H2020 5G-TRANSFORMER Project

Grant No. 761536

5G-TRANSFORMER Refined Architecture

Abstract

This deliverable reports the updates on the 5G-TRANSFORMER architecture. The main updates from the initial architecture designed provided in deliverable D1.2 are identified, providing a brief description of each one and the motivation and rationale of the changes. Additionally, it includes, as an annex, a full self-contained description of the updated 5G-TRANSFORMER architecture design.

Document properties

Document number	D1.3
Document title	5G-TRANSFORMER refined architecture
Document responsible	Carlos J. Bernardos (UC3M)
Document editor	Carlos J. Bernardos (UC3M)
Editorial team	Carlos J. Bernardos (UC3M)
Target dissemination level	Public
Status of the document	Final
Version	1.0

Production properties

Reviewers	Thomas Deiß (NOK-N), Luca Cominardi (UC3M), Carlos J. Bernardos (UC3M), Carlos Guimarães (UC3M)

Disclaimer

This document has been produced in the context of the 5G-TRANSFORMER Project. The research leading to these results has received funding from the European Community's H2020 Programme under grant agreement Nº H2020-761536.

Table of Contents

List of Contributors

Partner Short Name	Contributors
UC3M	Kiril Antevski, Arturo Azcorra, Carlos J. Bernardos, Jaime García, Carmen Guerrero, Alberto García
NEC	Xi Li, Andres Garcia Saavedra, Josep Xavier Salvat
TEI	Paola Iovanna, Fabio Ubaldi, Teresa Pepe
NOK-N	Thomas Deiß
TID	Alberto Solano Rodriguez, Luis M. Contreras
ORANGE	Philippe Bertin
BCOM	Céline Merlet, Farouk Messaoudi
NXW	Giada Landi, Marco Capitani
CTTC	Josep Mangues-Bafalluy, Jorge Baranda, Ricardo Martínez, Luca Vettori, Ramon Casellas
POLITO	Marco Ajmone Marsan, Carla Fabiana Chiasserini
EURECOM	Adlen Ksentini, Pantelis Frangoudis
SSSA	Barbara Martini

List of Figures

List of Tables

List of Acronyms

Acronym	Description
5GT-MTP	Mobile Transport and Computing Platform
5GT-SO	Service Orchestrator
5GT-VS	Vertical Slicer
ABNO	Applications-Based Network Operations
AD	Administrative Domain
AM	Abstraction Manager
API	Application Programming Interface
AppD	Application Descriptor
AS/PCE	Active Stateful Path Computation Element
BSS	Business Support System
CIM	Cooperative Information Manager
CN	Core Network
COP	Control Orchestration Protocol
CQI	Channel Quality Indicator
CSAR	Cloud Service Archive
CSMF	Communication Service Management Function
DCSP	Data Centre Service Provider
EM	Element Manager
EVS	Extended Virtual Sensing
E/WBI	Eastbound/Westbound Interface
E2E	End-to-end
GMPLS	Generalized Multi-Protocol Label Switching
GTP	GPRS Tunnelling Protocol
HMI	Human Machine Interface
HSS	Home Subscriber Server
HTTP	HyperText Transfer Protocol
IM	Information Model
JSON	JavaScript Object Notation
LCM	LifeCycle Management
LL	Logical Link
LSA	Link Selection Algorithm
MANO	Management and Orchestration
MEA	Multi-access edge application
MEAO	Multi-access edge application orchestrator
MEC	Multi-access edge computing
MEO	Multi-access edge orchestrator
MEP	Multi-access edge platform
MEPM	Multi-access edge platform manager
MEPM-V	Multi-access edge platform manager - NFV
MES	Multi-access edge service
MLPOC	Multiple Logical Point of Contact
MME	Mobility Management Element
MVNO	Mobile Virtual Network Operator
NBI	Northbound Interface
NF	Network Function
NF FG	NF Forwarding Graph
NFP	Network Forwarding Path
NFV	Network Function Virtualization

NFVIaaS	NFVI as a Service
NFV-NS	Network Service
NFV-NSI	Network Service Instance
NFV-NSO	Network Service Orchestrator
NFVI	Network Functions Virtualisation Infrastructure
NFVI-PoP	Network Functions Virtualisation Infrastructure Point-of-Presence
NFVIaaS	NFVI as a Service
NFVO	NFV Orchestrator
NFVO-RO	Resource Orchestrator
NNI	Network to Network Interface
NS	Network Slice
NS-OE	NS Orchestration Engine
NSaaS	Network Slice as a Service
NSD	Network Service Descriptor
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NSSI	Network Slice Subnet Instance
NSSMF	Network Slice Subnet Management Function
NST	Network Slice Template
OEM	Original Equipment Manufacturer
OF	OpenFlow
OSS	Operating Support System
PA	Physical Application
PGW-C	Packet Gateway Control Plan
PGW-U	Packet Gateway User Plan
PNF	Physical Network Function
PNFD	Physical Network Function Descriptor
QoS	Quality of Service
RAN	Radio Access Network
REST	Representational State Transfer
RMA	Resource Management Application
RNIS	Radio Network Information
ROOE	Resource Orchestration Engine
SBI	Southbound Interface
SDK	Software Development Kit
SDN	Software-Defined Networking
SGW-C	Serving Gateway Control Plan
SGW-U	Serving Gateway User Plan
SLA	Service Level Agreement
SLO	Service Level Objective
SLPOC	Single Logical Point of Contact
SM	Service Manager
SOE	Service Orchestration Engine
TD	Technology Domain
TMOP	5G-TRANSFORMER Mobile Transport and Computing Platform Operator
TMVS	5G-TRANSFORMER Managed Vertical Service
TOSCA	Topology and Orchestration Specification for Cloud Applications
TOR	Top of the Rack

TRF	5G-TRANSFORMER Resource Federation
TS	5G-TRANSFORMER Service
TSC	5G-TRANSFORMER Service consumer
TSF	5G-TRANSFORMER Service Federation
TSP	5G-TRANSFORMER Service Provider
TUVS	5G-TRANSFORMER Unmanaged Vertical Service
VA	Virtual Application
VA FG	VA Forwarding Graph
vCDN	virtual Content Delivery Network
VIM	Virtual Infrastructure Manager
VISP	Virtualization Infrastructure Service Provider
VL	Virtual Link
VLD	Virtual Link Descriptor
VNF	Virtualised Network Function
VNFFG	VNF Forwarding Graph
VNFC	Virtualised Network Function Component
VNFD	Virtualised Network Function Descriptor
VNFM	Virtual Network Functions Manager
VS	Vertical Service
VSaaS	Vertical Service as a Service
VSB	Vertical Service Blueprint
VSD	Vertical Service Descriptor
VSI	Vertical Service Instance
VXLAN	Virtual Extensible LAN
WAN	Wide Area Network
WIM	Wide area network Infrastructure Manager
XCI	5G-Crosshaul Control Infrastructure
XFE	5G-Crosshaul Forwarding Element
YAML	YAML Ain't Markup Language
YANG	Yet Another Next Generation

Executive Summary and Key Contributions

The 5G-TRANSFORMER project aims at simultaneously supporting the needs of different vertical industries (such as, eHealth, automotive, entertainment, eIndustry and MNVO) and, consequently, enriching the telecom network ecosystem. It leverages on the concept of network slicing for providing slices tailored to needs (e.g., networking and computing requirements) of different vertical industries and to allow per-slice management of virtualized resources. The 5G-TRANSFORMER architecture is based on three novel building blocks, namely the Vertical Slicer (5GT-VS), the Service Orchestrator (5GT-SO) and the Mobile Transport and Computing Platform (5GT-MTP).

The main contribution of this deliverable is the description and specification of the 5G-TRANSFORMER system, which updates the initial design (as described in D1.2 [2]). This updated design includes the following novelties and/or enhancements (compared to the one described in D1.2):

- Service scaling and arbitration;
- Policy management;
- Enhancement to vertical support;
- Vertical service composition;
- Enhanced monitoring architecture;
- NFV network service composition and federation;
- MEC support;
- Initial RAN abstraction;
- Advanced management of network slices composed of multiple slice subnets;
- Support for inter-NFVI-PoP orchestration;
- Enhanced placement algorithms (5GT-MTP), including functional splits;
- Enhanced placement algorithms (5GT-SO);
- Dynamic VNF configuration on instantiation.

This deliverable is divided into two parts. The first part, which is presented in the main body of this deliverable, focuses on the changes made since the initial design of the 5G-TRANSFORMER system reported in D1.2 [2]. It mainly addresses the changes on the baseline design, its building blocks (i.e., 5GT-VS, 5GT-SO and 5GT-MTP) and interfaces and reference points as well as on the aspects related with monitoring framework. The second part, which is presented as an annex of this deliverable, comprises a full self-contained updated description of the 5G-TRANSFORMER system. It provides an analysis of the 5G-TRANSFORMER stakeholders, services and architecture requirements, followed by an overview of the 5G-TRANSFORMER architecture including each building block regarding their key functionalities and the internal architecture design. The high-level workflows among the architectural building blocks for a set of basic service operations are also presented and described.

This approach allows the readers already familiar with 5G-TRANSFORMER to easily identify the main changes without having to go through the description of the whole architecture, while newcomers to 5G-TRANSFORMER have access to the full description of current status of the 5G-TRANSFORMER system.

The deliverables D3.3 [6], D4.3 [8] and D2.3 [4] provide more details about the internal operation of the main building blocks of the 5G-TRANSFORMER system (5GT-VS, 5GT-SO and 5GT-MTP, respectively). The above additional deliverables are meant as a complementary source of information to this deliverable.

In more details, the main contributions in this deliverable are the following:

- **Revision of the 5G-TRANSFORMER architecture requirements**, with the introduction of new business requirements, as well as new functional requirements regarding the discovery, fulfilment and assurance phases.
- **Refinement of the overall 5G-TRANSFORMER architecture design, main building blocks and interfaces**, to support the new features and enhancements enumerated above.
- High-level description on how 5G-TRANSFORMER supports **SLA management and service scaling** as well as **policy management**.
- **Revision of the high-level example workflows**, with the inclusion of two novel workflows addressing service and resource federation.
- **Dissemination** of 5G-TRANSFORMER architecture design and steering related standardization activities (activities collected by WP6), resulting in several contributions to **academic journals, conferences and magazines**, **standardization bodies** and **public demos** showcasing the use cases developed in the project. A full list of the dissemination activities can be found in D6.5 [10].

1 Introduction

5G-TRANSFORMER (also referred as 5GT for shorter references) is designing a flexible SDN/NFV-based platform to support next-generation mobile networks and novel vertical-oriented use cases. Namely, eHealth, automotive, entertainment and eIndustry use cases drive the design of 5G-TRANSFORMER's architecture targeting such heterogeneous service requirements. To achieve its purpose, the 5G-TRANSFORMER system defines three main functional blocks as described in the baseline architecture published in the previous deliverable D1.2 [2]: the Vertical Slicer (5GT-VS), the Service Orchestrator (5GT-SO) and the Mobile Transport and Computing Platform (5GT-MTP).

The Vertical Slicer (5GT-VS) is the entry point for vertical industries, which are the consumers of 5G-TRANSFORMER services. The main goal of 5GT-VS is three-fold. Firstly, to provide an interface that lets vertical players compose (exploiting an exposed catalogue of blueprints), instantiate, update, terminate and monitor their Vertical Service Instances (VSI). Secondly, to arbitrate resources among contending VSIs in case of resource shortage. Lastly, to create and manage Network Slices (NS). An NS is modelled as an extended ETSI NFV Network Service (NFV-NS) and described by extended ETSI NFV Network Service Descriptors (NSD) [30], which 5GT-SO can orchestrate when requested by 5GT-VS.

The Service Orchestrator (5GT-SO) is responsible for the management and orchestration of NFV-NSs. In doing so, the 5GT-SO can exploit resources from local domains, federate resources or services from domains that belong to different administrations. The 5GT-SO's main function is to process NFV-NS orchestration requests from 5GT-VS via (i) service orchestration (NFV-NSO), and ultimately (ii) resource orchestration (NFVO-RO). Regarding the service orchestration, the 5GT-SO shall manage the deployment of NFV-NSs requested by the 5GT-VS (also possibly across different domains), including their lifecycle management (on-boarding, instantiation, scaling, termination) and the management of the VNF forwarding graphs associated to the network service. Regarding the resource orchestration, the 5GT-SO is in charge of aggregating resources spanning across multiple domains, multiple technologies and/or multiple administrations (via federation). Clearly, both NFVO-NSO and NFVO-RO coordinate to provision enough resources to each deployed VNF and (virtual) links.

The Mobile Transport and Computing Platform (5GT-MTP) replaces traditional rigid "one-size-fits-all" deployments with a customizable SDN/NFV-based transport and computing platform capable of simultaneously supporting a diverse range of networking and computing requirements specific to the vertical industries. The 5GT-MTP hosts the physical and/or virtual mobile transport network and computing infrastructure within which vertical services are deployed.

This deliverable refines the 5G-TRANSFORMER architecture by focusing on the extensions introduced to enhance the baseline architecture as described in deliverable D1.2 [2]. These extensions were motivated by advanced research conducted in the project as well as the ongoing development of 5G-TRANSFORMER architecture building components.

The structure of this deliverable is as follows: Section 2 describes the extensions of the 5G-TRANSFORMER architecture, focusing on the baseline architecture followed by the details on the extensions and enhancements that were brought to each of the three main functional blocks (i.e., 5GT-VS, 5GT-SO, 5GT-MTP), but also to the monitoring framework as well as to interfaces and reference points. This deliverable then presents the main conclusions in Section 3. Finally, as an annex, namely Annex A, the overall updated description of 5G-TRANSFORMER system is included, providing a self-contained description of the refined 5G-TRANSFORMER architecture.

2 5G-TRANSFORMER Architecture Refinements

This chapter introduces the refinements made over the initial design of the 5G-TRANSFORMER reported in deliverable D1.2 [2]. The changes in the main building blocks of the 5G-TRANSFORMER system architecture (namely, 5GT-VS, 5GT-SO, 5GT-MTP, monitoring framework and interfaces and reference points) are described, together with the rationale of why they were required.

2.1 Baseline Design

The refined design of the 5G-TRANSFORMER architecture, as presented in Figure 1 (details in Section A.3.1), follows the same concepts and design defined in the initial version of the baseline architecture reported in D1.2 [2], and same function roles for the three main architecture components: Vertical Slicer (5GT-VS), Service Orchestrator (5GT-SO) and Mobile Transport and Computing Platform (5GT-MTP).

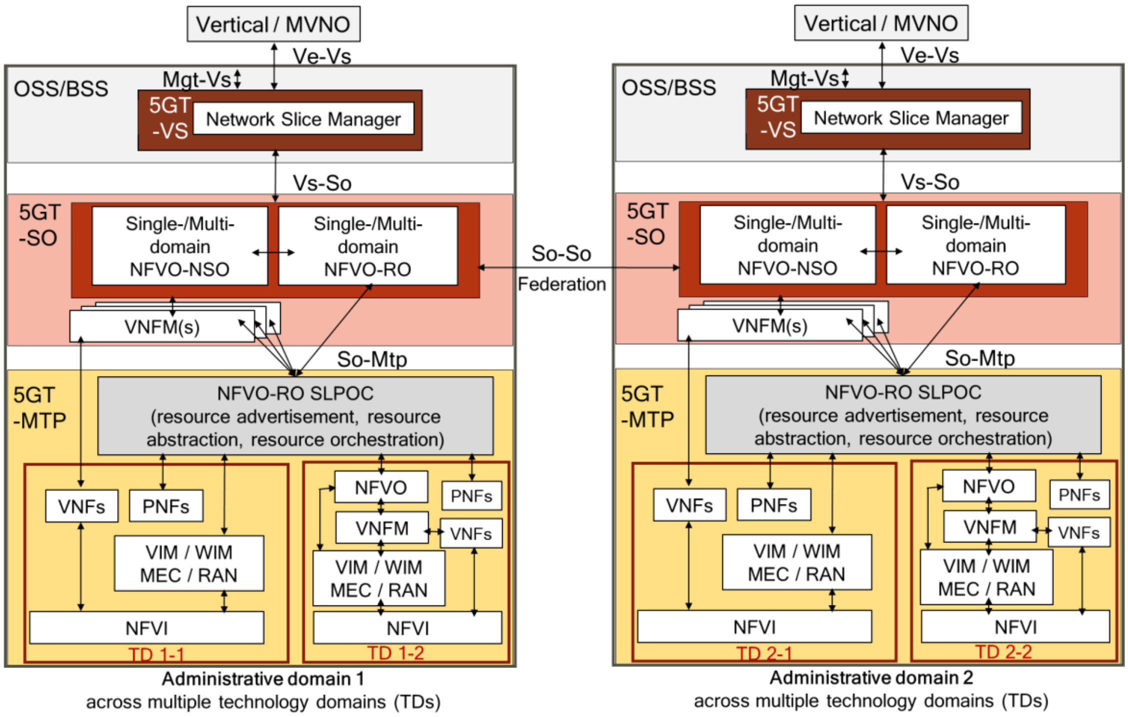


FIGURE 1: 5G-TRANSFORMER SYSTEM ARCHITECTURE

However, there are a number of added functionalities and extended features supported at the 5GT-VS, 5GT-SO and 5GT-MTP in the latest architecture design. They are described in the following subsections. In summary, the main new features supported by the 5G-TRANSFORMER architecture include:

- Service scaling to enable vertical services to be scaled dynamically and automatically to satisfy the SLA requirements in an efficient manner.
- Policy management to allow enforcing policies through the system to better support the SLA requirements and preferences of the verticals.
- Enhancement of the vertical support to allow creation of new instances of vertical services at runtime.

- Vertical service composition inside the 5GT-VS to allow verticals to compose vertical services from a set of basic building blocks.
- Refined monitoring architecture that allows (i) defining and encoding the monitoring requirements into the NSD onboarded from the 5GT-VS to the 5GT-SO; and, (ii) interaction to the 5GT-SO to support auto-scaling of NFV network services and automated SLA management.
- NFV network service composition and federation inside the 5GT-SO, to support decomposing of NSDs into several nested NFV-NS segments that can be deployed locally or across different technology domains or different administrative domains (through federation).
- Enhanced support for MEC through the 5G-TRANSFORMER system by providing MEC orchestration at the 5GT-SO level in terms of deciding optimum placement of MEC applications (AppDs) and their required resources and connectivity in the underlying MTPs, considering the location constraints of the MEC services, and then the MTP supports the deployment of the MEC applications and specific MEC services through the configuration and interaction with the MEC platforms/hosts via developed MEC plugins.
- Extension of the 5GT-MTP abstraction for MEC and RAN as well as development of different SBI plugins to manage the different WAN, MEC and radio resources.
- Extension of the interfaces at the northbound of the 5GT-VS to improve vertical support, update of the Vs-So interface to support policy management, and extension of the So-Mtp interface to restrict that a single 5GT-SO interacts only with a single 5GT-MTP. In addition, So-Mtp extensions shall enhance 5GT support for transport and the inter-NFVI-PoP connectivity as well as for the MEC and RAN abstraction.

Table 1 summarizes the main features included in the current refined architecture (described in this deliverable) in contrast to the features available in the initial design (as described in D1.2 [2]) and those planned for the final design (which will be described in D1.4). In the following sections, we provide detailed information on the updates performed at the 5GT-VS, 5GT-SO, 5GT-MTP, monitoring and interface level, respectively.

TABLE 1: SUMMARY OF FEATURES IN 5G-TRANSFORM DESIGN

Features in initial design	Features in refined design	Features planned for final design
Vertical Service Descriptor to Network Service Descriptor translation	Service scaling and arbitration	Enhanced RAN abstraction and orchestration
Initial interface towards verticals	Policy management	3GPP RAN slicing
Basic authentication and per-tenant authorization	Enhancement to vertical support	Feedback to vertical for KPI verification and SLA management
Basic arbitration mechanisms	Vertical service composition	
Basic lifecycle management (no service composition)	Enhanced monitoring architecture	

Initial Vs-So and So-Mtp interface (single PoP)	NFV network service composition and federation
NV-NS creation, instantiation, termination and query operational status operations	MEC support
Resource orchestration functions	Initial RAN abstraction
Basic monitoring platform	Advanced management of network slices composed of multiple slice subnets
Cloud abstraction	Support for inter-NFVI-PoP orchestration
Resource allocation and termination for VIM and WIM domains	Enhanced placement algorithms (5GT-MTP), including functional splits
	Enhanced placement algorithms (5GT-SO)
	Dynamic VNF configuration on instantiation

2.2 5GT-VS

The 5GT-VS has been extended regarding the following architectural aspects:

Policy management: Some SLA requirements such as preferred or prohibited infrastructure providers, energy reduction, or service creation time are difficult or impossible to map to deployment flavours or other parts of an NFV-NSD. Such requirements can better be handled by policies for the 5GT-SO. Throughout the first period of the project, ETSI NFV has already an interface for policy management at the Os-Ma-nvfo reference point [29], which is the basis for the interface between 5GT-VS and 5GT-SO. Based on the SLA requirements of a vertical service instance (VSI), 5G-TRANSFORMER extended this interface and contributed the extension to ETSI NFV such that the 5GT-VS can associate policies such that the 5GT-SO can act accordingly. For further details see A.3.4.

Feedback from vertical applications: Some vertical services require the creation of new instances of vertical services at runtime. This can be achieved by calling the REST API of the 5GT-VS from within the vertical applications. The vertical applications have to be configured with the IP address of the API of the 5GT-VS. This extension is used in the eHealth use case to create new instances of the service supporting the handling of emergencies. For further details see D3.3 [6], Section 5.7.3.

Service configuration: Instances of the same vertical service may differ regarding the geographical deployment environment or the profile of the verticals' customers accessing the service instance. This function allows the user to define per-instance application-level parameters during the service instantiation request. The type of parameters that can be configured are defined in the Vertical Service Blueprint (VSB) and their values, as provided by the vertical user, are encoded in the NFV-NS request sent to the 5GT-SO.

Vertical-driven service composition: Verticals are allowed to compose vertical services from a set of basic building blocks in addition to selecting and completing vertical service blueprints. This composition provides additional flexibility to the verticals without increasing the complexity of vertical services to them. The extension does not have impact on the other components of the 5G-TRANSFORMER system. We have based this extension on two complementing approaches of other 5G-PPP projects to describe services. For further details see D3.3 [6], Section 5.4.4.

The functionality of the 5GT-VS has been extended to handle more complex service definitions and scenarios as in its first release. Major extensions provide different ways to scale vertical services and to allow the handling of composite service definitions:

- **Autoscaling**: Vertical services need to be scaled dynamically to satisfy the SLA requirements in an efficient manner. Feedback from the development of the different use cases indicated that three types of scaling are needed: One type of scaling - automated NFV-NS scaling - is performed by the 5GT-SO based on directives from the 5GT-VS. These directives are captured in the Network Service Descriptor (NSD), in the form of autoscaling rules associated to the detection of resource shortage conditions, expressed in terms of thresholds on resource-related monitoring parameters or VNF indicators. Scaling actions are the change of the instantiation level, scale up/down decisions of specific VNFs, and creation/termination of VNF instances. The format of the autoscaling rules has been defined. For further details see A.3.3. The 5GT-VS has been extended to support two other types of scaling: vertical-driven service scaling and automated network slice scaling. In vertical-driven service scaling it is the vertical who explicitly requests the service scaling through the interface of the 5GT-VS. In automated network slice scaling, the 5GT-VS itself decides on the scaling of a slice, e.g. based on arbitrating decisions. In both types of scaling, the actual scaling is performed by the 5GT-SO.
- **Composite services**: the service provider can define a service as a composition of other services. E.g. this allows to describe functionality common to several services once and refer this description in other service definitions. The components of the 5GT-VS have been extended to handle composite services as well, as described in the next point.
- **Advanced management of network slices composed of multiple slice subnets**: A slice may consist of several slice subnets similar to the definition of composite services from constituent services. On the other hand, a slice subnet and its resources may be used by several upper-layer services, thus extending the concept and the advantages of resource sharing to slice sharing. This sharing is still compliant with the service requirements in terms of per-tenant isolation and defined vertical-driven policies. However, sharing a slice subnet across multiple end-to-end slices may impact its performance and, consequently, the performance of the vertical services running on top of it. Therefore, the size of a slice subnet instance must be dynamically adjusted based on the composite slices that are sharing it. The Arbitrator in the 5GT-VS has been extended to decide which slice subnets must be re-used and shared to instantiate new end-to-end network slices and how existing subnets must be modified to support the new traffic and processing load. The procedures to scale slice subnets and

instantiate composite network slices are handled within the functional entities responsible of the network slice management in the 5GT-VS.

2.3 5GT-SO

The main architectural updates concerning the 5GT-SO are the following:

Service scaling. 5GT Vertical Services are designed to be deployed and used in highly dynamic environments. Here the volume of service requests and the mobile traffic might change drastically and unexpectedly. Moreover, users might connect and disconnect from the service at a very high rate.

Provisioning a service with the size required to serve the highest anticipated traffic might be overly expensive, for example, the traffic might be as much as 10x higher during the day than at night. Moreover, such strategy might still lead to interruptions of the service or degradation of the QoS during unforeseen peaks.

To mitigate these issues, the 5GT-SO, in coordination with 5GT-VS, offers the capability of scaling the service at runtime, without service interruption, in order to be able to accommodate a very high number of peak users without wasting resources in periods where the service demand is lower.

In 5G-TRANSFORMER three different types of service scaling have been defined:

- The **Vertical-driven VSI Scaling** occurs whenever the scaling of a vertical service instance is triggered by the vertical, either manually through the GUI of the 5GT-VS or through the 5GT-VS REST API.
- The **Automated VSI Scaling** occurs whenever the 5GT-VS, usually its Arbitrator component, encounters the need to resize a vertical service or a network slice. This leads the 5GT-VS to trigger a scaling action on the corresponding NFV-NS instance. The most likely scenario is if the vertical customer has exhausted its resources quota, but it needs to instantiate a high priority service. In that case the 5GT-VS scales down another - lower priority - service in order to free the resources required to instantiate the high priority service.
- The **Automated NFV-NS** scaling is driven by the auto-scaling rules expressed in the NFV-NSD and it is managed entirely at the 5GT-SO level via the SLA Manager. Each auto-scaling rule contains a condition, based on the service's monitored data, and a reaction, i.e. a scaling action. Whenever the condition is verified, the scaling reaction is performed by the 5GT-SO.

More details on this can be found in Section A.3.3 and D4.3 [8].

The main architectural updates in the 5GT-SO to support service scaling are the **Monitoring Manager** and **the SLA Manager**. In short:

- **Monitoring Manager:** This functional block of the 5GT-SO is responsible for translating high-level, service-centric monitoring requirements of instantiated NFV-NSs into low-level, resource-centric tasks to be handled by the Monitoring Platform. It provides the details of these tasks to the SLA Manager for threshold configuration.
- **SLA Manager:** This block is responsible to assure SLA satisfaction through online SLA monitoring. To achieve this goal, the SLA Manager exploits relevant information from the Monitoring Platform and assesses SLA breaches and/or

anomalous behaviours and triggers corrective actions, such as scaling of services or resources following predefined auto-scaling rules, to ensure the target SLAs are met.

The details of these architectural updates in the 5GT-SO can be found in Section A.5.2 and in D4.3 [8].

Service Composition and Service Federation. The NFVO-NSO (Network Service Orchestration functional block) coordinates all NFV-NS deployment operations. An important architectural update of the NFVO-NSO is the support of Composite NSO, which provides the ability of decomposing NSDs into several NFV-NS segments that can be deployed locally or across different technology domains or different administrative domains (through federation). To this aim, three new functional components are introduced to augment the capabilities of the NS Orchestration Engine (NS-OE):

- **Constituent NSO:** This entity is responsible to handle nested NFV-NSs performing tasks such as lifecycle management and orchestration of nested NFV-NS segments.
- **Composite NSO:** The role of this functional block is to use NS-OE algorithms to decompose NSDs into several NFV-NS segments and to decide where to place each of them.
- **Link Selection Algorithm (LSA):** In case of service federation, that is, when different NFV-NS segments are deployed across different administrative domains, the LSA module shall compute the network links between them.

Service federation is the overall process of establishing, consuming or providing NFV network services (NFV-NS) by an administrative domain from/to other (peering/partner) administrative domain. The administrative domain that requests service is referred to as consumer domain while the peering administrative domain capable of providing service is a provider domain. The service federation is the combination of establishing business/service level agreements among the administrative domains and consuming/providing federated nested NFV-NSs that are part of composite End-to-end NFV-NS.

The reader can find more details in Section A.5.3 and in D4.3 [8].

Placement Algorithms (PAs). New algorithms to support multi-domain multi-technology service orchestration and, in particular, to accommodate location and latency constraints, have been embedded into the 5GT-SO. All the algorithms have been implemented and are integrated in the 5GT-SO software distribution. In order to have a unified way to access them, a single REST API over which they can be called has been specified using OpenAPI v2.0. Although all algorithms operate on the same basic system model, each one has its particularities, which are also captured in the API (e.g., Placement Algorithm A, if configured to optimize for availability, needs VNF- and NFVI-PoP-level reliability information; Placement Algorithm C requires gateway IP addresses). The PAs are expected to be running as micro-services, each exposing a REST API endpoint, which the RO-OE uses to request the calculation of resource allocations during the instantiation of NFV-NSs. All algorithms, which are written in python, use the python flask library to implement their API servers.

The details of these algorithms can be found in D4.3 [8] in Section 6.6.1.

2.4 5GT-MTP

The functional architecture of the 5GT MTP framework, documented in Section A.3.1.3, follows the same approach defined in the initial version of the 5GT architecture. The MTP manages all the infrastructures as a Single Logic Point of Contact (SLPOC) providing a common abstraction view of the managed resources to the SO. In the first phase of the project the resources managed and abstracted are the transport (WIM) and datacentre ones (VIM), while in the second phase the abstraction and the resource management are extended adding the handling of Radio and MEC resources.

The main updates are related to the introduction of new features and functional components and their relative interfaces. In more detail, the second release introduces:

- the support to Radio,
- the support to MEC,
- the slicing support in a scenario where heterogeneous technologies can be used for the transport network,
- the monitoring of the infrastructure resources controlled by the MTP, and
- local Placement Algorithm (PA) to handle resource selection inside Technology Domain.

The MTP Monitoring feature is compliant to the 5GT monitoring system architecture with a hierarchy of monitoring services deployed in the different layers of 5GT system (see Section 2.5). The MTP requests monitoring tasks to the 5GT monitoring platform relating to the infrastructure resources (for example, verify that the allocated infrastructure resources for a service can satisfy the service level agreement like delay, number of vCPU and memory needed by VNFs) and retrieve the processed monitoring data in order to take decisions about the lifecycle of the infrastructure resources (like service level agreement recovery by rerouting the connectivity in alternative paths or by migrating the VM).

Moreover, the interfaces defined in the first phase for the communication with the SO (MTP NBI) and for the communication with the infrastructure manager plugin (MTP SBI) were extended to retrieve Radio and MEC abstract parameters and to allocate/terminate Radio and MEC resources. The abstraction of RAN and MEC resources uses algorithms that manage the technology specific information and harmonize them to provide a common view with respect to the VIM/WIM abstract resource information. In this way SO can use same algorithms (or new one) to place the network function of vertical services in the different technology domain.

5GT adopts ETSI IFA005 [22] for the interfaces between 5GT-SO and 5GT-MTP and between 5GT-MTP and VIMs/WIMs. IFA005 defines standard interfaces to enable a consumer block at the VIM north bound interface to retrieve information about resources capability (e.g., maximum CPU), to allocate and reserve a resource, and to be notified about a change of a resource. The IFA005 interface is extended to handle radio and MEC resources. For MEC specific resources, the extension is based on the information model described in ETSI MEC 010-2 [36].

For the monitoring, instead, the 5G-TRANSFORMER monitoring framework has been adopted, thus the interfaces use the Rest API provided by the framework.

2.5 Monitoring

The functional architecture of the 5G-TRANSFORMER monitoring framework, documented in section A.3.1.4, follows the same concepts and principles defined in the initial version of the overall system architecture, with a hierarchy of monitoring services deployed in the different layers of 5G-TRANSFORMER system. Each of them collects and processes monitoring data related to the logical entities managed in each layer, retrieving elementary monitoring data from a variety of sources, aggregating and elaborating them into consolidated monitoring parameters. Such parameters are then made available for the respective layers of the 5G-TRANSFORMER system, in order to take decisions about the lifecycle of the logical elements managed at each layer. Moreover, monitoring parameters related to vertical service instances are also provided to the vertical user (or its applications) for further processing at the application layer or just to visualize the performance of the service through monitoring graphs.

During the second phase of the project, the focus has been on the implementation of a monitoring platform that implements the monitoring service for the 5GT-SO (documented in deliverable D4.3 [8]), as enabler for auto-scaling of NFV network services and automated SLA management. In this context, taking into account feedbacks from the development activities, the architecture has been refined on the following aspects:

- Definition of the monitoring requirements, scaling criteria and strategies for NFV network services, specified in the NSD onboarded from the 5GT-VS to the 5GT-SO. These requirements are defined through a set of monitoring parameters and autoscaling rules in the NSD.
- Procedures for SLA management and service autoscaling, based on performance monitoring. This mechanism is based on a publish/subscribe approach where the monitoring platforms receives subscriptions about monitoring alerts and generates notifications according to the values of the collected metrics.
- Interactions between the monitoring platform deployed at the 5GT-SO layer and the internal components of the 5GT-SO, for handling monitoring configuration and processing monitoring alerts. At the 5GT-SO level, the Monitoring Manager element is in charge of configuring the monitoring platform according to the parameters specified in the NSD, thus activating a custom monitoring process for each service instance. The elaboration of monitoring alerts is handled at the SLA Manager and it is part of the decision procedure within the whole autoscaling workflow.

In particular, as reported in section A.3.3, the monitoring platform is configured during the service instantiation phase in order to collect and process a set of monitoring parameters related to the given service instance, according to the monitoring requirements specified in the NSD. The 5GT-SO configures also a set of thresholds on the monitored data, allowing the monitoring platform to identify anomalous conditions in the service behaviour. Following this approach, the monitoring platform collects and elaborates the relevant monitoring data during the runtime of the service, aggregating data from different sources and generating notification alerts towards the 5GT-SO whenever the configured thresholds are passed. A new component within the 5GT-SO, the SLA Manager, is in charge of processing the monitoring alerts, reacting according to the autoscaling rules defined in the NSD and triggering the required service scaling actions.

2.6 Interfaces and Reference Points

In this section, the main changes regarding the design of the interfaces between the main building blocks of the 5G-TRANSFORMER architecture are described.

Northbound of the 5GT-VS. No changes have been made in this interface.

Vs-So Interface. Regarding the interface between the 5GT-VS and the 5GT-SO, a new reference point, named Vs-So(-POL), was introduced to enable policy management functionalities, in addition to the already existing reference points in the Vs-So interface (namely, the Vs-So(-LCM), the Vs-So(-MON) and the Vs-So(-CAT)). To achieve its purpose, the Vs-So(-POL) offers primitives to transfer, delete, activate, deactivate, associate, disassociate policies and receive notifications about policy conflicts. The Vs-So(-POL) was introduced due to the need of applying SLA policies for the instantiation of the NFV-NS. By applying the policies, the 5GT-SO can perform scaling/healing procedures according to the associated policies. More details can be found in Section A.3.1.5.2 of this deliverable and in Section 5.3.5 of D3.3 and Section 2.4.1 of D4.3.

So-Mtp Interface. In the refined 5G-TRANSFORMER architecture design, it was defined that the 5GT-SO and the 5G-MTP maintain a 1:1 relationship. In other words, a single 5GT-SO interacts with a single 5GT-MTP via the So-Mtp interface. A client-server relationship is created and all the interface contents were devised to consider that this interface enables the communication between the SO and a specific MTP. This approach differs from the initial 5G-TRANSFORMER architecture design where a 1:N relationship was defined (i.e., a single 5GT-SO could interact with multiple 5GT-MTPs). The reason for this change is related to the fact that the MTP takes care of all NFVI-PoPs in a single administrative domain. This allowed to simplify the 5GT system to contain a single 5GT-VS, single 5GT-SO and a single 5GT-MTP. More details about this change can be found in Section A.3.1.5.3.

The initial design of the So-Mtp interface was conceived to be based on existing set of standard documents being produced within the ETSI NFV framework, namely ETSI GS NFV-IFA 005 [22], ETSI GS NFV-IFA 006 [23] and ETSI GS NFV-IFA 008 [25]. In the refined 5G-TRANSFORMER design, this interface also takes into consideration the ETSI GR NFV-IFA 022 [31] from which were extracted some ideas. The reason behind this is related with the fact that specific functions and requirements exclusively in the scope of handling virtualized network resources between NFVI-PoPs (i.e., within the WAN interconnecting remote NFVI-PoPs) were not fully supported by the ETSI NFV framework documents taking into consideration in the initial design of the 5G-TRANSFORMER architecture. By extending this interface in this respect, it became possible to better cope with the intricacies and characteristics to manage the network resources in the WAN by means of a WIM controller. More details can be found in Section A.3.1.5.2.

So-So Interface. No changes have been made in this interface.

3 Conclusions

The aim of this document is to provide an updated version of the 5G-TRANSFORMER architecture design, which was refined since its initial design to better cope with the requirements and needs faced during the project execution and implementation.

More specifically, it has addressed the following novelties and/or enhancements: (i) service scaling and arbitration; (ii) policy management; (iii) enhancement to vertical support; (iv) vertical service composition; (v) enhanced monitoring architecture; (vi) NFV network service composition and federation; (vii) MEC support; (viii) initial RAN abstraction; (ix) advanced management of network slices composed of multiple slice subnets; (x) support for inter-PoP orchestration; (xi) enhanced placement algorithms in both the 5GT-MTP and 5GT-SO; and finally (xii) dynamic VNF configuration on instantiation.

This deliverable was structured in a way that could allow readers already familiar with 5G-TRANSFORMER to easily identify the main changes without requiring them to read the whole architecture definition and, at the same time, to provide newcomers to have access to the full description of the current status of the 5G-TRANSFOMER system. In the main body of this deliverable, namely in Section 2, the major changes regarding the main building blocks of the 5G-TRANSFORMER system and the interfaces between one another had been identified and described, providing the main motivation and rationale of each change. Finally, as an annex, a full self-contained description of the updated 5G-TRANSFORMER architecture design is provided, avoiding the need to cross-check with previous deliverables to have the whole view of the refinements done over the 5G-TRANSFORMER system. More details in what respects the changes in the 5GT-VS, 5GT-SO and 5GT-MTP building blocks can be found in D3.3 [6], D4.3 [8] and D2.3 [4], respectively.

The final 5G-TRANSFORMER system design will be provided in the future deliverable D1.4, which will address the enhancement of RAN abstraction and orchestration, RAN slicing and feedback to verticals for KPI verification and SLA management.

4 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, D2.1, Definition of the Mobile Transport and Computing Platform, March 2018.

[4] 5G-TRANSFORMER, D2.3, Final design and implementation report on the MTP, May 2019.

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

[6] 5G-TRANSFORMER, D3.3, Final design and implementation report on the Vertical Slicer, May 2019.

[7] 5G-TRANSFORMER, D4.1, Definition of service orchestration and federation algorithms, service monitoring algorithms, March 2018.

[8] 5G-TRANSFORMER, D4.3, Final design and implementation report on service orchestration, federation and monitoring platform, May 2019.

[9] 5G-TRANSFORMER, D5.1, Definition of vertical testbeds and initial integration plans, May 2018.

[10] 5G-TRANSFORMER, D6.5, Report on WP6 progress and update of CoDEP, May 2018.

[11] 3GPP TR28.801, V15.1.0, Telecommunication management; Study on management and orchestration of network slicing for next generation network, January 2018.

[12] 3GPP TS23.501, V16.0.2, System architecture for the 5G System (5GS), April 2019.

[13] 3GPP TR 23.714, V14.0.0, Study on Control Plane (CP) and User Plane (UP) separation of Evolved Packet Core (EPC) nodes, June 2016.

[14] 3GPP TS28.530, V0.4.0, Telecommunication management; Management of 5G networks and network slicing; Concepts, use cases and requirements, January 2017.

[15] 5GCity, D4.1: Orchestrator design, service programming and machine learning models, 2018.

[16] NHTSA, “Preliminary regulatory impact analysis - FMVSS No. 150 Vehicle-to-Vehicle Communication Technology for Light Vehicles,” December 2016.

[17] G. Avino, M. Malinverno, C. Casetti, C. F. Chiasserini, F. Malandrino, M. Rapelli, G. Zennaro “Support of Safety Services through Vehicular Communications: The Intersection Collision Avoidance Use Case”.

[18] 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.

[19] TeleManagement Forum. “TeleManagement Forum Information Framework (SID): GB922_Addendum_4SO_Service_Overview_R9-5_v9-7”.

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

[21] ETSI GS NFV 003, “Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV”, v1.2.1, Dec. 2014.

[22] 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.

[23] ETSI GS NFV-IFA 006, “Network Function Virtualisation (NFV); Management and Orchestration; Vi-Vnfm reference point - Interface and Information Model Specification”, v2.1.1, 2016.
[24] ETSI GS NFV-IFA 007, “Network Function Virtualisation (NFV); Management and Orchestration; Or-Vnfm reference point - Interface and Information Model Specification” v2.1.1, 2016.
[25] ETSI GS NFV-IFA 008, “Network Functions Virtualisation (NFV); Management and Orchestration; Ve-Vnfm reference point - Interface and Information Model Specification”, v2.1.1, 2016.
[26] ETSI GS NFV-IFA 010, “Management and Orchestration; Functional requirements specification”, v2.4.1, 2018.
[27] ETSI GS NFV-IFA 011, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; VNF Descriptor and Packaging Specification” v3.2.1, April 2019.
[28] ETSI GS NFV-IFA 012, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on Os-Ma-Nfvo reference point - application and service management use cases and recommendations”, v3.1.1, October 2018.
[29] ETSI GS NFV-IFA 013, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Os-Ma-Nfvo reference point - Interface and Information Model Specification” v3.2.1, April 2019.
[30] ETSI GS NFV-IFA 014, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Network Service Templates Specification” v3.2.1, April 2019.
[31] ETSI GR NFV-IFA 022, “Management and Orchestration; Report on Management and Connectivity for Multi-Site Services”, v0.8.2, 2018.
[32] ETSI GS NFV-IFA 028, “Management and Orchestration; Report on architecture options to support multiple administrative domains”, v3.1.1, 2018.
[33] ETSI GR NFV-EVE 012 V3.1.1, “Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework”, 2017.
[34] ETSI GS MEC 001 V1.1.1, “Mobile Edge Computing (MEC): Terminology”, March 2016.
[35] ETSI GS MEC 003, “Mobile Edge Computing (MEC); Framework and Reference Architecture”, v1.1.1, March 2016.
[36] ETSI GS MEC 010-2, “Mobile Edge Computing (MEC); Mobile Edge Management; Part 2: Application lifecycle, rules and requirements management”, v1.1.1, July 2017.
[37] ETSI GR MEC 017, “Mobile Edge Computing (MEC); Deployment of Mobile Edge Computing in an NFV Environment”, v1.1.1, February 2018.
[38] ITU-T TR Y.3011, “Framework of Network Virtualization for Future Networks, Next Generation Networks-Future Networks”, International Telecommunication Union-Telecommunication Standardization Sector (ITU-T), Tech. Rep., 01 2012, v14.0.0.
[39] 3GPP TR 23.799, “Study on Architecture for Next Generation System”, Third Generation Partnership Project (3GPP), Tech. Rep., 12 2016, v14.0.0.
[40] A. de la Oliva *et.al.*, “5G-TRANSFORMER: Slicing and Orchestrating Transport Networks for Industry Verticals", accepted by IEEE Communications Magazine, 2018.
[41] C. Casetti et.al, Network Slices for Vertical Industries, 1st Workshop on Control and management of Vertical slicing including the Edge and Fog Systems (COMPASS), Barcelona, IEEE, 2018.

