the physical and/or virtual mobile transport network and computing infrastructure within which vertical services are deployed.

Against this backdrop, the objective of this deliverable is to provide the first version of the 5GT-MTP design developed over the first ten-month period of the project since kick-off in June 2017. The deliverable addresses the following aspects of the 5GT-MTP, namely: the internal architecture of the 5GT-MTP; the 5GT-MTP northbound interface abstraction towards the service orchestrator (5GT-SO); the workflows between the 5GT-SO and the 5GT-MTP as well as workflows among the various components of the 5GT-MTP; and the mapping of the 5G-TRANSFORMER use cases to the 5GT-MTP. The deliverable is structured as follows.

Section 2 presents a high-level description of the 5G-TRANSFORMER system architecture highlighting the following key building blocks: The Vertical Slicer (5GT-VS), the Service Orchestrator (5GT-SO) and the Mobile Transport and Computing Platform (5GT-MTP).

Section 3 describes the technical requirements of the 5GT-MTP at different stages of the service lifecycle.

Section 4 describes the 5GT-MTP system architecture and identifies the core components, namely: Physical Network Functions (PNFs), Virtual Network Functions (VNFs), Virtual Infrastructure Manager (VIM), Wide area network Infrastructure Manager (WIM), VNF Managers (VNFMs) and NFV Orchestrator (NFVO) Resource Orchestrator (NFVO-RO) and Single Logical Point of Contact (SLPOC). Section 4 also defines 5GT-SO-5GT-MTP reference points and provides a detailed description of the resource abstraction exposed via the 5GT-MTP NFVO-RO SLPOC. Finally, Section 4 presents the 5GT-MTP innovations beyond the state-of-the-art and examples of the YANG information modelling for computational and storage resources.

Section 5 presents the 5GT-MTP workflow descriptions associated with the following lifecycle events: instantiating a non-nested network service, modifying a non-nested network service, terminating a non-nested network service, VNF instantiation, VNF termination and monitoring of virtual resources.

Section 6 presents a mapping of the vertical use cases to the 5GT-MTP. Particularly, each use case describes the resource abstraction exposed by the 5GT-MTP to the 5GT-SO.

Finally, in **Section 7**, a conclusion is presented to summarize the findings of this deliverable, as well as setting the prospects for future work.

In order to keep the main body of the document as short as possible, several annexes are included at the end, containing additional information and results.

2 5G-TRANSFORMER System Overview

To describe the 5GT-MTP within its context, we present in this section a summary[1] of the system architecture described in [1]. Relevant reference architectures for the 5G-TRANSFORMER system architecture are presented in Annex II.

The 5G-TRANSFORMER project explores how the network can better serve the needs of 5G-TRANSFORMER customers (i.e., vertical industries and M(V)NOs) by offering the abstraction, flexibility, and dynamic management capabilities they require. In terms of notation, it is important to differentiate between (vertical) service, i.e., that is requested by the customer of the 5G-TRANSFORMER system, from the underlying network service deployed to fulfill the requirements of the vertical. An example of the former is a car manufacturer requesting the deployment of an automotive intersection collision avoidance service. The latter will be deployed in the form of an NFV network service, in general.

The key architectural concept to support such adaptation to the needs of verticals and M(V)NOs is network slicing. The term network slice aligns network functionality to business needs [48], since it allows customers to request not just functions, but also business objectives (e.g., quality of service, security, etc.), as a sort of intent. The scope of a slice may be a single customer facing service (using TM Forum terminology [49]) or a group of such services. The system will also allow infrastructure providers to share the 5G mobile transport and computing infrastructure efficiently among verticals and M(V)NOs, hence enhancing 5G-TRANSFORMER provider network usage efficiency. In terms of deployment, network slices can be implemented by means of ETSI NFV network services.

The architecture is conceived to support multiple combinations of stakeholders by introducing open Application Programming Interfaces (API) among components [1]. Through these APIs, the system hides unnecessary details from the verticals, allowing them to focus on the definition of the services and the required Service Level Agreements (SLAs). As for interfaces, particularly relevant for the goals of the project are east-westbound interfaces, which enable service and resource federation across different administrative domains, allowing 5G-TRANSFORMER service providers to enhance their service offerings to their customers by peering with other providers.

We envision a system of three major components: vertical slicer (5GT-VS), service orchestrator (5GT-SO) and mobile transport and computing platform (5GT-MTP), see Figure 1. The 5GT-VS is the entry point for the vertical requesting a service and it handles the association of these services with slices as well as network slice management. The 5GT-SO is responsible for end-to-end orchestration of services across multiple domains and for aggregating local and federated (i.e., from peer domains) resources and services and exposing them to the 5GT-VS in a unified way. Finally, the 5GT-MTP provides and manages the virtual and physical IT and network resources on which service components are eventually deployed. It also decides on the abstraction level offered to the 5GT-SO.

---

[1] This is text common to [3][4], and this document.
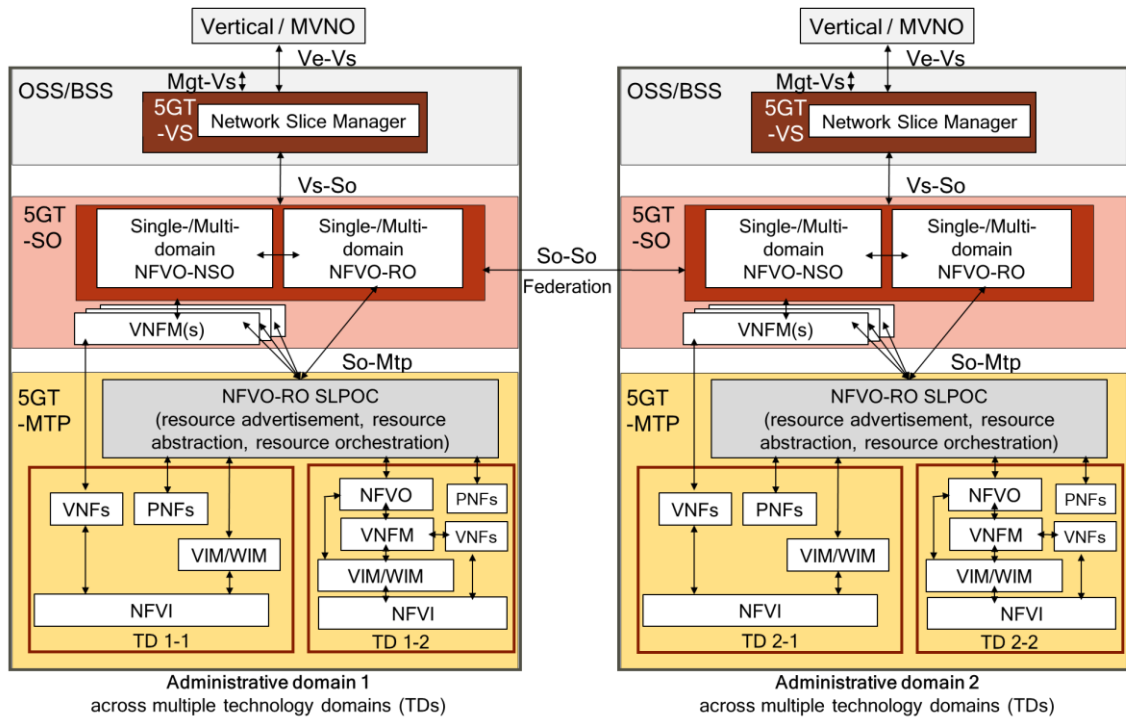
FIGURE 1: 5G-TRANSFORMER SYSTEM ARCHITECTURE

2.1 Vertical Slicer (5GT-VS)

The 5GT-VS is the common entry point for all verticals into the 5G-TRANSFORMER system. It is part of the operating and business support systems (OSS/BSS) of the 5G-TRANSFORMER service provider (TSP) [1]. Vertical services are offered through a high-level interface at the 5GT-VS northbound that is designed to allow verticals to focus on the service logic and requirements, without caring on how they are eventually deployed at the resource level. This latter issue would be up to TSP. Therefore, vertical services, will use those services offered by the TSP. In fact, the 5GT-VS offers a catalogue of vertical service blueprints, based on which the vertical service requests are generated by the vertical. The role of the 5GT-VS is to trigger the actions allowing the 5G-TRANSFORMER system to fulfil the requirements of a given incoming service request. After the appropriate translation between service requirements and slice-related requirements by the VSD/NSD Translator and Arbitrator, corresponding to the Communication Service Management Function (CSMF), as defined in [50], a decision is taken on whether the service is included in an already existing slice or a new one is created.

The vertical slicer is the component of the system that is conscious of the business needs of the vertical, their SLA requirements, and how they are satisfied by mapping them to given slices. Intra-vertical arbitration is also part of the vertical slicer, by which intra-vertical contention is resolved to prioritize those services that are more critical, according to the agreed SLA.

The VSI/NSI Coordinator and LC Manager is the core component of the 5GT-VS. It contains functionality that can be mapped to that of the Network Slice Management Function (NSMF) and Network Slice Subnet Management Function (NSSMF), as defined in [50]. More specifically, the NSMF is in charge of lifecycle management of network slice instances. All possible combinations between vertical services and

network slices exist; that is, a network slice can be shared by different vertical services, but a given vertical service may be mapped to multiple network slices as well. In turn, network slices may be composed by network slice subnets, which may offer part of the functionality required by a given network slice. And network slice subnets may be shared by multiple network slices.

The final result of all this process is a request sent by the 5GT-VS towards the 5GT-SO to create or update the NFV network services (NFV-NS) that implement the slices.

In summary, through this process, the 5GT-VS maps vertical service descriptions and instantiation parameters at the vertical application (VA) level into an NFV-NS (existing or new) implementing the network slice. In turn, such NFV-NS will be updated or created through a network service descriptor (NSD), which is a service graph composed of a set of virtual network functions (VNF) chained with each other, and the corresponding fine-grained instantiation parameters (e.g., deployment flavor) that are sent to the 5GT-SO. Given the operations carried out through it, the VS-SO interface (see Figure 1) takes ETSI GS NFV-IFA 013 [42] as basis.

2.2 Service Orchestrator (5GT-SO)

The NFV-NS from the 5GT-VS is received by the 5GT-SO through the VS-SO interface. The main duty of the 5GT-SO [43] is to provide end-to-end orchestration of the NFV-NS across multiple administrative domains by interacting with the local 5GT-MTP (So-Mtp reference point) and with the 5GT-SOs of other administrative domains (So-So reference point). If needed (e.g., not enough local resources), the 5GT-SO interacts with 5GT-SOs of other administrative domains (federation) to take decisions on the end-to-end (de)composition of virtual services and their most suitable execution environment. Even if a service is deployed across several administrative domains, e.g., if roaming is required, a vertical still uses one 5GT-VS to access the system, and so, the 5GT-SO hides this federation from the 5GT-VS, and thus, the verticals.

The 5GT-SO embeds the network service orchestrator (NFV-NSO) and the resource orchestrator (NFVO-RO) with functionalities equivalent to those of a regular NFV orchestrator and it may be used for single and multi-domains [8].

Since the network slices handled at the 5GT-VS will in general serve complex end-to-end services, in the general case, the corresponding network service will be a composition of nested NFV-NSs. The lifecycle management of this complex NFV-NS is the role of the NFV-NSO.

In case a network service is requested that must be distributed across multiple domains, the 5GT-SO receiving the request becomes the parent NFV-NSO and sends ETSI GS NFV-IFA 013 [42] requests for each of the constituent NFV-NSs to other NFV-NSOs. Therefore, a hierarchy of NFVO-NSOs is established. The child NFVO-NSOs may belong to the same 5GT-SO that received the request from the 5GT-VS or to a peer 5GT-SO, which, in turn, may establish an additional hierarchy, which is transparent for the parent NFVO-NSO. The child NFVO-NSO belonging to the same 5GT-SO would be in charge of the lifecycle management of the constituent service that is eventually deployed over the local 5GT-MTP, i.e., the 5G-MTP with which the 5GT-SO has a direct relationship through the So-Mtp interface. When a child NFVO-NSO belongs to a different domain, there is service federation.

Eventually, a resource-related request is generated towards the underlying NFVO-RO to assign virtual resources towards the deployment of the (constituent) NFV-NS. The NFVO-RO functionality of the 5GT-SO handles resources coming from the local 5GT-MTP (real or abstracted) and from the 5GT-SOs of other administrative domains (abstracted). The NFVO-RO will decide on the placement of the Virtual Network Functions (VNF) of the NFV-NS based on the information available in the NFVI resources repository and the NFV instances repository. Most likely, the information available in these repositories will be more detailed when coming from the local 5GT-MTP than from a federated domain.

As for the NFV infrastructure as a service (NFVIaaS) use case, the 5GT-VS will request the 5GT-SO for a set of virtual resources, as opposed to a complete E2E NFV-NS as before. Therefore, this request is directly handled by the NFVO-RO, which is in charge of allocating resources either from the local 5GT-MTP or from a peer 5GT-SO. The latter option corresponds to resource federation. In this case, the request from the local NFVO-RO will reach the NFVO-RO of the peering domain. In all cases, the interaction between NFVO-ROs is based on ETSI GS NFV-IFA 005 [24]. This also includes the interface with the 5GT-MTP, where an additional NFVO-RO lower in the hierarchy is embedded, as explained below.

Notice that the NFVI resources handled by the NFVO of the 5GT-SO based on which decisions are taken will have a higher or lower abstraction level depending on the policies applied in this respect by the 5GT-MTP and the peering 5GT-SO. In general, the NFVO-RO of the local 5GT-SO will take coarse-grained decisions and the 5GT-MTP and peer 5GT-SO will take finer-grained ones, since they are closer to the actual resources.

The 5GT-SO also embeds the Virtual Network Function Managers (VNFM) to manage the lifecycle of the VNFs composing the NFV-NS. ETSI GS NFV-IFA 006-based interfaces [44] will be used to allow the VNFM interacting with the NFVO-RO Single Logical Point of Contact (SLPOC) entity in the 5GT-MTP, as well as peer SOs for resource allocation requests involving the VNFs under its control. For managing the VNF instances, ETSI GS NFV-IFA 008-based interfaces [45] will be used to allow the VNFM to directly configure the VNF instances running in the 5GT-MTP.

2.3 Mobile Transport and Computing Platform (5GT-MTP)

The 5GT-MTP [46] is responsible for orchestration of resources and the instantiation of VNFs over the infrastructure under its control, as well as managing the underlying physical mobile transport network, computing and storage infrastructure. In general, there will be multiple technology domains (TD) inside a 5GT-MTP (e.g., data centres, mobile network, wide area network). The 5GT-MTP NFVO-RO acts as end-to-end resource orchestrator across the various technology domains of the 5GT-MTP. The computing and storage infrastructure may be deployed in central data centres as well as distributed ones placed closer to the network edge, as in MEC [37]. Therefore, the 5GT-MTP is in charge of managing the virtual resources on top of which the NFV-NSs are deployed.

In terms of resource orchestration, the NFVO-RO acts as single entry point, i.e., single logical point of contact (SLPOC) in ETSI GS NFV-IFA 028 [47] terminology, for any resource allocation request coming from the SO. The So-Mtp interface is based on ETSI GS NFV-IFA 005 [24]and ETSI GS NFV-IFA 006 [44]. The former allows the

NFVO-RO of the 5GT-SO to request resource allocations to the NFVO-RO of the 5GT-MTP, whilst the latter allows the VNFM of the 5GT-SO to request resource allocations for the VNF under its control.

In terms of managing VNF instances, the So-Mtp interface also consists of ETSI GS NFV-IFA 008-based interfaces [45] to allow the VNFM of the 5GT-SO to directly configure the VNF instances running in the 5GT-MTP.

Depending on the use case, the 5GT-MTP may offer different levels of resource abstraction to the 5GT-SO. However, the 5GT-MTP NFVO-RO has full visibility of the resources under the control of the Virtual Infrastructure Managers (VIM) managing each technology domain, since they belong to the same administrative domain. ETSI GS NFV-IFA 005-based interfaces [24]are deployed between the 5GT-MTP NFVO-RO and the 5GT-MTP VIMs. Therefore, when receiving a resource allocation request from the 5GT-SO, the 5GT-MTP NFVO-RO generates the corresponding request to the relevant entities (e.g., VIM or WAN Infrastructure Manager (WIM)), each of them providing part of the virtual resources needed to deploy the VNFs and/or configure the relevant parameters of the PNFs that form the NFV-NS. As a special case, a resource request may be translated into an ETSI GS NFV-IFA 013-based NFV-NS request [42] towards a mobile network technology domain [8]. This option is offered to hide the complexity of the mobile network to the rest of the system whilst keeping the required flexibility inside the mobile domain (e.g., to decide on the most appropriate functional split). Therefore, a full ETSI MANO stack is represented in technology domain 1-2 (see Figure 1) even if the focus of the 5GT-MTP is handling virtual resources and not NFV-NSs. In any case, this NFV-NS is hidden to the 5GT-SO, since it is abstracted as a virtual link.

2.4 Monitoring Architecture

In the 5G-TRANSFORMER framework, each architectural component (i.e. 5GT-VS, 5GT-SO, 5GT-MTP) includes a monitoring service able to provide performance metrics and failure reports targeting the logical entities managed by each component. Following this approach, the 5GT-MTP monitoring service will produce monitoring data about the local physical and virtual resources, the 5GT-SO monitoring service will produce monitoring data about the managed VNFs and NFV network services, while the 5GT-VS monitoring service will produce monitoring data about network slices and vertical services. This hierarchy of monitoring services is shown in Figure 2, where the arrows indicate a consumer-provider interaction. In particular, the 5GT-SO monitoring service can be a consumer of the monitoring service provided by the underlying 5GT-MTP or by a federated 5GT-SO, while the 5GT-VS can be a consumer of the monitoring service provided by the local 5GT-SO.

The monitoring data generated at each layer can be used to feed internal decisions within each architectural component or to serve external consumers of monitoring data. For example, the 5GT-SO monitoring service can elaborate performance metrics about an NFV network service, and these metrics can be used by the 5GT-SO to take scaling decisions for the involved VNFs. On the other hand, the performance metrics computed at the 5GT-SO monitoring service can be provided to the 5GT-VS monitoring service for further elaboration. When metrics and alerts are exchanged between two monitoring services, the level of visibility and disclosure of monitoring information should be regulated based on authorization policies and business agreements, in particularly

when monitoring data that belongs to different administrative entities. This may be the case, for example, between the 5GT-MTP and the 5GT-SO monitoring services when they are handled by different actors or between the monitoring services of federated 5GT-SOs.
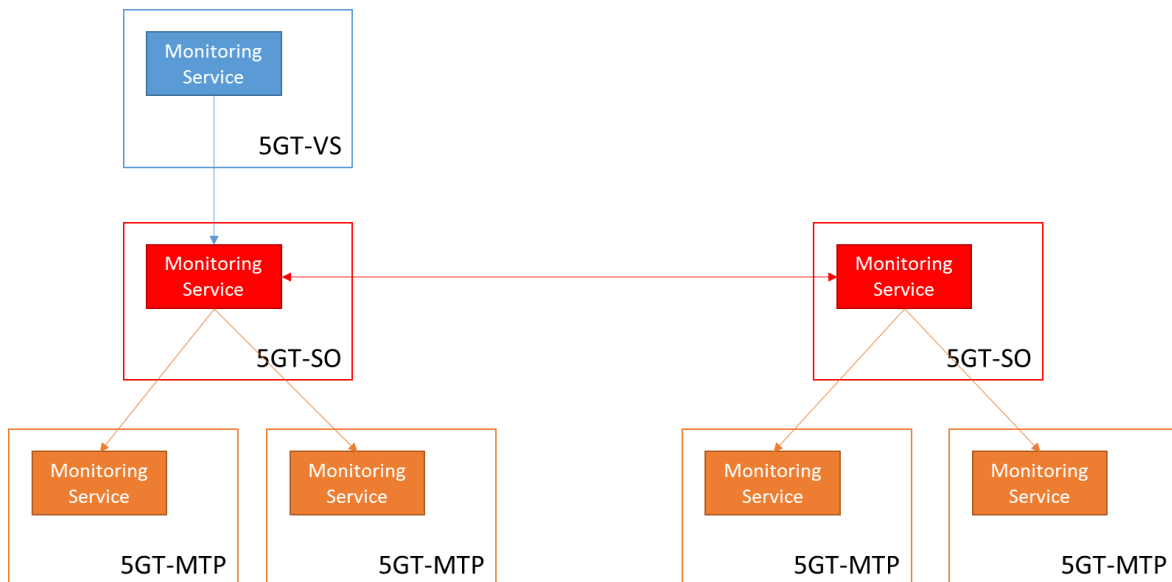
FIGURE 2: HIERARCHY OF MONITORING SERVICES IN 5G-TRANSFORMER ARCHITECTURE

It is important to highlight that the 5G-TRANSFORMER architecture does not impose any constraint on the monitoring platform implementation, but defines just the expected behavior of the service and the external APIs that each monitoring platform should expose to the consumers of its monitoring data. This means that each actor may implement its own specific monitoring platform and in case of overlapping roles, like in the 5GT-VS and 5GT-SO case where they are owned and managed by the same administrative entity, a single monitoring platform may be deployed for both of them.

3 Requirements on the 5GT-MTP

Technical requirements on the overall 5G-TRANSFORMER system have been defined in [1]. The requirements covered in [1] focus on properties related to vertical services and relevant use cases. General requirements related to the overall system are described in [2]. In this section, we define functional requirements specific to 5GT-MTP. The notation used to refer to the different requirements is described in Section 11 (Annex III).

The 5GT-MTP is involved in the service lifecycle at different stages. Thus, different requirements can be considered according to each stage, namely (1) Discovery, (2) Fulfilment, (3) Assurance, and (4) Decommissioning.

3.1 Discovery

During the discovery phase, the 5GT-MTP exposes the underlying infrastructure, following the appropriate abstraction levels, to the 5GT-SO. The following requirements are identified:

TABLE 1: REQUIREMENTS ON THE DISCOVERY PHASE

ID	Requirement	F/NF
ReqMTP.Di.01	The 5GT-MTP shall store a catalog of NFVI-PoPs available within the 5GT-MTP's administrative domain, and related resources (computing, storage, networking) in addition to available PNFs/VNFs.	F
ReqMTP.Di.02	The 5GT-MTP must provide the means to expose available resources with appropriate abstraction levels to 5GT-SO.	F
ReqMTP.Di.03	The 5GT-MTP shall provide the means to expose the catalog of PNFs/VNFs to the 5GT-SO	F
ReqMTP.Di.04	The 5GT-MTP shall keep up-to-date the catalog of related NFVI components	F
ReqMTP.Di.06	The 5GT-MTP shall expose the current state of available PNFs and should expose the history of states of available PNFs.	F
ReqMTP.Di.07	The 5GT-MTP shall certify the credentials of entities accessing its NFVI catalog.	F
ReqMTP.Di.08	The 5GT-MTP shall allow to create several instances of the same VNF	F
ReqMTP.Di.09	The 5GT-MTP shall store a catalog containing the service connection points along with some metadata, such as the location, etc.	F
ReqMTP.Di.10	The 5GT-MTP shall support to create, retrieve, update, and delete VNFDs	F
ReqMTP.Di.11	The 5GT-MTP must provide the 5GT-SO with the means to send detailed resource allocation requests	F

ReqMTP.Di.12	The 5GT-MTP must provide the 5GT-SO with the means to configure VNF instances	F

3.2 Fulfilment

During the service fulfilment phase, the 5GT-SO orchestrates (namely, creates and instantiates) network services requested by 5GT-VS, using the infrastructure abstraction provided by the 5GT-MTP. From the 5GT-MTP perspective this involves: appropriate configuration of the VNFs, Vas and PNFs, and allocation of resources in available NFVI-PoPs.

The following requirements are identified:

TABLE 2: REQUIREMENTS ON THE FULFILMENT PHASE

ID	Requirement	F/NF
ReqMTP.Fu.01	Depending on the modality of the contracted service, the 5GT-MTP could be required to offer proper configuration and management interfaces to instantiated VNFs or requested PNFs.	F
ReqMTP.Fu.02	The 5GT-MTP shall allow VNF scaling (up/down/in/out)	F
ReqMTP.Fu.03	The 5GT-MTP shall allow resource scaling (up/down/in/out)	F
ReqMTP.Fu.04	The 5GT-MTP shall provide appropriate isolation and access guarantees to available PNFs	F
ReqMTP.Fu.06	The 5GT-MTP shall certify the credentials of entities accessing its NFVI.	F
ReqMTP.Fu.07	The 5GT-MTP shall maintain information regarding the mapping between NSD, VNFs/PNFs and allocated resources.	F

3.3 Assurance

The 5GT-MTP is responsible for guarantying the performance agreements made with the 5GT-SO for orchestrated VNFs and allocated resources in NFVI-PoPs and PNFs, including sufficient monitoring information. The following requirements are identified:

TABLE 3: REQUIREMENTS ON THE ASSURANCE PHASE

ID	Requirement	F/NF
ReqMTP.As.01	The 5GT-MTP must provide the 5GT-SO tools to monitor the QoS attained to instantiated VNFs, and allocated PNFs and related resources.	F
ReqMTP.As.02	The 5GT-MTP shall certify the credentials of entities accessing its NFVI monitoring information.	F
ReqMTP.As.03	The 5GT-SO should provide isolation and performance guarantees among tenants sharing PNFs	NF
ReqMTP.As.04	The 5GT-MTP should be fault-tolerant and report failure	F

	events upstream to the 5GT-SO should the 5GT-MTP not be able to solve issues.	
ReqMTP.As.05	The 5GT-MTP shall commit to assure performance indicators of exposed resources.	F

3.4 Decommissioning

Once a service is decommissioned, the 5GT-MTP shall properly release the used resources and terminate the required VNFs as a response to the 5GT-SO termination operations.

The following requirements are identified:

TABLE 4: REQUIREMENTS ON THE DECOMMISSIONING PHASE

ID	Requirement	F/NF
ReqMTP.De.01	The 5GT-MTP must be able to identify the resources allocated to a VNF upon a VNF termination procedure	F
ReqMTP.De.02	The 5GT-MTP must be able to identify the monitoring mechanisms to be de-activated as a result of a VNF termination or resource deallocation	F
ReqMTP.De.03	The 5GT-MTP must be able to notify the 5GT-SO about a VNFs or resources terminated	F
ReqMTP.De.04	The 5GT-MTP must restore the state of available PNFs when its allocation is terminated	F

4 5GT-MTP Internal architecture and interfaces

4.1 5GT-MTP architecture description and main functionalities

This section describes the system architecture and key building blocks specified as guidelines for the development of the 5GT-MTP. The architectural design of the 5GT-MTP aims at providing a set of functionalities and operations to support the Service Orchestrator (through the 5GT-SO-5GT-MTP interface) to achieve efficient utilization of different NFVI infrastructure domains, following a NFVI-as-a-Service model. The design of the 5GT-MTP architecture is leveraging the works carried out in the 5GPP Phase 1 projects, 5G-Crosshaul in particular, and standard development organizations such as ETSI NFV.

The 5GT-MTP is responsible for orchestration of virtual resources and the instantiation of VNFs or VAs to deploy the network services (requested by the 5GT-SO) over the infrastructure under its control, as well as managing the underlying physical mobile transport network, computing and storage infrastructure. The architecture of the 5GT-MTP is depicted in Figure 3. The main building block of the 5GT-MTP is the NFVO-RO SLPOC that acts as a single point of contact towards the 5GT-SO providing the suitable abstract view and receiving the resource requests. Moreover, within the 5GT-MTP the NFVO-RO acts as resource orchestrator to select and configure the transport and radio resources compliant with the request from the 5GT-SO. More details are reported below.
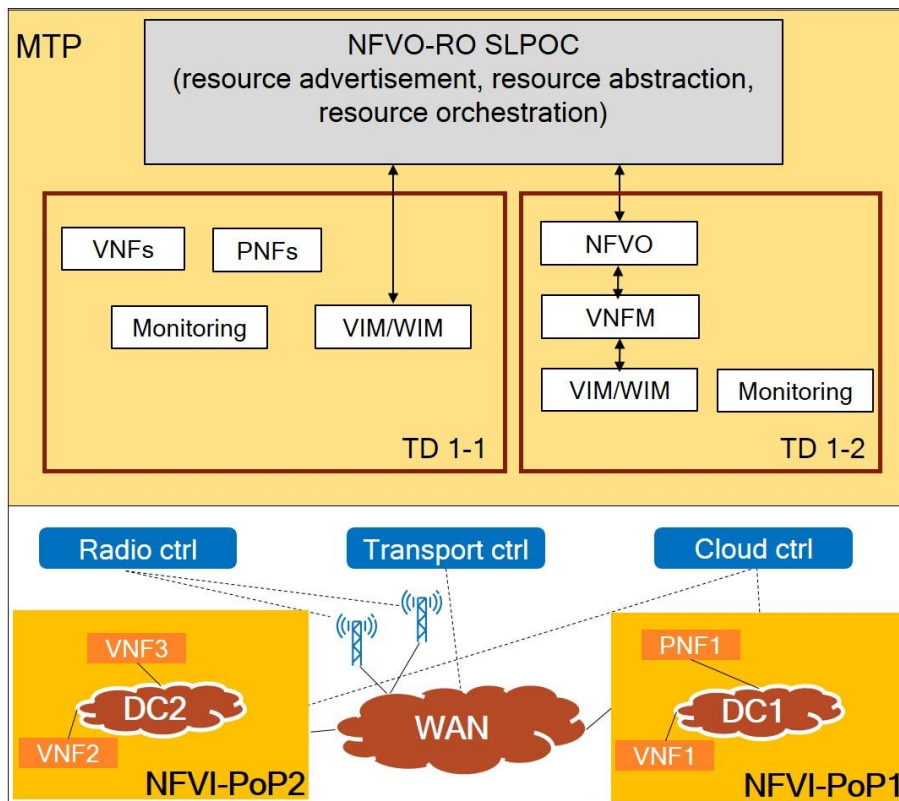
FIGURE 3: 5GT-MTP ARCHITECTURE

The computing and storage infrastructure may be deployed in central data centres as well as distributed, as in Multi-Access Edge Computing (MEC). Depending on the use

case, the 5GT-MTP may offer different levels of resources abstraction to the 5GT-SO via the 5GT-MTP resource abstraction component, which in turn forwards the 5GT-SO requests to the right entity accordingly (as single point of contact): VIM/WIM, VNFM or PNF, or NFVO. The monitoring block is responsible for collecting data from the different domains (transport, radio and cloud), monitoring the physical infrastructure and providing the needed monitoring information to the 5GT-SO.

