Deliverable D1.1

Design of infrastructure architecture and subsystems v1

Editor:	Dan Warren, Samsung
Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	31st December 2018
Actual delivery date:	
Suggested readers:	ICT-19 projects, industry verticals seeking to conduct Testing on 5G systems
Version:	1.0
Total number of pages:	81
Keywords:	5G, Architecture,

Abstract

To allow 5G-VINNI Facility Sites to implement their networks with some degree of consistency, and with the potential for inter-operability and interconnection between Facility Sites, a common architecture for 5G-VINNI is defined. This draws heavily on pre-existing work from Standards, as well as from previous 5G-PPP projects. This document is produced ahead of implementation, and so is to some extent a description of the expectations for implementing a standards compliant facility. Future versions will reflect this as a starting point, but also include lessons learned during implementation.

[End of abstract]

Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815279.

Impressum

Full project title	5G Verticals Innovation Infrastructure
Project acronym	5G-VINNI
Number and title of work-package	WP1 Architecture and Design of 5G-VINNI End-to-End Platform
Number and title of task(s)	T1.1 Infrastructure architecture and subsystems design
Document title	Design of infrastructure architecture and subsystems v1
Editor: Name, company	Dan Warren, Samsung
Work-package leader: Name, company	Dan Warren, Samsung

Copyright notice

Executive summary

5G-VINNI is built on the concept of a common architectural design that is then adapted to the specific use cases that are to be implemented on each Facility Site. The intent is to provide enough consistency and stability within 5G-VINNI Facility Sites to enable ICT-19 projects and other vertical industry proposals to be able to demonstrate their own capabilities, without the need to worry about whether the underlying 5G network will be reliable, or be changing as a result of its own experimental nature.

As with previous generations, 5G relies upon strong definitive standardisation of architecture and protocols. Ultimately these serve to underpin mass market adoption across the world, and foster a global ecosystem of vendors, operators and device manufacturers, that as a consequence enable economy of scale in all industry components and facilitate worldwide roaming and interconnect. This will develop over time with 5G, but at this point the 5G ecosystem is in its very earliest stage. However, with stability in Standards groups, the ability to build a standards-based test facility already exists.

This document sets the basis of that ambition in 5G-VINNI - to define the architecture that all 5G-VINNI Facility Sites will adhere to. The document seeks to identify the areas where full compliance is required, as well as the optional capabilities that a Facility Site may choose to employ for specific purposes. This deliverable represents a starting point for the technical design work on the Facility Sites themselves (WP2), the service implementation definitions (WP3) and the testing and monitoring of the networks that are built (WP4). Because the document is drafted in advance of the networks being implemented, it is some ways an exercise of understanding standards that are already available, and then considering the potential for practical implementation. It cannot, however, reflect the lessons that will undoubtedly be learned during implementation, which will result in practical knowledge, and where major issues are discovered, may provide feedback and result in modification of the Standards the design is based upon. A second version of this document (D1.4) will be produced later in 5G-VINNI's timeline, which will be able to reflect the updated architecture as the implementation of Facility Sites progresses.

List of authors

Company	Author
Samsung	Dan Warren
Simula	Ahmed Elmokashfi
Keysight	Andrea F. Cattoni
Telenor	Andres J. Gonzalez
Huawei	Artur Hecker
Altice Labs	Carlos Parada
SES	Christos Politis
University of Patras	Christos Tranoris
Athens University	Costas Kalogiros
Telefonica	Diego R. Lopez
Intracom Telecom	Dimitrios Kritharidis
Nokia	Duncan Silvey
Simula	Foivos Michelinakis
Athens University	George Darzanos
Athens University	George Stamoulis
Telenor	Håkon Lønsethagen
Huawei	Ishan Vaishnavi
Nokia	Joao Rodrigues
Altice Labs	Jorge Carapinha
Telefonica	Jose Ordonez-Lucena
Huawei	Joseph Eichinger

Telenor	Kashif Mahmood
Intracom Telecom	Konstantinos Chartsias
Intracom Telecom	Konstantinos Katsaros
SES	Konstantinos Liolis
Intracom Telecom	Konstantinos Stamatis
Huawei	Mohammed Gharba
Huawei	Osama Abboud
Universidad Carlos III de Madrid	Pablo Serrano
Telenor	Pål Grønsund
University of Patras	Panagiotis Papaioannou
Ericsson	Puneet Gupta
Huawei	Ramin Khalili
Huawei	Riccardo Trvisonno
Ericsson	Robert Lagerholm
Nokia	Shane Jackson
Nokia	Tom-Kristian Berg
Intracom Telecom	Vasileios Theodorou
Altice Labs	Victor Marques
Huawei	Wint Yi Poe

Table of Contents

List of figures

List of tables

Abbreviations

3GPP	3rd Generation Partnership Project
5G	Fifth Generation (mobile/cellular networks)
5G PPP	5G Infrastructure Public Private Partnership
5G-EIR	5G Equipment Identity Register
AF	Application Function
AKA	Authentication and Key Agreement
AMF	Access and Mobility Management Function
API	Application Programming Interface
ARPU	Average Revenue Per User
AUSF	Authentication Server Function
BGP	Border Gateway Protocol
BBU	Based Band Unit
BSS	Business Supporting Systems
CapEx	Capital Expenditure
cMTC	Critical Machine-Type Communications
CP	Control Plane
CPE	Customer Premises Equipment
CPRI	Common Public Radio Interface
CRM	Customer Relationship Management
CSC	Communication Service Customer
CSP	Communication Service Provider
CU	Centralised Unit
DCSP	Data Centre Service Provider
DÉCOR	Dedicated Core Network
DN	Data Network
DSRC	Dedicated Short Range Communications
DU	Distributed Unit
E2E	End-to-End
EAP	Extensible Authentication Protocol
eCPRI	Enhanced Common Public Radio Interface
eMBB	Enhanced Mobile Broadband
EMS	Element Management System
eNB	enhanced NodeB
EPC	Evolved Packet Core
EPG	Evolved Packet Gateway
E-UTRA	Evolved Universal Terrestrial Radio Access

FCAPS	Fault, Configuration, Accounting, Performance and Security
FEC	Forwarding Equivalence Class
FWA	Fixed Wireless Access
gNB	next generation NodeB
HPLMN	Home Public Land Mobile Network
HSS	Home Subscriber Server
IMSI	International Mobile Subscriber Identity
IoT	Internet of Things
IPSec	Internet Protocol Security
ITS	Intelligent Transportation Systems
KPI	Key Performance Indicator
LAN	Local Area Network
LSO	Lifecycle Service Orchestration
LTE	Long Term Evolution
LTE-M	LTE for Machines
MANO	Management and Network Orchestration
MEC	Multi-access Edge Computing
MEF	Metro-Ethernet Forum
MeNB	Master eNodeB
MIMO	Multi-Input Multi-Output
mIoT	Massive Internet of Things
MME	Mobility Management Entity
mMTC	Massive Machine-Type Communications
MN	Master Node
MNO	Mobile Network Operator
MP	Management Plane
MPLS	Multi-Protocol Label Switching
N3IWF	Non-3GPP Interworking Function
NAS	Non-Access Stratum
NB-IoT	Narrow Band Internet of Things
NEF	Network Exposure Function
NEP	Network Equipment Provider
NF	Network Function
NFP	Network Forwarding Path
NFVI	Network Function Virtualisation Infrastructure
NFVO	NFV Orchestrator
NGC	Next Generation Core
NMS	Network Management System

NOP	Network Operator
NR	New RAT, New Radio
NRF	Network Repository Function
NS FM	Network Service Fault Management
NS LCM	Network Service Life Cycle Management
NS PM	Network Service Performance Management
NSA	Non-Stand Alone
NSI	Network Slice Instance
NSSAI	Network Slice Selection Assistance Information
NSSF	Network Slice Selection Function
NSSI	Network Sub-Slice Instance
NWDAF	Networks Data Analytics Function
ONAP	Open Network Automation Platform
OpEx	Operational Expenditure
OSS	Operations Support System
PCF	Policy Control Function
PCRF	Policy and Charging Rules Function
PDU	Packet Data Unit
PGW	Packet Gateway
PLMN	Public Land Mobile Network
PNF	Physical Network Function
PON	Passive Optical Network
PoP	Point of Presence
PtMP	Point-to-Multi-Point
PtP	Point-to-Point
QAM	Quadrature Amplitude Multiplex
QCI	QoS Class Indicator
QoE	Quality of Experience
QoS	Quality of Service
R&D	Research and Development
RAB	Radio Access Bearer
RAM	Resource Advertising Management
RAN	Radio Access Network
RAT	Radio Access Technology
RLF	Radio Link Failure
RMM	Resource Management Monitoring
RRC	Radio Resource Control
RU	Remote Unit

SA	Stand Alone
SBA	Service Based Architecture
SCS	Sub-carrier Spacing
SDN	Software Defined Networks
SDO	Standards Definition Organisation
SEPP	Security Edge Protection Proxy
SES	Satellite Earth Stations and Systems
SFC	Service Function Chaining
SgNB	Secondary gNodeB
SGW	Serving Gateway
SLA	Service Level Agreement
SMF	Session Management Function
SN	Secondary Node
SUPI	Subscriber Permanent Identifier
TDD	Time Division Duplex
TLS	Transport Layer Security
TMF	TeleManagement Forum
TR	Technical Report
TS	Technical Specification
UDM	Unified Data Management
UDR	Unified Data Repository
UDSF	Unstructured Data Storage Function
UE	User Equipment
UHD	Ultra-High Definition
UICC	Universal Integrated Circuit Card
UP	User Plane
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communications
V2X	Vehicle to Everything
VAC	Value Added Connectivity
VIM	Virtualised Infrastructure Manager
VISP	Virtualisation Infrastructure Service Provider
VLD	Virtual Link Descriptor
VNF	Virtualised Network Function
VNFC	VNF Components
VNFD	VNF Descriptor
VNFFG	VNF Forwarding Graph
VNFFGD	VNFFG Descriptor

WAN	Wide Area Network
WIM	WAN Infrastructure Management
WLAN	Wireless Local Area Network
WP	Work Package
ZSM	Zero-touch network and Service Management

Definitions

This document contains specific terms to identify elements and functions that are considered to be mandatory, strongly recommended or optional. These terms have been adopted for use similar to that in IETF RFC2119, and have the following definitions.

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

Terminology used within the 5G-VINNI project draws a separation between ‘5G-VINNI Facility’ and ‘5G-VINNI Facility Site’. These terms are used as follows:-

- **5G-VINNI Facility** – this is the total of all parts of the 5G-VINNI network, comprising of the individual 5G-VINNI Facility Sites, and links that are used to interconnect between 5G-VINNI Facility Sites.
- **5G-VINNI Facility Site** – this is the infrastructure and capability built at a single location, associated with 5G-VINNI. This is limited to the infrastructure built under the control of a single 5G-VINNI operational partner.

1 Introduction

This document provides an architectural definition for a generalised individual 5G-VINNI site and the overall 5G- VINNI facility, based on requirements for support of the 5G-VINNI Release 0, 1, 2 and 3 implementations. The definition of each Release of the 5G-VINNI Facility is given in Chapter 2.

The document is structured as follows:-

- Chapter 2 identifies architectural principles that pervade throughout the 5G-VINNI concept and hence, form the basis for much of the definition of the 5G-VINNI Facility and each of the individual sites.
- Chapter 3 captures key aspects of pre-existing work that 5G-VINNI draws upon. This is greatly done by reference to 3GPP standards, but also takes into account other SDOs, forums and previous 5G-PPP work in projects and working groups. This is summarised to allow further definition for 5G-VINNI to build upon this work.
- Chapter 4 contains Top level architectures for a facility site, testing and monitoring and for Management and Orchestration. The subsections within the chapter provide detailed definition of a generalised 5G-VINNI Facility Site. Each Facility Site is expected to support all mandatory aspects of the definition whilst additional optional aspects may or may not be supported, dependent upon the capabilities and use cases that the Facility Site is intending to test. Multiple aspects of a 5G-VINNI Facility Site implementation are covered, these being End-User Devices, RAN, Core, Core to RAN interfaces, Transport, Management and Orchestration, External-facing interfaces, Security and Testing and Monitoring.
- Chapter 5 addresses topics around interconnection and interoperability between one 5G-VINNI Facility Site and another. This is of particular importance as the project moves towards Releases 1, 2 and 3 of the 5G-VINNI Facility where increasing levels of interconnectivity between Facility Sites are planned.
- Chapter 6 covers Research topics associated with Architectural aspects of 5G-VINNI.

The document forms the starting point of much of the work that is to be undertaken in the remainder of Work Package 1 and in the other Work Packages of the 5G-VINNI project. Throughout the document, forward reference is made to the deliverables, tasks and work packages that make up the project.

Direction is given to the evolution of the 5G-VINNI facility through the definition of a number of 'releases' of the end-to-end (E2E) architecture.

- Release 0 of E2E facility (Month 12, or M12) - For 5G-VINNI project internal validation of KPIs and specific use cases for E2E facility validation. Facility will consist of Non Stand Alone (NSA) 5G New Radio (NR) and 5G Core. Virtualization Infrastructure, NFV Orchestration and Service Orchestration will be implemented. E2E Slicing is implemented supporting basic life-cycle events.
- Release 1 of E2E facility (M18) - Ready for use by ICT-18-19-22 projects and other external use cases. The main facility sites (Norway, UK, Spain and Greece) will be 3GPP Rel15 compliant. At minimum one of the facility sites will include Stand Alone (SA) 5G NR and 5G Core. E2E slicing is implemented supporting all planned life-cycle events. Service orchestration across two interconnected main facility sites.
- Release 2 of E2E facility (M24) - Backward compatible with Release 1. Two main facility sites will include SA 5G NR and 5G Core. Service Orchestration across 3 interconnected main facility sites. Minimum 2 vertical use cases from ICT-18-19-22 project(s) and 3 use cases from other external verticals "customers" of the 5G-VINNI facilities are on-boarded and running on the 5G-VINNI facility.
- Release 3 of E2E facility (M30) - Backward compatible with Release 2. All main facility sites will include SA 5G NR and 5G Core. Service Orchestration across all interconnected main facility sites. Initial results from vertical use cases and KPIs validation and testing.

The progressive shift of each facility site from Non-Stand Alone (NSA) to Stand Alone (SA) architecture is a reflection on the timing of the implementation of 5G-VINNI and the relative maturity of technology that is available. 5G implementation is the goal of 5G networks as they are implemented, but in the initial stages, a 5G NSA implementation is expected. NSA architecture refers to the implementation of 5G-NR RAN tied to an Evolved Packet Core (EPC) from LTE and LTE-Advanced network implementation. This implementation is described fully in 3GPP TS23.501 [1], and expanded upon in sections 4.2.2 (RAN aspects) and 4.3.1 (Core network aspects).

Similarly, Stand-Alone architecture, as the basis of a ‘fully-5G’ deployment is described in detail in 3GPP TS23.501 [1], and this is further expanded upon in sections 4.2.3 (RAN aspects) and 4.3.2 (Core network aspects).

2 Architectural Principles

5G-VINNI is underpinned by key architectural principles identified in the 5G-VINNI proposal document. These principles are provided below.

- *Common and consistent sets of capabilities available at all 5G-VINNI facility sites.* All 5G-VINNI facility sites support a Core set of functional components and capabilities, meaning that partners that are to perform testing at any facility can be assured a baseline set of capabilities.
- *Interworking between facility sites*. A key strength is the ability to perform E2E testing between sites. This requires a common set of interfaces to be supported by all facility sites to allow the flow of traffic and interaction between control plane systems.
- *Optional elements to allow each site to be unique, and support specific use cases, testing and KPIs.* Whilst each facility site supports the Core set of functions, individual sites will be targeting support of individual (and different) use cases, industry verticals and KPI's. In order to do this, additional optional functions must be supported at sites that need them, to allow specific tailoring and customisation of a site to meet its specific requirements.
- *Support for Network Slicing*. A key component of 5G-VINNI is the support of network slicing. This is supported within each site, to allow parallel testing of multiple use cases on common infrastructure. It will also be supported across interconnect interfaces to demonstrate and support E2E slice establishment and operation, managed between sites.
- *Ability for each site to extend and evolve during the project*. 5G has an underlying principle of networks built using virtualised network functions (VNFs). The iteration and upgrade of functional network capabilities can change very quickly and with high frequency. Work is in progress on 5G specifications, and so within the timeline of the project it is likely that the Standards and State-of-the-Art implementation of a 5G network will potentially change significantly. Therefore it is important that 5G-VINNI facility sites are designed in such a way so as to allow implementations to evolve as technology evolves and improves. This is fundamental to maintaining the relevance of the 5G-VINNI facility to vertical industry partners' testing requirements throughout the duration of the project.
- *Openness and accessibility of the framework*. The design of the 5G-VINNI E2E facility aims to provide openness to allow the flexibility and accessibility for vertical industries to run and test the pilot use cases. The framework allows for different mappings of network functions to an E2E 5G infrastructure, according to different service requirements placed upon it by vertical industries.

With virtualisation, orchestration and network slicing all included as underpinning architectural principles and identified as the basis for implementation as early as 5G-VINNI Release 0, it is clear that a multi-layered implementation is expected with network management and orchestration as disciplines that are pivotal for the instigation of network functions, network slices within one facility site, slices that may traverse multiple facility sites, and within the service layer.

3 Pre-existing work

Whilst 5G-VINNI is a Research project, the nature of ICT-17 as a whole is to build test networks that are Standards-based, and show practical implementation of work that has previously been done in industry SDO's and previous 5G-PPP funded projects.

This chapter provides an overview of work that underpins the development of the 5G-VINNI Facility, and drives the architecture of individual 5G-VINNI Facility Sites.

3.1 3GPP - Release 15

3.1.1 TSG SA1 - Service

Release 14 study item 3GPP TR 22.891 [2] identified the market for verticals, the use cases and requirements that the 3GPP system would need to support. In particular, the use cases are categorized as

- Enhanced Mobile Broadband (eMBB)
- Critical Communications
- Massive Machine Type Communications
- Network Operation
- Enhancement of Vehicle-to-Everything

Figure 3.1: represents the five dimensions of services originally proposed by FS-SMARTER. Based on these service dimensions and use cases, Release 15 specifications 3GPP TS 23.501 [1], 3GPP TS 23.502 [2] and 3GPP TS 23.503 [4] addressed, in particular, eMBB Network Slicing. The study items for the Ultra Reliable Low Latency Communications (URLLC), massive Machine Type Communication and Vehicle to Everything (V2X) are addressed in Release 16.