[42] 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.
[43] 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.
[44] Horizon 2020, 5GPPP: 5G-Crosshaul project, Available at: http://5g-crosshaul.eu/
[45] D. King and A. Farrell, “A PCE-Based Architecture for Application-Based Network Operations”, IETF RFC 7491, March 2015.
[46] M. Björklund, Martin, “YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)”, IETF RFC6020, October 2010.
[47] A. Bierman et al, “Restconf Protocol”, IETF RFC 8040, January 2017.
[48] L. Qiang et al., “Technology Independent Information Model for Network Slicing”, draft-qiang-coms-netslicing-information-model-02, Work in progress, Expires July 30, 2018.
[49] RFC 2119, “Key words for use in RFCs to Indicate Requirement Levels”, IETF, 1997.
[50] Description of Network Slicing Concept by NGMN Alliance, NGMN 5G P1 Requirements & Architecture Work Stream End-to-End Architecture, version 1.0, January 2016.
[51] NGMN Alliance, “Description of Network Slicing Concept”, v.1.0.8, September 2016.
[52] NGMN Alliance, NGMN 5G White Paper, February 2015.
[53] Open Source MANO. https://osm.etsi.org/
[54] J. Baranda et.al., “Resource Management in a Hierarchically Controlled Multidomain Wireless/Optical Integrated Fronthaul and Backhaul Network”, IEEE NFV-SDN conference, November 2017.
[55] 5GEx, https://5g-ppp.eu/5gex/
[56] 3GPP TS 29.060 V15.1.0 (2017-12), General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface.
[57] IETF, RFC7348, Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks, August 2014.
[58] Multi-access Edge Computing, http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing
[59] IETF RFC3198, Terminology for Policy-Based Management, November 2001.
[60] L. Cominardi, "Multi-domain federation: scope, challenges, and opportunities," Workshop in 3rd IEEE Conference on Network Softwarization, Bologna, Italy, July 2017.

Annex A: Updated 5G-TRANSFORMER architecture

This annex presents a self-contained description of the updated 5G-TRANSFORMER architecture.

This annex is structured as follows:

- Section A.1 presents a high-level overview of 5GT scope, stakeholders and offered services.
- Section A.2 describes the architecture requirements, including business and functional requirements.
- Section A.3 presents the high-level 5GT system architecture design, where is not only presented the main building-blocks defined in the 5GT and the interfaces among them, but also discussed relevant aspects regarding service deployment in environments combining multiple technology domains, SLA management and service scaling, and policy management.
- Sections A.4, A.5 and A.6 summarizes the design of, respectively, the 5GT-VS, 5GT-SO and 5GT-MTP building blocks that compose the 5GT architecture.
- Section A.7 presents a set of high-level example workflows that are common to all of the services and provides details on the different interactions between the 5GT building blocks.
- Section A.8 defines the key terminology used in 5G-TRANSFORMER project.
- Section A.9 discusses how composed vertical services can be handled within the 5G-TRANSFORMER system architecture.
- Finally, Section A.10 provides summarizes how the federation across different 5GT systems is addressed.

A.1 5G-TRANSFORMER Overview

This section presents the high-level overview of 5G-TRANSFORMER scope, including the stakeholders and the services offered by the system. Finally, the selected vertical services and use cases to be implemented in the project are presented.

A.1.1 5G-TRANSFORMER Stakeholders

An initial analysis of the 5G-TRANSFORMER business ecosystem has been presented in [1]. Based on the definitions provided by 3GPP [11] and NGMN ([50],[51],[52]), this analysis identifies the different stakeholders and the basic relationships established among them in the next generation of network services. After the gap analysis presented in [1], 5G-TRANSFORMER business ecosystem and stakeholders modelling is closer to the treatment provided by NGMN because of its support of a multi-domain scenario and its more flexible approach to reflect the various possible customer-provider relationships between verticals and operators. Based on these considerations, Figure 2 presents the different stakeholders defined in 5G-TRANSFORMER and their relationship with the system architecture, which will be further described in Section A.3.

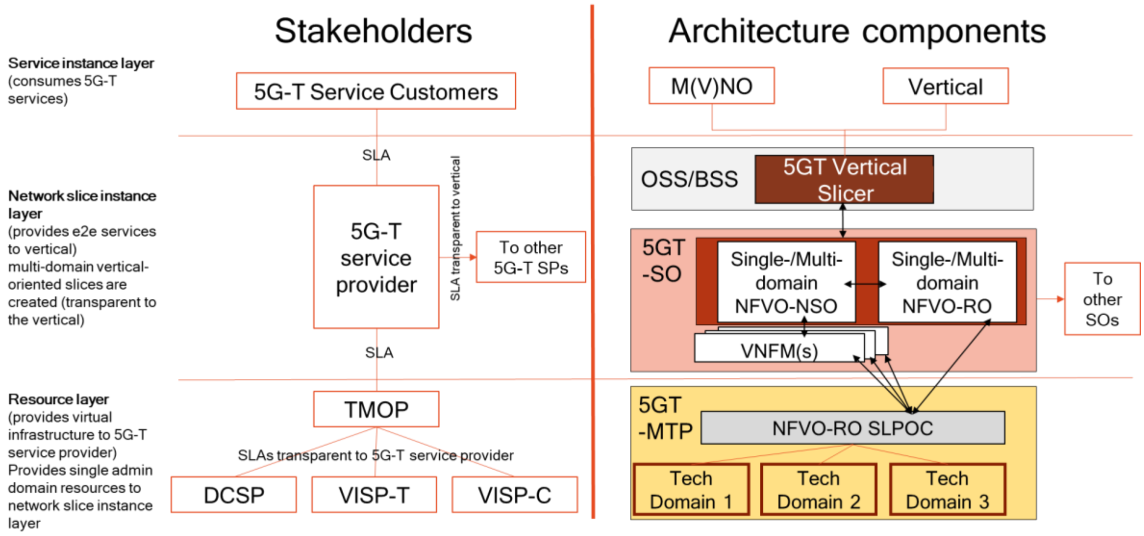


FIGURE 2: 5G-TRANSFORMER STAKEHOLDERS MAPPING WITH THE SYSTEM ARCHITECTURE

The different stakeholders for the 5G- TRANSFORMER system are defined in a top-down approach as follows:

- **5G-TRANSFORMER Service Consumer (TSC)**: uses 5G-TRANSFORMER services (see Section A.1.2 below for a definition) that are offered by a 5GT Service Provider. Note that a 5G-TRANSFORMER Service Provider can also be a TSC of another service provider through federation. In the context of 5G-TRANSFORMER, the main role considered as consumer of services is the vertical industry.
- **5G-TRANSFORMER Service Provider (TSP):** provides 5G-TRANSFORMER services (described in Section A.1.2). Designs, builds and operates its 5GT services.
- **5G-TRANSFORMER Mobile Transport and Computing Platform Operator (TMOP)**: in charge of orchestrating resources, potentially from multiple virtual infrastructure providers and offered to the TSP. In that sense, it acts as an aggregator of resources. The virtual infrastructure features transport and computing resources, potentially including those of datacentre service providers with which the TMOP has an agreement. The TMOP designs, builds, and operates the computing and network aggregated virtual infrastructure services and it has agreements[1] with Virtualization Infrastructure Service Providers (VISPs) (see below for a definition).
- **Virtualization Infrastructure Service Provider (VISP)**: Provides virtualized infrastructure services and it designs, builds and operates its virtualization infrastructure(s) [11]. A VISP can be further specialized depending on the kind of infrastructure it manages: a VISP-T provides virtual transport infrastructures while a VISP-C provides virtual computing infrastructures.
- **Data Centre Service Provider (DCSP):** Provides data centre services and it designs, builds and operates its data centres [11]. The difference between DCSP and VISP-C is that the former is closer to the raw resources (host servers) offering simple services of raw resource consumption. Additionally, these resources are

[1] Initially, it is assumed that TMOPs and VISPs belong to the same administrative domain. This might be different in a general scenario

located in a centralized location (datacentre). The latter offers access to a variety of virtual infrastructure resources created by aggregating multiple technology domains and by making them accessible through a single API for all of them. For instance, VISP-C may offer not only centralized datacentre resources, but also distributed computing resources available throughout the network.

A.1.2 5G-TRANSFORMER Services

From a business perspective, 5G-TRANSFORMER Services (TS) are services focused on a specific industry or group of customers with specialized needs (e.g., automotive services, entertainment services, e-health services, industry 4.0). From a technical point of view, it is a composition of general functions, denoted as Virtual Applications (VA), as well as network functions and defined by its functional and behavioural specification. Hence, a TS provides more general functionalities than just network functionalities.

TS are offered by a 5G-TRANSFORMER Service Provider (TSP) to 5G-TRANSFORMER Service Consumers such as verticals through its northbound interface or to other TSPs through the east-westbound interface (E/WBI). Such services can include a bundle of the different types of services, as explained in the following.

If service requests come through the northbound interfaces requested by the verticals or the Mobile Virtual Network Operators (MVNOs) (to ask for services such as the ones described in Section A.1.3), the following four types of TSs can be distinguished:

- **5G-TRANSFORMER Managed Vertical Service (TMVS)**: These vertical services are fully deployed and managed by the TSP and consumed as such by the vertical (i.e., without any interface available to modify the service logic, but only for getting operational information, at most).
- **5G-TRANSFORMER Unmanaged Vertical Service (TUVS)**: Vertical services are deployed by the TSP (i.e., VNFs and their connectivity), but their logic is partly or fully managed by the vertical. This includes the configuration of VNF internals to control the logic of the vertical services at service level, e.g., the algorithms for EVS (Extended Virtual Sensing) for the automotive use case. In this case, the lifecycle management of the NFV network service and its VNFs is still retained by the TSP.
- **Network Slice as a Service (NSaaS):** to provide a network along with the services that it may support. For instance, a TSP may provide a mIoT network slice as a service, which may support several services, including sensor monitoring, collision avoidance as well as traffic management, and warehouse automation. The TSC (i.e. the NSaaS customer) can, in turn, play the role of a provider itself, and offer to its own consumers its vertical services built on top of the services of the network slice (B2B2X). Based on a mutual agreement, the relevant network slice characteristics and some limited network slice management capability need to be exposed.
- **NFVI as a Service (NFVIaaS):** The tenant (e.g., a vertical or an MVNO) is offered a virtual infrastructure including associated resources (networking/computing/storage) under its full control, in which it can deploy and manage its own NFV network services on top of it. It is assumed that the vertical will deploy its own MANO stack. This is probably the most usual service consumed by M(V)NOs, since they have the knowledge and the need to customize their communication service offering for their

own customers. Resources could be virtual cores, storage, virtual nodes and links, etc.

  - NOTE: The tenant can deploy and connect under its own control VMs on these resources.

Additionally, TSPs can also consume TSs offered by peering TSPs. This interaction is done through the east-westbound interface (E/WBI) of 5GT-SOs. There are two types of service federation:

- **5G-TRANSFORMER Service federation (TSF)**: The consumer TSP uses NFV network services offered by the peer TSP. This may be the case when an end-to-end service is split into constituent services that are deployed in multiple TSP administrative domains.
- **5G-TRANSFORMER Resource federation (TRF)**: The consumer TSP uses NFV (abstracted) virtual network resources offered by the peer TSP. This may be the case when an end-to-end NFVIaaS service is built by combining virtual resources belonging to multiple TSP administrative domains.

A.1.3 Vertical Services

The 5G-TRANSFORMER consortium includes verticals from different industries. In this section we summarize the use cases (UC) selected for demonstration within the project and reasons behind this selection. They are mapped to the service types classified in Section A.1.2. Further details about all the use cases considered can be found in D1.1 [1].

A.1.3.1 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 towards 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 analyse 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 peculiar features outlining the key aspects that mostly impacts on 5G.

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

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 initially selected and proposed for implementation.

TABLE 2: AUTOMOTIVE USE CASES

ID	Name	Goal in context	General description
UC A.01, UC A.02	Extended Virtual Sensing (EVS)	Avoid possible collision crossing intersection.	The purpose of the EVS 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-Trough	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.

An initial analysis of the EVS is reported in the Section 5.1 of D1.1 [1] where a more detailed description of the use case including goals, involved actors, flow, UML UC and sequence diagrams are provided. The analysis of the See-through UC is reported in deliverable D1.2 [2], in the Annex VI: See-Through for Safety in Section 16.

The two use cases have been analysed providing a common draft architecture able to cope both.

Because the architecture is common to both use cases, it was decided to focus the Proof of Concept (PoC) on the implementation of the EVS, as representative of the whole scenario.

The selection of this application for the reference PoC has been done considering:

- the potential benefit that the application could provide to the Road Safety;
- the Governments C-ITS application roadmaps;
- the maturity of the application from an implementation point of view;
- challenges for 5G-TRANSFORMER.

The crash avoidance effectiveness of the EVS has been evaluated, on average, in the order of 50% [16]. This is indeed one of the main reasons why it is listed as priority application of European and North America Market [17]. Moreover, the See-Through application presents some difficulty and complexity due to effective video real-time processing. Finally, the EVS has requirements particularly challenging for the project, especially in terms of latency and MEC functionalities required for realizing the deployment of the applications and the related components in a cost-efficient way, satisfying the QoS requirements set by the vertical.

For the aforementioned reasons, the EVS application is the candidate application to assess the 5G infrastructure and architecture provided within the project. According to the TSs types, introduced in Section A.1.2, EVS is classified as TUVS.

A.1.3.2 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 the society has been experiencing with the explosion of Internet. The amount of users grows daily and the users demand progressively media-rich contents and a better quality of experience, imposing great challenges to the network infrastructures (in terms of data rates, number of connections, quality of experience, etc.) not present before.

For the last years, the entertainment industry has been working on improving fan engagement solutions on sport venues. In particular, the ability to setup smart venues where ultra-high definition interactive media can be delivered massively to mobile devices with minimum latency has been challenging because current wireless communication technologies do not allow to cost-efficiently support a dense concentration of connected devices with intensive data traffic consumption.

The planned demonstrations will focus on the On-site live experience (OLE) and Ultra-high fidelity media use cases, which aim to provide an immersive experience together with an ultra-high definition content. The selection of these two use cases is due to the increasing demand from the customers for this kind of solutions. This makes them particularly interesting for sport event organizers, since they provide promising opportunities for new revenues. At the same time, these use cases are aligned to state of the art expectations from sport fans regarding quality and features of interactive media.

From the technological perspective, the two use cases are also the most challenging ones, presenting the higher impact over the project platform and over the services themselves. First, because the goal of the demonstrations is to expand the network slices to be as close as possible to the sport fans (i.e., using federation, expanding the slice over multiple data centres, etc.) and integrate the services with the MEC. Second, because the OLE UC requires a seamless composition of vertical services, and finally because the services involved require to be scaled automatically and be capable of serving a high amount of users. We refer to D1.1 [1] for further details about both UCs, and to D5.1 [9] for further details about the demonstrations.

In terms of service classification UC.E01 and UCE02 are both classified as TUVS, as established in Section A.1.2, since the TSP in this case aims to keep certain control over the service logic and rely on the platform for the service management.

TABLE 3: ENTERTAINMENT USE CASES

ID	Name	Goal in context	General description
UC.E01	On-site live event experience	To provide a better fan experience to	Large scale event sites, such as stadiums are more and more being connected in order to give

	(OLE)	users attending (on-site) an event	better experience to their customers (replay, choose a specific camera, language, augmented reality to bring additional information, etc.)
UC.E02	Ultra-high fidelity media	To guarantee a high-quality of content experience for Ultra High Fidelity Media in both closed and open venues.	The next generation of media consumption will be driven by the high definition. Consumers (fans) will demand 4k and 8k quality in their media consumption through their user devices. Both linear (e.g. live programming, streaming) and non-linear (e.g. on-demand) content will be used for providing this Ultra High Fidelity Media experience.

A.1.3.3 eHealth

The eHealth is one of the industries that can effectively take advantage of the future 5G networks to improve the quality of life and medical assistance of people in emergency situations. Hence, we consider two main targets: e-Infrastructure and eHealth application. Both are considered as TUVS type of vertical services.

On one hand, the e-Infrastructure use case focuses on how the current municipality infrastructure based on TETRA can be replaced exploiting the novel 5G features. This will allow not only emergency alarms to be received with smaller delay and thus be processed in a small amount of time, but also to access in real time the clinical history of the patient from the place of the incident to give the patient a better medical attention. In addition, the eHealth use case will need a high-priority and low latency service in the 5G-TRANSFORMER system. To address this, the 5G-TRANSFORMER system will allow to have access to the resources of the emergency system in extreme cases where the network is overloaded by users like in big events.

On the other hand, the eHealth application aims to study how new technologies such as Edge Computing can help improving the speed and the quality of response. This application tries to reduce the response time and automate the processes of communicating between the patient and the medical personnel and among the medical personnel.

TABLE 4: EHEALTH USE CASES

ID	Name	Goal in context	General description
UC.H.01	Heart attack emergency	To provide a better medical assistance in emergency cases	Emergencies that requires 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.

This use case has been selected as it poses the key challenges for eHealth and provides immediate advantages from the use of 5G. It covers the main requirements from the eHealth partner of the project, SAMUR, summarized next: (i) voice communications substituting TETRA system, (ii) improved location of the ambulances, and, (iii) ambulance activation and tracking system, capable of exchanging with the ambulance service commands, patient's medical history, information about recommended navigation routes , information about accident site, environmental risks, and receiving remote monitoring information from the ambulance activity (e.g., speed, use of acoustic signals and priority warning lights, breakdowns).

A.1.3.4 eIndustry

The production and manufacturing industry are 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 processes.

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.

In such ever-changing environment, wired connectivity would mean complex cabling and high operational expenditure for upgrading that cabling. Thus, a new wireless connectivity that allows to reduce costs and infrastructures and meets the tight requirements in terms of bandwidth and latency becomes fundamental. In particular, 5G wireless connectivity, with its standardized networking capabilities, built-in security, guaranteed grades of service as well as network slicing concepts, is therefore a perfect tool for advanced industries that want to take advantage of digital transformation.

Data rates, latency, reliability and positioning accuracy requirements on industrial wireless communications are increasing continuously. The upcoming 5G mobile network will be able to meet enhanced requirements as it targets data rates of up to ten gigabits per second, latencies under one millisecond, extremely high communication reliability, and/or elevated accuracy in positioning.

For the maximum flexibility in a production plant, new kinds of cloud driven robots and a massive number of connected sensors will be deployed. The manufacturing processes will be continuously monitored through wireless connectivity 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 (5G-MTP).

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.

Among the identified eIndustry use cases, Table 5 presents the cloud robotics as candidate for implementation in 5G-TRANSFORMER project.

The main reason behind this decision is that this use case is the most challenging from a communication network point of view. Indeed, the Cloud Robotics use case poses severe requirements on the underlying communication network, making it essential for the industrial environment to be equipped with 5G solutions. The increasing need for customization of manufacturing process will require more flexible production plant and, as a consequence, centralised control functionalities in the cloud will be required to optimize processes and implement lean manufacturing.

Moving the control into the cloud, it is possible to utilize its massive computing power, but at the same times very low latency and high bandwidth will be required to transfer instantaneously a huge amount of information.

According to the TSs types introduced in Section A.1.2, Cloud Robotics is classified as TUVS.

TABLE 5: EINDUSTRY USE CASES

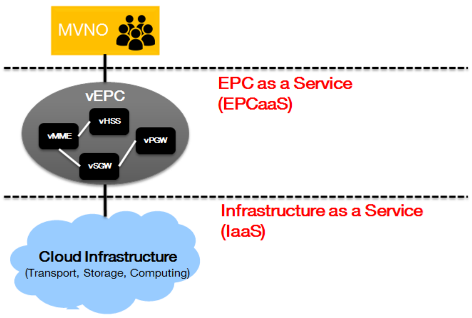
ID	Name	Goal in context	General description
UC I.02	Cloud Robotics	Highly automation of the factory plant is provided moving the control of the production processes and of the robots functionalities in cloud, exploiting wireless connectivity to minimize infrastructure, optimize processes, implement lean 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.

A.1.3.5 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 Mobile Network Operator's (MNO) business toward new markets and a large variety of tailored services. This evolution is especially 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 Mobile Virtual Network Operator (MVNO) business model emerges from this evolution through the creation of a new business model with new players that disrupts the traditional mobile value chain, capturing the interest of 5G-TRANSFORMER. The vEPC use case reflects perfectly the concept of network slicing as defined in 3GPP TS 28.530 [14] in terms of business service types. Actually, a Virtual Service Provider (VSP), i.e., a MNO, can offer to its customers (MVNOs) LTE services in the form of a Network Slice as a Service (NSaaS), exposing also a set of specific management functions. MVNOs can in turn provide their own services on top of the vEPC services: this use case matches with the B2B2X business service type. In addition, it highlights the difference between the two types of clients interacting with the 5GT-VS: MVNO vs Vertical. Unlike the MVNO, a Vertical customer has no knowledge of the underlying network characteristics that is being deployed to support its business services, which makes the vEPC an essential use case to demonstrate the 5G-TRANSFORMER features.

In D1.1 [1], several use cases have been identified as relevant for the MNO/MVNO domain in 5G-TRANSFORMER. We chose to focus here on the vEPCaaS use case presented in Table 6. However, it is anticipated that 5G will cover several telco and vertical use cases, not targeting pure mobile services only but also fixed, IoT, and convergent ones'. This requires flexibility to provision, deploy and manage core networks adapted the specific target services. For example, a vEPC tailored to support fixed/mobile broadband access requires specific functions and dimensioning, different from another one supporting only IoT data, and then a third one dedicated to emergency services. Even when the three network instances are supported on the same / virtualized infrastructure. This motivates the choice of focusing on the vEPC use case to experiment how it can be instantiated on demand in a dedicated network slide, to serve different types of MNO/MVNO and relevant services.

TABLE 6: MNO/MVNO USE CASES

ID	Name	Goal in context	General description
UC M.01	vEPCaaS	Creation of an MVNO service through the deployment and operation of a network slice with a vEPC in “as a Service” mode. 	The vEPC can be instantiated as a virtualized Control plane only or as a complete virtualized Control and User planes core network. Based on mutual agreement, limited relevant management are also exposed.

A.2 5G-TRANSFORMER Architecture Requirements

Technical requirements on the 5G-TRANSFORMER system and the vertical services have been defined already in [1]. These requirements have focused on properties of the vertical services and the corresponding network slices. Additional general requirements on the 5G-TRANSFORMER system, emerged during the lifetime and development of the project, which are described in this deliverable. In Section A.2.1 we present requirements from the business perspective point of view, in Section A.2.2 we present functional requirements corresponding to the different phases of the lifecycle of vertical services. The definition of the requirements follows the methodology described in Annex I in Section 11 of deliverable D1.2 [2].

A.2.1 Business Requirements

The 5GT-VS forms part of the business front-end of the 5G-TRANSFORMER system. It is a component that directly interacts with the customer through the service request, which is internally transformed into a network slice instantiation.

The initially identified business related requirements are identified in Table 7.

TABLE 7: INITIAL BUSINESS REQUIREMENTS

ID	Requirement	F/NF
ReqT.B.01	The 5G-TRANSFORMER system should include a portal to interface with the vertical customer.	F
ReqT.B.02	The 5G-TRANSFORMER system shall support different kinds of vertical customers.	NF
ReqT.B.03	The 5G-TRANSFORMER system shall be open and extensible to support any new kind of vertical customer.	NF
ReqT.B.04	The 5G-TRANSFORMER system should be able to accept the service specification from the vertical customer including both functional and non-functional requirements expected for the requested service.	F
ReqT.B.05	The 5G-TRANSFORMER system shall be able to deploy a vertical service instance using resources of multiple administrative and technological domains [2]. This shall be transparent to the vertical customer.	F
ReqT.B.06	The 5G-TRANSFORMER system shall expose appropriate interfaces to external customers to allow them to consume services offered by the 5G-TRANSFORMER system.	F
ReqT.B.07	The 5G-TRANSFORMER system shall adhere to industry multi-tenancy requirements including isolation, scalability, elasticity and security, where security is meant to provide protection to prevent attacks, denial of service or information leaking.	NF
ReqT.B.08	The 5G-TRANSFORMER system shall allow to negotiate and monitor service SLAs, with appropriate granularity according to the final service characteristics.	F
ReqT.B.09	The 5G-TRANSFORMER system shall provide vertical customers with service catalogue information about available service offers and capabilities, in order to facilitate the automated provision of services.	F
ReqT.B.10	The 5G-TRANSFORMER system shall provide a mechanism to perform vertical service accounting and charging. This information should be available internally and externally (for the vertical	F

[2] To use resources of multiple administrative domains federation among these domains is needed. Even with a single administrative domain, multiple technological domains might have to be used.

	customer).	
ReqT.B.11	The 5G-TRANSFORMER system should be able to support long-live and short-lived services.	F
ReqT.B.12	The 5G-TRANSFORMER system should be reliable.	NF
ReqT.B.13	The 5G-TRANSFORMER system should be available (as carrier class component providing 5 nines availability).	NF
ReqT.B.14	The 5G-TRANSFORMER system should keep responsiveness for vertical customer requests. The 5G-TRANSFORMER system should provide response times as an interactive system.	NF
ReqT.B.15	The 5G-TRANSFORMER system shall allow blueprints composed of other blueprints.	F
ReqT.B.16	The 5G-TRANSFORMER system shall allow VSDs composed of other VSDs.	F
ReqT.B.17	The 5G-TRANSFORMER system shall allow a vertical to specify which vertical service instance to use in a composed vertical service.	F
ReqT.B.18	The 5G-TRANSFORMER system shall support the specification of preferred, non-preferred, and prohibited virtual infrastructure providers.	F
ReqT.B.19	The 5G-TRANSFORMER system shall allow a vertical to define whether a child service of a composed service instance has the same lifecycle as the parent service instance or whether it has a lifecycle of its own.	F
ReqT.B.20	The 5G-TRANSFORMER system shall allow a vertical to define whether a vertical service instance can be a child service of several composed services, i.e., whether it can be shared among other vertical services.	F
ReqT.B.21	The 5G-TRANSFORMER system shall support to specify the deployment area based on KPIs[3] of another service.	F

Complementary to the requirements above, new requirements have been incorporated to the 5G-TRANSFORMER design and solution, which are presented in Table 8.

TABLE 8: ADDITIONAL BUSINESS REQUIREMENTS

ID	Requirement	F/NF
ReqT.B.22	The 5G-TRANSFORMER system, in its different components and modules, should be supported by different implementations to avoid solution lock-in, either proprietary or open source.	NF
ReqT.B.23	The 5G-TRANSFORMER system should allow vertical-driven service composition.	F

[3] As an example, intersection collision avoidance should cover critical intersections, where ‘critical’ is defined in terms of occurrence of abrupt braking maneuvers in the past.

A.2.2 Functional Requirements

The 5G-TRANSFORMER system is involved in the service lifecycle at different phases, each having different requirements. The phases are discovery, fulfilment, assurance, and decommissioning. All the requirements in this section are functional requirements. This section includes also new requirements for the distinct phases referred before.

A.2.2.1 Discovery

The discovery phase facilitates the 5G-TRANSFORMER system to understand what capabilities and services are supported. That information will be exposed to the vertical customers for 5G-TRANSFORMER service offering.

The initially identified requirements regarding the discovery phase are presented in Table 9.

TABLE 9: INITIAL REQUIREMENTS ON THE DISCOVERY PHASE

ID	Requirement
ReqT.Di.01	The 5G-TRANSFORMER system must provide vertical customers with the means to submit detailed requests including information regarding the location of resources and service points, QoS, charging options.
ReqT.Di.02	Service catalogue entries may contain a service manifest and a price tag (or an indicative price range from which the exact price can be extracted at run-time).
ReqT.Di.03	The 5G-TRANSFORMER system shall provide the customer with the ability to request a service from the catalogue along with the expected SLA.
ReqT.Di.04	Service catalogue entries and satisfied service requests should result in an SLA commitment for the respective service.
ReqT.Di.05	The service request (and associated SLA) must contain a parameter describing the service aging or Time To Live (TTL).
ReqT.Di.06	The 5G-TRANSFORMER system must support both private (i.e., towards specific vertical customer(s)) and public dissemination of service offers.
ReqT.Di.07	The 5G-TRANSFORMER system should provide a mechanism to set-up, re-size and terminate services.
ReqT.Di.08	The 5G-TRANSFORMER system shall support a vertical to create several instances of the same vertical service.
ReqT.Di.09	The 5G-TRANSFORMER system shall allow a vertical to store its service descriptions persistently, and to create, retrieve, update, and delete vertical service descriptions[4].
ReqT.Di.10	The 5G-TRANSFORMER system shall allow the 5G-TRANSFORMER

[4] A vertical may provide the location of virtual machine images of its virtual applications as part of its service descriptions. These images may have to be certified by the 5G-TRANSFORMER system provider before actually onboarding them to the 5G-TRANSFORMER system.

	service provider to define vertical service blueprints.
ReqT.Di.11	The 5G-TRANSFORMER system shall store and keep up-to-date a catalogue of NFVI-PoPs available within its administrative domain and of related resources (computing, storage, networking) in addition to available PNFs/VNFs.
ReqT.Di.12	The 5G-TRANSFORMER system shall monitor the (current) state of available PNFs and keep track of the history of states of available PNFs.
ReqT.Di.13	The 5G-TRANSFORMER system shall certify the credentials of entities accessing its NFVI catalogue.
ReqT.Di.14	The 5G-TRANSFORMER system shall allow to create/delete different instances of VNF(s) and update/retrieve VNFDs

In addition to the requirements above, some additional requirements are identified for the discovery phase, which are presented in Table 10.

TABLE 10: ADDITIONAL REQUIREMENTS ON THE DISCOVERY PHASE

ID	Requirement
ReqT.Di.15	The 5G-TRANSFORMER system should integrate diverse technological environments such as RAN and MEC.
ReqT.Di.16	The 5G-TRANSFORMER system should facilitate different levels of federation, such as service and resource federation, involving multiple parties.

A.2.2.2 Fulfilment

During the service fulfilment phase, the 5G-TRANSFORMER system produces the necessary mapping from the customer request to the network slice templates, i.e., network service descriptors, to properly instruct the 5GT-SO.

The initially identified requirements regarding the fulfilment phase are presented in Table 11.

TABLE 11: INITIAL REQUIREMENTS ON THE FULFILMENT PHASE

ID	Requirement
ReqT.Fu.01	Depending on the modality [5] of the contracted service, the 5G-TRANSFORMER system could be required to offer proper configuration and management interfaces to the slice capabilities honouring the service request, in order to manage them similarly as if they were owned and dedicated resources (tenant-managed slices).
ReqT.Fu.02	Depending on the modality of the contracted service, the 5G-TRANSFORMER system could accommodate a vertical customer service request in some existing slice (provider-managed slices).

[5] Modality refers to the possibility of requesting tenant-managed network slices (i.e., request of control and management capabilities of the allocated resources functions) or provider-managed network slices (i.e. the control and management is retained by the provider and the tenant simply uses the network slice).

ReqT.Fu.03	The 5G-TRANSFORMER system shall support the vertical to express policies[6]. The 5G-TRANSFORMER system shall enforce such policies.
ReqT.Fu.04	The 5G-TRANSFORMER system could allow the vertical customer to specify[7] policies associated to the service to describe e.g. elasticity rules to be enforced when the service requires re-deployment/re-configuration.
ReqT.Fu.05	The 5G-TRANSFORMER system shall support a vertical to manage the lifecycle of each of its vertical service instances separately.
ReqT.Fu.06	The 5G-TRANSFORMER system shall provide configuration and management interfaces to instantiated VNFs or available PNFs.
ReqT.Fu.07	The 5G-TRANSFORMER system shall allow VNF scaling (up/down/in/out).
ReqT.Fu.08	The 5G-TRANSFORMER system shall allow resource scaling (up/down/in/out).
ReqT.Fu.09	The 5G-TRANSFORMER system shall provide appropriate isolation and access guarantees to available PNFs.
ReqT.Fu.10	The 5G-TRANSFORMER system shall certify the credentials of entities accessing its NFVI.
ReqT.Fu.11	The 5G-TRANSFORMER system shall maintain information regarding the mapping between NSD, VNFs/PNFs and allocated resources, along with the state of PNFs and VNFs.

Complementing the previous requirements, a number of new requirements has been identified for the fulfilment phase, which are presented in Table 12.

TABLE 12: ADDITIONAL REQUIREMENTS ON THE FULFILMENT PHASE

ID	Requirement
ReqT.Fu.12	The 5G-TRANSFORMER system should permit the extension of connectivity among different NFVI-PoPs from multiple providers at both intra- and inter-domain levels.
ReqT.Fu.13	The 5G-TRANSFORMER system should allow system-driven service composition (i.e., the decision on composition is taken by the 5G-TRANSFORMER system itself).
ReqT.Fu.14	The 5G-TRANSFORMER system should permit the dynamic on-boarding of NSDs.

A.2.2.3 Assurance

The 5G-TRANSFORMER system informs the vertical customer about events and performance of the vertical service instances. As a consequence, the 5G-TRANSFORMER system should gather monitoring information and expose it to the

[6] QoS policies, charging policies, security policies, and regulatory policies where needed.
[7] As an example of such policies, it could be possible to specify the concentration of resources by night for an efficient management of energy consumption during low workloads.

vertical, who could take different actions accordingly. In addition, the monitoring information can be also consumed internally by the 5G-TRANSFORMER system to trigger events like arbitration.

The initially identified requirements regarding the assurance phase are presented in Table 13.

TABLE 13: INITIAL REQUIREMENTS ON THE ASSURANCE PHASE

ID	Requirement
ReqT.As.01	The 5G-TRANSFORMER system must provide the vertical customer with tools to monitor the QoS attained for the requested service.
ReqT.As.02	The 5G-TRANSFORMER system should provide a mechanism for the vertical for reward/penalty in service provisioning in case of SLA conformance/failure.
ReqT.As.03	This 5G-TRANSFORMER system should be able to map the associated SLA into KPIs[8] to be monitored during the slice execution lifecycle.
ReqT.As.04	The 5G-TRANSFORMER system should be able to support lawful interception mechanisms.
ReqT.As.05	The 5G-TRANSFORMER system should prevent DoS attacks from a malicious behavior from vertical customers.
ReqT.As.06	The 5G-TRANSFORMER system should provide isolation among vertical customers' procedures and requests.
ReqT.As.07	The 5G-TRANSFORMER system shall allow dynamically set-up a traffic monitoring service in any given 5G network slice.
ReqT.As.08	The 5G-TRANSFORMER system shall arbitrate resources among vertical service instances of one vertical based on priorities and SLA requirements.
ReqT.As.09	The 5G-TRANSFORMER system shall be able to collect performance information and manage fault information, even when vertical services are deployed across multiple administrative and technological domains.
ReqT.As.10	The 5G-TRANSFORMER system shall monitor the QoS attained to instantiated VNFs, allocated PNFs, and related resources.
ReqT.As.11	The 5G-TRANSFORMER system shall provide isolation and performance guarantees among tenants, even when sharing PNFs.
ReqT.As.12	The 5G-TRANSFORMER system shall be tolerant to failures and report failure events whenever issues cannot be solved.

Apart from those requirements, some new requirements have been elicited for the assurance phase, which are presented in Table 14.

[8] The SLA will be expressed in terms of service parameters, while the KPIs represent technical parameters. For example, service latency (as part of the SLA) could require to monitor transport network latency but also VM execution latency (as two different KPIs to monitor).

TABLE 14: ADDITIONAL REQUIREMENTS ON THE ASSURANCE PHASE

ID	Requirement
ReqT.As.13	The 5G-TRANSFORMER system must provide means for optimizing the deployment of network services in terms of resources, performance indicators, etc, by means of intelligent capabilities and artifacts for function placement (including multi-domain environments), arbitration, etc.
ReqT.As.14	The 5G-TRANSFORMER system must support service scalability for assuring proper service delivery.

A.2.2.4 Decommissioning

The 5G-TRANSFORMER system also participates in the proper release of resources and services to the vertical once a service instance is no longer required.

The identified requirements regarding the decommissioning phase are presented in Table 15.

TABLE 15: REQUIREMENTS ON THE DECOMMISSIONING PHASE

ID	Requirement
ReqT.De.01	The 5G-TRANSFORMER system should be able to identify the slice(s) to be decommissioned as result of service termination.
ReqT.De.02	The 5G-TRANSFORMER system should be able to identify the monitoring mechanisms to be deactivated as result of service termination.
ReqT.De.03	The 5G-TRANSFORMER system should have means for receiving acknowledgement of releasing resources.
ReqT.De.04	The 5G-TRANSFORMER system should be able to notify the vertical customer about service instance termination.
ReqT.De.05	The 5G-TRANSFORMER system should be able to keep track of charging and accounting information even after service termination for periods of time according to local laws.
ReqT.De.06	The 5G-TRANSFORMER system shall be able to identify and deallocate resources allocated to VNFs upon a VNF termination procedure.
ReqT.De.07	The 5G-TRANSFORMER system shall be able to identify the monitoring mechanisms to be deactivated as a result of a VNF termination procedure or resource deallocation procedure

A.3 5G-TRANSFORMER System Architecture

This chapter introduces the proposed 5G-TRANSFORMER system architecture. Section A.3.1 presents a high-level design of the main components and introduces their key functional roles as well as the interfaces among them. The details of the updated design of individual components have been reported in the 5G-TRANSFORMER D2.3 [4], D3.3 [6] and D4.3 [8] deliverables. They are summarized at the architecture level in this deliverable in the following sections in order to give a complete view of the whole system architecture design. Besides, the reference architectures from 3GPP, ETSI NFV and ETSI MEC have been studied in D2.1 [3], D3.1 [5] and D4.1 [7]. A glossary on the relevant terminologies is provided in Section A.8.

A.3.1 Baseline Architecture Design

The 5G-TRANSFORMER project explores how the network can better serve the needs of 5G-TRANSFORMER customers (i.e., vertical industries including 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., the service 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 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 [18], 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 [19]) or a group of such services. By means of network slicing, the system creates different slices tailored to the needs of different vertical services, which share the underlying 5G mobile transport and computing infrastructure, hence enhancing 5G-TRANSFORMER provider network 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 (see Section A.1.1) by introducing open Application Programming Interfaces (API) among components. 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 (E/WBI), which enable service and resource federation across different administrative domains, allowing 5G-TRANSFORMER service providers (TSP) 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 3. The 5GT-VS is the entry point for the vertical requesting a service, it is the building block in the 5GT system to provide vertical support. More specifically, it provides a vertical front-end to handle service requests. It also handles the translation of vertical service requirements to network slice requirements and mapping of these services to slices, as well as responsible for network slice management. The 5GT-SO is the orchestrator to offer the 5GT-VS the network slice services by provisioning and orchestration the end-to-end services within the local domain or across multiple domains 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 cloud and network resources over the infrastructure on which service components are eventually deployed. It also decides on the abstraction level offered to the 5GT-SO, enabling hierarchical orchestration decisions and network management.

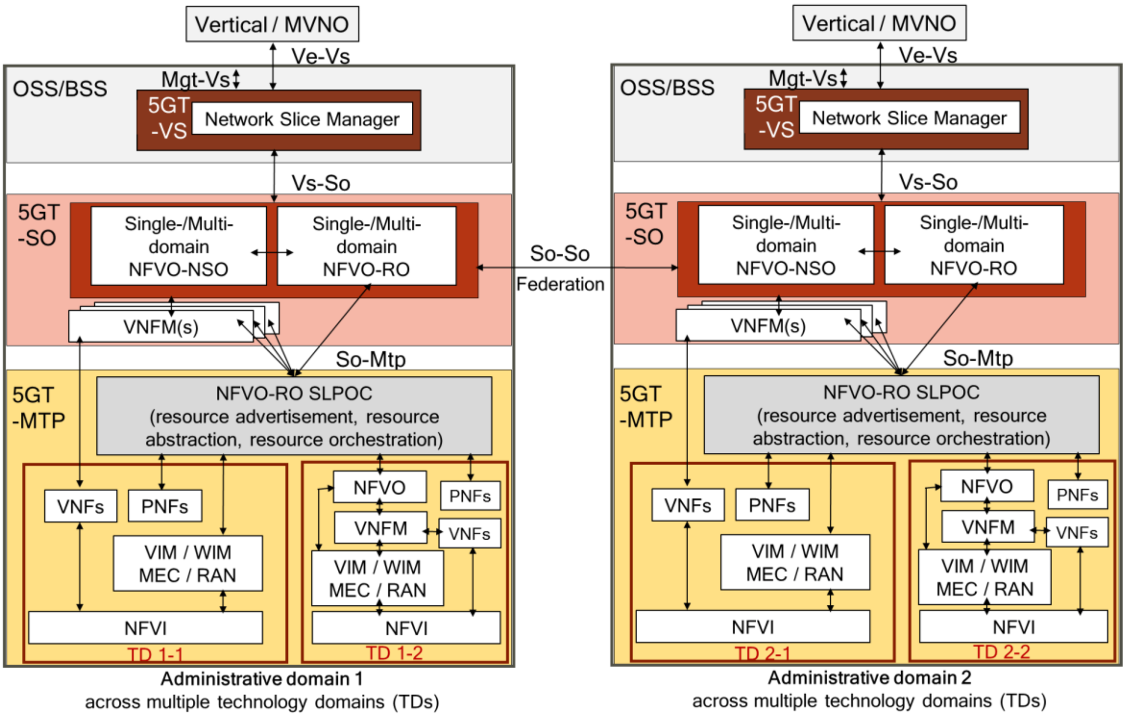


FIGURE 3: 5G-TRANSFORMER SYSTEM ARCHITECTURE

A.3.1.1 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-TSP [1]. Vertical services are offered through a high-level interface at the 5GT-VS northbound, which is designed to allow verticals to focus on the service logic and requirements, without caring about how their services are eventually deployed at the resource level. This latter issue would be up to the TSP. Therefore, vertical services will use those services offered by the TSP. In fact, the 5GT-VS offers a catalogue of vertical service blueprints (VSB), based on which the vertical service requests are generated by the vertical. Additionally, the 5GT-VS offers a set of basic services as building blocks, which the vertical can combine itself to a vertical service. 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 [11], a decision is taken on whether the service can be provided through an already existing slice or a new one needs to be created.

The vertical slicer is the component of the system that is aware 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 central 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 [11]. More specifically, the NSMF is in charge of lifecycle management of network slice instances. All possible relations 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 this slice-related 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 3) takes ETSI GS NFV-IFA 013 [29] as basis.

A.3.1.2 Service Orchestrator (5GT-SO)

The request of instantiating or managing a NFV-NS from the 5GT-VS is received by the 5GT-SO through the Vs-So interface. The main duty of the 5GT-SO [42] is to provide end-to-end orchestration of the NFV-NS along with their life-cycle management 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., in case of not enough local resources), the 5GT-SO interacts with peer 5GT-SOs located in 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 single 5GT-VS to access the system. The 5GT-SO hides this federation from the 5GT-VS and thus from the verticals.

The 5GT-SO embeds the network service orchestrator (NFVO-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 [20].

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 NFVO-NSO.

In case the NFVO-NSO decides to deploy a network service across multiple administrative domains (requiring service and/or resource federation), the 5GT-SO receiving the request becomes the parent NFVO-NSO and sends ETSI GS NFV-IFA 013 [29] requests for each of the constituent NFV-NSs to other NFVO-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 peer 5GT-SOs of other administrative domains (abstracted). The NFVO-RO will decide on the placement of the VNFs of the NFV-NS and their required infrastructure resources based on the information available in the NFVI resources repository and the NFV-NS instances repository. The NFVI repository will include information about the infrastructure controlled by the local 5GT-MTP as well as resources offered by peer domains. Most likely, the information available in this repository will be more detailed when coming from the local 5GT-MTP rather than from a federated domain. The interaction to the 5GT-MTP is based on ETSI GS NFV-IFA 005 [22]. This also includes the interface with an additional NFVO-RO inside the 5GT-MTP (i.e., TD 1-2 in Figure 3), as explained below.

Notice that the NFVI resources handled by the NFVO-RO of the 5GT-SO based on which decisions are taken will have a higher or lower abstraction level depending on the policies applied 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, as they are closer to the actual resources.

The 5GT-SO also includes the Virtual Network Function Managers (VNFM) to manage the lifecycle of the VNFs composing the NFV-NS. ETSI GS NFV-IFA 006-based interfaces [23] 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 5GT-SOs for resource allocation requests involving the VNFs under its control.