The design of the 5GT-MTP architecture is based on the system architectures defined within the H2020 5G-Crosshaul project, which leveraged the standard and reference specifications of the SDN and NFV architectures. Specifically, on the data plane, the 5G-Crosshaul architecture includes two types of nodes: the Crosshaul Forwarding Elements (XFEs) responsible for forwarding data traffic, and the Crosshaul Processing Units (XPUs), which are in charge of computing operations. The XFEs can also cope with different link and physical-layer technologies thanks to the introduction of an innovative common framing to transport both backhaul and fronthaul traffic. XPUs instead can host VNFs and support C-RAN related operations. The main component of the 5G-Crosshaul control plane is the Crosshaul Control Infrastructure (XCI), which integrates the SDN control in the ETSI/NFV MANO architecture. The XCI also provides an abstracted view of the available resources, states and functions through the Northbound Interface (NBI). The Southbound Interface (SBI) connects the XCI to the data plane nodes and allows the execution of control and management functions on the hardware elements. Within the XCI structure there is the controller layer, composed of the network, computing, and storage controllers, enabling the allocation and configuration of the different resources composing the NFVI. The 5G-TRANSFORMER project extends the 5G-Crosshaul transport solution with MEC and dynamic creation of slices and placement of VNFs to take into account the needs of vertical industries. Figure 4 presents the 5GT-MTP TD1-1 and TD1.2 mapping with the ETSI NFV MANO architecture, highlighting three architectural alternatives:

- Case 1: the 5GT-MTP exposes virtual resources and the possibility to instantiate entire VNFs through the VNFM;
- Case 2: the 5GT-MTP exposes PNFs that can be configured but not instantiated (e.g. a physical BTS). At the VIM/WIM level the 5GT-MTP only instantiates virtual resources related to networking;
- Case 3: the 5GT-MTP abstracts an entire network service to the 5GT-SO and it takes care internally about how to orchestrate it, through the NFVO - VNFM - VIM/WIM stack.

It is worth noting that case 1 and case 2 correspond to the 5GT-MTP TD1-1 while case 3 corresponds to the 5GT-MTP TD1.2.
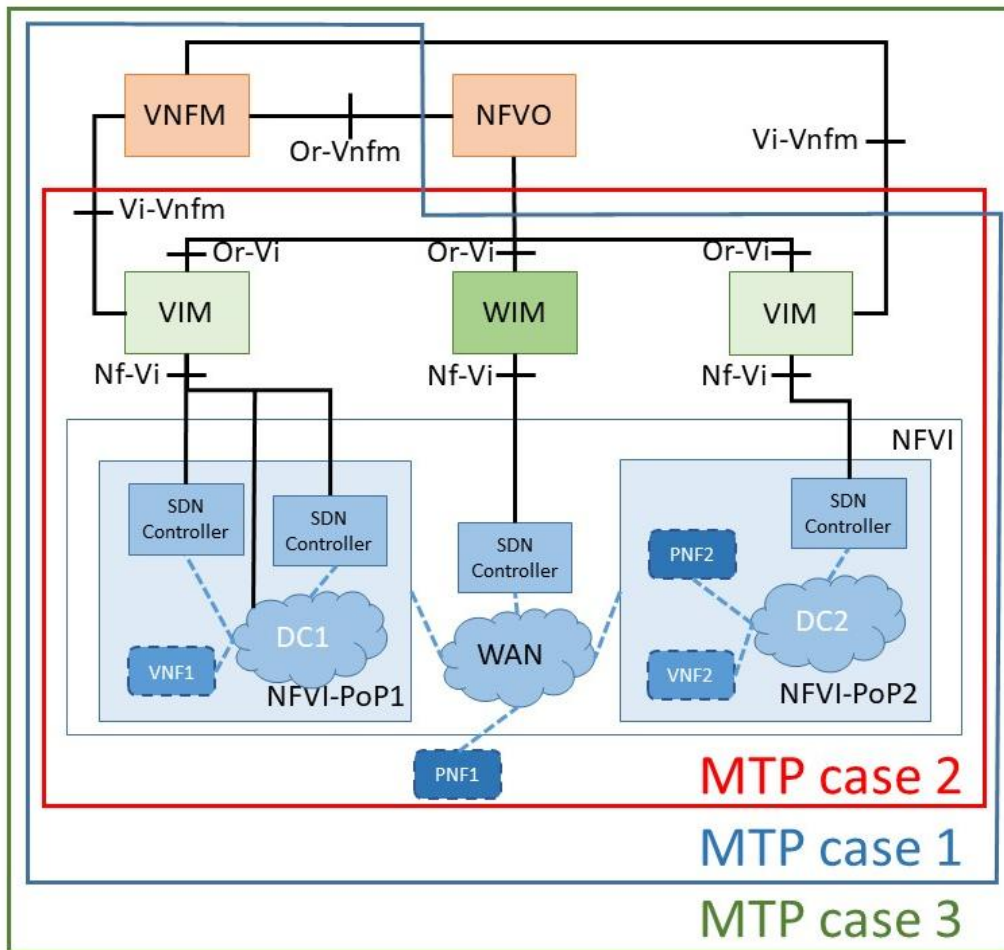
FIGURE 4: 5GT-MTP MAPPING WITH ETSI NFV MANO

Independently from the kind of service exposed by the 5GT-MTP to the 5GT-SO, as shown in Figure 3, the 5GT-MTP should contain the following components:

Virtual Infrastructure Manager (VIMs)

VIMs are in charge of managing storage, networking and computational resources in its respective NFVI-PoP administrative domain. The VIM is typically handled by a cloud platform, like e.g. OpenStack. In addition, each NFVI-PoP/administrative domain under the VIM's responsibility may include one or more SDN Controllers (e.g. OpenDaylight) in charge of establishing the transport connectivity between VNFs deployed within an NFVI-PoP. In case of multi-layer or multi-technology network infrastructures, SDN Controllers can also be deployed in a hierarchical model to handle the heterogeneity of the technological domains through dedicated child controllers.

WAN Infrastructure Manager (WIMs)

WIMs are in charge of providing inter-domain links, which will be translated into configurations of the transport network between NFVI-PoPs gateways through the proper SDN Controller.

Network Function Virtualization Infrastructure (NFVI)

NFVI provides all the hardware (e.g. compute, storage and networking) and software (e.g. hypervisor) components that constitute the infrastructure where VNFs are

---

deployed. Eventually, also sharing PNFs among different NFV-NSs can be taken into consideration for the virtualization infrastructure.

A VIM or a WIM can interface with the underlying SDN Controllers to request virtual connectivity services through the Nf-Vi reference point or establish directly the connectivity services by configuring the network nodes. In the latter case, SDN Controllers become part of the VIM itself, controlling directly virtual entities such as virtual switches or network functions within the related NFVI PoP. This kind of hierarchy in management and orchestration of heterogeneous resources provided by the NFVI brings the benefit of different layers of abstraction, where, from the bottom to the upper layer of the 5GT-MTP inner architecture, each component provides the proper NBI to request services. With the aim of offering NFV MANO services across multiple administrative domains, the NFVI pool of resources can be provided as a service. In the NFVIaaS paradigm, we can identify the consumer as a service provider which wants to run VNF instances inside an NFVI provided as a service by a different administrative entity: the NFVIaaS provider. This means that the NFVIaaS consumer has the control of the VNF instances, but it does not control the underlying infrastructure. In particular, since the provider's NFVI is structured in several VIMs, the provider can offer the access to the service following two different types of interactions between the two administrative entities:

- Multiple Logical Point of Contact (MLPoC), where the consumer has the visibility of the different VIMs within the provider's administrative domain and communicates directly with each of them.
- Single Logical Point of Contact (SLPoC) (see Figure 5), where the VIMs are hidden to the consumer and the provider's administrative domain contains a SLPoC function in charge of acting as a single unified interface offered to the consumer.
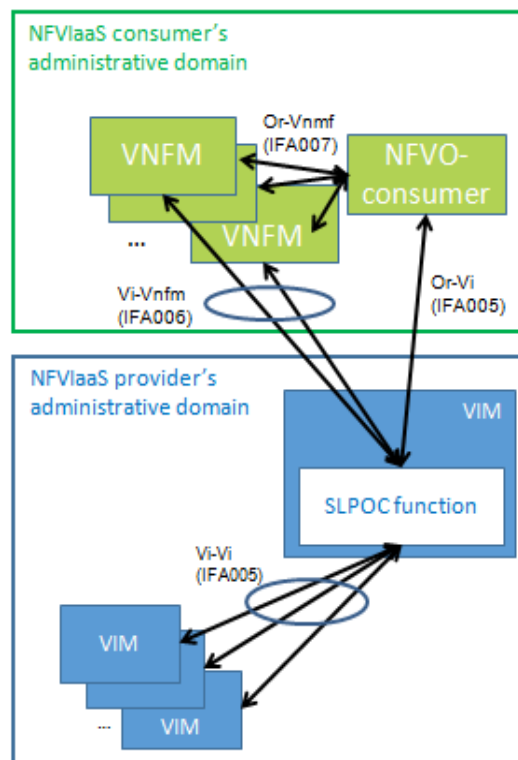
FIGURE 5: SLPOC FUNCTION

To enable the deployment of vertical use cases with mobile applications that require very low latency, the 5GT-MTP architecture should be extended to deal with the Multi-access Edge Computing (MEC) technology. In fact, the possibility of reducing the latency, by bringing IT and cloud computing capabilities near to the mobile access side, allows the deployment of use cases in different industry's branches, such as the automotive and the cloud robotics, where the "instantaneous" processing of the data is a key factor. An example of possible integration between MEC and NFV MANO architecture is provided in [31]. On the MEC side, we can identify the following components for the 5GT-MTP MEC extension.

Mobile Edge Platform (MEP) is a VNF deployed at the 5GT-MTP NFVI-PoP or NFVI edge. It offers services, such as Radio Network Information Service (RNIS), and location API for ME VNF applications. The latter are deployed in the same NFVI-PoP. The MEC applications use the MEC service to adapt the application to user context or run low-latency applications at the edge.

Mobile Edge Platform Manager – NFV (MEPM-V) corresponds to the MEP Element Manager. It is in charge of managing the application rules and requirements. The lifecycle management in the context of 5GT-MTP is delegated to the VNFM-MEC.

VNFM-MEC is in charge of Life Cycle Management of the MEC application VNF as well as the Mobile Edge Platform. It is connected to the 5GT-MTP NFVO via the well-defined Or-Vnfm interface, while it uses the Ve-Vnfm-em and Ve-Vnfm-vnf interfaces to communicate with the MEPM and MEC application VNF, respectively. At the 5GT-MTP level, the VNFM-MEC communicates with the VIM in order to manage the needed resources for the deployment of the MEC Apps, where the VIM uses the Nf-Vi interface to manage the NFVI Edge resources, e.g. supporting containers.

4.2 5GT-MTP innovations

The 5GT-MTP, as the overall 5G-TRANSFORMER architecture, has been designed to be aligned with the ETSI NFV specifications. However, in some cases the ETSI specifications should be extended to support the goals of the project. This section aims to describe such extensions or innovations.

One innovation is the fact that the 5GT-MTP decouples the VIM from the NFVO and VNFM through a REST-API interface that covers both the Or-Vi and Vi-Vnfm interfaces defined in ETSI-NFV. This decoupling allows future developments where an orchestrator may interface with more than one 5GT-MTP and also a 5GT-MTP could accept requests from multiple service orchestrators. This decoupling also facilitates further independent development of 5GT-MTP and 5GT-SO.

The above-mentioned decoupling also facilitates developing an 5GT-MTP architecture where one 5GT-MTP can integrate several VIMs and WIMs from different technological domains and expose a unified view to the upper layers (the 5GT-SO in the 5G-TRANSFORMER project). The integration of several VIMs and WIMs allows a single VNFM and NFVO to control several technological domains.

In order to allow the integration of several VIMs and WIMs in one 5GT-MTP, the 5GT-MTP includes an abstraction layer, which in turn is able to provide different levels of abstraction at both cloud computing and networking levels. Depending on the level of details exposed to the upper layer, the 5GT-MTP may take autonomous decisions

about resource orchestration (also considering radio network related constraints) or these decisions may be taken directly by the 5GT-SO.

As the second innovation, the integration of MEC in the 5G-TRANSFORMER project has also its reflection in the 5GT-MTP which is able to support the deployment of MEC applications and services providing the following features: (i) advertisement of MEC hosts, including their characteristics (locations, capabilities, network connectivity to RAN and WIMs); (ii) deployment of MEC applications and configuration of the related traffic steering; (iii) advertisement of MEC services running in each MEC hosts; (iv) support of network interfaces towards the RAN and the data plane in generals to enable MEC services like Radio Network Information Service (RNIS).

And lastly, the 5GT-MTP can also deploy, manage and provide a Connectivity Service, including the combination of network functions and connectivity from the RAN up to the vEPC. This kind of service is offered as an NFVI resource to the upper layer (i.e. the 5GT-SO) and is managed autonomously by the 5GT-MTP itself. This means that the 5GT-MTP is able to select, deploy and configure the most suitable RAN Split, as well as to decide the internal decomposition of such service, e.g. using physical or virtual network functions, and its dimensioning. This functionality is enabled by the 5GT-MTP TD1-2.

4.3 5GT-SO-5GT-MTP reference points

In the 5GT-MTP two set of interfaces (i.e., reference points) are defined: an external Northbound interface (NBI) between 5GT-MTP and 5GT-SO and an internal Southbound Interface (SBI) between 5GT-MTP VIM/WIN and NFVI.

4.3.1 5GT-MTP NBI/5GT-SO SBI

The 5GT-MTP northbound interface (NBI) addresses the interworking between the 5GT-SO and the 5GT-MTP building blocks of the 5G-TRANSFORMER architecture. The 5GT-MTP NBI coincides with the 5GT-SO SBI as defined in D4.1 [4]. Thus, the description of the 5GT-SO SBI is reported here for ease of consultation. It is worth mentioning that 5GT-SO and 5GT-MTP may follow a 1: N relationship. That is, a single 5GT-SO may interact via multiple SBI instances towards N 5GT-MTPs which handle the configuration and programmability of a number of domains including heterogeneous virtualized resources for compute, storage and networking. In the following we are also assuming that a 5GT-MTP is managed by a single 5GT-SO. Besides managing the utilization (i.e., de/allocation) of the virtualized resources, the 5GT-SO SBI/5GT-MTP NBI also encompasses the required functionalities for deploying (updating and terminating) demanded VNFs by a given NFV-NS. In the 5G-TRANFORMER project all these operations are supported by the so-called So-Mtp interface.
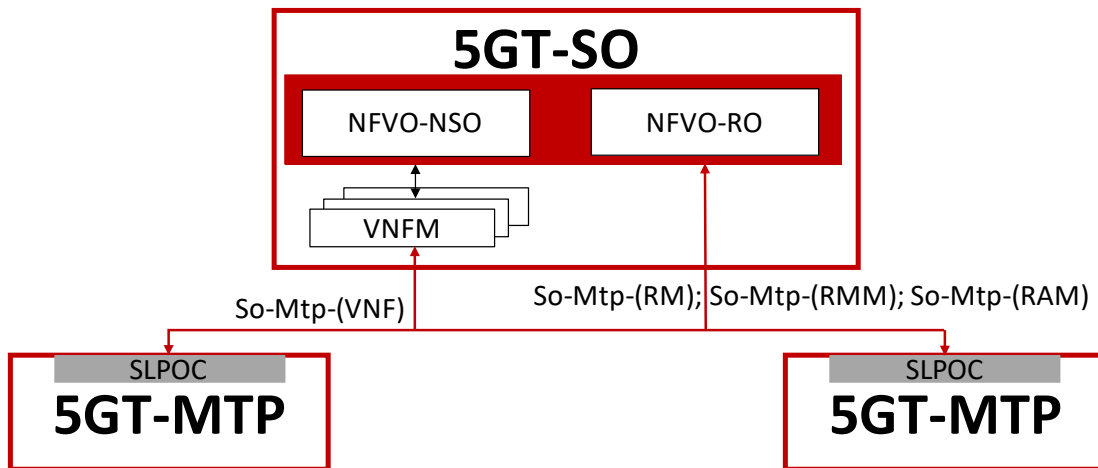
FIGURE 6: REFERENCE POINTS FOR 5GT-SO SBI (I.E., SO-MTP INTERFACE)

Figure 6 illustrates the targeted 5GT-SO SBI/5GT-MTP NBI and its key reference points. Similar to the 5GT-SO NBI, the 5GT-SO SBI/5GT-MTP NBI is mostly based on a set of standard documents being produced within the ETSI NFV framework, namely ETSI GS NFV-IFA 005 [24], ETSI GS NFV-IFA 006 [44] and ETSI GS NFV-IFA 008 [45]. In a nutshell, the 5GT-SO SBI/5GT-MTP shall provide the operations and functions, supported by a well-defined set of messages and workflows, for: (i), providing abstracted information (e.g., capacities, availability, connectivity, etc.) of the virtualized resources managed by each 5GT-MTP; (ii) managing (i.e., instantiation, reservation, allocation, scaling up/down and release) of the virtualized resources required to support an NFV-NS; (iii) enabling the fault management and performance monitoring aiming at recovering interrupted services or ensuring the targeted SLAs demanded by each NFV-NS; and (iv) supporting the lifecycle management (i.e., creation, configuration, modification and termination) along with related performance and fault management of the VNFs instantiated over the virtualized (compute and storage) resources. The SO-MTP interface enables communicating the specific entities of the 5GT-SO (VNFM and NFVO-RO) with a single logical point of contact (SLPOC) at each 5GT-MTP entity. Accordingly, four reference points for the 5GT-SO SBI are conceived:

- **So-Mtp(-RAM).** It provides the Resource Advertisement Management functions. That is, it allows feeding the 5GT-SO's NFVI repository with information regarding the virtualized resources that will accommodate requested NFV-NSs. Such information can be delivered by using different levels of details/abstraction. Thus, the adopted abstraction, and vision of the resources, will notably impact on the 5GT-SO NFVO-RO algorithms used for the VNF placement and/or networking computation. The mechanism used by the 5GT-MTP(s) to update the 5GT-SO's NFVI could be achieved also via different mechanisms such as immediate update when a change in any (abstracted) virtualized resource occurs (e.g., allocation or reservation), upon an explicitly demand sent by the 5GT-SO, or even applying predefined periodic updates.

- **So-Mtp(-RM).** This encompasses the Resource Management operations over the virtualized resources. Basically, it contains the set of operations used for reserving, allocating, updating (in terms of scaling up or down) and terminating (i.e., release) the resources handled by each 5GT-MTP. In short, and according to the abstracted information managed by the 5GT-SO, the So-Mtp(-RM)

coordinates the involved 5GT-MTPs to manage the utilization of a selected set of virtualized resources. This entails the VNF placement and triggering the reservation / allocation of the network resources constituting the demanded NFV-NS.

- **So-Mtp(-RMM).** This provides the Resource Management and Monitoring operations. Basically, it provides the required interworking procedures including the primitives and parameters for supporting the 5GT-SO Monitoring Service capability. This entails a twofold functionality: i) fault management to recover / restore the interrupted NFV-NS by a failure (e.g., link failure, VNF host crashes, etc.); and ii) permanent performance monitoring crucial to ensure the demanded SLA for the existing NFV-NS. Obviously, the involved information needed for such Monitoring Service functions are also related to the adopted granularity and abstracted information (i.e., level of detail) within the 5GT-SO.

- **So-Mtp(-VNF).** This takes over the general VNF lifecycle management (e.g., scaling up/down a particular VNF instance, fixing VNF malfunctions, etc.) commanded by the 5GT-SO VNFM. Moreover, the VNF configuration (i.e., specifying targeted parameters defining the targeted VNF behaviour) is also supported over this reference point. Last but not least, the So-Mtp-VNF reference point supports performance and fault management functionalities to monitor the VNF operations and adopt/apply (if needed) necessary actions to revert/solve VNF failures and performance degradations.

As anticipated above, the implementation of the 5GT-SO SBI/5GT-MTP NBI operations and, particularly, those driven into the four reference points, leverage the procedures (i.e., interworking between entities, messages and basic contents) described in ETSI GS NFV-IFA 005, 006 and 008 [24], [44], and [45], resp., as much as possible. Note that other functions/deviations required to be covered within the 5G-TRANFORMER framework but not supported by those standard documents could be eventually added. Nonetheless, focusing exclusively on the operations currently supported in [24], [44], and [45], the following operations are identified as essential for encompassing the implementation of the 5GT-SO SBI/5GT-MTP NBI reference points:

- For the **So-Mtp(-RAM),** a set of pairs of request/response messages is considered. This is divided into two main top sets/groups, namely, Virtualized Resources Information Management and Virtualized Resources Capacity Management. These two subsets of messages (see [24] and [44]) grant 5GT-SO with multiple functionalities such as subscription to specific (filtered) resource information, query of and update (for changes) resource information, specific resource capacity, etc. Specifically, the resource capacity can be provided with respect to its current status (i.e., available, allocated or reserved) as well as specifying the amount of resources. Moreover, the resource information can be retrieved for a particular 5GT-MTP governing a certain Resource Zone (e.g., geographic NFVI-PoP specifying the reachable endpoints). With respect to the amount of resources, this clearly depends not only on the type of virtualized resource but also on the abstraction policy. For the compute resources, it can be delivered the total amount of available or allocated virtual memory and virtual CPU; for storage, the information is related to the size of storage and type (e.g., volume, object) or the support of remote direct memory access; for networking resources, the provided attributes

regarding the link type (e.g., VLAN or GRE), the supported link QoS parameters (e.g., latency), the total link bandwidth, the IP addressing for a specific (sub-)network, the port types connected to specific network elements (e.g., Layer 1, 2 or 3), etc.

- For the **So-Mtp(-RM)**, ETSI GS NFV-IFA 005 [24] defines a set of request/response messages to allocate, query, update, migrate, and terminate virtualized resources are specified. Such defined sets are tailored to the operations to be made over a particular virtualized resource set. For instance, for compute resources the set named Virtualized Compute Management Interface is defined. This provides the specific allocation of compute resource over a particular Resource Zone, with a determined set of virtual CPUs and memory, as well as informing about the software image to be set on the Virtual Machine. For network resources, the set Virtualized Network Resource Management Interface takes over of all the operations to be made over the network resources. Non-exhaustively, this includes the allocation of selected bandwidth over a network entity (e.g., link) or the utilization of an entire data port. In addition, the operations to create Network Forwarding Paths (NFPs) to accommodate VNFFGD are supported. The request messages should include the list of virtual networks and ports forming the NFP. In addition, the set Virtualized Storage Resource Management Interface entails the set of operations (mapped to pair of messages) handling the storage resources. Similar to the compute and networking resources, these operations enable the selection of an amount of storage to be allocated over, e.g., a particular Resource Zone. Finally, ETSI GS NFV-IFA 005 also defines a top interface to reserve virtualized resources referred to as Virtualized Resource Reservation Interfaces. This interface is used to (pre-)book virtualized resources which eventually may be needed and used.

- For the **So-Mtp(-RMM)**, ETSI GS NFV-IFA 005 [24]also specifies the interfaces (messages and contents) supporting fault management and performance monitoring. Specifically, the Virtualized Resource Fault Management Interface defines the messages enabling the 5GT-SO to subscribe for notifications from 5GT-MTP about containers crashes, virtual network port errors or reserved resources unavailable or exhausted. To this end, such interface supports detailed alarms. The Virtualized Resources Performance Management Interface describes a set of messages used for collecting measurements within notifications that will feed the 5GT-SO's Monitoring Service. These messages include resource consumption, memory oversubscription, disk latency, etc. In general, the collection of such information is controlled by a Performance Monitoring (PM) job. The interface is oriented on handling the management of PM jobs (creation, subscribe, update, query, etc.). For a given PM job, it can be specified the object to be monitored (e.g., CPU power consumption in VM), the performance metric, the frequency for capturing the measurements, threshold to send notifications, etc.

- For the **So-Mtp(-VNF)**, ETSI GS NFV-IFA 008 [44] describes the messages and contents supporting the operations for the creation / configuration / termination, scaling (up / down), monitoring and fault management of the VNFs being deployed in a specific NFVI-PoP and handled by the SO-MTP. In this context, [44] firstly addresses the necessary set of messages (as request/respond pairs)

used for both initially configuring and modifying (e.g., deleting) a VNF (or Component). For the sake of completeness, this specific interworking is triggered by the 5GT-SO VNFM. In general, the messages providing a VNF operation must carry a unique identifier to unambiguously determine over which particular VNF (or Component) the action will be conducted. Moreover, configuration data or parameters are also included specifying the amount of required memory, CPU capacity, storage size, connection points (address and ports), software image of the VNF container, etc. That is, the set of parameters providing the description of the targeted VNF (Component) is referred to as VNF Descriptor (VNFD). In general, VNFDs are on-boarded in the so-called VNF package and they are assigned by an identifier which allows determining which descriptor is followed by a VNF instance being created. Exhaustive details about the VNFDs and the configurable parameters are provided in ETSI GS NFV-IFA 011. Another interworking operation supported by **So-Mtp(-VNF)** is the VNFM Indication actions. These indicators are used to notify the 5GT-SO VNFM about a VNF behaviour which can be eventually used by the VNFM to trigger auto-scaling operations. Finally, the performance and fault management interworking enable creating PM jobs to impose the generation of notifications sending specific VNF parameters status (e.g. CPU) which are then gathered and processed. To this end, the interface entails the creation of thresholds to manage the notification message creation or the periodicity when such notifications are composed and sent. Last but not least, for the fault management purposes, the 5GT-SO VNFM is able to indicate the subscription demanding for specific alarms generated by the VNFs when, for instance, to react when a fault occurs.

Although the considered ETSI GS NFV-IFA 005 [24], ETSI GS NFV-IFA 006 [44] and ETSI GS NFV-IFA 008 [45] documents cover most of the interworking operations to be deployed in the targeted 5GT-SO SBI/5GT-MTP NBI, as mentioned above, some specific capabilities required in 5G-TRANSFORMER may not be entirely supported, and therefore appropriate extensions will be required. In this regard, the support of multiple 5GT-MTPs to a single 5GT-SO is a notable objective within 5G-TRANSFORMER. Another relevant capability regards how to manage a pool of PNFs shared by multiple 5GT-MTPs. In other words, the goal is to investigate how the 5GT-SO SBI/5GT-MTP NBI can allocate, update, terminate, etc. PNFs resources that are not bounded to a single 5GT-MTP. We note that scenarios such as these need to be carefully studied with the focus on how the 5GT-SO SBI/5GT-MTP NBI is impacted and whether ETSI GS NFV-IFA 005 can support it. These and other open issues that may arise will be explored recurrently during the 5GT-SO SBI/5GT-MTP NBI interface design, implementation and validation.

4.3.2 5GT-MTP abstraction towards the 5GT-SO

An important and exclusive function of the 5GT-MTP is to decide the abstraction (i.e., the level of details) with which the resources are exposed to the 5GT-SO, the utilized information model, and data model. Possible solutions are summarized in the following sections.

In 5GT-TRANSFORMER, the 5GT-MTP includes both abstraction and virtualization functions. Resources are abstracted, in the sense that they are exposed in a simplified view to the 5GT-SO. Moreover, they are virtualized, that is, they are offered to the 5GT-

SO as they were a subset of resources dedicated to a specific VNF-NS. For sake of simplicity in the rest of the chapter the term "abstraction" is used to refer to both abstracted and virtualized resources exposed from 5GT-MTP to 5GT-SO.

Resource abstraction is a key element for the interworking of the 5GT-MTP and the 5GT-SO that has to meet the following needs: i) keep the independency between the technology deployed in transport and radio and the information model to describe its resources so that the same information model may be maintained even if the technology changes; ii) simplify the tasks of transport, radio and exposition view to 5GT-SO with a clear separation of role and responsibility, to facilitate fault locations; iii) decouple the radio and transport solutions that can evolve to different releases independently; iv) associate radio and transport to different providers that can be combined to each other in N:M relationship, where N and M are positive integers. For example, there are situations where the same transport provider shares the transport infrastructure to more radio providers, or the case where a radio provider makes use of more transport providers. In this approach, the transport layer works on the physical infrastructure resources and provides a suitable abstract view to the radio. Such abstract view can be organized in different views according several criteria. In turn, the radio layer works on the abstract view exposed by the transport to manage the radio resources and provide the connectivity services to the 5GT-SO. Again, also at the radio level a view has to be provided according different criteria. In case of Cloud Radio Access Network (C-RAN), the radio functions are virtualized and provided as generic processing functions deployed on data-centres and servers. In this case, the 5GT-MTP has to manage the C-RAN and coordinate with the transport to provide the connectivity required by the 5GT-SO.