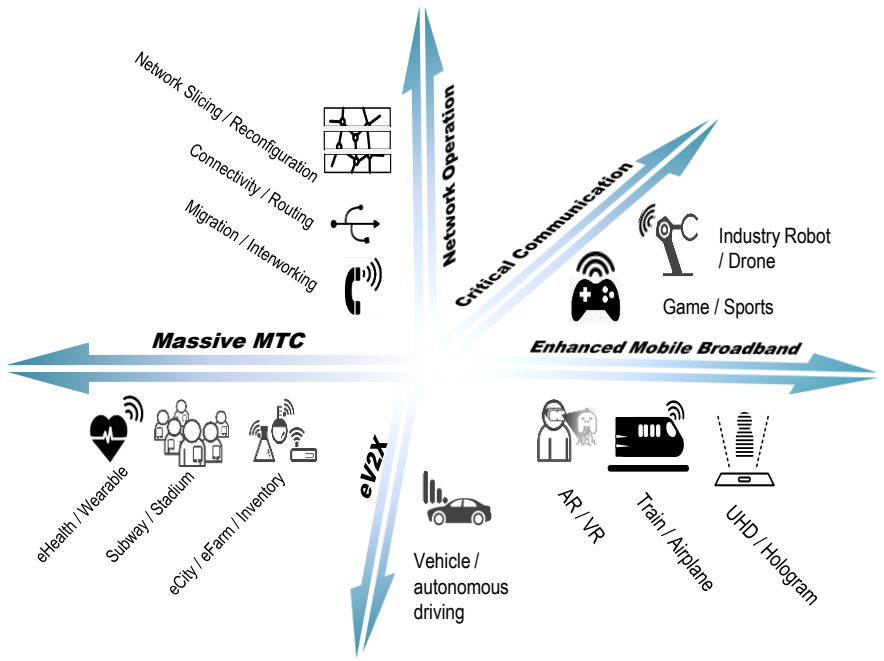


Figure 3.1: Service dimensions proposed by FS_SMARTER

The complete list of use cases identified and studied can be found in 3GPP TR 22.891 [2].

The majority of ongoing 3GPP SA1 study items reflect high industry interest towards vertical requirements and vertical integration in mobile networks. Some relevant Release 16 SA1 study items are summarized as below.

FS_CAV (3GPP TR 22.804 [5]) analyses how 5G systems will extend mobile communication services beyond mobile telephony, mobile broadband, and massive machine-type communication into vertical application

domains. Communication for industrial automation in vertical domains comes with demanding requirements in terms of high availability, high reliability, low latency, and, in some cases, high-accuracy positioning, isolation of communication flows, and guaranteed assurance.

Communication for automation in vertical domains has to support the applications for production in the corresponding vertical domain, for instance, industrial automation, energy automation and transportation. This focus, together with regulations specific for vertical domains, have led to tailored communication concepts in vertical domains such as dependable communication as well as specific security standards and mechanisms.

FS_5GLAN (3GPP TR 22.821 [6]) describes new use cases and potential requirements for a 3GPP network operator to support 5G Local Area Network (LAN)-type services over the 5G system (i.e. UE, RAN, Core Network, and potential application to manage the LAN-style service). In this context, 5G LAN-type services with 5G capabilities (e.g. performance, long distance access, mobility, security) allow a restricted set of UEs to communicate with each other.

FS_V2XIMP (3GPP TR22.886 [7]) has the objective to identify use cases and potential service requirements to enhance 3GPP support for V2X service in the following areas:

- Support for non-safety V2X services (also, referred to as "comfort service") (e.g. connected vehicle, mobile high data rate entertainment, mobile hot-spot/office/home, dynamic digital map update)
- Support for safety-related V2X services (e.g. autonomous driving, car platooning, priority handling between safety-related V2X services and other services)
- Support for V2X services in multiple 3GPP Radio Access Technologies (RATs) (e.g. LTE, New RAT (NR)) and networks environment, including aspects such as interoperability with non-3GPP V2X technology.
- The identification of use cases and potential requirements covers both evolved LTE RAT and new 3GPP RAT (e.g. NR) and also covers V2X operation using 3GPP RATs where there are non-3GPP V2X technologies in use.

FS_Lucia (3GPP TR 22.904 [8]) studies the introduction of an optional, user-centric authentication layer on top of the existing subscription authentication mechanism, supporting authentication mechanisms and interactions with external authentication systems. The new authentication layer is supposed to complement and not to replace existing subscription credentials.

3.1.2 TSG SA2 - Architecture

The specification of 3GPP Release 15 5G system architecture is included in 3GPP TS 23.501 [1], 3GPP TS 23.502 [3] and 3GPP TS 23.503 [4]. Release 15 provides the specification of the '5G phase 1 system', including the key features and functionalities needed to deploy a commercial 5G network. The definition of 5G system architecture includes some prominent 5G-specific features:

- Architecture Modularisation
- Service Based Architecture (SBA)
- Network Slicing
- Multi-access Edge Computing (MEC) support
- Common 5G Core
- 3rd Party Application Function (AF) Integration

3.1.2.1 Architecture Modularisation

To improve flexibility, 5G system architecture differs from the more traditional 4G reference model, where clusters of functions are gathered under network elements, and instead introduces a set of finer granularity Network Functions (NFs) with looser implementation restrictions. 5G system NFs are:

- Authentication Server Function (AUSF)
- Access and Mobility Management Function (AMF)
- Data Network (DN), e.g. operator services, Internet access or 3rd party services
- Unstructured Data Storage Function (UDSF)
- Network Exposure Function (NEF)
- Network Repository Function (NRF)
- Network Slice Selection Function (NSSF)
- Policy Control Function (PCF)

- Session Management Function (SMF)
- Unified Data Management (UDM)
- Unified Data Repository (UDR)
- User Plane Function (UPF)
- Application Function (AF)
- User Equipment (UE)
- (Radio) Access Network ((R)AN)
- 5G-Equipment Identity Register (5G-EIR)
- Security Edge Protection Proxy (SEPP)
- Network Data Analytics Function (NWDAF)

3.1.2.2 Service Based Architecture

5G system architecture introduces a Service-Based reference model, where NFs provide services to each other. The interaction between NFs is represented in two ways.

- A service-based representation, where NFs within the Control Plane (CP) enable other authorized NFs to access their services. This is referred to as the Service-Based Architecture (SBA).
- A reference point representation, showing the interaction between the NF services in the NFs described by point-to-point reference points between any NFs pair.

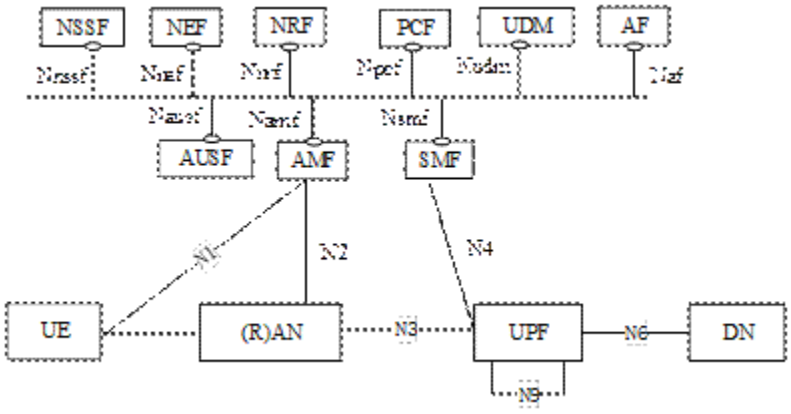


Figure 3.2: 5G System architecture, service based representation for the non-roaming case from 3GPP TS 23.501 [1]

Figure 3.2, taken from 3GPP TS 23.501 [1] illustrates the reference model in the service-based representation for the non-roaming case. Other variants of both the service-based and reference point defined architecture can be found in 3GPP TS 23.501 [1].

3.1.2.3 Network Slicing

A key feature of 5G system architecture is Network Slicing. 5G system network slice refers to a complete E2E logical network (i.e. including both Access and Core networks) providing telecommunication services and network capabilities. Network slicing enables the network operator to deploy multiple, independent PLMNs where each is customized by instantiating only the features, capabilities and services required to satisfy the subset of the served users/UEs or a related business customer needs. Multiple instances of the same Network Slice (Network Slice Instances (NSI)s) MAY be deployed over virtual and physical infrastructures, e.g. by the same operator willing to provide separately the same telecommunication service to different groups of users.

In order to identify NSIs, and to implement selection procedures allowing UEs to register to and utilise services provided by NSIs, the Network Slice Selection Assistance Information (NSSAI) parameter has been defined. A NSSAI, which can have standard values as well as PLMN-specific values, identifies the Core NFs common to a group of NSIs. Within a group of NSIs identified by an NSSAI, each NSI is uniquely identified by the Subscribed NSSAI (S-NSSAI) parameter.

SBA, together with softwarization and virtualization, provides network agility that will enable an operator to respond to customer needs quickly. Dedicated and customized network slices can be deployed with functions, features, availability and capacity as needed.

3.1.2.4 Edge Computing Support

5G system architecture includes enhancements to support MEC. MEC enables operators and 3rd party services to be hosted close to the UE's access point of attachment, to achieve efficient service delivery through reduced E2E latency and load on the transport network.

The 5G Core Network includes enhanced capability to select UPF close to the UE and MAY execute traffic steering from the UPF to the local Data Network via N6 interface shown in Figure 3.2. This MAY be based on the UE's subscription data, UE location, and information from the Application Function (AF). Additionally, 5G Core Network MAY expose network information and capabilities to an Edge Computing Application Function.

Edge Computing is supported by the combination of the following 5G enablers:

- User plane (re)selection: the 5G Core Network (re)selects UPF to route the user traffic to the local Data Network;
- Local Routing and Traffic Steering: the 5G Core Network selects the traffic to be routed to the applications in the local Data Network;
- Session and service continuity to enable UE and application mobility;
- An Application Function MAY influence UPF (re)selection and traffic routing via PCF or NEF;
- Network capability exposure: 5G Core Network and Application Function are able to provide information to each other via NEF or directly;
- Quality of Service (QoS) and Charging: PCF provides rules for QoS Control and Charging for the traffic routed to the local Data Network;
- Support of Local Area Data Network: 5G Core Network provides support to connect to the Local Area Data Network in a certain area where the applications are deployed.

3.1.2.5 3rd Party Application Function Integration

The flexibility of its architecture, together with the network slicing feature makes the 5G system suitable to tackle vertical industries, satisfying diverse and dynamic requirements. Towards this direction, 5G System has also been enriched by enhanced capabilities allowing Application Function (AF) from 3rd parties to interact and influence 5G system behaviour. This will further enhance the possibility of customising the architecture and features of a communication system. One notable example included in Release 15 is where the AF can influence traffic routing - an AF MAY send requests to influence SMF routeing decisions for traffic of Packet Data Unit (PDU) Session. The AF requests MAY influence UPF (re)selection and allow routeing of user traffic to a local Data Network. The AF requests are sent to the PCF via N5 or via the NEF. The AF requests that existing or future PDU Sessions of multiple UE(s) are sent via the NEF and MAY target multiple PCF(s). The PCF(s) transform(s) the AF requests into policies that apply to PDU Sessions.

3.1.2.6 Deployment Options

To facilitate a gradual deployment of 5G System, to ensure smooth integration with existing legacy 4G systems, and to allow independent deployments of 5G RAN and 5G Core, 3GPP specified a set of architecture options. The numbering of options is not incremental and reflects the 3GPP defined numbering.

Option 2 (Stand-alone) defines a full 5G System, including 5G Core with gNB, as shown in Figure 3.3.

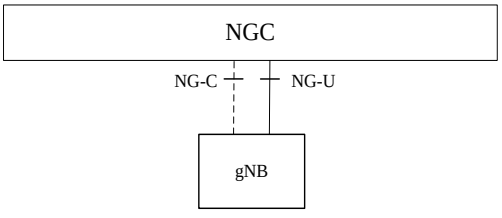


Figure 3.3: 5G System architecture, Option 2 (standalone)

Options 3, 3a and 3x (Non-Standalone) allow NR deployments reusing EPC with the support of LTE eNB. With these options, the LTE eNB is connected to the EPC with Non-standalone NR. The NR user plane connection to the EPC goes via the LTE eNB (Option 3) or directly (Option 3A). In Option 3x, the solid line shown between LTE-eNB and gNB is used for user plane data transmission terminated at the gNB, i.e., S1-U data from EPC is split at the gNB. These options are shown in Figure 3.4.

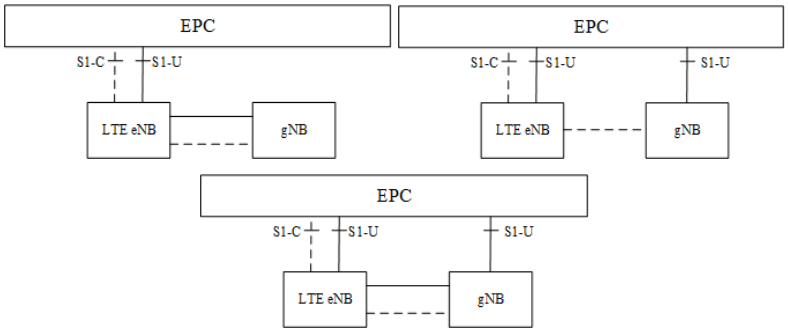


Figure 3.4: 5G System architecture, Option 3, 3a and 3x (Non standalone)

Options 4 and 4a (Non Standalone): with this option, the gNB is connected to the Next Generation Core (NGC) with Non-standalone E-UTRA. The E-UTRA user plane connection to the NGC goes via the gNB (Option 4) or directly (Option 4A), as shown in Figure 3.5.

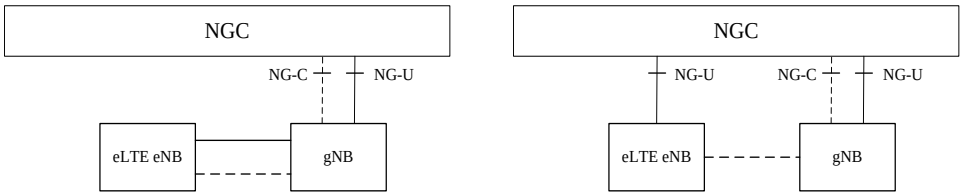


Figure 3.5: 5G System architecture, Option 4 and 4a (Non standalone)

Option 5: with this option eLTE eNB is connected to NGC, as shown in Figure 3.6.

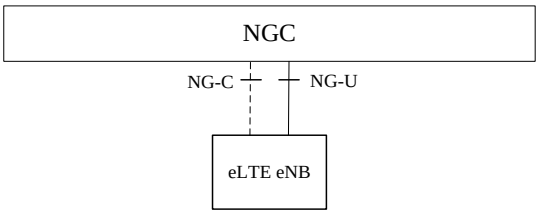


Figure 3.6: 5G System architecture, Option 5

Options 7 and 7A: with this options, the eLTE eNB is connected to the NGC with Non-standalone NR. The NR user plane connection to the NGC goes via the eLTE eNB (Option 7) or directly (Option 7A), as shown in Figure 3.7.

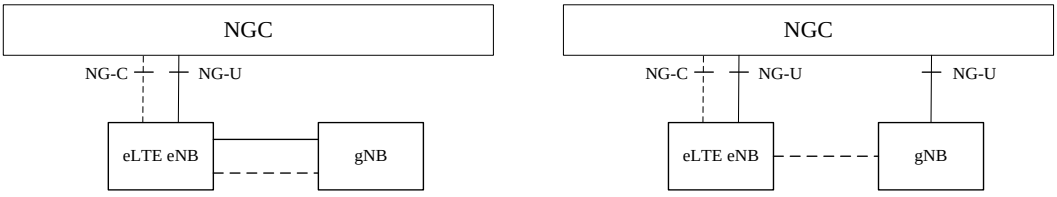


Figure 3.7: 5G System architecture, Option 7 and 7a

3.1.3 TSG SA3 – Security

5G brings a number of new possibilities for Mobile Network Operators (MNOs) to enable new use cases and business models, by application of mobility to new business sectors and device types. However, this in turn has the potential to open up new threats to security of the customer and the network. 3GPP SA3 has completed

work to offset this risk in a number of areas. Table 3.1 below summarises the incremental security measures that have been put in place in 5G, when compared to 4G.

Table 3.1: Summary of security enhancement in 5G

Security Feature	4G	5G
Access Agnostic Authentication	Not Access Agnostic	Unified Authentication for all access
Authentication Credentials	Only AKA credentials	AKA Credentials or Certificate for Internet of Things (IoT)/Private networks
Authentication Protocol	EPS-AKA over 4G-NAS	5G-AKA over 5G NAS or EAP-AKA'/EAP-TLS over 5G NAS
Security Platform for Authentication Credentials	UICC	UICC or Non-removable UICC
Home Control for Authentication	Not Supported	Supported (HPLMN involved in Authentication and holds a key)
Integrity Protection of UP traffic	Not Supported	Supported (optional to use)
Subscription Identity Protection	IMSI is not protected, if there is no security context	SUPI is always protected using Asymmetric Cryptography
Network Domain Security	IPSec (point-to-point architecture)	TLS/Application layer protection (SBA)
Security Visibility	Visibility to User	Visibility to User and Application (e.g. via API), per PDU session granularity

3.1.4 TSG SA5 - Telecom Management

3GPP TS 28.530 [9] defines use cases to be supported by the Release 15 3GPP management system. It also introduces the key concepts relating to slicing and services. One of the key discussions is the foreseen roles in the 3GPP system.

3.1.4.1 Roles in the 5G network

The roles identified by 3GPP within a 5G network are illustrated in Figure 3.8, and defined as;-

- Communication Service Customer (CSC) – the final user of the communication service. For example a vertical company.
- Communication Service Provider (CSP) – the actual entity that provisions the communication service for the CSC. This could be an operator that specializes in providing the V2X service.
- Network Operator (NOP) – The traditional network operator that provides the network services to host the communication service. This could be a PLMN owning operator.
- Network Equipment Provider (NEP) – Vendor role that build the network equipment.
- Virtualisation Infrastructure Service Provider (VISP) – Specialized vendor for virtualized services
- Data Centre Service Provider (DCSP) – Cloud provider where the VNF MAY be hosted
- Network Function Virtualisation Infrastructure (NFVI) Supplier – Supplies NF virtualization infrastructure to its customers.
- Hardware Supplier – Supplies hardware.

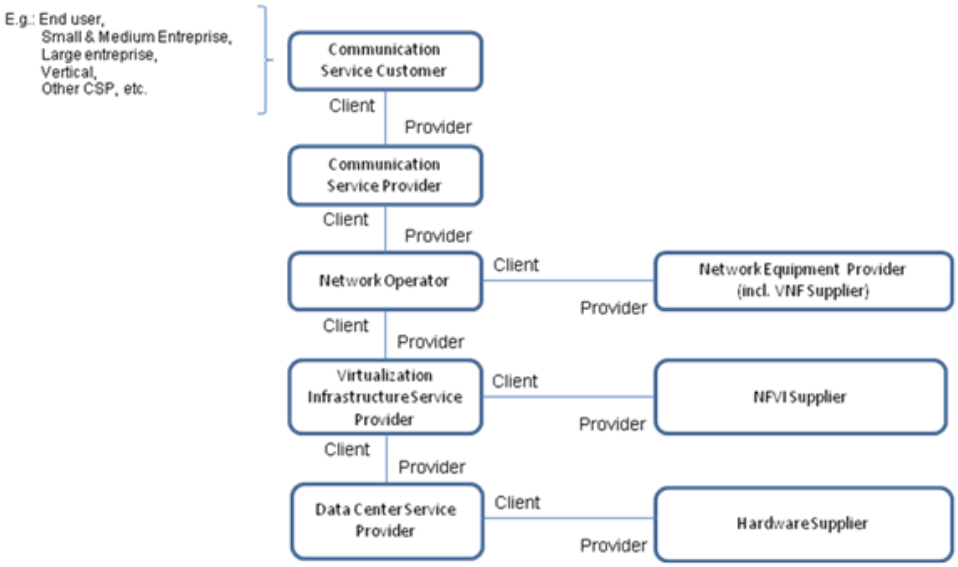


Figure 3.8: The high level roles in 5G (for TS28.530 [9])

5GEx and SLICENET projects have defined actor role models that are in line with the 3GPP defined roles, in [69] and [70] respectively. 5GEx has also defined some additional roles that harmonically complement the actor role model defined by 3GPP. The complementary roles that were defined and are of high interest for 5G-VINNI, were identified as *Support roles*. These supporting roles include the following:

- NFVI Operations Support Provider: Offering operations support services for the readiness, operations, administration and maintenance of NFVI.
- VNF Deployment Support Provider: Offering VNF (SWaaS) support service for on-boarding, deployment and related readiness tasks.
- VNF Operations Support Provider: Offering VNF (SWaaS) in-use operations support services. This can depend on the specific function of the VNF and can cover a range of support tasks, e.g. data analytics.