A.3.1.3 Mobile Transport and Computing Platform (5GT-MTP)

The 5GT-MTP [43] 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. Therefore, the 5GT-MTP is in charge of managing the virtual resources on which the NFV-NSs are deployed. The deployment in the first phase manages resources of transport and data centre sites, while in the second phase it also handles MEC and Radio resources.

The NFVO-RO acts as single entry point, i.e., single logical point of contact (SLPOC) in ETSI GS NFV-IFA 028 [32] terminology, for any resource allocation request coming from the 5GT-SO. The So-Mtp interface is based on ETSI GS NFV-IFA 005 [22] and ETSI GS NFV-IFA 006 [23]. 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 VNFs under its control. Such interfaces are extended to provide radio and MEC abstract resources as well as the allocation/termination operation over MEC and radio domains.

In addition, an additional interface based on IFA032 is introduced to manage network connectivity between VNF deployed in different Datacentre sites

In terms of managing VNF instances, the So-Mtp interface also consists of ETSI GS NFV-IFA 008-based interfaces (i.e., the Ve-Vnfm-vnf reference point) [25] to allow the VNFM of the 5GT-SO to directly configure the VNF instances deployed in the 5GT-MTP.

Depending on the use case, the 5GT-MTP may offer different levels of resource abstraction to the 5GT-SO. The abstraction is unified and in a common view shows all the heterogeneity of the resources (compute, storage of datacentre sites, transport, mobile radio resources). 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 [22] 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, interconnect them, and/or configure the relevant parameters of the PNFs that form the NFV-NS. IFA005 is used also to allocate resources in MEC and mobile edge domains. As special case, the mobile and transport network are abstracted in combined way to offer an unified and homogenous view of the infrastructure resources (both mobile and transport) for the vertical services. In order to represent in the abstraction the concept of mobility in a certain geographical area and report the corresponding infrastructure resources (mobile and transport), the Virtual Link concept used for the transport, is extended with the concept of coverage area. This option is offered to hide the complexity of both transport and mobile network to the rest of the system whilst keeping the required flexibility inside the mobile domain (e.g., to let the specific domain decide on the most appropriate functional split of the base station functions). Therefore, a full ETSI MANO stack is represented in technology domain 1-2 (see Figure 3) 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 an abstract link. To note that the management of technology domain 1-2 has been completed in the second phase by means of the local placement algorithm defined in the MTP and detailed in the D2.3 [4].

A.3.1.4 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-NSs, 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 4, 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-NS, and these metrics can be used by the 5GT-SO to take scaling decisions for the involved VNFs. This procedure is adopted, for example, in the case of SLA Management (Section A.3.3), where the SLA Manager component within the 5GT-SO triggers auto-scaling procedures based on rules defined in the Network Service Descriptor and depending on monitoring metrics collected and processed by the 5GT-SO monitoring service.

On the other hand, the performance metrics computed at the 5GT-SO monitoring service can be also 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 related to monitoring data belonging 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.

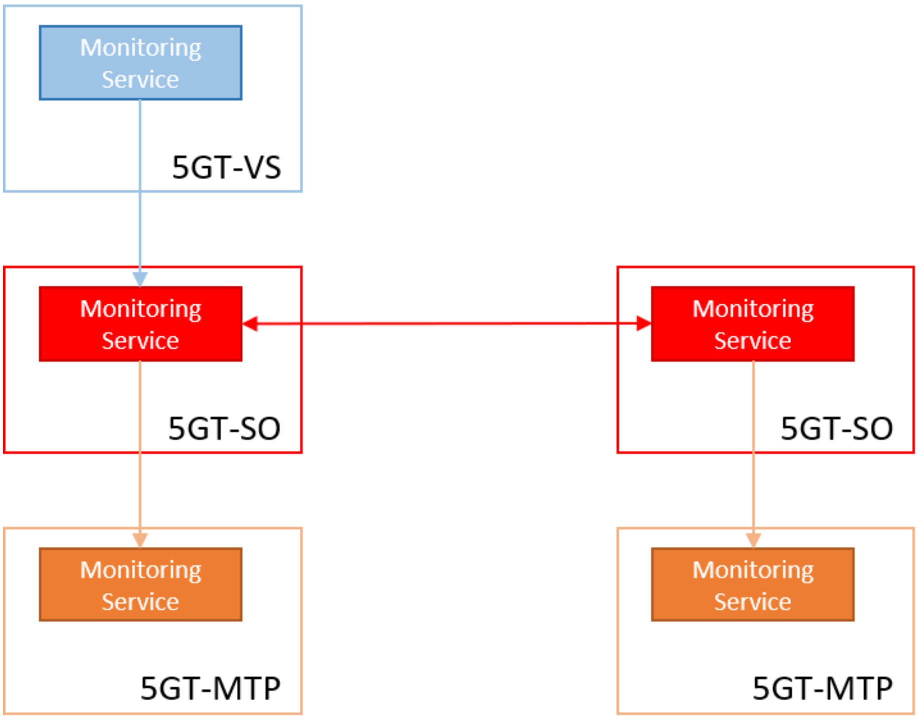


FIGURE 4: HIERARCHY OF MONITORING SERVICES IN 5G-TRANSFORMER ARCHITECTURE

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 behaviour 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 cases where the 5GT-VS and the 5GT-SO are owned and managed by the same administrative entity, a single monitoring platform may be deployed for both of them.

A.3.1.5 Interfaces and References Points

This section introduces the design of the interfaces between the main building blocks, describing their interactions and the corresponding reference points in the 5G-TRANSFORMER architecture.

A.3.1.5.1 Northbound of the 5GT-VS

The 5GT-VS provides the interface of the 5G-TRANSFORMER system to the customer as well as to the TSP. Therefore, the northbound interface (NBI) of the 5GT-VS is as well the NBI of the 5G-TRANSFORMER system. Two reference points are defined at the northbound of the 5GT-VS (see Figure 5):

- **Ve-Vs**, between a vertical and the 5GT-VS. This reference point provides the mechanisms to allow the vertical to retrieve VSBs, to manage Vertical Service Descriptors (VSD), to request operational actions on vertical service instances (VSI), like instantiation, termination, modification, and to monitor performance and failures of instantiated vertical services.
- **Mgt-Vs**, between the TSP's OSS/BSS Management Platform and the 5GT-VS. This reference point provides primitives to manage tenants, SLAs and VSBs. It is used mainly for management and administrative issues and it is handled internally within the 5G-TRANSFORMER service provider.

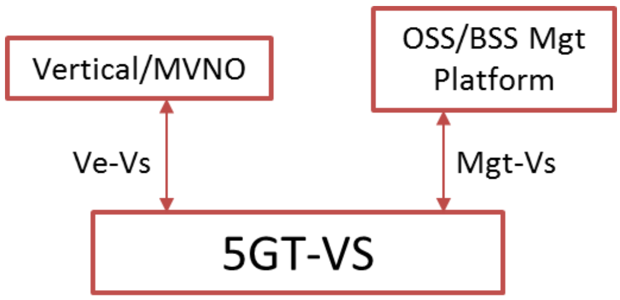


FIGURE 5: REFERENCE POINTS ON THE NORTHBOUND OF THE 5GT-VS

The 5GT-VS NBI implementing both reference points is a REST API based on HTTP and JSON messages, where the 5GT-VS acts as REST server and the verticals and the OSS/BSS Management Platform act as REST clients. The interface should also support asynchronous notifications from the 5GT-VS, for example based on web sockets. Suitable mechanisms for authentication and authorization of the entities issuing the requests should also be supported. However, this section focuses only on the main functionalities supported at each reference points, while the full specification of the primitives and their abstract messages is documented in D3.3 [6] and the protocol messages encoding will be specified during the software design and implementation phase.

Note, the 5GT-VS NBI specified in this section focuses on the provisioning of Vertical Services. For NSaaS we consider network slices of 5G access and core networks. Management of these slices can be done by the vertical or MVNO through a dedicated management service access point. For NFVIaaS we consider 5G core networks, where specific functionality such as Home Subscriber Server (HSS) is provided by the MVNO. These subsets of the more generic NSaaS and NFVIaaS can be described by vertical service blueprints and descriptors and handled within the 5GT-VS as described below.

The Ve-Vs reference point identifies the following operations:

- Query VSBs.
- Create, query, update, and delete VSDs.
- Instantiate, query, terminate, modify Vertical Service Instances (VSI).
- Notifications about vertical service lifecycle events.
- Query monitoring parameters for VSIs.
- Subscriptions/notifications about vertical service monitoring parameters.
- Notifications about vertical service failures.

As an example, we provide the definition of the Query VS blueprints operation. All operations at the reference point are specified in D3.3 [6]. The Query VS blueprints operation allows a vertical to retrieve one or more VSBs from the 5GT-VS catalogue. The blueprints are then used by the vertical to create the VSDs for the vertical services to be instantiated.

The Query VS blueprints messages are specified in Table 16.

TABLE 16: QUERY VS BLUEPRINTS MESSAGES

Message	Direction	Description	Parameters
Query VS blueprint request	Vertical → 5GT-VS	Request to retrieve one or more VSBs matching the given filter.	• Filter (e.g. VSB ID, …) Vertical ID.
Query VS blueprint response	5GT-VS → Vertical	Response including the details of the requested VSBs.	• List<VSB>

The Mgt-Vs reference point between the OSS/BSS Management Platform (Mgt in the following) and the 5GT-VS identifies the following operations:

- Create, query and delete tenants.
- Create, query, modify and delete SLAs.
- Create, query and delete VSBs.

As for the Ve-Vs reference point, all the operations at this reference point are specified in D3.3 [6].

Beyond these operations, the OSS/BSS Management Platform has also access in read mode to the information related to all the entities managed by the 5GT-VS, i.e., VSDs, VSIs, NSIs and NSSIs, which can be retrieved from the related catalogues and records.

A.3.1.5.2 Vs-So Interface

The 5GT-SO NBI refers to the interface between the 5GT-VS and the 5GT-SO, based on the ETSI NFV IFA 013 interface (reference point *Os-Ma-nfvo* between the OSS/BSS and the NFVO in the NFV MANO architecture) [29]. This interface is used for: (i) network service lifecycle management and forwarding of information related to the status of the NFV network services; (ii) management of NFV NS Descriptors, VNF packages and PNF Descriptors (PNFDs); (iii) monitoring of Network Service Instances (NFV-NSI); and (iv) policy management.

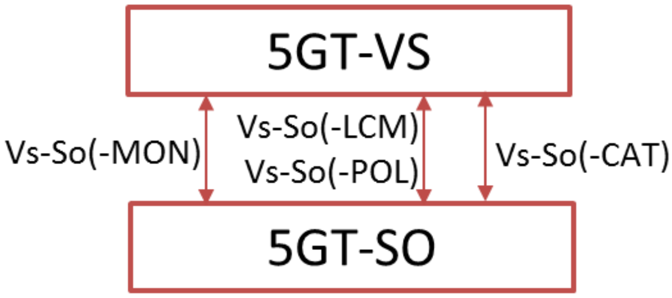


FIGURE 6: REFERENCE POINTS BETWEEN 5GT-VS AND 5GT-SO

The interactions between the 5GT-VS and 5GT-SO implement the four reference points shown in Figure 6 and described below:

- **Vs-So (-LCM - LifeCycle Management)** is used for operations on NFV-NSs. It offers primitives to instantiate, terminate, query, and reconfigure NFV network service instances or receive notifications about their lifecycle.
- **Vs-So (-MON - MONitoring)** is used for the monitoring of NFV-NS and VNF instances through queries or subscriptions and notifications about performance metrics. Additionally, this interface provides APIs for the fault management.
- **Vs-So (-CAT - CATalogue)** is used for the management of NFV NSD, PNFDs, VNF and MEC Application package descriptors, including the onboarding, removal, updates and queries.
- **Vs-So (-POL - POLicy)** is used for the management of policies. It offers primitives to transfer, delete, activate, deactivate, associate, disassociate policies and receive notifications about policy conflicts.

The mapping between the four Vs-So reference points and the specific interfaces defined in the ETSI NFV IFA 013 is reported in deliverable D3.3 [6], together with the required extensions in terms of descriptors' information model (e.g. in support of MEC applications integrated in NFV-NSs) and interfaces' information elements (e.g. in support of geographical or latency constraints specification).

A.3.1.5.3 So-Mtp Interface

The So-Mtp Interface (5GT-SO southbound interface (SBI) as seen by the 5GT-SO and 5GT-MTP northbound interface (NBI) as seen by the 5GT-MTP) addresses the interworking between the 5GT-SO and the 5GT-MTP building blocks of the 5G-TRANSFORMER architecture. In the refined architecture, a single 5GT-SO interacts with a single 5GT-MTP via the defined So-Mtp interface. In a nutshell, such an interface allows eventually handling the configuration and programmability of a number of resource domains (including virtualized resources for compute, storage and networking) being coordinated by the underlying 5GT-MTP. Besides managing the utilization (i.e., de/allocation) of the virtualized resources, the So-Mtp interface provides the required functionalities for deploying (updating and terminating) demanded VNFs by a given NFV-NS.

As depicted in Figure 7, the So-Mtp interface can be split into different subsets of interfaces providing different functions. In this regard, the So-Mtp interface has been designed to effectively support within a single and common API multiple operations and functionalities derived from the required interworking and communication between 5GT-SO elements and the 5GT-MTP entity.

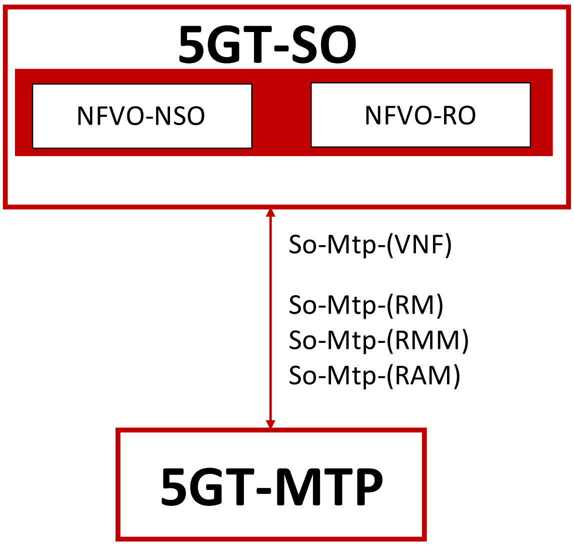


FIGURE 7: REFERENCE POINTS FOR 5GT-SO SBI (I.E., SO-MTP INTERFACE)

Formerly, the design of the So-Mtp interface was conceived to be based on (or leverage at most) existing set of standard documents being produced within the ETSI NFV framework, namely ETSI GS NFV-IFA 005 [22], ETSI GS NFV-IFA 006 [23] and ETSI GS NFV-IFA 008 [25]. However, we realized that specific functions and requirements exclusively in the scope of handling virtualized network resources between NFVI-PoPs (i.e., within the WAN interconnecting remote NFVI-PoPs) were not fully supported by these standard documents. In this context, some efforts were devoted to better cope with the intricacies and characteristics to manage the network resources in the WAN by means of a WIM controller. Documents such as ETSI GR NFV-IFA 022 [31] provided some ideas in this respect. Regardless of the considered implementation, the following high-level operations and functions have been defined and deployed to be supported by the So-Mtp interface in terms of a set of messages and workflows:

- Providing abstracted information (e.g., capacities, availability, connectivity, etc.) of the virtualized resources managed by each 5GT-MTP. This entails both internal NFVI-PoP resources (e.g., aggregated amount of available CPU or memory, etc.) as well as inter-NFVI-PoP network resources. For the latter, networking links interconnecting a given pair of NFVI-PoPs are referred to as logical links. For each derived logical link handled in the 5GT-SO and exposed by the 5GT-MTP specific attributes such as the available bandwidth and total delay are provided. Observe that this information will be necessary in the Placement Algorithm execution at the 5GT-SO.

- Managing (i.e., instantiation, allocation, scaling up/down and release) of the virtualized resources required to support both 5GT-SO operations. For instance, this entails allocating network resources in the logical links connecting a pair of NFVI-PoPs, or allocating compute resources within a NFVI-PoP that will eventually host a specific VNF.
- Supporting the lifecycle management (i.e., creation, configuration, modification and termination) of the VNFs to be placed in selected NFVI-PoPs.
- Supporting fault management and performance monitoring jobs that involves the resources and infrastructures allocated in VIM, WIM, Radio domains for a service.

According to the previous operations to be covered by the So-Mtp interface, these are divided into the following subset of interfaces: **So-Mtp(-RAM)** provides the Resource Advertisement Management functions; **So-Mtp(-RM)** encompasses the Resource Management operations over the virtualized resources; **So-Mtp(-RMM)** provides the Resource Management and Monitoring operations; **So-Mtp(-VNF)** takes over the general VNF lifecycle management (e.g., scaling up/down a particular VNF instance, fixing VNF malfunctions, etc.). For more details, please, refer to D2.3 [4].

In light of the above defined pool of functionalities and operations to be supported over the So-Mtp interface, its implementation (i.e., API) was designed to leverage on:

1. selected messages compliant with the current standard ETSI IFA 005 specification. Such messages support the operations for intra-NFVI-PoP virtualized resource management (e.g., allocating / terminating virtualized compute resources, instantiating VNFs, etc.).
2. the use of a proprietary 5GT-designed set of messages for handling the interactions related to inter-NFVI-PoP operations that are better tailored to the intricacies and features of the 5GT solution (e.g., the management and abstraction used by the 5GT-SO in terms of logical links between a pair of NFVI-PoPs' gateways). It is worth stating that the design of those inter-NFVI-PoP-oriented messages was done inspired and reusing (as much as possible) some information elements actually supported in the ETSI IFA 005.

A.3.1.5.4 So-So Interface

The 5GT-SO provides the interface of the 5G-TRANSFORMER system to another external 5G-TRANSFORMER system. Therefore, the eastbound/westbound interface (E/WBI) of the 5GT-SO is as well the E/WBI of the 5G-TRANSFORMER system. Six reference points are defined at the E/WBI of the 5GT-SO (see Figure 8):

- **So-So**(-**L**ife **C**ycle **M**anagement), between consumer 5GT-SO NFVO-NSO and provider 5GT-SO NFVO-NSO. This reference point provides the mechanisms to instantiate, terminate, query or re-configure nested NFV-NS or receive notifications for federated nested NFV-NS.
- **So-So**(**-MON**itoring), between consumer 5GT-SO NFVO-NSO and provider 5GT-SO NFVO-NSO. This reference point provides monitoring of nested NFV-NS through queries or subscription/notification of performance metrics, VNF indicators and NFV-NS failures.
- **So-So(-Cat**alogue**)**, between consumer 5GT-SO NFVO-NSO and provider 5GT-SO NFVO-NSO. This reference point provides primitives to subscribe/notify for

changes, queries of descriptors (NSDs and AppDs) and packages (MEC Application Packages).

- **So-So(-R**esource **M**anagement**),** between consumer 5GT-SO NFVO-RO and provider 5GT-SO NFVO-RO. This reference point provides the operations for configuration of the resources, configuration of the network paths for connectivity of VNFs/VMs. These operations mainly depend on the level of abstraction applied to the actual resources.
- **So-So(-R**esource **M**onitoring **M**anagement**),** between consumer 5GT-SO NFVO-RO and provider 5GT-SO NFVO-RO. This reference point provides monitoring of different resources, computing power, network bandwidth or latency, storage capacity, VMs, MEC hosts provided by the peering administrative domain. The details level depends on the agreed abstraction level.
- **So-So(-R**esource **A**dvertising **M**anagement**),** between consumer SO-SO Resource Advertisement and provider SO-SO Resource Advertisement. This reference point provides the mechanism for advertising available resource abstractions to/from other 5GT-SOs. Periodic or event-triggered updates for near real-time availability of resources.

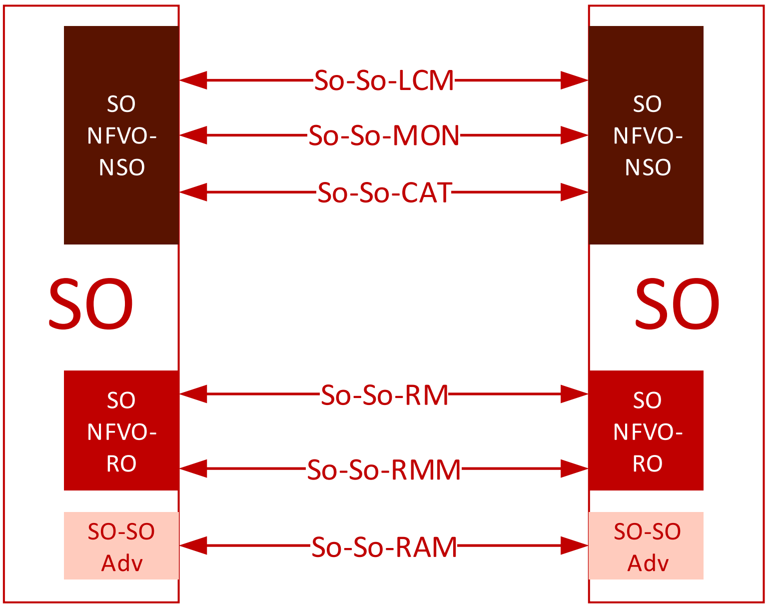


FIGURE 8: REFERENCE POINTS FOR 5GT-SO E/WBI (I.E.,SO-SO INTERFACE)

The 5GT-SO E/WBI implementing all reference points is a REST API based on the HTTP protocol and JSON messages, where the provider 5GT-SO acts as REST server and the consumer 5GT-SO acts as REST client. The interface should also support asynchronous notifications from the provider 5GT-SO, for example based on web sockets. Suitable mechanisms for authentication and authorization of the entities issuing the requests should also be supported. However, this section focuses only on the main functionalities supported at each reference points, while the full specification of the primitives and their abstract messages is documented in D4.3 [8]. The 5GT-SO E/WBI specified in this section focuses on the provisioning of the Network Service as a Service (NFV-NSaaS) and the NFVI as a Service (NFVIaaS) cases.

The **So-So-LCM** reference point implements the following operations specified in IFA 013 [29]:

- Instantiate, query, terminate, modify nested NFV-NSs.
- Subscribe/Notify about nested NFV-NSs lifecycle change notification.

The **So-So-MON** reference point implements the following operations specified in IFA 013 [29]:

- Subscribe/notify, query about nested NFV-NSs performance information.
- Create, delete, query nested NFV-NSs threshold operation.
- Subscribe/notify about nested NFV-NSs fault alarms.

The **So-So-CAT** reference point implements the following operations specified in IFA 013 [29]:

- Query, update, subscribe/notify changes on NSDs.
- Query, update, subscribe/notify changes of MEC Application Package

The **So-So-RM** reference point implements the following operations specified in IFA 005/006/008 [22][23][25]:

- Add, query, update, delete software image operations.
- Allocate, query, update, terminate, operate on compute/network/storage resources. Due to abstraction of the resources, operations and information are limited to the abstracted level.
- Create, instantiate, scale, scale to level, change, terminate, delete, query, operate, modify, get operation status of VNFs. Due to abstraction of the resources, operations and information are limited to the abstracted level.

The **So-So-RMM** reference point implements the following operations specified in IFA 005/006/008 [22][23][25]:

- Query, subscribe/notify, create/delete/query threshold operation for performance information on compute/network/storage resources. Due to abstraction of resources, operations and information are limited to the abstracted level.
- Subscribe/notify for lifecycle change of VNFs.
- Subscribe/notify, query, create/delete/query threshold performance operation of VNFs/VMs.
- Subscribe/notify for fault alarms of VNFs/VMs.

The **So-So-RAM** reference point implements the following operations specified in IFA 005/006 [22][23]:

- Query quota information for compute/network/storage resources. Due to abstraction of resources, information is limited to the abstract level.
- Subscribe/notify for change of compute/network/storage resources. Due to abstraction of resources, information is limited to the abstract level.

The reference points on the E/WBI are based on operations defined in the ETSI NFV IFA documents, however there are some gaps not covered there. ETSI NFV-EVE 012 [33] addresses the lack of support for service availability levels, priority handling for virtual resource assignment as well as service continuity during healing procedures of a service. These gaps are currently out of scope of the project. However, as a future work the So-So-RAM reference point should be extended to enable dynamic exchange of resource abstractions (available for federation) either by implementing APIs for connection to a central point or mechanisms to establish peer-to-peer network.

A.3.1.6 Services Mappings to Network Slices and NFV Network Services

This section presents the process of instantiating a vertical service in the 5G-TRANSFORMER system through the different building blocks.

Figure 9 shows the high-level workflow for instantiation of a vertical service through these modules, more detailed workflows are shown in Section A.7.2. During service onboarding phase, the TSP defines a set of vertical service blueprints in a vertical service catalogue offered to the tenants (verticals or MVNOs) through service advertisement.

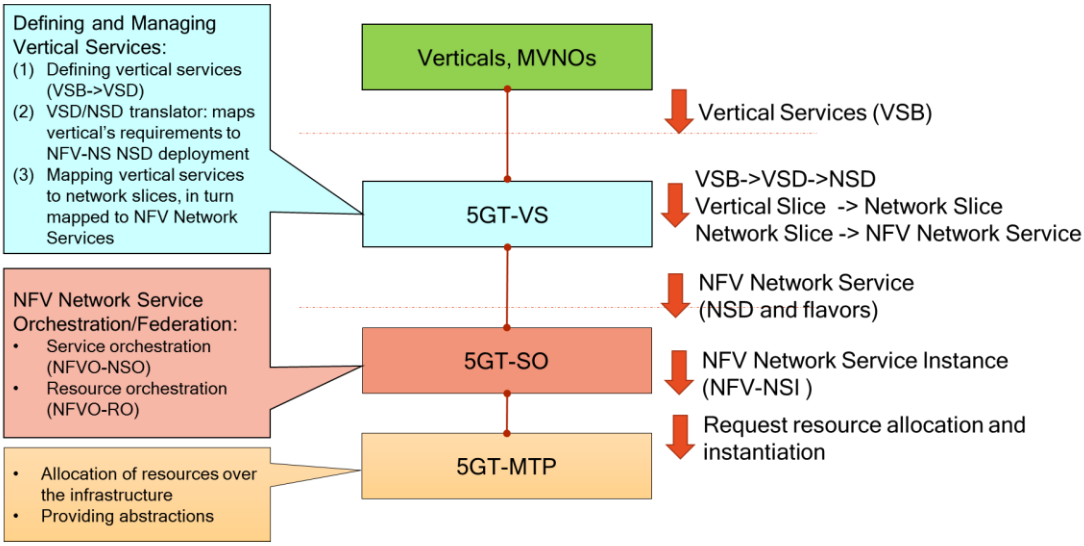


FIGURE 9: FROM VERTICAL SERVICE TO NETWORK SLICE TO NFV NETWORK SERVICE

As an alternative, the vertical may compose a vertical service from predefined building blocks. By providing missing parameters of the building blocks a VSD is created.

To instantiate a vertical service based on the VSD, the tenant selects a VSB from the catalogue of the 5GT-VS via its NBI (i.e., Ve-Vs as described in Section A.3.1.5.1). Thereafter, the tenant completes the VSB to a high-level service description of the vertical service, the VSD. To instantiate this vertical service, the 5GT-VS translates the high-level requirements to network slice level requirements and maps the vertical service instance to (existing or new) network slice instance(s). In the 5G-TRANSFORMER system, a network slice would be directly mapped to a NFV network service instance (NFV-NSI), and the network slice characteristics will be described by a NFV Network Service Descriptor. In this context, creating a network slice instance will involve defining a NFV network service descriptor including its deployment flavours. This is performed by the VSD/NSD translator inside the 5GT-VS. In turn, the process for instantiating the requested NFV-NS is initiated from the 5GT-VS by sending a request to the 5GT-SO via the Vs-So interface (see Section A.3.1.5.2), requesting the 5GT-SO to instantiate the NFV-NS according to the contents of the NSD. This request typically contains the pointer to the NSD and the deployment flavours as well as additional input parameters (e.g. IP addresses to be assigned to some of the network functions) and constraints (e.g. location where to deploy all or some of the network functions). The instantiated flavour may be determined by the 5GT-VS or the 5GT-SO. In this latter case, it will be done based on the autoscaling rules that are present in the NSD.

The 5GT-SO is the main entity to provision and manage the NFV NSI. It is in charge of receiving the network service instantiation requests from the vertical slicer, which characteristics are defined through the network service descriptor and instantiation parameters (e.g., deployment flavour). It is also in charge of handling the service and resource orchestration workflow in a coordinated way. In this direction, the 5GT-SO receives the service structure and requirements from the 5GT-VS and requests the available infrastructure (e.g., network topology, available computation power) from the 5GT-MTP, and through the placement algorithm (PA), which maps these requirements to the infrastructure. Eventually, the service components, VNFs and virtual links (VLs), requested by an NS are deployed over the underlying infrastructure in locations where they fulfil the requirements towards a successful service delivery. Furthermore, whenever there is an agreement with a peering administrative domain offering services to the local domain, the 5GT-SO interacts with other 5GT-SOs via the So-So interface (see Section A.3.1.5.4) to request service federation across multiple administrative domains.

Thereafter, the 5GT-SO will request the 5GT-MTP to perform the actual resource allocation and instantiation. The 5GT-MTP is mainly responsible for the actual allocation/instantiation/control/configuration of virtual resources (including networking, computing and storage resources) over the underlying infrastructure.

Figure 10 shows examples of mapping the vertical services to NFV network services in the 5G-TRANSFORMER system. Different VSIs compose one or multiple vertical services, which are mapped to different network slices (existing or new) and in turn mapped to different NFV network service instances. Different VSIs (e.g., VSI 2 and VSI 3) can map to instances of the same type of NFV-NS with different deployment flavours (VSI 2 -> NFV-NSI 3 and VSI 3 ->NFV-NSI 4 respectively) or of instances of different types of NFV-NSs. Scenarios where a VSI is mapped to a set of concatenated or nested NFV-NSs can also be envisioned (VSI 1 -> NFV-NSI 1 connecting to NFV-NSI 2). The scenario of composed services is studied in the Section A.9.

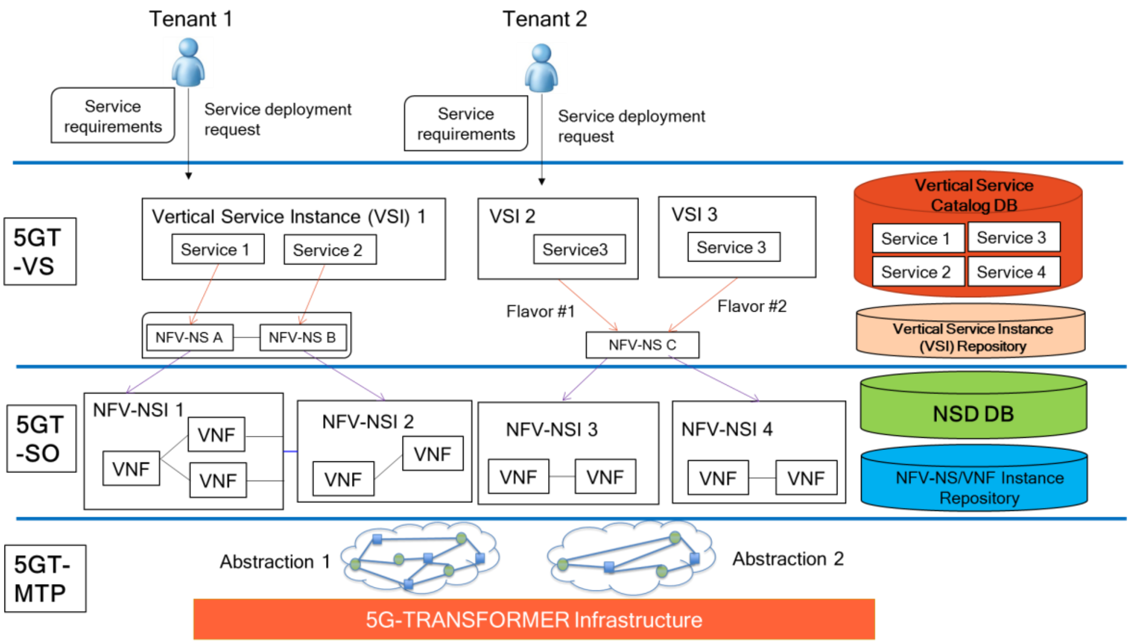


FIGURE 10: EXAMPLES OF SERVICE MAPPING

A.3.1.7 Network Slice and Network Slice Management

Network slicing has captured much attention within the research community and the industry, as well as the Standards Development Organizations (SDOs) such as Next Generation Mobile Network (NGMN), 3GPP, and International Telecommunication Union - Telecommunication Standardization Sector (ITU-T).

The NGMN considers network slicing as a central part of 5G network architecture. NGMN has refined the concept from a top-down design that relies on a 3-layer perspective; namely the service instances layer, the network slice layer, and resources layer. The network slice instances are built with the combination of sub-network instances, eventually shared among multiple network slices. To describe this mapping, NGMN uses network slices blueprints (templates). On top of a network slice instance, multiple service instances can run (typically, verticals with similar characteristics). 3GPP defines the network slicing as a technology that “enables the operator to create networks, customized to provide optimized solutions for different market scenarios which demand diverse requirements, e.g. in terms of functionality, performance and isolation” [38]. From the ITU-T perspective, the network slicing is seen as logical isolated network partitions composed of virtual resources, isolated and equipped with a programmable control and data plane [39].

To sum up; a network slice is considered a complete logical network including Radio Access Network (RAN) and Core Network (CN). It enables the operator to create customized networks to provide optimized solutions for different scenarios that demand different requirements, especially about the functions, performance, and isolation. A network slice describes a system behaviour implemented by a network slice instance (NSI), built from a network slice template (NST), which represents the network functions and resources to provide telecommunication services. Within 5G-TRANSFORMER we use NFV NSDs to define NSTs.

According to 3GPP, a network slice has to verify several requirements, namely 1) completeness of a NSI in the sense that it should include all functionalities and resources necessary to support some set of communication services; 2) a NSI contains network functions (VNFs and/or PNFs) and all the connections information between NF (e.g. topology connections, individual link requirements (e.g. QoS attributes)); 3) resources used by a NSI, a NSI is realized via the required physical and logical resources; 4) instance-specific policies and configuration are required when creating a NSI. Ultra-low latency and ultra-reliability are examples of network characteristics; and 5) a NSI can be fully or partially isolated from another NSI. The isolation can be logical and/or physical.

The 5GT-VS implements the Network Slice Management Function (NSMF) and Network, Slice Subnet Management Function (NSSMF) and performs the mapping between VS instances and network slice instances, as well as the coordination between their lifecycles. The NSMF and the NSSMF manage respectively, the lifecycle of network slice and network slice subnet instances. It consists of several phases:

1. **Preparation Phase:** includes the design templates, the planning capacity, the on-boarding, and evaluation of the network slice requirements, as well as the preparation of the network environment.
2. **Commissioning Phase:** consists of NSI provisioning, which includes the NSI creation, wherein all the needed resources for the slice whether shared or not are allocated and configured. The creation of an NSI can include the creation and/or the modification of the NSI constituents.
3. **Creation Phase:** is a composition of several activities, namely: activation, supervision, performance reporting, resource capacity planning, modification, and de-activation. The activation makes the NSI ready to support the traffic for vertical services. The resource capacity planning activity includes any action that calculates resource usage based on an NSI provisioning, and performance monitoring and generates modification polices as a result of the calculation. The NSI modification may include the capacity or topology changes. The modification represents the creation or modification of NSI constituents. NSI modification can be triggered by receiving new network slice requirements or as the result of supervision/reporting. The deactivation includes actions that make the NSI inactive and stops the vertical services.
4. **Decommissioning Phase:** consists of the decommissioning of the non-shared constituents and removing the NSI specific configuration from the shared constituents. After the decommissioning phase, the NSI is terminated and does not exist anymore.

To realize the aforementioned phases of lifecycle management, the NSMF and the NSSMF uses operations of the Vs-So-Lcm interface.

A.3.1.8 Federation across Multiple Administrative Domains

Federation is a mechanism for integrating multiple administrative domains at different granularity into a unified open platform where the federated resources and services can trust each other at a certain degree. An administrative domain is a collection of network services and resources operated by a single organization. The administrative domain is viewed as a single compact entity, where its internal structure is hidden or unimportant from outside. The resources and services inside the administrative domain operate with high degree of mutual trust among themselves, but the interaction with other administrative domains is subject to stringent trustworthiness constraints, with a default high level of alert. The federation is formed to increase the degree of trust among different administrative domains with a goal of better interoperability of services and resources. Embodiment of a service/business-level agreement or partnership between two administrative domains is a federation of trust.

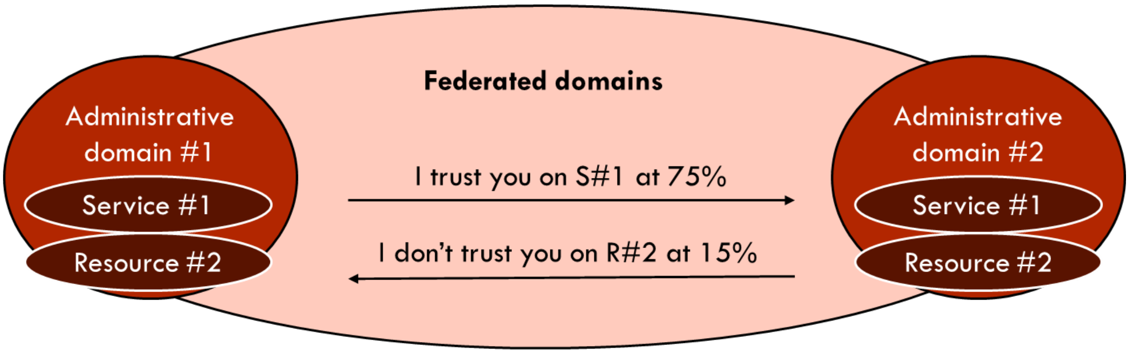


FIGURE 11: FEDERATION AS A DOMAIN UNIFIED BY MUTUAL TRUST

A.3.1.8.1 Federation across 5G-TRANSFORMER systems

From an architectural perspective, the federation in 5G-TRANSFORMER occurs only at the 5GT-SO level, where all the connections with the federated domains are established by the 5GT-SO via EBI/WBI.

For the NSaaS, the 5GT-SO NFVO-NSO is the enabling point of NSaaS for consuming or providing external (federated) NFV-NSs. The external connection between 5GT-SO NFVO-NSOs (belonging to different administrative domains) is realized by using the first three reference points of the E/WBI (So-So-LCM, So-So-MON, So-So-CAT) addressed in Section A.3.1.5.4. These reference points are based on the IFA 013 interface. Two types of NSaaS can be established: service delegation and service federation.

For the NFVIaaS, the federation is done between two 5GT-SO NFVO-ROs of two different ADs through the next three reference points of the E/WBI (So-So-RM, So-So-RMM, So-So-RAM) described in Section A.3.1.5.4. These reference points are based on the interfaces IFA 005 and IFA 006 used for allocation/orchestration of resources. The closest option to the 5G-TRANSFORMER approach is the SLPOC (Single logical point of contact) approach referred as SLPOC-F (Single logical point of contact for federation). The 5GT-SO NFVO-RO of the provider AD has the role of the SLPOC of the provider AD. The consumer 5GT-SO NFVO-RO can request NFVI resources to the provider 5GT-SO NFVO-RO via the three mentioned interfaces (So-So-RM, So-So-RMM, So-So-RAM), dedicated to resource federation operations. Both approaches, direct and indirect (described in Section A.10.1), are feasible for implementation. In direct mode the consumer 5GT-SO NFVO-RO and the consumer VNFM will use the reference points of the E/WBI to communicate with the provider SLPOC-F or specifically the provider 5GT-SO NFVO-RO. In the indirect case, the consumer 5GT-SO NFVO-RO would act as a proxy to the consumer VNFM using IFA 007 interface (as described in Section A.10.1). In that case the communication between different ADs for resource federation procedures is only done between the consumer 5GT-SO NFVO-

RO and the provider 5GT-SO NFVO-RO (which is the provider SLPOC-F) on the E/WBI.

Compared to the MLPOC case (as described in Section A.10.1), the SLPOC-F functionality of the provider 5GT-SO NFVO-RO allows to act as a proxy point to the underlying provider MTP (which can be referred as the VIM in the use case described in Section A.10.1) both in direct and indirect mode.

The choice of the VNF management mode is a trade-off between architectural complexity and overall performance of the VNFM. In direct mode, the connection from the consumer VNFM to the provider 5GT-SO is implemented using procedures defined in ETSI NFV IFA 006. In parallel the connection between the consumer 5GT-SO NFVO-RO and provider 5GT-SO NFVO-RO is defined using the ETSI NFV IFA 005 procedures. In the indirect case, as described above, the consumer 5GT-SO NFVO-RO has the proxy role. In that case, the connection between the consumer VNFM and the consumer 5GT-SO NFVO-RO is re-using the procedures defined in the ETSI NFV IFA 007. The consumer 5GT-SO NFVO-RO processes the VNFM requests received on the IFA 007 based internal interface and adapts them to send on the existing (IFA 005 based) interface to the provider 5GT-SO NFVO-RO. The direct mode solution increases the interface complexity and implementation of the E/WBI APIs. The consumer VNFM is exposed to the external administrative domains that demands application of higher security (e.g., MTP proxy to ensure VNFM protection). In indirect mode, high computational overhead is introduced in the consumer 5GT-SO NFVO-RO. This overhead would increase linearly with the number of consumed federated services which threats the 5GT-SO NFVO-RO of being a bottleneck in the 5G-TRANSFORMER system. Outlined, for low number of consumed federated services the indirect mode seems like a good solution, however for higher number of consumed federated services the direct mode increases the performances (e.g., with lower delays) at a cost of higher security and higher interface complexity.

A.3.1.8.2 Federation with non 5G-TRANSFORMER systems

To cover a large set of clients, the 5G-TRANSFORMER system should be able to interact with different systems to provide a wide range of services. In the federation case, we distinguish the 5G-TRANSFORMER system as a provider or a consumer of services.

A.3.1.8.2.1 5G-TRANSFORMER system as a consumer

Figure 12 depicts the case of federation between a 5G-TRANSFORMER Administrative Domain (AD) and a non 5G-TRANSFORMER AD, with the 5G-TRANSFORMER AD acting as a consumer that requests NFV-NS services or virtualized resources from the non 5G-TRANSFORMER AD.

According to the conclusions stated in Section A.3.1.8.1, the 5G-TRANSFORMER AD is able to federate with a non 5G-TRANSFORMER AD through the 5GT-SO. This later can interact with an NFV Orchestrator (NFVO) and/or VIM via an SLPOC functionality. In case of NSaaS, the federation is done between the NFVO-NSO of the 5GT-SO and the NFVO of the provider via the reference point Or-Or using the IFA 013 interface. Regarding the NFVIaaS case, we distinguish the two cases of VNF management. In the direct mode, the VNFM inside the 5T-SO may interact either with the NFVO or the VIM/WIM (according to the location of the SLPOC functionality) through the reference point Vi-Vnfm using the interface IFA 006. Whilst, for the indirect mode, the VNFM invokes virtualized resource management operations on the NFVO (inside the 5GT-SO) through IFA 007 interface, which in turn invokes them towards the SLPOC on the provider side via the reference point Or-Or using IFA 005 interface. Concerning the resource orchestration, this is done between the NFVO-RO on the 5G-TRANSFORMER AD and the NFVO of the provider, through the reference point Or-Or via the IFA 005 interface.