To allow a suitable deployment of resources, the 5GT-MTP has to provide the 5GT-SO with an abstract view of the available resources, providing an adequate level of detail. For this reason, one of the 5G-TRANSFORMER objectives is to provide new abstraction models for infrastructure resources. In the context of 5G-TRANSFORMER, the 5GT-MTP resources consists of the RAN and core network, transport network, MEC infrastructure, compute, and storage resources. Thus, the 5GT-MTP will provide a scalable and efficient abstraction that takes into account all these aspects. In particular, to allow a correct selection of the resources for a specific service, the 5GT-MTP will expose (with the suitable level of abstraction) information about:

- availability of NFVI-PoP resources, identifying also the geographical location of the servers for a correct placement of the V(N)Fs,
- type and characteristic of available connectivity.

As far as the connectivity is concerned, two possible approaches may be followed. In the first, the 5GT-MTP exposes the connectivity hiding the complexity of the radio and transport interaction. The 5G-MTP provides as a logical link characterized with performance parameters (e.g., latency, bandwidth). The network functions to implement this logical connectivity (e.g., evolved NodeB, Evolved Packet Core --- EPC),) as either physical or virtual network functions, and the type of functional split to be adopted, are entirely decided and managed by the 5GT-MTP that will orchestrate them not only on the bases of the service requirements but also taking into account the constraints of the underlying transport network. Using the second approach, the 5GT-MTP exposes only networking and computing resources, optionally combined with specific PNFs, allowing the 5GT-SO to decide how to compose the desired mobile

communication service selecting the most suitable combination of VNFs, PNFs and their inter-connections. To better understand the main functionalities and requirements of the 5GT-MTP and highlight the importance of the resource abstraction, some examples are provided in the following.

Figure 7 shows an example of physical infrastructure managed by the 5GT-MTP composed of some eNodeBs and four NFVI_PoPs (or data centers) connected to each other through a transport network. The NFVI_PoP_telco data centers provide compute and storage resources for network functions, while NFVI_PoP_cloud data centers provide compute and storage resources for the Vertical Application (VA). For sake of simplicity, the access network (AN) for mobile is simplified (tree topology) and based on Long Term Evolution (LTE) but extensions to more complex AN are straightforward.
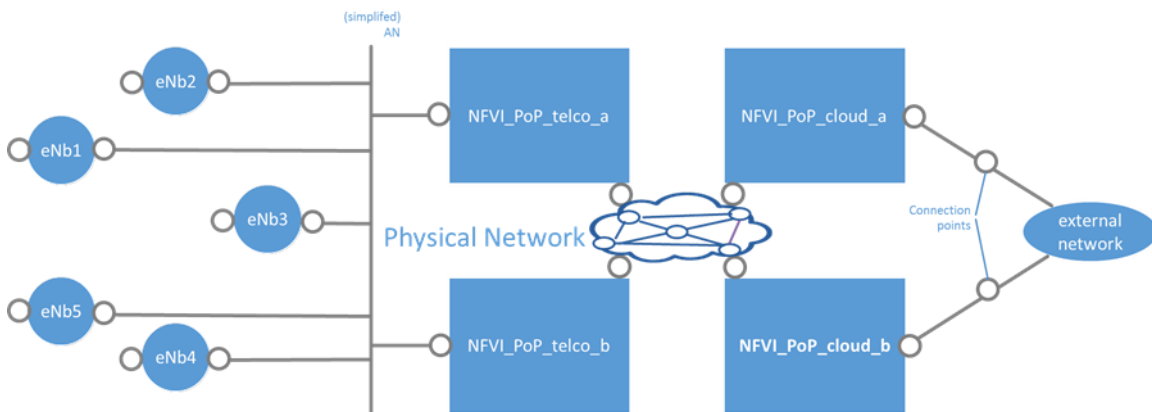
FIGURE 7: EXAMPLE OF PHYSICAL INFRASTRUCTURE

Figure 8 describes a service (i.e., NFV-NS) deployed on the physical infrastructure described above. EPC functions are deployed in NFVI_PoP_telco_a, while an application server (AS) is deployed in NFVI_PoP_cloud_b. The choice of the particular NFVI_PoPs to use for deploying the EPC functions and AS may be based on the VM images available at the different data centres and/or geographical position of the data centres. An alternative deployment is shown in Figure 9, where control and data forwarding planes related to the S-GW and P-GW are split. In this case the intelligence and decision making are centralized in the P-GW and S-GW controllers (P-GW-C and S-GW-C).

Considering the scenario described in three different alternatives of 5GT-MTP abstraction model with different level of details are considered.
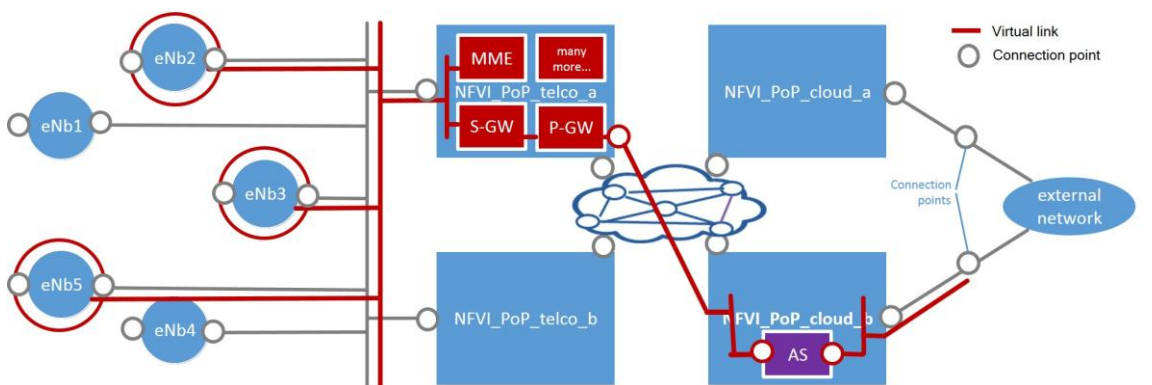
FIGURE 8: VERTICAL SERVICE COMPONENTS
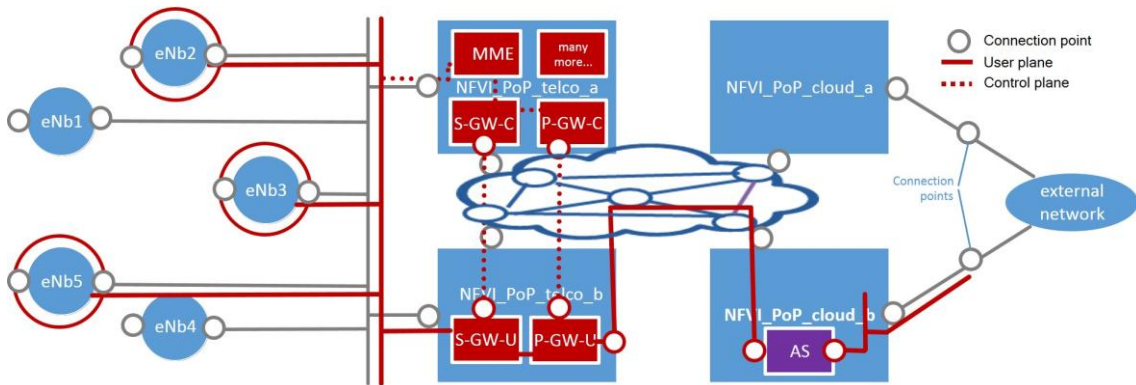
FIGURE 9: VERTICAL SERVICE COMPONENTS IN THE CASE OF EPC SPLIT

Alternative 1

5GT-MTP exposes all physical resources (mobile, transport, storage, and compute) to the 5GT-SO (see Figure 10). That is to say, the 5GT-SO has a full view of all physical resources and takes decisions on how to orchestrate the NFV-NS based on a full view of physical resources. Clearly, this alternative has severe scalability and resources ownership issues (as data centres may belong to different providers).
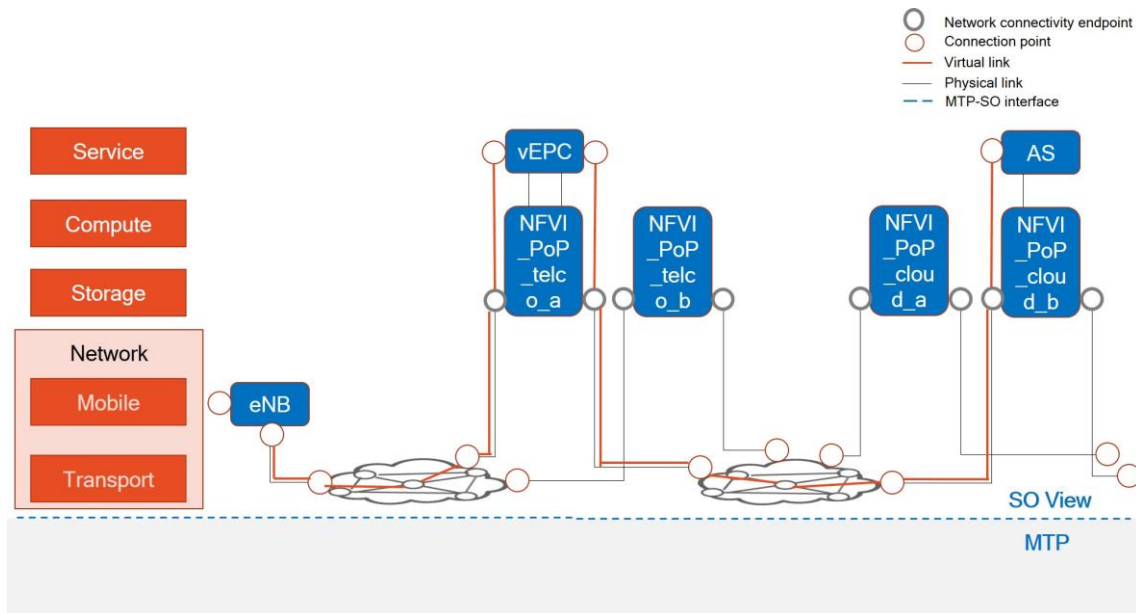
FIGURE 10: 5GT-SO VIEW OF FINAL SERVICE FOR ABSTRACTION ALTERNATIVE 1

Alternative 2

In this abstraction model the MTP exposes cloud resources availability and an abstract view of transport and mobile resources. The 5GT-SO is not fully aware of the physical network topology. As shown in Figure 11, the 5GT-MTP presents an abstract network topology of the transport network, which is composed by a set of logical links interconnecting some service access points (connection point identifying a specific eNB) with telco data centers (NVFI_PoP_telco_a and NVFI_PoP_telco_b) as well as a set of logical links interconnecting telco data centers with cloud data centers (NVFI_PoP_telco_a and NVFI_PoP_telco_b). The 5GT-SO takes decision on which data centers to deploy the network functions (i.e., vEPC) and the AS, selects the access point of the service and then decides how to interconnect the access point,

network functions and AS together in order to compose the final NFV-NS using the logical links reported in the abstract view of resources provided by the 5GT-MTP. As shown in Figure 12, the final service is seen from 5GT-SO as a logical link connecting the service access point to the left-side network connectivity endpoint of NVFI_PoP_telco_a (where the vEPC is deployed) as well as a logical link from the right-side network connectivity endpoint of NVFI_PoP_telco_a to the left-side network connectivity endpoint of NVFI_PoP_cloud_b (where the AS is deployed).
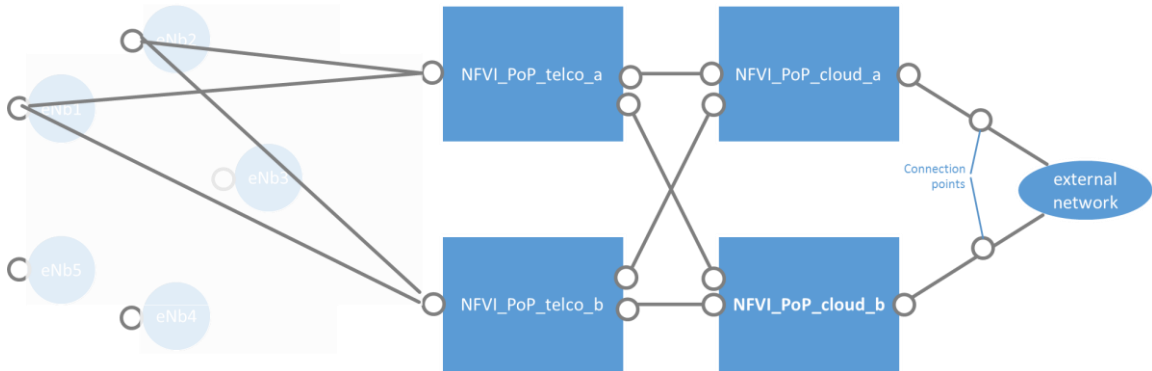
FIGURE 11: 5GT-SO VIEW FOR ABSTRACTION ALTERNATIVE 2
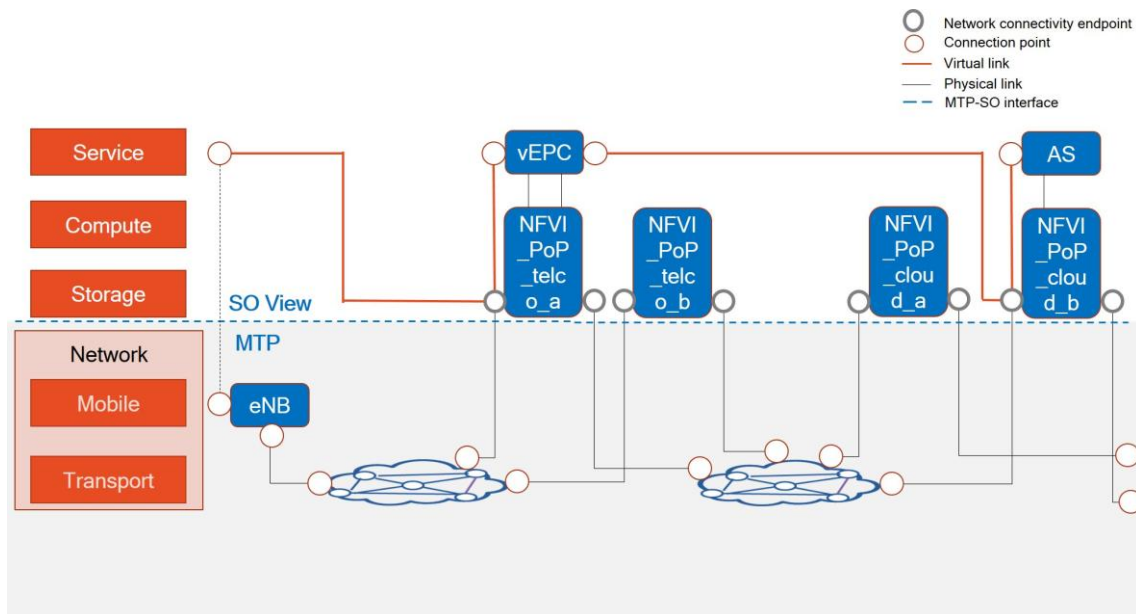
FIGURE 12: 5GT-SO VIEW OF FINAL SERVICE FOR ABSTRACTION ALTERNATIVE 2

Alternative 3

In this abstraction model the 5GT-MTP exposes cloud data center resources availability, an abstract view of transport and mobile resources, and completely hides telco data center resources. The 5GT-SO is not fully aware of the physical network topology. As shown in Figure 13, the 5GT-MTP presents an abstract network topology of the transport network, which is composed by a set of logical links interconnecting some service access points (connection point identifying a specific eNB) with cloud data centers (NVFI_PoP_cloud_a and NVFI_PoP_cloud_b). In this alternative, the 5GT-SO takes decision on which data center to deploy the AS, chooses the access point of the service and then decides how to interconnect the access point and the AS in order to compose the final NFV-NS using the logical links shown in the abstract view

of resources provided by the 5GT-MTP. As shown in Figure 14, the 5GT-SO sees the final service as composed by a logical link from the service access point (connection point above the eNB) to the network connectivity endpoint of NVFI_PoP_cloud_b (where the AS is deployed).
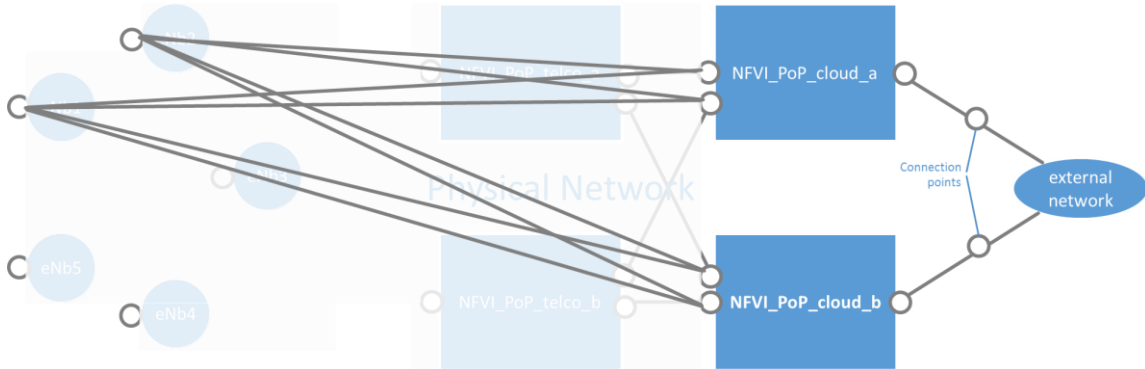
FIGURE 13: 5GT-SO VIEW FOR ABSTRACTION ALTERNATIVE 3
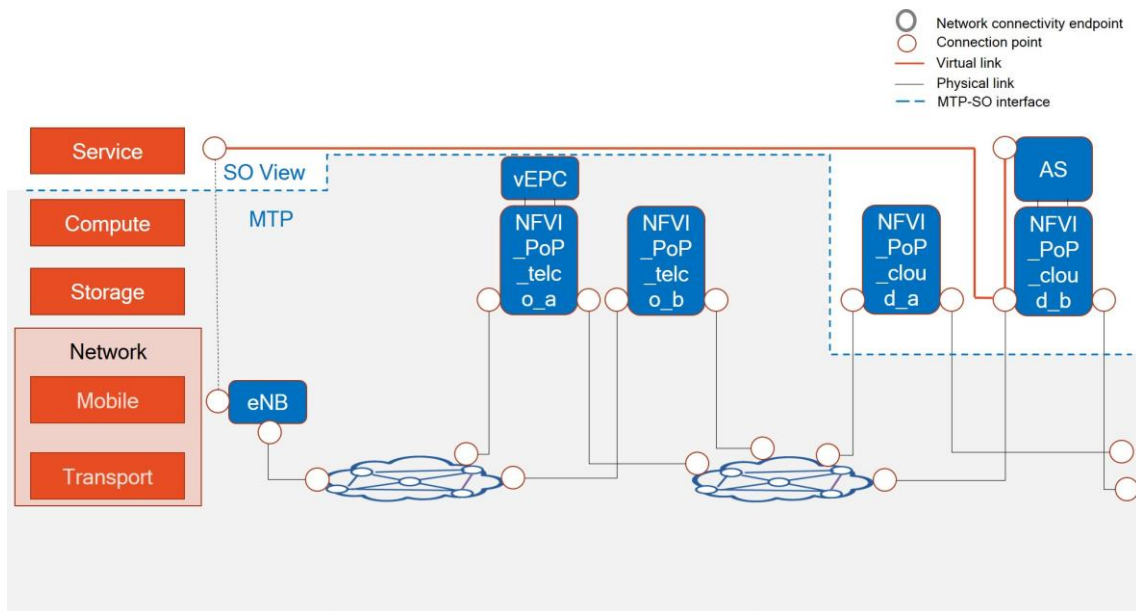
FIGURE 14: 5GT-SO VIEW OF FINAL SERVICE FOR ABSTRACTION ALTERNATIVE 3

4.3.3 Abstraction, Information model, and 5G-MTP SBI

This subsection focuses on the abstraction and modelling of the information exchanged between 5GT-MTP and 5GT-SO. With reference to IFA005 as described above, the following information models will account for maximum, available, and allocated resources. We assume that allocated resources coincide with reserved resources and we also assume that the knowledge of two among maximum, available, and reserved can give the knowledge of the third quantity. Abstraction of the following entities is required to enable 5GT-SO for orchestration:

- Logical Link
- Computation capabilities
- Storage capabilities

Logical link abstraction enables an exchange of information, between 5GT-MTP and 5GT-SO, related to logical link. This information exchanged between the two control

entities should be technology agnostic (thus should be independent from the fact that the logical link is, for example, an optical link or an MPLS link). However, the 5GT-MTP must keep track of the whole information describing the connectivity, also considering the technology that could make some parameter relevant (e.g., the central frequency for an optical link) in the 5G-MTP southbound interface (SBI). Then, such an information, when passed to the 5GT-SO, should be properly filtered, exchanging only the parameters required by the 5GT-SO (e.g., source and destination of the logical link, latency, and so on). The Annex (Table 13) shows, as an example, the information modelling related to a logical link in an optical network related to the southbound interface between 5G-MTP and the SDN controller, thus including all the parameters relevant for such a technology (e.g., transponders). However, the information required by the 5GT-SO is less, thus several parameters describing the logical link can be filtered when exchanged between 5GT-MTP and 5GT-SO, e.g. depending on the algorithm for orchestration deployed within the 5GT-SO, specific information could be required or not. At the moment, we can assume that the logical link parameters required by the 5GT-SO are the source and the destination nodes, a measure of the link capacity (*available rate*), and latency as shown in the Table 5.

TABLE 5: ASSUMED LOGICAL LINK PARAMETERS TO BE EXCHANGED WITH THE 5GT-SO

Identifier
Source node
Destination node
Available bit rate

Then, abstraction is required for computational resources. Based on [24], we define computational resources including CPU, memory (RAM), as summarized in Table 6.

Finally, abstraction is required for storage resources as summarized in Table 7.

TABLE 6: INFORMATION MODELLING TO DEFINE A COMPUTATIONAL RESOURCE

Identifier
Max CPU
Used CPU
Max RAM
Used RAM
IP address

TABLE 7: INFORMATION MODELLING TO DEFINE A STORAGE RESOURCE

Identifier
Max Memory
Available Memory
IP address

Once the list of parameters identifying a given resource is defined (e.g., a logical link) a data modelling is required to express such parameters and to enable the exchange of such information. As an example, YANG data modelling could be assumed as data modelling language and NETCONF as protocol to exchange this information. Indeed, in the last years, NETCONF is emerging as an SDN protocol [25][26] attracting interest from service providers and network operators since it presents several advantages: it operates on data encoded in XML, thus commonly used XML tools can be adopted to process NETCONF content; it is based on YANG data modelling [27] instead of bit encoding. YANG is a highly readable language enabling the description of data plane

devices in a vendor-neutral way [28] [29]. The interest on YANG is demonstrated by the active work of several consortiums and projects including the presence of operators, service providers, and vendors [30]-[32], which are developing YANG models describing network elements. According to IFA005, the query of information regarding consumable virtualized resources that can be provided by the VIM is a requirement. This could be accomplished with the NETCONF <get> message. Moreover, NETCONF supports a feature named *Subtree Filtering* [25] allowing an application to select specific parameters from the whole set of parameters describing a network entity. Thus, such a feature is particularly indicated to enable the 5GT-SO to request for a limited set of parameters from the whole set (e.g., the one in Table 13). Still according to IFA005, notifications to the consumer of changes to information regarding consumable virtualized resources that can be provided by the VIM is another requirement. In this case, the NETCONF <notification> message can be exploited. In summary, with reference to the interfaces/operations such as: "Allocate Resource", "Query Resource", "Update Resource", "Scale Resource", "Migrate Resource", "Operate Resource", "Release Resource", "Create Resource Reservation", "Query Resource Reservation", "Update Resource Reservation", "Release Resource Reservation", "Virtualized resources performance management", "Virtualized resources fault management" - the following NETCONF messages may support these operations:

- <get>: queries about resources (e.g., Query Resources, Query Resource Reservation)
- <edit-config>: allocation/reservation of resources (e.g., Allocate resource, Create Resource Reservation - we can assume "allocate" and "create resource reservation" to be the same -, Scale Resource)
- <notification>: alarms and monitoring information exchange (Virtualized resources performance management, and Virtualized resources fault management)
- <delete-config>: release of resources and tear down (Release Resource, Release Resource Reservation)

Figure 35 in the Annex shows the tree representation of a YANG data modelling of logical link assuming the information modelling for the southbound interface, thus the data modelling for the 5GT-SO will be a sub-set.

The Figure 15 shows the tree representation for a YANG data modelling of computational resources assuming the information modelling in Table 6. The CPU is expressed in MHz while the RAM in GB.

```
module: cpt
  +--rw computationalresources
     +--rw computationalresource [id]
        +--rw id          computationalresource-id-type
        +--rw max-cpu     cpu-type
        +--rw max-ram     ram-type
        +--rw used-cpu    cpu-type
        +--rw used-ram    ram-type
        +--rw ip?         inet:ip-address
```

FIGURE 15: YANG TREE REPRESENTATION OF COMPUTATIONAL RESOURCES

```
module: mem
   +--rw storageresources
      +--rw storageresource [id]
         +--rw id                 storageresource-id-type
         +--rw max-storage        memory-type
         +--rw available-storage  memory-type
         +--rw ip?                inet:ip-address
```

FIGURE 16: YANG TREE REPRESENTATION OF STORAGE RESOURCES

Figure 16 shows the tree representation for a YANG data modelling of storage resources assuming the information modelling in Table 7. The storage resource is expressed in GB.

5 5GT-MTP workflow descriptions

This section describes the internal workflows within the 5G-TRANSFORMER Mobile Transport and Computing Platform (5GT-MTP). The workflows capture the interaction among 5GT-MTP components during the following scenarios; instantiation of a non-nested network service, modification of a non-nested network service, termination of non-nested network service and monitoring of a non-nested network service. We present the operations executed by the 5GT-MTP after receiving the following requests from the 5G-TRANSFORMER Service Orchestrator (5GT-SO), namely; Resource Allocation Request, Vertical Service Update Request, Termination Resource Request and Create Performance Monitoring Job Request. It is assumed that the 5GT-MTP components include; VIM/WIM, SDN controllers, NFVI-PoPs, VNFs, PNF and PNFM. It is worth noting that under the 5G-TRANSFORMER architecture, the VNFMs are within the 5G-TRANSFORMER Service Orchestrator (5GT-SO). For the sake of clarity and ease of comprehension in the flow diagrams also part of 5GT-SO internal workflows are reported.
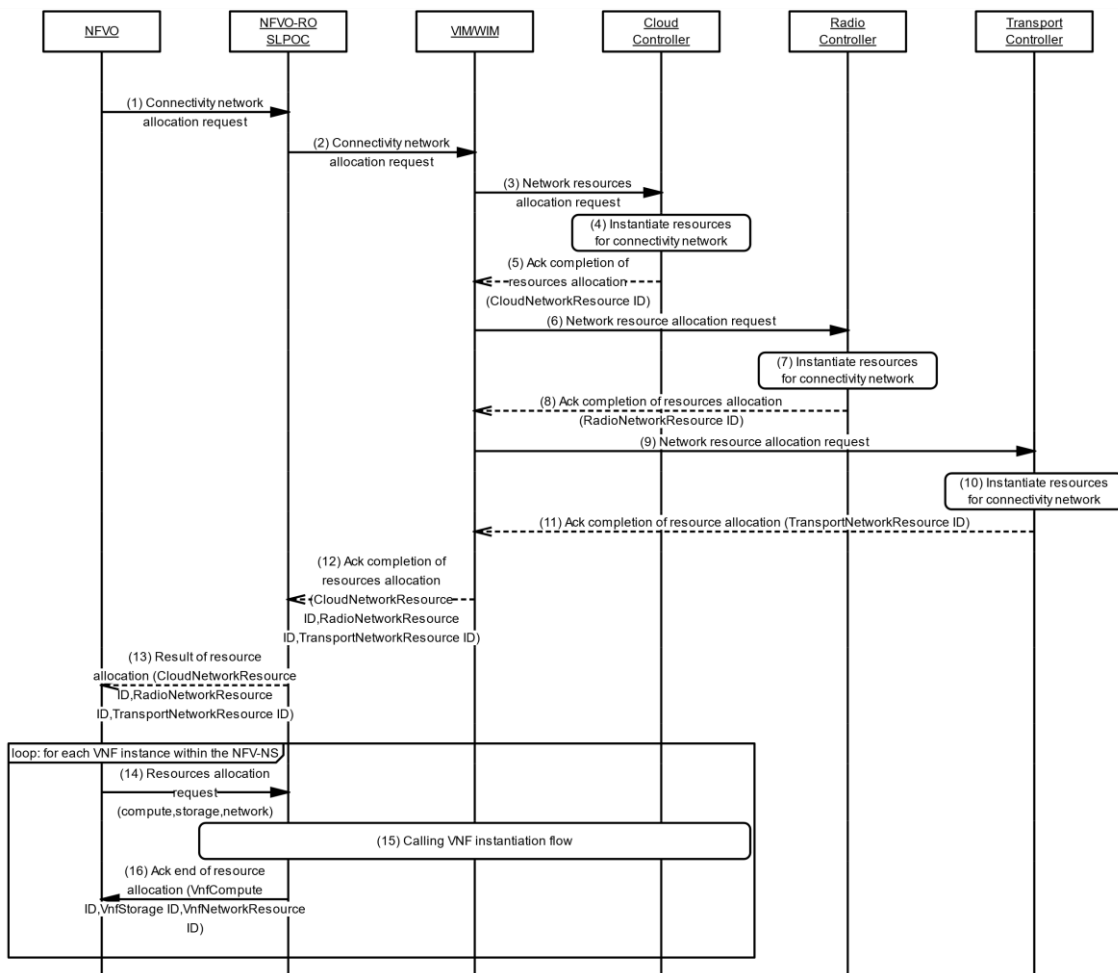
FIGURE 17: NON-NESTED NETWORK SERVICE INSTANTIATION FLOW

5.1 Instantiate a non-nested network service

Figure 17 presents the 5GT-MTP internal workflow for instantiating a non-nested NFV-NS. It describes the allocation of NFVI resources and instantiation of VNFs for a NFV-

NS instance. The flow is composed of two steps. The first step is to allocate the connectivity network needed for the NFV-NS. The second step is to instantiate VNFs.