The 3GPP defined roles, along with the support roles defined by 5GEx project, are key for 5G-VINNI architecture definition and will be taken into account. An extended analysis of 5G ecosystem actors, how these roles are mapping to 5G-VINNI and what are the new roles mandated by 5G-VINNI are of WP5 interest and such an input will be delivered in D5.1.

3.1.4.2 The Concept of the Network Slice Instance (NSI)

In alignment with the roles, 3GPP SA5 has also defined the concept of a NSI shown in Figure 3.9 and its lifecycle shown in Figure 3.10.

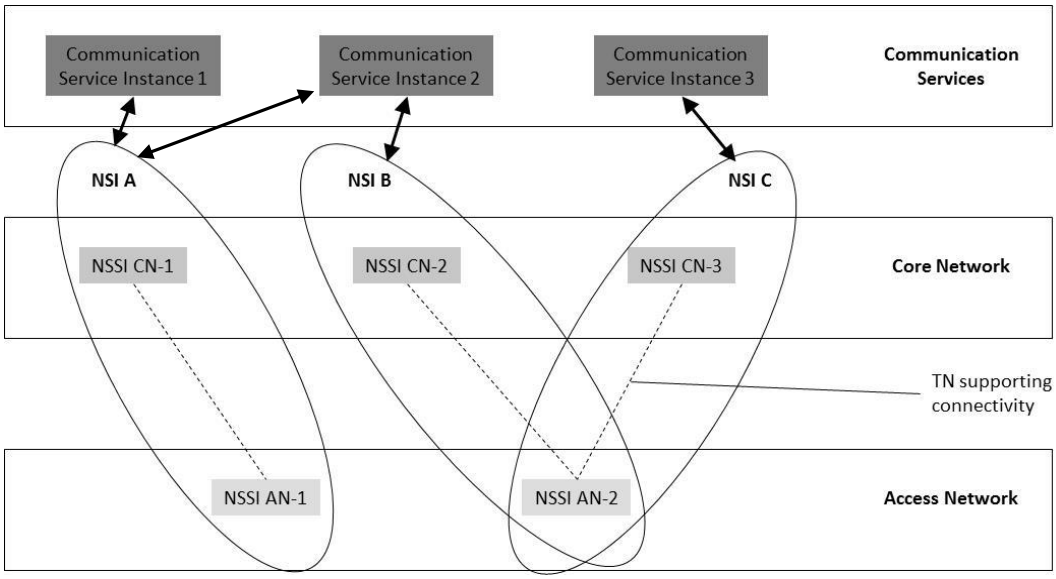


Figure 3.9: The NSI composed of NSSI to provide communication services (from 3GPP TS 28.530 [9])

An NSI can be composed of multiple Network Sub-Slice Instances (NSSI) which could be shared across multiple NSI. One of multiple NSI could be used to provide one or more communications services. Before deploying the NSI, the NSI has to be prepared, which includes designing, on-boarding and other network-related processes. The commissioning phase is when the NSI is provisioned in the network. Following this there is activation and then the operation phase is initiated. The operation phase includes the performance and fault management of the NSI. Finally, when it is no longer needed, the NSI can be terminated. This process is illustrated in Figure 3.10.

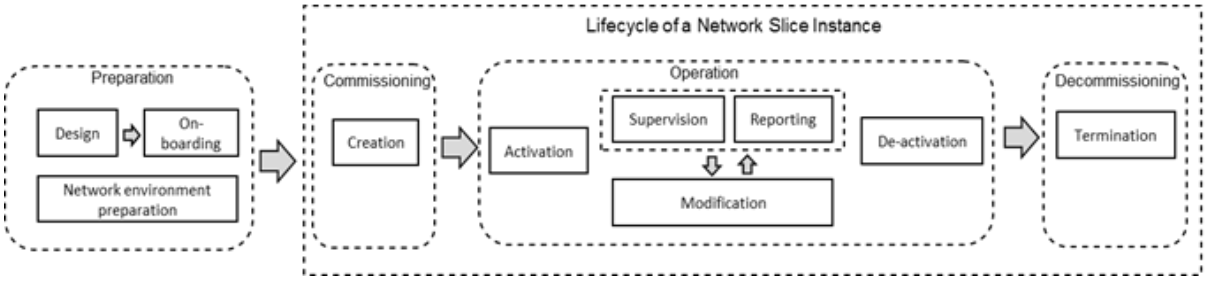


Figure 3.10: The life cycle of an NSI (from 3GPP TS 28.530 [9])

An NSI is made of one or more NSSI which, in turn, can recursively be composed of one or more NSSI. As such the relationship between the NSI and NSSI becomes unclear. While the need for sub-management based on different domains is clear, the need for an overall NSI is unclear. 3GPP SA5 resolves this issue by specifying that an NSI and an NSSI are both derived from the same parent class and they both have a significant number of common attributes. The difference is that NSI is that the highest level of NSSI that is provided externally to host a communication service.

3.1.4.3 The Management plane (MP) SBA

3GPP SA5 specifications for R15 have moved away from an Integration Reference Point definition-based architecture to a service oriented framework in the interest of modularity and scalability of the management services. The key concept and design principles of management services are defined in 3GPP TS 28.533 [10], while the generic list of management services and their example usage is defined in 3GPP TS 28.532 [11]. This, however, highlights a gap in the specification as now the management services themselves would needs a service management framework, but the requirements for this are currently not discussed in the standard.

Using the concept of SBA as specified in 3GPP TS 28.533 [10], 3GPP SA5 in 3GPP TS 28.532 [11] specifies the generic list of management services that an operator management system would have. These include provision, fault and performance management services for NSI, NSSI and NFs.

3.1.4.4 E2E KPIs

With the introduction of slicing and the use cases to support verticals (Communication services consumers), there is a need to provide verticals with details on the management data of the E2E service. 3GPP SA5 has concluded that the fragmented collection of KPIs of existing infrastructure and services would not be of interest to the communication services consumer, and has instead created a new specification on the standardized E2E KPIs in 3GPP TS 28.554 [12]. Some of the key KPIs that have been specified are:

- E2E Latency
- Downlink latency
- Throughput: upstream and downstream
- Upstream throughput at N3
- Downstream at N3

3.1.5 RAN decomposition

In some implementations of LTE, RAN networks have implemented a separation between the Radio Unit (RU) and the Baseband Unit (BBU), by exposing the Common Public Radio Interface (CPRI). This has been conceived to allow the BBU to be at a network consolidation point, which results in operational efficiencies and cost savings. However, the CPRI interface is constrained by the need to meet very low delay requirements, and being of very high bandwidth, resulting in a need for high capacity transport links. It is also noted that while CPRI is widely recognised as an implementation option, it is not acknowledged in 3GPP LTE standards.

In 5G, however, 3GPP TR 38.801 [33] included work to study possible options for decomposition of the RAN environment, which resulted in the identification of eight options. These are illustrated in Figure 3.11.

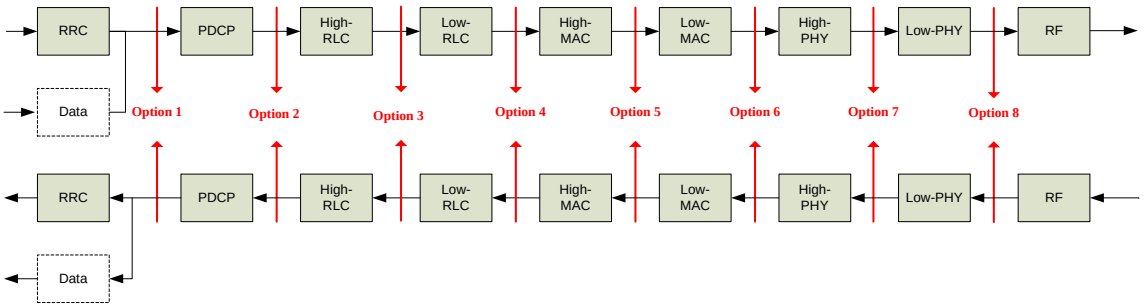


Figure 3.11: RAN decomposition options (3GPP TR 38.801 [33])

Option 8 from Figure 3.11 is the exposure of the CPRI interface, while Option 7 is referred to as enhanced CPRI (eCPRI). Both amount to the separation of the RU from the BBU. Of the remaining options, only Option 2 has resulted in significant further work. Option 2 results in the separation of a Distributed Unit (DU) and a Centralised Unit (CU), with the F1 Reference point defined to connect the CU and DU.

Further to this, work was conducted to separate the CU into UP and CP components. This work was documented in 3GPP TR 38.806 [67] and is illustrated in Figure 3.12.

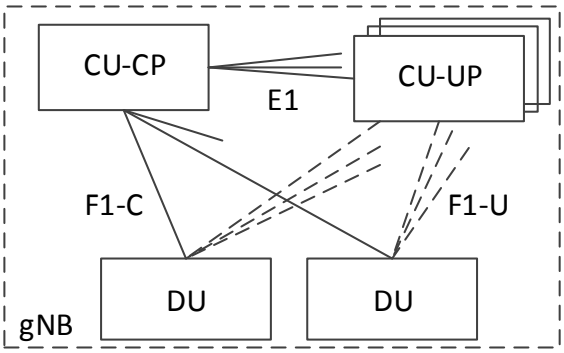


Figure 3.12: CP/UP split of gNB (3GPP TR 38.806 [67])

As can be seen, this results in the decomposition of the F1 interface into control and user plane parts, and the exposure of the E1 reference point between the CP and user plane functions of the CU.

3.2 5GPPP Architecture Working Group

The 5GPPP Architecture working group (WG) has acted as the point of consolidation for 5GPPP Projects in Phase 1, for topics relating to 5G Architecture. This has led to the production of two White Papers, the most recent of which – 'View on 5G Architecture, Version 2.0' [14] was published in December 2017.

The white paper looks at 'capturing novel trends and key technological enablers for the realization of the 5G architecture. It also targets at presenting in a harmonized way the architectural concepts developed in various projects and initiatives (not limited to 5GPPP projects only) so as to provide a consolidated view on the technical directions for the architecture design in the 5G era.' As such, the scope of the white paper is somewhat aligned with that of this document. The key difference is that this deliverable is the basis for actual realization and implementation of 5G Architectures, over which specific use cases will be enable, whereas the Architecture WG paper is a retrospective look at what has been done.

In some areas, the Architecture White Paper is a major reference point. This is particularly true for Orchestration and Management topics, where the White Paper captures the consolidated outputs from multiple 5G-PPP projects, and documents how this work has been successful in influencing external SDOs. This is reflected in this document through direct references included in Sections 4.6 and 4.7, and throughout Chapter 5.

5G-VINNI will take the lessons described in the White Paper as potential 'best practise', but equally not be tied completely to the work of the Architecture WG. Indeed, as 5G-VINNI facility and site implementation commences, it is possible that practical experience provides evidence that Architecture WG recommendations need to be revisited and modified. 5G-VINNI will play an active role in the Architecture WG.

3.3 ETSI NFV

The ETSI Industry Specification Group for Network Functions Virtualization (ETSI ISG NFV) is the group responsible for the standardization of NFV technology. NFV decouples NFs from purpose-built (proprietary, closed, costly) hardware appliances, and move them to software images that can run on commodity (general-purpose, industry standard, inexpensive) hardware. Leveraging IT virtualization technologies, NFV enables traditional NFs to be virtualized, resulting in the so-called VNFs. These VNFs can be deployed and operated with great agility on top of commodity hardware, and can be flexibly chained to define network services. To manage them, ETSI ISG NFV has defined a reference framework, with different functional blocks and interfaces to exchange information. In this section we provide an overview on these concepts, as they provide the foundations for the operation of the 5G-VINNI facility in virtualized environments.

3.3.1 The Concept of NFV Network Service

An NFV network service (NS) is a composition of NFs. A NF is a processing function in the network which has defined functional behaviour and external interfaces. According to ETSI NFV, an individual NF may be implemented as a VNF, i.e. a software image running on commodity hardware, or as a Physical Network Function (PNF), i.e. a purpose-built hardware appliance. However, when arranged into a NFV NS, at least one of the involved NFs SHALL be a VNF.

The components of a NS include NFs (with at least one VNF), virtual links and VNF Forwarding Graphs (VNFFGs). Virtual links are abstractions of physical links that logically connect the connection points (CPs) exposed by the different NFs, providing connectivity between them. To specify how these connections are made along the entire NS, one or more VNF Forwarding Graphs (VNFFGs) are defined. A VNFFG describes the connectivity topology of (all or part of) the NS, and optionally includes one or more Network Forwarding Paths (NFPs) to specify how traffic will flow across the NFs defined in this topology. To this end, each NFP includes (i) a list of CPs forming a VNF chain, and (ii) forwarding rules applicable to those CPs.

For a fine-grained control of its performance, scalability, security and reliability, a VNF can be decomposed into one or more VNF components (VNFCs), each performing a well-defined task within the VNF functionality. Each VNFC is hosted in a single virtual deployment unit (e.g. virtual machine, docker container, unikernel) and connected with other VNFCs through internal virtual links, as described in Adamuz-Hinojosa *et al* [15]. These

virtual links logically connect the internal CPs exposed by the VNFCs, hence providing connectivity at the VNF level.

Figure 3.13 and Figure 3.14 show an example of a NFV NS. As seen from Figure 3.13, this NS is composed of four NFs (three VNFs and one PNF) connected with five virtual links. To provide an insight on the internal composition of a VNF, one of the NS's VNFs have been expanded and analysed in detail. In Figure 3.14, two VNFFGs have been proposed to illustrate some of the (tentative) goals for VNFFG differentiation, these being service chains for traffic from different network planes. In this example, VNFFG1 is intended for user plane traffic, and VNFG2 for CP traffic. Note that VNFFG1 in turn includes two NFPs for traffic steering, enabling the definition of two user plane traffic flows for distinct processing, e.g. for different QoS treatment.

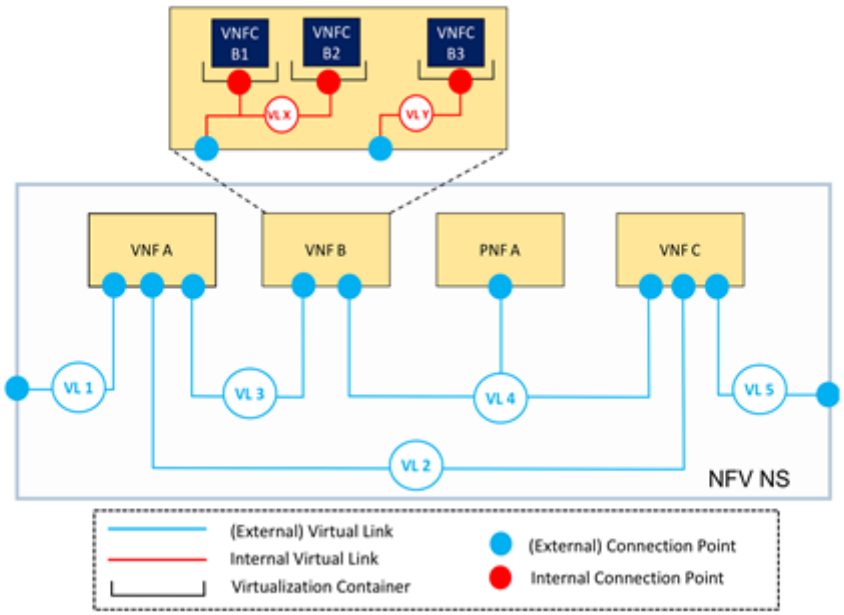


Figure 3.13: Example of a NFV NS

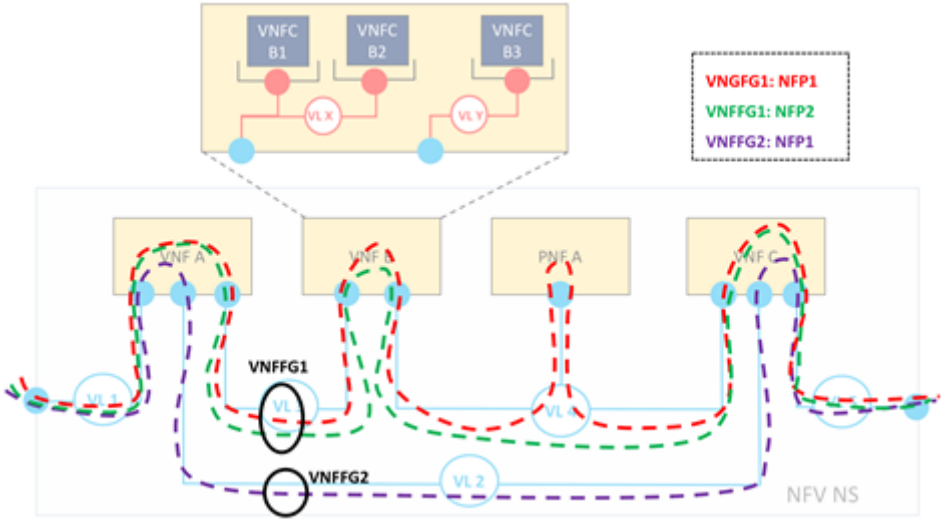


Figure 3.14: VNFFGs and NFPs in a NFV NS

3.3.2 ETSI NFV reference architectural framework

The deployment and operation of NSs in (highly diverse) virtual environments bring multiple points of management that need to be carefully addressed. Most of them highly depend on the NF composition mechanism used for NS definition. This mechanism creates a set of relationship/dependencies between a given NS and its components, which exerts a strong influence on their management, and on the orchestration of the underlying resources. To deal with these issues in an effective manner, ETSI ISG NFV has defined a reference

architectural framework for NFV in ETSI GS NFV-MAN 001 [16]. The ETSI NFV framework provides architectural foundations that allow consistency and uniformity in NS deployment and operation.

A high-level overview of the ETSI NFV architectural framework is graphically depicted in Figure 3.15. As can be seen, this framework consists of three main working domains: (i) infrastructure and NF layers, (ii) NFV Management and Orchestration, and (iii) Network Management Systems (NMS). Each of these domains are briefly described below.