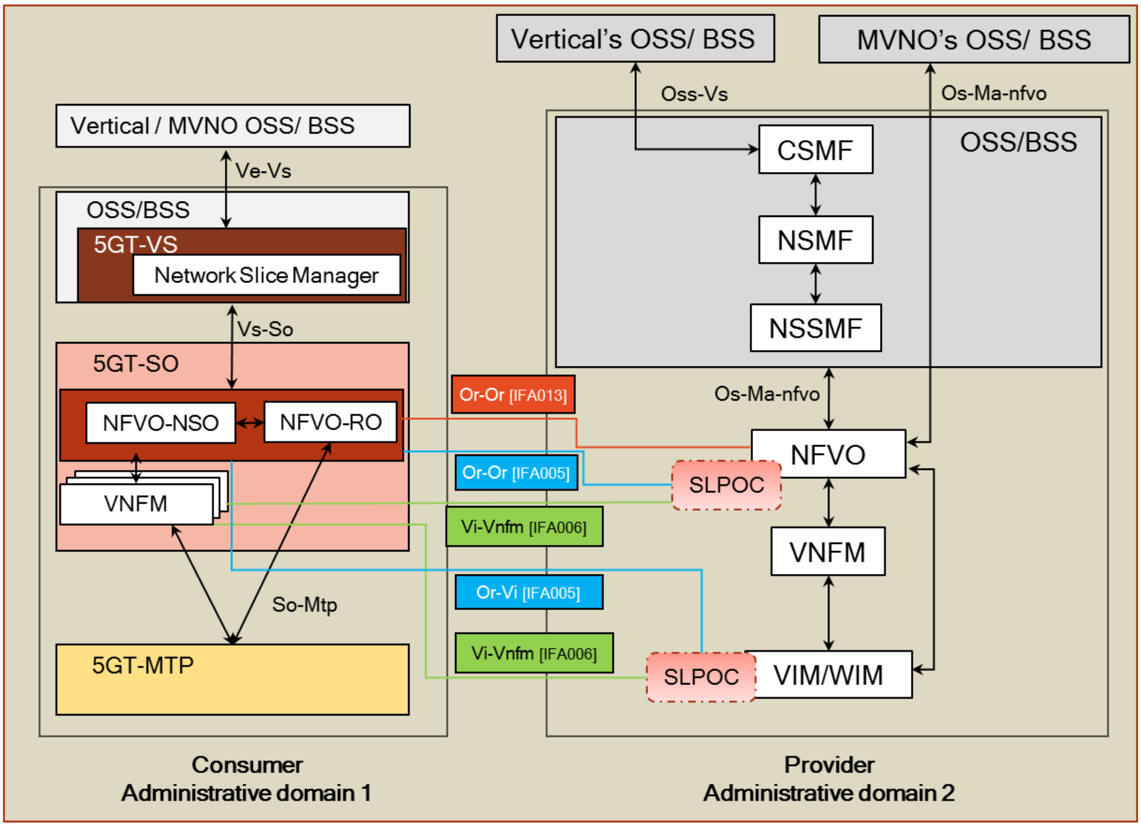


FIGURE 12: FEDERATION WITH NON-5GT ADMINISTRATIVE DOMAIN (5G-TRANSFORMER AD AS CONSUMER)

A.3.1.8.2.2 5G-TRANSFORMER system as a provider

We highlight the possible cases of federation between a 5G-TRANSFORMER system and a non 5G-TRANSFORMER system on Figure 13, in which the 5G-TRANSFORMER AD is considered as a provider for the non 5G-TRANSFORMER AD. We categorize these possibilities of federation into three classes:

- **OSS/BSS - 5GT-SO federation:** This case of federation is done between the network slice management functions (NS(S)MF) in one hand and the 5GT-SO on the other hand. This kind of federation could be requested for instance, in case of roaming for M(V)NOs or to maximize the availability for mission-critical communications. The NS(S)MF interacts with the 5GT-SO, which exposes a set of network services, through Os-Ma-Nfvo IFA 013 interface to instantiate a network service or to update its deployment flavour.

- **NSaaS:** The federation is done between the NFVO on the consumer AD and the NFVO-NSO on the 5GT-SO through the reference point Or-Or using the IFA 013 interface.
- **NFVIaaS:** the resource federation is done between the NFVO of the consumer AD and the NFVO-RO on the 5GT-SO via IFA 005 interface. For the VNFs management in the direct mode, it is done via Vi-Vnfm reference point using IFA 006 interface. Whilst for the indirect mode, this is done through IFA 005 interface (i.e., VNFM on the consumer interacts with the NFVO on the same AD using IFA 007 interface, and the NFVO forwards the requests to the 5GT-MTP through the SLPOC on the 5GT-SO using the IFA 005 interface).

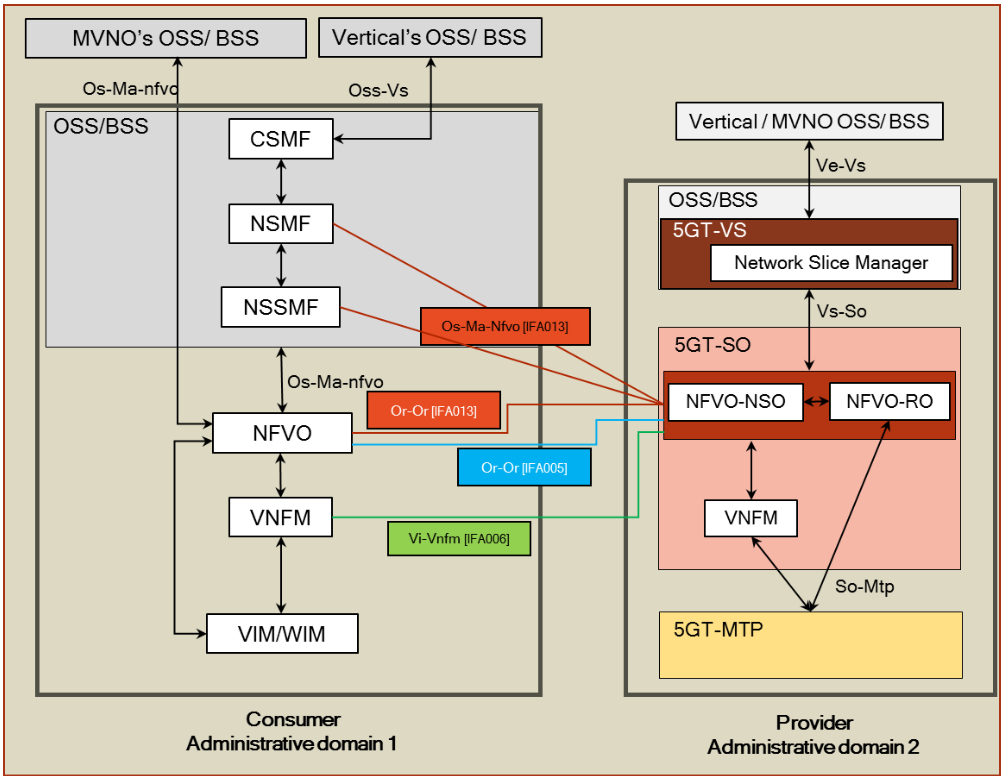


FIGURE 13: FEDERATION WITH NON-5GT ADMINISTRATIVE DOMAIN (5G-TRANSFORMER AD AS PROVIDER)

A.3.1.9 Integration of MEC

To integrate Multi-access Edge Computing (MEC) in 5G-TRANSFORMER, we follow the MECinNFV [37] document, which defines a new reference architecture of MEC. This architecture, shown in Figure 14, facilitates the integration of MEC elements (Mobile Edge Platform - MEP, Mobile Edge Orchestrator - MEO, Mobile Edge Platform Manager - MEPM) in the ETSI NFV environment. Note that the MEP and the MEPM are run as a VNF. The MEO became the Mobile Edge Application Orchestrator (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 management functions have been moved to the VNFM. Moreover, the VNFM is in charge of the lifecycle management of the MEP as well as the MEPM. Another important difference between the original reference architecture in [35] and this NFV-oriented one is the new set of interfaces (Mv1, Mv2 and Mv3) and the use of interfaces defined by the ETSI NFV. However, [37] mentions several issues to be further studied. When integrating MEC in the 5G-TRANSFORMER reference architecture, three points need to be addressed:

- How to integrate the MEC application Descriptor inside a Network Service Descriptor (NSD).
- The role of the MEAO by report to the Service Orchestrator (5TG-SO) of 5G-TRANSFORMER.
- How to implement traffic redirection, knowing that the mp2 interface's focus has been updated.

These are described in more detail subsequently. Additional topics related to the support of network slicing in a MEC system are described in [MEC024], driven by 5G-TRANSFORMER.

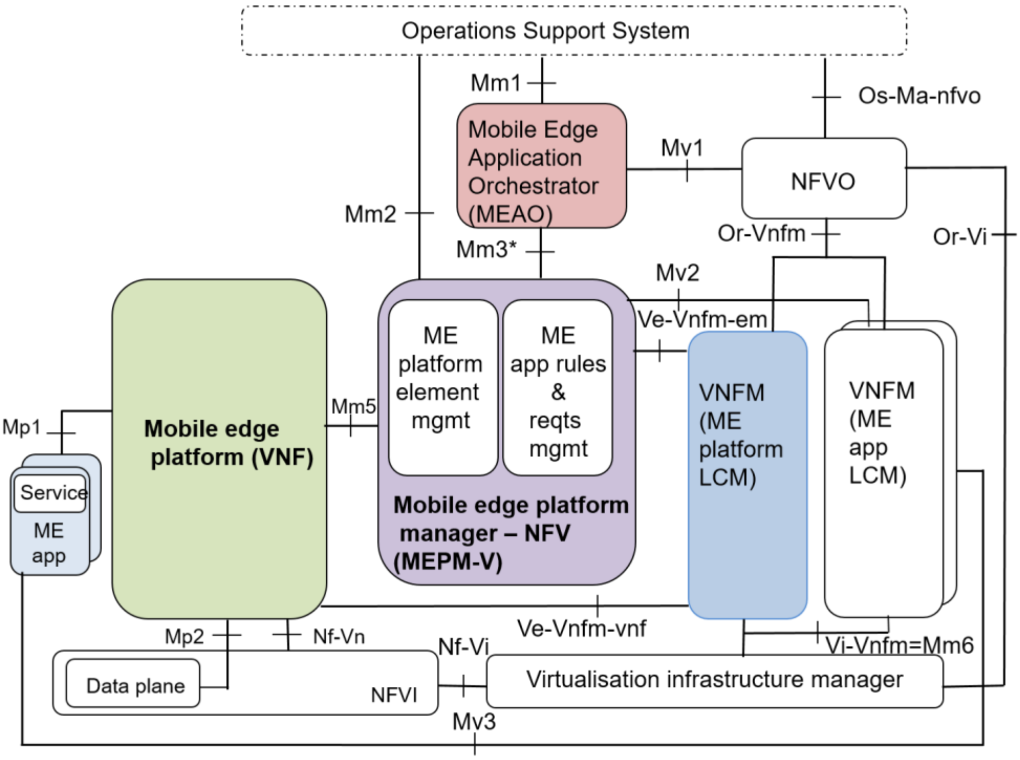


FIGURE 14: MEC IN NFV

A.3.1.9.1 Integration of MEC application in the NSD

We identified two solutions to integrate a MEC application descriptor inside the Network Service Descriptor (NSD). The first solution involves in the extension of the NSD to integrate AppD (see Figure 15). Similar to a Virtual Network Function Descriptor (VNFD), the AppD has been defined in [36] to specifically describe a MEC application. It contains several fields that represent the requirements of the MEC application; particularly *appTrafficRule* and *appDNSRule* that concern the traffic redirection requirements of an application, *appServiceRequired* that indicates the required MEC service to run the MEC application, and *appLatency* that indicates the latency requirement of a MEC application.

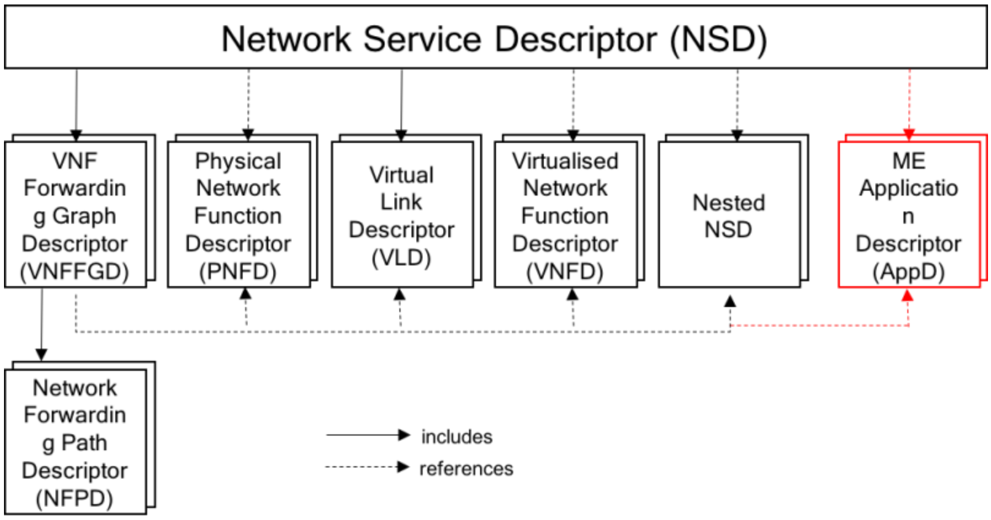


FIGURE 15: INTEGRATION OF APPD INTO A NSD

Unlike a classical VNF, a MEC application is not allowed to scale resources (scale up/down); therefore, the AppD does not include information on scalability. Instead, a migration of a MEC application between two hosts is allowed.

The second solution consists in extending the VNFD, by integrating MEC-specific fields, like *appTrafficRule* and *appLatency*. In this case, VNFs will leave these fields empty, while MEC applications use these fields to indicate specific constraints, such as latency, MEC services to consume, and traffic redirection. This solution avoids modifying the NSD, but requires changing the VNFD.

For 5G-TRANSFORMER we decided to use Solution 1, which assumes the extension of the NSD with AppDs, in order to avoid changing the VNFD and keep compatibility with the current VNFs. Figure 16 illustrates the relationship between the NSD and AppD. It should be noted that the choice of one solution has no impact on the other components described later in this document.

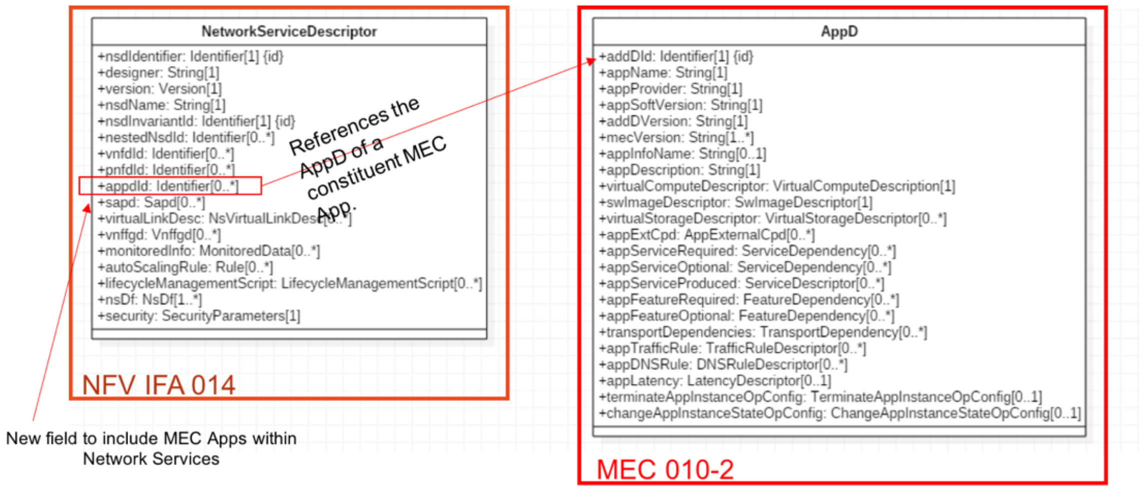


FIGURE 16: INTEGRATION OF APPD INTO A NSD

A.3.1.9.2 The role of MEAO

To recall, the MEA(O) is in charge of the placement of MEC applications, and their instantiation using the NFVO function. In 5G-TRANSFORMER, we propose to integrate the MEAO within the 5GT-SO. To guarantee that MEC applications will be hosted at the edge, the placement algorithm used by the 5GT-SO will take the location constraint as the main constraint to place the AppD. When a NSD is instantiated, a location where to host the MEC application will be included. According to the abstraction information provided by the 5GT-MTP layer, particularly the location of the different NFVIs used by the 5GT-MTP, the placement algorithm of the 5GT-SO will place the MEC application by ensuring that the selected NFVIs supports MEC services and is located in a region covering the requested location. Indeed, we envision that the 5GT-MTP will provide information on the available NFVIs, geographical location, the support of MEC and the available resources.

A.3.1.9.3 Instantiation of MEC applications and MEC platform configuration

The initial solutions mentioned in D1.2 [2] have been updated according to the 5GT-MTP functions and definition. The updated solution described in this deliverable assumes that traffic redirection and other MEC-level functionality is handled at the 5GT-MTP level. Indeed, the 5GT-MTP defines a set of plug-ins, i.e., VIM, WIM and MEC, which are in charge of enforcing the NSD and placing the VNF and MEC applications in the NFVI (central and edge) according to the 5GT-SO decisions. Among these plug-ins, the MEC plug-in is in charge of implementing traffic redirection, DNS rule management, and other MEC platform configuration for the MEC applications. In the considered scenario, we assume that the MEP and the vEPC functions (particularly the SPGW-U) are already deployed at the edge NFVI. We assume a multi-tenancy environment where the MEP and vEPC are shared among the 5G-TRANSFORMER network slices. Furthermore, we have opted for a solution where MEC application descriptors are also onboarded and maintained by the MTP. Therefore, when the 5GT-SO requests the deployment of a MEC application at a specific edge NFVI, it includes in the request a reference to the corresponding AppD. The 5GT-MTP may then extract the relevant information elements from the AppD to apply MEC-specific configurations. These elements are *appTrafficRule, appDNSRule, appServiceRequired, appServiceProduced, and appLatency.* The first indicates the type of traffic that needs to be offloaded to the MEC application, the second includes a set of DNS rules to be added to the MEC DNS service,[9] while the third and fourth indicate the MEC service(s) that the MEC application requires and provides, respectively. The *appLatency* element represents the maximum latency tolerated by the MEC application and can be used by the 5GT-MTP for intra-PoP application placement or other optimizations. The exact semantics of *appLatency* are not yet specified in the relevant ETSI MEC standard [36]. In 5G-TRANSFORMER, we consider this field to refer to the maximum latency from the UE to the MEC application.

Figure 17 presents a high-level view of the process of instantiating a NS that includes MEC applications. The 5GT-VS initiates the process of network service instantiation[10]

[9] This can be used for DNS-based traffic offloading, by resolving DNS requests from UEs to the newly created MEC application instance.
[10] For simplicity, the message to create a NSI identifier that precedes *InstantiateNs* is not shown.

by communicating to the SO the NSI identifier that corresponds to an onboarded NSD which includes AppD references (1). Location constraints for the placement of MEC applications are included in the instantiation request, and in particular in the *SapData* information element.

Upon the reception of this message, the 5GT-SO executes the necessary steps required to derive a resource allocation and placement decision for the VNFs and MEC applications included in the NS. These steps include the execution of a placement algorithm that also satisfies location constraints and considers in its decision the MEC capabilities of the underlying NFVI-PoPs, as these are exposed by the 5GT-MTP. For a MEC application, the 5GT-SO sends to the 5GT-MTP a deployment request (2) which includes, among others, the MEC application's AppD identifier and the desired location.

The 5G-MTP then requests (3) the instantiation of the MEC application to the VIM plug-in, by indicating the location (i.e., the selected edge NFVI, also including the precise location information specified in the NS instantiation request) and the image of the application. It is assumed that the image is already on-boarded at the edge NFVI. Note that, at this level, the 5GT-MTP uses an intra-PoP internal placement algorithm that selects the appropriate NFVI to host the MEC application instance, which supports the maximum tolerated latency and provides the MEC service(s) requested by the MEC application. This information is obtained from the fields found in the AppD stored at the 5GT-MTP, specifically *appLatency* and *appServiceRequired*. The VIM plug-in requests (4) the creation of the application instance at the selected edge NFVI and receives (5) the identifier and the IP address of the instantiated application. Then, it acknowledges (6) the creation of the instance to the 5GT-MTP, indicating this IP address and application instance identifier. This confirmation is eventually returned to the 5GT-SO (7).

With an additional message (8), the 5GT-SO requests the 5GT-MTP to set up any necessary MEC-level configuration, as this is specified in the application's AppD. In this message, the AppD identifier, the application instance identifier, and the IP address information of the MEC application are included. The 5GT-MTP then retrieves the AppD that corresponds to the given identifier, extracts the *appTrafficRule*, *appDNSRule*, *appServiceProduced* and *appServiceRequired* fields, complements them with information about the IP address of the instance, which is not available at the AppD at onboarding time, and requests the configuration of the MEP for the new MEC application to the MEC plug-in (9).

The MEC plug-in, which acts as a MEC platform manager (MEPM), as per the ETSI MEC architecture specification [35], applies the requested configuration to the MEP over the Mm5 reference point. This involves a number of message exchanges between the MEC plug-in and the MEP. First, the MEC plug-in requests the set up of traffic rules (10). In response, the MEP uses the Mp2 interface to apply traffic redirection rules to the SPGW-U deployed at the edge (11). In a similar fashion, the MEC plug-in configures the MEC DNS service appropriately (14) based on the presence of the *appDNSRule* field in the request, and notifies the MEP of potential services provided or required by the MEC application (16). After the successful execution of the above steps, the MEC plug-in confirms the successful MEP configuration to the 5GT-MTP (18). The latter then acknowledges the deployment and configuration of the MEC application to the 5GT-SO (19).

The above procedure is repeated for each AppD referenced in the NSD that corresponds to the NSI that is instantiated. The successful creation of the NSI is eventually signalled to the 5GT-VS (20).

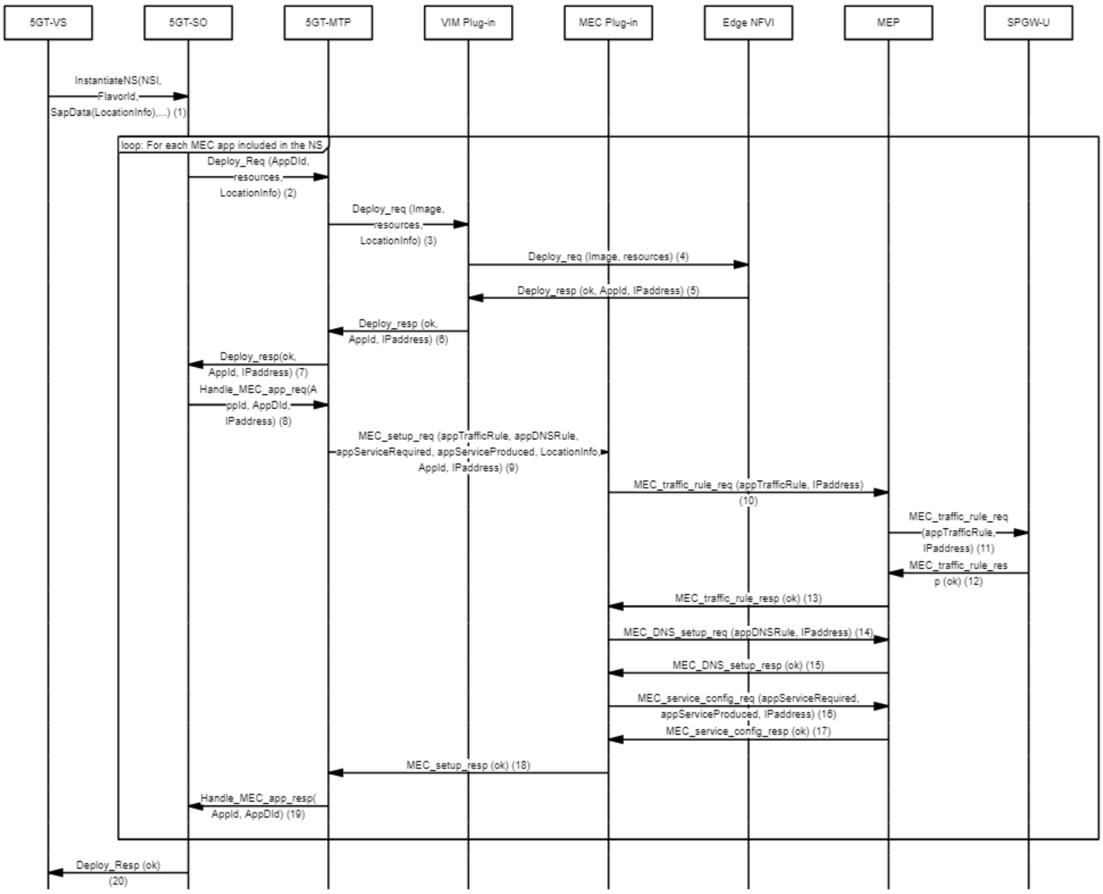


FIGURE 17: WORKFLOW OF DEPLOYING AN INSTANCE OF A NSD INCLUDING AN APPD

A.3.2 Service Orchestration over Multi-site Multi-technology Domains

The main purpose of this section is to discuss implementation-specific design decisions for the 5G-TRANSFORMER architecture and issues related with service deployment in the environments combining multiple technology domains.

There are some solutions proposed in the past to solve the multi-domain issues, such as the Multi-VIM Solutions proposed in Open Source MANO [53] and solutions used in 5G-Crosshaul for network resource provisioning in multi-technology domain transport networks [54]. However, some required functionality was missing (e.g., inter-NFVI-PoP connectivity in [53] or integration of cloud resources in [54]) to solve the needs of 5G-TRANSFORMER.

During the design phase some assumptions were made:

- A single administrative domain is considered, i.e., overall end to end infrastructure is managed by a single administrative entity. However, this administrative domain consists of multiple technology domains. Each of these technology domains is managed by an appropriate software component, generally referred to as VIM or WIM, depending on the nature of this domain. These are the blocks considered in the discussion provided below. However, in a general case, other specific controllers may also be deployed depending on the scenario (e.g., RAN controller).

- Trust relationships between peering technology domains are pre-arranged.
- In case of 5GT integrated stack, implementation of APIs will be based on the appropriate IFA specifications.

A general use case for the current discussion is presented in Figure 18.

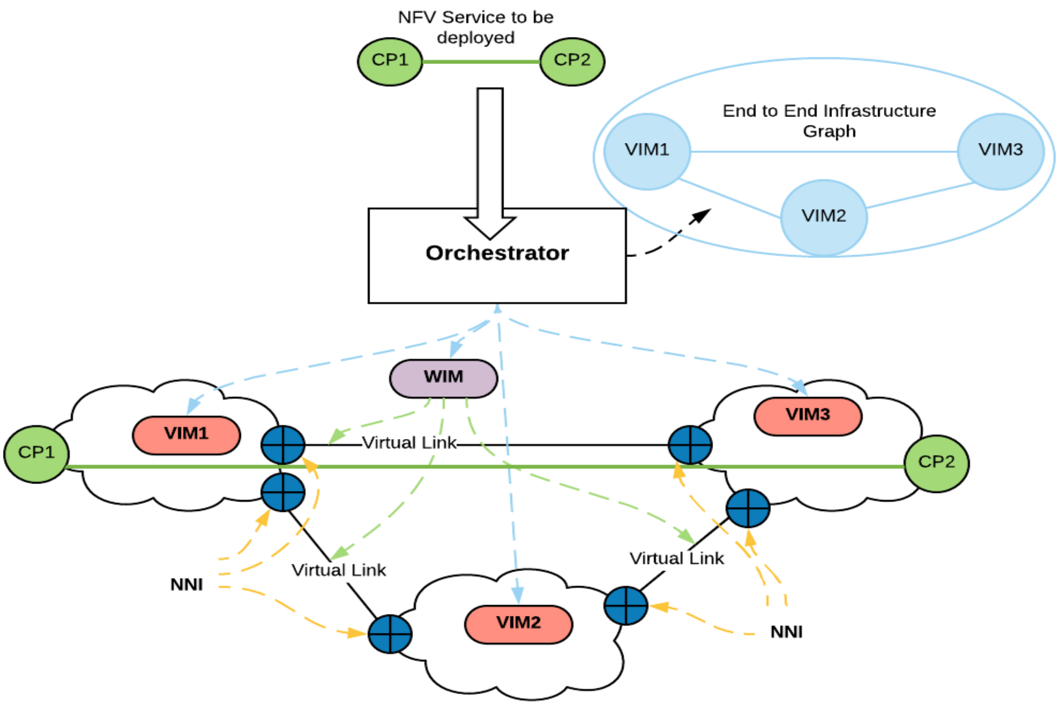


FIGURE 18: MULTI DOMAIN USE CASE PRESENTATION

According to this scenario, in a general case, an NFV network service will be deployed over heterogeneous multi domain infrastructure. In this particular case, a connectivity service is requested between 2 connection points (CP1, CP2), which might represent arbitrary service entities (e.g., origin video server, extended virtual sensing entities). Basically, this is a “Network as a Service” case, where network connection is provisioned on demand. In a more general setting, arbitrary resources, like networking, compute, or storage, might be requested by the customer.

The underlying infrastructure consists of several technology domains - cloud and networking, where each of these domains is managed by an appropriate VIM or WIM entity. There is a generic service orchestrator that is capable of deploying the requested services on top of the managed infrastructure. This orchestrator has an internal representation of the underlying infrastructure as abstracted by the 5GT-MTP.

Upon service instantiation, the 5GT-SO handles the incoming request and starts the deployment process. During the deployment phase it should first decide on the placement of VNFs and, together with the 5GT-MTP, it should allocate consistent resources across all domains according to the incoming service request.

A.3.2.1 End-to-End infrastructure graph

As it was stated above, to process a service instantiation request, the 5GT-SO has an internal representation of the underlying resources, i.e., an end-to-end infrastructure graph (Figure 19) as abstracted by the 5GT-MTP. Once at the 5GT-MTP level, a detailed view of the underlying infrastructure is available by interacting with the underlying VIMs and WIMs. Inside each technology domain, the corresponding VIM or WIM has an internal representation of the managed infrastructure. For example, OpenFlow SDN controllers have capabilities to generate the topology automatically for all managed switches using specialized discovery mechanisms based on appropriate protocols. In the same way, interaction with the VIMs (e.g., OpenStack) allow recovering the underlying computing and storage resources available in datacentres under their control.

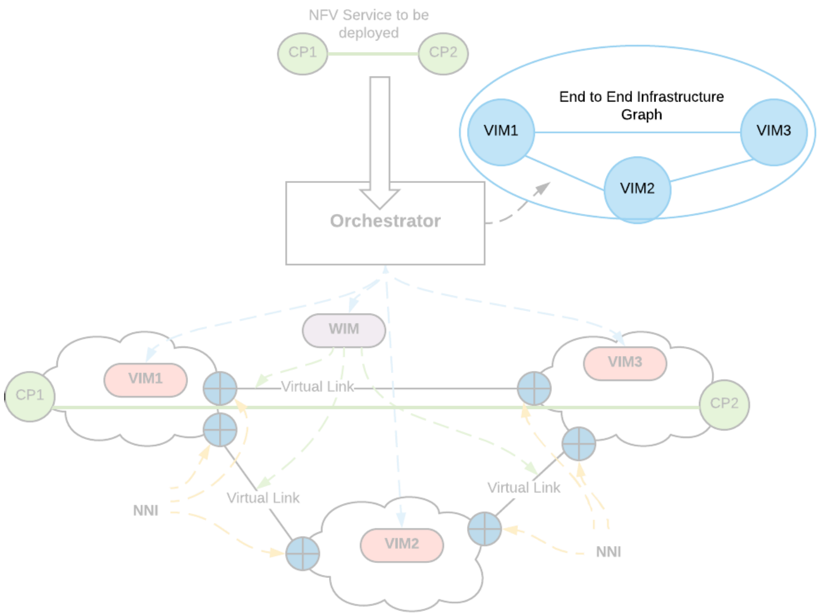


FIGURE 19: END TO END INFRASTRUCTURE GRAPH

In summary, the 5GT-SO has an abstracted view of the topology with two elements, namely NFVI-PoPs and logical links interconnecting them. NFVI-PoPs are characterized by aggregating their computing and storage capacity and logical links by their bandwidth and delay. These links abstract the multiple transport domains that allow interconnecting the NFVI-PoPs. This abstracted information is used by the placement algorithm of the 5GT-SO to take placement decisions of VNFs in NFVI-PoPs and to select the logical links to interconnect the deployed VNFs. This placement decision eventually results in requests to the MTP to deploy the VNFs in NFVI-PoPs, and so, the 5GT-MTP will contact the appropriate VIM for that. In the same way, it will contact the WIMs of the WAN transport network involved in offering the interconnection of the NFVI-PoPs where VNFs were deployed. Since in a general case the WAN transport will consist of multiple technological domains (e.g., wireless transport, multi-layer optical transport), the 5GT-MTP is in charge of contacting each of them and calculating the E2E path, but in this case, with the detailed topology.

A.3.2.2 E2E path calculation and resource allocation

The 5GT-SO is capable to calculate, according to some criteria/constraints, an optimal placement and end-to-end path and to request resource allocations to the 5GT-MTP based on its exposed abstracted view. At the 5GT-MTP level, a representation of the internal structure of each technology domain, combined with inter-domain connectivity options, has been defined. This challenge is presented in Figure 20.

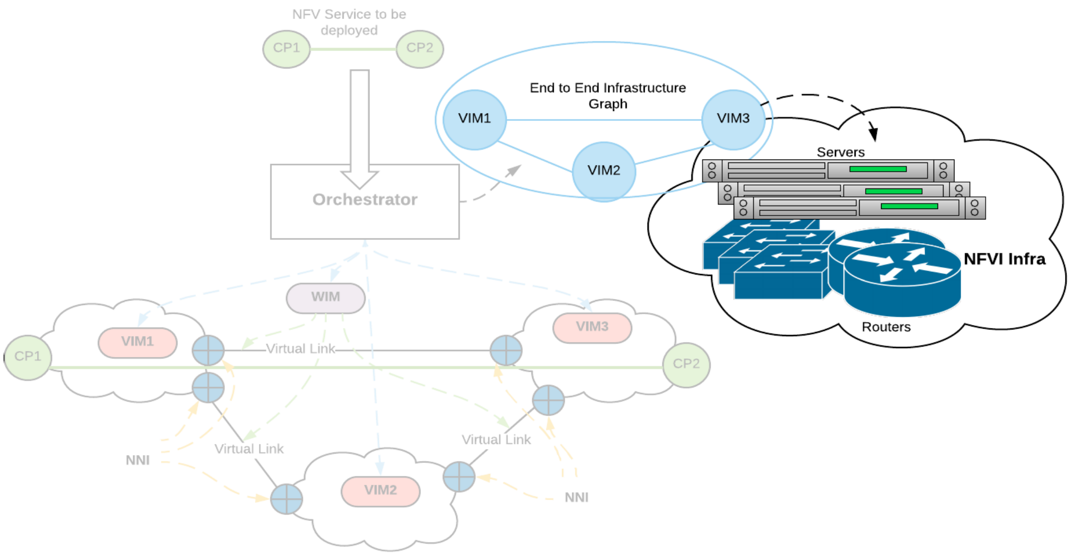


FIGURE 20: PRESENTATION OF THE TECHNOLOGY DOMAIN CONNECTIVITY

Inside of each technology domain, typically there is a complex infrastructure managed by its own control plane. And this internal intra-domain structure is known by the 5GT-MTP by interacting with the VIMs and WIMs.

Therefore, path calculation is done in a two stage process. First, the 5GT-SO selects the logical link for inter-NFVI-PoP connectivity. And second, the 5GT-MTP, when receiving the 5GT-SO request to establish a path through a given logical link, de-abstracts the logical link in the sense that it actually calculates the physical path followed by the traffic and establishes it by contacting all the technology domains involved in the calculated path.

A.3.2.3 Network to Network Interface (NNI) specification

Interconnected technology domains typically have logical or physical points where one domain ends and another starts. For example, a transport network managed by a WIM is connected to the Telco cloud, managed by a particular VIM. We refer to this point between technology domains, as usual, as the Network to Network Interface (NNI). However, it should be noted that in our case this definition is broader than used in the typical Telco environment, as NNI might referred not just as “network to network” interconnection, but “network to cloud” or “cloud to cloud” connections (Figure 21).

While deploying virtual links across multiple technology domains, configuration and consistent setting of these NNIs has to be ensured. Particularly, several interconnection scenarios are possible, such as:

- There is a border device (router, switch, VPN concentrator, etc.), which explicitly splits domains and has interfaces connected to both technology domains. Thus, consistency between the domains is managed at the border device via appropriate configuration and there is a database at the 5GT-MTP level for stitching the different technology domains.
- There is no specific device deployed on purpose for interconnecting two technology domains, as a sort of exchange. This can be the case, for example, when a link coming from the provider edge node in the transport network is attached to the switch acting a gateway of the NFVI-PoP in the Telco cloud. In this case, 5GT applies consistent configuration to the VIM/WIM, which are managing the corresponding entities at each domain.

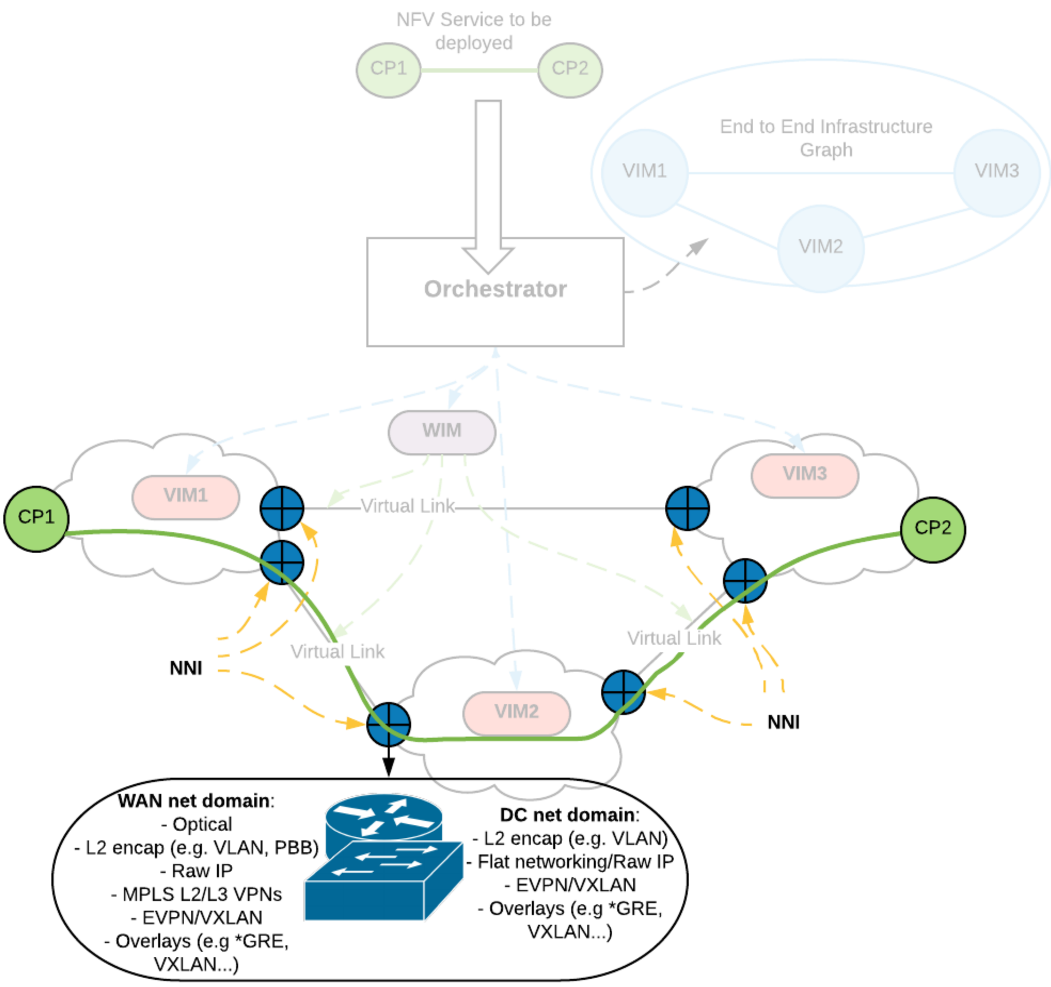


FIGURE 21: NETWORK TO NETWORK INFRASTRUCTURE SPECIFICATION

An additional challenge is presented in Figure 21. In general, each technology domain might have its own transport (networking) technology. For example, in the WAN domain, MPLS L2/L3 VPNs are the dominant transport technologies. To attach an MPLS VPN link to the Telco cloud, this VPN should be either terminated at the border device or there should be some entity capable to do VPN termination inside the NFVI-PoP. On the other hand, inside the Telco cloud, a completely different networking technology, like VXLAN [57], might be used. Thus, the 5GT-MTP is supplied with information to ensure consistent resource allocation and NNI configuration.

A.3.3 SLA Management and Service Scaling

Some of the SLA requirements of a vertical service can be continuously validated and verified during the execution of the service itself, monitoring specific metrics related to the average consumption of its computing or networking resources in relevant time intervals. This is the typical case of services where the resource consumption is a mirror of other service-level performance (e.g. request load, service response time, end to end delay, etc.) and, on the other hand, directly influences the behaviour and the perceived quality of the service. In such cases, scaling up or out the service to higher instantiation levels allows to mitigate, resolve or even prevent breaches in the SLAs. As example, we can consider a virtual Content Delivery Network (vCDN) service, where high consumption of resources at the caches indicates their overloading and leads to several issues in the streaming of the video to the final users. Increasing the resources assigned to the existing caches or instantiating new caches for a more distributed service load allows to solve the temporary resource limitation, guaranteeing the continuous delivery of the service with the required quality.

In 5G-TRANSFORMER architecture, this type of service scaling in support of automated SLA management is handled by the 5GT-SO, as main responsible of all the resource-driven orchestration decisions and procedures. However, it should be noted that the relationship between resource consumption and service performance is specifically dependent on the particular service and cannot be generalized for all the different types of services. As a consequence, the 5GT-SO needs some directives to handle the scaling in the most suitable manner to guarantee the desired performance for the given service. These directives are captured in the Network Service Descriptor (NSD), in the form of autoscaling rules associated to the detection of resource shortage conditions, expressed in terms of thresholds on resource-related monitoring parameters. In particular, an autoscaling rule includes two major items:

- The condition to be matched for triggering the execution of the autoscaling rule, described as a set of scaling criteria related to thresholds on the monitored data or VNF indicators defined in the NSD.
- The scaling action(s) to enforce when the condition is detected. Different kinds of scaling actions can be performed, like the scaling to a different instantiation level of the whole network service, the scale up or down of specific VNFs and the creation or the removal of VNF instances.

At the 5GT-SO level, autoscaling is handled through the joint cooperation of different functional entities, in particular the Monitoring Platform, the SLA Manager and the Service Manager. The detailed internal procedure will be further described in D4.3 [8], while here we report only a summary of the concept.

The collection and processing of the monitoring parameters defined in the NSD is handled by the Monitoring Platform of the 5GT-SO Monitoring Service, through performance monitoring jobs activated on specific virtual resources (e.g. VMs or virtual networks) created during the service instantiation phase. The Monitoring Platform is also responsible to detect the “triggering condition”, generating alerts whenever threshold-based scaling criteria are matched. These alerts are notified to the SLA Manager, which in turn takes decisions about the scaling actions to be performed, based on the directives expressed in the NSD autoscaling rules. The actual enforcement of the scaling actions is performed by the Service Manager, when triggered by the SLA Manager. In this final step, the Service Manager coordinates the internal procedures of the 5GT-SO to scale the service resources and, at the end of the process, notifies the 5GT-VS with the information about the new instantiation level applied to the service instance.

A.3.4 Policy Management

The SLA requirements of vertical services, as defined in D1.1 [1] cannot be treated in a uniform way by the 5G-TRANSFORMER system. Some of the SLA requirements such as latency, user data rate, or mobility can be translated into SLA requirements or deployment flavours within NFV NDSs. Other SLA requirements such as preferred or prohibited infrastructure providers, energy reduction, or service creation time are better handled by policies for the 5GT-SO. Here, ‘policy’ is understood as an additional rule or optimization target for the 5GT-SO. E.g. the 5GT-SO should orchestrate a vertical service such that energy consumption is minimized.