The instantiation workflow is triggered by the Resource Allocation Request for the NFV-NS's connectivity network from the NFVO in the 5GT-SO to the 5GT-MTP- NFVO-RO SLPOC also referred to as the Network Functions Virtualization Orchestrator - Resource Orchestrator (NFVO-RO). The subsequent operations of the 5GT-MTP are listed below.

1. The NFVO sends a request to the 5GT-MTP- NFVO-RO SLPOC to allocate the connectivity network needed for the NFV-NS using the operation Allocate Resource from Virtualized Network Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20] for the different domains (Cloud, Transport and Radio).

2. 5GT-MTP- NFVO-RO SLPOC forwards the request to instantiate the network connectivity to the VIM/WIM.

3. VIM/WIM requests from Cloud Controller the allocation of network resources for the connectivity network using the operation Allocate Resource from Virtualized Network Resources Management Interface. The request includes the parameters reported in [20].

4. Cloud Controller instantiates the network resources for the connectivity network.

5. Cloud Controller acknowledges completion of network resource allocation to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (CloudNetworkResource ID) of allocated resources.

6. VIM/WIM requests from Radio Controller the instantiation of the network resources for the connectivity network using the operation Allocate Resource from Virtualized Network Resources Management Interface. The request includes the parameters reported in [20].

7. Radio Controller instantiates network resources for the connectivity network.

8. Radio Controller acknowledges completion of radio network resources allocation to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (RadioNetworkResource ID) of allocated resources.

9. VIM/WIM requests from Transport Controller the allocation of the network resources using the operation Allocate Resource from Virtualized Network Resources Management Interface [20]. The request includes the parameters reported in [20].

10. Transport Controller instantiates the network resources for the connectivity network.

11. Transport Controller acknowledges completion of transport network resources allocation to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (TransportNetworkResource ID) of allocated resources.

12. VIM/WIM acknowledges the completion of connectivity network allocation back to 5GT-MTP- NFVO-RO SLPOC.

13. 5GT-MTP- NFVO-RO SLPOC returns result of connectivity network allocation back to NFVO. The acknowledge message includes the parameters reported in [20] for the different domains (Cloud, Transport and Radio). Such parameters

include the identifiers (CloudNetworkResource ID, RadioNetworkResource ID and TransportNetworkResource ID) of allocated network resources.

14. NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC for allocation of resources (compute, storage and network) for each VNF to be instantiated for the NFV-NS using the operation Allocate Resource from Virtualized Network Resources Management Interface, Virtualized Compute Resources Management Interface and Virtualized Storage Resources Management Interface based on ETSI NFV IFA 005 [20].

15. 5GT-MTP- NFVO-RO SLPOC allocates resources (compute, storage and network) for the VNF to be instantiated by calling the VNF instantiation flow.

16. 5GT-MTP- NFVO-RO SLPOC returns result of resources (compute, storage and network) allocation back to NFVO. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of allocated resources.
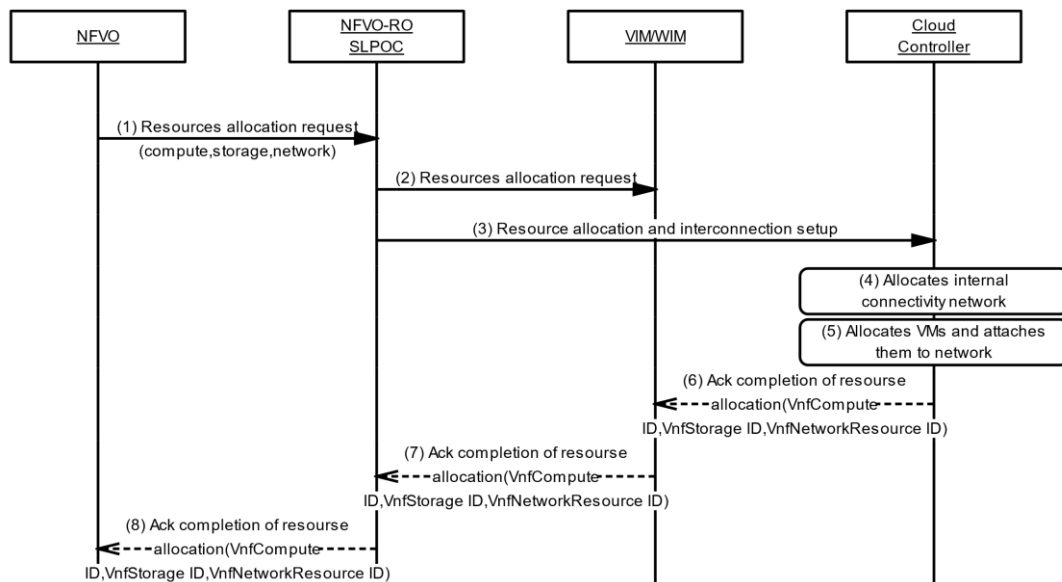
FIGURE 18: VNF INSTANTIATION FLOW

5.1.1 VNF instantiation

Figure 18 presents the 5GT-MTP workflow for instantiating a VNF. It describes the allocation of NFVI resources (storage, compute and network) needed for the different VMs composing the VNF and instantiation of the internal (network that is only used as internal to the VNF instance) connectivity network to interconnect VMs. The VNF instantiation workflow is triggered by the Resource Allocation Request for a new VNF instance from the NFVO to the 5GT-MTP- NFVO-RO SLPOC. The subsequent operations of the 5GT-MTP are listed below.

1. NFVO sends a requests to the 5GT-MTP- NFVO-RO SLPOC for allocation of resources (compute, storage and network) needed for the various VMs of the VNF instance using the operation Allocate Resource from Virtualized Network Resources Management Interface, Virtualized Compute Resources Management Interface and Virtualized Storage Resources Management Interface based on ETSI NFV IFA 005 [20].

2. 5GT-MTP- NFVO-RO SLPOC forwards the resource allocation request to the VIM/WIM.

3. VIM/WIM requests from Cloud Controller the allocation of resources (compute, storage, network) for the connectivity network.

4. Cloud Controller allocates the internal connectivity network.

5. Cloud Controller allocates the needed compute and storage resources, instantiates the VMs and attaches instantiated VMs to internal connectivity network.

6. Cloud Controller acknowledges the completion of resource allocation back to VIM/WIM. The acknowledge message includes the parameters reported in [20] The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of allocated resources.

7. VIM/WIM acknowledges the completion of resource allocation back to 5GT-MTP- NFVO-RO SLPOC.

8. 5GT-MTP- NFVO-RO SLPOC acknowledges the completion of resource allocation back to NFVO. The acknowledge message includes the parameters reported in [20]. The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of allocated resources.
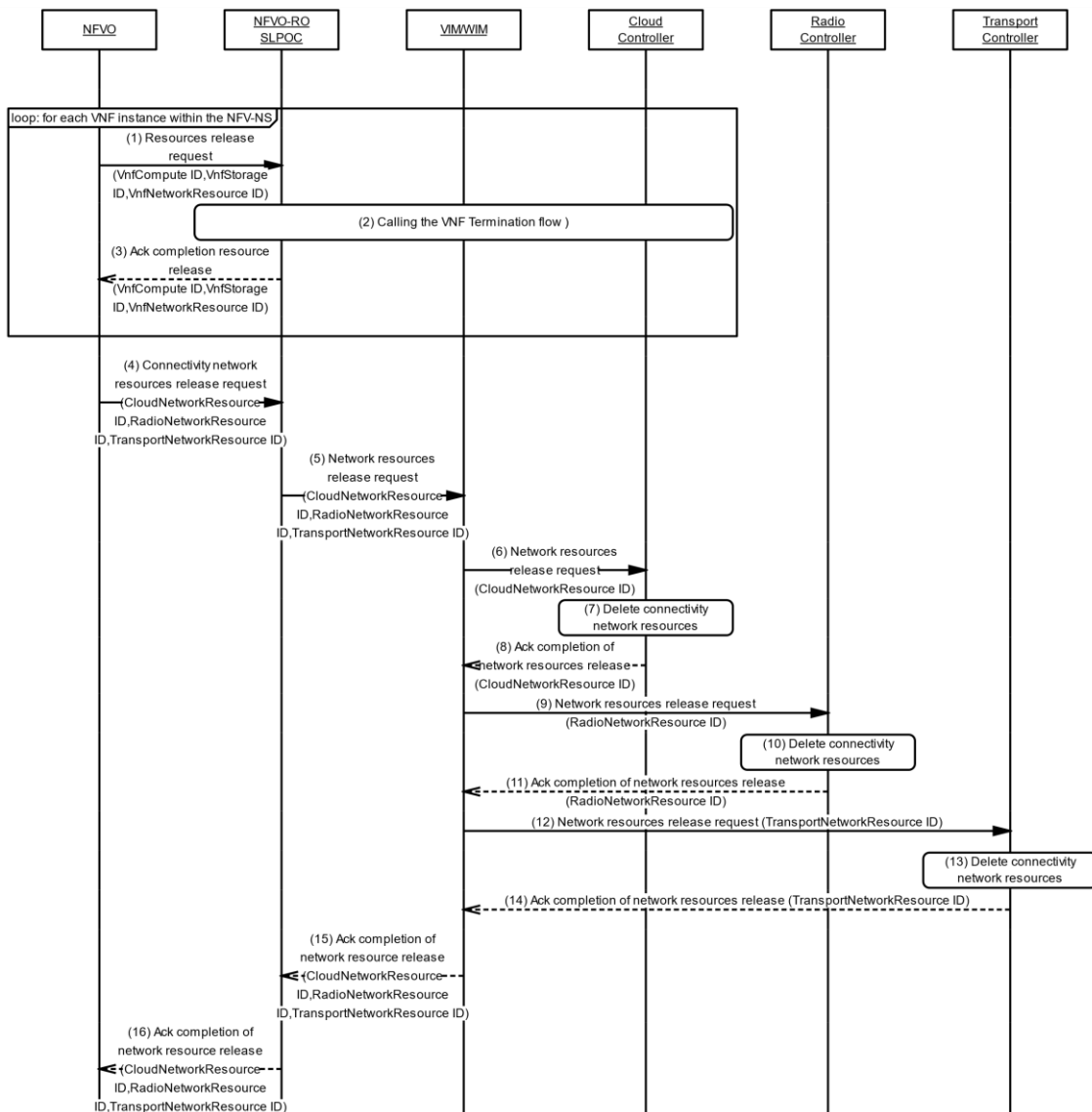
FIGURE 19: NON-NESTED NETWORK SERVICE TERMINATION FLOW

5.2 Terminate a non-nested network service

Figure 19 presents the 5GT-MTP workflow for terminating a non-nested network service. This flow is composed of two steps. The first step is to terminate the VNFs of the NFV-NS. The second step is to delete the network resources allocated for the connectivity network of the NFV-NS. The termination workflow is triggered by the Resource Termination Request for a NFV-NS from the NFVO to the 5GT-MTP- NFVO-RO SLPOC. The subsequent operations of the 5GT-MTP are listed below.

1. For each VNF instance that was instantiated along with the NFV-NS instantiation, NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC to release resources (compute, storage and network) used by the various VMs of the VNF instance using the operation Terminate Resource from Virtualized Compute Resources Management Interface, Virtualized Storage Resources Management Interface and Virtualized Network Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20]. The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of resources to be released.
2. 5GT-MTP- NFVO-RO SLPOC terminates the VNF by calling the VNF Termination flow.
3. 5GT-MTP- NFVO-RO SLPOC returns result of resources termination back to NFVO. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of released resources.
4. Using information kept for this NFV-NS instance, NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC to release the resources of the connectivity network of the NFV-NS instance using the operation Terminate Resource from Virtualized Network Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20] for the different domains (Cloud, Transport and Radio). The parameters include the identifiers (CloudNetworkResourceID, RadioNetworkResource ID and TransportNetworkResource ID) of released network resources.
5. 5GT-MTP- NFVO-RO SLPOC forwards the request to the VIM/WIM.
6. VIM/WIM requests from Cloud Controller deletion of network resources of the connectivity network.
7. Cloud Controller terminates the network resources of the connectivity network.
8. Cloud Controller acknowledges completion of network resource termination to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (CloudNetworkResource ID) of terminated resources.
9. VIM/WIM requests from Radio Controller termination of radio network resources of the connectivity network.
10. Radio Controller terminates the radio network resources of the connectivity network.
11. Radio Controller acknowledges completion of network resources termination to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (RadioNetworkResource ID) of terminated resources.
12. VIM/WIM requests from Transport Controller termination of the network resources for the connectivity network.

13. Transport Controller terminates the network resources of the connectivity network.
14. Transport Controller acknowledges completion of network resources release to the VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifier (TransportNetworkResource ID) of released resources.
15. VIM/WIM acknowledges the completion of network resources release back to 5GT-MTP- NFVO-RO SLPOC.
16. 5GT-MTP- NFVO-RO SLPOC returns result of network resources release back to NFVO. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifiers (CloudNetworkResource ID, RadioNetworkResource ID and TransportNetworkResource ID) of released network resources.
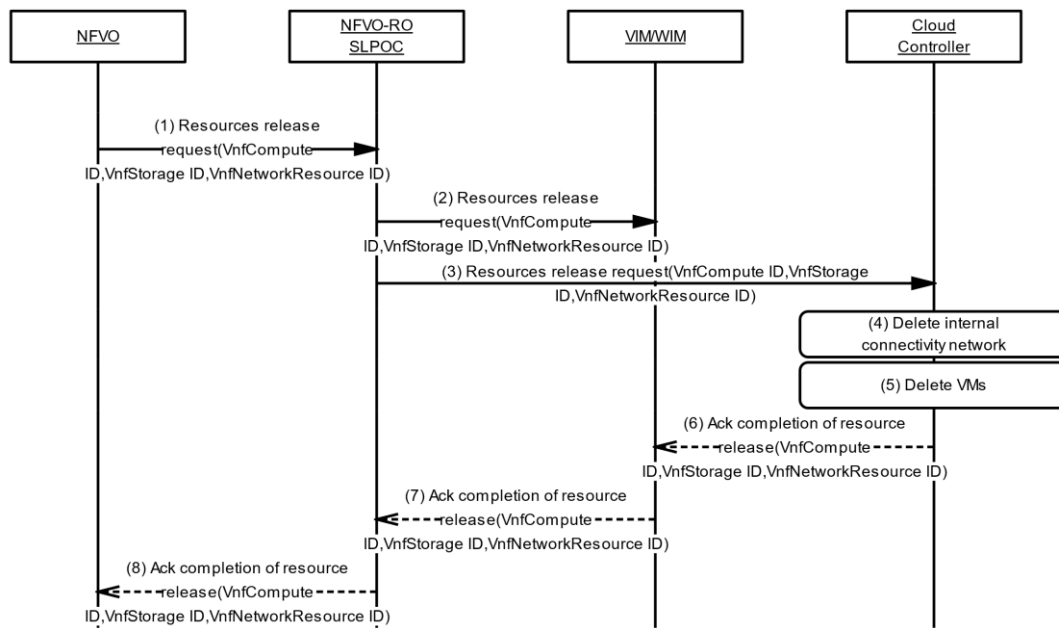
FIGURE 20: VNF TERMINATION WORKFLOW

5.2.1 VNF Termination

Figure 20 presents the 5GT-MTP workflow for terminating a VNF. It describes the deletion of VMs and connectivity resource composing the VNF. This workflow is triggered by the Resource Deletion Request from the NFVO to the 5GT-MTP- NFVO-RO SLPOC. The subsequent operations of the 5GT-MTP are listed below.

1. NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC for deletion of resources (compute, storage and network) used by the VNF instance using the operation Terminate Resource from Virtualized Network Resources Management Interface, Virtualized Compute Resources Management Interface and Virtualized Storage Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20]. The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of resources to be terminated.
2. 5GT-MTP- NFVO-RO SLPOC forwards the resources termination request to the VIM/WIM.
3. VIM/WIM forwards the resources termination request to the Cloud Controller.
4. Cloud Controller deletes the internal connectivity network.

5. Cloud Controller deletes the compute (VMs) and storage resources of the VNF instance.

6. Cloud Controller acknowledges the completion of resources release back to VIM/WIM. The acknowledge message includes the parameters reported in [20]. Such parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of released resources.

7. VIM/WIM acknowledges the completion of resources release back to 5GT-MTP-NFVO-RO SLPOC.

8. VIM/WIM acknowledges the completion of resources release back to 5GT-MTP-NFVO-RO SLPOC.

   The acknowledge message includes the parameters reported in [20]. Such parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of released resources.
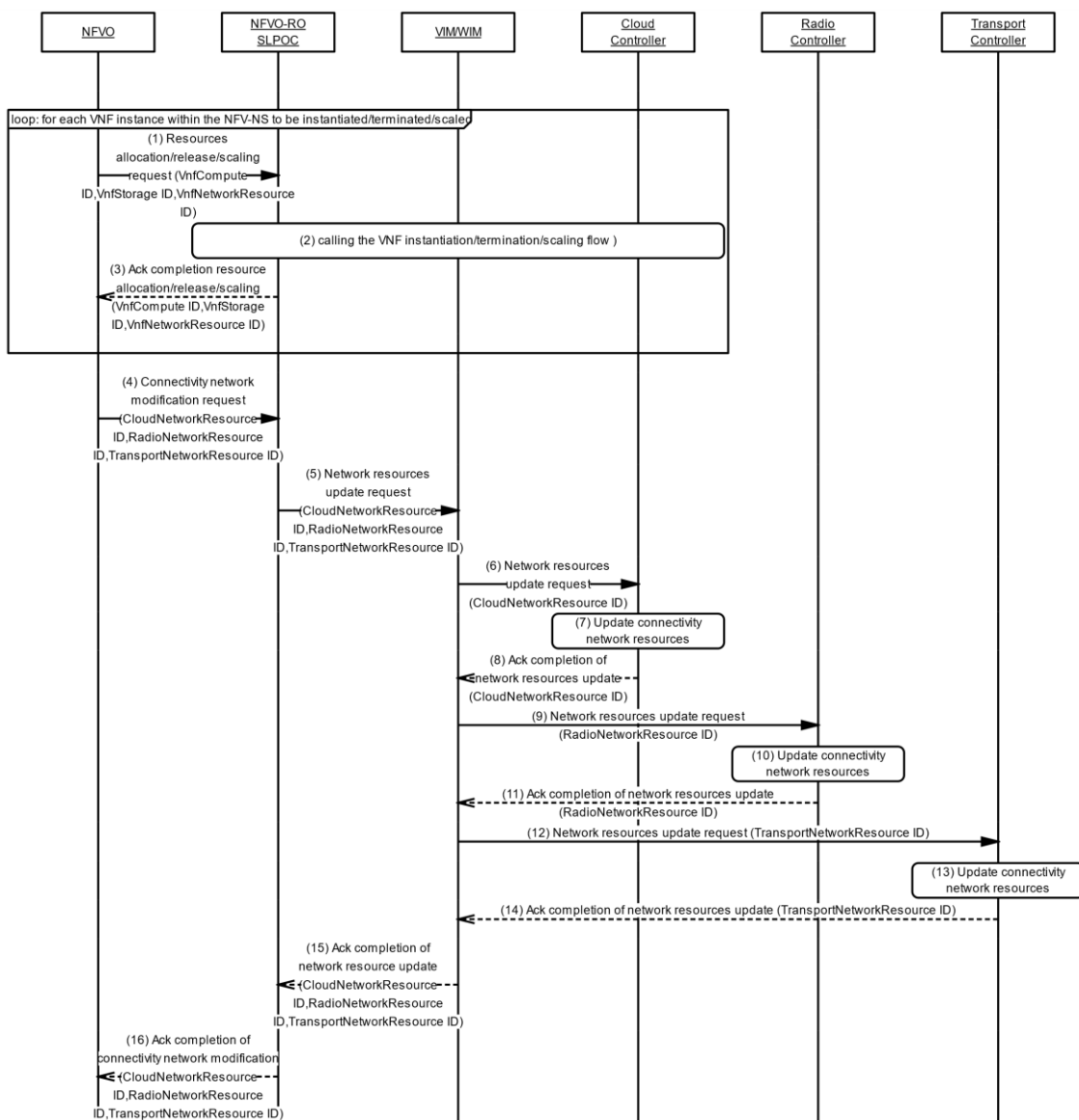
FIGURE 21: NON-NESTED NETWORK SERVICE MODIFICATION FLOW

5.3 Modify a non-nested network service

Figure 21 presents the 5GT-MTP workflow for modifying a non-nested network service. In the most general case the workflow could involve the scaling/instantiation/ termination of VNFs and the modification of the NFV-NS's connectivity network.

The modification workflow is triggered by the Add/Remove/Update VNF Resource Request from the 5GT-SO to the 5GT-MTP- NFVO-RO SLPOC. The subsequent operations of the 5GT-MTP are listed below.

1. For each VNFs to be instantiated/terminated/scaled, NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC for allocation/release/scaling of resources (compute, storage and network) using the operation Allocate/Terminate/Scale Resource from Virtualized Network Resources Management Interface, Virtualized Compute Resources Management Interface and Virtualized Storage Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20]. The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of resources to be allocated/released/scaled.

2. 5GT-MTP- NFVO-RO SLPOC allocates/terminates/scales the VNF resources (compute, storage and network) by calling the VNF instantiation/termination/scaling flow.

3. 5GT-MTP- NFVO-RO SLPOC returns result of resources (compute, storage and network) allocation/termination/scaling back to NFVO. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of allocated/released/scaled resources.

4. NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC for modification of the NFV-NS's connectivity network using the operation Update Resources from Virtualized Network Resources Management Interface based on ETSI NFV IFA 005 [20]. The request includes the parameters reported in [20]. The parameters include the identifiers (CloudNetworkResource ID, RadioNetworkResource ID and TransportNetworkResource ID) of resources to be updated.

5. 5GT-MTP- NFVO-RO SLPOC forwards the request to the VIM/WIM.

6. VIM/WIM sends a request to Cloud Controller for update of network resources for the connectivity network.

7. Cloud Controller updates the nework resources for the connectivity network.

8. Cloud Controller acknowledges the completion of resource update back to the VIM/WIM. The acknowledge message includes the identifier (CloudNetworkResource ID) of updated resources.

9. VIM/WIM sends a request to Radio Controller for update of network resources for the connectivity network.

10. Radio Controller updates the nework resources for the connectivity network.

11. Radio Controller acknowledges the completion of resource update back to the VIM/WIM. The acknowledge message includes the identifier (RadioNetworkResource ID) of updated resources.

12. VIM/WIM sends a request to Transport Controller for update of network resources for the connectivity network.

13. Transport Controller updates the nework resources for the connectivity network.

14. Transport Controller acknowledges the completion of resource update back to the VIM/WIM. The acknowledge message includes the identifier (TransportNetworkResource ID) of updated resources.

15. VIM/WIM acknowledges the completion of connectivity network update back to 5GT-MTP- NFVO-RO SLPOC.

16. 5GT-MTP- NFVO-RO SLPOC returns result of connectivity network update back to NFVO. The acknowledge message includes the parameters reported in [24]. Such parameters include the identifiers (CloudNetworkResource, RadioNetworkResource and TransportNetworkResource IDs) of updated resources.
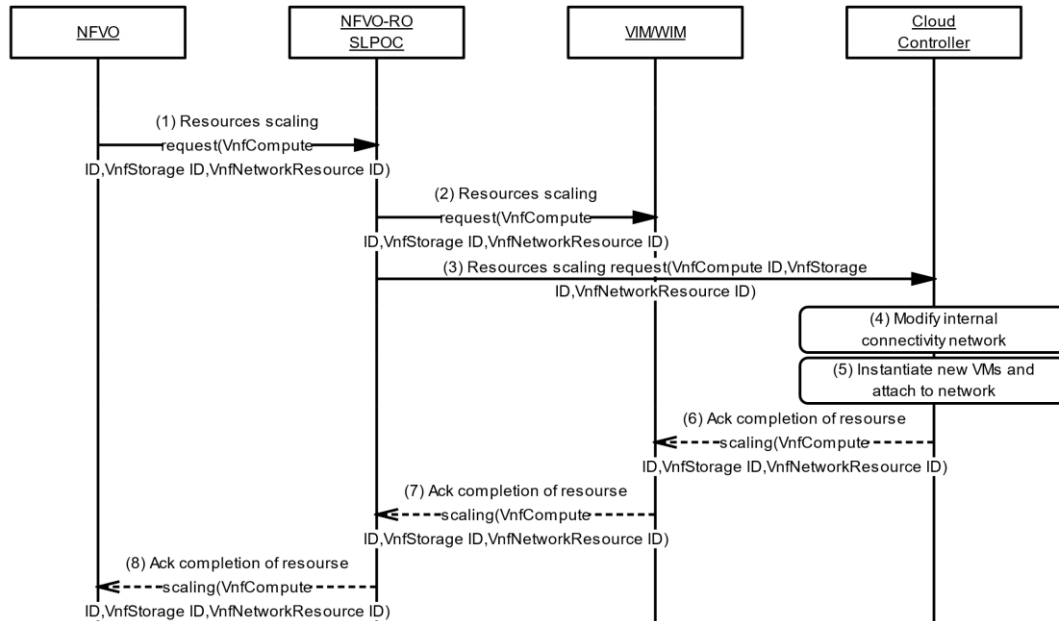
FIGURE 22: VNF INSTANCE SCALING FLOW

5.3.1 VNF instance scaling

Figure 22 presents the 5GT-MTP workflow for scaling a VNF. It describes the scaling of VNF storage and compute resources and the update of VNF network resources to connect eventually new created VMs to the internal connectivity network. The VNF instance scaling workflow is triggered by the VNF Resource Scaling Request from the NFVO to the 5GT-MTP- NFVO-RO SLPOC. The subsequent operations of the 5GT-MTP are listed below.

1. NFVO sends a request to 5GT-MTP- NFVO-RO SLPOC for modification of VNF resources (compute, storage and network) to the 5GT-MTP- NFVO-RO SLPOC using the operations Scale Resource from Virtualized Compute Resources Management Interface and Virtualized Storage Resources Management Interface, and Scale Resource from Virtualized Network Resources Management Interface [20]. The request includes the parameters reported in [20]. The parameters include the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of resources to be scaled/updated.

2. 5GT-MTP- NFVO-RO SLPOC forwards the resource scaling/update request to the VIM/WIM.

3. VIM/WIM requests from Cloud Controller the scaling/update of VNF resources.

4. Cloud Controller modifies as needed the internal connectivity network.

5. Cloud Controller creates and starts the needed compute (VMs) and storage resources and attaches new instantiated VMs to internal connectivity network.

6. Cloud Controller acknowledges the completion of resource change back to VIM/WIM. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of changed resources.

7. VIM/WIM acknowledges the completion of the scaling request back to 5GT-MTP- NFVO-RO SLPOC. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of changed resources.

8. 5GT-MTP- NFVO-RO SLPOC acknowledges the completion of the scaling request back to NFVO. The acknowledge message includes the identifiers (VnfCompute ID, VnfStorage ID and VnfNetworkResource ID) of changed resources.

5.4 Monitoring of virtual resources

Description:

The workflow describes the mechanisms used to request 5GT-MTP performance metrics from the 5GT-SO Monitoring Service to the 5GT-MTP Monitoring Service.

Prerequisites: The NFV-NS is already established and its NSD includes some monitoring parameters that the 5GT-SO Monitoring Service needs to elaborate as "5GT-SO Performance Metrics" (e.g. total consumption of RAM for the whole network service) starting from a collection of "5GT-MTP Performance Metrics" (e.g. consumption of RAM for each of the VMs instantiated for that network service).

Assumptions: The 5GT-MTP Monitoring Service collects raw monitoring data from VIM, Transport WIM and Radio WIM through dedicated monitoring agents that handle the VIM- or WIM-specific APIs used by the particular controller to expose monitoring data. Since these interfaces depend on the specific implementation, the exchange of messages between agents and VIM-/WIM- is not shown. For example, in case of a VIM based on OpenStack, the VIM monitoring agent could use the REST APIs of the Ceilometer service or it could connect directly to the OpenStack message queue; the Transport WIM monitoring agent could use the REST APIs of the SDN controller to retrieve OpenFlow statistics, etc.

The objective of the diagram is not to show all the possible sources of monitoring information for the 5GT-MTP Monitoring Service, but it is just providing some examples. Actually, the 5GT-MTP Monitoring Service could also collect monitoring data from VMs or containers or PNFs and so on. The mechanisms are always based on the same approach of defining dedicated monitoring agents to collect monitoring data and report them to the 5GT-MTP Monitoring Service, where they are further elaborated.

Workflow:

1. The 5GT-SO Monitoring Service, which acts as consumer of the 5GT-MTP Monitoring Service, identifies the 5GT-MTP performance metrics that are needed to elaborate the 5GT-SO performance metrics specified in the NSD of the network service just established.

2. For each of the required 5GT-MTP performance metrics, the 5GT-SO Monitoring Service requests the 5GT-MTP Monitoring Service to create a Performance Monitoring job (PM job), indicating the target resource and the desired performance metric. Further parameters, like collection and reporting period may be specified.