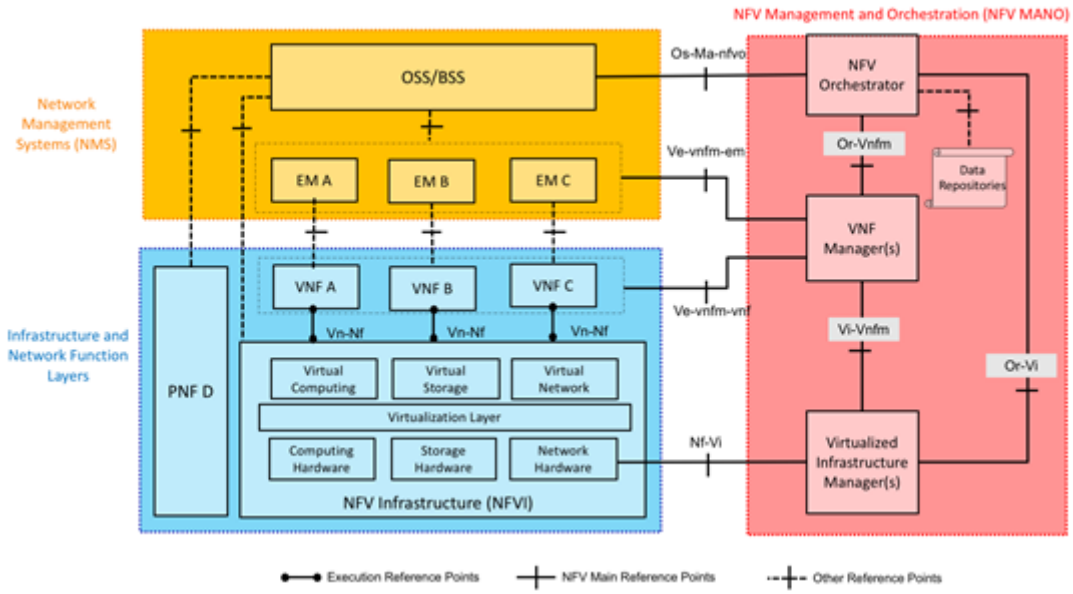


Figure 3.15: ETSI NFV reference architectural framework

3.3.2.1 Infrastructure and NF Layers

The infrastructure and NF layers include the NFVI and the different NS's NFs - VNFs and (if required) PNFs. While PNFs run on purpose-built, closed hardware equipment, VNFs are deployed and executed on the NFVI. The NFVI is the collection of all hardware and software resources that build up the cloud environment on top of which VNFs (and their constituent VNFCs) run. By means of a virtualization layer, the underlying physical resources are abstracted and logically partitioned into a set of virtual resources, used for VNF(C) hosting and VNF(C) connectivity. The set of resources that make up the NFVI could be defined within a single Point of Presence (PoP), but also might be distributed across multiple PoPs. In such a case, VNFs could be deployed and executed across geographically remote areas, enabling multi-site NFV scenarios.

3.3.2.2 NFV Management and Orchestration

This domain focuses on all the virtualization-specific tasks that are needed for NS deployment and operation in NFV environments, and is made up of the NFV Management and Network Orchestration (NFV MANO) architectural framework. NFV MANO is a software stack consisting of three types of functional blocks, each with a well-defined functionality - Virtualised Infrastructure Manager (VIM), VNF Manager (VNFM), and NFV Orchestrator (NFVO). Figure 3.15 shows their logical disposition in the NFV-MANO stack. As seen, these NFV-MANO functional blocks operate at different abstraction layers, and exchange information using standardized reference points. In the following lines, a brief description of their functionality is provided.

At the lowest abstraction layer, there is the **VIM**. Using the Nf-Vi reference point, the VIM controls and manages the NFVI resources on top of which VNF(C)s are deployed and executed. Depending on the NFVI resource distribution, the management scope of a VIM could include one or more PoPs.

At the intermediate abstraction layer, there is the **VNF Manager (VNFM).** The VNFM performs management actions over VNF instances. These instances can be from the same or different VNF types. The management scope of the VNFM includes the lifecycle management of VNFs (e.g. instantiation, scaling, healing, termination), and their performance and fault management at the virtualized resource level. The latter requires collection and correlation of performance and fault data from the specific NFVI resources that host the VNF instances. To

do this, the VNFM communicates with the VIM using the interfaces specified in the Vi-Vnfm reference point in ETSI GS NFV-IFA 006 [17].

Finally, there is the **NFVO**. Operating above the VIM and the VNFM, the NFVO is the heart of the NFV-MANO architecture. It combines VNFs and (if required) PNFs to deploy and operate NS. To do this, the NFVO functionality can be divided into two broad categories: (i) resource orchestration, and (ii) NS orchestration. Resource orchestration enables access to NFVI resources (independently of any VIM) in an abstracted manner, and the governance of VNF instances sharing the same NFVI. NS orchestration chains NFs to create one or more NSs, and manages instances of these NSs during their life cycle. A more detailed summary of resource and NS orchestration functions is graphically depicted in Figure 3.16, and thoroughly discussed in Gonzalez *et al* [18]. As seen, most of these functions assume that the NFVO needs to exchange information with the VNFM and VIM. For this end, the interfaces specified in the Or-Vnfm (as per ETSI GS NFV-IFA 007 [19]) and Or-Vi (as per ETSI GS NFV-IFA 005 [20]) reference points are defined.

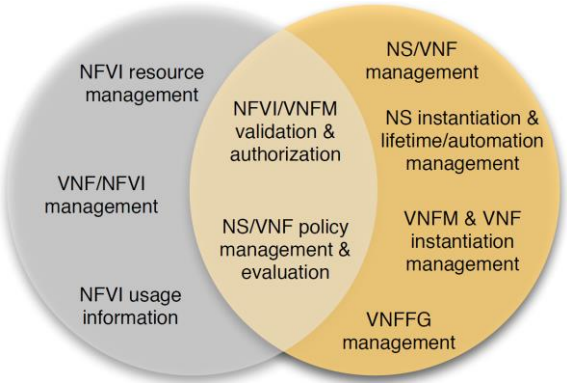


Figure 3.16: The NFVO functionality split between Resource Orchestration (grey circle) and NS orchestration (yellow circle) – Gonzalez et al [18]

To assist the above-referred functional blocks with their tasks, the NFV MANO contains a set of data repositories that keep different types of information. Up to four different data repositories can be considered in NFV MANO:

- **Network Service Catalogue***:* stores one or more NS Descriptors (NSDs), each including Virtual Link Descriptors (VLDs) and VNFFG Descriptors (VNFFGDs). An NSD is a readymade template that contains machine-readable information used by the NFVO to deploy instances of a given NS, and operate them throughout their lifetime.
- **VNF Catalogue:** stores one or more VNF Packages, each describing the deployment and operational behaviour of a given VNF. A VNF package is composed of two parts: (i) a VNF Descriptor (VNFD), and (ii) one or more software images. On one hand, VNFD contains information on how instances of the VNF need to be created (i.e., how many virtual deployment units are needed to host the instances of the VNF's VNFCs, and which are their resource requirements) and operated (i.e. under which conditions different life cycle management actions can be triggered). Used by the VNFM, VNFDs are similar to NSDs, but at the VNF level. On the other hand, the software images run the code executing VNF application. The EM is responsible to install the software image(s) into the virtual deployment units hosting VNF's VNFC instances. **NFV Instances Repository**: holds run-time information about all the VNF and NS instances that are executed in the NFVI. This information can be used by the VNFM and the NFVO to trigger appropriate life cycle management operations over the instances under their management domain.
- **NFVI Resources Repository:** keeps updated information about the state (available / reserved / allocated) of NFVI resources. The NFVO makes use of this information to perform resource orchestration functions.

3.3.2.3 NMS

The NMS domain focuses on traditional (i.e. non-virtualization-related, vendor-agnostic) tasks. Unlike NFV MANO, which is focused on NS (and VNF) management and orchestration at the virtualized resource level, NMS focuses on application-aware NS (and VNF) configuration and management, and is responsible for the deployment and maintenance of PNFs. NMS comprises two types of functional blocks:

Element Manager (EM): anchor point responsible for the fault, performance, configuration, accounting, and security management (FCAPS) of one or more VNFs. Unlike a VNFM, an EM focuses on the VNF application layer management. This means collecting application-specific VNF data rather that NFVI-related VNF information.

Operations/Business Supporting Systems (OSS/BSS): a collection of subsystems and management applications that network operators use to provision and operate their NSs.

Despite their orthogonal management scopes, NMS and NFV MANO domains need to interact to enable NS deployment and operations, where the two management levels (i.e. management at the virtualized resource level and the application level) are considered. The interplay between the two working domains is realized via two reference points: Ve-Vnfm-em reference point, which is defined in ETSI GS NFV-IFA 008 [21], and is used for the exchange of information between VNFM and EM for VNF management, an Os-Ma-Nfvo reference point, defined in ETSI GS NFV-IFA 013 [22], and used for the exchange of information between NFVO and OSS for NS management.

3.3.3 Ongoing work in ETSI ISG NFV

Since its definition in 2013, ETSI NFV reference architectural framework has been continually enhanced with new features. To provide network operators, solution providers, and the research community with a stable platform for implementation while introducing new features, ETSI ISG NFV organizes its work into two-year phases, with each phase defining a new release. Releases 1 and 2 are already completed, and NFV Release 3 is on progress. The main focus on the new release is the enrichment of the NFV architectural framework, making NFV technology ready for expansive deployment and enhanced operation. A brief recap on the NFV Releases is graphically depicted in Figure 3.17.

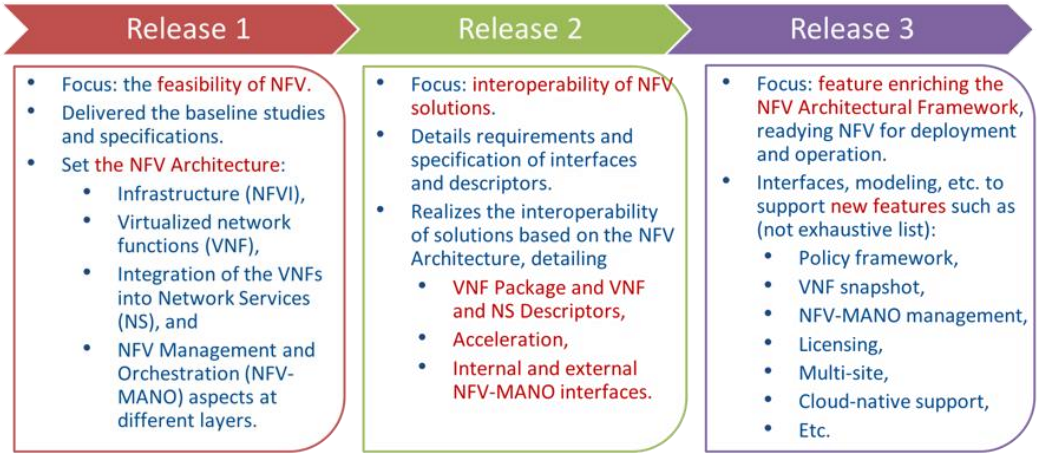


Figure 3.17: Overview on NFV Releases. Source: ETSI ISG NFV

As described in Chapter 4, NFV MANO is a key component in 5G-VINNI facility, as it enables efficient and cost-effective management and orchestration in virtual environments. Keeping track of advancements conducted by these Working Group with respect to the NFV MANO, and progressively introducing them in the deployed NFV MANO stack(s) will be key to enhance 5G-VINNI facility capabilities and to guarantee interoperability across different sites.

3.4 ETSI ZSM

ETSI Zero-touch network and Service Management (ZSM) group was conceived as a next-generation management system that leverages the principles of Network Functions Virtualization (NFV) and Software Defined Networking (SDN). It will be designed for the new, cloud-based network infrastructures and functions, and **based on cloud-native principles to address zero-touch** (fully automated) management and operation. The challenges introduced by the deployment of new network foundations such as NFV and new architectures such as 5G trigger the need to accelerate network transformation and radically change the way networks and services are managed and orchestrated.

Within ETSI GS ZSM 002 [23], ETSI ZSM decouples network domain management and E2E cross-domain service management. This eliminates monolithic systems and reduces overall service management overhead, whilst enabling independent evolution of the domain and E2E management capabilities. The principle is illustrated in Figure 3.18.

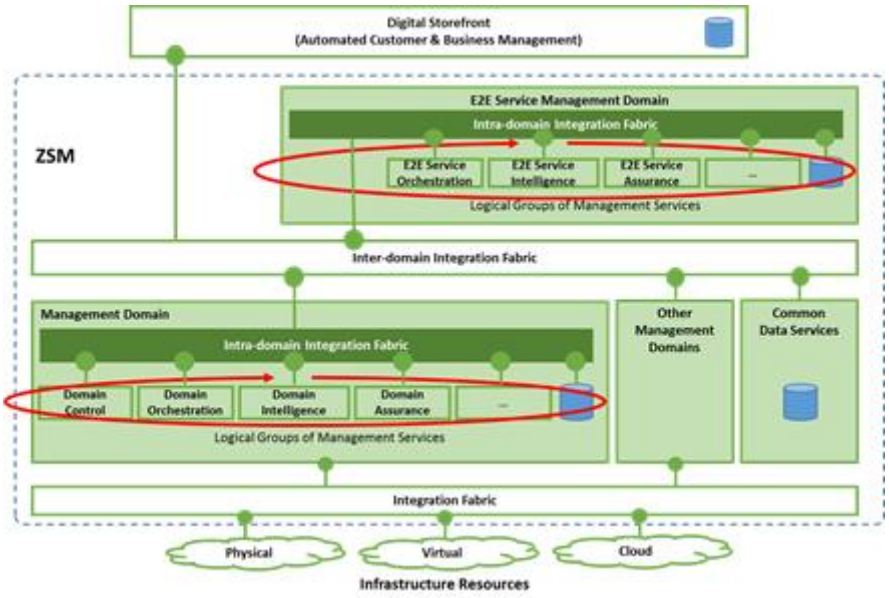


Figure 3.18: ZSM framework reference architecture (from ETSI GS ZSM 002 [23])

This principle is used within 5G-VINNI's framework architecture, as described in Chapter 4.

3.5 Satellite Integration into 5G

Integration of Satellite networks into 5G is considered a crucial endeavour to fully satisfy the challenging 5G connectivity requirements. In this context, the outputs from the EU H2020 5GPPP Phase 2 project SaT5G [24] and the ESA ARTES project SATis5 [26], as well as the works in Timeola *et al* [26] and Corici *et al* [27], aim at fostering seamless integration of satellite usage within terrestrial 5G networks. With this way, they will enable high-value attractive solutions for Satellite communications and terrestrial actors.

The positioning of the satellite link in 5G system architecture, as defined in the 3GPP TS 23.501 [1], is depicted in Figure 3.19 including two main possibilities:

- **Direct access:** satellite-capable UE has a direct access to the 5G network through a satellite link;
- **Indirect access or backhaul:** UE connects to (R)AN via 3GPP or non-3GPP access technologies. (R)AN is connected to the 5G Core through a satellite link.

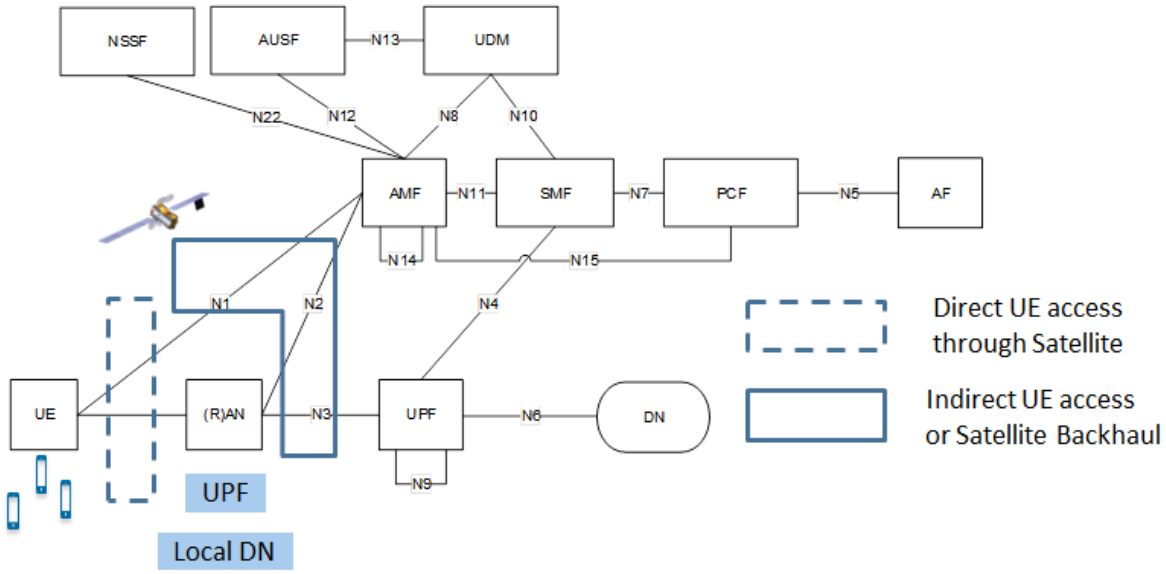


Figure 3.19: Satellite positioning into 5G System Architecture (Source: SaT5G [24])

For each integration type, several "extensions" can be derived, depending on specific implementation options. Figure 3.20 provides a taxonomy of the various implementation options as well as support of some additional features (MEC and Multilink for instance) identified by the EU H2020 5GPPP Phase 2 project SaT5G [24], [26] project in terms of satellite integration into 5G.

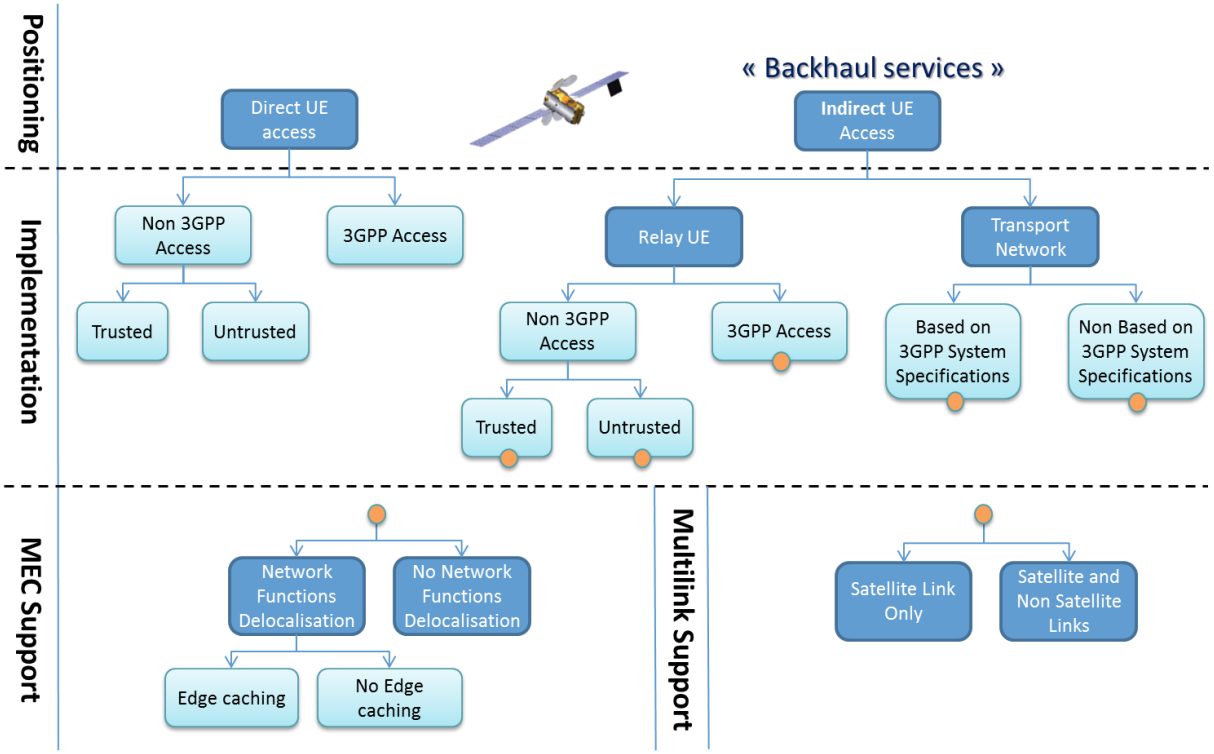


Figure 3.20: Taxonomy of Various Implementation Options for Satellite Integration into 5G System (Source: SaT5G [24])

Management and orchestration (MANO) is of paramount importance for achieving a seamless integration, and support of Network Slicing, is a "must have" feature introduced for 5G, as was described in section 3.1. Two main approaches can be envisaged in this domain – either a single orchestrator for both satellite access network and terrestrial network or, as a more likely option, specific orchestrators are developed for each segment, and appropriate interfaces need to be defined to enable overall orchestration.