The first class of SLA requirements described above simply has to be satisfied by the 5GT-SO. Still, the 5GT-SO may have several choices how to satisfy the SLA requirements. Here, the policies can be understood as directives to the 5GT-SO to make decisions in favour of a specific target, where the targets correspond to the second class of SLA requirements. As an example, the amount of gNBs used to provide coverage for a vertical service could be minimized. Additionally, features supporting discontinuous transmission (DTX) or putting cells of unused carriers to sleep mode could be activated to reduce energy consumption of the RAN.

ETSI NFV has defined already an interface for policy management at the Os-Ma-nvfo reference point (IFA013 [29]). This interface as defined by ETSI NFV provided operations for transferring, deleting, (de)activating, and querying policies. The project extended this interface via change requests[11] with operations to associate policies with specific network service instances and to disassociate them again.

Based on the SLA requirements of a VSI, the 5GT-VS can associate policies with the NFV NSI used for this VSI. The 5GT-SO acts based on the associated policies. Either, it uses the policies in its own decision making or it can activate a corresponding policy in an MTP for decision making closer to the infrastructure. Continuing the example above, the 5GT-SO might activate a policy for energy consumption on the infrastructure manager of the RAN.

Further details of policy management and the corresponding interface are described in D3.3 [6].

[11] NFVIFA(18)0001001, IFA013ed321 CR add policy associate disassociate operations, NFVIFA(18)0001002, IFA007ed321 CR add policy associate disassociate operations, NFVIFA(18)0001003, IFA008ed321 CR add policy associate disassociate operations, NFVIFA(18)0001012, IFA010ed321 CR add policy associate disassociate operations.

A.4 Vertical Slicer Design

This chapter summarizes the design of the Vertical Slicer (5GT-VS), as well as its main functional blocks and the algorithms therein. The summary is provided in a high level in this deliverable, in order to give a complete view and offer a better understanding of the whole system design. The details on the design of individual components and their internal workflows and interfaces can be found in D3.3 [6]. In this chapter, we describe its architecture in Section A.4.1 and its main components in subsequent sections.

A.4.1 Vertical Slicer Overview

The 5GT-VS is the common entry point for all verticals into the 5G-TRANSFORMER system, being part of the OSS/BSS of the administrative domain of a 5G-TRANSFORMER service provider (TSP). The 5GT-VS coordinates and arbitrates the requests for vertical services. The vertical services are offered through a high-level interface focusing on specifying the logic and the needs of the vertical services.

Specifically, the 5GT-VS allows defining vertical services by selecting from a set of vertical-oriented service blueprints or by composing building blocks. The selected blueprints or composed building blocks, along with instantiation parameters, will result in Vertical Service Descriptors (VSDs). Then, the 5GT-VS maps the VSD and requirements defined therein onto a network slice, which we describe with extended ETSI NFV Network Service Descriptors (NSDs) [30]. NSDs define forwarding graphs composed of a set of VNFs or Virtual Applications (VAs) connected with Virtual Links (VLs), where some of the VAs have specific characteristics and constraints of MEC applications. Importantly, the 5GT-VS allows mapping several vertical service instances (VSIs) onto one network slice, handling vertical-dependent sharing criteria or strategies and taking care of the necessary capacity changes in existing slices. In conclusion, the most fundamental tasks of the 5GT-VS are to provide the functionality for creating the vertical service descriptions, and to manage the lifecycle and the monitoring of VSIs and of the corresponding network slice instances (NSIs) to which they were mapped.

In addition to such tasks, the 5GT-VS provides arbitration among several vertical service instances in case of resource shortage in the underlying infrastructure and based on global budgets for resource utilization, as defined in the SLAs between the verticals and the TSP. The arbitration component in the 5GT-VS maps priorities of vertical services and SLA requirements to ranges of cardinalities of resources. These cardinalities are used by the 5GT-SO while deploying the NFV network services (NFV-NS) and, in case of actual resource shortage, to assign resources to the most important vertical service instances.

The architecture of the 5GT-VS is shown in Figure 22. As mentioned, the 5GT-VS is part of the provider's OSS/BSS, and interacts with the vertical (including the M(V)NO) through its northbound interface (NBI), and with the service orchestrator through its southbound interface (SBI). The same REST API is used by VAs triggering actions of the 5GT-VS for instantiating additional vertical services, such as the emergency services in the eHealth use case. The 5GT-VS interacts also with the OSS/BSS Management Platform of the TSP. This, in turn, interacts with the TSP, but this interaction between management platform and TSP is out of the scope of the 5GT-VS. Here, the TSP can manage tenants (tenant management), manage the resource budgets of tenants (SLA management), and onboard Vertical Service Blueprints (VSB) (VS Blueprints catalog).

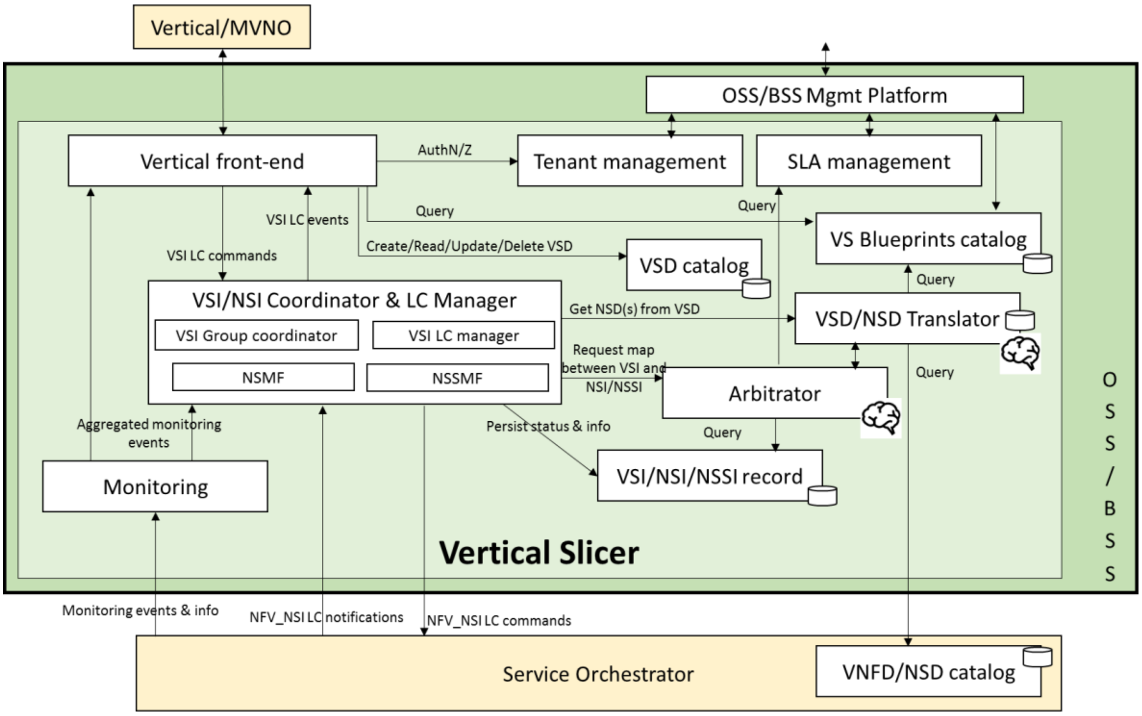


FIGURE 22: THE VERTICAL SLICER ARCHITECTURE

The Vertical Front-end is the entry point to receive requests from the verticals/MVNO on vertical service provisioning, management and monitoring. Importantly, the Vertical Front-end presents the vertical with Vertical Service Blueprints (VSB), stored and handled through the related catalogue (VS Blueprints catalog), and supports the vertical in providing the parameters therein to obtain the corresponding Vertical Service Descriptor (VSD). VSBs and VSDs are described in D3.3 [6]. Additionally, the Vertical Front-end allows verticals to compose vertical services from a set of building blocks.

The lifecycle (LC) of Vertical Service Instances (VSIs) and the corresponding Network Slice Instances (NSIs) are handled through the component called “VSI/NSI Coordinator & LC Manager”. This component, described in Section A.4.5, is the central engine of the 5GT-VS. It manages the association between VSIs and NSIs, regulating the sharing of network slices among different vertical services and their decomposition into network slice subnet instances (NSSIs). Moreover, it handles the finite state machines for VSIs’ and NSIs’ lifecycle, coordinating commands and events associated with them. The network slice management is handled through requests for instantiation, modification and termination of the corresponding NFV-NSs, interacting with the 5GT-SO. The status and the current characteristics of VSIs and NSIs/NSSIs are stored persistently in the VSI/NSI/NSSI records.

While the “VSI/NSI Coordinator & LC Manager” manages the lifecycle of slices and vertical services, the complete 5GT-VS decision logic is implemented in the VSD/NSD Translator and in the Arbitrator. The VSD/NSD Translator, described in Section A.4.2, selects the descriptors of the NFV network services that are able to support the requested vertical services. The VSD/NSD Translator also identifies the network service deployment flavor (DF) and instantiation level (IL) (e.g., number of VNF instances, amount of vCPU or RAM for VNFs, bandwidth of VLs) to guarantee the performance and the characteristics defined in the VSD. The VSD/NSD Translator determines as well, which policies - see Section A.4.2 - to associate with a vertical service instance. The Arbitrator, described in Section A.4.3, decides about the sharing of network slices among different vertical services and the scaling of vertical services based on service priority and resource budget in verticals' SLAs.

The 5GT-VS Monitoring Service, described in Section A.4.4, interacts with the 5GT-SO to collect monitoring data about the established NFV network services and aggregates these data to produce metrics and KPIs for the corresponding network slices and vertical services. These metrics can be used for SLA verification, or to make decisions about the lifecycle of a network slice, e.g. to trigger a scale-up action in case of decreasing performance.

A.4.2 The VSD/NSD Translator Module

The VSD/NSD Translator module is an entity within the 5GT-VS that maps the vertical's requirements into technical specifications needed by the 5GT-SO to perform a NFV-NS deployment. For vertical services derived from VSBs, the vertical first selects the VSB, which contains parts that have to be filled in with the requirements it demands. Once the vertical provides actual values for the missing parts in the VSB, the latter becomes a Vertical Service Descriptor (VSD): a high-level specification of a network service, based on a service logic perspective rather than a resource-based perspective.

For vertical services composed from building blocks, the building blocks may have corresponding parameters to be provided by the vertical. Once these parameters are provided, there is again a VSD for this composed service, which is handled in the same way as VSDs derived from VSBs.

A VSD is then mapped onto a Network Service Descriptor (NSD) through the following steps: (1) the Translator queries the Virtual Network Function Descriptor (VNFD) catalogue within the 5GT-SO for the VNFs referenced in the VSD; (2) the Translator queries the NSD catalogue within the 5GT-SO for the NSDs referenced in the VSD; (3) the Translator fills the fields present in the deployment flavour of each VNFD and of each Virtual Link Descriptor (VLD), using the QoS restrictions of the VSD (such fields will be finalized by the Arbitrator later on), (4) the Translator determines which policies have to be associated with this vertical service.

A.4.3 The Arbitrator Module

SLA management is a key aspect in service provisioning to a vertical. Any degradation of the SLA can impact not only the technical behaviour of the vertical service but also the reputation or business leadership of the vertical itself. The vertical, as 5G-TRANSFORMER service consumer, will specify Service Level Objectives (SLOs) adapted to its service needs. An example of SLOs can be a maximum end-to-end latency of 20ms for a vertical service in the automotive domain. Vice versa, the TSP's OSS/BSS can state different service level classes, representing distinct guarantees on resource and/or service availability. The matching between the SLOs desired by the vertical and the service level classes offered by a provider defines an SLA. The contents and terms of an SLA are used as directives for the configuration and orchestration of services and resources in network slices, where the orchestration is done by the 5GT-SO.

The Arbitrator within the 5GT-VS implements these mechanisms and provides the 5GT-SO with support for service deployment. It operates based on the SLA, as well as the information on each service provided by the vertical, namely: (1) the service priority level, (2) the VNFFG representing the service, (3) bandwidth requirements of the VLs and the relative CPU requirements of the VNFs in the VNFFG, as well as their memory/storage requirements, and (4) the vertical's quality of service (QoS) requirements (e.g., end-to-end latency, service availability or reliability level). Note that such information is described in the NSD created by the VSD/NSD Translator for the received VSD.

The Arbitrator's tasks are twofold:

1. Decide how to map new VSIs in NSIs/NSSIs, allowing multiple vertical services to share one or more NSIs or NSSIs;
2. Determine the deployment flavors associated with each service, thereby meeting the vertical's QoS requirements and accounting for the priority level.

When instantiating a new vertical service and performing the second of these tasks, the Arbitrator may detect that there are insufficient resources in a resource budget for all VSI of this vertical. This situation is reported to the vertical, which may then decide to a) cancel the instantiation request, b) negotiate additional resources with the TSP, i.e. increase the resource budget, and retry the instantiation, or c) confirm the instantiation request. In the last case, the new VSI would be instantiated at the expense of lower-priority VSIs.

A.4.3.1 Sharing of network slices among vertical services

During the instantiation of a vertical service, the Arbitrator is responsible for making decisions about the sharing of NSIs and NSSIs among different vertical service instances. A typical example could be the sharing of the same vEPC among different vertical service instances with similar requirements in terms of mobile access, or the sharing of a service component for the collection of vehicle messages among multiple automotive services that make use of the same information. At present, we focus on arbitration among vertical services of one vertical.

The decision about slice sharing is made by the Arbitrator analysing the pair <NSD; DF> initially computed by the VSD/NSD Translator, which describes the structure, the characteristics and the cardinalities of a potential NSI able to support the requested VSI. Starting from the NSD, the Arbitrator verifies if one or more existing NSIs can be re-used to accommodate the new VSI or at least part of it, e.g. one or more nested NFV-NSs that may be included in the NSD. This decision is affected by the constraints specified by the vertical in the VSD, in particular by the constraints about isolation.

If no suitable NSI is available, the Arbitrator decides to create a new NSI, based on the NSD originally computed by the VSD/NSD Translator. It will proceed with the update of the deployment flavor and instantiation level initially set by the VSD/NSD Translator (see Section A.4.3.2). The "VSI/NSI Coordinator & LC Manager" manages the instantiation of the new network slice and its slice subnets, triggering the 5GT-SO to deploy the required NFV-NSIs.

On the other hand, if the Arbitrator finds existing NSIs that can be used, it determines if and how they should be modified to accommodate the additional service. For this, it interacts with the VSD/NSD Translator, providing as input the VSDs of all the VSIs that will share resources from the given set of network slices. The VSD/NSD Translator combines the requirements of the various VSDs and returns a suitable set of pairs <NSD; DF> defining the characteristics and the size of the target NSIs. Assuming that the elaborated solution is compliant with the vertical's SLAs, the Arbitrator identifies the NSIs to be modified and new ones to be created, according to the result provided by the VSD/NSD Translator. As in the previous case, the "VSI/NSI Coordinator & LC Manager" manages the procedures to execute the required actions.

A.4.3.2 Computation of deployment flavors

Regarding deployment flavours, we consider firstly the case where a new NSI is deployed for a requested vertical service. The Arbitrator assigns resources according to the priority level of the VSIs of a vertical, starting from the highest-priority VSI. The Arbitrator assigns memory and storage resources to this VSI, whereas for CPU and bandwidth allocation a more complicated procedure is needed, as this allocation can be adjusted and affects placement decisions as well. Here, service latency is the main performance metric, i.e. the maximum latency that a VSI is allowed to experience. Two components contribute to the service latency: (i) the processing time, due to the execution of the VNFs in the VNFFG, and (ii) the network travel time, which is due to the time needed to transfer data from one VNF to the next in the VNFFG, when adjacent VNFs are deployed on different servers. Ideally, all VNFs are placed in the same NFVI-PoP to reduce network travel time and the complete latency budget can be used for processing time, reducing the amount of required processing resources. On the other hand, having all VNFs in one NFVI-PoP might be overly restrictive. Therefore, the Arbitrator determines a second, more flexible, deployment flavour. As a smaller portion of the latency budget would be available for processing, the required amount of processing resources is higher.

In some cases, an alternative or fallback service chain can be defined, providing less-optimal results, but requiring fewer resources while providing. In this case, the Arbitrator can define a third deployment flavour corresponding to this fallback service.

The Arbitrator will update the NSD created by the VSD/NSD Translator with the information for the deployment flavours explained before. Eventually, when instantiating the VSI, the 5GT-VS will pass a list of initial deployment flavours to the 5GT-SO, which will try to instantiate them in order until one deployment flavour could be instantiated successfully. This requires an extension of the life cycle management interface of IFA013 to pass such a list of initial deployment flavours instead of just a single one.

Let us now consider the case where the vertical service to be processed can be provisioned on an existing NSI or using an existing NSSI. In this case, the Arbitrator adds the traffic load due to the newly requested vertical service to the load of the existing VNFs/VLs, and re-computes the necessary CPU and bandwidth, as described above. Again, the Arbitrator will use the CPU and bandwidth values obtained for the best case (and, if needed, those obtained for the worst case), to update the deployment flavor of the involved VLDs and VNFDs.

In case of resource shortage, some lower-priority services may not be accommodated, or may be terminated due to the need to re-allocate resources to higher priority services.

To summarize, the Arbitrator module allows the 5GT-SO to make deployment decisions meeting the vertical's indications even if (i) the 5GT-SO is unaware of higher-layer information like the SLA between the vertical and OSS/BSS, and (ii) the total amount of available resources is not sufficient to adequately deploy all requested services.

A.4.4 The Monitoring Service

The 5GT-VS Monitoring Service is responsible for producing monitoring data about network slices and vertical services and for elaborating more elementary monitoring data about VNFs and NFV network services, as retrieved from the underlying 5GT-SO Monitoring Service. It provides monitoring data about the VSIs as input for internal decisions related to resource arbitration at the Arbitrator. The 5GT-VS provides also monitoring data about the deployed VSIs to the verticals, to feed the internal processing of vertical applications, where needed. The 5GT-VS monitoring service should be able to expose the monitoring information it receives from the 5GT-SO Monitoring Service to the VS Front-end. Actually, this exposure can be part of the SLA.

The 5GT-VS Monitoring Service receives and shares monitoring data from the 5GT-SO via the Vs-So-Mon reference point, or data generated internally. The VSI LC Manager and/or NS(S)MF send a request to the 5GT-VS Monitoring Service indicating the monitoring metrics to subscribe to. The 5GT-VS Monitoring Service relays the request to the 5GT-SO Monitoring Service. During service runtime, the 5GT-SO indicates the availability of monitoring data to the 5GT-VS Monitoring Service. The 5GT-VS Monitoring Service retrieves the data and informs the VSI LC Manager and NS(S)MF. It can also send this information directly to the 5GT-VS Front-end to provide the vertical with it, depending on the SLA and what was defined in the VSD.

The 5GT-VS Monitoring Service can supervise alarms related to an NSI and/or VSI only if it has subscribed to them. The NS(S)MF sends a corresponding request for subscription to NSI alarm notifications towards the 5GT-VS Monitoring Service to receive the alarm notifications related to the NSI. As a result of this operation, whenever an alarm is raised, a notification will be sent to the 5GT-VS Monitoring Service, which will transmit it to the NS(S)MF. The NS(S)MF can request the 5GT-VS Monitoring Service to create, modify, and delete a threshold on a specified performance metric (for NFV-NSI(s)) for which notifications will be generated when crossed. The VSI LC Manager can do the same for vertical service(s).

A.4.5 VSI/NSI Coordinator & LC Manager

This component coordinates the activities of the other 5GT-VS components. To do so, it uses several subcomponents and interacts with the other components and databases of the 5GT-VS. The subcomponents are the VSI Group Coordinator to handle the VSIs of one vertical, the VSI LC Manager to handle the lifecycle of single VSIs, and the Network Slice (Subnet) Management Functions (NSMF, NSSMF) to handle the lifecycle of NSIs and NSSIs.

A.4.5.1 VSI Group Coordinator

One or several resource budgets may be defined for the VSIs of a vertical. These resource budgets define the maximum amount of e.g. storage, vCPU, memory, and bandwidth. The resource consumption of a specific VSI is checked against one such resource budget. The VSI Group Coordinator component maintains the relation between resource budgets and VSIs for each individual vertical. The actual checking and arbitration among different VSIs is done by the Arbitrator (see Section A.4.3).

A.4.5.2 VSI LC Manager

The VSI LC Manager proposes vertical services to customers via the catalogue of VSBs, which may include also information on the services' cost. Internally, the VSI LC Manager keeps a record of the requested VSIs and maintains their references to the associated supporting NSIs. It also performs accounting of vertical service resource consumption per NSI.

The 5GT-VS SBI is used to link the NSMF in the 5GT-VS with the 5GT-SO, e.g. for network slice management: discovery, allocation and LCM of NSIs, policy management, performance monitoring, fault management and accounting. Due to the monitoring and the fault management of NSIs, the VSI LC Manager is aware of the operational state of each NSI and able to verify whether it matches the expected SLAs of the corresponding VSIs. If the outcome indicates that a vertical service SLA has not been fulfilled, remediation is required.

In addition, the VSI LC Manager supports LCM operations, pricing and charging, performance monitoring and fault reporting of VSIs. If a customer changes the requirements of a VSI, the VSI LC Manager triggers the Arbitrator to update the network slice requirements via recalculation of the NSD deployment flavor. Eventually, the VSI LC Manager asks the NSMF for a modification of the NSI capacities.

As mentioned, the VSB catalogue includes information on the SLA costs. To enable the charging of the verticals for the VSIs, the VSI LC Manager traces their resource consumption according to information (e.g., monitoring data, events) reported by the 5GT-SO to the 5GT-VS Monitoring Service. Thus, the VSI LC Manager subscribes to alarm notifications, using the 5GT-VS Monitoring Service. Some service related alarm data may be provided, if required, to the verticals. Also, the VSI LC Manager subscribes to performance measurement notifications. Again, some service related performance measurement data may be provided to the verticals.

A.4.5.3 NSMF and NSSMF

The NSMF manages NSIs; correspondingly the NSSMF manages NSSIs, see also Section A.3.1.7. As mentioned, the VSI LC Manager relies on the NSMF and NSSMF to assess the feasibility of providing NSIs and NSSIs, respectively. As the functionalities of NSMF and NSSMF are rather similar, we describe in the following the NSMF functionality only. The NSMF and NSSMF are based on [11].

Specifically, the NSMF checks which PNFs and/or VNFs are referenced in the NSDs (the NSDs are exposed by the NSD catalogue inside the 5GT-SO to the 5GT-VS), then the NSMF provides the VSI LC Manager with the NSIs, through which a VSI could be deployed. The NSMF also keeps record of all the network slice requirements (e.g. number of CPU, storage) per NSI. This information can be used by the Arbitrator to recalculate the deployment flavours of the NFV-NSs when new VSIs are instantiated or existing ones are updated or terminated. However, the NSMF is not aware of the VSIs but only of the requirements received from the Arbitrator. Due to the monitoring and the fault management of NFV-NS instances, the NSMF is aware of the operational state of the latter and able to verify the expected functionalities and whether its SLAs are met or not. In case of SLA violations, the NSMF can trigger corresponding alarms.

A.5 Service Orchestrator Design

This chapter summarizes the design of the Service Orchestrator (5GT-SO), as well as its main functional blocks and orchestration algorithm framework. A detailed description of 5GT-SO architecture as well as the review of the state of the art solutions for the 5G-SO platform is provided in the 5G-TRANSFORMER deliverable D4.3 [8]. The aim is to provide an overview of the 5GT-SO, presenting a general insight of the proposed 5GT-SO approach and architecture solutions, which complement the whole 5G-TRANSFORMER system architecture design.

A.5.1 Key functionalities of 5GT-SO

The 5GT-SO is in charge of end-to-end (E2E) orchestration of the NFV-NS across one or multiple administrative domains by interacting with the local 5GT-MTP and/or with other 5GT-SOs, respectively, while addressing and managing the allocation of different vertical slices. 5GT-SO receives the service requirements from 5GT-VS via the Vs-So interface (see Figure 3 in Section A.3.1) in the shape of a Network Service Descriptors (NSD). A NSD describes a Network Service (NFV-NS) that 5GT-SO is able to provide (either on its own or by leveraging neighbouring SOs); it is expressed in terms of chaining of VNF components (i.e., constituent VNFs) described by VNF Descriptors (VNFD). A VNFD describes a VNF in terms of its deployment and operational behaviour as well as connectivity (i.e., virtual links) requirements.

The 5GT-SO processes the NSD in order to assign virtual networking, computing and storage resources across its local 5GT-MTP domain and, if applicable, across federated domains, while addressing service requirements specified in the NSD. To this purpose, the 5GT-SO embeds a series of modules, most notably, the Network Service Orchestrator (NFV-NSO) and the Resource Orchestrator (NFV-RO) with functionalities equivalent to those of ETSI NFV Orchestrator [20]. The NFV-NSO has the responsibility of coordinating the deployment of NFV-NSs along with their lifecycle management. The NFV-RO is in charge of orchestrating virtual resources under the control of the underlying 5GT-MTP. Additionally, the 5GT-SO provides multi-domain service orchestration through service federation. In this direction, the 5GT-SO interacts with 5GT-SOs of other administrative domains through the So-So interface (federation) to take decisions on the end-to-end (de)composition of network services and their most suitable execution environment, i.e., service federation.

Service orchestration involves the management and instantiation of VNFs and/or VAs at local, edge and cloud NFVIs. The problem of mapping VNFs to (virtual) computing entities (nodes, NFVI-PoPs) and the mapping of virtual links between VNFs into (virtual) paths, depending on the granularity of abstraction offered by the 5GT-MTP, can be tackled by different optimization strategies, namely heuristics or mixed-integer linear programming. Moreover, automatic network service management and selfconfiguration algorithms (e.g., failure recovery) are also required to adapt deployments to network changes and/or the infrastructure resource utilization (e.g., virtual machine CPU load) in order to prevent service degradations or SLA violations due to concurrent usage of resources (e.g., through NFV-NS scaling). To this purpose, the collection and aggregation of monitoring data is foreseen to trigger self-adaption actions.

Summarizing, the 5GT-SO key functionalities are the following:

- decides the optimal service (de)composition for the whole NFV-NS based on local resource capabilities exposed by local 5GT-MTP;
- performs the lifecycle management of the whole NFV-NS as well as of each VNF composing the NFV-NS through the VNFM, including service catalogue management;
- decides the optimal placement of VNFs/VAs[12] along with the optimal deployment of virtual links connecting VNFs through mapping operations, thereby enabling the execution of the NFV-NS into the local 5GT-MTP;
- requests the needed services to federated 5GT-SO(s) to address the execution of a portion(s) of the NFV-NS in other administrative domains;
- performs monitoring tasks and SLA management functions to enable the triggering of self-adaptation actions (e.g., healing and scaling operations) thereby preventing service performance degradations or SLA violations.

A.5.2 5GT-SO architecture

Figure 23 presents a high-level overview of 5GT-SO subsystems and their interactions designed to achieve the essential 5GT-SO functionalities described before.

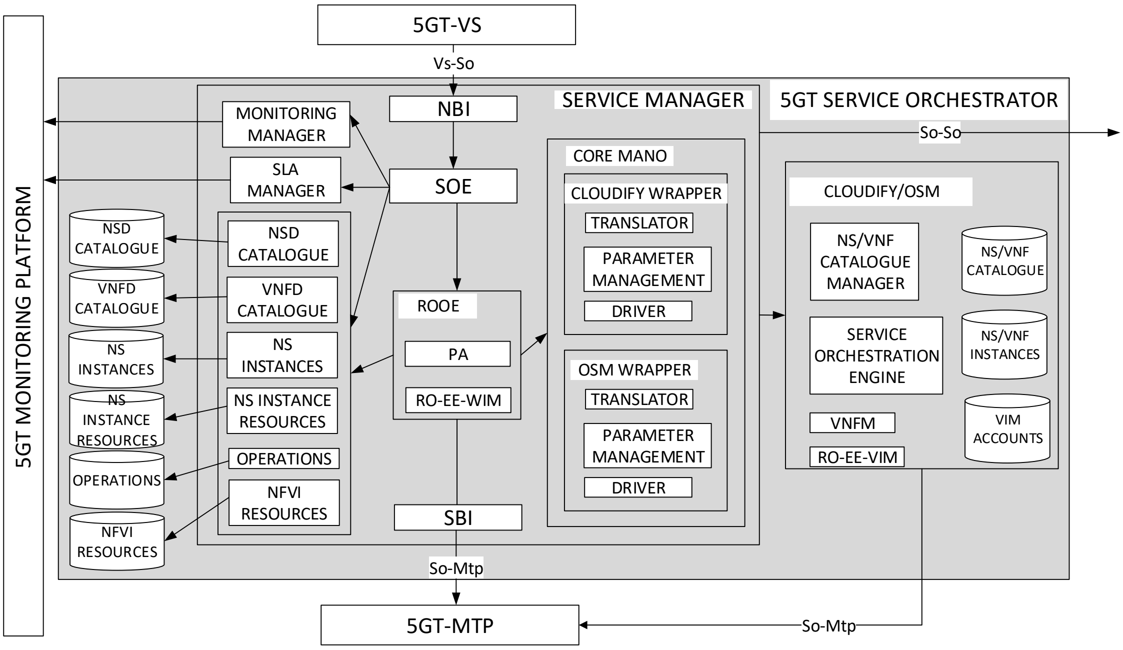


FIGURE 23: THE 5GT-SO ARCHITECTURE

[12] The granularity that 5GT-SO has when placing functions (NFVI-PoPs, servers, etc.) depends on the level of abstraction offered by the 5GT-MTP.

The aim of 5GT is to design a flexible service orchestrator architecture capable of integrating, and so, exploiting, the functionality already available from well-known and production-level MANO platforms (e.g., OSM and Cloudify) as far as the high-level functionality described above is concerned. In this way, the focus is put on the innovative functionality brought by the project (e.g., federation).

At a high-level, its architecture is organized in four main building blocks. First, the service manager (SM) is a novel building block introduced by 5GT, which is the brain of the 5GT-SO, as it handles service requests, and has an end-to-end view of what is offered by the 5GT-MTP. Thanks to this global view, it is also in charge of stitching NFVI Point-of-Presence (NFVI-PoP) and WAN network connections in an automated way when multiple NFVI-PoPs are involved in providing a network service, which is a functionality hitherto not offered by open source MANO platforms. It also embeds the service orchestration and resource orchestration logic and dispatches relevant tasks to the other building blocks according to the operational workflow it handles. Second, the Core MANO platform (i.e., OSM or Cloudify) will handle the computing resources by interacting with the 5GT-MTP. Since the SM uses the API of each MANO platform and they are different, a wrapper is needed for this translation and adaptation. It will also handle the specificities of each orchestration platform in a way that does not alter the internal workflow. Third, even if the placement algorithm is part of the resource orchestrator of the 5GT-SO, it is highlighted here as one main building block, since it is accessible through a well-defined REST API. In this way, external algorithmic modules can be easily tested and their performance compared. Finally, the databases are in charge of maintaining the state of the system in terms of service offering, instantiated network services, 5GT-MTP resources used, etc.

This 5GT-SO design has several advantages. It brings flexibility in terms of development. In fact, large software platforms are powerful, but they are complex and heavy given the huge amount of functionality brought. By introducing the SM, 5GT exploits the power of these platforms whilst allowing adding new functionality easily, hence focusing on the novel concepts to be evaluated. Furthermore, we believe this approach makes the 5GT code more survivable and evolvable, since it can be adapted to new MANO platform releases by just modify the wrapper to adapt to the new API. In this way, the 5GT functionality can also be exploited with newer releases. Moreover, other core MANO platforms may be integrated in the future by developing the corresponding wrappers.

In the figure, we can find the main building blocks of the Service Manager component, which acts as the brain of the service orchestrator module of the 5GT architecture. These building blocks are[13]:

- **Database (DB) module**: this module contains several submodules to interact with each of the external databases containing the system status information. Currently, the available external databases are:
  - **NSD (Network Service Descriptor) catalogue:** stores the information of NS descriptors.
  - **VNFD (VNF Descriptor) catalogue:** stores the information of VNF descriptors.
  - **NS Instance Database:** contains information of NS instances.

[13] Northbound/Southbound Interface sub-modules act as interfaces with the corresponding block of the 5G-TRANSFORMER architecture

  - **NS Instance Resources Database:** stores information of resources used by an NS Instance.
  - **Resources Database:** stores information of the status of the different resources (networking, computing and storage) of the underlying transport Infrastructure.
  - **Operations database:** stores information about NS operations, i.e., instantiate, terminate an NS.

- **Service Orchestration Engine (SOE):** this module receives the requests from the NBI. For the NS query-related operations it interacts with the appropriate databases to fulfil the request and retrieve NS instance information status. If the requests involve 5GT-MTP resources management (e.g. for NS instantiation or termination requests), the SOE delegates the operation to the ROOE module. It also handles composite network services and service federation through the So-So interface.
- **Resource Orchestration Engine (ROOE):** it handles the requests that are related to 5GT-MTP resources management, i.e., NS instantiation and termination. It interacts with the Placement Algorithm (PA) submodule, the coreMANO module and the Resource Orchestration Execution Engine WIM (RO-EE-WIM) submodule to accomplish the request, where WIM stands for wide area network infrastructure manager. The ROOE submodules perform the following functionalities:
  - **Placement Algorithm (PA):** determines the placement of the virtual network functions and their required allocation of resources during the instantiation of NSs based on its characteristics.
  - **Resource Orchestration Execution Engine WIM (RO-EE-WIM):** it is submodule in charge to communicate with the SBI to handle network resources management operations at the 5GT-MTP.
- **CORE MANO:** it allows the interaction between the SM and the available open source orchestration platforms. Current supported orchestration platforms are Cloudify and OpenSource MANO. The different wrappers manage the requests coming from the ROOE to handle compute and storage resources operations at the 5GT-MTP. Among others, they include the VNF Manager (VNFM). It is in charge of the lifecycle management of the VNFs deployed by the 5GT-SO.
- **Monitoring manager:** it is in charge of requesting the performance monitoring jobs to the 5GT Monitoring platform based on the needs of the NFV-NS. These needs are contained in the NSD of the requested NFV-NS.
- **SLA Manager**: it is in charge of configuring the alerts in the 5GT Monitoring platform. This may entail configuring thresholds of monitored metrics that, when exceeded, trigger an alert from the monitoring platform towards the SLA manager. Once there, the SLA manager handles those alerts so that SLA compliance is guaranteed. In the event a requested SLA is not met, the SLA Manager may trigger scaling actions to prevent or recover from SLA violations. The behaviour of the 5GT-SO in case of alert is encoded as rules in the NSD of the NFV-NS (e.g., autoscaling rules).

Additionally, and as mentioned above, the 5GT-SO is capable of integrating multiple MANO platforms through the wrappers in the CoreMANO module. These platforms embed the usual MANO components (NFVO, VNFM, catalogue, etc.). In the 5GT architecture, the integrated MANO platforms are in charge of handling the deployment of VNFs in the computing infrastructure of the system. Inter-NFVI-PoP interconnection is handled by the SM.

A.5.3 Service Orchestration and Federation

Decisions for service orchestration and/or federation are triggered when 5GT-VS requests the instantiation of a network service or modifications of an existing network service. In both cases, the 5GT-SO has two main decisions to make, namely:

1. How to decompose the network service graph (i.e., NSDs) into several network service segments and, consequently, where to implement each segment, whether in the local 5GT-MTP domain or by leveraging neighbour 5GT-SOs (using service federation).
2. How to map a service segment into a set of virtual resources, i.e., where to run the component VNFs in the virtual infrastructure based on specified resource demand.

The first decision includes service decomposition and the possible use of service federation, where the second step consists of resource orchestration and possible use of resource federation.

A.5.3.1 Service Orchestration

A.5.3.1.1 Information provided to 5GT-SO

The 5GT-SO interacts with the 5GT-VS and the 5G-MTP, where both provide information that is used to make orchestration decisions. The 5GT-VS is using NSD to define information about the set of VNFs and their interconnection for all the NFV-NS instances that will be requested for that NSD. At the service instantiation time, the 5GT-VS provides information about the needed resources for each VNF in the specific NFV-NS instance in the form of an NSD deployment flavour.

The input data from 5GT-MTP consists of a set of available resources that are stored in the 5GT-SO's NFVI resource repository. In general, the 5GT-VS asks the 5GT-SO to instantiate a certain network service, and it is the 5GT-SO's task to decide how to provide it, by using the virtual resources exposed by the 5GT-MTP and/or through service federation.

A.5.3.1.2 Service orchestration decision steps

Service orchestration decisions consist of a preliminary step and three major steps.

The preliminary step of the 5GT-SO is when all input data is translated into an *instance-level system model*, synthetically representing all the elements to account for, categorized in two categories: 1) VNF-related: VNFs to place, capabilities they require, amount of data they exchange; 2) infrastructure-related: NFVI-PoPs able to run VNFs and their capabilities, the set of logical links [14] connecting NFVI-PoPs and their capacity/delay.

Both the VNF-related and the infrastructure-related elements of the system model can be represented as graphs, namely *the VNFFG* and a virtual *infrastructure* graph.

[14] In this context, logical links are the links exposed by the 5GT-MTP to the 5GT-SO. A logical link is a path at the infrastructure level connecting two NFVI-PoPs. As such, they are distinct from IFA013 virtual links.

An example of the virtual infrastructure graph is shown on the Figure 24. The graphs consists of the NFVI-PoPs h_x with capabilities $\kappa(h)$ and the logical links connecting the NFVI-PoPs with the capacity $C(h_1,h_2)$ and delay $\delta(h_1,h_2)$.

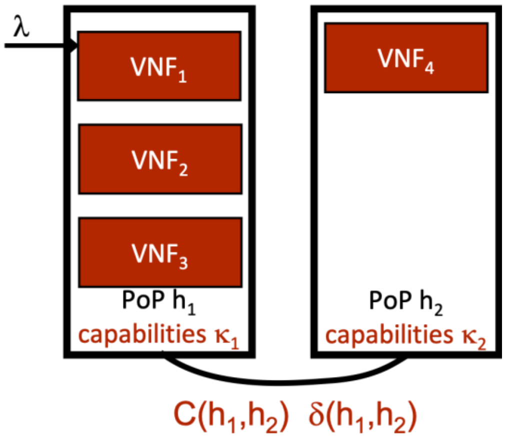


FIGURE 24: EXAMPLE OF VIRTUAL INFRASTRUCTURE GRAPH

The three major decision steps that result in three decision variables are:

- VNF[15] placement, i.e., location where in the virtual infrastructure exposed by the 5GT-MTP each VNF shall be hosted. VNF placement decisions are represented through binary variables A(h,v), each expressing whether a given VNF v is assigned to a certain NFVI-PoP h. Using a binary variable here reflects the fact that VNFs cannot be split across multiple NFVI-PoP. If we extend the model to consider a VNF composed of multiple VNFCs, then, it could be that a VNF can be split across multiple NFVI-PoPs;
- Resource assignment, i.e., how much of virtual resources (computational power, memory, storage, etc) each VNF shall be assigned. For resources such as memory, storage, bandwidth, etc., VNFs shall get the exact amount they need, while for other resources (e.g., CPU) can be allocated in a more flexible fashion using minimum and maximum thresholds;
- Traffic routing, i.e., which logical links at the infrastructure level should be used to implement the virtual link instances for transferring data between different VNFs, and their bandwidth assignment. The traffic and the delays on the logical links must not exceed the capacity of the logical links and the service time requirements expressed in the virtual link descriptors.

The three decision variables described above impact one another. The Placement Algorithms as part of the RO-OE is in charge of taking placement decisions and selecting the logical links to use. After that, the RO-OE will handle the actual deployment of this decision by interacting with the MANO platform, which in turn interacts with the 5GT-MTP, for the VNFs, and the 5GT-MTP through the SM SBI for the WAN link creation. Details of the Placement Algorithms supported by the 5GT-SO can be found in D4.3 [8].

[15] For simplicity, we refer to VNFs only in the description of our algorithms. Notice however that they can be easily extended to VNFCs.

A.5.3.1.3 Constraints

Along with the decisions that have to be made, there are four major constraints that the 5GT-SO has to honour. The first one, NFVI-PoP capabilities, is the total amount of resources assigned to VNFs must not exceed the capabilities of the NFVI-PoP that is running the VNFs. The second is the minimum VNF resources or that for each VNF must be assigned at least the minimum amount of resources required. The third constraint, the link capacity, meaning that the total amount of traffic produced between VNFs placed at NFVI-PoP h1 and VNFs placed at h2 must not exceed the logical link capacity between h1 and h2. The fourth constraint is the service time which mainly depends on the processing time (in the VNFs) and the network delays (sum of the delays on the traveling logical links). Additional constraint is added for each service-specific KPI.

A.5.3.1.4 Objective

The objective is to make placement (of VNF), assignment (of resources) and network routing that satisfies the described constraints while maintaining the service KPIs with optimized network usage, power consumption, or cost.

A.5.3.2 Federation

In 5G-TRANFORMER there are two types of federation: federation of services and federation of resources. Both federation procedures occur in the 5GT-SO. To enable federation, the administrative domains (implementing the 5G-TRANSFORMER system) need to define mutual business/service agreements among each other (federation level), initial catalogue of services and the network connections between administrative domains (So-So interface, see Section A.3.1.5.4).

A.5.3.2.1 Federation levels

The federation relationships among administrative domains are ranked on different federation levels, depending of the business/service agreements. The federation levels indicate the mutual degree of trust among administrative domains and they can be from lower to higher, defined as bronze, silver, gold, platinum, etc. For higher federation levels (e.g., platinum), significantly broader options and specific information parameters about resources and/or services are exchanged between the administrative domains. For the lower levels (e.g., bronze), the offering of resources and/or services is limited and information for parameters is more abstract and descriptive.

A.5.3.2.2 Service federation

The decision for service federation occurs upon service decomposition. The 5GT-VS issues a request for instantiation of new NFV-NS or modification of existing NFV-NS (using NSD). The requested composite NFV-NS is decomposed into one or more segments. These segments are nested NFV-NSs that can be instantiated on another administrative domain. The decomposition of the composite NFV-NS can be performed statically or dynamically (e.g., by using an algorithm on instantiation procedure). The dynamic decomposition procedure is for further study.

The 5GT-SO NFVO-NSO is in charge of performing the service federation procedure. Pre-requirements for service federation is the service catalogue and the network connections between different administrative domains. We assume that network connections are established upon business/service agreements are made. The catalogue of services is used to collect all the federation capabilities of the other domains. The catalogue of services can be formed dynamically or in a static manner. In the dynamic manner the 5GT-SOs exchange their capabilities to provide service federation among themselves, while the 5G-TRANSFORMER system is running. Each 5GT-SO broadcasts (periodically or event-triggered) the capabilities to the different groups of peers, grouped according to the federation levels. In the static manner, the catalogue of services is defined during the business/service agreements among the administrative domains and contains more generic and pre-defined services.

When the 5GT-SO NFVO-NSO decides to consume federated NFV-NS from external domain, first it checks the catalogue of services for available NFV-NS. Based on certain criteria (e.g., federation level, footprint of the service, closest peer), which is still not defined, the 5GT-SO NFVO-NSO would send a request for instantiation of the NFV-NS (segment or nested service) to the provider 5GT-SO NFVO-NSO. If the NFV-NS is available, the provider 5GT-SO NFVO-NSO instantiates and provides the NFV-NS to the consumer 5GT-SO NFVO-NSO. All the details (e.g., VNFs, VLs, monitoring parameters, etc.) are hidden away from the consumer 5GT-SO NFVO-NSO, depending on the federation level between both administrative domains. The consumer 5GT-SO NFVO-NSO would use the So-So-LCM of the E/WBI to send requests for lifecycle operations of the provided NFV-NS.

A.5.3.2.3 Resource federation

The federation of resources occurs in the 5GT-SO NFVO-RO. The federation of resource is split in two phases: 1) advertisement phase and 2) allocation and management phase. As in the federation of services, the pre-requirements would be business/service agreements (federation levels) and setup of network connections.