3. (a-b-c) Starting from the specification of the target resource, the 5GT-MTP Monitoring Service identifies the 5GT-MTP entity that is able to provide the raw monitoring data needed to compute the desired 5GT-MTP performance metrics. For example, in case of performance metrics related to computing resources, the source of monitoring data could be the VIM (case a), while for performance metrics associated to transport connections or radio connections the source could be the Transport (case b) or the Radio WIM (case c). In order to activate the collection of the raw monitoring data, the 5GT-MTP Monitoring Service sends a request to the related agent, indicating the elementary performance metric (i.e. the desired raw data).

4. (a-b-c) The agent activates the collection of the raw monitoring data. This may imply an interaction with the VIM/WIM for subscriptions or the starting of a thread for periodical polling, depending on the specific implementation and characteristics of the monitoring source.

5. (a-b-c) The agent provides a reply to acknowledge the result of the request.

6. The 5GT-MTP Monitoring Service creates an internal job to aggregate the raw monitoring data into the 5GT-MTP performance metric requested by the 5GT-SO Monitoring Service and assigns a unique ID to the PM job.

7. The 5GT-MTP Monitoring Service returns the PM job identifier to the 5GT-SO Monitoring Service.

8. The 5GT-SO Monitoring Service stores the returned identifier in its internal repository.

9. The 5GT-SO Monitoring Service sends a subscription request to the 5GT-MTP Monitoring Service, in order to activate the notifications from the 5GT-MTP Monitoring Service when new 5GT-MTP performance metrics are available. The subscription request includes a filter to identify the type of information that the 5GT-SO Monitoring Service wants to receive.

10. The 5GT-MTP Monitoring Service generates a unique subscription identifier, stores and returns it to the 5GT-SO Monitoring Service.

11. The 5GT-SO Monitoring Service stores the received identifier in its internal repository.

This step terminates the workflow which allows the 5GT-SO Monitoring Service to subscribe for receiving specific monitoring data from the 5GT-MTP Monitoring Service. During the service runtime, the 5GT-MTP Monitoring Service collects the raw monitoring data from the different 5GT-MTP entities, elaborates them and generates the 5GT-MTP performance metrics that are then collected from the 5GT-SO Monitoring Service. The detailed workflow for this second phase is the following:

12. (a-b-c) The VIM or WIM monitoring agent has collected new raw monitoring data related to the target resource and sends it to the 5GT-MTP Monitoring Service.

13. The 5GT-MTP Monitoring Service elaborates one or more raw monitoring data (e.g. through aggregation and correlation) generating a new 5GT-MTP performance metric, which is stored in the internal database.

14. The 5GT-MTP Monitoring Service notifies the 5GT-SO Monitoring Service about the presence of a new 5GT-MTP performance metric for the target resource.

15. The 5GT-SO Monitoring Service requests the desired 5GT-MTP performance metric (this message could be sent to receive multiple values with a single message; a filter is used to specify the requested metrics).

16. The 5GT-MTP Monitoring Service replies with the requested values.
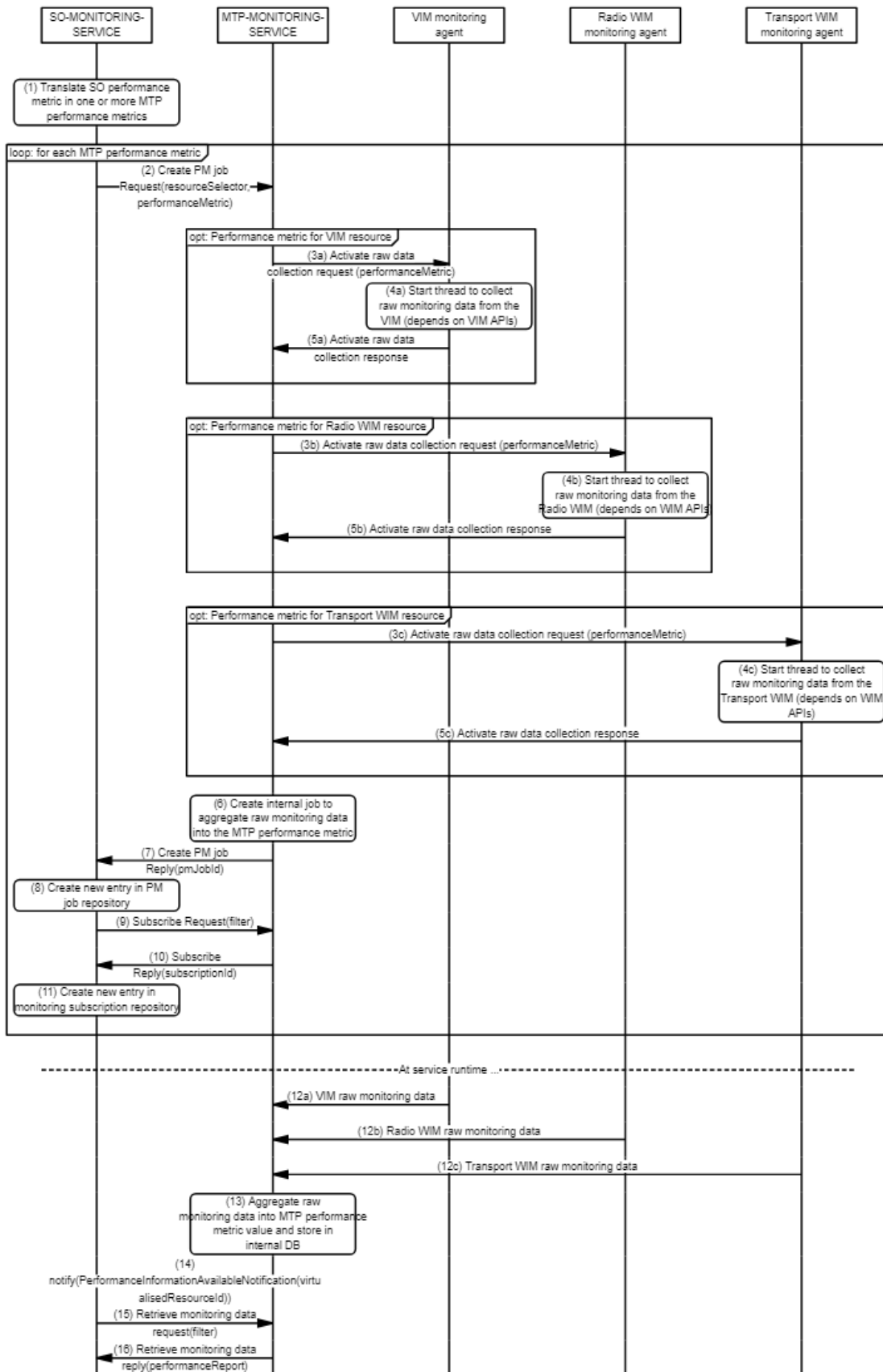
FIGURE 23: WORKFLOW FOR 5GT-MTP MONITORING

6 Use cases mapping to 5GT-MTP

6.1 vEPCaaS

The 5GT-MTP provides Network as a service solutions tailored to the needs of MNO and MVNOs. Several use cases are foreseen, and we focus on the provision of vEPCaas, providing core components to build a mobile network service offer. Both control plane and user planes can be virtualized while additional network functions packaged by the MVNO may be added to the relevant network service graph, and managed together with the provided vEPC VNFs. A rationale for this is they are involved in the same logical links or in the same control procedures. Another rationale is consistency between deployment favours of VNFs from different types. Slicing may be used to distinguish network functions packaged by the MVNO from the ones from the 5GT-MTP catalo. But it seems unneeded to give up the benefit of a layered architecture making the 5GT-MTP aware of a fundamental OSS construct. Instead, the OSS and the 5GT-SO system may, for instance, allow the 5GT-MTP to consider logical links which can cross slice subnet borders.

Verticals can be seen as customers of MNO/MVNO, but also as customers of MVNOs, including MVNOs that are themselves consumers of network service as described above. For this to be achievable, the 5GT-MTP must supply the appropriate means to offer services to verticals. From an access right and containment point of view, this 'recursiveness' is probably mainly achieved by design in the 5GT-VS and the 5GT-SO, e.g. embedding roles, making entities under control of a vertical being a subset of entities controlled by the MVNO. For example, in the OSS, an MVNO offering services to verticals would be provided with slices rather than communication services.

The 5GT-MTP will provide the necessary abstractions so that a given MVNO could be provided "on demand" with a full vEPC network. The abstraction levels allow to leverage on the possibility to aggregate several access, transport and control network functions as a logical link offered by the 5GT-MTP, see Figure 24.
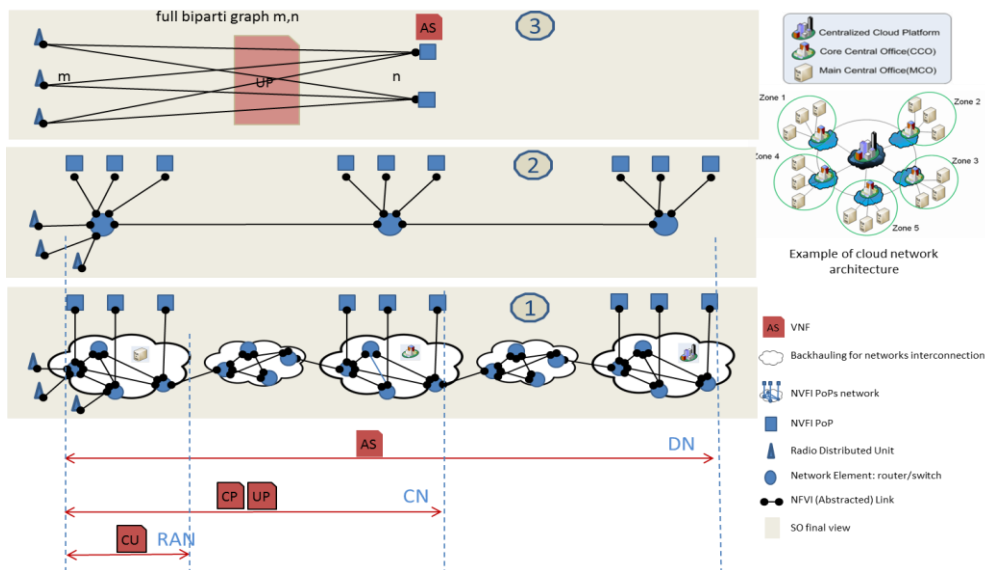
FIGURE 24: LOGICAL LINK ABSTRACTION EXAMPLE FOR VEPCAAS USE CASE

6.2 Connected Car

As mentioned in section 9.1, future connected vehicles pave the way to the development of several services where the automotive industry and mobile networks play a fundamental role. Among such services, Vehicle-to-Everything (V2X) safety applications have a prominent social and economic impact. The exchange of information on the vehicle dynamics, their processing as well as the delivery of warnings, can greatly reduce the risk of accidents involving vehicles as well as vulnerable roads users (pedestrians, cyclers). While the traditional way to deploy such services foresees vehicle-to-vehicle communication and safety applications implemented at the vehicle, within the 5GT MTP we consider an infrastructure-based deployment taking the collision avoidance between vehicles as a use case. In this scenario, an example of infrastructure-based service deployment is presented in Figure 25.
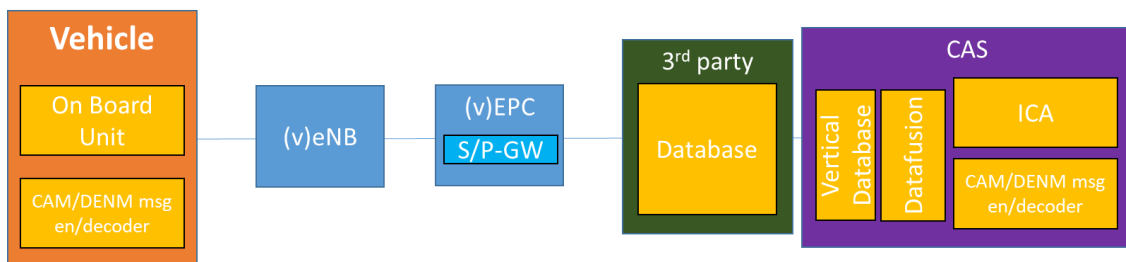
FIGURE 25: COLLISION AVOIDANCE USE CASE

According to the ETSI specifications in [33], each vehicle periodically (e.g., every 0.1 s) generates Cooperative Awareness Messages (CAMs), including the vehicle position, speed, heading, among others. As depicted in Figure 25, in our example CAMs are transmitted as V2I unicast messages to the (v)eNodeB covering the area of interest (e.g., urban intersection). Messages are then forwarded towards the Packet Gateway (P-GW) within the (v)EPC, which then hands them to a 3rd party-trusted database. The 3rd party-trusted database should store the most recent CAMs sent by the vehicles travelling over the geographical area of interest. The collision detection algorithm, which is run by the automotive vertical in a Collision Avoidance Server (CAS) and which we refer to as ICA hereinafter, selects and processes the information which are considered useful (through the Data fusion VA) from the 3rd party database and detects the risk of collisions between the vehicles, if any. Upon detecting a possible collision, the application generates a warning message, following the Decentralized Environmental Notification Message (DENM) format specified by ETSI [32], which is delivered to the involved vehicles through the (v)EPC and the (v)eNodeB. Vehicles receiving the warning can then display it to the driver, or execute a proper action (e.g., braking) if they are automated vehicles.

Importantly, the infrastructure supporting the ICA application should comply with several requirements, among which:

- reliable coverage over the monitored area (in order for the alarm messages to be delivered correctly and for the cars to receive the desired quality of service level),
- highly reliable positioning accuracy,
- strict latency, in order to take action when a dangerous situation is detected.

Due to the aforementioned latency constraint, the ICA application is a strong candidate for a Multi-access Edge Computing (MEC)-based implementation. Over the abstracted view of the of the physical and virtual resources presented by the 5GT-MTP, the 5GT-SO places the different Virtual Applications (VAs) of the ICA, e.g., the 5GT-SO instantiates the 3rd party database or the CAS where the ICA runs as an application server. The 5GT-MTP can present the abstraction of the resources to the 5GT-SO mainly in two ways: with the (v)EPC already deployed as a network service and handled transparently by the 5GT-MTP, or without the (v)EPC. In the latter case, the 5GT-SO is the entity that should instantiate and handle the (v)EPC.

In both cases, the 5GT-MTP presents the resources at hand with the corresponding capabilities:

- The access mobile resources, i.e., all the (v)eNodeBs. The 5GT-MTP presents each eNodeB with its coverage and mobility support capabilities. The 5GT-SO has the task of selecting which subset of eNodeBs to include in the collision avoidance application in the target area.
- The transport resources, i.e., all the (virtual) links. The 5GT-MTP hides the complexity of the network connectivity showing only the logical connectivity between the access network and the NFVI-PoPs, or between NFVI-PoPs. Each (virtual) link is presented by the 5GT-MTP with its latency, reliability and bandwidth so that the 5GT-SO can decide, in order to comply with the delay constraint, where to place all the VAs composing the collision avoidance application.
- The computation and storage capabilities, both at the MEC and at the Cloud NFVI-PoP.

Figure 26, presents an example of the abstraction described above.
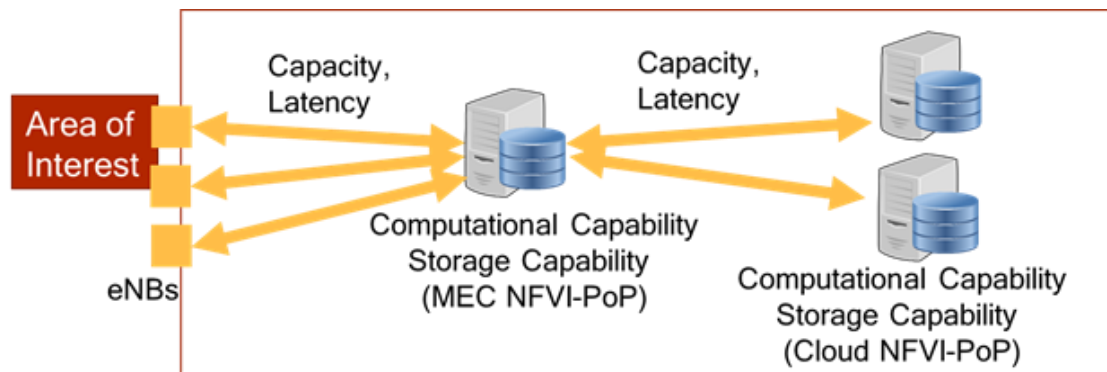
FIGURE 26: A POSSIBLE EXAMPLE OF ABSTRACTION FOR THE ICA APPLICATION

The main difference between the two abstractions, with or without the (v)EPC, is that the end-to-end connectivity is ensured by the 5GT-MTP only if the 5GT-MTP handles the (v)EPC. Otherwise, the 5GT-SO has to place, other than the VAs of the ICA application, also the entities of the EPC, as for example the S/P-GW, the MME and the HSS. When the (v)EPC is handled by the 5GT-MTP, the latency of each link presented in the abstraction takes into account that the control plane introduces some additional delay, mainly for bearer instantiation and handover procedures. If the (v)EPC is instead placed by the 5GT-SO the 5GT-SO autonomously accounts for the impact on the end-to-end connectivity of the EPC placement.

Finally, the decision of the 5GT-SO of placing the VAs of the ICA application in the MEC or in the Cloud NFVI-PoP is scenario-dependent, since the latency constraint is of utmost importance in this use case. If the Cloud NFVI-PoP presents a latency from the target area, which a sum of the transmission delay (due to the distance) and of the processing delay (due to the available processing capabilities) which is larger than the delay budget the collision avoidance requires, then the 5GT-SO places the ICA at the MEC, as in Figure 27. Furthermore, the collision avoidance application requires at least two different slice instantiations, one for the 3rd party database, which collects all the required CAMs from the target area, and one for the automotive vertical, where the ICA runs.
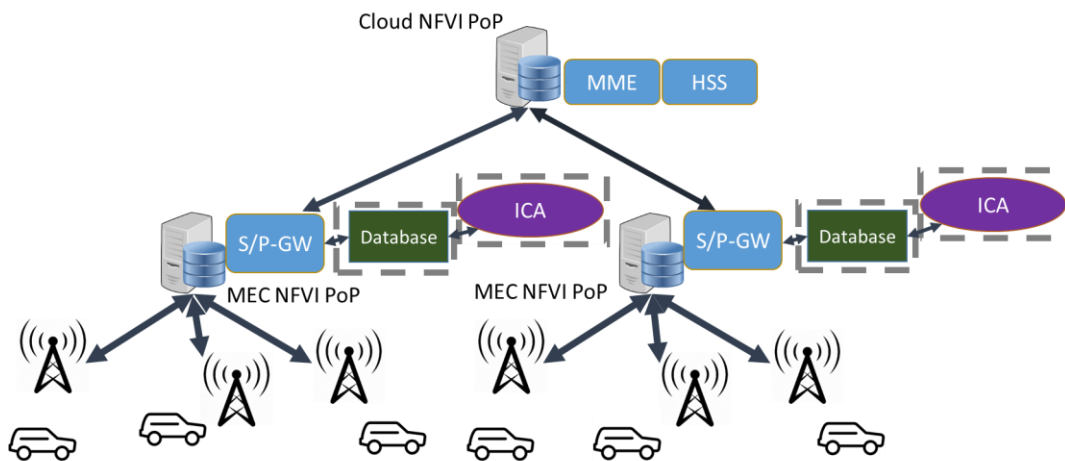
FIGURE 27: A POSSIBLE VAS PLACEMENT BY THE 5GT-SO

6.3 Cloud robotics

Cloud Robotics (CR) is a paradigm that leverages on cloud technologies and mobile communication to enhance the capabilities of robots. Control services are moved into the cloud running on dedicated hosts or data centres allowing the development of smart robotic systems with unlimited computing capacity. Offloading computation-intensive tasks to the cloud, only the necessary sensors, actuators, and basic processing power are kept on the robots. To allow the interaction among robots and the external environment in real-time, huge amounts of information will have to be transferred instantaneously. The mobile communication must satisfy specific requirements in terms of data rates, latency, reliability, density of connections, coverage, etc.

To allow a suitable deployment of the resources, the 5GT-MTP must provide an abstract view of the available resources with an adequate level of detail.

The 5GT-MTP will expose information about the availability of data centre resources, identifying also the geographical location of the servers for a correct placement of the V(N)F, thus the access point of the service (connection points) and the available connectivity among them in terms of logical link with specific parameters.

An example of abstract view for the Cloud Robotics use case is reported in Figure 28.
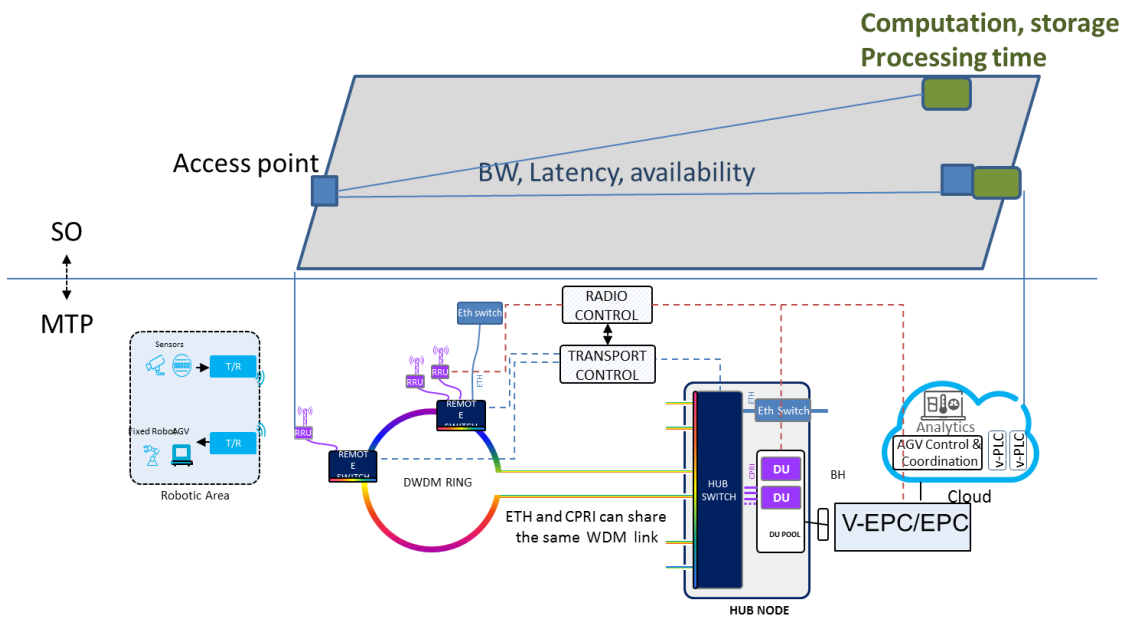
FIGURE 28: ABSTRACTION FOR CLOUD ROBOTICS USE CASE

According to the proposed abstraction model, the exposed abstraction hides the complexity of the underlay physical network, reporting only the logical link connecting a source and destination node. Each logical link can be associated to one or more physical paths. For instance, it can correspond to two disjoint wavelengths.

Each logical link is described in terms of QoS parameters that take into account the specific needs of the particular use case:

-   Available bit rate
-   Latency
-   Availability

As far as the DC resources are concerned, in the abstraction will be reported information about:

-   Available computation and storage resources (e.g., max CPU, max memory)
-   Processing time

Note that, the exposed parameters take into account both the contribution of the radio and transport network segment. For instance, latency represents the overall E2E latency.

6.4 Entertainment

The 5G-T project includes use cases coming from the Entertainment industry. The key objective within the project is to use cloud, NFV and SDN technologies in combination with the mobile communication to improve the experience of the fans attending to a sport event. With this new approach all the network functions and applications will become VNFs and VA running in the cloud but complying with strict requirements in terms of data rates, latency, etc.

In the 5G-T all the orchestration and management decisions taken at the 5GT-SO level will be based on the abstraction provided by the 5GT-MTP, and therefore it is critical to determine the level of detail of the abstraction required by the entertainment use cases.

In this deliverable we focus on the abstraction level required for the 3rd alternative as described in section 4.3.2, since this alternative represents the ground scenario (as it also happens in other use cases). The detail required for the other two alternatives can be built on top of this ground scenario and will be subject of study in future deliverables.

The abstraction in this case will expose: the data center resources, links between the resources and service access points (along with some additional information such as the geographical location). This way all the complexity from the physical network is hidden. In Figure 29 we show an example of how this abstraction can be used by the 5GT-SO to deploy the service associated with UC.E01 "On-Site Live Experience" (OLE).
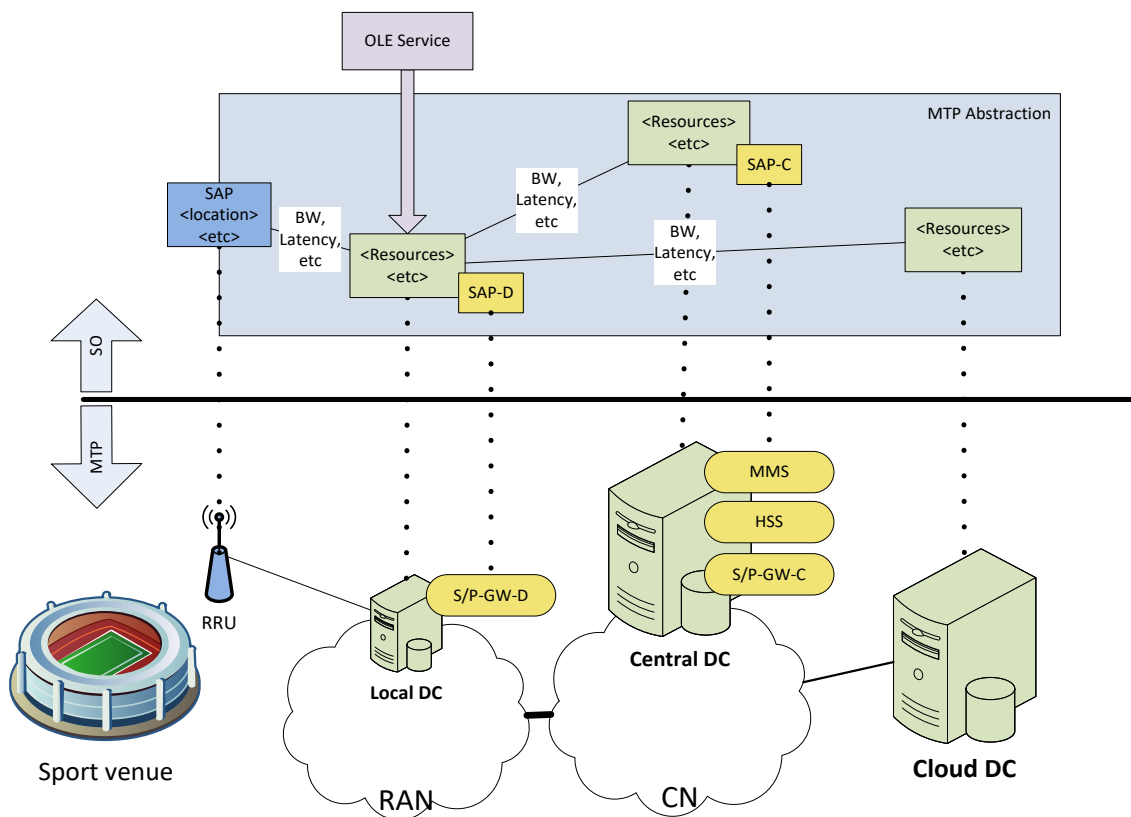
FIGURE 29: ENTERTAINMENT 5GT-MTP ABSTRACTION MAPPING

6.5 eHealth

The eHealth use case is one of the most critical verticals we have in the 5G-TRANSFORMER project due to the intrinsically characteristics of health emergency services on unexpected high demand of traffic with low latency requirements and synchronization of several and heterogeneous sources of information in short time.

eHealth can be defined as the delivering of health services by means of information and communication technologies. There are different kinds of services that can be provided by eHealth systems: from wearable devices connected to servers, to electronic health records and health information networks. The actors (both human and machines) involved in the use cases are also diverse in roles and dynamism. Focusing in the networking infrastructures, we can have a static scenario where the patient is

monitored at home or in a hospital or a mobile one where there is a need of mobile networking infrastructure or even the deployment ad-hoc emergency one.

There are two main targets on the eHealth vertical, the e-Infrastructure use case and the eHealth application. The objective in the first use case is to migrate the current TETRA based communication system to a 5G network. The main requirements are high-priority and low latency services. In the case of the eHealth application, the objective is take advantage of the MEC for automating the collection of data for wearables, detection of problems and automatic calling to medical services. In that case a part from the management of the collected data and low latency, there is a need of high quality real time video to interconnect the different actors: patient, ambulance, doctor at hospital.

We are considering the specific use case on Heart Attack Emergency, where the patient (a mobile user) has a smart wearable which can monitor precise cardiac, respiratory, sleeping and activity data. In a none emergency situation, the raw collected data is transmitted to the cloud via the 5G RAN. On the cloud takes place the processing, analysis and monitoring of the user. The 5G-MTP needs to provide a low latency cloud server close to the user to reduce the latency communication to the wearable. In a case of an emergency situation, the requirements for the location of the server that collects and monitors the raw data is more critical because there is a need of include the monitoring via real time video and even a remote surgery controlled remotely by the doctor in a hospital.

To allow a suitable deployment of the resources for all the cases, the 5GT-MTP must provide an abstract view of the available resources with an adequate level of detail. The 5GT-MTP will expose information about the geographical location of the user and the data centre resources, for a correct placement of the server that collect the wearable data in both an emergency and normal cases.