Considering the 5G reference system architecture defined in the 3GPP TS 23.501 [1], for the integration of Satellite Networks into 5G system, some existing interfaces may need to be adapted and new interfaces may need to be defined. In that perspective the EU H2020 5GPPP Phase 2 project SaT5G is considering satellite networks according to several scenarios of integration in the 5G system, each leading to a specific Capital Expenditure (CapEx) and/or Operational Expenditure (OpEx) impact compared to existing satellite communication (SatCom) solutions. The outcome of the SaT5G project is captured in the ETSI TR 103 611 [28].

SatCom integration in the 5G-VINNI project will focus on the indirect access or backhaul scheme described in ETSI TR 103 611 [28]**,** whereby the UE accesses to (R)AN via 3GPP or non-3GPP access technologies, and the (R)AN is connected to the 5G Core through a satellite link. The implementation options of this SatCom positioning into 5G can be based on two main approaches:

- **Satellite Terminal as a 3GPP Relay Node:** Based on a concept already introduced in 4G, the satellite terminal is managed by the terrestrial 5G Core as a 3GPP relay node (node considered as a UE with specific roles, that is aggregating the traffic from other UEs and relaying it to the Core network). Different access may provide sub-categories of relay node.
    - Based on 3GPP access: the 5G protocol stack, including NR is implemented over the satellite link.
    - Based on trusted non-3GPP access: the envisaged satellite access technology is identified as trusted by the 5G Core.
    - Based on untrusted non-3GPP access: the envisaged satellite access technology is identified as untrusted by the 5G Core.
- **Satellite as Transport Network**: this corresponds to the natural evolution of the current solution of satellite backhaul. The transport network used transparently provides transport features from one point to another. For 5G, some advanced interfaces are foreseen in order to adapt the transport to the traffic characteristics (e.g. required QoS, priority…) and/or grant some management responsibilities of the transport network to the 5G network. To ease this integration of future transport network, the development of such transport network can be based on 5G specifications.

These architecture options aim to provide innovative evolution paths with respect to the current state of the art of satellite backhaul, which is mainly based on considering the satellite links as pure transport network with limited flexibility and configurability. For 5G integration, innovative adaptation functions are therefore required with advanced features such as backhaul resourcing on-demand, QoS adaptation, support of 5G network slicing etc. to bring real added value satellite-based backhauling and edge delivery solutions. These adaptation functions are currently being investigated as part of the new 3GPP work item documented in 3GPP TR 23.737 [29].

3.6 Interfaces for cross-domain service orchestration

An E2E network infrastructure may consist of multiple administrative domains, each including resources and management and orchestration systems operated by a given partner or administrative authority (e.g. a set of partners). The deployment and operation of services in an E2E infrastructure could bring the involvement of two or more administrative domains, some of them even operating at different abstraction levels. To guarantee effective and coherent service provisioning for this kind of scenario, coordination and exchange of information across management and orchestration systems from different domains is required. To this end, inter-domain secure interfaces may be used.

Many works have addressed the problem of service orchestration involving more than a single administrative domain. Despite the heterogeneity of the solutions proposed so far, two main approaches have been identified: cross-domain service orchestration at the resource and NF layer (section 3.6.1), and cross-domain service orchestration at the application layer (section 3.6.2). The former focuses on E2E orchestration of *resource-facing services* (i.e. NFV network services) enabling communication between NFVOs from different administrative domains. The latter also deals with E2E orchestration, but applied to *customer-facing services*. A customer-facing service is a service that can be directly consumed in the context of an application scenario or use case, and is composed of a set of service functions, including NFs (arranged into one or more resource-facing services) and AFs. Service functions rely on the capabilities delivered by the NFV framework, while AFs provide value-added capabilities focused on the service application. From a deployment viewpoint, a customer-facing service consists of one or more resource-facing services, i.e. one or more NFV NSs. Unlike in resource-facing services, cross-domain orchestration in customer-facing services brings inter-domain communication

between management and orchestration systems that are logically placed on top of the NFVO. Indeed, while NFVO operates at the resource and NF layer (focused on service deployment), the aforementioned systems operate at the application layer (focused on service semantics).

3.6.1 Cross-domain service orchestration at the resource and NF layer

The deployment and operation of NFV NSs across sites that extend two or more administrative domains is key to effectively overcome the potential restrictions (e.g. in terms of coverage, network capabilities, resource capacity, geographical footprint) of a single administrative domain. The orchestration of NFV network services in multi-domain environments brings new implications, particularly in the definition of eastbound / westbound interfaces across NFVOs to securely exchange information and expose management capabilities across domains. Several previous projects have drawn the attention on the definition of these interfaces. Among them, the most significant progress has been captured in ETSI GS NFV-IFA 030 [38], 5GEx Deliverable D2.2 [39], and 5G-Transformer Deliverable D1.2 [40].

ETSI GS NFV-IFA 030 [38] defines the Or-Or reference point. This reference point enables NFVOs from different administrative domains to communicate with each other. The Or-Or reference point assumes that NSs managed by one NFVO (nested NFVO, NFVO-N) are included as constituents of the NSs managed by the other NFVO (composite NFVO, NFVO-C). The processing of nesting NSs (managed by NFVO-N) into composite NSs (managed by the NFVO-C) means triggering network service orchestration functions across both NFVOs. The NFVO-C and NFVO-N exchange information and management services using the following (external facing) interfaces:

- **NSD management interface (NSD management)**: allows NFVO-N to announce NSDs towards the NFVO-C, so that the later could make use of them to build composite NSs.
- **Network Service Life Cycle Management interface (NS LCM)**: defines the set of NS life cycle operations that the NFVO-C can invoke over a network service instance managed by the NFVO-N, when this instance has been nested into a composite NS instance.
- **Network Service Fault Management interface (NS FM)**: defines the fault-related alarms and events that the NFVO-C can collect from a NS instance managed by the NFVO-N, when this instance has been nested into a composite NS instance.
- **Network Service Performance Management interface (NS PM)**: defines the performance information/actions that the NFVO-C can collect/trigger from/towards a NS instance managed by the NFVO-N, when this instance has been nested into a composite NS instance.
- **Network Service Life Cycle Management Granting Operation interface (NS LCM Granting Operation):**allows the NFVO-N managing a NS instance to request a grant for authorization of a life cycle operation over that instance, when it is part of a composite NS instance. This interface enables NFVO-C to approve/reject a NS life cycle operation from the NFVO-N when this may negatively impact the behaviour of the composite NS instance.
- **Network Service Instance Usage notification interface (NS LCM Usage Notification)**: allows NFVO-N managing a NS to receive notifications from the NFVO-C, indicating that this instance has been nested/detached into/from a composite NS instance.

The operations defined in the first four interfaces above can be mostly reused from the Os-Ma-Nfvo reference point defined in ETSI GS NFV-IFA 030 [38], (see Figure 3.15) with the NFVO-C taking the role of an OSS. However, the last two interfaces are completely new, introduced by the dependencies existing between nested and composite network services.

As an alternative, 5G Transformer project defines the So-So reference point in 5G-Transformer Deliverable D1.2 [40] for the interconnection of Service Orchestrators belonging to different administrative domains. The concept of Service Orchestrator (SO) as defined in 5G Transformer is closely aligned with the NFVO functional block. Unlike Or-Or, So-So assumes that not only network service orchestration functions are involved in cross-domain NFV operation, but also resource orchestration functions. Additionally, So-So does not necessarily requires composition/nesting relationships in multi-domain NFV NSs. The set of interfaces defined in the So-So reference point are the following:

- **So-So-Catalogue interface (So-So-CAT):** allows SOs to announce their NSDs and VNF Packages, and trigger certain management actions over them.
- **So-So-Life Cycle Management interface (So-So-LCM):** allows a SO to invoke life cycle management operations on a NS instance in another administrative domain (i.e., managed by another SO).

- **So-So-Monitoring (So-So-MON):** allows a SO to collect performance and fault data from an NS instance in another administrative domain, and trigger monitoring jobs towards that instance.
- **So-So-Resource Management interface (So-So-RM):** allows a SO to invoke management operations over the resources that accommodates a NS instance in another administrative domain.
- **So-So-Resource Management Monitoring interface (So-So-RMM):** offer monitoring means to collect/provide information on the resources across administrative domains.
- **So-So-Resource Advertising Management interface (So-So-RAM):** allows the exchange of network topology and resource information between administrative domains. Although included in the So-So reference point, the capabilities offered in this interface are out of scope of the NFVO's functionality (i.e. is out of the scope of network service orchestration and resource orchestration functionality). This is the point that set the difference between 5G Transformer's SO and ETSI NFV MANO's NFVO.

Finally, 5GEx Deliverable D2.2 [39] defines the I2 reference point. Similarly to So-So, the I2 reference point offer network service orchestration, resource orchestration, and resource advertising capabilities for cross-domain NFV management operations. However, unlike Or-Or/So-So, it assumes the participation of functional blocks beyond the NFVO/SO, including VNFMs and other specific modules.

The interfaces defined in I2 reference point are:

- **I2-Catalog Management interface (I2-C),** allows administrative domains to announce their catalogues, including NSDs and VNF Packages.
- **I2- Service Management interface (I2-S):** allows a NFVO to request network service instances to/from other administrative domain (i.e. to/from other NFVO).
- **I2- Life Cycle Management (I2-F):** allows a NFVO/VNFM to invoke life cycle management operations on a NS/VNF instance in another administrative domain.
- **I2- Resource Control (I2-RC):** allows a NFVO/VNFM to invoke management operations over the resources that accommodates a NS/VNF instance in another administrative domain.
- **I2-Resource Topology interface (I2-RT):** allows the exchange of network topology and resource information between administrative domains.
- **I2- Monitoring (I2-MON):** allows exchange of monitored data between administrative domains.

Despite their differences, the interfaces defined in ETSI GS NFV-IFA 030 [38], 5GEx Deliverable D2.2 [39], and 5G-Transformer Deliverable D1.2 [40] present some similarities. Their tentative mapping is summarized in Table 3.2.

Table 3.2: Cross Domain NFVO Interfaces

Scope	Or-Or reference point (ETSI NFV) [38]	So-So reference point (5G Transformer) [40]	I2 reference point (5GEx) [39]
NS Orchestration	NSD Management	So-So -CAT	I2-C
	NS LCM	So-So-LCM	I2-F
	NS LCM Operation Granting		I2-S
	NS Instance Usage Notification		
	NS PM	So-So-MON	
	NS FM		
Resource Orchestration	Out of scope	So-So-RM	I2-RC
		So-So-RRM	I2-MON
Resource Advertisement	Out of scope	So-So-RAM	I2-RT

3.6.2 Cross-domain service orchestration at the application layer

Discussion on cross-domain orchestration of customer-facing services goes beyond the scope of NFV framework, since it takes into account management and orchestration systems operating at a higher abstraction level than NFVO does. Little progress has been achieved in the context of these systems so far. The most significant work has been carried out by the Metro Ethernet Forum (MEF), with the definition of the Lifecycle Service Orchestration (LSO) Reference Architecture (see Figure 3.21).

The Core element of this framework is the Service Orchestration Functionality defined in the E2E MANO, which is responsible for the orchestration of customer-facing services in a single administrative domain. E2E MANO extends service orchestration at the *application layer*, considering the semantics of the different service functions rather than their realization (i.e. how the underlying virtualized resources supporting these functions are deployed and executed). The latter, known to as resource-facing service orchestration, includes orchestration at *resource* and *NF* layers.

A key role of the E2E MANO is to connect the business needs from the verticals (or other types of clients) consuming the customer-facing services at the application layer to the resource and network capabilities exposed by underlying management domains and partner provider domains. In order to do this it must be able to expose the resource service capabilities of the underlying resource domains and service capabilities of partners and external providers northwards to the customers. It does this by functioning as an abstraction layer, exposing a generic set of interfaces while hiding specific technical and implementation details. These generic interfaces are the business and management interfaces that are exposed northwards to customers and east-west to partners and other operators.

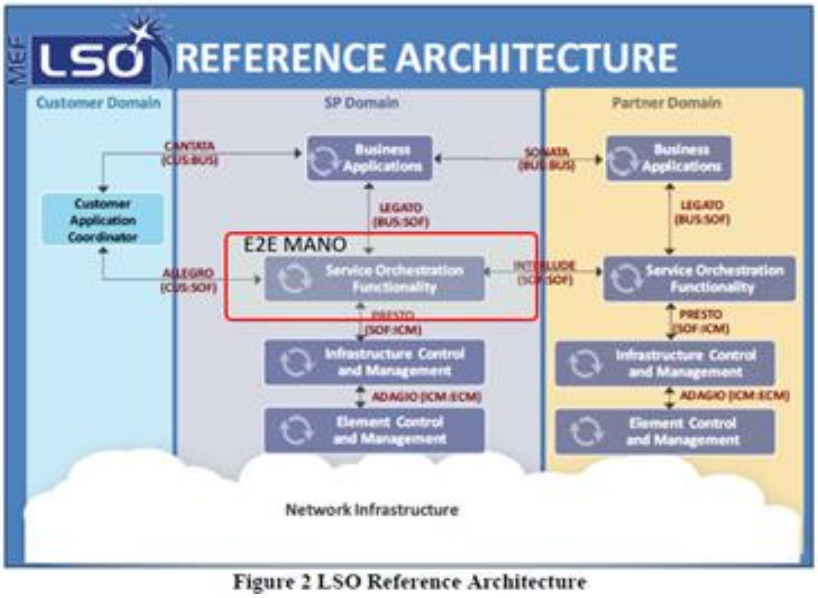


Figure 2 LSO Reference Architecture

Figure 3.21: MEF LSO Reference Architecture

The purpose of the MEF LSO Reference Architecture is to automate the entire lifecycle for customer-facing services orchestrated across multiple domains. For this end, the MEF LSO reference architecture describes the reference points between the Service Orchestration Functionality and the Customer (Customer Application Coordinator), partners and external providers (Partner Domain SOF), and the NFV MANO and domain-specific managers (Infrastructure Control and Management). These reference points are:

- E2E MANO ALLEGRO (CUS:SOF): comprises the services to the Customer Application Coordinator supporting supervision and control of services. This will include request, modify, manage, control, and terminate Products or Services. The services defined in this reference points are implemented as northbound interfaces.
- INTERLUDE (SOF:SOF): comprises the services for the coordination of a portion of LSO services within the partner domain that are managed by a Service Provider's Service Orchestration Functionality within the bounds and policies defined for the service. Includes requests for technical operations or

dynamic control behaviour associated with a Service. The services defined in this reference points are implemented as eastbound/westbound interfaces.

- PRESTO (SOF:ICM): comprises the services needed to manage the network resources and services within the different management domains within the operators network. This would include the NFV MANO and the domain-specific managers. The services defined in this reference points are implemented as southbound interfaces.

This architecture also supports recursivity, whereby the services exposed by one E2E MANO can be consumed by E2E MANO of another operator, which functions as a higher level within a hierarchy of recursive multiple domain-specific management and orchestration entities. This effectively creates a hierarchy of E2E MANO instances, each providing a multi-domain orchestration and management entity to coordinate E2E service creation across the domains that they manage.

3.6.3 Business interfaces

Apart from the interfaces being defined to enable the cross-domain orchestration at the network and application layer, there are also interfaces that facilitate business related processes; we refer to these as business interfaces. Notable examples are the MEF LSO SONATA (BUS:BUS) and CANTATA (CUS:BUS), the TMF Product Catalogue Management/Product Inventory Management APIs and 5GEx's I2-C and I1-C interfaces being enhanced to support the proposed "5GEx Business Layer".

- **LSO CANTATA (CUS:BUS)** is a customer-facing interface that belongs to the family of LSO Reference Points as defined in MEF 55 [41]. LSO CANTATA is the management reference point that provides the customer with business-level capabilities related to the customer's services via a Service Portal. Contrary to LSO ALLEGRO discussed in section 3.6.2, LSO CANTATA supports non-control related management interactions between the LSO defined Service Provider domain and the Customer domain. Examples of interactions that could be performed through LSO CANTATA are:
    - Customer browses the Product Catalog for Product Offerings that are available for selection by the Customer.
    - Based on Product Offerings the Customer develops, places, tracks and changes Product Orders.
    - Customer requests modification of Product Instances.
    - Customer provides and views customer acceptance testing information.
    - Customer views Product Instance performance and fault information.
    - Customer receives information about the scheduled maintenance that may impact their Product Instances.
    - Customer places and tracks trouble reports.
    - Customer queries and views usage and billing information.
- **LSO SONATA (BUS:BUS)** is cross administrative domain interfaces that also belongs to the family of LSO Reference Points as defined in MEF 55 [41]. LSO SONATA defines the management reference point supporting business-level management and operations interactions between two network providers. LSO SONATA supports non-control cross domain interactions between the Service Provider who is facing a Customer and a Partner network provider. LSO SONATA connects between the BSS functions of the Service Provider domain and the Partner domain respectively. Examples of interactions that could be performed through LSO SONATA are:
    - Service Provider browses the Partner's product catalogue for Product Offerings that are available for the Service Provider to select. This may include some geographical and service information to support availability queries of a Product Offering in a specific geographical area.
    - Service Provider develops, places, tracks, and changes Product Orders with the Partner.
    - Service Provider requests modification of Product Instances.
    - Service Provider receives Product Instance performance and fault information provided by the Partner.
    - Service Provider receives information from the Partner about the scheduled maintenance that may impact their Product Instances.
    - Service Provider places and tracks trouble reports.
    - Service Provider exchanges usage and billing information.

- **5GEx's I2-C and I1-C** are the catalogue interfaces defined by the 5GEx project, for cross-domain and customer-facing management of service offerings respectively. The interfaces has been initially defined in [39] and extended in [69] to enable the business process required by the "5GEx Business Layer". Note that LSO SONATA and LSO CANTATA has been promoted as implementation candidates.
    - I1-C enables Customers to consume services offerings that have been pushed by the Service provider and management of slices by the vertical Customer.
    - I2-C enables Service Providers to exchange information about their service offerings, and also facilitates the coordination and negotiation for the creation of multi-provider service offerings.