The advertisement phase of the federation of resources has the objective for providing each 5GT-SO NFVO-RO with an updated view of the available resources in the other administrative domains. This phase contains calculation of resource abstractions and broadcasting of the resource abstractions towards peering 5GT-SOs. The abstraction of resources is applied twice. First, the 5G-MTP applies resource abstraction on the underlying NFVI infrastructure to hide some details of NFVI resources from the local 5GT-SO. The second abstraction is applied by the 5GT-SO NFVO-RO to hide away more parameters of the underlying resources (e.g., number of CPU cores, type of storage, etc.). The outcomes of the calculations are categorized in federation levels. The categorized information is then broadcast by the SO-SO Advertisement block to the peering 5GT-SOs according to the established federation level. On the receiving end, the categorized information is stored in the NFVI database, as an updated view of peering 5GT-SOs resources.

The second phase, allocation and management of resources, is triggered in the 5GT-SO NFVO-RO upon decision to use federated resources. Once the decision is made, the request for resources is sent from the consumer 5GT-SO NFVO-RO to the chosen provider 5GT-SO NFVO-RO. The algorithm for choosing the available resource abstractions of an external provider 5GT-SO NFVO-RO is a work in progress, satisfying certain criteria (e.g., resource capacity, delay, federation level, location, etc.). Once the provider 5GT-SO NFVO-RO processes the request and allocates the resources on its constituent 5GT-MTP, it provides positive response to the consumer 5GT-SO NFVO-RO, with details of the allocated federated resources. The consumer 5GT-SO NFVO-

RO includes the allocated federated resources as its own, managing and monitoring their performance through the monitoring primitives of the So-So interface. The limitations on the management operations and monitoring parameters directly depend on the federation level between the two 5GT-SOs. The provider 5GT-SO NFVO-RO is a forwarding point between the consumer 5GT-SO NFVO-RO and the provider 5GT-MTP.

A.6 Mobile Transport and Computing Platform Design

The Mobile Transport and Computing Platform (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. The following sections summarizes the 5GT-MTP key functionalities and architecture. This summary presents a general insight of the proposed 5GT-MTP architecture solutions to complement the 5G-TRANSFORMER system architecture design. The details on the design of the 5G-MTP and its internal workflows and interfaces, as well as a survey of the state of the art solutions for the 5G-MTP platform are reported in D2.3 [4].

A.6.1 Key functionalities of 5GT-MTP

The 5GT-MTP is an SDN/NFV-based component capable of simultaneously supporting a diverse range of networking and computing requirements specific to the vertical industries. It provides and manages the virtual and physical compute, storage and network resources on which service components are eventually deployed. The primary functionalities of the 5GT-MTP are two-fold.

The first functionality is the coordinated allocation and provisioning of radio, transport, storage and computational resources required by the vertical services. The 5GT-MTP can integrate several Virtual Infrastructure Managers (VIM) and WAN Infrastructure Managers (WIM) from different technological domains and expose a unified view to the upper layer (the 5GT-SO in the 5G-TRANSFORMER project). Each VIM or WIM, in turn, can interface with the underlying infrastructure to request virtual resources. By resorting to the NFVIaaS paradigm, we can identify the 5GT-SO as a service consumer which wants to run VNF instances inside an NFVI provided as a service by the NFVIaaS provider, namely, the 5GT-MTP. This means that the 5GT-SO has the control of the VNF instances, but it does not control the underlying infrastructure.

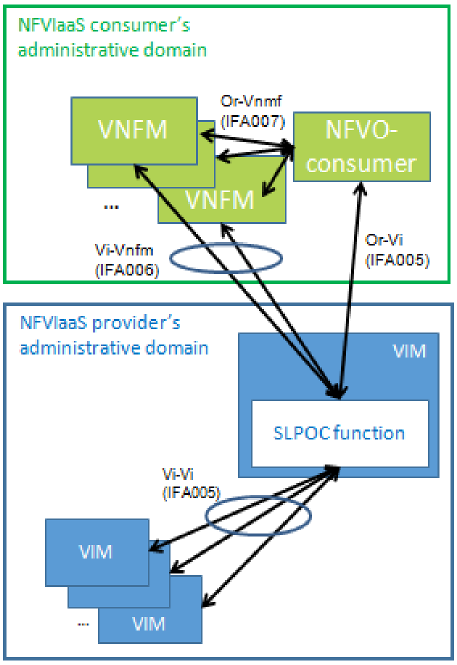


FIGURE 25: SLPOC FUNCTION

In particular, since the 5GT-MTP is structured in several VIMs/WIMs, it acts as a single entry point (see Figure 25), i.e., single logical point of contact (SLPOC) in ETSI GS NFV-IFA 028 [32] terminology, for any resource allocation request coming from the 5GT-SO. In this case the NFVIaaS provider's VIMs/WIMs are hidden from the NFVIaaS consumer and unified interfaces are exposed by the SLPOC and offered to the NFVIaaS consumer.

The second functionality is providing an abstracted view of the resources to the 5GT-SO, thus hiding the complexity of the specific underlying technologies. 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 VNFs,
- type and characteristic of available connectivity provided in the form of logical links.

Depending on the use case, the 5GT-MTP may offer different levels of resource abstraction to the 5GT-SO. However, the 5GT-MTP has full visibility of the resources under the control of the VIMs or WIMs managing each technology domain, since they belong to the same administrative domain. 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.

A.6.2 5GT-MTP architecture

The design of the 5GT-MTP architecture leveraged the works carried out in the 5GPP Phase 1 projects, 5G-Crosshaul [44] in particular, and standard development organizations such as ETSI NFV. The architecture (see Figure 26) of the 5GT-MTP aims at providing a set of functionalities and operations to support the 5GT-SO to achieve efficient utilization of resource allocations for the VNFs under its control.

The main building block of the 5GT-MTP is the Single Logical Point of Contact for resource orchestration (5GT-MTP NFVO-RO SLPOC) that acts as a single point of contact towards the 5GT-SO providing the suitable abstract view of the resources managed by the 5GT-MTP and receiving the resource requests (see Figure 25). Moreover, the NFVO-RO SLPOC acts as resource orchestrator to select and configure the transport, radio and compute/storage resources compliant with the requests from the 5GT-SO. 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 [58]. In particular, when receiving a resource allocation request from the 5GT-SO, the 5GT-MTP NFVO-RO SLPOC generates the corresponding request to the relevant entities (e.g., VIM or WIM), each of them providing part of the needed virtual resources.

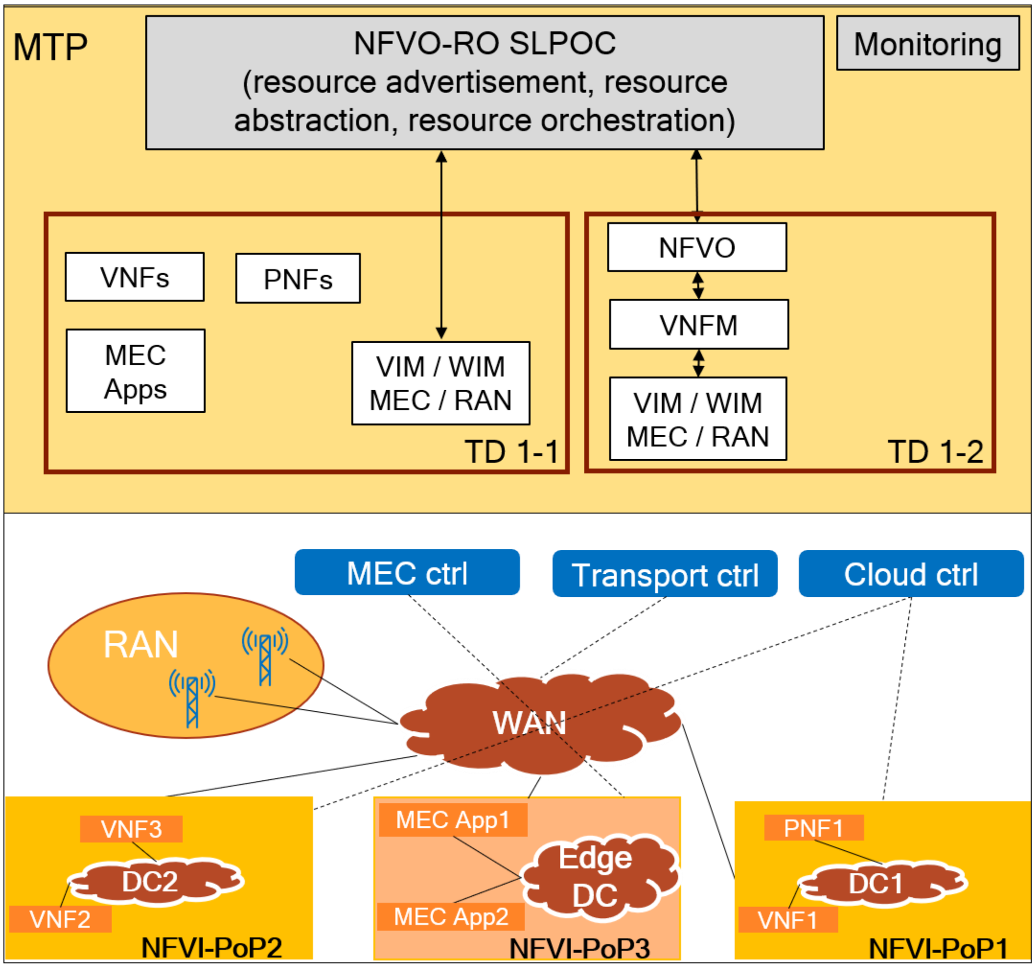


FIGURE 26: 5GT-MTP ARCHITECTURE

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 is mostly based on a set of standard documents being produced within the ETSI NFV framework, namely ETSI GS NFV-IFA 005 [22] and ETSI GS NFV-IFA 006 [23]. The former allows the NFVO-RO of the 5GT-SO to request resource allocations to the 5GT-MTP, whilst the latter allows the VNFM of the 5GT-SO to request resource allocations for the VNFs under its control. In terms of managing VNF instances, the So-Mtp interface also consists of ETSI GS NFV-IFA 008-based interfaces [25] to allow the VNFM of the 5GT-SO to directly configure the VNF instances running in the 5GT-MTP.

At the internal southbound interface between the 5GT-MTP NFVO-RO SLPOC and the technological domains (TDs), the SLPOC might request resources to each TD. For example, it might interact with VIM through ETSI GS NFV-IFA 006 or with PNFs and WIM. Moreover, as a special case, a resource request may be translated at the SLPOC into an ETSI GS NFV-IFA 013-based NFV-NS request [29] towards a mobile network technology domain. 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 26) 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 logical link. A logical link is a path connecting two physical interfaces. It is different from the virtual links defined in IFA 013 as an abstracted representation of the connection between the VNFs (characterized by a given bandwidth and latency), independent of the physical interfaces.

Other main building blocks of the 5GT-MTP are the VIMs, WIMs, Network Function Virtualization Infrastructure (NFVI) and Monitoring component:

- **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 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.
- **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.
- **MEC Controller** (MEC in the figure): it configures and manages MEC-specific services (e.g., traffic redirection, DNS update, Radio Network Information) at the MEC platforms/hosts.
- **RAN Controller** (RAN in the figure): it controls the radio infrastructure as well as their connectivity to the mobile core.
- **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.

- **Monitoring component:** the monitoring block is responsible for collecting data from the different domains (transport, radio and cloud), monitoring the physical infrastructure and virtual resources, and providing the needed monitoring information to the 5GT-SO.

A.6.3 MTP Abstraction

As highlighted in Section A.6.1, an important and exclusive function of the 5GT-MTP is to decide the abstraction (i.e., the level of details) related to the resources exposed to the 5GT-SO. The abstraction includes the definition of an information model and the related data model. According to the general architecture defined in 5G-TRANSFORMER, abstraction of logical links (i.e., a physical path connecting two physical node interfaces), computation and storage capabilities is required to enable 5GT-SO for orchestration.

Three different alternatives of 5GT-MTP abstraction with different levels of details are considered.

Alternative 1

5GT-MTP exposes all physical resources (mobile, transport, storage, and compute) to the 5GT-SO. An example is provided in Figure 27 where 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 as well as resources ownership issues as data centers may belong to different providers.

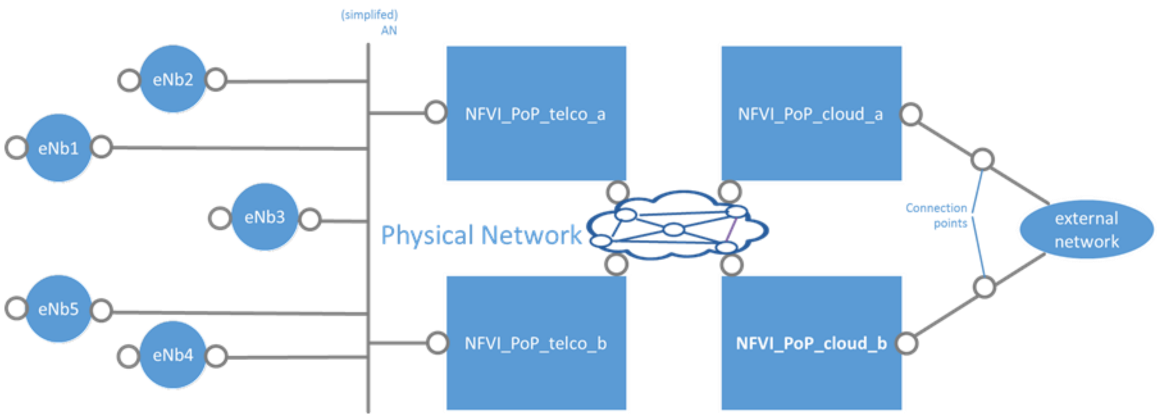


FIGURE 27: 5GT-SO VIEW FOR ABSTRACTION ALTERNATIVE 1

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 28, 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_cloud_a and NVFI_PoP_cloud_b).

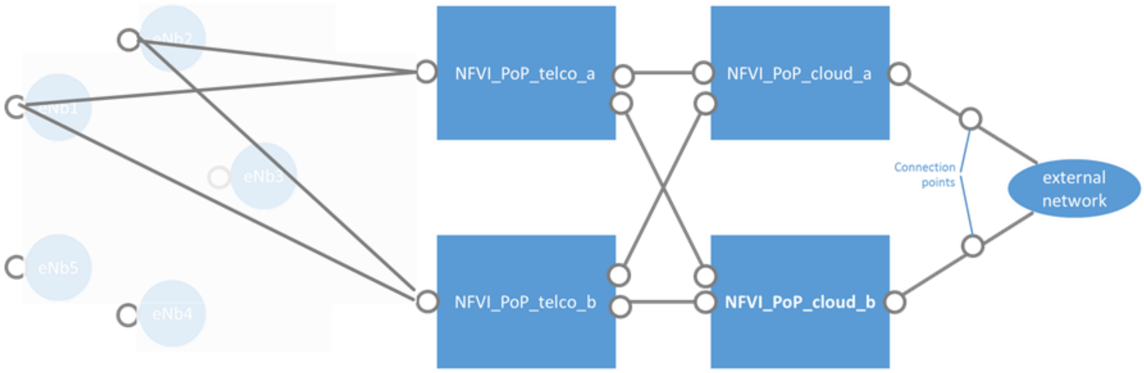


FIGURE 28: 5GT-SO VIEW FOR ABSTRACTION ALTERNATIVE 2

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 29, 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 points identifying a specific eNB) with cloud data centers (NVFI_PoP_cloud_a and NVFI_PoP_cloud_b).

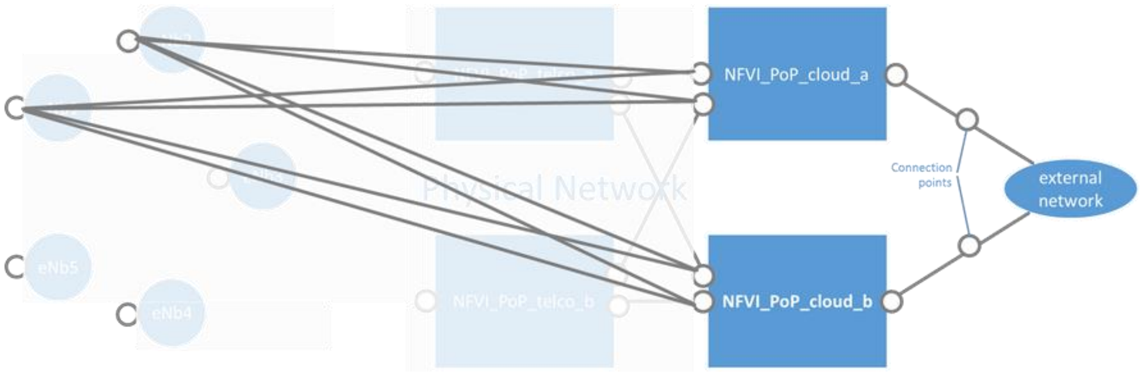


FIGURE 29: 5GT-SO VIEW FOR ABSTRACTION ALTERNATIVE 3

The following part defines the information model for the aforementioned resources, i.e. the list of parameters describing such resources exposed to the 5GT-SO, which is thus enabled for orchestration. Such information is communicated by the 5GT-MTP to the 5GT-SO through the 5GT-SO SBI/5GT-MTP NBI defined in [4] and [8].

In the 5GT-MTP network resources are represented by Logical Links (LL) defined as the link interconnecting two IP endpoints. Table 17 shows the information model for a logical link, including the IP address of terminating nodes and information related to the bandwidth and the latency induced by such connectivity. Table 18 shows the information model of computational resources including info about CPU, RAM, and address. Table 19 shows the information model of storage resources including info about memory and address.

TABLE 17: ASSUMED LOGICAL LINK PARAMETERS TO BE EXCHANGED WITH THE 5GT-SO

Identifier
Source node
Destination node
Available bit rate
Max bit rate
Latency

TABLE 18: INFORMATION MODELLING TO DEFINE A COMPUTATIONAL RESOURCE

Identifier
Max CPU
Used CPU
Max RAM
Used RAM
IP address

TABLE 19: INFORMATION MODELLING TO DEFINE A STORAGE RESOURCE

Identifier
Max Memory
Available Memory
IP address

Finally, based on the information model, a data model has to be defined to express relevant information parameters, thus enabling the exchange of this information through the adoption of proper protocol. YANG is one candidate language for data modelling [3]. Recently, an IETF informational draft provides a technology independent information model for transport network slicing based on YANG [48]. Figure 30, Figure 31 and Figure 32 show the tree representation of YANG models for logical links, computational resources, and storage resources, respectively. The tree representations reflect the information models of Table 17, Table 18 and Table 19. Data modelling, for each resource type, presents a list: e.g., a list, thus a set, of logical links, each one identified by an identifier. For each information parameter, then, the data model expresses the "type" of parameter. As an example, in Figure 32, the parameter "ip" (which refers to the IP address of the resource) is expressed with the "ip-address" type. For more details on information and data models, the reader can refer to the 5GT D2.3 [4].

```
module: LLD
   +--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
```