An example of abstract view for the Hearth Attack Emergency use case is reported in Figure 30.
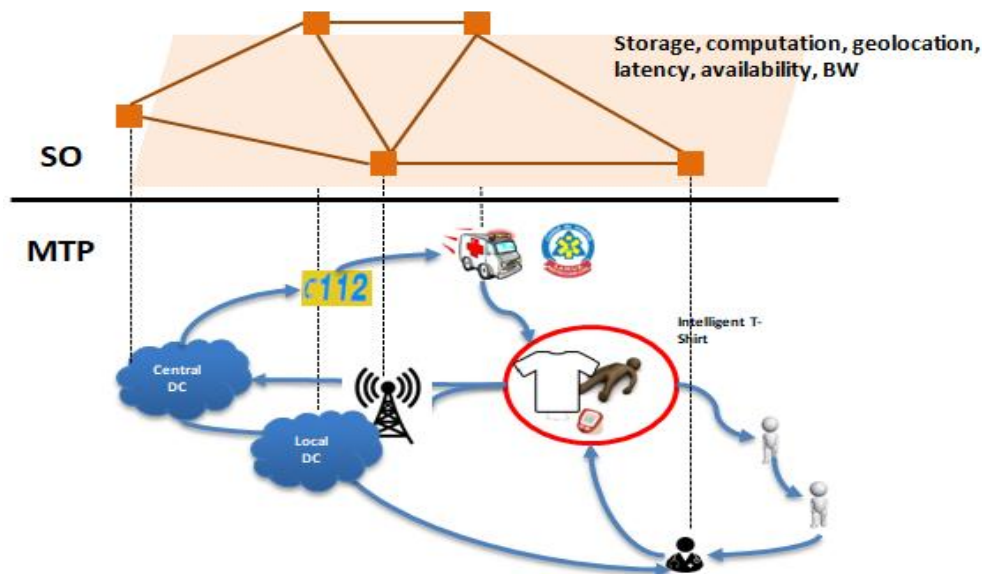
FIGURE 30: 5GT-MTP ABSTRACTION MAPPING FOR EHEALTH USE CASE

6.6 ETSI

6.6.1 ETSI NFV

Work is also ongoing inside ETSI NFV on how the NFV architecture in general, but more specifically, the ETSI MANO components can support network slicing.

In this respect the Evolution and Ecosystem (EVE) working group has carried out activities that map NFV and 3GPP network slicing concepts (see EVE012 [55]). On the one hand, ETSI NFV EVE012 [55] establishes the correspondence between a network slice (3GPP) and a network service (ETSI NFV). There, ETSI describes that an NFV Network Service (NFV-NS) can be regarded as a resource-centric view of a network slice, for the cases where a NSI would contain at least one virtualized network function. According to 3GPP 28.801 [6], an NSSI can be shared by multiple NSIs. The virtualized resources for the slice subnet and their connectivity to physical resources can thus be represented by the nested network service concept defined in ETSI NFV-IFA 014 [56] (right hand side of ¡Error! No se encuentra el origen de la referencia.), or one or more VNFs and PNFs directly attached to the Network Service used by the network slice. Figure 31¡Error! No se encuentra el origen de la referencia. illustrates the relationship between 3GPP's network slice and ETSI NFV network service.
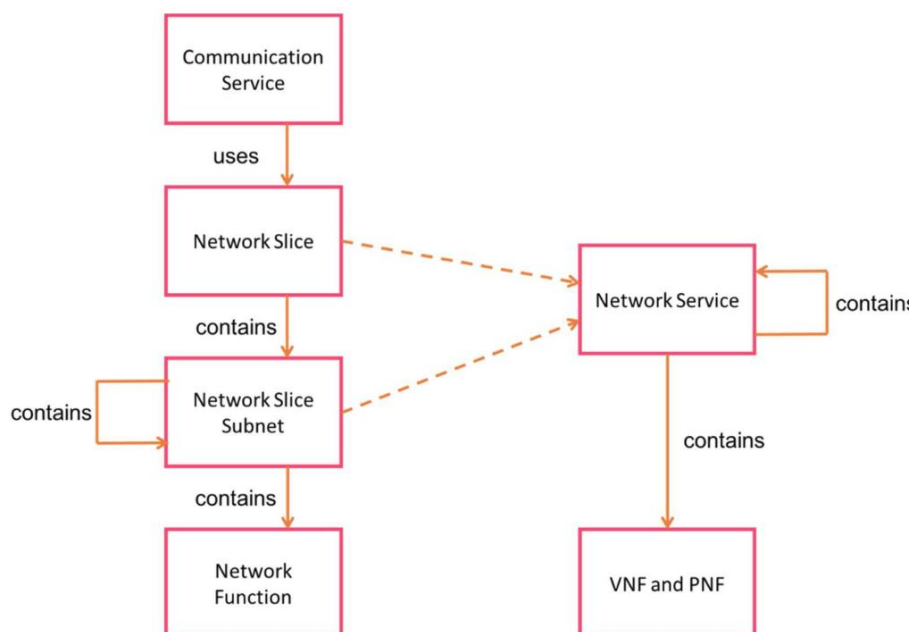
FIGURE 31: RELATION BETWEEN 3GPP AND ETSI INFORMATION MODELS (FROM [55])

As mentioned before, 3GPP 28.801 [6] identifies three management functions related to network slicing management: Communication Service Management Function (CSMF), Network Slice Management Function (NSMF), and Network Slice Subnet Management Function (NSSMF).

¡Error! No se encuentra el origen de la referencia.The Os-Ma-Nfvo reference point can be used for the interaction between 3GPP slicing related management functions and NFV MANO. To properly interface with NFV MANO, the NSMF and/or NSSMF need to determine the type of network service or set of network services, VNF and PNF that can support the resource requirements for a NSI or NSSI, and whether new instances of these network services, VNFs and the connectivity to the PNFs need to be created or existing instances can be reused.

From a resource management viewpoint, NSI can be mapped to an instance of a simple or composite network service instance or to a concatenation of such network service instances. From a resource management viewpoint, different NSIs can use instances of the same type of network service (i.e. they are instantiated from the same Network Service Descriptor or NSD) with the same or different deployment flavors. Alternatively, different NSIs can use instances of different types of network services. The first approach can be used if the NSIs share the same types of network functions (or a large common subset) but differ in terms of the performance expected from these network functions (and from the infrastructure connectivity service (ICS)s connecting them) and/or the number of instances to be deployed for each of them. If slices differ more significantly, mapping to different network services, each with its own NSD can be considered. The same mapping principles might apply to NSSIs.
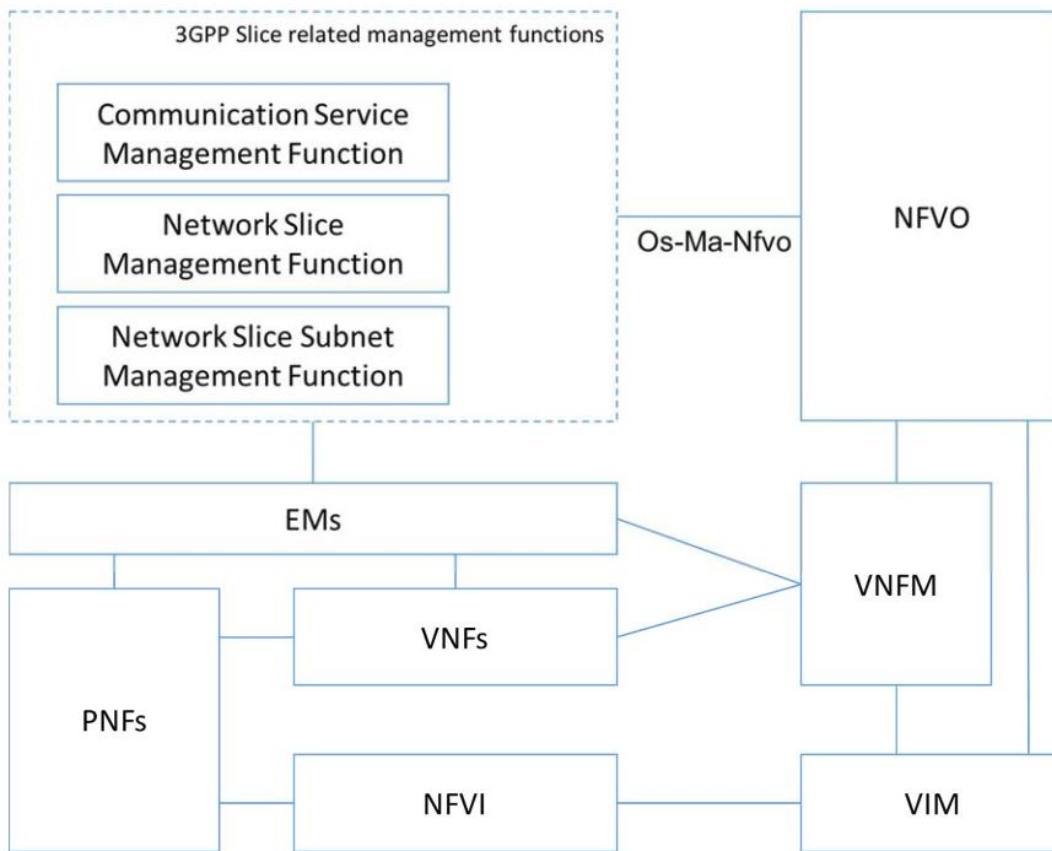
FIGURE 32: NETWORK SLICE MANAGEMENT IN AN NFV FRAMEWORK (FROM [55])

Also, as described before, 3GPP 28.801 [6] describes the lifecycle of a network slice, which is comprised of the four following phases: (i) Preparation; (ii) Instantiation, configuration and activation; (iii) Run-time; and (iv) Decommissioning.

The preparation phase includes the creation and verification of Network Slice Template(s) (NST(s)). From an NFV perspective, the resource requirement for a NST can be realized by one or more existing NSDs that have been previously on-boarded on the NFVO. The creation of a new NST can lead to requiring update of an existing NSD or generation of a new NSD followed by on-boarding the new NSD if the slice requirements do not map to an already on-boarded NSD. Indeed, the NFV-NS for the multiple NSIs may be instantiated with the same NSD, in order to deliver exactly the same optimizations and features but dedicated to different enterprise customers. On

the other hand, a network slice intended to support totally new customer facing services is likely to require a new NS and thus the generation of a new NSD. The network slice instantiation step in the second phase triggers the instantiation of the underlying NSs. NFV-MANO functions are only involved in the network slice configuration phase if the configuration of virtualisation-related parameters is required on one or more of the constituent VNF instances. Configuration of the network applications embedded in the constituent network functions involves the NSMF or NSSMF and/or other parts of the OSS/BSS, and the element managers (if any) associated to these functions. NFV-MANO functions can be triggered during the network slice activation step. If explicit activation of VNFs is required, the NSMF or the NSSMF can change the operational state of those VNFs through an Update NFV-NS operation defined in ETSI NFV-IFA 013 [42]. The involvement of NFV-MANO in the run-time phase is limited to the operations related to the performance management, fault management, and lifecycle management of virtualised resources (e.g., scaling an underlying NFV-NS to expand a NSI). The decommissioning phase triggers the termination of the underlying network service instances.

Additionally, and given the multiple administrative boundaries of the 5G-TRANSFORMER architecture, the Interfaces and Architecture (IFA) working group is of particular interest for our project. ETSI NFV IFA028 [47] reports on potential architecture options to support the offering of NFV MANO services across multiple administrative domain. NFV-MANO services can be offered and consumed by different organizations, e.g. by different network operators or by different departments within the same network operator. Administrative domains as defined in ETSI NFV IFA010 [57] can be mapped to such different organizations. Examples of use cases for NFV-MANO service offerings across multiple administrative domains are described in ETSI NFV 001. Furthermore, ETSI NFV IFA022 [58] reports on the functional architecture needed to provision and manage multi-site network services. To this end, a set of multi-site use cases are studied.

Furthermore, compliance with widely accepted standards of the 5G-TRANSFORMER architecture is also relevant to maximize its impact. Therefore, in a more general architectural context than that defined by the previous documents (which focus on specific issues) the interfaces already defined in ETSI NFV MANO are also relevant:

- ETSI GS NFV-IFA 013 [42] defines the interfaces supported over the Os-Ma-nfvo reference point of the NFV MANO architectural framework as well as the information elements exchanged over those interfaces;
- ETSI GS NFV-IFA 005 [24] defines the interfaces supported over the Or-Vi reference point of the NFV MANO architectural framework as well as the information elements exchanged over those interfaces.

ETSI GS NFV-IFA 006 [44] defines the interfaces supported over the Vi-Vnfm reference point of the NFV MANO architectural framework as well as the information elements exchanged over those interfaces.

6.6.1.1 ETSI MEC

Multi-Access Edge Computing (MEC) is one of the key concepts for fulfilling some of the requirements of vertical services, and therefore its integration in the 5G-TRANSFORMER architecture is nexus in its design. MEC and its integration in an NFV

context was studied in ETSI MEC017 [18] document and a reference architecture is provided with the following key observations:

- The mobile edge platform is deployed as a VNF and therefore the procedures defined by ETSI NFV for this means are used;
- ETSI NFV MANO sees mobile edge applications as regular VNFs allowing for reuse of ETSI MANO functionality (with perhaps some extensions);
- The virtualization infrastructure is deployed as a NFVI and its virtualized resources are managed by the VIM. For this purpose, the procedures defined by ETSI NFV infrastructure specifications, i.e. ETSI NFV INF003 [59], ETSI NFV INF004 [60] and ETSI NFV INF005 [61] can be used.

7 Conclusions

In this deliverable, we presented the initial design of the 5G-TRANSFORMER Mobile Transport and Computing Platform (5GT-MTP). The deliverable addressed the following aspects of the 5GT-MTP: the internal architecture of the 5GT-MTP; the 5GT-MTP Northbound Interface (NBI) and abstraction towards the service orchestrator (5GT-SO); the workflows between the 5GT-SO and the 5GT-MTP as well as workflows among the various components of the 5GT-MTP; and the mapping of the 5G-TRANSFORMER use cases to the 5GT-MTP. The following highlights the key achievements in this deliverable:

- An analysis of the 5GT-MTP innovations beyond the state-of-the-art, namely: the integration of a MEC platform; the ability to compose a connectivity service and expose it to the 5GT-SO; and the decoupling of the VIM from the NFVO and VNFM.
- A comprehensive description of the 5GT-MTP system architecture including: Physical Network Functions (PNFs), Virtual Network Functions (VNFs), Virtual Infrastructure Manager (VIM), Wide area network Infrastructure Manager (WIM) and 5GT-MTP Single Logical Point of Contact for resource orchestration (5GT-MTP NFVO-RO SLPOC).
- A detailed description of the 5GT-MTP workflows associated with the following lifecycle events: instantiating a non-nested network service, modifying a non-nested network service, terminating a non-nested network service, VNF instantiation, VNF termination and monitoring of virtual resources.
- An exhaustive characterization of the 5GT-MTP technical requirements at different stages of the service lifecycle, i.e. service discovery, assurance, fulfilment and decommissioning.
- An in-depth description of the 5GT-MTP North Bound Interface (NBI) towards the Service Orchestrator (5GT-SO) specifying the abstraction of resources exposed by 5GT-MTP via NFVO-RO SLPOC.
- A detailed mapping of the vertical use cases (i.e. automotive, entertainment, e-Health, e-Industry and MNO/MVNO) to the 5GT-MTP.
- Baseline examples of the YANG information modelling for computational and storage resources.

In a nutshell, the main contribution of this deliverable is an initial design of the 5GT-MTP as an evolved VIM that abstracts the underlying infrastructure (i.e. radio, transport network, compute and storage) in order to dynamically support the specific requirements of the vertical industries. Future work will extend the present 5GT-MTP design to include: definition of data models for the abstracted resources, specification of resource orchestration algorithms and extension of standard interfaces to support the 5GT-MTP and 5GT-SO interaction.

8 References

[1] 5G-TRANSFORMER, D1.1, Report on vertical requirements and use cases, November 2017.

[2] 5G-TRANSFORMER, D1.2, 5G-TRANSFORMER initial system design, May 2018.

[3] 5G-TRANSFORMER, D3.1, Definition of vertical service descriptors and SO NBI, March 2018.

[4] 5G-TRANSFORMER, D4.1, Definition of service orchestration and federation algorithms, service monitoring algorithms, March 2018.

[5] NGMN Alliance, "Description of Network Slicing Concept", NGMN 5G P1 Requirements & Architecture Work Stream End-to-End Architecture, v1, Jan. 13, 2016.

[6] "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Study on management and orchestration of network slicing for next generation network (Release 15)", 3GPP TR 28.801 V15.0.0 (2017-09).

[7] ETSI GS NFV-INF 001 V1.1.1 (2015-01)

[8] ETSI GS NFV-MAN 001, "Network Functions Virtualisation (NFV); Management and Orchestration", v1.1.1, 2014.

[9] 5G!Pagoda, "D2.3: Initial report on the overall system architecture definition", https://goo.gl/2Z7PWu, June 30, 2017

[10] SONATA D2.2. Architecture Design, [Online] http://www.sonata-nfv.eu/content/d22-architecture-design-0

[11] D. King, Y. Lee, "Applicability of Abstraction and Control of Traffic Engineered Networks (ACTN) to Network Slicing", draft-king-teas-applicability-actn-slicing-01, Exp. Jan. 3, 2018.

[12] 5G PPP Architecture Working Group, "View on 5G Architecture (Version 2.0)", 2017-07-18.

[13] X. de Foy, A. Rahman, "Network Slicing - 3GPP Use Case", draft-defoy-netslices-3gpp-network-slicing-01, Exp. October 28, 2017.

[14] 5G Americas, "Network Slicing for 5G Networks & Services", Nov. 2016.

[15] ETSI, "Network Function Virtualisation – White Paper on NFV priorities for 5G", Feb 21, 2017.

[16] ONF, "TR-526 Applying SDN Architecture to 5G Slicing", Issue 1, April 2016.

[17] "Network Slicing Supported by Dynamic VIM Instantiation" - Presentation at NFVRG; IETF100; Nov 2017; https://datatracker.ietf.org/meeting/100/materials/slides-100-nfvrg-3-network-slicing-support-by-dynamic-vim-instantiation/

[18] "Network Slicing Use Cases" draft 02; IETF100; Nov 2017; https://datatracker.ietf.org/doc/draft-netslices-usecases/

[19] "Technology Independent Information Model for Network Slicing" draft 01; IETF100; Nov 2017; https://datatracker.ietf.org/doc/draft-qiang-coms-netslicing-information-model/

[20] "Interconnecting (or Stitching) Network Slice Subnets "draft 01; IETF100; Oct 2017; https://tools.ietf.org/html/draft-defoy-coms-subnet-interconnection-01

[21] "Problem Statement of Supervised Heterogeneous Network Slicing" draft ; IETF100; Oct 2017; https://tools.ietf.org/html/draft-geng-coms-problem-statement-01

[22] ORCHESTRA project: http://www.orchestraproject.eu/

[23] N. Sambo *et al.*, "Sliceable transponder architecture including multiwavelength source," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 6, no. 7, pp. 590-600, July 2014.

[24] ETSI GS NFV-IFA 005, "Network Function Virtualisation (NFV); Management and Orchestration; Or-Vi reference point - Interface and Information Model Specification", v2.1.1, 2016.

[25] R. Enns, M. Bjorklund, J. Schoenwaelder, and A. Bierman, *Network Configuration Protocol (NETCONF)*, IETF RFC 6241, Jun. 2011.

[26] M. Dallaglio, N. Sambo, F. Cugini, and P. Castoldi, "Control and management of transponders with NETCONF and YANG," *J. Opt. Commun. Netw.*, vol. 9, no. 3, pp. B43-B52, Mar. 2017.

[27] M. Bjorklund, YANG–A Data Modeling Language for the Network Configuration Protocol (NETCONF), IETF RFC 6020, Oct. 2010.

[28] A. Shaikh, T. Hofmeister, V. Dangui, and V. Vusirikala, "Vendor-neutral network representations for transport SDN," in *Proc. Opt. Fiber Commun. Conf.*, 2016, pp. 1-3.

[29] Network Functions Virtualisation (NFV) Release 2; Protocols and Data Models; NFV descriptors based on TOSCA specification.

[30] Network Functions Virtualisation (NFV); Protocols and Data Models; NFV descriptors based on YANG Specification.

[31] B. Li et al., "An MEC and NFV Integrated Network Architecture", April 2017, http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2017/2/articles/201706/P020170613285738133203.pdf

[32] ETSI TS 102 637-3 V1.1.1 (2010-09), Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service.

[33] ETSI TS 102 637-2 V1.2.1 (2011-03) - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications Part 2: Specification of Cooperative Awareness Basic Service.

[34] Sauter Martin, "From GSM to LTE-Advanced", 2nd Ed., John Wiley & Sons, Ltd, 2014.

[35] "5G radio access", Ericsson White paper, UEN 284 23-3204 Rev C, April 2016.

[36] "5G systems", Ericsson White paper UEN 284 23-3251 Rev B, January 2017.

[37] ETSI MEC, [Online], http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing

[38] ETSI GS MEC 001 V1.1.1, Mobile Edge Computing (MEC): Terminology, March 2016.

[39] ETSI GR MEC 017, Mobile Edge Computing (MEC); Deployment of Mobile Edge Computing in an NFV environment, Draft 2017/10

[40] 5G-PPP, "5G: Serving vertical industries", https://5g-ppp.eu/2nd-5g-vertical-workshop/

[41] 3GPP. "Study on management and orchestration of network slicing for next generation network v15.1.0 (Release 15)." 3GPP Technical Report TR28.801, Jan. 2018.

[42] ETSI GS NFV-IFA-013, V2.3.1, Management and Orchestration; Os-Ma-Nfvo reference point - Interface and Information Model Specification, 2017.

[43] Xi Li et.al., Service Orchestration and Federation for Verticals, 1st Workshop on Control and management of Vertical slicing including the Edge and Fog Systems (COMPASS), Barcelona, IEEE, 2018.

[44] ETSI GS NFV-IFA 006, "Network Function Virtualisation (NFV); Release 2; Management and Orchestration; Vi-Vnfm reference point - Interface and Information Model Specification", v2.4.1, February 2018.

[45] ETSI GS NFV IFA 008, "Ve-Vnfm reference point - Interface and Information Model Specification." V2.4.1, Release 2, Feb. 2018.

[46] P. Iovanna et.al., 5G Mobile Transport and Computing Platform for verticals, 1st Workshop on Control and management of Vertical slicing including the Edge and Fog Systems (COMPASS), Barcelona, IEEE, 2018.

[47] ETSI GS NFV-IFA 028, "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on architecture options to support multiple administrative domains", v3.1.1, January 2018.

[48] B. Chatras, U. S. Tsang Kwong and N. Bihannic, "NFV enabling network slicing for 5G," 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), Paris, 2017, pp. 219-225.

[49] TeleManagement Forum. "TeleManagement Forum Information Framework (SID): GB922_Addendum_4SO_Service_Overview_R9-5_v9-7".

[50] 3GPP TR28.801. Telecommunication management; Study on management and orchestration of network slicing for next generation network.

[51] 3GPP TS23.501, V15.0.0, System Architecture for the 5G System; Stage 2, 2017.

[52] 3GPP TS 38.300, "NR; NR and NG-RAN Overall Description; Stage 2; (Release 15)", v15.0.0, December 2017.

[53] 3GPP TS28.530, V0.4.0, Telecommunication management; Management of 5G networks and network slicing; Concepts, use cases and requirements, 2017.

[54] 3GPP TR 28.531, "Management and orchestration of networks and network slicing; Provisioning; Stage 1 (Release 15)", v0.3.0, February 2018.

[55] ETSI GR NFV-EVE 012, "Network Functions Virtualisation (NFV) Release 3; Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework" v3.1.1, December 2017.

[56] ETSI GS NFV IFA 014, "Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Network Service Templates Specification" v2.3.1, August 2017.

[57] ETSI GS NFV-IFA 010, Management and Orchestration; Functional requirements specification", v2.4.1, 2018.

[58] ETSI GR NFV-IFA 022, Management and Orchestration; Report on Management and Connectivity for Multi-Site Services", v0.8.2, 2018.

[59] ETSI GS NFV-INF 003, Infrastructure; Compute Domain", v1.1.1, 2014.

[60] ETSI GS NFV-INF 004, Infrastructure; Hypervisor Domain", v1.1.1, 2015.

[61] ETSI GS NFV-INF 005, Infrastructure: Network Domain, V1.1.1, 2014.

9 Annex I: Vertical Services

The 5G-TRANSFORMER consortium includes partners from several Vertical Industries identified in the market portfolio of the 5G-PPP [35], namely: automotive, entertainment, healthcare and manufacturing, as well as representatives from the MNO/MVNO industry. The following sections summarize the use cases established in D1.1 [1] and detail the expectations within the 5G-TRANSFORMER project in order to ease the reading of the rest of the document.

9.1 Automotive

The automotive industry is currently undergoing key technological transformations as more and more vehicles are connected to the Internet and to each other, and advances are being made toward higher automation levels. In order to deal with increasingly complex road situations, automated vehicles will have to rely not only on their own sensors, but also on those of other vehicles, and will need to cooperate with each other rather than make decisions on their own.

These trends pose significant challenges to the underlying communication system, as information must reach its destination reliably within an exceedingly short time frame - beyond what current wireless technologies can provide. 5G, the next generation of mobile communication technology, holds promise of improved performance in terms of reduced latency, increased reliability and higher throughput under higher mobility and connectivity density.

Vehicle domain features differ across the target operative scenarios which are strongly characterized by their own peculiarities. In order to better analyze the needs of the automotive domain versus the incoming communication technology, we considered four main scenarios (urban, rural, highway and transversal) and several use cases quite different for their unique features outlining the key aspects that have the most impact on 5G.

Typical automotive use cases are various and can address heterogeneous domains. In D1.1 [1] more than 25 use cases from those most popular in the literature have been described; the identified use cases are grouped in 6 domains: safety, mobility, entertainment, e-road, digitalized vehicles and automated vehicles.

In the 5G-TRANSFORMER project, we focus on the safety domain where, thanks to 5G capabilities, the vehicle can outline/foresee dangerous situations and properly react on time. In particular, two use cases have been selected and proposed for implementation: Intersection Collision Avoidance and See-Through as described in Table 8.

TABLE 8: AUTOMOTIVE USE CASES

ID	Goal In context	General description
UC A.01/ UC A.02 ICA (Intersection Collision Avoidance)	Avoid possible collision crossing intersection	The purpose of the ICA system is to alert drivers about the existence of any possible obstacles and eventually activate the emergency braking system. The communication infrastructure facilitates a real-

		time exchange of data between the involved entities.
UC A.04 See-Through (Safety)	Vehicles are able to see through obstacles, thanks to cooperation among them achieving bilateral awareness of road conditions	Thanks to the cooperation between vehicles, streaming information is provided to all the vehicles that want/need to access to it. This information can be used to identify potential obstacles that cannot be detected through on-board sensors.

9.2 Entertainment

The Media and Entertainment (M&E) industry is one of the industries most affected by the deep changes in terms of user habits and expectations that society has been experiencing with the explosion of Internet. The number of users grows daily and the users demand progressively media-rich content and a better quality of experience.

While all these changes provide a great economical fuel for the industry, they also impose challenges to the network infrastructures not present before, e.g., data rates, number of connections, quality of experience, etc. 5G PPP already identifies the entertainment industry as one of the key interested parties. This is because one of the key objectives of 5G is to open operators' networks for new services, and this is the key enabler to support the data rates and the latency required to give an immersive experience. Furthermore, 5G also aims to provide the services with network information not available before (i.e. packet losses, signal level, etc.) to better adapt the service to the network conditions.

The 5G-TRANSFORMER project focuses on the M&E services, in particular, targeting sports events. The aim is to encompass these services to the "fan engagement" trend, which envisions smarter venue services by means of providing targeted and high-quality content and following fans along the journey with contextualized information. This trend also envisions fans as content producers (i.e. to share videos, photos, emotions, opinions, comments, etc.), and captures the explosion of IoT devices by including them as additional content producers. The final goal is to give the fans a more interactive, immersive and participative experience like never before.

The following use case is considered in the project, in order to address the needs for the different actors and scenarios identified in D1.1 for the Entertainment vertical industry:

TABLE 9: ENTERTAINMENT USE CASE

ID	Goal In context	General description
UC E.01 On-site live event experience	To provide a better fan experience to users attending (on-site) an event	Large scale event sites, such as stadiums are more and more being connected in order to give better experience to their customers (replay, choose a specific camera, language, augmented reality to bring additional information, etc.)

9.3 eHealth

The eHealth use case is one of the most critical Vertical Industries in the 5G-TRANSFORMER project. This industry can effectively take advantage of the future 5G networks to improve the quality of life and medical assistance of people in emergency situations. It aims to be able to detect and assist people in emergencies in the minimum possible time in order to assure the maximum probability of people surviving the emergency.

5G networks will be able to support high demands of traffic with low delay requirements, thus it is very valuable for the eHealth use case because it facilitates discovery and attendance in short time.

On one hand, the e-Infrastructure use case focuses on how the current municipality infrastructure based on TETRA can be replaced based on the 5G features. This will allow emergency alarms to be received with smaller delay and thus be processed in a short amount of time to send an ambulance to the place of the emergency. This will also allow access in real time to the clinical history of the patient from the place of the incident to provide the patient with better medical attention. In addition, to have a better e-Infrastructure, the eHealth use case will need a high-priority and low latency service in the 5G-TRANSFORMER system. To address that, the 5G-TRANSFORMER system will allow access to the resources of system in extreme cases where the network is overloaded by users such as in big events.

On the other hand, the eHealth application aims to leverage on new technologies such as MEC, improving response time. This application aims to reduce the response time and automate processes of communication among medical personnel and between patient and medical personnel. The idea is to have an application based on MEC for automating the collection of data from wearables, detection of problems, and automatic ambulance requests, all of which require mechanisms for patient feedback (call back). If possible, it is important to provide video feed between emergency teams and off-site doctors because ambulance personnel are not necessarily specialized in some emergencies, such as urgent surgery. In this case they would need to contact a doctor which would monitor and guide the process over real time 4K video.