Interfaces that facilitate the business processes are of high interest for 5G-VINNI, mostly of the vertical Customer side. This is a topic that falls within WP5's interests and will be further investigated in T5.2 and elaborated in D5.2.

4 Template Facility Architecture

At a top level, the architecture for a 5G-VINNI facility site is as shown in Figure 4.1.

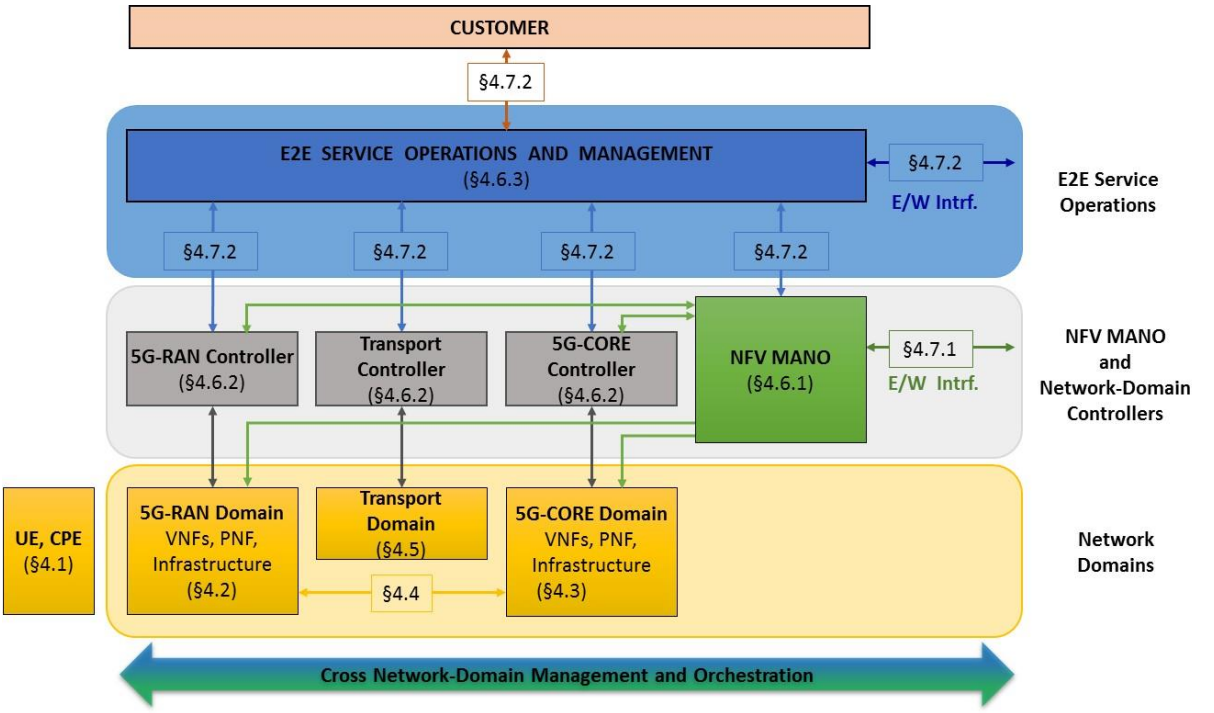


Figure 4.1: Architecture of a 5G-VINNI Facility Site

In the lower layer of the figure are the infrastructure, VNFs and PNFs in the RAN and Core, which are explained in detail on Sections 4.2 and 4.3 respectively, and the infrastructure for the transport domain which is described in Section 4.5.

Above the Network Domains is NFV-MANO and the respective controllers of each domain. NFV-MANO is focused on virtualization-specific tasks (i.e. management at the virtualized resource level), while domain controllers focus on non-virtualization-related operations (i.e. management at the application level). The NFV-MANO is responsible for managing VNFs, combines them in order to set up one or more network service, and in general take care of instantiation, scaling, updating, and terminating VNFs and NSs. Further details of this block will be provided in section 4.6.1.

The Domain Controllers at the RAN and Core are in charge of managing the different NFs at the application level (independently of their deployment), and in general to provide control on all the non-virtualization-related operations. The Domain Controllers at the transport include components such as SDN controllers or Multi-Protocol Label Switching (MPLS) management and control components. Further details of this block will be provided in section 4.6.2.

The E2E service operations and management level is in charge of coordinating the different domain controllers and the network services provisioned by the NFVO, in order to have a harmonic service across RAN, transport and Core. Further details of this block will be provided in section 4.6.3. Interfaces exposed by the orchestration and management layers are described in section 4.7.

Not addressed by this diagram are Security (section 4.8) and Testing and Monitoring (section 4.9).

The architecture can be implemented in a range of ways, which are captured within Chapter 4. Each section identified in the diagram contains aspects that are mandated and aspects that are optional. This allows individual facility sites to implement an architecture that can deliver a set of KPIs and use cases that are specific to that particular site, whilst adhering to the principles and architecture of 5G-VINNI as a whole.

4.1 End-user device

To some extent, end-user devices are outside of the scope of 5G-VINNI. Devices tend to be tied to vertical use cases and as such are more in the domain of prospective ICT-19 projects. It is therefore not of 5G-VINNI, or of this document to be overly prescriptive regarding the characteristics that a device ought to support. However, it is clear that for a device to connect to a 5G-VINNI facility site's infrastructure, it SHOULD support the reciprocal interface requirements to that mandated upon the RAN (see section 4.2).

It is noted that a device has greater levels of optionality inherent within its requirements than the RAN. Where, in section 4.2, the RAN is specified to support multiple options of certain capability, the device only needs to support one of the available options in order to successfully attach to the network. Supporting one such option in the device where that device option is a RAN mandate, will allow the device to connect to any 5G-VINNI facility site. Even if the device supports a capability that is described in section 4.2 as optional in the RAN, there is still a likelihood that the device may be able to attach to some of the 5G-VINNI facility site deployed networks.

Any individual device SHOULD meet the requirements placed upon it to attach to a specific 5G-VINNI facility site, as described in the Facility Site specific design.

Where required, 5G-VINNI Facility Sites MAY support devices that are specific to Use Cases and hence need to provide specific capabilities. Theses MAY include (but are not limited to):-

- devices supporting LTE-M
- devices supporting Narrow Band IoT (NB-IoT)
- devices providing FWA
- devices for V2x
- devices exposing alternative air interface connectivity, using 5G as a back/front haul connection technology
- experimental devices offering functionality for future capabilities (R16, R17 and beyond)
- devices specifically for the purpose of air interface testing.

Where any of these options, or any other alternative access or device type is to be supported, the individual facility site SHALL describe their implementation within WP2 documentation for that Facility site.

4.2 Radio Access Network (RAN)

Initial 5G-VINNI RAN architecture is aligned with what is being planned for early launches of commercial 5G networks in Europe. This will create a 5G research environment that, in some ways, closely reflects functionality and performance offered by early commercial 5G networks. Some of the 5G-VINNI facility sites MAY implement Standalone (SA) architecture from start while other facilities MAY align with the development related to early commercial networks and implement NSA architecture from start and then migrate to SA architecture during the later Releases of the 5G-VINNI project. This mix of SA and NSA-based facility sites will shift according to the timeline of 5G-VINNI and the relevant 5G-VINNI Release.

3GPP Standards for 5G are defined to flexibly handle any use cases under the broad umbrella of eMBB, Massive Machine-Type Communication (mMTC) and Critical Machine-Type Communication (cMTC). 3GPP Standards allow backward as well as forward compatibility.

The 5G-VINNI RAN will initially offer service opportunities in the 3.5 GHz frequency band (3GPP n77/n78, referred to here as 'mid-band') and in subsequent phases also in the 26 GHz frequency band (3GPP n258, referred to here as 'high-band').

The non-standalone (NSA) architecture deployment will be based on Option 3, as identified in section 3.1.2.6.

For NSA implementation the frequency band used by the LTE eNB **MUST** be aligned with available frequencies at the different facilities as well as LTE anchor band support in the 5G non-standalone supporting devices.

The majority of the functional requirements described below are related to the NSA architecture. RAN requirements applicable for SA implementation are briefly indicated in the text.

4.2.1 RAN Functionality

The functions listed below are a collection of requirements that SHALL or MAY be supported in 5G-VINNI project. Specific details of what will be supported at the different facility sites depends on NSA or SA implementation as well as spectrum availability and device capabilities. Many of the requirements listed are vital to fulfil the targeted 5G KPIs and use cases.

LTE – NR Dual Connectivity

LTE – NR Dual Connectivity SHALL be supported at facility sites implementing NSA architecture. LTE – NR Dual Connectivity (EN-DC) is a mandatory part of 5G NSA operation that allows early introduction of 5G by overlaying NR to existing LTE networks, connecting to the 5G enabled Core Network (CN) through the S1 interface. This is shown in Figure 4.2.

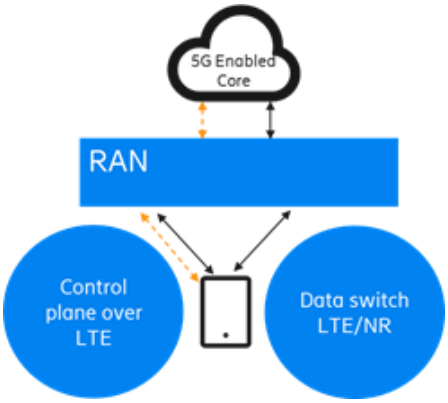


Figure 4.2: LTE-NR Dual Connectivity.

Dual Connectivity enables the UE to have two separate connections, referred to as Split Bearers, one to a Master Node (MN) and one to a Secondary Node (SN). In NSA the MN is always the eNB or Master eNB (MeNB) and the SN is always the gNB or SgNB. The control signalling towards the UE and the CN is handled by the Master eNB.

The X2 interface is used between the MeNB and SgNB and when a Split Bearer is established the user-plane towards UE and CN is terminated in the SN (Secondary gNB) while the LTE user data is transferred via the X2 interface, this is also referred to as Option 3x.

A Split Bearer is established as soon as the UE is in NR coverage and removed when the UE leaves NR coverage or when the Radio Access Bearer (RAB) is released from the Mobility Management Entity (MME), or due to inactivity. To find NR coverage the UE SHOULD support to be configured to do measurements while in LTE coverage, or the eNB can be configured, on a per cell level, to try to do a blind establishment of Split Bearers. Loss of NR coverage is triggered by Radio Link Failure (RLF) detected in the gNB or in the UE.

Mobility

In 5G-VINNI, connected and idle mode mobility SHOULD be supported for both NSA and SA architectures.

Uplink/downlink Decoupling

5G-VINNI NSA architecture SHALL support uplink/downlink Decoupling allowing LTE spectrum (using lower frequency bands than NR) with superior coverage to be used for uplink traffic while NR spectrum with superior peak data rate and latency is used for downlink.

By combining the high speed and low latency of the NR downlink with the larger coverage and high reliability of the LTE uplink the coverage provided by NR spectrum is extended.

LTE-NR Aggregation

LTE-NR Aggregation increases user peak bitrates and app coverage by combining one or more LTE carriers with one or more NR carriers to allow users to experience higher speeds everywhere in the network.

This functionality SHALL be supported for NSA. It ensures that end user bit-rates are always better for a 5G user compared to a 4G user.

Multiple Input Multiple Output (MIMO)

Multi-antenna transmission (also known as MIMO) with a massive number of antenna elements is a key feature for NR systems. Especially for high band, introduction of multi-antenna with a massive number of antennas and highly directional beamforming brings many benefits.

Highly directional beamforming improves data transmission coverage and link performance. When both elevation and horizontal beamforming are available, the maximum beamforming gain can be obtained in each direction. Interference is minimized by highly directional beamforming, which also helps to further improve throughput for cell-edge users. Consequently, beamforming with massive MIMO (M-MIMO) improves network capacity by increasing cell throughput. Furthermore, receiver beamforming helps to improve the link SINR performance.

Massive MIMO and beamforming SHAL**L** be supported in 5G-VINNI for both mid-band and high-band frequencies. It enhances both capacity and data rate coverage and utilizes cell shaping to deliver flexible cell coverage for different deployment environments.

Numerology support

A sub-carrier spacing (SCS) in accordance with 3GPP Rel 15 flexible numerology SHALL be supported. It enables 5G-VINNI RAN system to deliver lower latency compared to today's commercial LTE networks. Flexible numerology allows for other SCS options that MAY be implemented as well.

Carrier Bandwidths NR

Carrier bandwidths in low- and mid-frequency bands are expected to be narrower than in high frequency bands, hence 20, 40, 50, 60, 80 and 100 MHz carrier bandwidths SHALL be supported for mid-band. High-band SHALL support carrier bandwidths of 50 and 100 MHz. Actual implementation needs to be aligned with spectrum availability at the different 5G-VINNI facility sites. Other 3GPP standardized NR carrier bandwidths MAY be implemented.

Modulation

64QAM in uplink and downlink SHALL be supported. 256QAM for downlink transmission MAY be implemented.

Network slice aware RAN

Slicing is an E2E concept defined by 3GPP to allow differentiation of groups of users.

For RAN, a network slice provides additional information on the policy for handling traffic in addition to the regular QoS for radio bearers.

A gNB MAY serve several slices and each slice MAY (or may not) contain users with different QoS requirements.

RAN integration with NR standalone (SA) Core architecture

When 5G-VINNI facility sites implement SA architecture, RAN functionality SHALL be adopted to fulfil applicable requirements. 5G Core has new interfaces and protocols towards RAN and mobile devices. This requires that the RAN and device strategies need to be coordinated with the introduction of 5G Core.

NR QoS Framework

For facility sites implementing architecture with 5G Core, the NR QoS framework SHALL be supported.

5G Fixed Wireless Access (FWA)

Functionality to support FWA application at mid-band and/or high-band frequencies MAY be implemented.

Dynamic spectrum sharing

To allow NR to use same frequencies as LTE, dynamic spectrum sharing MAY be implemented at the different facility sites using legacy LTE frequency bands. Radio access node will dynamically determine if an NR or LTE user is to be served.

Energy Efficiency

Energy efficiency is a key metric within 5G definition. Functionality that will allow for energy efficient operation MAY be implemented at 5G-VINNI facility sites.

Satellite backhaul

Functionality supporting backhaul with extended delay e.g. satellite links, MAY be activated.

Operational efficiency

Network management functions enabling efficient and novel operation of 5G RANs MAY be implemented.

Public safety

Functionality to support mission critical communication use cases in 5G-VINNI MAY be implemented.

cMTC and mMTC

Support for MTC MAY be offered at the facilities supporting NSA architecture by having support for LTE access and the standardized NB-IoT and Cat M radio bearers.

NR Side Link

Side link is an important feature for V2x and is targeted to be standardized in 3GPP Rel-16. 5G-VINNI facility sites implementing use cases for V2x and having access to supporting devices MAY implement NR side link.

Network Slice aware RAN

SA implementations MAY support CN instance and slice selection.

Decomposition of RAN

5G-VINNI implementations MAY employ decomposition of the gNB according to Option 2, Option 7 or Option 8, identified in 3GPP TR 38.801 [33].

4.3 Core

As previously discussed, 5G System as defined in 3GPP Release 15 can be realized in two ways, a) 5G-NR connected to EPC, also referred to as NSA & b) 5G-NR connected to 5G Core, also referred as SA. While EPC used in NSA is an evolution of the current LTE Core network with additional features, 5G Core is a new service-based architecture with new functional nodes.

4.3.1 5G EPC/NSA Core Architecture

NSA implementation SHALL be based on Option 3 architecture, as described in section 3.1.7.2 and expanded upon for RAN specific topics in section 4.2. Figure 3.4 in section 3.1.7.2 illustrates Option 3. This section describes 5G-VINNI specific Core Network aspects of NSA

4.3.1.1 5G EPC/NSA Architecture overview

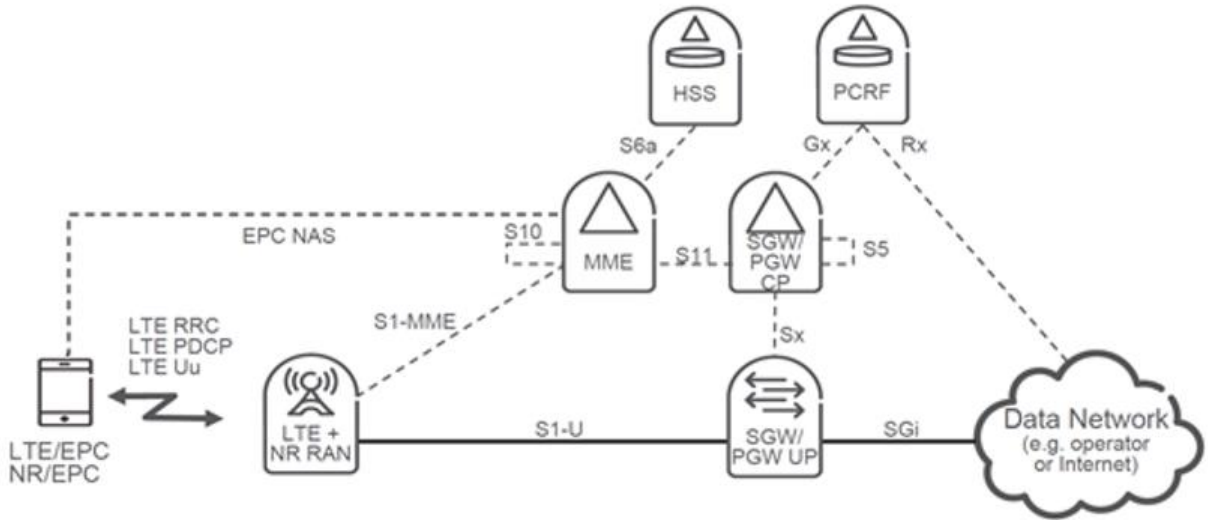


Figure 4.3: functional view of the non-roaming RAN – 5G EPC interaction supporting Option 3

Figure 4.3 above presents the functional view of the non-roaming RAN – 5G EPC interaction supporting Option 3.

The 5G EPC Core components include:

- MME
- Serving Gateway (SGW)
- PDN, Packet data Network Gateway (PGW)
- Home Subscriber server (HSS)
- Policy and Charging Rules function (PCRF)

Apart from the basic features supported by the existing LTE EPC Core, additional features that SHOULD be supported by the LTE EPC to enable the 5G EPC network are as follows:-

5G EPC General

- Option 3 volume usage control and reporting of network performance, statistics, network optimization and recording

5G Enabled SGW/PGW

- Separation of CP and UP for maximum topology flexibility and completely independent capacity scaling.
- Support for individual high data rates for 5G (5/10 Gbps)

5G Enabled MME & HSS

- 5G data rate support
- 5G NR Admission and Resource control
- 5G subscription handling
- Dynamic mobility switching (NR/LTE)
- Network slicing - DECOR, 5G GW selections

5G Enabled PCRF

- Support for policy & QoS control on LTE/NR access

4.3.1.2 Key MME and Evolved Packet Gateway (EPG) features for 5G EPC

As an overview, Table 4.1 presents key functions considered in EPC for 5G Option 3 operation. The symbol (■) indicates affected NFs, implying function/features needed per component.