FIGURE 30: YANG TREE REPRESENTATION OF LOGICAL LINKS

```
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 31: 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 32: YANG TREE REPRESENTATION OF STORAGE RESOURCES

A.7 Workflows

Through this section we present the workflows that should be used by verticals to on-board, instantiate, modify, and terminate simple, non-nested vertical services in a single-domain scenario. In addition, the workflows regarding monitoring and federation are also presented. These workflows rely on the three main components of the 5G-TRANSFORMER system, namely 5GT-VS, 5GT-SO, and 5GT-MTP.

A.7.1 NFV Network Service Service On-boarding

Description: The workflow describes the on-boarding of VNFs, VAs, and NFV-NS that are respectively, initiated by the 5GT-SO admin, the verticals, and the 5GT-VS.

Prerequisites: The mapping between vertical service and the NSD has been already performed.

Assumptions: None.

Workflow: The workflow is shown in Figure 33. It can be split into three actions, which are started at separate times by different actors.

Action 1: On-boarding of VNFs

This first action is triggered by the 5GT-SO administrator or operator, which requests the 5GT-SO to on-board and provides the VNF package (including, the VNFD, and additional configuration files) as inputs (1). The 5GT-SO requests the software provider to provide the VNF package (2). The software provider replies by sending the VNF package (3). At the reception of the VNF package, the 5GT-SO extracts the VNFD and stores it in its VNFD/AppD catalogue (4). At the end of this action, the 5GT-SO sends a successful on-boarding message to the 5GT-SO service provider (5).

Action 2: On-boarding of VAs

This action is triggered by a vertical, which requests the 5GT-VS to on-board a VA (Virtual Application that is developed or used by the verticals which can be a VNF or a vertical application also including MEC applications), in which the VA package (6) is included. When received, the 5GT-VS sends the 5GT-SO a request to on-board the VA package, including the VNFD (7) to which the 5GT-SO replies with a request for the VA package itself (8). The 5GT-VS replies with the VA package, as well as the VNFD or AppD and related additional configuration files (9). Next, the 5GT-SO extracts the VNFD or AppD from the VA package and stores it into its VNFD/AppD catalogue (10). In case of MEC applications, the 5GT-SO may request to on-board the AppD into the 5GT-MTP, where it is stored in the local AppD DB (11, 12, 13). AppDs need to be processed at the infrastructure (5GT-MTP) level during the service instantiation time to instantiate MEC specific services and perform related configurations (such as traffic redirection, see Section A.3.1.9.3). The 5GT-SO sends a successful uploading of VAs message to the 5GT-VS (14), which in turn confirms, to the vertical, the successful on-boarding of VAs requests (15).

Action 3: On-boarding of VNF-NSs

This last action is triggered by the 5GT-VS. It sends an on-boarding request to the 5GT-SO (16), in which the NSD info is included. The 5GT-SO checks that all the VNFs referenced in the NSD are already stored in its VNFD/AppD catalogue (17), then stores the NSD in its NSD catalogue (18). If these two last operations succeed (i.e., 17 & 18), the 5GT-SO replies with a success message to the original request from the 5GT-VS (19).

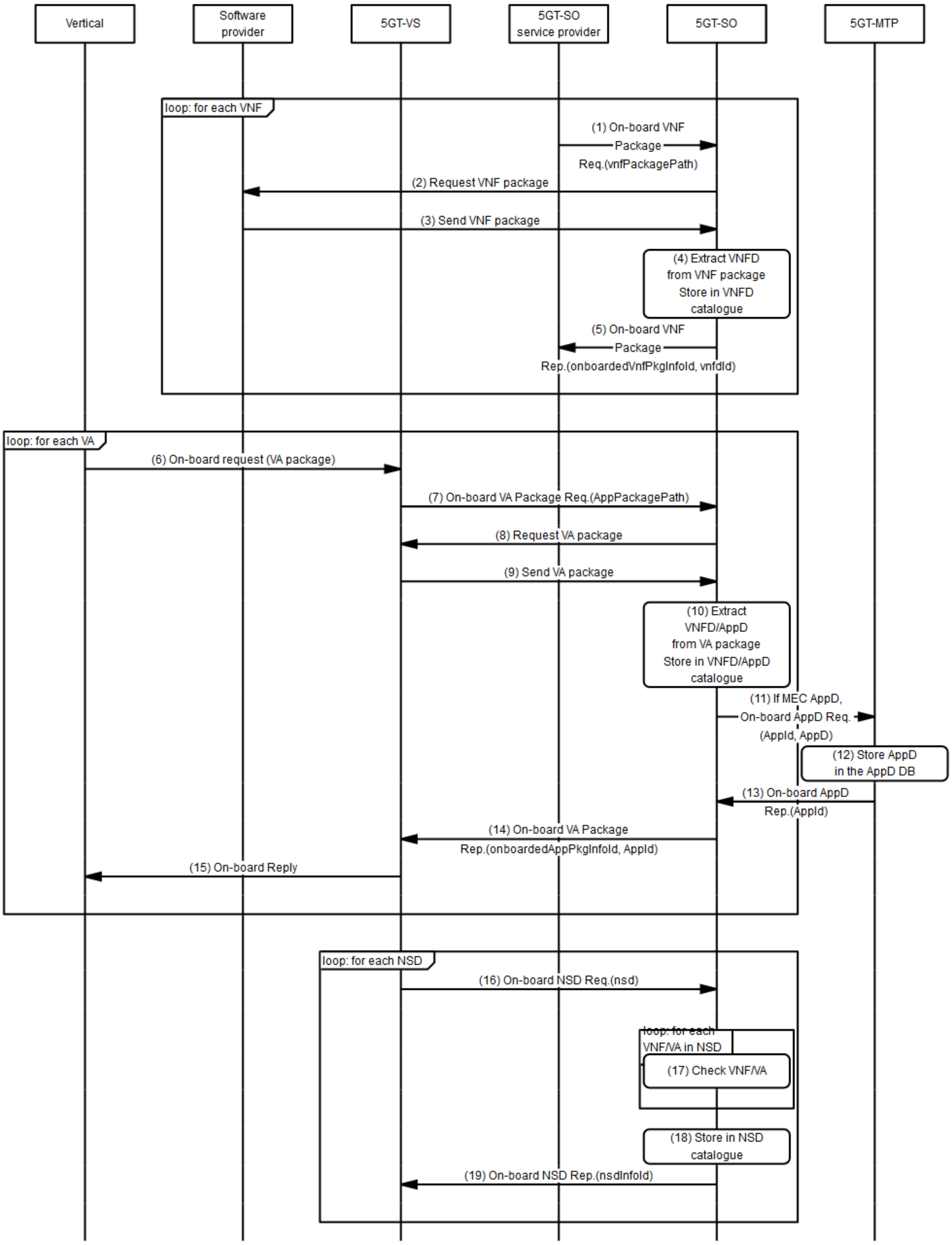


FIGURE 33: SERVICE ON-BOARDING WORKFLOW

A.7.2 Vertical Service Instantiation

Description: The workflow describes the instantiation of a vertical service, triggered by the vertical.

Prerequisites: The vertical has already selected the appropriate blueprint and prepared a vertical service description from it.

Assumptions: It is assumed that the requested vertical service is simple and deployable in a single new network slice instance. Besides that, the NSD to which the vertical service is mapped and which describes the network slice should be already on-boarded to the 5GT-SO and the same for the VSD on the 5GT-VS.

Workflow: The workflow is presented in Figure 34. The vertical requests for the instantiation of a vertical service identified by a vsdid (1). Triggered by this request, the vertical should proceed to authentication and authorization (2). Next, the 5GT-VS creates and stores an entry for the vertical service instance (VSI) in its repositories (3). After that, the VSI ID is sent with a response message to the vertical (4). At this stage, the service is not yet deployed, this is an internal bookkeeping of the 5GT-VS.

Next, the 5GT-VS maps the VSD to an NSD and checks whether a new VSI would be possible according to the vertical's resource budget (5). If this budget has not been exhausted, then the 5GT-VS requests for a new network service identifier from the 5GT-SO (6), which replies with a nfv_nsInstanceId (7). Note that, the VSI is mapped to a network slice instance (NSI), which is further mapped to an NFV network service instance (NFV-NSI). This latter is described by an NSD on-boarded to the 5GT-SO.

After receiving the NFV-NSI ID, the 5GT-VS requests the 5GT-SO to instantiate the NFV-NS with a specific deployment flavour (8). The 5GT-SO makes the orchestration decisions on NFV-NS, VNF placement and resource allocation (9), then informs the 5GT-VS that the NFV-NS instantiation is started (10).

Based on the resource decisions made by the 5GT-SO, for each VNF the 5GT-SO sends a "resource allocation request" to the 5GT-MTP to ask for the allocation of virtual resources (storage, compute, and network) inside the 5GT-MTP (11). After successful instantiation and allocation of resources for all the VNFs (12), the 5GT-MTP responds to the 5GT-SO, with the allocated resources information and their IDs (13). Afterwards the 5GT-SO updates its NFVI resource repository with the information provided by the 5GT-MTP (14). Alternatively, the 5GT-SO may also request an update of the resource abstraction from the 5GT-MTP with certain mechanisms, e.g., via periodical polling.

Afterwards, the 5GT-SO instantiates the NFV-NSI by instantiating the VNFs included in this NSD (15), then updates its NFV-NSI record with the last instantiated service (16). Then, it informs the 5GT-VS (17), which in turn updates its VSI, and NSI records with the instantiated vertical service and network slice, respectively (18). Finally, a notification is sent to the vertical about the instantiation of his vertical service (19).

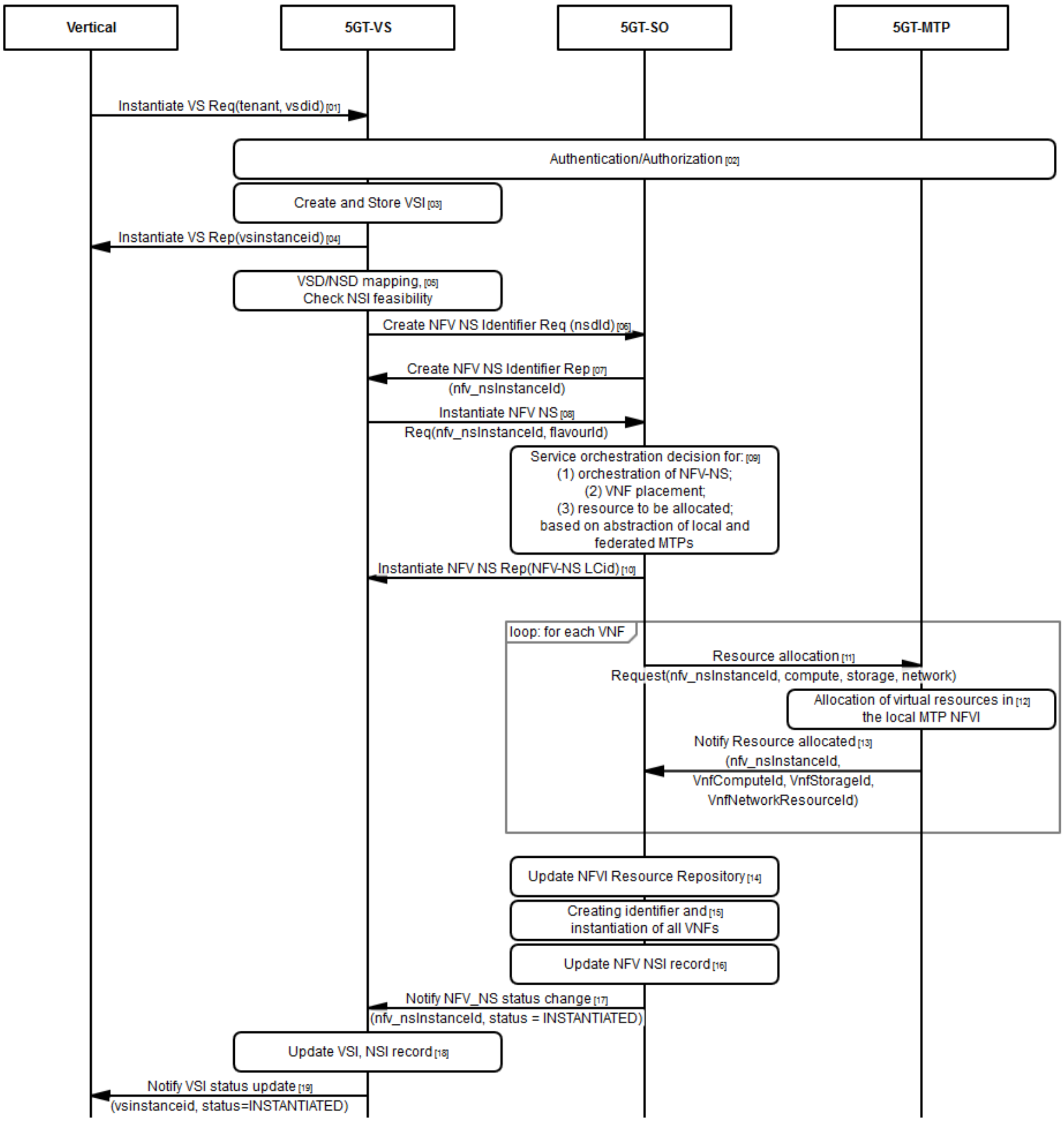


FIGURE 34: VERTICAL SERVICE INSTANTIATION WORKFLOW

A.7.3 Vertical Service Termination

Description: The workflow describes the termination of a vertical service, triggered by the vertical.

Prerequisites: The vertical service instance has been instantiated.

Assumptions: Single vertical service instance is mapped to a single network slice instance, which is mapped to a single network slice instance.

Workflow: This workflow is described in Figure 35. When a vertical want to terminate its VSI, it requests the 5GT-VS about its motivation of terminating the VSI, which includes the vsInstanceId (1). The 5GT-VS proceeds to the authentication of the Vertical, thereafter, the 5G-VS queries its database to get the nsInstanceId that corresponds to the vsInstanceId (2). The 5GT-VS checks the status of the obtained nsInstanceId, maps the nsInstanceId to the corresponding nfv_nsInstanceId (3). The vertical is informed that the termination has started, but is still ongoing (4). The 5GT-VS sends the 5GT-SO a termination request for the corresponding nfv_nsInstanceId (5). The 5GT-SO initiates VNFs termination, Virtual Links (VLs) and VNFFG deletions and release of orchestrated resources (6). The 5GT-SO responds to the termination of the lifecyle for the nfv_nsInstanceId including the LCid or lifecycleOperationOccuranceId (7). The 5GT-SO sends a Termination request to the 5GT-MTP for resources that were held by the NFV Network Service instance in the 5GT-SO (nsi_id) and the VNFs (vnf_id) (8). The 5GT-MTP performs release of resources and decoupling of the resources with the corresponding nsi_id and vnf_id (9). The 5GT-MTP responds with successful release of the corresponding nsi_id and vnf_id (10). The 5GT-SO updates the NFVI Resource Repository affected with the recently released resource (11). The 5GT-SO sends a notification message with the terminated status of the nfv_nsInstanceId (12). The 5GT-VS concludes termination of the nsInstanceId that corresponds with the terminated nfv_nsInstanceId (13). The 5GT-VS notifies the vertical that the VSI has terminated (14).

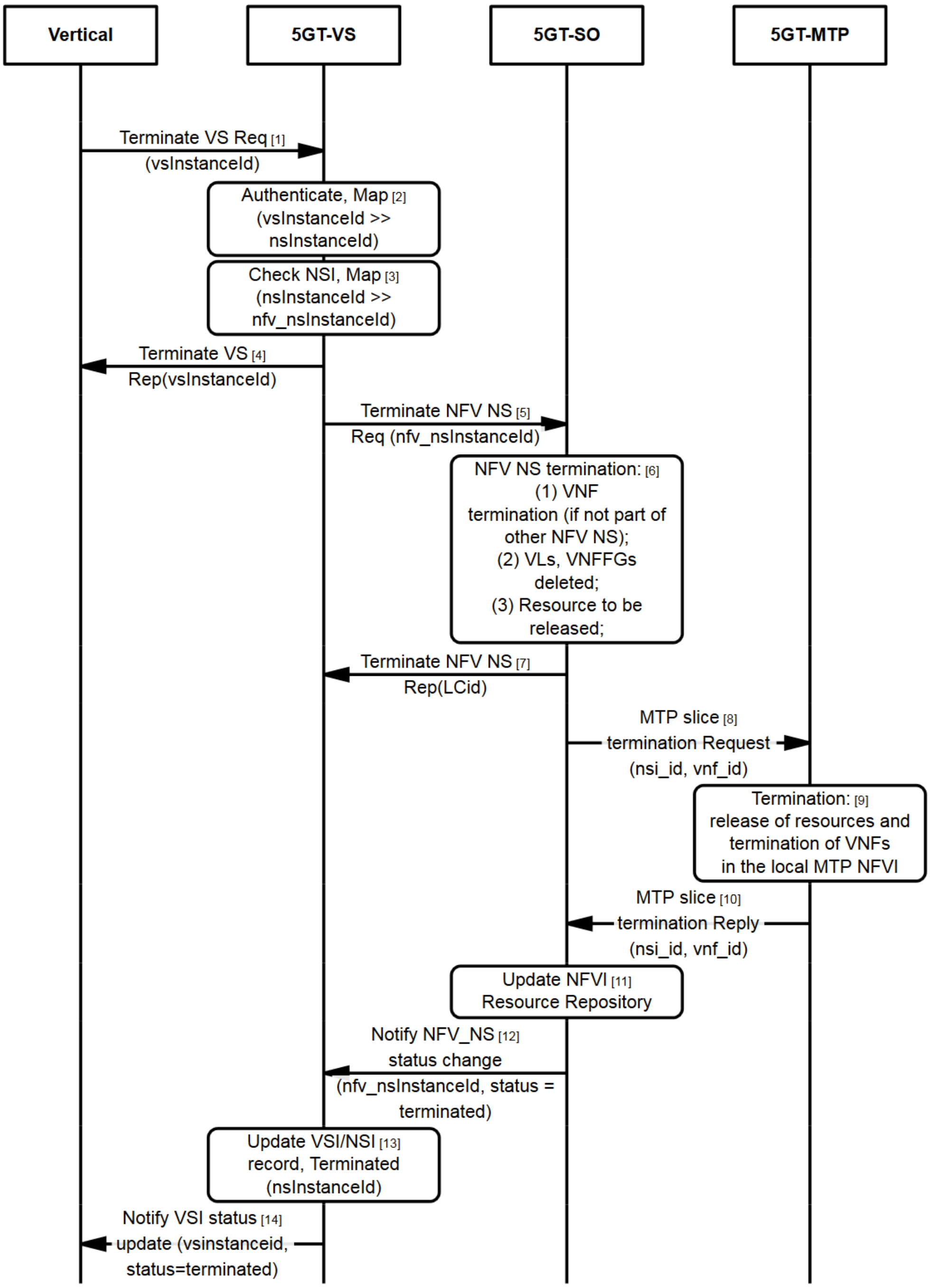


FIGURE 35: VERTICAL SERVICE TERMINATION WORKFLOW

A.7.4 Vertical Service Modification

Description: The workflow describes the modification of a vertical service, requested by a vertical.

Prerequisites: The vertical service instance has been instantiated.

Assumptions: It is assumed that the vertical service is deployed in a single non-shared network slice.

Workflow: The workflow is described in Figure 36. The workflow is split into two steps.

Step 1: the VSI is changed

The vertical provides a modified description for the vertical service (1). The new VSD is stored in the VSD record (2), and a new vsdId is provided for the vertical (3).

Step 2: the VSD of a specific VSI is changed

In this second step, the vertical applies the new VSD to a VSI. The 5GT-VS maps the vsinstanceid to the used version of the VSD (11). The operation returns immediately, while the actual processing continues. Based on the previous VSD, the 5GT-VS can determine the necessary changes to the NSD and check whether the modified version is feasible within the resource budget of this vertical (21). Thereafter the 5GT-VS triggers the changes to the network slice by scaling its defining NSD. In our example, this could be a change of the VNFs cardinality or a change of the deployment flavour. Note that depending on the severity of the NSD changes, the NSDs changes could be also more severe and might trigger more complex interactions among 5GT-VS and 5GT-SO. The 5GT-SO makes the placement decisions corresponding to these changes (23). The NFV_NS Scale operation returns (24), but the activities continue with applying the changes to the 5GT-MTP. Depending on these changes, the 5GT-SO triggers add/remove/update requests for VNFs, identified by their vnfId, at the 5GT-MTP (31, 32, 33). The 5GT-SO updates its record of used resources and of instantiated network services (41, 42) and notifies the 5GT-VS that the modification has finished (43). In turn, the 5G-VS updates the VSI/NSI record (44) and notifies the vertical that the modification is finished (45).

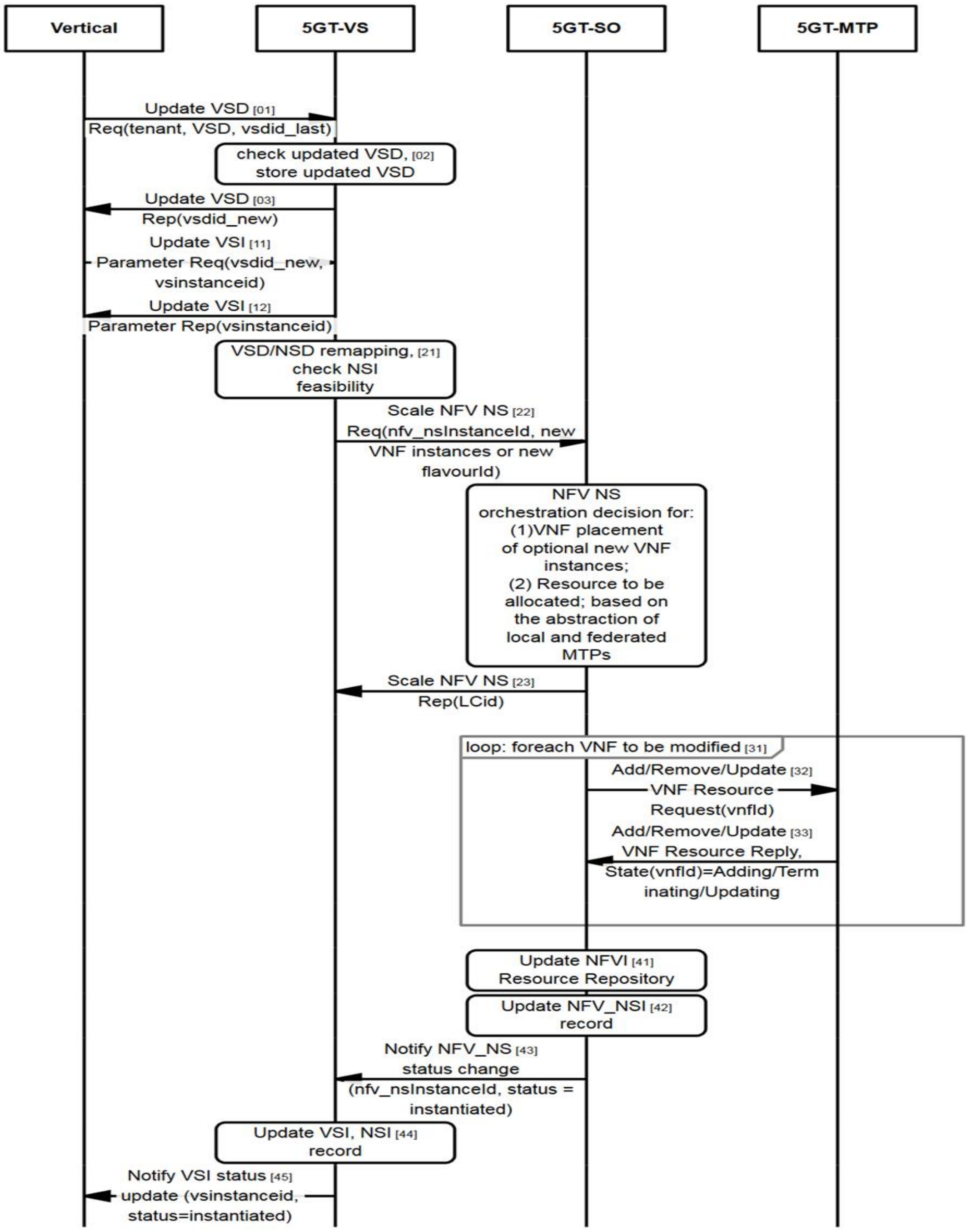


FIGURE 36: VERTICAL SERVICE MODIFICATION WORKFLOW

A.7.5 Vertical Service Monitoring

We distinguish two cases, according to whether we monitor NFV-NS instances or VS instances.

Case 1: Monitoring of NFV-NS instances.

This is done by the 5GT-SO, so it can take internal decisions including the scaling.

Description: This workflow describes how the 5GT-SO collects monitoring data of an NFV-NSI.

Prerequisites: A nsi_Id has been created by the 5GT-VS, which assumes that all the descriptors have been already on-boarded and a vertical has requested the creation of a vertical service.

Assumptions: We suppose that:

i. A vertical requests for a service in a single domain scenario.
ii. The MonitoredData in the NSD includes only MonitoringParameters, not VnfIndicatorInfo.
iii. NS Performance Management Interface is exposed by the 5GT-SO monitoring service via the OS-Ma-Nfvo interface [29].
iv. Virtualized Resource Performance Management Interface is exposed by the 5GT-MTP monitoring service via the Vi-Vnfm interface [23].
v. We consider that the 5GT-SO, 5GT-SO monitoring service, and 5GT-MTP monitoring service are in client-server relationship.

Workflow: This workflow is represented by Figure 37 and Figure 38. The actions (01) and (02) and the procedures (03) and (04) from Figure 37 summarize the vertical service instantiation procedure. For more details about this procedure, please refer to Section A.7.2 regarding the Vertical Service Instantiation Workflow.

After the resource allocation phase is concluded at the 5GT-MTP and the 5GT-SO is aware of the created virtual resources, the 5GT-SO starts the monitoring procedure. For each monitoring parameter, encoded in the NSD *monitoredData* field of the instantiated NFV-NS, the 5GT-SO creates a PM job and subscribes with the 5GT-SO Monitoring (05). Each of these monitoring parameters is a performance metric to be monitored at the NFV-NS or the VNF level. We define this metric as *5GT-SO performance metric*.

For each of the created PM jobs, the 5GT-SO Monitoring Service needs to register with the 5GT-MTP Monitoring Service to receive the required *elementary* monitoring data (defined as *5GT-MTP performance metrics*). Indeed, each of the 5GT-SO performance metrics is translated into one or more 5GT-MTP performance metrics (06). For each of the required 5GT-MTP performance metric, the 5GT-SO Monitoring Service requests the 5GT-MTP Monitoring Service to create a PM job (07), (08), which is then stored in the internal repository (09). In addition, the 5GT-SO Monitoring Service subscribes to the 5GT-MTP Monitoring Service to receive monitoring notifications with a specific filter (10), (11), and stores the subscription notification (12).

Once all the subscriptions are completed, the 5GT-SO Monitoring Service is able to receive all the required monitoring data from the 5GT-MTP Monitoring Service. Hence, it instantiates a new internal job to process the monitoring data and generates the target 5GT-SO performance metrics (13). This step concludes the creation of the PM job and the related ID is returned to the 5GT-SO (14). Finally, in order to receive notifications about the 5GT-SO performance metric, the 5GT-SO subscribes with the 5GT-SO Monitoring Service (15), (16) and stores the subscription information (17).

The collection of monitoring data at the service runtime is shown in Figure 38. At the end of (17), the 5GT-SO Monitoring Service has subscribed to notifications coming from the 5GT-MTP Monitoring Service. When the latter produces new monitoring data for which the 5GT-SO Monitoring Service has an active subscription, it sends a notification (18) and the 5GT-SO Monitoring Service sends a request to retrieve the data (19), (20). If needed, the 5GT-SO Monitoring Service aggregates the data (21) and stores the result as a new performance metric value in its repository (22). Then the 5GT-SO is in turn notified about the availability of a new performance report related to the given NFV-NSI (23) and the 5GT-SO retrieves the new report (24), (25). This may trigger a new action at the 5GT-SO (e.g. a scaling procedure (26)).

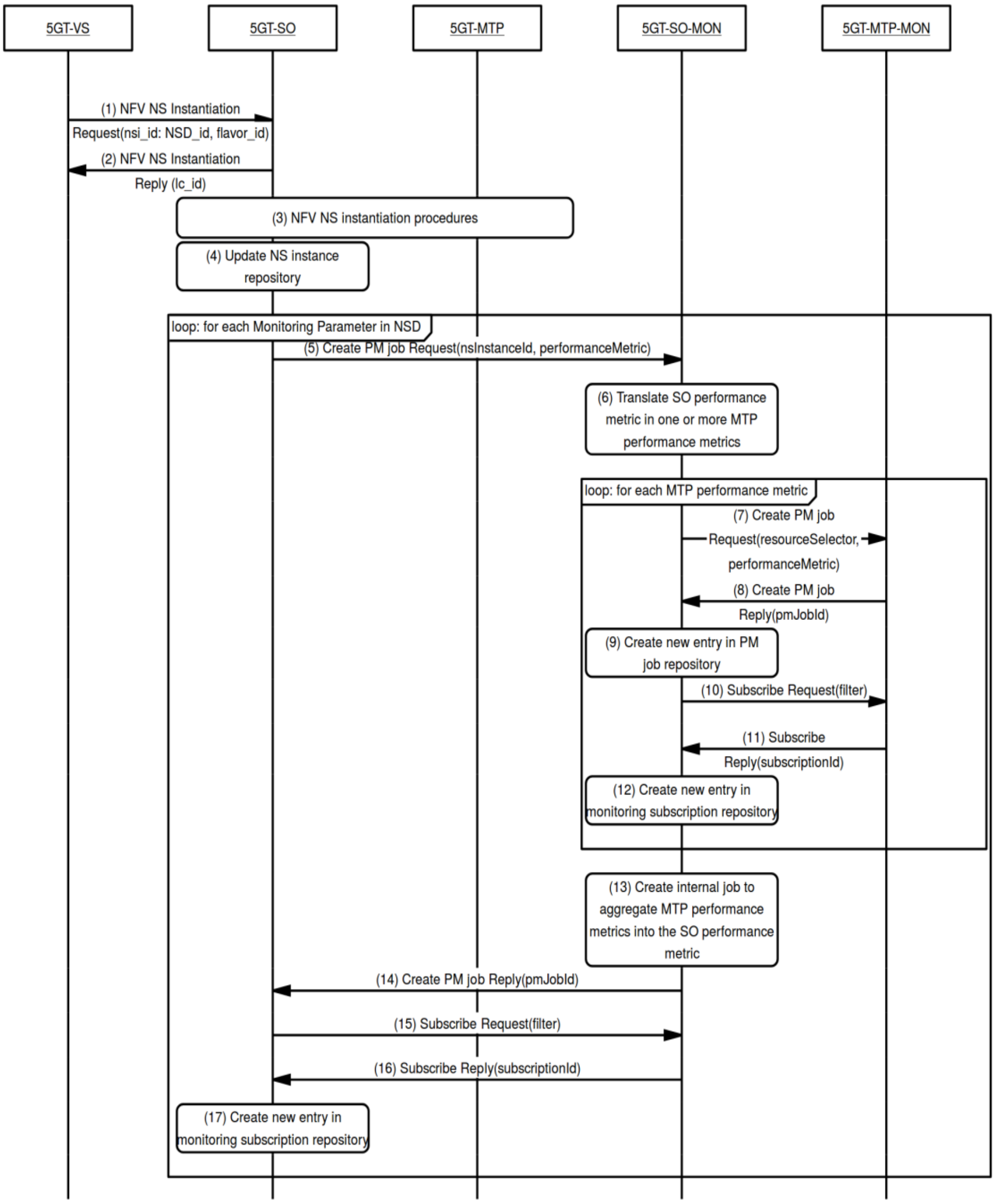


FIGURE 37: VERTICAL SERVICE MONITORING WORKFLOW BY 5GT-SO (1)

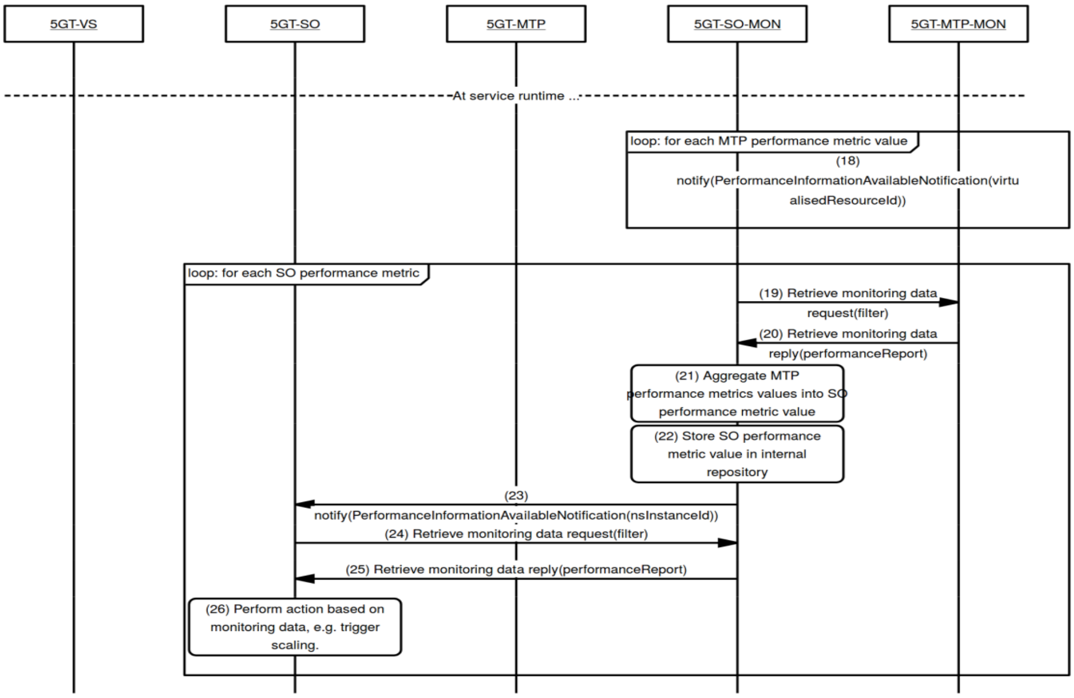


FIGURE 38: VERTICAL SERVICE MONITORING WORKFLOW BY 5GT-SO (2)

Case 2: Monitoring of VS instances.

Description: This workflow describes how the 5GT-VS collects monitoring data about an NFV-NSI.

Prerequisites: All the descriptors have been on-boarded, the vertical has requested the instantiation of a new Vertical Service and the 5GT-VS has already created a nsi_Id.

Assumptions: We suppose that:

i. A vertical requests for a service in Single-domain scenario.
ii. The MonitoredData in the NSD includes only MonitoringParameters, not VnfIndicatorInfo.
iii. *One-to-One* relationship between a Vertical Service instance and Network Service instance (no sharing or nesting)
iv. NS Performance Management Interface is exposed by the 5GT-SO monitoring service via the OS-Ma-Nfvo interface [29].
v. Virtualized Resource Performance Management Interface is exposed by the 5GT-MTP monitoring service via the Vi-Vnfm interface [23].
vi. We consider that the 5GT-SO, 5GT-SO Monitoring service, and 5GT-MTP monitoring service are in client-server relationship.

Workflow: The main difference with the previous case is that the entity requesting for monitoring parameters about a NFV-NSI is the 5GT-VS instead of the 5GT-SO. This workflow is represented by Figure 39 and Figure 40.

Through Figure 39, a network service is instantiated (1-4) and the required PM jobs at the 5GT-SO Monitoring Service are created (5-14) in the same way as in the previous case. However, now the entity that subscribes to the 5GT-SO Monitoring Service to receive the 5GT-SO performance metrics is the 5GT-VS (16-17). Obviously, this can be done only after the 5GT-VS has been notified by the 5GT-SO about the successful instantiation of the NFV-NS (15).

At runtime (Figure 40), the procedure to collect 5GT-MTP performance metrics and aggregate them into 5GT-SO performance metrics at the 5GT-SO Monitoring Service is again the same as in the previous case (19-23). However, since the subscription has been performed by the 5GT-VS, the notification about the availability of a new performance information is sent to the 5GT-VS itself (24), which in turn sends a request to retrieve the monitoring report (25-26).

This workflow can be extended to any kind of 5GT-SO Monitoring Services' consumer, with suitable authorization rights (e.g. a federated 5GT-SO).

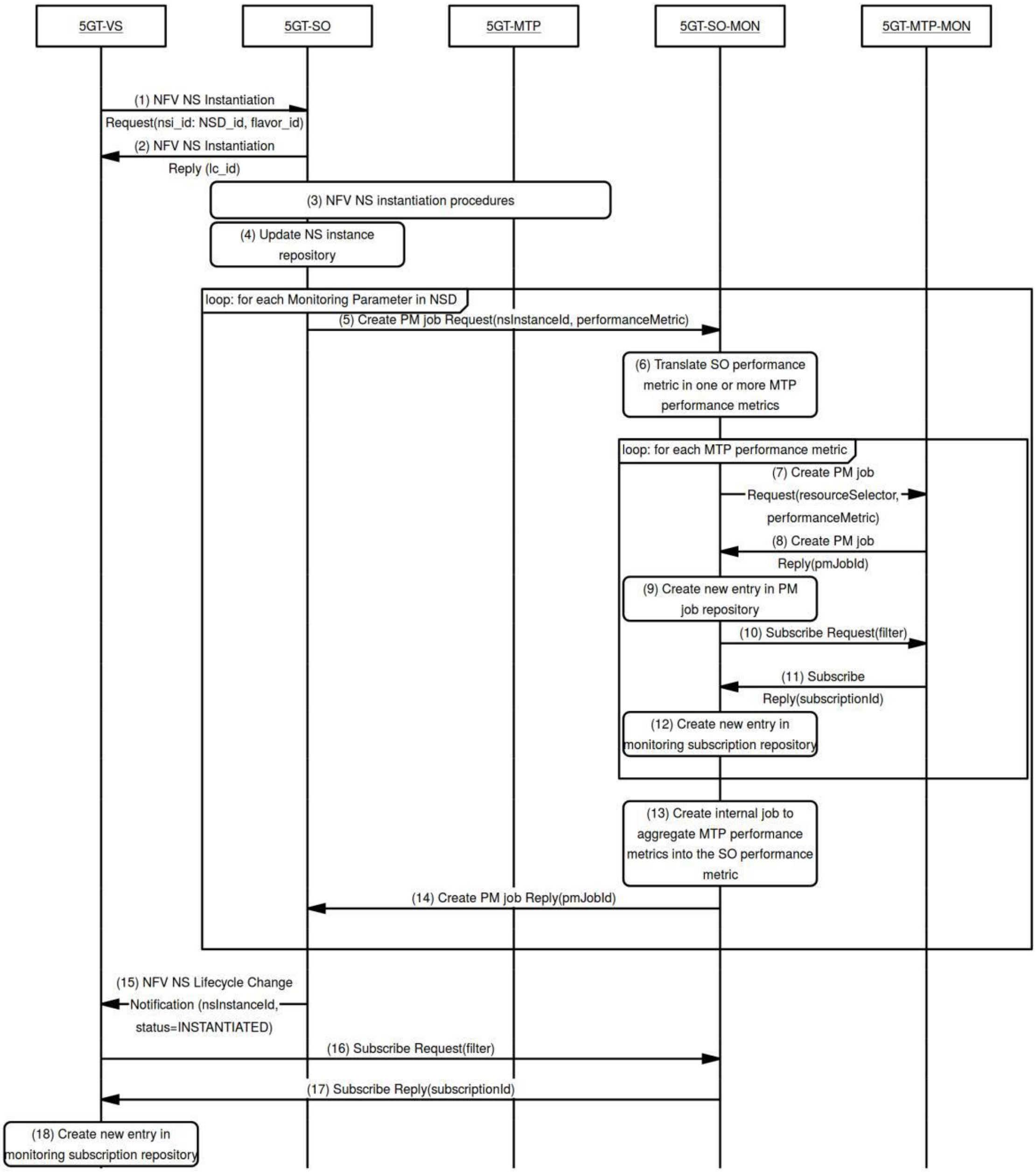


FIGURE 39: VERTICAL SERVICE MONITORING WORKFLOW BY 5GT-MTP (1)

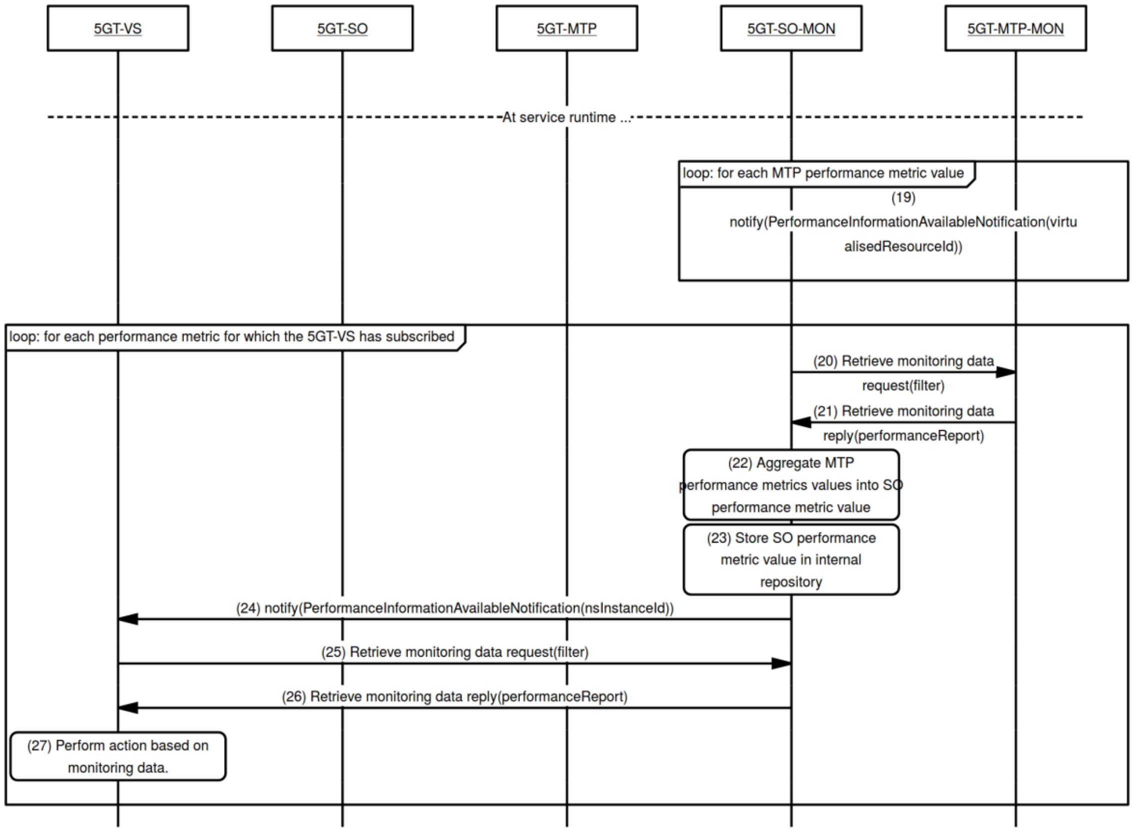


FIGURE 40: VERTICAL SERVICE MONITORING WORKFLOW BY 5GT-MTP (2)

A.7.6 Service Federation

Description: The workflow describes federation of NFV-NSs (NSaaS) between different administrative domains.

Prerequisites: Different administrative domains need to have some business relations.

Assumptions: 5GT-SO receives a request for instantiation of a certain NFV-NS (with the NSD) that the 5GT-SO has no sufficient resource capabilities or capacity, it can exploit the service federation for accommodating the requested service.

Workflow: This workflow is represented by Figure 41. The agreed administrative domains (white and grey in Figure 41) create a set of NFV-NSs that each domain can offer to each peering administrative domain. A set of agreed and offered NFV-NSs create a *white-list* of services which is referred to as catalogue of services. The pre-phase is concluded when the catalogues from each peering administrative domain is stored into the Catalogue DB (1).

NFV-NS instantiation request received from the 5GT-VS by the 5GT-SO (2). The 5GT-SO performs decomposition of the requested NFV-NS. The decomposition of requested NFV-NS can result in a single or set of multiple nested NFV-NSs. In this workflow, a decision is made to consume one of the set of decomposed nested NFV-NSs through federation of NFV-NS or by consuming NFV-NSaaS (3). The consumer 5GT-SO sends a request for NFV-NS instantiation to the selected provider 5GT-SO (4). The peering 5GT-SO check the availability of the requested nested NFV-NS and starts the instantiation (5). The provider 5GT-SO completes the instantiation procedure involving its constituent 5GT-MTP and positive feedback is sent back to the consumer 5GT-SO (6). The consumer 5GT-SO integrates the federated NFV-NS into the end-to-end NFV-NS by inter-nesting (stitching) with the other nested NFV-NS instantiated locally (7, 8).

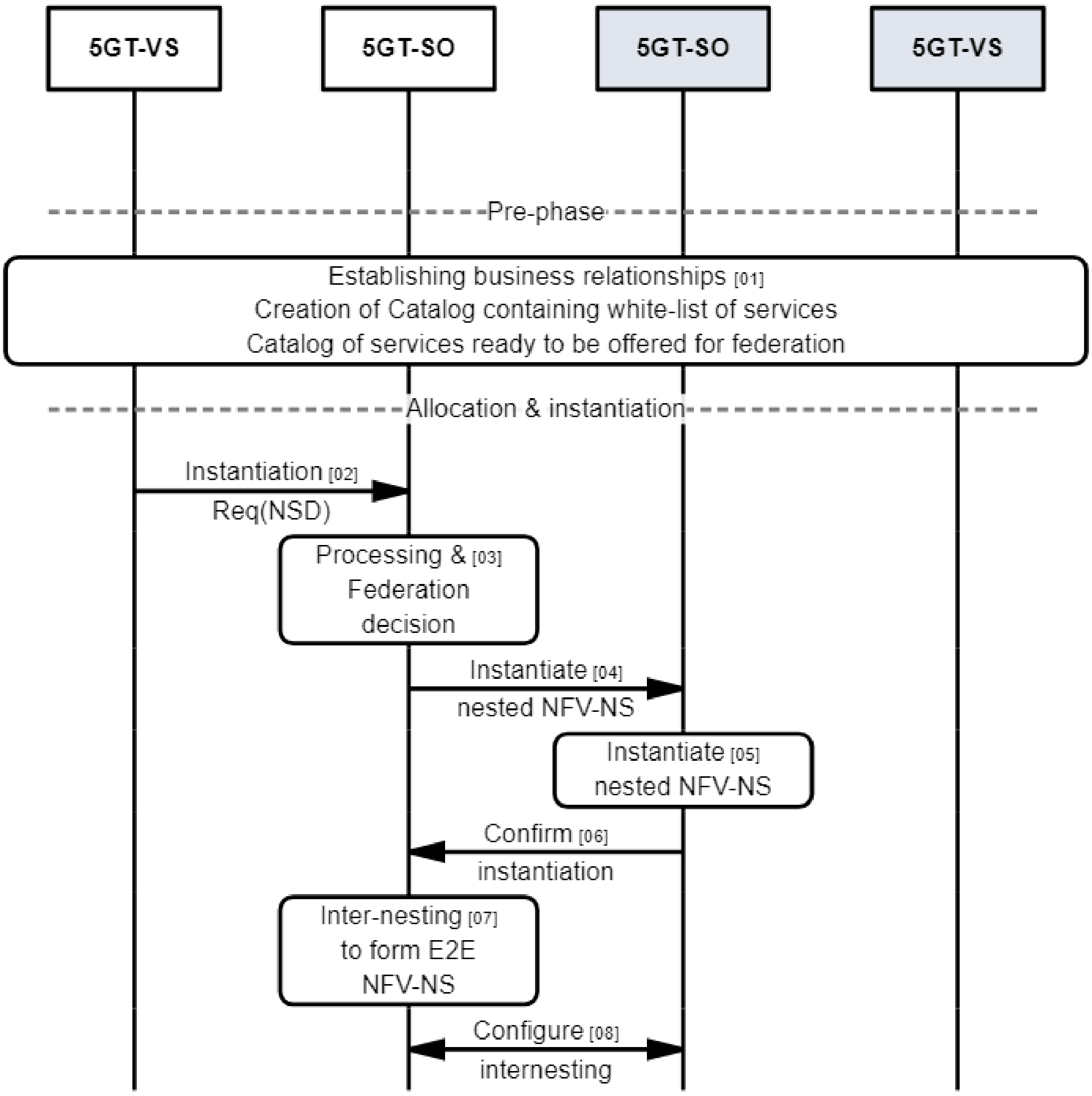


FIGURE 41: SERVICE FEDERATION WORKFLOW

A.7.7 Resource Federation

Description: The workflow describes federation of NFVI resources or resource abstractions (NFVIaaS) between different administrative domains.

Prerequisites: Different administrative domains need to have already established business relations.

Assumptions: There are lack of NFVI resources at the local (constituent) 5GT-MTP.

Workflow: This workflow is represented by Figure 42. The pre-phase usually is part of the same business agreements between administrative domains, held "offline" as in the NFV-NSaaS. All the terms and conditions are agreed about how the interactions between the agreed administrative domains will proceed (1).

The 5GT-MTP sends/advertises resource abstractions to the 5GT-SO (Advertising resource abstractions) (2). The 5G-SO stores the received resource abstractions and calculates the available resource abstractions (ready to be offered) for federation to potential consumer 5GT-SOs (3). The 5GT-SO broadcasts/advertises the available resource abstractions. A process of exchanging (sending and receiving) calculated resource abstractions among different domains (4). When an advertisement from other peering 5GT-SO is received, the carried information is then filtered and stored (5).

In the 5GT-SO, federation of resources is triggered for instantiating/scaling-out a NFV-NS/VNF by using resources from external 5GT-SO (6). The 5GT-SO issues a request for the actual allocation of the available resource abstractions at the peering 5GT-SO (7). The peering (provider) 5GT-SO processes the request (8). The provider 5GT-SO sends it to its constituent 5GT-MTP (9). The provider 5GT-MTP allocates and instantiates the resources that correspond to the requested resource abstractions (10). Once the resources are allocated and instantiated, the provider 5GT-MTP sends back a positive response to the provider 5GT-SO (11). At the same time, the provider 5GT-MTP updates the provider 5GT-SO with the newly abstracted view of the resources as result of the performed actions (12). The provider 5GT-SO provides a positive response to the consumer 5GT-SO with some details of the newly allocated resources (13). The consumer 5GT-SO includes the allocated federated resources in a NFV-NS/VNF and establishes tunneling connection. The connection from the consumer 5GT-SO to the federated resources will pass through the provider 5GT-SO (as a forwarding point) up to the peering provider 5G-MTP as an end-point (14). The provider 5GT-SO re-calculates its abstraction capabilities and triggers the advertisement phase (15). The 5GT-SO broadcasts the new resource abstractions to all peering 5GT-SOs (16). The consumer 5GT-SO receives the new resource abstractions and stores the newly received resource abstractions (17).

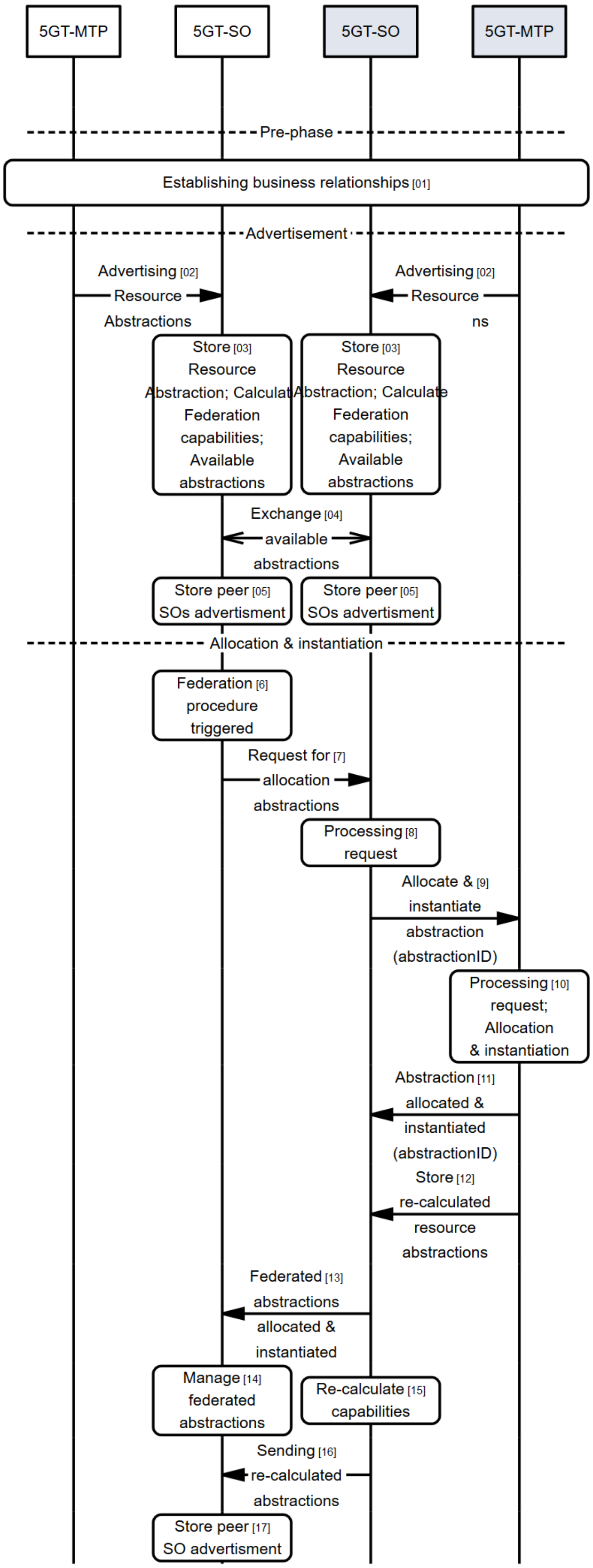


FIGURE 42: RESOURCE FEDERATION WORKFLOW

A.8 Glossary

This section defines the key terminology used in 5G-TRANSFORMER. The definitions of terms are structured according to their area, such as virtualization related or business related. The terms defined here are the most relevant ones, especially those that have different definitions by various standardization developing organizations.

A.8.1 General Terms

Data model: [59] a mapping of the contents of an information model into a form that is specific to a particular type of data store or repository. A "data model" is basically the rendering of an information model according to a specific set of mechanisms for representing, organizing, storing and handling data. It has three parts:

- A collection of data structures such as lists, tables, relations, etc.
- A collection of operations that can be applied to the structures such as retrieval, update, summation, etc.
- A collection of integrity rules that define the legal states (set of values) or changes of state (operations on values).

Information model (IM): an abstraction and representation of the entities in a managed environment, their properties, attributes and operations, and the way that they relate to each other. It is independent of any specific repository, software usage, protocol, or platform. [59]

Policy: [1] policy can be defined from two perspectives:

- A definite goal, course or method of action to guide and determine present and future decisions. "Policies" are implemented or executed within a particular context (such as policies defined within a business unit).
- Policies as a set of rules to administer, manage, and control access to network resources.

NOTE: [59] These two views are not contradictory since individual rules may be defined in support of business goals.

Service: the behaviour or functionality provided by a network, network element or host. To completely specify a "service", one must define the "functions to be performed ..., the information required ... to perform these functions, and the information made available by the element to other elements of the system". Policy can be used to configure a "service" in a network or on a network element/host, invoke its functionality, and/or coordinate services in an interdomain or end-to-end environment. [59]

A.8.2 Network function virtualization related

The central concepts around network function virtualization and network services are based on the definitions of ETSI NFV.

Network Function (NF): functional block within a network infrastructure that has well-defined external interfaces and well-defined functional behaviour. [21]

Network Service (NFV-NS): composition of Network Functions and defined by its functional and behavioural specification. [21]

NOTE: "The Network Service contributes to the behaviour of the higher layer service, which is characterized by at least performance, dependability, and security specifications. The end-to-end network service behaviour is the result of the combination of the individual network function behaviours as well as the behaviours of the network infrastructure composition mechanism." [21]

NOTE: A network service can be seen as a set of VNFs or PNFs, connected by VLs as defined in a VNFFG.

Network Service Descriptor (NSD): template that describes the deployment of a Network Service including service topology (constituent VNFs and the relationships between them, Virtual Links, VNF Forwarding Graphs) as well as Network Service characteristics such as SLAs and any other artefacts necessary for the Network Service on-boarding and lifecycle management of its instances. [21]

NOTE: The NSD includes a number of deployment flavours, each referencing deployment flavours of all or a subset of the NFV-NS's constituent VNFs and Virtual Links. The NSD also provides a list of pointers to the descriptors of its constituent VNFs (i.e. VNFDs) and additional information on the connectivity between them together with the traffic forwarding rules.

Network Service Instance (NFV-NSI): refers to an instance of a network service (NFV-NS).

NOTE: In our case the process for instantiating a Network Service instance is initiated from 5GT-VS by sending a request to the 5GT-SO. This request typically contains a pointer to a NSD, a flavour selector and additional input parameters (e.g., IP addresses to be assigned to some of the network functions) and constraints (e.g. location where to deploy all or some of the network functions). The flavour mechanism not only enables selecting a subset of VNFs and virtual links to be instantiated but also the actual flavour for each of the selected objects and the number of instances to be created for each selected VNF.

NFVI as a Service (NFVIaaS): the tenant (e.g., a vertical or an MVNO) is offered a virtual infrastructure including associated resources (networking/computing/storage) under its full control in which it can deploy and manage its own NFV network services on top of it. It is assumed that the vertical will deploy its own MANO stack. This is probably the most usual service consumed by M(V)NOs, given that they have the knowledge and need to customize their communication service offering to their own customers. Resources could be virtual cores, storage, virtual nodes and links, etc.

NOTE: The tenant can deploy and connect VMs on these resources under its own control.

NOTE: NFVIaaS includes the provision of network slices or network slice subnets as a service.

Network Service as a Service (NSaaS): provide to a tenant the possibility to define and instantiate a network service.

NF forwarding graph (NF FG): graph of logical links connecting NF nodes for the purpose of describing traffic flow between these network functions. [21]

Physical Application (PA): implementation of a VA via a tightly coupled software and hardware system.

NOTE: analogous to PNF.

NOTE: may include devices such as cameras, smart city sensors, etc.

Physical Network Function (PNF): implementation of a NF via a tightly coupled software and hardware system. [21]

VA Forwarding Graph (VA FG): forwarding graph among VA, VNF, PA, PNF nodes.

Virtual Application (VA): more general term for a piece of software which can be loaded into a Virtual Machine. [21]

Virtual link (VL): set of connection points along with the connectivity relationship between them and any associated target performance metrics (e.g. bandwidth, latency, QoS). [21]

Virtualised Network Function (VNF): implementation of an NF that can be deployed on a Network Function Virtualisation Infrastructure. [21]

Virtualised Network Function Component (VNFC): internal component of a VNF providing a defined subset of that VNF's functionality, with the main characteristic that a single instance of this component maps 1:1 against a single Virtualisation Container. [21]

Virtualised Network Function Descriptor (VNFD): configuration template that describes a VNF in terms of its deployment and operational behaviour, and is used in the process of VNF on-boarding and managing the lifecycle of a VNF instance. [21]

VNF Forwarding Graph (VNF FG): NF forwarding graph where at least one node is a VNF. [21]

VS as a Service (VSaaS): provide to a vertical the possibility to define and instantiate a vertical service.

NOTE: similar to NSaaS, with vertical instead of network service.

A.8.3 Network slice related

Network slice (**NS**): a network slice is a complete logical network with specific services offered to customers over a shared compute, storage and network infrastructure. E.g. a network operator can build a network slice including an Access Network (AN) and a Core Network (CN) to enable communication services.

Network slice instance (NSI): a set of network functions and the resources for these network functions which are arranged and configured, forming a complete logical network to meet certain network characteristics [11].

NOTE: A Network slice instance is the realization of a network slice.

NOTE: In ETSI NFV parlance a network slice instance would typically be deployed as a NFV Network Service instance (NFV-NSI). Different slices can map to instances of the same type of NFV-NS with different deployment flavours or instances of different types of NFV-NS. In an NFV framework, creating a network slice will typically involve filling a NSD and requesting the NFV

Orchestrator to instantiate a NFV-NS according to the contents of its NSD and selected deployment flavour.

Network slice subnet instance (NSSI): a set of network functions and the resources for these network functions which are arranged and configured to form a logical network (sub-network) [11].

NOTE:

- A NSI may include one or more NSSIs, which can include one or more VNFs or PNFs.
- A NSSI can be shared by multiple NSIs.
- Both NSI and NSSI can be mapped to a nested NFV Network Service.

A.8.4 Vertical service related

Vertical: 5G-TRANSFORMER stakeholder belonging to a specific industry or group of customers and consuming 5G-TRANSFORMER services (defined in Section A.8.6). MVNOs are considered a special type of vertical in 5G-TRANSFORMER.

NOTE: The existence of network slices is transparent to the vertical and it is fully under the control of the 5G-TRANSFORMER Service Provider how to handle them, including, for instance, mapping services into network slices.

Vertical Service (VS): from a business perspective, it is a service focused on a specific industry or group of customers with specialized needs (e.g., automotive services, entertainment services, e-health services, industry 4.0). The 5G-TRANSFORMER system is designed to fulfil the requirements of vertical services coming from the 5G-TRANSFORMER Service consumers (Section A.8.6). M(V)NO requests are also handled as a special kind of vertical service.

From a technical point of view, it is a composition of general functions as well as network functions and defined by its functional and behavioural specification.

NOTE: The vertical service behaviour is the result of the combination of the individual VA behaviours as well as the behaviours of the network infrastructure composition mechanism.

NOTE: A vertical service is similar to a network service, but provides more general functionality than network functionality.

Vertical Service Blueprint (VSB): a parameterized version of a VSD, where parameters have to be provided to provide a complete VSD, which is ready to be instantiated.

NOTE: There can be a wide range of parameters. The parameters can be used to express requirements of the vertical service, but also management related parameters such as file locations of virtual machine images or the priority of a service. A subset of parameters to express requirements are: Bitrate of VAs and the connecting links, round-trip time among two VAs, geographical area to be covered by the vertical service.

Vertical Service Descriptor (VSD): a description of the deployment of a vertical service including service topology (constituent VAs and the relationships between them, Virtual Links, VNF Forwarding Graphs) as well as vertical service characteristics such as SLAs and any other artefacts necessary for the vertical service on-boarding and lifecycle management of its instances.

> NOTE: A VSD may still contain instance specific parameters to be provided at instantiation time. This is similar to parameters provided at instantiation time of VNFs.

A.8.5 Multi-access edge computing related

The central concepts around multi-access edge computing are based on the definitions of ETSI MEC [34] and recent draft integrating NFV and MEC [37]. Following the renaming of mobile edge computing to multi-access edge computing, the definitions have been changed accordingly.

Multi-access edge application (MEA): application that can be instantiated on a multi-access edge host within the multi-access edge system and can potentially provide or consume multi-access edge services. [34]

Multiple-access Edge Application Orchestrator (MEAO): it has the same functions as MEO, excepting that it should use the NFVO to instantiate the virtual resources for the MEA as well as for the MEP.

Multiple-access Edge Host (MEC Host): it provides the virtualization environment to run MEC applications, while it interacts with the mobile network entities, via the MEP platform, to provide MES and offload data to MEA.

Multiple-access Edge Orchestrator (MEO): the MEO is in charge of the orchestration and the instantiation of MEA.

Multiple-access Edge Platform Manager (MEPM): it is in charge of the lifecycle management of the deployed MEA. The MEPM is in charge of the MEP configuration, such as the MEC application authorization, the traffic type that needs to be offloaded to the MEC application, DNS redirection, etc.

Multiple-access Edge Platform Manager - NFV (MEPM-V): the virtualized version of the MEPM delegates the LCM of MEA to one or more VNFMs, and keeps the MEP configuration.

Multi-access edge platform (MEP): collection of functionalities that are required to run multi-access edge applications on a specific multi-access edge host virtualisation infrastructure and to enable them to provide and consume multi-access edge services, and that can provide itself a number of multi-access edge services. [34]

Multi-access edge service (MES): service provided via the multi-access edge platform either by the multi-access edge platform itself or by a multi-access edge application. [34]

A.8.6 Business logic/stakeholder related

5G-TRANSFORMER Service (TS): services offered by a 5G-TRANSFORMER Service Provider (TSP) to 5G-TRANSFORMER Service Consumers, such as verticals, through the 5GT-VS northbound interface or to other TSPs through the east-west interface (E/WBI). As for the former, it includes the following four types of services: 5G-TRANSFORMER Managed Vertical Service (TMVS), 5G-TRANSFORMER

Unmanaged Vertical Service (TUVS), NSaaS (sect. A.8.2), and NFVIaaS (sect. A.8.2). Additionally, TSPs can also consume TSs offered by peering TSPs. This interaction is done through the east-west interface (E/WBI) of 5GT-SOs, and so, it is not a slice-related interaction but an NFV network service-related one. There are two types of services offered in this way: federated services and federated resources (sect. A.8.7). A service offered by a 5G-TRANSFORMER Service Provider can include a bundle of such services.

5G-TRANSFORMER Managed Vertical Service (TMVS): vertical services that are fully deployed and managed by the TSP and consumed as such by the vertical (i.e., without any interface available to modify the service logic, but only for getting operational information, at most).

5G-TRANSFORMER Unmanaged Vertical Service (TUVS): vertical services that are deployed by the TSP (i.e., instantiating VNFs and their connectivity), but their logic is partly or fully managed by the vertical. This includes the configuration of VNF internals to control the logic of the vertical services at service level, e.g., the algorithms for EVS (Extended Virtual Sensing) for the automotive use case. In this case, the lifecycle management of the NFV-NS and its VNFs are still retained by the TSP.

5G-TRANSFORMER Service Consumer (TSC): uses 5G-TRANSFORMER services that are offered by a 5G-TRANSFORMER Service Provider. Note that a 5G-TRANSFORMER Service Provider can also be a TSC of another service provider.

5G-TRANSFORMER Service Provider (TSP): provides 5G-TRANSFORMER services. Designs, builds and operates its 5G-TRANSFORMER services.