TABLE 10: EHEALTH USE CASE

ID	Goal In context	General description
UC H.01 Heart attack emergency	To provide better medical assistance in emergency cases	Large scale event sites where a lot of groups are deployed to cover the emergencies and have to communicate between them in real time. Emergencies that require real time communication between the ambulances and doctors. Improvement of the current infrastructure to guarantee the real time exchange of information to detect early the emergencies.

9.4 eIndustry

The production and manufacturing industry is currently undergoing important changes mainly driven by the ongoing introduction of new emerging technologies, including mobile network, cloud computing, robotics, machine intelligence and big data. Nowadays we are facing a new industrial revolution, commonly referred to as Industry 4.0, whose aim is to provide mass customization with costs comparable to those of mass production. This can be achieved leveraging on full digitalization and automation of industrial process.

The major ingredient to ensure full digitalization and automation is the virtualization of control, allowing to centralize all the intelligence of the operations in order to increase flexibility and facilitate the changes of the manufacturing plants. Moreover, it is essential to monitor all the elements of an industrial manufacturing plant through wireless connectivity (in order to avoid cabling that further increases complexity) and information processing (including big data and analytics technologies). These enhanced functionalities introduce strict requirements on data rates, latency, reliability, etc., all of which are addressed in the 5G mobile transport and computing platform.

The role of 5G in industry 4.0 extends to large area logistics (i.e., in the optimization of maritime, ground, air transportations, as well as to optimize port operations and goods production processes) where there is a similar need to increase the productivity and the efficiency of the processes to cut production costs and become more and more competitive.

In D1.1 [1] several use cases have been identified for the e-Industry vertical, namely monitoring in production line, cloud robotics, automated logistics, electric power generation, electric power transmission and electric power distribution. Several use cases can coexist in different scenarios. For example, in an automated factory both monitoring application and cloud robotics solutions can be in use. All use cases presented in D1.1 involve enabling more efficient manufacturing and lean production which poses severe requirements on the underlying communication network making it essential that the industrial environment be equipped with 5G solutions.

Among all the e-Industry use cases, the cloud robotics has been selected as the candidate for implementation in the final demonstrators of 5G-TRANSFORMER project.

TABLE 11: EINDUSTRY USE CASE

ID	Goal In context	General description
UC I.02 Cloud robotics	High automation of the factory plant is provided moving the control of the production processes and of the robots' functionalities into cloud, exploiting wireless connectivity to minimize infrastructure, optimize processes, implement lean manufacturing.	The controlling functionality of the robots is moved to the cloud, in order to utilize its massive computing power. Huge amounts of information will have to be transferred instantaneously. With lower latency and higher bandwidth than other forms of wireless connectivity, 5G is the optimal choice.

9.5 MNO/MVNO

Increasing the capacity and the elasticity of mobile network operators' networks is one of the most important challenges foreseen in 5G networks, as it will allow opening MNOs business toward new markets and a large variety of tailored services. This evolution is brought through the convergence of mobile networks and cloud infrastructures, which provides the capability for mobile operators to use network function virtualization (NFV) concepts and cloud-based infrastructures in order to virtualize and decentralize their network entities. Hence, the MVNO business model emerges from this evolution through the creation of a new business model that disrupts the traditional mobile value chain. In the MVNO model, new players can participate in the mobile value chain and extract value to leverage their valuable assets.
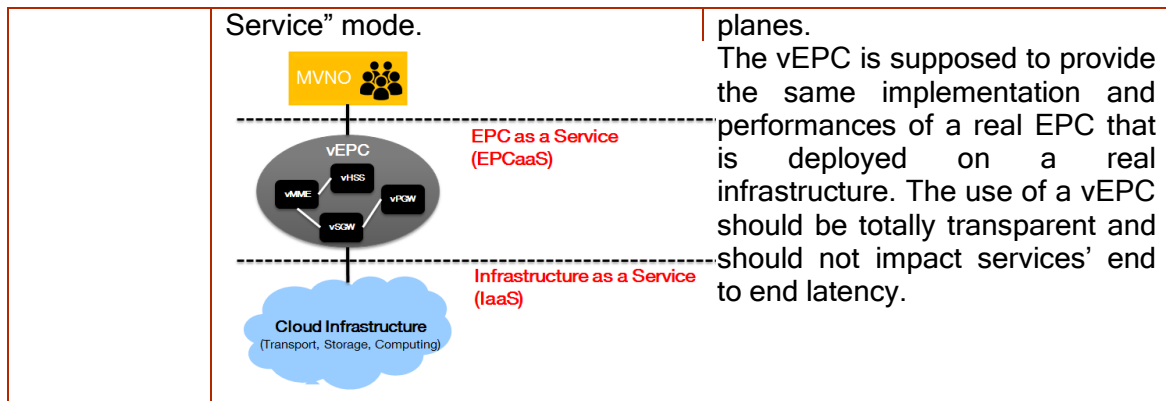
In 5G-TRANSFORMER, the MNO/MVNO industry is especially relevant and interesting because of the new roles it injects into the mobile value chain as well as the nature of services it offers compared to the other studied vertical industries in the project. For instance, offering the Network as a Service (NaaS) or the Infrastructure as a Service (IaaS), which are types of services that are challenging for MNOs and MVNOs, in order to reach real on demand and finely tailored services for their customers.

Thus, the MNO/MVNO player has a different role compared to the other verticals. In fact, the relation between the MNO/MVNO and the 5G-TRANSFORMER system depends on the chosen MVNO business model. For instance, in the case of a Full MVNO or an MVNE business model as described in D1.1 [1], the role of the MVNO exceeds that of a simple vertical service provider and has almost the same role as an MNO acting as a Network service provider. Likewise, the role of the MNO hosting an MVNO is built on the offering of a Network as a Service (NaaS); for instance, the MNO would rely on network slicing combined with services like EPCaaS and IaaS in order to set up an MVNO network and provide Network services like connectivity to the MVNO. In addition, verticals can be seen as customers of an MNO or an MVNO.

In [1], several use cases have been identified as relevant for the MNO/MVNO domain in 5G-TRANSFORMER. We chose to focus here on the use case UC M.01 vEPCaaS. This use case describes how the MNO/MVNO can offer Network as a Service (NaaS) for its customers by offering a dedicated and on demand core network. In the same context of offering Network as a Service, we also investigate the particular case of NFVIaaS as an example of IaaS. In this case of IaaS the challenge resides in the fact that the correspondent MVNO business model may imply the ownership by the MNO/MVNO of an OSS/BSS that provides the capability to create and configure its own network slice instances for its customers and the non-ownership of an NFVI infrastructure. In this case, 5G-TRANSFORMER will offer NFVIaaS for the MNO/MVNO. One possible question in this case is whether it is possible for an MNO/MVNO to request a network slice with specific Network Functions Virtualization Infrastructure (NFVI) resources, which would be possible through an interface or service catalogue.

TABLE 12: MNO/MVNO USE CASES

ID	Goal in context	General description
UC M.01 vEPCaaS	Build of an MVNO service through the deployment and operation of a network slice with a vEPC in "as a	The vEPC can be instantiated as a virtualized Control plane only or as virtualized Control and User

	Service" mode.	planes.
	MVNO EPC as a Service (EPCaaS) vEPC vHSS vMME vPGW vSGW Infrastructure as a Service (IaaS) Cloud Infrastructure (Transport, Storage, Computing)	The vEPC is supposed to provide the same implementation and performances of a real EPC that is deployed on a real infrastructure. The use of a vEPC should be totally transparent and should not impact services' end to end latency.

10 Annex II: Reference architectures

A few Standard Development Organizations (SDOs) and fora are contributing to the design of management systems for 5G that have many common design principles: (i) flexibility, (ii) adaptability, and (iii) cost-efficiency. Despite the many common design objectives across these working groups (as presented below), there are various relevant architectural concepts (e.g., slicing, federation/multi-domain, edge computing) that are specific to individual groups. In this context, 5G-TRANSFORMER strives to bridge the gaps across such heterogeneous ecosystem in order to harmonically integrate these concepts under a single architecture.

This section introduces ongoing architectural work at 3GPP and ETSI NFV and MEC that fulfills two objectives:

- serve as inspiration to define the 5G-TRANSFORMER architecture;
- set the framework in which 5G-TRANSFORMER must be integrated to maximize its impact, i.e., by seeking as much as possible compliance with what is already defined.

Despite the fact that there are other organizations discussing about slicing and architectural concepts related with 5G-TRANSFORMER, we focus on the ones below because they are the ones with a more complete definition of their architecture and building blocks, and so, they go well beyond requirements and high-level concepts.

10.1 3GPP

The most relevant working groups inside 3GPP related to 5G-TRANSFORMER are SA2 (Architecture) and SA5 (Telecom management).

10.1.1 3GPP SA2

The 3GPP SA2 Working Group (WG), responsible for overall system architecture, is currently working on specifying the 5G Core (5GC) architecture with network slicing being a main feature of 5GC. Technical Specification (TS) 23.501 [51] defines Stage-2 system architecture for the 5G system which includes network slicing. **¡Error! No se encuentra el origen de la referencia.** depicts an example of network slicing from 3GPP's perspective.
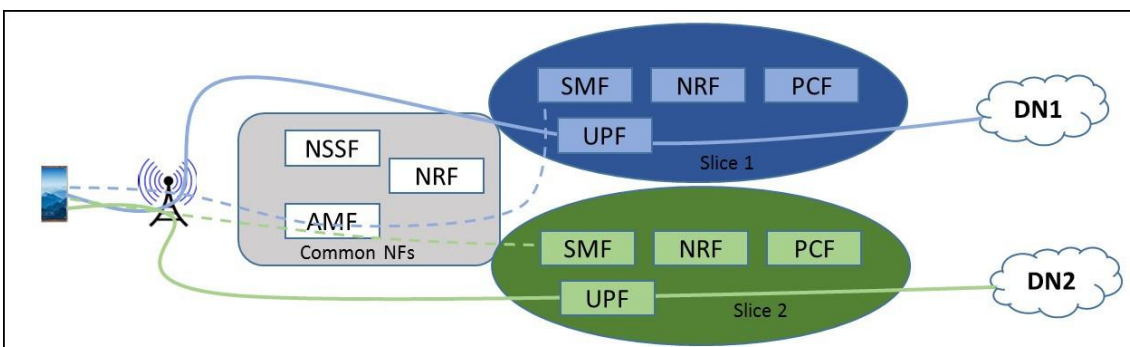

FIGURE 33: EXAMPLE OF NETWORK SLICES FROM 3GPP SA2 PERSPECTIVE

A network slice is viewed as a logical end-to-end network that can be dynamically created. A given User Equipment (UE) may access to multiple slices over the same Access Network (e.g. over the same radio interface). Each slice may serve a particular

service type with agreed upon SLA. In the following, we provide highlights of 3GPP network slicing as being defined in TS 23.501 [51] in SA2.

A network slice is defined within a Public Land Mobile Network (PLMN) and includes the Core Network Control Plane and User Plane Network Functions as well as the 5G Access Network (AN). The 5G Access Network may be a Next Generation (NG) Radio Access Network described in 3GPP TS 38.300 [52] or a non-3GPP Access Network.

TS 23.501 [51] defines Network Function, Slice, and Slice Instance as follows:

- Network Function: A 3GPP adopted or 3GPP defined processing function in a network, which has defined functional behavior and 3GPP defined interfaces. (Note: a network function can be implemented either as a network element on a dedicated hardware, as a software instance running on a dedicated hardware, or as a virtualized function instantiated on an appropriate platform, e.g. on a cloud infrastructure.);
- Network Slice: A logical network that provides specific network capabilities and network characteristics;
- Network Slice instance: A set of network function instances and the required resources (e.g. compute, storage and networking resources) which form a deployed network slice.

10.1.2 3GPP SA5

3GPP SA5 Working Group (WG) is the 3GPP telecom management working group. 3GPP SA5 specifies the requirements, architecture and solutions for provisioning and management of the network, including Radio Access Network (RAN) and Core Network (CN) and its services.

SA5 has completed a study on management and orchestration on network slicing (3GPP 28.801 [6]) and started the normative specification work for release 15 based on this study. It is expected to be completed by the second quarter of 2018, including:

- Network slice concepts, use cases and requirements (3GPP 28.530 [53]);
- Provisioning of network slicing for 5G networks and services (3GPP 28.531 [54]);
- Assurance data and performance management for 5G networks and network slicing;
- Fault supervision for 5G network and network slicing.

The following description highlights management and orchestration aspects of network slicing in 3GPP 28.801 [6]. However, these may be updated in the SA5 normative specifications based on the ongoing development of the SA2 technical specifications.

- General management and orchestration aspects of network slicing defined in 3GPP 28.801 [6]. Based on 3GPP 23.501 [54], SA5 has defined different management aspects for network slices in 3GPP 28.801 [6] as listed below:
  - Managing a complete Network Slice Instance (NSI) is not only managing all the functionalities but also the resources necessary to support certain set of communication services.
  - An NSI not only contains Network Functions (NFs), e.g belonging to AN and CN, but also the connectivity between the NFs. If the NFs are interconnected, the 3GPP management system contains the information relevant to connections between these NFs such as topology of

connections, individual link requirements (e.g. QoS attributes), etc. For the part of the Transport Network (TN) supporting connectivity between the NFs, the 3GPP management system provides link requirements to the management system that handles the part of the TN supporting connectivity between the NFs.
- o NSI can be composed of network slice subnets of physical network functions and/or virtualized network functions.
- Network Slice Instance lifecycle management. 3GPP 28.801 [6] has introduced the network slice instance lifecycle management as depicted below in **¡Error! No se encuentra el origen de la referencia.**, considering it independent of the network service instance which is using the network slice instance. Typically, a network slice instance is designed (preparation phase), then it is instantiated (Instantiation, Configuration and Activation phase), then it is operated (Run Time phase) and finally it may be decommissioned when the slice is no longer needed (Decommissioning phase). 3GPP 28.801 [6] introduces 3 management logical functions:
  - o Communication Service Management Function (CSMF): Responsible for translating the communication service related requirement to network slice related requirements.
  - o Network Slice Management Function (NSMF): Responsible for management and orchestration of NSI and derive network slice subnet related requirements from network slice related requirements.
  - o Network Slice Subnet Management Function (NSSMF): Responsible for management and orchestration of network slice subnet instances (NSSI).
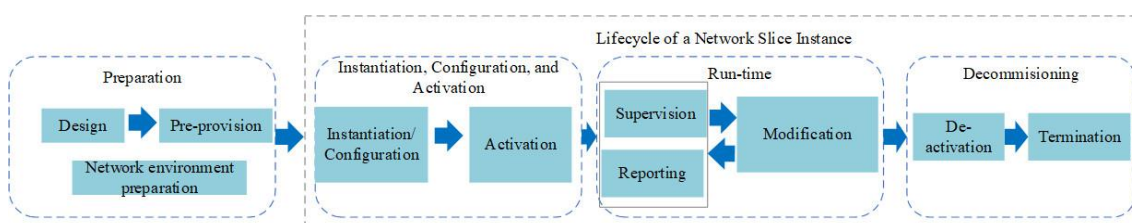
FIGURE 34: 3GPP VIEW ON NETWORK SLICE INSTANCE LIFECYCLE

10.2 Radio Access Network (RAN)

10.2.1 LTE RAN

LTE uses Orthogonal Frequency Division Multiplexing (OFDM) that transmits the data over many narrowband carriers of 180kHz each. Instead of a single fast transmission, a data stream is split into many slower data streams that are transmitted simultaneously. As a consequence, the attainable data rate compared, for example to the previous 3G Radio Access (e.g. UMTS), is similar in the same bandwidth but the multipath effect is greatly reduced because of the longer transmission steps.

Several bandwidths have been specified for LTE: from 1.25MHz up to 20MHz. All LTE devices must support all bandwidths, and which one is used in practice depends on the frequency band and the amount of spectrum available to a network operator.

Unlike previous radio standards, the baseline for LTE device has been set very high. In addition to the flexible bandwidth support, all LTE devices must support Multiple Input

Multiple Output (MIMO) transmissions, a situation which allows the base station to transmit several data streams over the same carrier simultaneously.

Another major change of LTE compared to previous systems has been the adoption of an all-Internet Protocol (IP) approach. While UMTS used a traditional circuit-switched packet core for voice services, for Short Messaging Service (SMS) and other services inherited from GSM, LTE solely relies on an IP-based core network.

Also, all interfaces between network nodes in LTE are now based on IP, including the backhaul connection to the radio base stations. Again, this is a great simplification compared to earlier technologies that were initially based on E-1, ATM and frame relay links, with most of them being narrowband and expensive.

The standard leaves the choice of protocols to be used below the IP layer open, which means that the physical infrastructure becomes completely transparent and interchangeable. To further simplify the network architecture and to reduce user data delay, fewer logical and physical network components have been defined in LTE. In practice, this has resulted in round-trip delay times of less than 25-30 milliseconds.

Optimized signalling for connection establishment and other air interface and mobility management procedures have further improved the user experience. The time required to connect to the network is in the range of only a few hundred milliseconds and power-saving states can now be entered and exited very quickly.

10.2.2 5G RAN

5G radio access technology will be a key component of the Networked Society. It will address high traffic growth and increasing demand for high-bandwidth connectivity. It will also support a massive number of connected devices and meet the real-time, high-reliability communication needs of mission-critical applications.

5G will provide wireless connectivity for a wide range of new applications and use cases, including wearables, smart homes, traffic safety/control, critical infrastructure, industry processes and very-high-speed media delivery. As a result, it will also accelerate the development of the Internet of Things.

The overall aim of 5G is to provide ubiquitous connectivity for any kind of device and any kind of application that may benefit from being connected.

A key point is that LTE will evolve in a way that recognizes its role in providing excellent coverage for mobile users, and 5G networks will incorporate LTE access (based on Orthogonal Frequency Division Multiplexing (OFDM)) along with new air interfaces in a transparent manner toward both the service layer and users.

5G wireless access must extend far beyond those of previous generations of mobile communication. These capabilities will include massive system capacity, very high data rates everywhere, very low latency, ultra-high reliability and availability, very low device cost and energy consumption, and energy-efficient networks.

5G should support data rates exceeding 10Gb/s in specific scenarios, such as indoor and dense outdoor environments, but hundreds of megabits per second data rates should be guaranteed to everyone, in any radio channel conditions and for any radio deployment.

Some envisioned 5G use cases, such as traffic safety and control of critical infrastructure and industry processes, may require much lower latency compared with what is possible with the mobile-communication systems of today.

To support such latency-critical applications, 5G should allow for an application end-to-end latency of 1ms or less, although application-level framing requirements and codec limitations for media may lead to higher latencies in practice. Many services will distribute computational capacity and storage close to the air interface

10.3 Core Network

10.3.1 LTE core network functions

The EPS system is made up of the Evolved Packet Core (EPC) and the E-UTRAN.

The EPC provides access to external data networks (e.g., Internet, Corporate Networks) and operator services (e.g., MMS, MBMS). It also performs functions related to security (authentication, key agreement), subscriber information, charging and inter-access mobility. The core network also tracks the mobility of inactive terminals (i.e., terminals in power saving states).

E-UTRAN, formed by base stations (eNodeB) and user equipment (UE) performs all radio related functions for active terminals.

Between the EPC and E-UTRAN there is an interface called S1 and between the eNodeBs there is an interface called X2.

Like previous Radio Access Technologies (e.g. UMTS), EPS supports a bearer concept for end-user data services. The EPS Bearer is defined between the User Equipment (UE) and the P-GW node in the EPC (which provides the end users IP point of presence towards external networks). The EPS bearer is further sub-divided into an E-UTRAN Radio Access Bearer (E-RAB) over the radio interface and S1 between the UE and SGW, and an S5/S8 bearer between S-GW and P-GW (S8 when S-GW and P-GW belong to different operators). End-to-end services (e.g. IP services) are multiplexed on different EPS Bearers.

There is a many-to-one relation between End-to-end services and EPS Bearers. The concept of bearers allows to manage in a segregated way and on several layers, many point-to-point virtual connections to and within the core network. Each bearer has its own class of service and its own priority, according to the offered service and purpose.

More information can be found in [34].

10.3.2 5G core network functions

Network applications such as Evolved Packet Core (EPC), voice over LTE (VoLTE), and future 5G core network functions will be cloud enabled: that is, they will have the ability to execute in the SDN/NFV cloud environment. Consequently, the applications will have the advantage of being automatically scalable as well as flexible in terms of where in the network they can be deployed (centrally, distributed or a combination of the two).

Cloudification of 5G core functions enriches and, in some cases, exceeds the concept of bearers in LTE.

For example, the complete core network can be deployed in a local server in a factory to support exceptionally short response times. At the same time, it should be possible

to support the factory with communication services from a centrally placed VoLTE installation.

Moving 5G Core functions to the cloud allows infrastructure to scale in or out and automatically; in other words, when an application needs more resources, the cloud automatically spins up another instance of that application and removes an instance when load decreases. Such flexible scaling is impossible to achieve when the application is implemented with dedicated hardware.

In addition to this, the infrastructure application may be deployed in a virtual data centre (VDC) that is distributed between many physical data centres. One way to simplify provisioning and control of an application across a VDC will be to define a network slice as a virtual network and let the owner of the slice manage it themselves, within constraints placed by the network provider.

This is essential, not only to have a more efficient core network, but also to ensure the extreme reliability of many 5G applications, such as autonomous driving or industrial automation and, at the same time, provide multiple instances of the same core network (slices) that fit to the various 5G use-cases.

More information can be found in [35], [36].

10.4 Logical Link Abstraction

This section is focused on the abstraction for a logical link in optical networks in relation with the southbound interface between 5G-MTP and the SDN controller, as shown in Table 13. The description of a logical link in optical networks has been elaborated in conjunction with the EU H2020 ORCHESTRA project [22] and can be assumed as an innovation of the 5G-TRANSFORMER project. As mentioned before, it is expected that 5GT-MTP, which is responsible for trigging the SDN-controller for logical link set up, keeps track of the whole set of information. Among all the parameters, for example, the set includes information about the used transponders and also the used reconfigurable optical add and drop multiplexers (ROADMs), since a node can have more ROADM modules, e.g. one per port, and information related to the rate. In an optical network, the available rate information is useful because a given transponder could be underutilized. For instance, a flexible optical transponder can support 200Gb/s PM-16QAM to serve 2x100GbEthernet interfaces and 100Gb/s PM-QPSK to serve 1x100GbEthernet [23]. Thus, if the IP layer only requests for a 100Gb/s, the transponder is underutilized for 100Gb/s. This means that the logical link has a spare rate of 100Gb/s.

TABLE 13: EXAMPLE OF INFORMATION MODELLING FOR A LOGICAL LINK IN OPTICAL NETWORKS

Identifier
Source node (IP address)
Destination node (IP address)
Path in the network (i.e., list of traversed nodes and links)
Latency
Allocated spectrum, defined with a central frequency and a width of bandwidth
Maximum supported bitrate
Used bit rate (or available bit rate)
Identifier of the used transponder at the source node
Identifier of the used transponder at the destination node
Identifier of the used ROADMs

Figure 35 shows the tree of a YANG data model for logical link in optical networks in relation with the southbound interface between 5G-MTP and the SDN controller. In particular, the tree shows that an array of logical links, denoted by the operator "[id]" of type "logical-link-id-type", which has been defined in the code as "uint64".

```
+--rw logical-links
   +--rw logical-link [id]
      +--rw id                 logical-link-id-type
      +--rw max-rate           bit-rate-type
      +--rw available-rate     bit-rate-type
      +--rw latency            latency-type
      +--rw src-node           node-id-type
      +--rw dst-node           node-id-type
      +--rw frequency-slot
      |  +--rw nominal-central-frequency-granularity?   frequency-ghz-type
      |  +--rw slot-width-granularity?                   frequency-ghz-type
      |  +--rw n                                         int16
      |  +--rw m                                         int16
      +--rw nodes
      |  +--rw node [id]
      |     +--rw id                 node-id-type
      |     +--rw ip?                inet:ip-address
      |     +--rw type               node-type
      |     +--rw transponders
      |     |  +--rw transponder [id]
      |     |     +--rw id                        transponder-id-type
      |     |     +--rw subcarrier-modules
      |     |        +--rw subcarrier-module [id]
      |     |           +--rw id     subcarrier-id-type
      |     +--rw roadms
      |        +--rw roadm [id]
      |           +--rw id     roadm-id-type
      +--rw links
         +--rw link [id]
            +--rw id     link-id-type
```

FIGURE 35: YANG TREE REPRESENTATION OF LOGICAL LINK IN OPTICAL NETWORKS

11 Annex III: Notation for Requirements

In this deliverable we follow – with slight adaptions – the notation for requirements used already in [1]. For each requirement, the following fields should be provided:

ID	Requirement	F/NF
ReqX.XX	e.g. The vehicle shall be connected to a 5G router	F/NF

The meanings of the fields are as follows:

➢ **ID**: is the identifier of the requirement (written in the form ReqX.XX).
➢ **Requirement**: a complete sentence explaining the requirement.
➢ **F/NF**: if the requirement is Functional (F) or Non Functional (NF).

NOTE: The requirement field is written following the approach followed by IETF documents, included next. The key words "must", "must not", "required", "shall", "shall not", "should", "should not", "recommended", "may" and "optional" in this document are to be interpreted as described in [44].

1. **MUST** This word, or the terms "**REQUIRED**" or "**SHALL**", mean that the definition is an absolute requirement of the specification.
2. **MUST NOT** This phrase, or the phrase "**SHALL NOT**", mean that the definition is an absolute prohibition of the specification.
3. **SHOULD** This word, or the adjective "**RECOMMENDED**", mean that there may exist valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighted before choosing a different course.
4. **SHOULD NOT** This phrase, or the phrase "**NOT RECOMMENDED**" mean that there may exist valid reasons in particular circumstances when the particular behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
5. **MAY** This word, or the adjective "**OPTIONAL**", mean that an item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because the vendor feels that it enhances the product while another vendor may omit the same item. An implementation which does not include a particular option MUST be prepared to interoperate with another implementation which does include the option, though perhaps with reduced functionality. In the same vein, an implementation which does include a particular option MUST be prepared to interoperate with another implementation which does not include the option (except, of course, for the feature the option provides).

12 Annex IV: State of the Art and challenges related to 5GT-MTP

12.1 ETSI NFV, interfaces, databases, information model, and data model

In the interaction between 5GT-SO and 5GT-MTP an important role is taken by the interfaces between the two layers. Thus, some of the interfaces already defined in ETSI NFV MANO are relevant. They are, specifically, the ETSI-MANO reference architecture [8], IFA013 [42], IFA005 [24], and IFA006 [44]. In addition, the interfaces between the VIM and the NFVI, defined other than in the ETSI-MANO reference architecture [8], are defined in ETSI GS NFV-INF 001 [7] and in the related infrastructure documents (i.e., ETSI GS NFV-INF 003 [59], ETSI GS NFV-INF 004 [60], and ETSI GS NFV-INF 005 [61]).