Table 4.1: key functions considered in EPC for 5G Option 3

Function	MME	EPG
Support of all Options 3/3a/3x	■	
User Throughput Enhancement		■
NR Access Control	■	
Dual Connectivity LTE/NR	■	
CP/UP separation	■	■
NR Volume Reporting	■	■
GW selection for NR usage	■	
DECOR Slicing using NR Capability	■	
QCI support for V2X applications	■	■
Management and Orchestration (NFV-MANO)	■	■

- General support of Option 3/3a/3x. MMEs need to support necessary signalling procedures for mobility and session management for Options 3/3a/3x.
- User Throughput Enhancement. To support higher capacity peak rates, particularly in 5G, EPG SHALL support high bit rate per user/UE.
- NR access control. This includes the 5G-enhanced attach and mobility accept procedures, based on reported UE capability (NR), UE's subscription information provided by the HSS, IMSI range and other selection criteria. UEs not fulfilling these criteria will be allowed to attach to LTE only.
- Dual Connectivity LTE/NR. This function is needed for 5G Option 3 operation. Here, MME assists RAN with necessary UP tunnel switching procedures, for instance at UE mobility or congestion situations.
- CP/UP separation: CUPS is a central part of optimizing the UP architecture in Network Slicing for 5G EPC LTE/NR. The EPG MAY be deployed as a GW-C or as a GW-U by configuration, which enables a flexible network architecture for intended service needs. The MME provides the EPG/GW-C with UE 5G capability (NR) information. The EPG/GW-C MAY add this information to the general criteria for GW-U selection. This adds access specific flavours and more service granularity to the slicing concept.
- NR Volume Reporting. With the latest evolution in 3GPP, RAN will report UE reporting per UE and per bearer. This information is much wanted, for network performance monitoring, statistics, trouble shooting and optimizations. The reports are sent from RAN over the CP to EPC. Reporting is made at mobility or when the UE goes to idle, that is, in conjunction with normal signalling. By this, no signalling load is added.
- GW selection for NR usage enables the possibility of - using a common 5G EPC/MME for all kinds of services - select GWs considering access (LTE/NR). This is an addition to traditional GW selection procedure for the S-GW and the P-GW. It is a simplification compared with the full Core Network division by DECOR but does not provide full isolation of MMEs for different services.
- Dedicated Core Network (DÉCOR) slicing using UE NR Capability. This is an enhancement to the DECOR function, implying that the eNB together with the MME selects dedicated Cores also considering access (LTE/NR). This decision is based on UE 5G Capability indication, added subscription data in the HSS, and DNS record enhancements.
- QCI support for Vehicle-to-Everything (V2X): The next generation of intelligent transport systems (ITS) will include V2X. Such systems MAY involve remote management of a fleet of vehicles, where a remote operator or an automated system is tracking, monitoring and potentially exercising some control over vehicles. This requires special QoS requirements, including support of dedicated QCI 75 (Guaranteed Bit Rate (GBR)) and 79 (non-GBR). See also 3GPP TS 23.203 [32].
- NFV-MANO – requirements for support of NFV-MANO are covered in section 4.6.

4.3.1.3 Other NSA Core Functions

The below described functions MAY be implemented by individual sites based on the needs/requirements, of the service use cases offered by the site and based on the kind of the architecture supported by the vendor partner in the site.

Service capability exposure function (SCEF): as mainly driven from IoT use cases, 3GPP SCEF could additionally be introduced for facility sites focused on serving IoT use cases. This node could eventually be evolved to 5G Core NEF.

4.3.2 5G Core/SA Core Architecture

5G Core architecture, as defined in the 3GPP Release 15, is described in section 3.1.2.2 and illustrated in Figure 3.2.

5G Core components are listed in section 3.1.2.1. Not all of the components mentioned are mandatory for all facility sites as some of the functions SHOULD be introduced based on the site facility use case fulfilment. It Is expected that individual facility sites will define their architecture in WP2 to meet intended use cases, and, where needed, incorporate sufficient flexibility to add other 5G Core functions if ICT-19s require.

4.4 Core to RAN interfaces

5G definitions of interfaces

5G System as per 3GPP Release 15 (R15) and later releases consists of Next Generation RAN and Core Network, and foresees NSA and SA operation. Details can be found in 3GPP TR 38.801 [33].

The Next Generation RAN (NG-RAN) represents the newly defined RAN for 5G, and provides both NR and LTE radio access. NG-RAN node is either:

- A gNB (i.e. a 5G base station), providing NR user plane and CP; or
- An ng-eNB, providing LTE/E-UTRAN services.

3GPP TR 38.801 [33] identifies the options for RAN to CN connectivity, which are described in 3.1.2.6. As discussed previously, initial 5G-VINNI implementations will use Option 3/3a/3x.

Table 4.2 illustrates the mandatory interfaces required by each 5G-VINNI operational option.

Table 4.2: Interface requirements for architectural options

Mode	**Options**	**Interfaces**											
		UE- RAN		UE - CN		RAN-RAN		RAN -CN					
		Uu	NG-Uu	NAS	N1	X2	Xn	S1-C	S1-U	N2	N3	SBI	PtP
SA	Option2		M		M		O			M	M	O	O
NSA	Option 3	M	M	M			M CP, UP	M eNB	M eNB				
NSA	Option 3.a	M	M	M			M CP	M eNB	M eNB, gNB				

4.5 Transport infrastructure

Transport is the part of the network that comprises the intermediate links between the Core Network or backbone and the small sub-networks at the "edge" of the entire hierarchical network. Backhaul plays a vital role in mobile networks by acting as the link between RAN and the Core.

In 5G-VINNI, wireless, optical and satellite links will be used for the backhaul section of the network.

4.5.1 Wireless backhaul

Wireless technologies with systems operating at microwave and lately millimetre-wave frequencies, provide operators the effective options when it comes to implementing combined macro-cell and small-cell backhaul. Wireless-based backhaul can reach any area, while offering:

- High Capacity – Up to multiple Gbit/s per link;
- High Reliability – Up to 99.999% link availability;
- Low CapEx;
- Fast deployment;
- High levels of flexibility, using the right topology mix tailored to network needs.

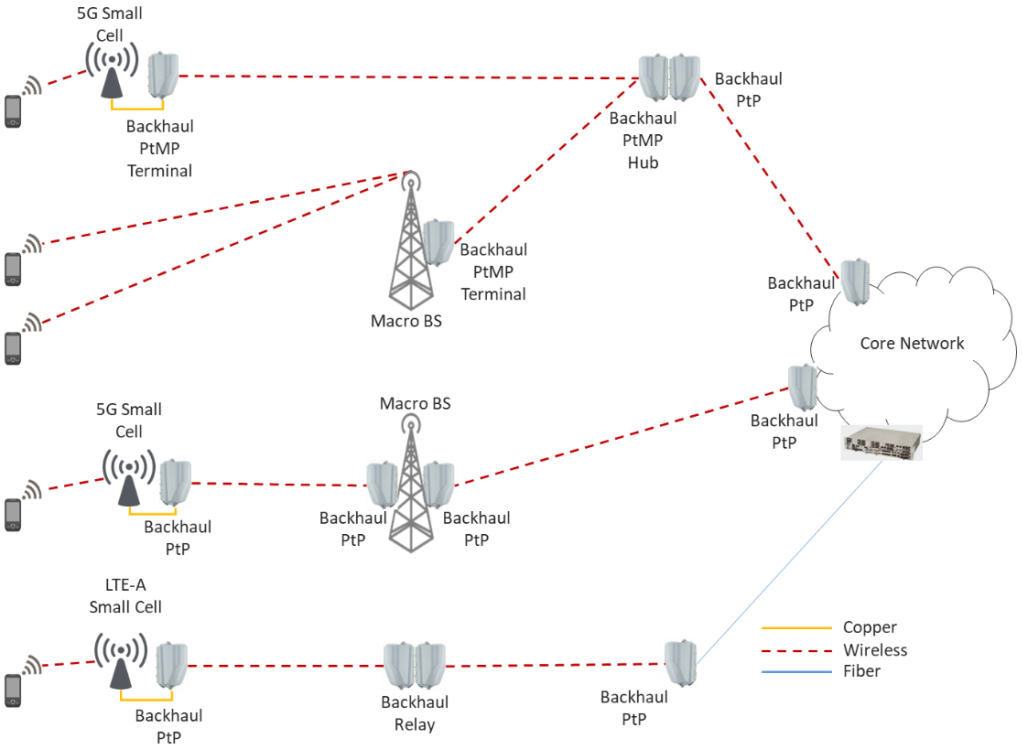


Figure 4.4: Wireless Backhaul Network with PtP and PtMP Configurations

Packet microwave or mm-wave technologies are increasingly prevalent elements of Heterogeneous Networks (HetNets), offering ample IP capacity for backhaul. In Figure 4.4 two different backhaul topologies are shown. Point-to-Point (PtP) systems can also be used as relays (repeaters) in a multihop backhaul network. Point-to-multi-Point (PtMP) systems have a Central Station (Hub) that connects to a number of Terminals. Each backhaul equipment is connected to an access Base Station, as shown in Figure 4.4, either a small cell or a macro-BS.

Although the backhaul network could be logically abstracted as a straight pipe between the access and the Core network, in reality it constitutes an aggregation domain which can have a complex hierarchical structure, considering that a high number of access base stations are aggregated at a certain point of the Core network. Various topologies can be used in a backhaul deployment, depending on each specific network's need, e.g. point-to-point line, tree structure, mesh, ring and combinations of these. Reliability can be achieved by using redundancy in equipment and appropriate topologies, e.g. a ring.

In 5G-VINNI, the wireless transport network can provide a versatile solution for 4G/5G RAN backhaul as a fibre substitute, as well as transport for IoT applications. A common orchestration platform for RAN, transport and Core will provide E2E network slice establishment, enabling the support of multiple use cases over a common infrastructure.

4.5.2 Optical Backhaul, Midhaul and Fronthaul

In 5G-VINNI an optical transport solution, a PON solution, based on NG-PON2, can be used to support backhaul, midhaul and fronthaul alternatives. Figure 4.5 shows how the 5G entities (CU, DU RU), as identified by splitting the gNB according to the options defined in 3GPP TR 38.801 [33], and described in section 3.1.5, MAY be placed on the network. Different splits in the gNB functions result in different requirements for bandwidth and latency, and different transport alternatives. These are shown in Figure 4.5.

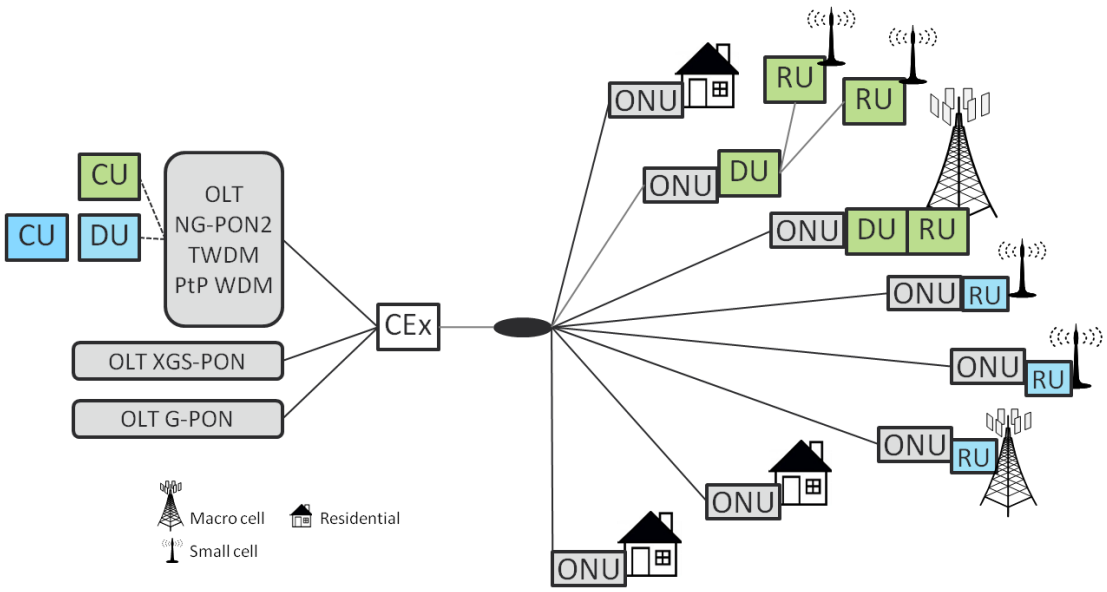


Figure 4.5: PON Architecture for 5G Fronthaul and Midhaul Transport (Source: ITU-T SG15 – ZTE contribution to "PON use cases for 5G wireless fronthaul")

Different combinations of elements MAY be done, resulting in different transport alternatives, as per Figure 4.6.

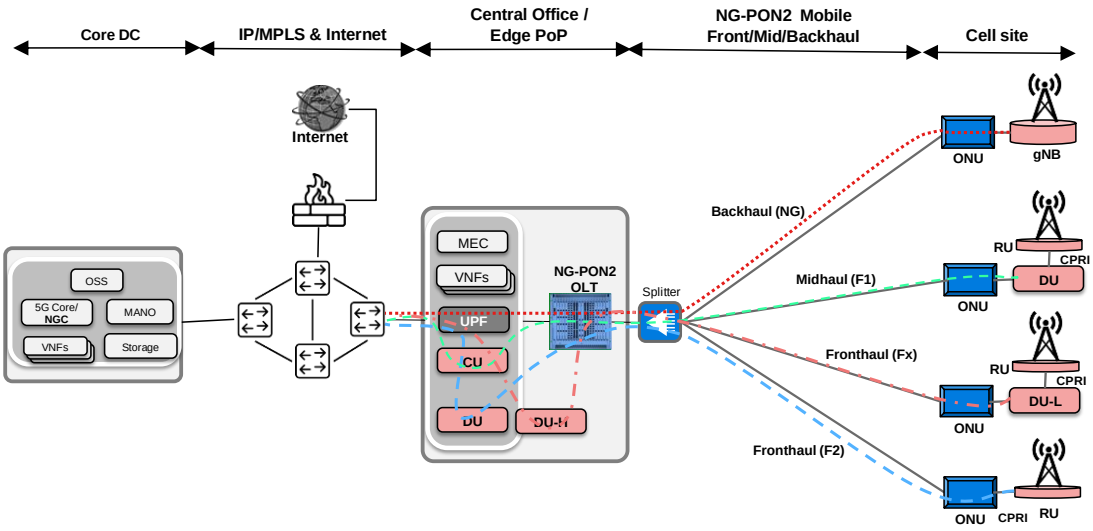


Figure 4.6: 5G Deployment Reference Architecture over a NG-PON2 infrastructure

In 5G-VINNI, NG-PON2 infrastructure is available to cover the Backhaul transport option, as described above. The Midhaul and Fronthaul (Fx) options are currently under development (outside the scope of 5G-VINNI) and are foreseen to be integrated in the facility infrastructure along the course of the project.

4.5.3 Satellite backhaul

MNOs likely perceive satellite backhauling services as islands of costly, difficult to integrate and inflexible connectivity. However, with the significant recent advances in satellite technology (e.g., High-Throughput Satellite) satellites can deliver cost-effective high-performance solutions to unserved and underserved areas and MNOs can leverage satellite solutions even more efficiently than before.

Moreover, the lack of standardised Ethernet service support, including the visibility, control and performance usually associated with terrestrial backhaul solutions, inhibits an MNO's ability to ensure consistent and uniform architectures and rollout planning. It also deprives operators of certainty in subscriber take up rate, and a standard, high-quality experience across the entire network to lay the groundwork for increased Average

Revenue Per User (ARPU). However, by adopting industry standard Ethernet service constructs and orchestration, it is possible for a satellite-based backhaul solution to plug seamlessly into an MNO's backhaul landscape – in the same manner as any terrestrial solution does. With the inter-carrier visibility and automation solution, coupled with the use of Metro Ethernet Forum (MEF) compliant LSO, it is possible for an MNO to turn every bit transported over satellite into a productive bit with no stranded capacity. Satellite players will 'plug into' the MNO ecosystem and become a true enabler of value-based outcomes.

As can be also seen from Figure 4.7, this is the approach adopted by SES Networks.

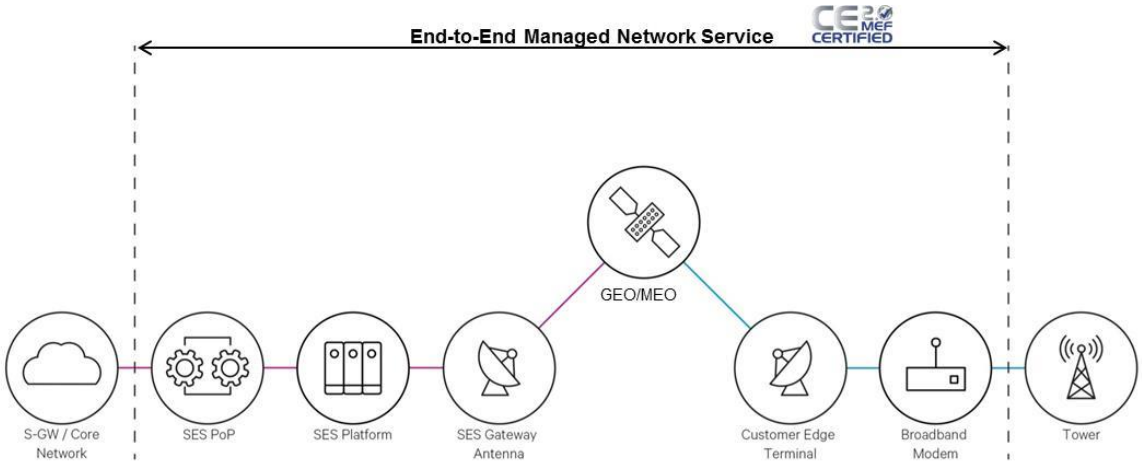


Figure 4.7: SES's GEO/MEO-based satellite backhaul offerings

In terms of SatCom integration into 5G, the 5G-VINNI project will focus primarily on the satellite backhaul solutions**.** A satellite backhaul connectivity deployment includes an edge node and a central node connected using a satellite backhaul link. The UEs connect to the edge node which connects to the central node through a GEO/MEO backhaul link. The backhaul is seen as a transport layer for the messages between the edge and the central node. Because of this, the backhaul SHOULD be as transparent as possible, while at the same time being able to assure a guaranteed communication quality. This can be configured statically or dynamically through the specific management interfaces. The backhaul connectivity architecture along with the relevant NFs is depicted in Figure 4.8.