5G-TRANSFORMER Mobile Transport and Computing Platform Operator (TMOP): in charge of orchestrating resources, potentially from multiple virtual infrastructure providers (VISP) and offered to the TSP. In that sense, it acts as an aggregator of resources. The virtual infrastructure features transport and computing resources, potentially including those of datacentre service providers with which the TMOP has an agreement. Designs, builds, and operates the computing and network aggregated virtual infrastructure services. It has agreements with Virtualization Infrastructure Service Providers (VISPs) (see below for a definition).

Virtualization Infrastructure Service Provider (VISP): provides virtualized infrastructure services. Designs, builds and operates its virtualization infrastructure(s) [11]. VISP-T provides virtual transport infrastructure and VISP-C virtual computing infrastructure.

Data Centre Service Provider (DCSP)[16]: provides data centre services. Designs, builds and operates its data centres. [11]

[16] The difference between DCSP and VISP-C is that the former is closer to the raw resources (host servers) offering simple services of raw resource consumption. Additionally, these resources are located in a centralized location (datacentre). The latter offers access to a variety of virtual infrastructure resources created by aggregating multiple technology domains and by making them accessible through a single API for all of them. For instance, VISP-C may offer not only centralized datacentre resources, but also distributed computing resources available throughout the network.

A.8.7 5G-TRANSFORMER specific terms

Abstracted Resource / Resource abstraction: limited description of a resource with intention to hide certain parameters (such as quantity, vendors, location of the resource, etc.) and secure enough to be shared with other administrative domains.

Abstracted Service / Service abstraction: limited description of a service with intention to hide certain parameters (such as used resources, virtual links, interconnections etc.) and secure enough to be shared with other administrative domains.

Administrative domain: is a collection of resources operated by a single organization. The internal structure (collection of resources) operates with significant degree of mutual trust among them. The domain's resources interoperate with other administrative domains in a mutually suspicious manner and they are viewed as a cohesive entity from the outside.

Available Resources for Federation: set of resources offered by a provider domain under pre-agreed terms and conditions; available resources potentially to be used by a consumer domain, with certain pre-agreed terms and conditions.

Available Services for Federation: available services offered by a provider domain to other potential consumer domains, under pre-agreed terms and conditions.

Consumer domain: administrative domain that demands resources or services from other administrative domains.

Federated Resources: resources managed by a consumer domain, but owned by a provider domain (operator). The consumer domain is allowed (by the provider domain) to manage and use the resources based on pre-agreed terms and conditions (SLAs). In this case, the consumer TSP uses NFV (abstracted) virtual resources offered by the peer TSP. This may be the case when an end-to-end NFVIaaS service is built by combining virtual resources belonging to multiple TSP administrative domains.

Federated Services: services managed by a consumer domain, but owned by a provider domain. The consumer domain is allowed (by the provider domain) to manage and use the services based on pre-agreed terms and conditions (SLAs). In this case, the consumer TSP uses NFV network services offered by the peer TSP. This may be the case when an end-to-end service is split into constituent services that are deployed in multiple TSP administrative domains.

Federation: is a mechanism for integrating multiple administrative domains at different granularity into a unified open platform where the federated resources and/or services can trust each other at a certain degree [60].

Mobile transport and computing platform (5GT-MTP): a component of the 5G-TRANSFORMER system, see Section A.3.1.3.

Local Repository: database (in an administrative domain) that holds information for available resources for federation, catalogue of services/abstracted services, provided by other provider domains.

Provider domain: administrative domain that offers resources or services to other administrative domains.

Service catalogue: composed set of services and/or service abstractions offered by a provider domain to other potential consumer domains using mutual taxonomy and agreed usage terms (SLAs). The composed service catalogue is shared among pre-agreed peering administrative domains.

Service Orchestrator (5GT-SO): a component of the 5G-TRANSFORMER system, see Section A.3.1.2.

Technology domain: is a collection of resources that are part of a single technology (system) and belong to a single administrative domain. The internal structure is defined and operated according to the technology definitions and standards. One or more technology domains can be part of an administrative domain.

Vertical Slicer (5GT-VS): a component of the 5G-TRANSFORMER system, see Section A.3.1.1.

NOTE: The prefix '5GT' of the acronym is used to avoid a clash with 'VS' standing for vertical service.

A.9 Composed Services

The complexity of vertical services ranges from one or a few VNFs deployed in the same network slice to vertical services composed of several other vertical services (called "child" services), deployed in multiple network slices, where in some cases some of service components are even shared with other vertical services. Such composed services occur in the automotive, entertainment, and eHealth use cases, see Sections A.1.3.1, A.1.3.2, and A.1.3.3. In the entertainment use case, the child services may have quite different characteristics and lifetimes, whereas in the automotive use case the child services may have different owners. Here, we describe how composed vertical services can be handled within the 5G-TRANSFORMER system architecture, starting from Vertical Service Blueprints (VSB) of composed vertical services to support at execution time.

As an abstract example, we consider a composed service consisting of a parent service A and two child services B and C. In Figure 43 two instances of this parent vertical service are depicted (A1 and A2). Each of them uses dedicated instances B1 and B2, respectively, of the child service B and a shared instance C12 of the child service C.

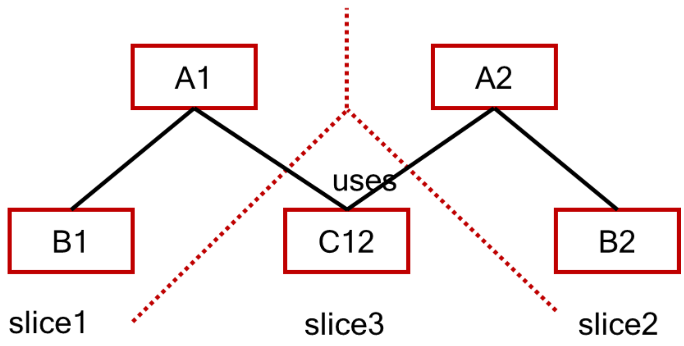


FIGURE 43: EXAMPLE OF COMPOSED VERTICAL SERVICE

The vertical service instances (VSI) are deployed to three network slice instances (NSI), one for the shared VSI C12, one for A1 and B1, and one for A2 and B2. Note, use of a dedicated NSI for the shared VSI C12 is just for illustration purposes, a shared VSI may as well be deployed in an NSSI of a parent VSI. This abstract example fits to the entertainment use cases, e.g. A could be the overall service for sports events, B could be a video distribution service, and C could be a general ticketing service. Similarly, in the automotive case, C could be a car manufacturer independent model of vehicle movements, whereas A could be a manufacturer specific enhanced sensing service and B could be another car manufacture specific service analysing the outcome of the enhanced sensing algorithm.

A.9.1 Vertical Service Blueprints for Composed Services

As defined in D3.1 [5], a Vertical Service Blueprint (VSB) has fields for its atomic functional components, service sequence, connectivity service, external, and internal interconnections. We generalize the atomic functional components to a hierarchy of functional components, where each level includes a list of lower level or of atomic components, and a description of how to connect them consisting again of service sequence, connectivity service, external, and internal connections. The functional components and the forwarding graph among them can thus be described in a VSB similar to nested NFVI-NSs defined in ETSI NFV IFA 014 [30].

Additionally, a functional component may also be a reference to another VSB. This allows to compose a vertical service from several other vertical services.

Note, both the child VSBs and the parent VSB are defined and offered by the 5G-TRANSFORMER service provider (TSP) to the verticals. This is a first step towards vertical-driven service composition; which is described in more detail in D3.3 [6].

Exemplary VSBs for the abstract services A, B, and C are shown in Table 20, Table 21 and Table 22, respectively, focusing on the information for service composition. Additional details on the service sequence and on the service itself would be provided in the blueprint, but have been omitted here for the sake of brevity. Note, for vertical services B and C different requirements on isolation and lifecycle dependencies have been provided.

TABLE 20: VSB OF VERTICAL SERVICE A

Field	Description
Name	A
Description	Abstract service a
Version	1.0
Identity	0123_vsbA
Parameters	n/a
Atomic functional components involved	VNF: Vnf_a Used VSB: B, C

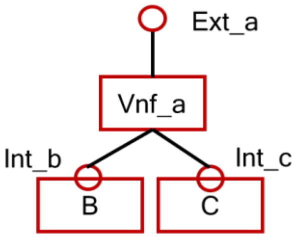
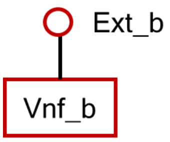
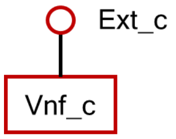
Service sequence	
Connectivity service	n/a
External interconnection	Ext_a
Internal interconnection	Int_b, Int_c

TABLE 21: VSB OF VERTICAL SERVICE B

Field	Description
Name	B
Description	Abstract service B
Version	1.0
Identity	0123_vsbB
Parameters	n/a
Atomic functional components involved	VNF: Vnf_b
Service sequence	
Connectivity service	n/a
External interconnection	Ext_b
Internal interconnection	n/a
Service constraints	• Security: Sharable[17] • Lifecycle independence: same lifecycle as parent • Etc.

[17] An instance of this vertical service can be in the same network slice as its parent.

TABLE 22: VSB OF VERTICAL SERVICE C

Field	Description
Name	C
Description	Abstract service C
Version	1.0
Identity	0123_vsbC
Parameters	n/a
Atomic functional components involved	VNF: Vnf_c
Service sequence	
Connectivity service	n/a
External interconnection	Ext_c
Internal interconnection	n/a
Service constraints	• Security: Isolated[18] • Lifecycle independence: independent • Etc.

A.9.2 Vertical Service Descriptors for Composed Services

In a VSD, a reference to a child VSB is not sufficient. Additionally, it should be possible to refer to another VSD or even a specific VSI. E.g., there could be several instances of a vertical service to support sports event, in this case the corresponding child VSIs for each sports event have to be composed. It has to be possible to refer specific VSIs either by the identifier given by the 5G-TRANSFORMER system, as well as a name provided by a vertical. Using a name allows to refer to a VSI before even an identifier is requested and assigned. This is relevant in cases where parent and child service have different lifecycles. We allow to use the following specific references in a VSD:

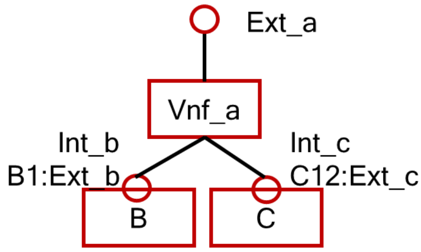
- VSB_name:VSI_name, VSB_Id:VSI_name: These named VSIs are instances of some VSD derived from the given VSB. This kind of references is inherited from VSBs.
- VSD_name:VSI_name/VSI_Id, VSD_Id:VSI_name/VSI_Id: A specific VSI of a specific VSD, both VSI and VSD can be identified either by name or identifier.

[18] An instance of this vertical service has to be deployed in its own network slice.

- VSD_name:*, VSD_ID:*: an arbitrary VSI of a specific VSD. Depending on the requirements on sharing or isolation, a shared VSI may be used or a new VSI has to be created.

In the example, VSB A refers to VSBs B and C, the VSDs refer to specific VSIs already. From the VSB A two VSDs a1 and a2 are derived, referring to specific child vertical services, named VSIs B1, B2, and C12. The relevant fields for a1 are shown in Table 23, here the specific instances are used in the list of functional components and in the service sequence.

TABLE 23: VSD FOR VERTICAL SERVICE A1

Field	Description
Name	A1
Description	Abstract service a1
Version	1.0
Identity	0123_vsda1
Parameters	n/a
Atomic functional components involved	VNF: Vnf_a Uses: B:B1, C:C12
Service sequence	
Connectivity service	n/a
External interconnection	Ext_a
Internal interconnection	Int_b, Int_c

A.9.3 Translation of Composed Services

When translating VSDs of composed vertical services to NSDs, the VSD/NSD Translator module (see Section A.4.2) will translate such a composed VSD to corresponding nested NSDs. This is the most straightforward approach and there is no benefit in either flattening the structuring or even adding additional levels of nesting in the NSDs by the 5GT-VS. If a VSD contains references to VSIs of other VSDs, then the Arbitrator also will make the decisions on mapping several VSIs to one network slice instance or to multiple ones, by adding their translated VSDs to the same NSD or by creating separate NSDs for them.

When a composed VSI is mapped to several NSIs, described by nested NSDs, some child NSDs can be mapped to the same NSI as the parent NSD, whereas others have to be mapped to a separate NSD. These separate NSDs are not referred from the parent NSD, they are NSDs on their own.

An additional NSD has to be created to describe how these NSDs are concatenated, see ETSI NFV IFA 012 [28]. This additional NSD describes communication among VSIs mapped to different network slices.

A.9.4 Instantiation of Composed Services

For the instantiation of composed services, we consider two separate cases. In the first case, the service composition is merely a means to structure the service description, all child services are in the same NSI as its parent and have the same lifetime. The second case is the more generic one with child services in separate NSIs and with different lifetimes.

A.9.4.1 Single Slice, Same Lifecycle

In the most simple case of a composed VSD, the structural information (VNFs and their connections) are provided as a kind of nested NSD. Parent and child NFV-NSs have the same lifecycle and are deployed in the same network slice instance. The VSD/NSD Translator has translated the VSD to a composite NSD including one or more nested NSDs.

Description: Instantiation of a vertical service with a composite service description.

Prerequisites: The vertical has selected a blueprint and prepared a vertical service description from it.

Assumptions: Although the vertical service has a composite description, it can be deployed in one network slice. We also assume that this service is deployed in a new network slice instance. We assume that the NSDs, to which the vertical service is mapped and which describe the network slice have been onboarded to the 5GT-SO before.

Workflow: In general, the instantiation of NFV-NS consists of two steps, firstly to get an identifier for the instance to be created and secondly to create the instance. In the first step, the 5GT-VS requests identifiers for the parent and all child NSDs. These requests can be made in arbitrary order or even in parallel, as an NSD according ETSI NFV IFA 014 [30] references other NSDs in the nestedNsdId field, but not specific instances.

In the second step, the 5GT-VS requests instantiation of the NFV-NSs in a top-down manner. In case location constraints are provided as part of instantiation parameters of the NFV-NSI, constraints on parent level can be considered already for placement decisions. The identifiers of the child NFV-NSIs can be provided as part of the Instantiate NS Operation according to ETSI NFV IFA 013 [29]. These child NFV-NSIs will be deployed to the same NSI as its parent. Siblings can be instantiated in arbitrary order or even in parallel.

This behaviour is described in more detail in the workflow in Figure 44 and Figure 45. The workflow is similar to the instantiation of a non-nested service as described in Section A.7.2. The workflow is triggered by the vertical, which is authenticated and its authorization checked (01, 02).

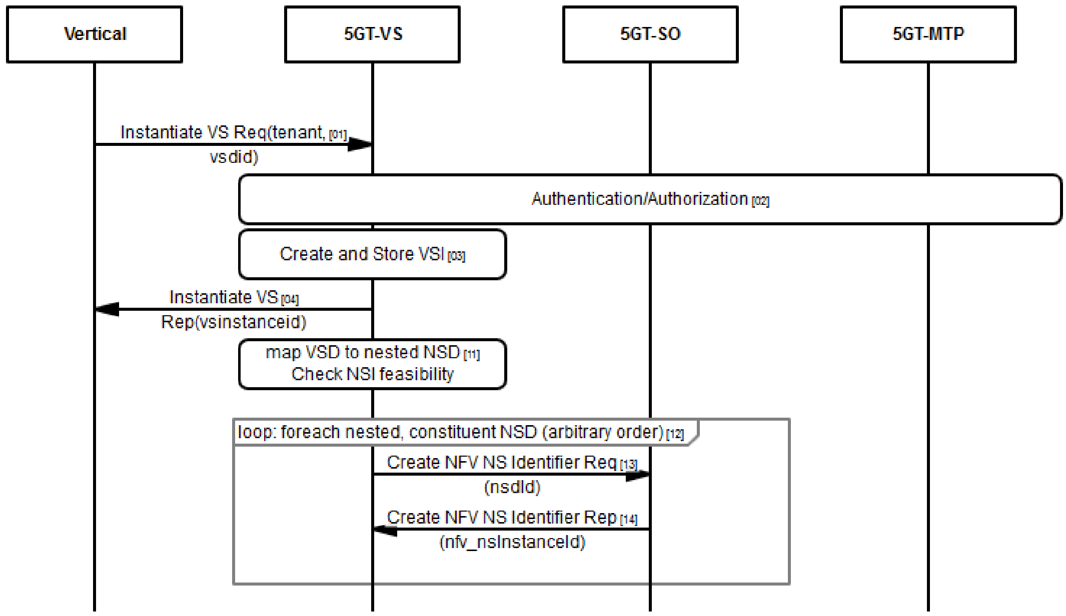


FIGURE 44: WORKFLOW SERVICE INSTANTIATION OF COMPOSED SERVICE, PART 1

The VSD is translated to a set of NSDs, including the composite NSD and one or more nested NSDs (11), and the composite NSD is checked for feasibility against the resource budget of the vertical. For each of the parent and child NSDs an identifier for an instance is requested (12, 13, 14). As described above, the identifiers can be requested in arbitrary order.

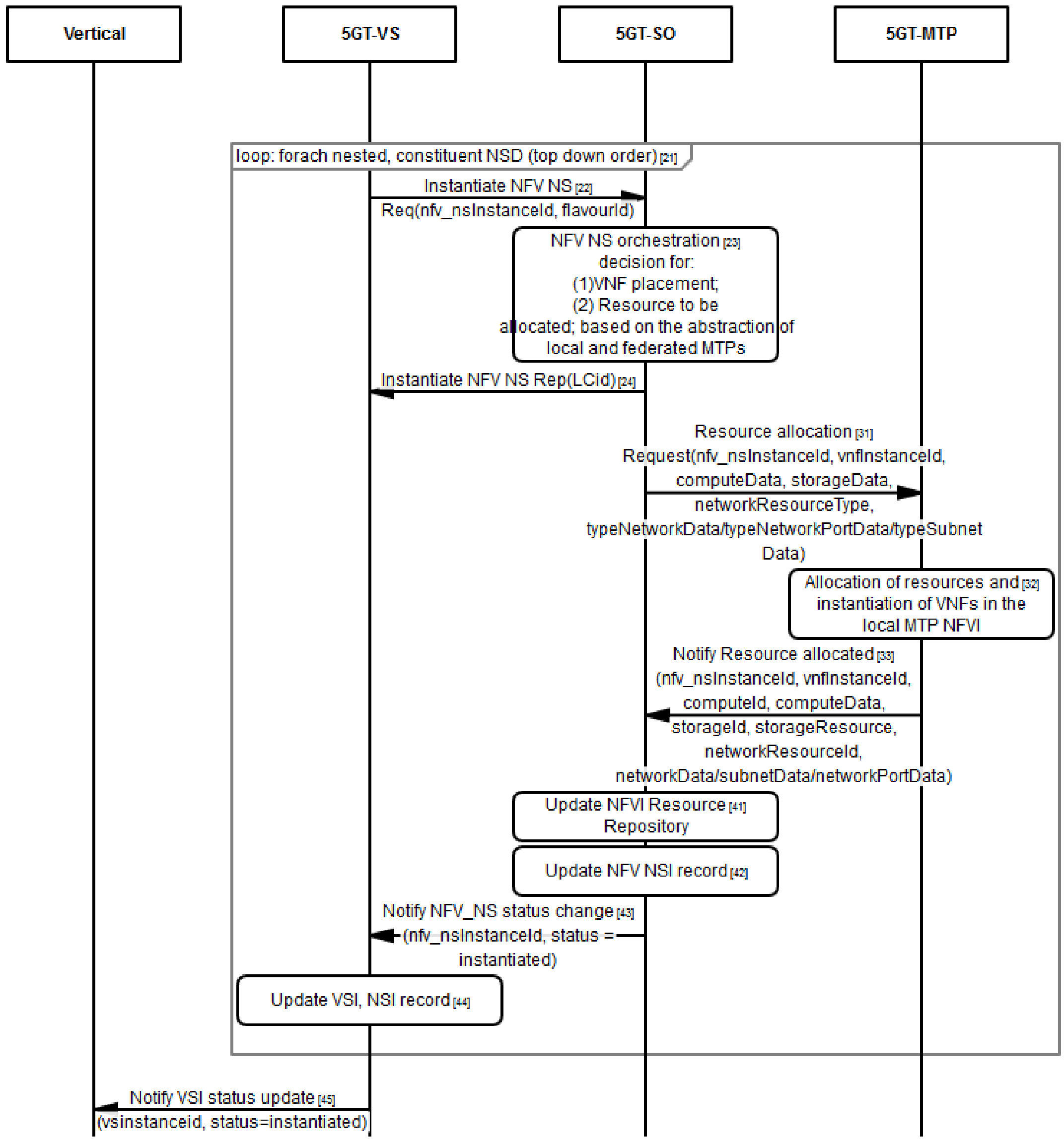


FIGURE 45: WORKFLOW SERVICE INSTANTIATION OF COMPOSED SERVICE, PART 2

Thereafter, the actual instances are created. Here, a top-down order is used to consider high-level placement constraints first. Identifiers of (yet to be instantiated) child NFV NSIs are available from the previous step and can be provided at the instantiation time of the parent. To ensure that all placement decisions can be done, the 5G-VS waits for the notification of the status change before triggering the instantiation of a child NFV NSI. Note, siblings can be instantiated in parallel to reduce the duration of the workflow.

A.9.4.2 Multiple Slice, Different Lifecycle

For this generic case, we consider in our example that VSIs A1, B1, and C12 have been created already and that VSIs A2 and B2 are to be created. A2 and B2 have been mapped to the same network slice, therefore this can be handled as the case described in Section A.9.4.1. It remains to connect the corresponding new NSI with the NSI of C12. This is done by the 5GT-VS updating the NSD describing how different NSIs are concatenated and triggering the corresponding modification at the 5GT-SO.

A.9.5 5GT-SO support for Composed Services

The support of composed services requires corresponding support at runtime such that VNFs deployed to different NSIs can communicate with each other at all and that they can find their communication peers.

A.9.5.1 Application-level Service Registry

Each VA or VNF provides an application service and may require other application services. Therefore, a VA or VNF instance needs to find another VA or VNF instance providing the requested services. MEC already allows an application to register a service, see ETSI MEC 003 [35]. This service registry was intended for infrastructural services such as the radio network information service. But this service registry can be extended to application services. Already the VSB should contain a MEP with such a service registry as a VNF. This service registry can be used by the vertical services mapped to different NSIs to discover each other at runtime.

A.9.5.2 Connecting Network Slices

Different child VSIs of a composed VSIs may be deployed to different NSIs. Nevertheless, traffic is flowing among child VSIs and the corresponding NSIs. When tunnelling techniques, e.g. GTP [56] or VXLAN [57], are used to separate NSIs, then the tunnel header of the source NSI has to be replaced with a tunnel header of the target NSI when a packet is flowing from the source to the target NSI. Such header operations require computational resources and may impact transmission latency, therefore it is recommended that high-bandwidth or low-latency traffic flows are kept within one NSI as much as possible. These header operations could be performed by dedicated VNFs, in this case these VNFs belong to the NSD used to described connectivity among NSIs, see Section A.9.3.

A.10 Federation across 5G-TRANSFORMER Systems

A.10.1 Resource Federation (NFVI-aaS)

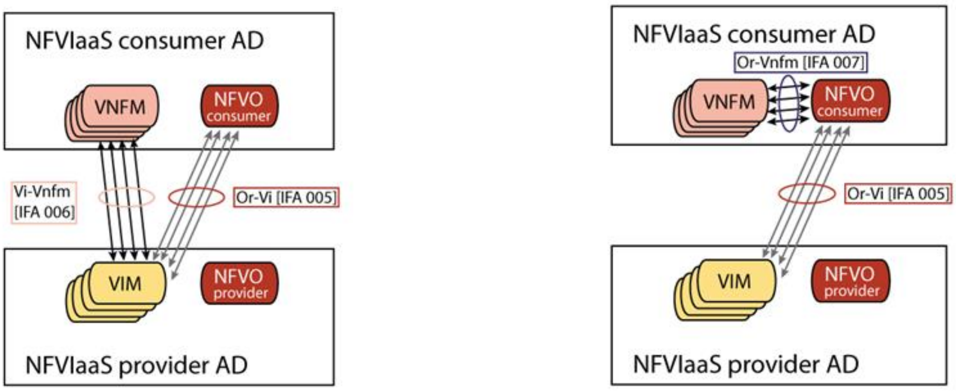
The infrastructure is owned by one party and leased to another one under an NFVI as a Service (NFVIaaS) model, wherein Virtual Network Functions (VNFs) can be remotely deployed and run inside the NFVI provided as a service. In an NFVIaaS model, we distinguish two actors that belong to different administrative domains (ADs), namely; *NFVIaaS provider* and *NFVIaaS consumer.* The NFVIaaS provider is responsible for resource control in his AD; it offers interfaces for resource management requests from NFVIaaS consumer, selects NFVI resources according to NFVIaaS consumer's request (considering the placement decision and SLA), manages NFVI resources (including resource management, reservation, quota management, fault management, and performance management), provides an overall view of NFVI resources (such as, capabilities and monitoring) and, lastly, manages software images for VNFs based on NFVIaaS consumer's inputs. The *NFVIaaS consumer* is in charge of VNFs and NFV-NS control in its AD and requests resources to the NFVIaaS provider to run its VNFs and map them to NFV-NSs. The NFVIaaS consumer manages the lifecycle of VNFs and NFV-NSs (including VNF package management, NSDs, granting of VNF's LCM operations), requests resources from NFVIaaS provider, gets information from NFVIaaS provider about NFVI resources (e.g., capacity information, usage information, and performance information) and, finally, distributes SW images to the NFVIaaS provider for VNFs that shall run on its NFVI.

In general, there are four interworking options in which NFVIaaS provider offers interfaces to the NFVIaaS consumers:

- Access via Multiple Logical Point of Contact (MLPOC): In MLPOC, the NFVIaaS consumer has visibility of the NFVIaaS provider's VIMs and may need to interface with different VIM implementations and versions. Regarding VNF management, there are two options:
    1. VNF-related resource management in *direct* mode.
    2. VNF-related resource management in *indirect* mode.
- Access via Single Logical Point of Contact (SLPOC): In SLPOC, the NFVIaaS provider's VIMs are hidden from the NFVIaaS consumer. Unified interfaces are exposed by the SLPOC and offered to the NFVI consumer. Similarly, to MLPOC, the VNF management has two variants:
    1. VNF-related resource management in *direct* mode.
    2. VNF-related resource management in *indirect* mode.

A.10.1.1 MLPOC: Multiple Logical Point of Contact

Figure 46 describes the direct and indirect modes of MLPOC regarding the VNF management. In the direct mode the VNFM in the NFVIaaS consumer AD invokes virtualized resource management operations on the VIM(s) in the NFVIaaS provider AD, (Figure 46a). In the indirect mode the VNFM in the NFVIaaS consumer AD invokes virtualized resource management operations on the NFVO-consumer in the NFVIaaS consumer AD, which in turn invokes them towards the VIM(s) in the NFVIaaS provider AD, (Figure 46b). The two figures highlight an NFVIaaS provider that allows access to MLPOC in its AD, while the NFVIaaS consumer issues NFVIaaS service requests using interfaces provided by the VIMs. It is assumed that VIMs provide logical points of contact for the virtualized resource management requests through existing interfaces (*Or-VI [NFV IFA005]* and *Vi-Vnfm [NFV IFA006]*), and both the NFVIaaS provider and NFVIaaS consumer have a business relationship, and exchange information about infrastructure tenants, resource groups, and access to VIMs.



a. Direct mode **b. Indirect mode**

FIGURE 46: NFVIAAS FEDERATION (MLPOC)

Table 24 compares the two variants of NFVIaaS architecture with MLPOC access. A discussion of the options preferred for the 5G-TRANSFORMER architecture is given in Section A.3.1.8.1.

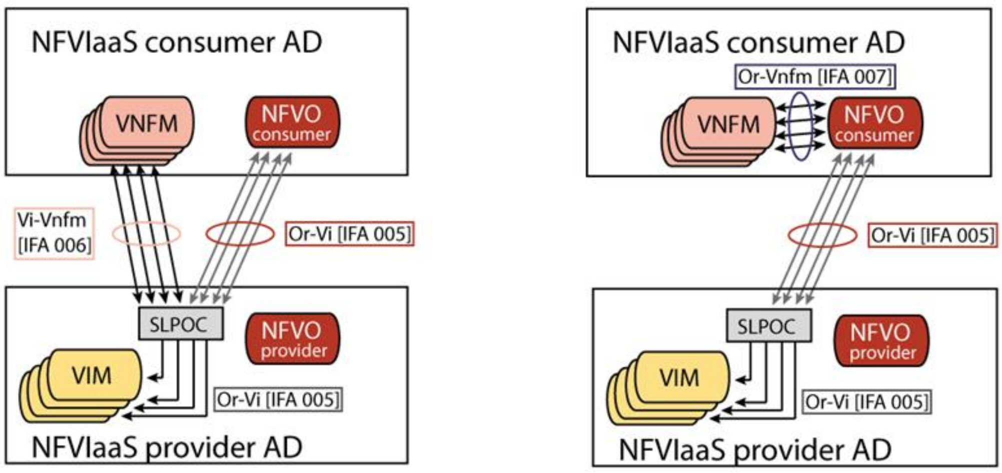
TABLE 24: NFVIAAS ARCHITECTURE WITH (DIRECT VERSUS INDIRECT) VNF MANAGEMENT FOR MLPOC ACCESS

	Direct mode	Indirect mode
NFV NS management	- NFVO-consumer is responsible for the NFV NS's LCM (including, NSD management, and VNF packages). - NFVO-consumer manages VNF-FGs and VLs for NFV NSs. - NFVO-consumer issues network resource management operations towards respective VIMs.	
VNF management	- VNFM requests resource management needed for VNF's LCM from the identified VIM(s).	- VNFM requests resource management needed for VNF LCM from the NFVO-consumer, which issues the requests towards the VIM(s).
	- VNFMs are responsible for VNF's LCM. - Before computing VNF lifecycle operation, VNFM requests an operation granting from NFVO-consumer. - NFVO-consumer collects information on consumable resource and virtualized resources capacity from VIMs. - It uses such information to identify and select target VIM(s) from which virtualized resources will be provided for VNF. - NFVO-consumer maintains and enforces permitted allowance at various granularity levels (VNFM, VNF, NS, etc.) to control resource consumption by VNFMs in relation with VNF lifecycle operation granting. - NFVO-consumer cannot guarantee resource availability during the granting of a VNF lifecycle request if the resource needed to accommodate such lifecycle operation have not been reserved in the VIM. - NFVO-consumer performs VNF package management and distributes the SW images of VNFs to the VIM(s) on which they will be deployed.	

<table>
<tr><td rowspan="2">Virtualized resource management</td><td>- VIMs are responsible for the virtualized resource management and providing interfaces to VNFMs and NFVO-consumer.</td><td>- VIMs are responsible for the virtualized resource management and providing interfaces to NFVO-consumer.</td></tr>
<tr><td colspan="2">- VIMs manage infrastructure tenants and infrastructure resource groups, and limit the scope of operations to the requesting infrastructure tenant to:
▪ Get only information related to infrastructure resource groups assigned to the tenant.
▪ May only initiate virtualized resource management related to infrastructure resource groups assigned to the tenant.
▪ May only request quota related to infrastructure resource groups assigned to this tenant.
▪ May only reserve virtualized resource belonging to infrastructure resource groups assigned to the tenant.</td></tr>
</table>

A.10.1.2 SLPOC: Single Logical Point of Contact

Figure 47 describes the NFVIaaS architecture with a SLPOC. Figure 47a describes the architecture in which the VNF related-resource management is done in direct mode, that is, the VNFM invokes virtualized resource management operations on the SLPOC, while Figure 47b depicts the indirect management mode of VNFs, wherein the VNFM invokes virtualized resource management operations on the NFVO-consumer, which in turn invokes them towards the SLPOC. The two figures depict the NFVIaaS consumer which is allowed to issue NFVIaaS service requests and access to the NFVIaaS provider's AD through the SLPOC, which provides interfaces (Vi-Vnfm[IFA 006] and Or-Vi[IFA 005]). The NFVIaaS provider hides its VIMs from the NFVIaaS consumer. However, unified interfaces are exposed by the SLPOC and offered to the NFVIaaS consumer. It is assumed that the NFVIaaS provider and the NFVIaaS consumer have a business relationship to exchange information regarding the infrastructure, tenants, resource groups, and access to the SLPOC.



a. Direct Mode **b. Indirect Mode**

FIGURE 47: NFVIAAS FEDERATION (SLPOC)

Table 25 compares between the two variants of NFVIaaS architecture with the SLPOC access.

TABLE 25: NFVIAAS ARCHITECTURE WITH (DIRECT VERSUS INDIRECT) VNF MANAGEMENT FOR SLPOC ACCESS

<table>
<tr><th></th><th>Direct mode</th><th>Indirect mode</th></tr>
<tr><td>NFV NS management</td><td colspan="2">- NFVO-consumer is responsible for the NFV NS's LCM (including, NSD management, and VNF packages).
- NFVO-consumer manages VNF-FGs and VLs for NFV NSs.
- NFVO-consumer issues network resource management operations towards SLPOC.</td></tr>
<tr><td rowspan="2">VNF management</td><td>- VNFM requests resource management needed for VNF's LCM from the identified SLPOC.</td><td>- VNFM requests resource management needed for VNF LCM from the NFVO-consumer, which issues the requests towards the SLPOC.</td></tr>
<tr><td colspan="2">- VNFMs are responsible for VNF's LCM.
- Before computing VNF lifecycle operation, VNFM requests an operation granting from NFVO-consumers.
- NFVO-consumer collects information on consumable resource and virtualized resources capacity from the SLPOC.
- It uses such information for the VNF's LCM decisions.
- NFVO-consumer maintains and enforces permitted allowance at various granularity levels (VNFM, VNF, NFV NS, etc.) to control resource consumption by VNFMs in relation with VNF lifecycle operation granting.
- NFVO-consumer cannot guarantee resource availability during the granting of a VNF lifecycle request if the resource needed to accommodate such lifecycle operation have not been reserved by the SLPOC.
- NFVO-consumer performs VNF package management and distributes the SW images of VNFs to the SLPOC, which forwards</td></tr>
</table>

	them to the VIMs.	
Virtualized resource management	- VIMs are responsible for the virtualized resource management and providing interfaces to VNFMs and SLPOC, which interfaces with the VNFM and NFVO-consumer.	- VIMs are responsible for the virtualized resource management and providing interfaces to VNFMs and SLPOC, which interfaces only with the NFVO-consumer.
	- SLPOC hides the VIM interfaces. - SLPOC maintains information about the infrastructure resource organization, availability, and utilization from various VIMs in the infrastructure domain. - All virtualized resource management requests from the NFVIaaS consumer go to the SLPOC, which forwards them to the VIM(s). - Existing interfaces Or-Vi [IFA 005] reference point can be reused for interfaces between SLPOC and VIMs. - SLPOC manages infrastructure tenants and infrastructure resource groups, and limit the scope of operations to the requesting infrastructure tenant to: ▪ Gets only information related to infrastructure resource groups assigned to the tenant. ▪ May only initiate virtualized resource management related to infra. Resource groups assigned to the tenant. ▪ May only request quota related to infra resource groups assigned to this tenant. May only reserve virtualized resource belonging to infra resource groups assigned to the tenant.	

To sum up, the MLPOC and SLPOC look like VIM(s) from the NFVIaaS consumer's NFVO and VNFMs view. For the architecture option using NFVIaaS provider's VIMs with MLPOC access, no new reference point is needed to integrate it into MANO functional blocks. Regarding the architecture options using SLPOC functionality, two options are possible: the SLPOC may be integrated into a VIM or into the NFVO of the NFVIaaS provider, which introduces new reference points and reuses the IFA005/IFA006 interfaces. As described in Section A.3.1.8, the SLPOC solution integrated into the NFVO is chosen for the 5G-TRANSFORMER architectural solution. These options enable the interaction for both NSaaS and NFVIaaS to remain on the same 5GT-SO level.

Regarding the different options, when the SLPOC is integrated into a VIM, this latter needs to support the SLPOC functionality and interfaces to: (i) forward requests from NFVIaaS consumer's NFVO and VNFM(s) to the respective VIMs; (ii) maintain an overall view over virtualized resources, quotas, reservations, and capacity per infrastructure resource group; (iii) connect between different VIMs, SLPOC, NFVO, and VNFMs in the NFVIaaS Provider's AD using IFA005 and IFA006 interfaces.

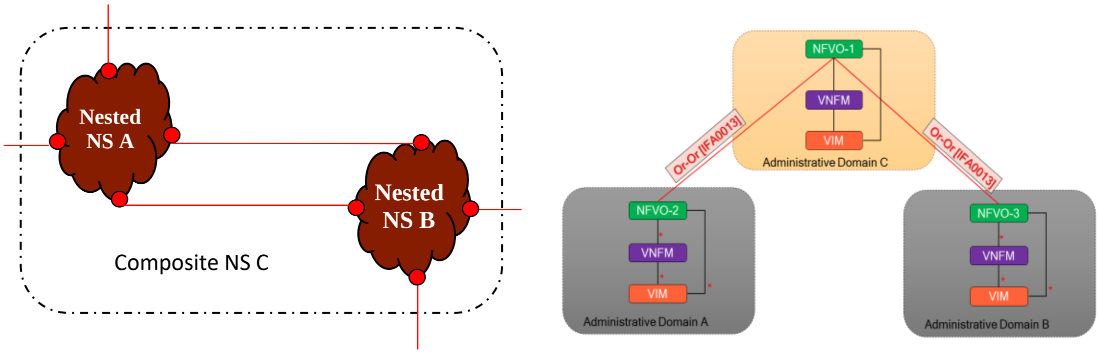
While, with the solution of the SLPOC is integrated into the NFVO inside the NFVIaaS provider's AD, the NFVO needs to support the SLPOC functionality and interfaces to: (i) manage the infrastructure resource groups and tenants; (ii) provide IFA005 interfaces towards the NFVIaaS consumer's NFVO and forward the requests to the respective VIM(s); (iii) provide IFA006 interfaces towards the NFVIaaS consumer's VNFM(s) and forward the requests to the respective VIM(s); (iv) limit the scope of operations to the infrastructure resource groups assigned to the requesting infrastructure tenants; (v) finally, maintain an overview of the virtualized resources, quotas, reservations, and capacity per infrastructure resource group.

A.10.2 Service Federation (NSaaS)

The NSaaS describes the use case of network services (NFV NS) provided by a network operator to different departments within the same operator, as well as to other network operators. Each administrative domain is seen as one or more NFVI-PoPs, VIMs, and VNFMs together with their related VNFs. The NFVO in each AD allows distinct specific set of NFV NSs, which are hosted and offered by each AD.

To better understand this use case, we rely on an example of composite NFV NS and nested NFV NSs depicted in Figure 48a. In this example, the composite NFV NS C is built from the two nested NFV NS A and B. The nested NFV NSs A and B, and the composite NFV NS C are provided by distinct ADs A, B, and C, respectively. For the management of these NFV NSs, a hierarchical architecture is shown in Figure 48b, which provides an example of multiple ADs, each one offering a set of NFV NSs. The NFVO-1 in the AD C is managing the composite NFV NS C, while the NFVO-2 (NFVO-3, respectively) in the AD A (B, respectively) manages the nested NFV NSs A (B, respectively) and exposes the nested NFV NSs to the NFVO-1. An SLA between the providers of the nested NFVs NS A or B and the provider of the composite NFV NS C should be negotiated, and to check if this SLA is met, monitoring resource usage by the nested NSs is needed. Each NFVO is responsible for a set of actions according to the NFV NS that is provided by its AD.

The NFVO-1 (for AD C) has to perform the instantiation of the composite NFV NS, and if needed, to trigger the instantiation of the nested NFV NSs A and B corresponding to NFVO-2 and NFVO-3. The NFVO-1 is also responsible for the LCM of other operations, including scaling and healing of the nested NFV NSs, in collaboration with NFVO-2 and NFVO-3. Note that the NFVO-1 is not aware of the virtualized resources in the ADs A and B. NFVO-2 and NFVO-3 provide NFVO functionalities for the nested NFV NSs. They receive the NFV NS LCM request from NFVO-1 and provide NFV NSs LCM for the nested NFV NSs. Besides managing NFV NS tenants and service groups and their association for the nested NFV NSs, NFVO-2 and NFVO-3 provide the NFVO-1 with information about the nested NSs (including, fault information, performance, and capacity information) belonging to service resource groups assigned to the related NS tenants, in order to limit the scope of operations to the requesting NS tenant.



a. Composite NS and nested NSs b. Composite NS and nested NSs

FIGURE 48: NSAAS USE CASE

To support this use case of NSaaS in the NFV-MANO architecture framework, a new reference point between NFVO-2 and NFVO-1, and between NFVO-3 and NFVO-1 called *Or-Or* based on *IFA013* is added to the NFV-MANO framework to support the NFV NS LCM actions provided through several ADs. In each AD, VNFMs interact with the NFVO on the same AD. Therefore, the reference point *Or-Vnfm* keeps unchanged in each AD. None of the NFVO-1, NFVO-2, and NFVO-3 is aware of the constituent VNF instances of the NS outside its AD. Therefore, none of these NFVOs interact with a VNFM from another AD.

A.10.3 Service Delegation vs. Service Federation

For implementation of the 5G-TRANSFORMER system, two different terms are used for federation of services: service delegation and service federation

The service delegation is a procedure where an NFV-NS is simply delegated to be instantiated in an external domain, without performing any modification of the request. In more details, the constituent 5GT-SO is receiving a request from the 5GT-VS for instantiating a NFV-NS. The requested NFV-NS is not a composite NFV-NS and it cannot be decomposed to several nested NFV-NSs. In the case that the 5GT-SO is not able to instantiate the requested NFV-NS over the local 5GT-MTP resources (e.g., not enough resource capacity, incompatible resources for instantiation the requested NFV-NS, etc.), the request can be redirected as it is to an external 5GT-SO. The constituent 5GT-SO would redirect the instantiation request without any modification to a provider peering 5GT-SO. The constituent 5GT-SO is acting like a proxy towards the peering administrative domain. The external provider 5GT-SO instantiates the requested NFV-NS and returns the service information towards the consumer/constituent 5GT-SO. As for redirecting the instantiation request, the constituent 5GT-SO acts as a proxy and redirects the service information towards the 5GT-VS. The instantiated service at the external domain, using the 5GT-SO as a proxy, is called a delegated NFV-NS. For the life-cycle duration of the delegated NFV-NS, the consumer/constituent 5GT-SO acts like a proxy and redirects bi-directional traffic from the 5GT-VS to the provider 5GT-SO and vice-versa.

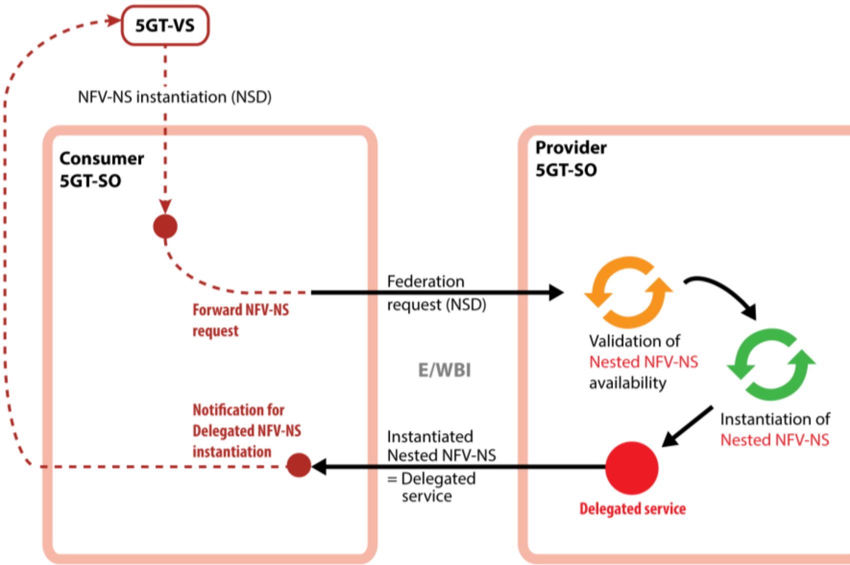


FIGURE 49: SERVICE DELEGATION

The service federation is a procedure of instantiating composite NFV-NS distributed in multiple administrative domains. Service federation occurs only for composite NFV-NS. As first step after receiving a request for instantiation of a composite NFV-NS, the 5GT-SO decomposes the composite NFV-NS into several nested NFV-NSs. Each nested-NFV-NS can be instantiated in local or external domain. The criteria for instantiating a nested-NS in an external domain can be pre-determined (static) or decided on instantiation time for various reasons (e.g., lack of local 5GT-MTP resources, geo-location, reduced latency, etc.). Once the 5GT-SO decomposes the composite NFV-NS, it issues requests for certain nested NFV-NSs to be instantiated to an external provider 5GT-SO (in other domain). The provider 5GT-SO receives the request and proceeds with instantiation procedure in same way as instantiating any single NFV-NS. Once the nested NFV-NS is instantiated, the provider 5GT-SO notifies the consumer 5GT-SO of the instantiated service. The service information (e.g., connection points, IP addresses, etc.) are obtained by the consumer 5GT-SO using a polling procedure. Finally, the consumer 5GT-SO performs “stitching” or chaining of the instantiated nested NFV-NS in both local and external domains, in order to form the composite/end-to-end (E2E) NFV-NS.

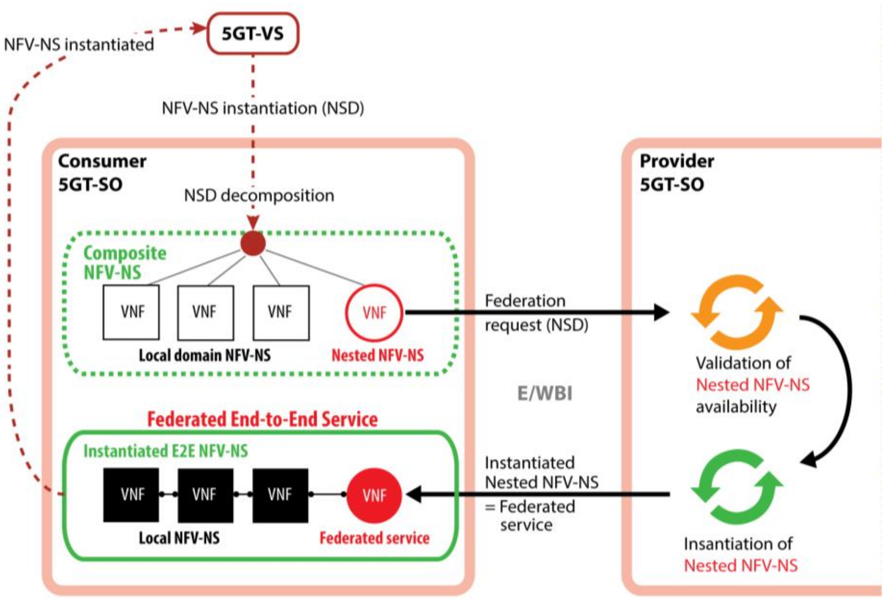


FIGURE 50: SERVICE FEDERATION