In the ETSI-MANO reference architecture, depicted in Figure 37, some interfaces are defined for the interaction between NFVO and VIM/WIN (i.e., Or-Vi reference point), the interaction between VNFM and VIM (i.e., Vi-Vnfm reference point), and the interaction between VIM and the NFVI (i.e., Nf-Vi reference points). Many of the interfaces related to the Or-Vi and Vi-Vnfm reference points are the same. Indeed, the Or-Vi reference point is used for exchanges between the NFV Orchestrator and the VIM and supports the following: NFVI resource reservation/release, NFVI resource allocation/release/update, VNF software image addition/deletion/update, forwarding of configuration information, events, measurement results, and usage records regarding NFVI resources to the NFV Orchestrator. The Vi-Vnfm reference point is used for exchanges between the VNFM and the VIM and supports the following: NFVI resources reservation information retrieval, NFVI resources allocation/release, exchanges of configuration information between reference point peers, and forwarding to the VNF Manager such information for which the VNFM has subscribed (e.g. events, measurement results, and usage records regarding NFVI resources used by a VNF). In addition, an *NFVI resource repository/catalogue/database,* depicted in Figure 36, is defined as a holder of information about available/reserved/allocated NFVI resources as abstracted by the VIM across operator's Infrastructure Domains, thus supporting information useful for resources reservation, allocation and monitoring purposes.
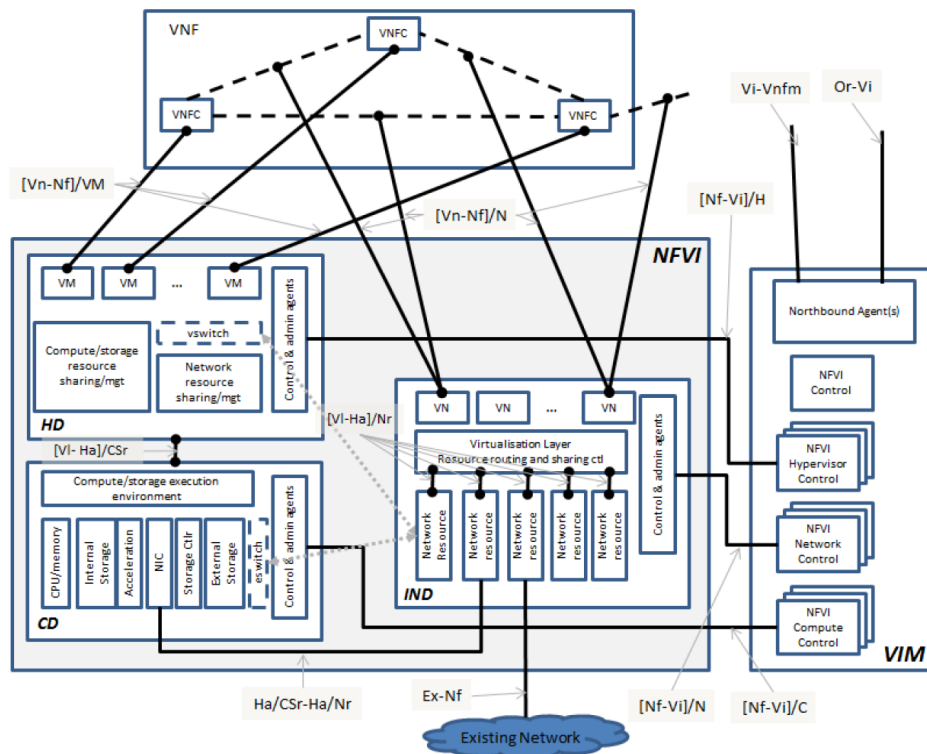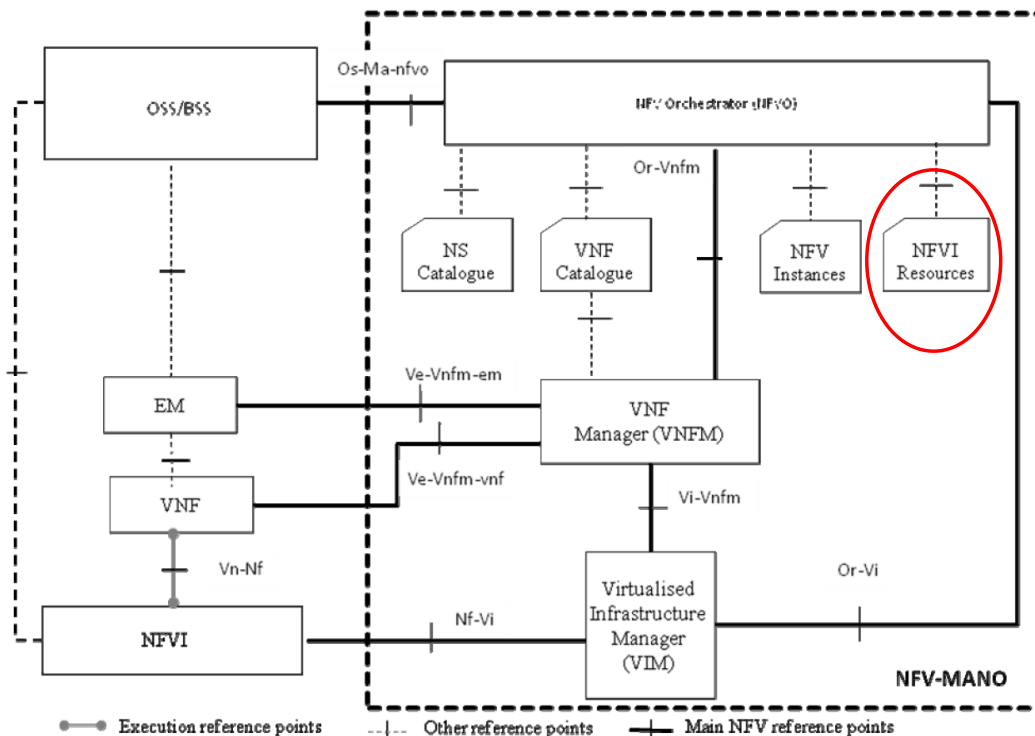
FIGURE 36: NFVI RESOURCE DATA BASE

FIGURE 37: ETSI INFRASTRUCTURE OVERVIEW

IFA005 further details the interfaces supported over the Or-Vi reference point of the NFV MANO architectural framework as well as the information elements exchanged over those interfaces. IFA006 further details the interfaces supported over the Vi-Vnfm

reference point of the NFV MANO architectural framework as well as the information elements exchanged over those interfaces. ETSI GS NFV-INF 001 [7] and the related infrastructure documents (i.e., ETSI GS NFV-INF 003 [59], ETSI GS NFV-INF 004 [60], and ETSI GS NFV-INF 005 [61]) further define interfaces for the Nf-Vi reference points.

In the ETSI-MANO reference architecture, interfaces concerning *virtualized resources* are defined and they can be classified in information interfaces, resource management interfaces, performance management interfaces, and fault management interfaces. The information interfaces are the *Virtualized resources catalogue management* interface and the *Virtualized resources capacity management* interface. The *Virtualized Resources Catalogue Management* interface is common between the Or-Vi and the Vi-Vnfm reference points. It allows an authorized consumer functional block to query the catalogues of virtualized resources and get notifications about their changes. The related information is produced by the VIM and consumed by the NFVO and the VNFM. The *Virtualized Resources Catalogue Management* interface includes operations such as "Query Resource Catalogue" and "Notify Resources Catalogue Changes". The *Virtualized resources capacity management* interface is exclusive of the Or-Vi reference point and allows an authorized consumer functional block to perform operations related to NFVI-PoP capacity and usage reporting. The interface allows retrieving information about: NFVI-PoP total resources capacity over which virtualized resources are provisioned, Virtualized resources capacity and density (e.g. how many virtualized resources can be created from existing NFVI-PoP resources), statistics and mapping of NFVI-PoP resources usage to virtualized resources global usage at NFVI-PoP level, and per deployed virtualized partition. The interface enables the capture of information for reporting NFVI-PoP total resources usage and executing analytics for capacity planning, capacity changes, and consequently for Network Service planning. The information is produced by VIM and consumed by the NFVO.

In [8] another interface is defined for *Virtualized Resource Management* which is common for both Or-Vi and Vi-Vnfm reference points. This interface allows an authorized consumer functional block to perform operations on virtualized resources available to the consumer functional block. The interface includes common operations for creating, querying, updating and terminating compute, storage and network isolated virtualized resources, or a composition of different types in a resource grouping, as well as managing virtualized resource reservations. For example, the "Scale Resource" operation allows scaling a virtualized resource by adding or removing capacity, e.g. adding vCPUs to a virtual machine. For this interface, VIM and NFVO can be information producers whilst the NFVO and VNFM are information consumers.

In [8] other interfaces are defined for *Virtualized resources performance management* and *Virtualized resources fault management*. They are both in common between the Or-Vi and the Vi-Vnfm reference points. The *Virtualized resources performance management* interface provides performance management (measurement results collection and notifications) related to virtualized resources including, but not limited to resource consumption level, e.g. vCPU power consumption, VM memory usage oversubscription, VM disk latency, etc. The *Virtualized resources fault management* interface provides fault information related to the resources visible to the consumer functional block, including virtual containers (VMs) crashes, virtual network ports errors, virtual container's to storage disconnection, etc. The interface also provides information about faults related to the pools of resources, for instance, reserved resources unavailable, resource exhaustion, etc.

In addition, other interface are defined in [8] concerning *VNF Software Image Management* (common between Or-Vi and Vi-Vnfm), *Policy administration* (Or-Vi only), *Network Forwarding Path management* (Or-Vi only). This latter interface provides the facility to have policy-based linkages on a VNF Forwarding Graph as expressed by a Network Forwarding Path. Here, the information producer is the VIM and the information consumer is the NFVO.

The following interfaces towards the NFVI, defined in [8] as well, are also of interest for the 5GT-MTP. The *NFVI hypervisor management* interface allows an authorized consumer functional block to request a producer functional block to perform operations on hypervisor-accessed resources (e.g. compute, storage and networking) in the NFVI. The information producer is the NFVI while the information consumer is the VIM. This interface needs to be exposed by NFV participating hypervisors. This interface maps to the Nf-Vi/H interface. The *NFVI Compute Management* interface allows an authorized consumer functional block to request a producer functional block to perform management operations on physical compute and storage resources in the NFVI. In [8], it is assumed that NFV-MANO supports the use of this interface in order to provide notifications from NFVI to the VIM regarding changes in physical resources (e.g. fault information, inventory information), without defining the specific interface operations. However, the description and operations of the interface are not complete. This interface maps to the Nf-Vi/C interface. The information is produced by the NFVI and consumed by the VIM. The *NFVI networking management* interface allows an authorized consumer functional block to request a producer functional block to perform management operations on networking resources in the NFVI. This interface focuses on functionality exposed by network (SDN) controllers comprised in the NFVI, via appropriate abstractions. In this case, the information producer is the NFVI while the information consumer is the VIM. This interface maps to the Nf-Vi/N interface. A sample operation that can be performed through this interface is the "Create virtual network" that creates virtual networks (L2/L3 overlay or infrastructure) for inter-VNF or inter-VNFC interconnectivity.

However, in [8], neither an information model nor a data model are provided. In [8] a sample information model for the Virtual Network Function Descriptor (VNFD) and Virtual Link Descriptor (VLD) and a sample data model based on Yang are provided that can be utilized for the aforementioned interfaces.

Within ETSI, ETSI/NFV IFA005 better specifies the interfaces defined in [8] at the Or-Vi reference point and also specifies an information model. VIM exposed interfaces are divided in: Software Image Management Interface for managing Virtual Machine (VM) images; Virtualized Compute Interfaces, dealing with the virtualized compute resources, Virtualized Network Interfaces dealing with virtualized network resources; and virtualized storage interfaces dealing with virtualized storage.

For each category, interfaces are defined for: providing information about resources (e.g., *Virtualized Compute Resources Information Management Interface*, *Virtualized* and *Compute Resources Capacity Management Interface*); managing (e.g., reserving, deleting, migrating, etc.) resources (e.g., *Virtualized Compute Resources Management Interface* and *Virtualized Compute Flavour Management Interface*), and notification (*Virtualized Compute Resources Change Notification Interface*).

In addition, interfaces are defined for *Virtualized Resource Fault Management Interface*, *Virtualized Resources Performance Management Interface*, *Virtualized Resource Reservation Interfaces*, and *Virtualized Resource Quota Interfaces*.

Moreover, information models for the information elements exchanged are provided. They are classified as: *Information elements related to software images*, *Information elements and notifications related to Consumable Virtualized Resources Information*, *Information elements and notifications related to Virtualized Resources*, *Information elements and notifications related to Virtualized Resources Performance Management*, *Information elements and notifications related to Virtualized Resources Fault Management*, *Information elements and notifications related to Virtualized Resources Capacity Management*, *Information elements and notifications related to Reservation*, *Network Forwarding Path (NFP) information element*, *Information elements related to NFVI-PoP*, and *Information elements and notifications related to Quota*. For example, the "Virtual Compute Resource Information element" is reported in Figure 38.

Attribute	Qualifier	Cardinality	Content	Description
computeResourceTypeId	M	1	Identifier	Identifier of the consumable virtualised compute resource type.
virtualMemory	M	0..1	VirtualMemoryResourceInformation	It defines the virtual memory characteristics of the consumable virtualised compute resource (see note).
virtualCPU	M	0..1	VirtualCpuResourceInformation	It defines the virtual CPU(s) characteristics of the consumable virtualised compute resource (see note).
accelerationCapability	M	0..N	AccelerationCapability	Acceleration capabilities (e.g. crypto, GPU) for the consumable virtualised compute resources from the set of capabilities offered by the compute node acceleration resources. The cardinality can be 0, if no particular acceleration capability is provided (see also
Attribute	Qualifier	Cardinality	Content	Description
				note).
NOTE: Cardinality can be "0" if the attribute refers to a characteristic that is not being reported on a specific query or notification, e.g. through a InformationChangeNotification.				

FIGURE 38: VIRTUAL COMPUTE RESOURCE INFORMATION ELEMENT

The information model for the "Virtual Network Resource Information" is depicted in Figure 39.

Attribute	Qualifier	Cardinality	Content	Description
networkResourceTypeId	M	1	Identifier	Identifier of the network resource type.
bandwidth	M	1	Number	Minimum network bandwidth (in Mbps).
networkType	M	0..1	String	The type of network that maps to the virtualised network. Examples are: "local", "vlan", "vxlan", "gre", etc.
networkQoS	M	0..N	NetworkQoS	Element providing information about Quality of Service attributes that the network shall support. See clause 8.4.4.3.

FIGURE 39: VIRTUAL NETWORK RESOURCE INFORMATION

In [29] and [30] NFV descriptors based on TOSCA specification and on YANG specification are reported, respectively.

Similarly, ETSI defines in IFA006 [44] the interface between VNFM and VIM (i.e., the Vi-Vnfm reference point) and an information model for it. This interface is used by the VNFM to retrieve information about resources capability (e.g., maximum CPU), to allocate and reserve a resource, and to monitor a given performance metric with the object of maintaining a vertical service. IFA005 and IFA006 specify very similar interfaces as highlighted in Figure 40.

Interface	Resource database	Reservation	Maintenance	Present in IFA005	Present in IFA006
Virtualised Resources Information Management	x			yes	yes
Virtualised Resources Capacity Management	x			yes	no
Virtualised Resources Management	x			yes	yes
Virtualised Resources Change Notification	x			yes	yes
Virtualised Resources Quota Management	x	x		yes	yes
Virtualised Resources Quota Change Notification	x			yes	yes
Virtualised Resources Reservation Management		x		yes	yes
Virtualised Resource Reservation Change Notification		x		yes	yes
Virtualised Resources Performance Management			x	yes	yes
Virtualised Resources Fault Management			x	yes	yes

FIGURE 40: INTERFACES STANDARDIZED IN IFA005 AND IFA006

Moreover, ETSI defines in IFA008 [45] the interface between VNFM and VNF (i.e., the Ve-Vnfm-vnf reference point). This interface is utilized for exchanges between VNF and VNF Manager, and supports the following functions: VNF Lifecycle Management (produced by VNFM, consumed by VNF), VNF Performance Management, resulting from virtualized resource performance information (produced by VNFM, consumed by VNF), VNF Fault Management, resulting from virtualized resource fault information (produced by VNFM, consumed by VNF), VNF Indicator (produced by VNF, consumed by VNFM), and VNF Configuration (produced by VNF, consumed by VNFM).

12.2 ETSI MEC

Multi-access or Mobile Edge Computing (MEC) is about hosting computation and storage (Edge Node) close to end users. Typically, Edge Nodes are highly distributed in the mobile network, and typically located close to the eNodeBs. MEC enables two types of services: (i) low-latency services, which require that the end-user application accesses in a low latency way to the remote server and (ii) context-aware service, which needs to access end-user contexts, such as the user channel quality conditions, in order to adapt the delivered service. MEC is being standardized within ETSI, via the MEC ISG group. The first released document [38] defines the MEC reference architecture. A simplified version of this architecture is shown in Figure 41¡**Error! No se encuentra el origen de la referencia.**.
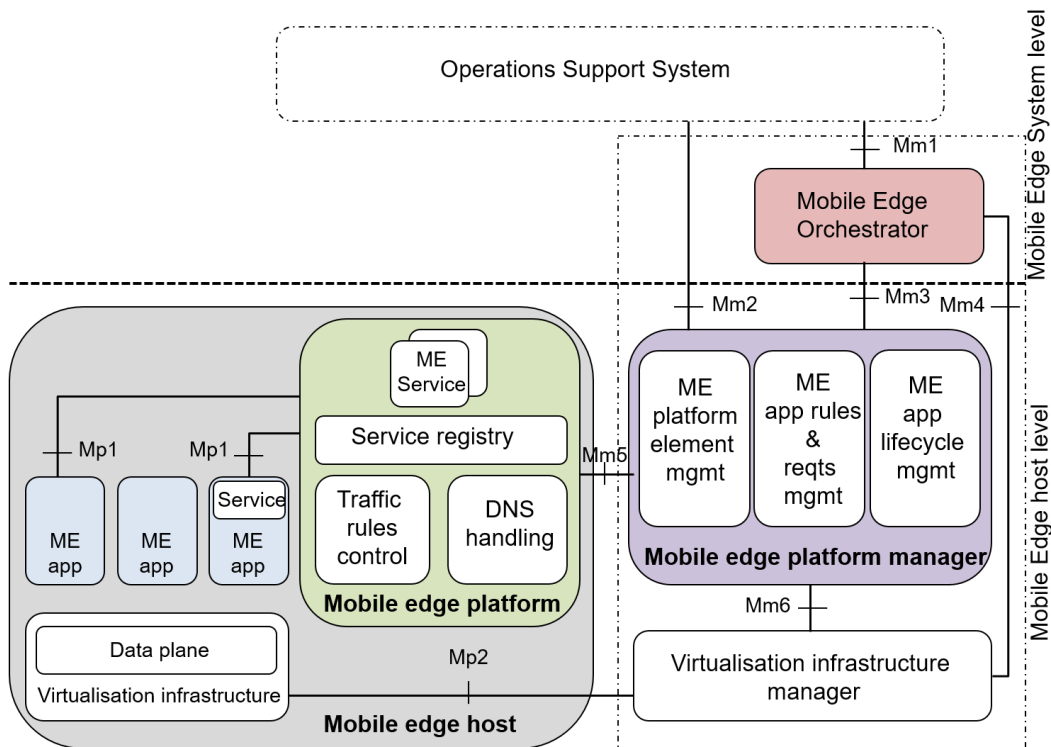
FIGURE 41: MEC REFERENCE ARCHITECTURE

The main components of the architecture include: the MEC host, the MEC application and the MEO (MEC orchestrator). The MEC host is the key element; it provides the virtualization environment to run MEC applications, while it interacts with the mobile network entities, via the MEC platform (MEP), to provide MEC services and offload mobile traffic to MEC applications. Respectively, Mp2 and Mp1 interfaces are used by the MEP to interact with the mobile network elements and provide MEC services, like the RNIS (Radio Network Information Service) API and traffic control to the MEC applications. In addition, two MEC hosts can communicate together via the Mp3 interface aiming at managing user mobility via MEC application migration among MEC hosts. Moreover, the MEP allows MEC applications to discover available MEC services available at the MEC host; it also allows to register a service provided by a MEC application.
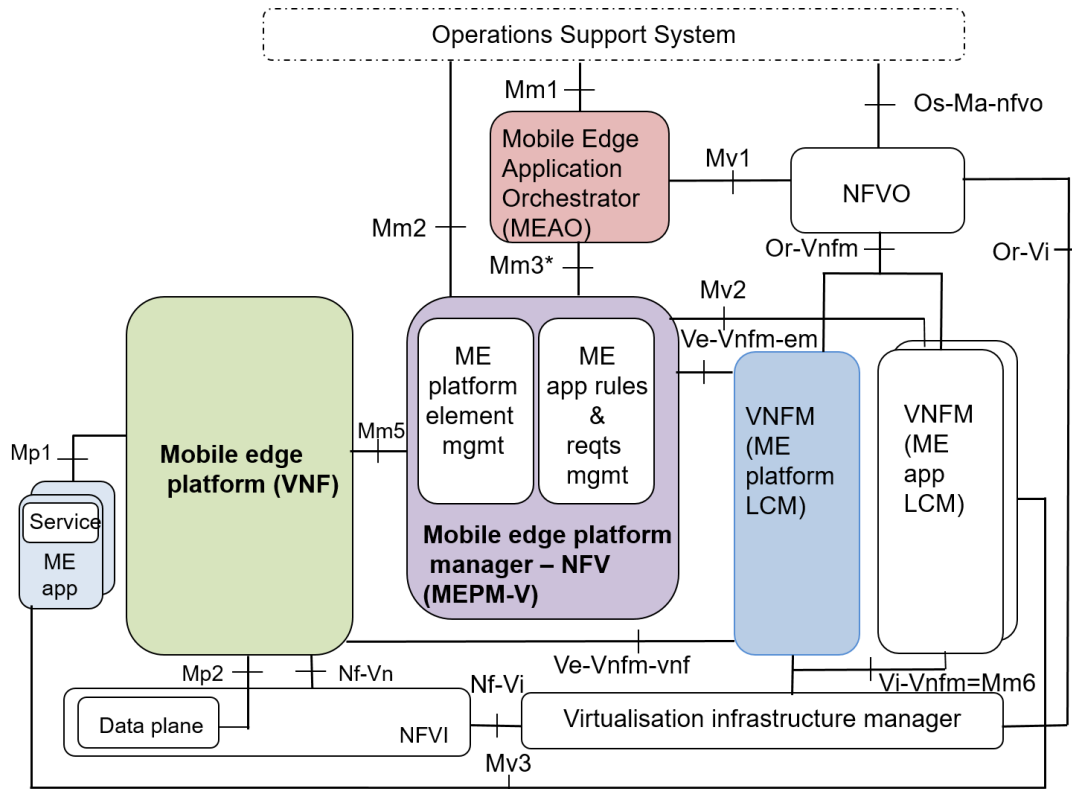
FIGURE 42: MEC REFERENCE ARCHITECTURE IN AN NFV ENVIRONMENT

The MEO is in charge of the orchestration and the instantiation of the MEC applications. Mm1 is used to communicate with the OSS/BSS in order to allow the latter to request the deployment of a MEC application, which is available in the catalogue. The MEO uses the Mm4 to upload image of the application into the edge VIM, which is in charge to instantiate the application on the MEC host. The MEPM element is in charge of the life-cycle management of the deployed MEC applications. The MEPM is in charge of the MEC platform configuration, via the Mm5, such as the MEC application authorization, the type of the traffic that needs to be offloaded to a MEC application, DNS redirection, etc. The MEPM uses the Mm3 interface to configure the MEC applications and communicate with the VIM to obtain information on the virtual resources used by a MEC application. This information is used by the MEO to check the MEC application resources status, and if deemed appropriate to decide if more resources are needed for the MEC applications. This information is also exposed to the OSS/BSS through the Mm2 interface.

As stated earlier, the MEC architecture is defined to run independently from the NFV environment. However, the advantage brought by NFV, and aiming to integrate and run all MEC entities in a NFV environment, has led the MEC ETSI group to update the reference architecture. The proposed draft [39] updates the reference architecture as shown in Figure 42**¡Error! No se encuentra el origen de la referencia.**. As could be noticed, the MEC platform and the MEPM are run as a VNF. The MEO became the MEAO; it keeps the main functions, except that it should use the NFVO to instantiate the virtual resources for the MEC applications as well as for the MEP. The MEC application life-cycle function has been moved to the VNFM. Moreover, the VNFM is in charge of the lifecycle management of the MEP as well as the MEPM. Table 14**¡Error! No se encuentra el origen de la referencia.** gives a comparison between the functions

of the MEO and MEAO. Another important difference between the reference architecture and the NFV-oriented one is the appearance of new interfaces (Mv1, Mv2 and Mv3) and the usage of interfaces defined by the ETSI NFV group.

TABLE 14: COMPARISON BETWEEN MEO AND MEAO

Function	MEO	MEAO
Maintaining an overview of the MEC, available resources, available MEC hosts, topology.	Yes	Yes
Selecting appropriate MEC host based on constraints (latency, available resources and services)	Yes	Yes
Triggering application instantiation and termination.	Yes	Via the NFVO
Triggering application relocation as needed when supported (migration due to mobility)		

12.3 EU 5G-PPP H2020 Projects

The design of the 5GT-MTP architecture draws mainly on the system architectures defined within the H2020 5G-Crosshaul project, which leverages the standard and reference specifications of the SDN and NFV architectures.

Mobile Transport platforms for 5G are being defined with the aim to support convergence between backhaul and fronthaul both in the data and control planes. Two H2020 5G-PPP phase-1 projects address this specific area, namely 5G-Crosshaul and 5G-XHaul. Whilst 5G-XHaul focuses more on the integration across heterogeneous transmission technologies such as optical and wireless, 5G-Crosshaul looks at the wider picture of the transport network at scale, from the edge up to the core, and develops the SDN/NFV tools for unified orchestration and control along with common forwarding plane across all underlying technologies.

The 5G-Crosshaul transport network integrates the backhaul and fronthaul segments of the network and implements a reconfiguration of all the networking elements allowing flexibility and software-defined reconfiguration in a multi-tenant and service-oriented unified management environment. The transport network includes high capacity switches and Remote Radio Heads interconnected by heterogeneous transmission links as well as 5GPoAs, cloud-processing units and points-of-presence of the core networks of one or multiple service providers. The implementation of two novel building blocks will flexibly interconnect distributed 5G radio access and core network functions: Crosshaul Control Infrastructure (XCI), "a control infrastructure using a unified abstract network model for control plane integration" and Crosshaul Packet Forwarding Element (XFE), "a unified data plane encompassing innovative high-capacity transmission technologies and novel deterministic-latency switch architectures".

This transport network implements an adaptive, sharable and cost-efficient solution that aims to switch and transport ultra-broadband signals that satisfy the synchronization, latency and speed requirements in a 5G network. It will simplify network operations and facilitate network densification regardless of growing technological diversity [29].

5G-TRANSFORMER will extend the work of 5G-Crosshaul and deliver a complete scalable 5GT-MTP by adding the support of: i) integrated MEC services, ii) dynamic

placement and migration mechanisms of virtual functions, iii) new mechanisms for sharing of VNFs by multiple tenants and slices, iv) new abstraction models for vertical services, and iv) customized profiles for the C-RAN functional split considering the requirements from verticals.

H2020 ORCHESTRA [22], started in 2015, is a project focused on the design and the demonstration of a network based on the observe-decide-act control loop, which makes the transport infrastructure observable, then subjected to optimization and characterized by high flexibility and reliability. The project spans from the data to the control plane including the design of performance monitors as well as the management of monitoring information. Key people in ORCHESTRA are present in 5G-TRANSFORMER and outputs related to the transport infrastructure properties and abstraction can be exploited by 5G-TRANSFORMER, with a specific attention to the state-of-the-art related to information and data modelling (e.g., for transponders and monitors) based on YANG.

The H2020 5G Exchange (5GEx) project has the goal to enable cross-domain orchestration of services over multiple administrations or over multi-domain single administrations. 5G-TRANSFORMER evolves the 5GEx cross-domain orchestration and federation mechanisms and applies to the mobile transport network, considering interworking with the domains of vertical industries.

12.4 5GT-MTP Challenges

The 5GT-MTP includes multiple heterogeneous infrastructures together with their management functions. In addition to this, an abstraction layer is desirable, whose role is to offer any overlaying orchestrator a homogeneous and usage-oriented view.

The design of this abstraction layer is a challenge in the 5GT-MTP system because the diverse infrastructures and their associated management functions can be different in many ways at the same time.

The infrastructures (referred to as NFVI by ETSI) may differ:

- in their resource role, especially intra-PoP, DC-like resources vs. inter-POP connectivity resources;
- in their locations, typically centralized abundant resource pool vs edge scarce resource pools;
- in their containment technologies, typically hosting virtual machines vs. hosting lighter containers, like Docker;
- in the administrative entities that hold or operate the NFVIs, i.e. some belonging to an operator vs. others of the same 5GT-MTP belonging to another one;
- in the low-level facilities that are available in an NFVI area, e.g. acceleration technologies.

The infrastructure management functions (referred to as VIM and WIM by ETSI) inherit from a great part of these differences:

- that the VIMs are in charge of resources one can find in data centres while WIMs rather include SDN controllers used to established connectivity between any two POPs where the datacentre resources are located;

- there are VIMs like Openstack, whose logic assumes the set of resources in a POP can be regarded as infinite. And there is the more recent question of how to use resources in POPs located at the edge, that are by nature either scarce or access technology specific;
- there are VIM technologies, e.g. Openstack, which are made for deploying virtual machines on hypervisors, and others, e.g. Kubernetes, which are designed to manage lighter containers that may be hosted on a same OS;
- that the same application, e.g. corresponding to a given VNFC graph, may be hosted on more than one infrastructure and therefore may require heterogeneous VIMs and WIMs to be involved simultaneously in a consistent way.

The main consequence of this wide heterogeneity is that an ideal 5GT-MTP would have to include an abstraction layer which is powerful enough to absorb most of these differences, despite the change that may occur in time. Moreover, the abstraction layer, which can be perceived as the common part of the 5GT-MTP management subsystem, should know the 5GT-MTP resources availability in real time, to address concurrent accesses and potential failures while maintaining the SLAs. It also must implement the NFVO-RO SLPOC concepts to offer a customized view for each tenant.

Therefore, we set the serving API (North of 5GT-MTP) as ETSI IFA005, and choose an approach based on information modelling, and standard operation flows unwinding.

According to 5G paradigm, the support of Vertical services requires the mobile infrastructure to work in cooperation with the transport network to assure the end user mobility while the service is utilized. Moreover, mobile services and part of radio access network functions could be provided as virtualized. Therefore, they could be not the final services (e.g. automotive, eHealth, cloud robotics) for the verticals, but they need orchestration functions to be deployed. This makes the scenarios more complex and require the 5GT-MTP to be designed accordingly. The 5GT-MTP should be an evolved VIM that interacts with complex domains of transport and mobile, where each one, at its turn, is composed by a heterogenous entity in terms of providers and technology in order to guarantee the decoupling of vertical applications, mobile application, and transport while allowing efficient interactions with each other.

Another challenge to tackle is the isolation of virtualized resources. In fact, isolation permits enhancement of security by protecting the confidentiality of tenant data when they are being processed and stored in a virtual host or being conveyed in a logical link. A tenant should not be able to access to the other collocated consumers information or to assess their workloads. Without this protection, one can carry out, intentionally or not, side channel attacks or information leakage. A second aspect of isolation is the control of the resource utilization. Once virtualized resources are allocated to a tenant, an excessive resource consumption by the latter should not starve another resource granting. The 5GT-MTP should nonetheless ensure the QoS it has committed for consumers (e.g. the guaranteed bit rate, the link latency, the amount of RAM and logical CPU…). For a better utilization of idle resources, it can tolerate some over-utilizations, which will not have negative effects on the QoS. The last impacting aspect of isolation for the 5GT-MTP is the addressing space; a virtualized infrastructure should own its resource numbering and identification. Even when multiple tenants are sharing a physical resource, each of them should have the perception of being a unique user of

the latter. In addition, actions and rules configured in an addressing space should not interfere with the actions and rules of another addressing space.