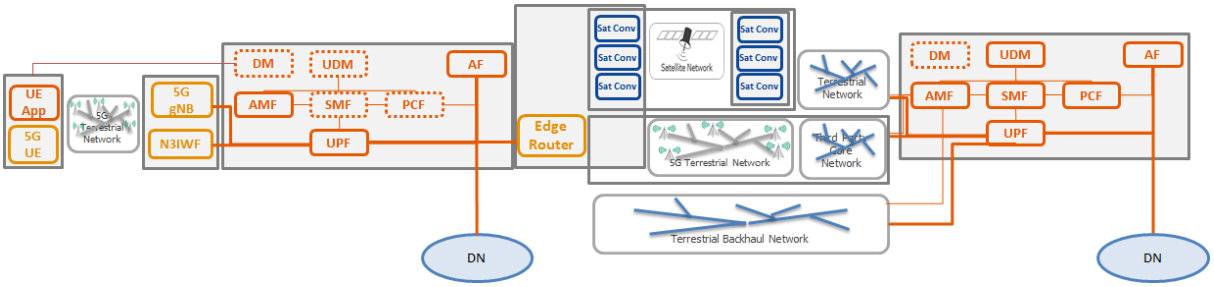


Figure 4.8: Backhaul Connectivity Architecture

The UE connects to a local terrestrial access network which is represented by either a 5G gNB (for 3GPP access) or by the Non-3GPP Interworking Function (N3IWF). In order to assure the connectivity to the local terrestrial access network though the different backhauls, a 5G Core Network is deployed with a functional split between the edge and the central nodes. The satellite convergence functionality includes all the functions which enable an efficient communication across the satellite network such as TCP PEP, QoS classification, compression, packets' formatting, broadcast, etc. The satellite convergence functionality is present in the remote satellite terminal which is part of the edge node as well as part of the satellite hub platform.

One of the NFs associated to the backhaul connectivity architecture is the Edge Router. The architecture includes this NF which acts as an SD-WAN router able to determine which backhauls are available and to steer the data traffic according to a set of routing policies over the backhauls. As this intermediary node represents the interaction across different types of backhaul, the Edge Router SHOULD include the security associations to be able to forward messages across the backhaul as well as other mechanisms necessary for proper data

packet packaging and reliability of communication. The Edge Router is able to connect to the different backhauls using the specific communication technologies: mobile wireless terrestrial backhaul; wired and fixed wireless terrestrial backhaul; and satellite backhaul.

4.6 Management and Orchestration

Management and Orchestration is a key capability for managing network slices. Once the underlying programmable infrastructure has been established, there is a need for components that can use the programmability exposed by those components to manage and orchestrate the resources within and across domains, as described in 5GPPP Architecture Working Group, 5G Architecture White Paper [14]. These domains constitute a collection of systems and networks that interoperate in a stable manner, and that are managed by a single organisation or administrative function as per ETSI GS NFV-MAN 001 [16]. The administrative function could be a network domain delineated by a technology boundary within a network or service operator, or it could be an operator domain delineated by the organisational ownership of network domains.

The 5G-VINNI reference architecture will provide orchestration and management functions within and across domains to facilitate zero-touch network and service management. These orchestration and management functions are:

- NFV MANO: focuses on the management and orchestration tasks that have an impact on the resources on top of which 5G-VINNI services are deployed and executed. The scope of NFV-MANO in a 5G-VINNI facility will also include the management and orchestration of the network services running on the resources.
- 5G-RAN Management Systems
- Transport Management Systems
- 5G-Core Management Systems
- E2E Service Operations and Management

E2E Service Operation and Management, within the 5G-VINNI architecture, encompasses the management and operations of E2E services that span multiple domains. It translates service definition / Service Design into configuration of resources (physical and virtualized) needed for service establishment, using orchestration to coordinate between domain controllers, the NFV-MANO, and other providers. In this context, E2E services MAY span multiple network domains provided by different administrative entities (e.g., different network service providers or external partners). Additionally, the E2E Service Operations and Management triggers the components of the domain management systems and NFV-MANO to apply the configuration of the required resources. In some cases, this could result in their actual allocation, as described in 3GPP TR 28.801 [34].

The 5G-VINNI reference architecture establishes a separation between network domain management (Network Domain Controllers and NFV MANO) and E2E service management (E2E Service Operations). Network domain controllers and NFVO MANO are grouped within network domain management and E2E Service Operations and Management are grouped within E2E Service Operations.

4.6.1 NFV MANO

From a resource management point of view, 5G-VINNI services are deployed and operated as NFV NSs. This makes NFV MANO a key component of the 5G-VINNI facility. With NFV MANO, facility operators are able to deploy and execute 5G-VINNI services in the NFVI with great agility and flexibility. The NFVI in 5G-VINNI facility span across seven facility sites (four main sites and three experimental sites), and consists of different PoPs for VNF deployment and execution, ranging from cell sites to edge/regional data centers.

Each facility site will have its own NFV MANO stack to manage and orchestrate NFV NSs within its administrative domain. As seen from Figure 4.1, NFV MANO exposes a set of SBIs and NBIs to facilitate the interaction with the rest of facility site components. On one hand, NFV MANO makes use of the SBI to manage (and collect information from) the site's NFVI resources on top of which 5G-RAN and 5G-Core VNFs run. On the other hand, NFV MANO uses the exposed NBIs to provide/receive information and operations that have an impact on the NSs under its management. NBIs facilitate the interplay between the NFVO and the rest of management blocks, including 5G-RAN and 5G-Core controllers (Section 4.6.2) and E2E Service Operations and Management (Section 4.6.3). From the perspective of NFV MANO, these management blocks take the role of NMS in the NFV reference framework (see Figure 3.15), so NBIs can take the role of Os-Ma-nfvo and Ve-Vnfm-em reference points.

In addition to the above-referred NBIs/SBIs, the NFV MANO stack deployed at each facility site SHALL have the three types of functional blocks (NFVO, VNFM, and VIM) and the four data repositories (NS Catalog, VNF Catalog, NFV Instances Repository, and NFVI Resources Repository) specified in Section 3.3.2. This guarantees interoperability across 5G-VINNI facility sites at NFV level (see section 4.7.1), regardless of any implementation-specific issues. Indeed, the implementations of NFV MANO could vary from site to site. The fact that ETSI NFV defines NFV MANO stack as a reference framework not only allows network operators and vendors to implement NFV MANO components using open source or vendor-specific solutions, but also to introduce some architectural modifications/extensions, as long as they do not conflict with ETSI NFV specifications. Significant points of discussion have emerged from these modification/extensions, leading to different architectural options for NFV MANO implementation. Some of these architectural options are briefly discussed below:

- **Multi-VIM support**: multiple VIMs (from different vendors, or of the same vendors with different versions) can coexist within the same facility site. This capability allows defining different types of hypervisors and environments for VNF deployment, ranging from classical virtualization to cloud-native environments.
- **SDN integration**: To introduce dynamicity and agility in connectivity management, at both the VNF level (i.e., VNFC-VNFC connectivity) and at the NS level (i.e., VNF-VNF/PNF connectivity), a VIM can also include an infrastructure SDN controller. This SDN controller SHALL enable changing infrastructure behavior on-demand, in a programmatic manner, according to the instructions sent from the VIM.
- **VNFM architectural approach**: The VNFM includes some variants, depending on its management scope (i.e. how many different VNF types the VNFM is able to serve) and management capabilities (i.e. the life cycle/performance/fault management operations the VNFM is able to carry out over the VNF instances). Today, two main architectural approaches are envisioned for a VNFM: *generic VNFM (VNFM-G),* able to manage the life cycle of a wide range of VNF types (potentially from different vendors) with standard basic management capabilities; and *application-specific VNFM (VNFM-S)*, that manages the lifecycle of a given VNF type (usually a vendor-specific VNF), but with advanced management capabilities, based on specific knowledge of the VNF applications. The selection of a VNFM-G or a VNFM-S depends on various factors, including VNF complexity, operational readiness, and skills availability.
- **NFVO functional split:** NFVO can be decomposed into its two main functionalities (see Figure 3.16), each implemented as a different orchestration block. On one hand, there is the Resource Orchestrator, responsible for executing NFVO's resource orchestration functionality. On the other hand, the NS Orchestrator, performing the NFVO's NS Orchestration functionality.

4.6.2 Network-Domain Controllers

4.6.2.1 RAN and Core Domain Controllers

The Domain Controllers at the RAN and Core are in charge of managing the different NFs at the application level (independently of their deployment), and in general to provide control on all the non-virtualization-related operations. Core & RAN domain controllers can be mapped to the EMS function group of FCAPS (Fault, Configuration, Accounting, Performance and Security management) for the functional/application part of the underneath deployed PNFs and VNFs. In this way we can say that these domain controllers perform changes at the application level in a "vertical" way on functions "horizontally" deployed by the NFVO. For instance, signalling issue at the mobile Core, abnormal releases at the RAN, or in general any potentially anomaly detected in terms of accessibility, retainability, integrity, availability, or mobility as defined in ETSI TS 132 450 [35], will be addressed by the RAN and Core Domain Controllers.

4.6.2.2 Transport Domain Controllers.

The Domain Controllers at the transport include components such as SDN controllers or MPLS management and control components. In 5G-VINNI the Transport network will provide advanced backhaul functionality, to interconnect for instance different locations from a common site.

MPLS networks for instance are well-known solutions used for several years in the transport network. The process of sending a packet through the MPLS network is the following: First, the Forwarding Equivalence Class (FEC) is defined. The FEC is a group of packets with similar characteristics (IP address source, IP address destination, Port, etc). Second, a path between the source and the destination is found using routing algorithms. This path is distinguished on each router through the use of one specific label. Finally, the first

router on the MPLS domain labels each of the incoming packets in order that they can be clearly identified and sent through the predefined path. MPLS allows the visualization of the E2E path instead of a hop by hop vision.

Figure 4.9 presents a simplified sketch of the SDN architecture based on the concepts presented in the IETF RFC 7426 [36].

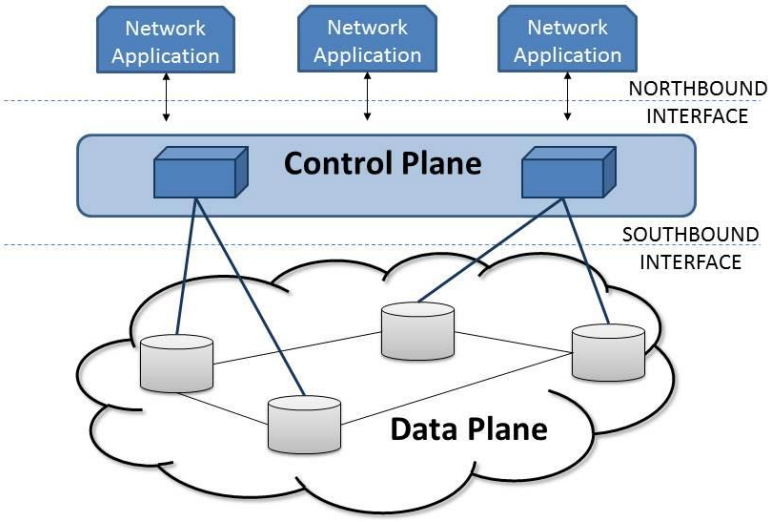


Figure 4.9: SDN architecture

The CP and the data plane are separated from each other. In addition, the CP is logically centralized in a software-based controller defined as the "network brain", while the data plane is composed of network devices ("network arms") that forward packets. The CP includes both northbound and southbound interfaces. The northbound interface provides a network abstraction to network applications, and the southbound interface (e.g., OpenFlow) standardizes the information exchange between the control and data planes.

4.6.3 E2E Service Operations and Management

Delivering a network slice will involve multiple domains both within an operator and across operators and providers, as illustrated in Figure 4.10. Delivering the slice will involve the coordination of service requests to these domains. E2E service operations and management (E2E MANO) manages the coordination across these domains for the delivery and management of network slices.

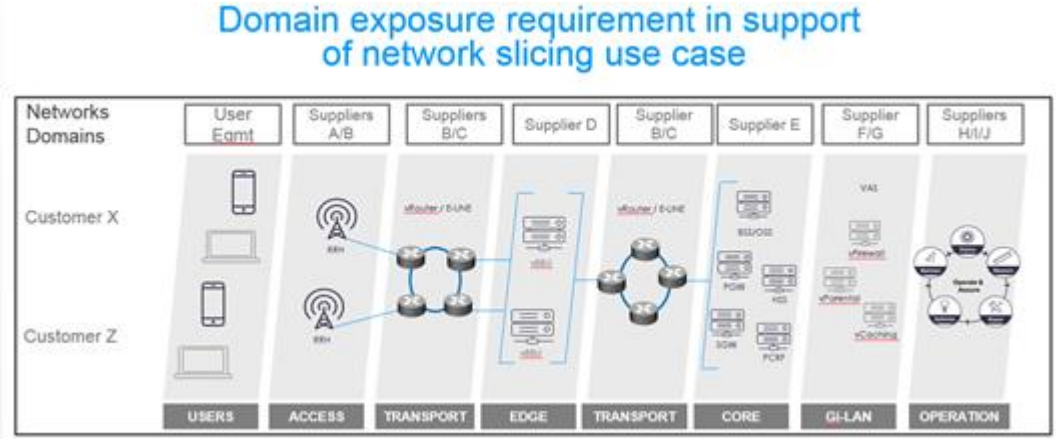


Figure 4.10: Domain Exposure requirements

Effectively, E2E MANO connects the business needs of verticals to the resource and network capabilities exposed by management domains and partner provider domains. It thereby enables orchestration and management across multiple domains and providers. It does so by offering business, or customer facing, services northwards via standardised API's to fulfil the service requirements of customers and verticals. The E2E MANO translates these business services into their component network services and requests the services from the appropriate domain. Where the component network service is exposed by a southbound

management domain, the E2E MANO interfaces to the management domain via southbound interfaces to request the network service; where the component network service is exposed by a partner or another operator the E2E MANO interfaces east/west to request the service from the service provider. In this way the E2E MANO hides how the business services are realised at the resource level, thereby allowing verticals to focus on their service requirements, Service Level Agreements (SLAs) and etc, rather than upon service implementation.

There is no single standard definition of E2E MANO, although its functions are described in multiple standards including 3GPP, MEF, ZSM and TMF. The E2E MANO within the 5G-VINNI reference architecture builds upon these references, and upon the work of other H2020 projects (such as H2020 5G-Transformer Project's Vertical Slicer)

Within 5G-VINNI the E2E MANO resides in the E2E Service Operations layer between the Customer layer and the NFV-MANO and Network Domain controllers. Although the E2E MANO can provide multiple functions the 5G-VINNI reference architecture will concentrate on the E2E Service Orchestration function.

The E2E Service Orchestration function is the set of service management layer functions supporting automation of the service lifecycle management capabilities providing coordinated E2E management and control of 5G network slice services. It is [MEF, Lifecycle Service Orchestration (LSO): Reference Architecture and Framework (2016) [37]. It is responsible for the coordination of the provisioning, configuration, and lifecycle management of the various resources across management domains in the E2E network that make up an E2E service, as described in ETSI GS ZSM 002 [23].

These E2E Service Orchestration functions have to provide process support within the processes associated with E2E lifecycle management of the business service (as opposed to network services), as per MEF, Lifecycle Service Orchestration (LSO): Reference Architecture and Framework (2016) [37]. These include:

- Product and Service development lifecycle management
- Partners on-boarding
- Order Fulfilment, Service Configuration and Activation
- Service Design and Delivery
- Service Control and Management
- Service Testing

From an information perspective E2E Service Orchestration manages E2E service specifications that are exposed to verticals and the hierarchy of service abstraction and decomposition of these into their constituent service components and the mapping of these to their respective management domains.

Based on ETSI GS ZSM 002 [23], *"The services in E2E service orchestration are responsible for the catalog-driven E2E orchestration of multiple management domains to create / modify / delete cross-domain customer-facing services. A service model defines how the various pieces of a service are linked together and map to management domains"*. Thus, the E2E Service Orchestration can be logically divided into sub-components that MAY include:

- E2E Service Catalogue
    - SHALL onboard component specifications from third party sources (e.g. resource layer), assembling the design itself (e.g. layout, parameters, transactions, policies, etc.) to validating and publishing to northbound BSS functions such as Customer Relationship Management (CRM) and Web portal is carried out here. The tool allows for collaboration across multiple designers.
    - To enable automation in such orchestration and management for service/slice deployment and (re)configuration functionalities, the MP SHALL be able to ***receive from the customer and describe the services and their requirements*** in sufficient detail to reflect the requirements of the customer. Based on the service requirements provided, Network Service Chaining or Service Function Chaining (SFC) can be employed to create a complex service provisioned jointly by a series of connected NFs. In a service chain, the NFs need to be joined in the required sequence according to the work flow and configured carefully in order to meet the defined specific service requirement. Building such a service chain without softwarisation requires considerable efforts for the administrator or a non-softwarised MP. Furthermore, reconfiguration and over-provisioning are often required to deal with growing demand over the time. 5GPPP

- E2E Active Inventory
    - A service repository which SHALL persist service instance representations and pertaining resources, to achieve this task the module includes a model that represents the most common elements of services and their virtual and physical resources. The model is meant to be a flexible, easily adaptable (e.g. online) to future demands. There is an acknowledgement that there are potentially multiple sources of significant volumes of information dispersed across the architecture. It is a futile exercise to duplicate this data to centralize it in a single persistence entity. Instead, the Service Repository holds references to the sources of information and when requested, it collects and merges them providing a uniform response to the caller. This technique is known as federation
    - When a service has been instantiated, a representation is held in 5G-VINNI e2e Service Operations domain. This information MAY be accessible to the end user through a UI. It contains details concerning general statistics of the services themselves (e.g. subscriber, allocated services, subscription types, etc.)
- E2E Service Process Manager
    - The SOM module plays a vital role in the E2E orchestration. It provides the functional component for orchestrating activities related to service operations and management. These include
    - Scheduling, assigning and coordinating Customer provisioning related activities including service creation / modification / move / deletion request(s) based on specific Customer orders;
    - Tracking of the execution process including assigning and tracking Service Component provisioning activities
    - Enriching or modifying request/order information under execution
    - Cancelling a Customer order when the initiating sales request is cancelled;
    - Monitoring the jeopardy status of Customer orders
    - Indicating completion of a Customer order
    - Verifying whether specific Service Request sought by Customers are feasible;
    - Decomposition of the Service into Service Components;
    - Allocating the appropriate specific service parameters within each Service Component to support service requests, control requests, or requests from other processes;
    - Reserving specific service related resources (if needed) for a given period of time until the initiating Customer order is confirmed, or until the reservation period expires (if applicable);
    - Configuring specific services, as appropriate;
    - Recovery of specific services;
    - Updating of the Service state information to reflect that the specific service has been allocated, modified or recovered;
    - Coordinating execution of the service delivery orchestration plan, delegating and tracking the actual Service Components implementation to Network Domain Controllers (e.g., sub-network connectivity), NFV MANO and external providers or partners

4.7 External facing interfaces

5G-VINNI facility consists of multiple administrative domains, each within the scope of a given facility operator. E2E service provisioning in 5G-VINNI facility MAY bring the involvement of different administrative domains, some of them even operating at different abstraction levels. To address this issue, external facing interfaces are needed. The definition of external facing interfaces brings multi-domain support and cross-domain interoperability in 5G-VINNI. External facing interfaces allow functional blocks from different 5G-VINNI facility domains to exchange information and services in the 'producer-consumer' communication model. As seen from Figure 4.1, and compliant with the approaches presented in Section 3.6, external facing interfaces can be defined at two levels: at the NFV MANO level (enabling cross-domain service orchestration at the resource and NF layer), and at the Service Operations and Management Level (enabling cross-domain service orchestration at the application layer). Further details on these interfaces will be addressed in Sections 4.7.1 and 4.7.2.

4.7.1 Interfaces at the NFV MANO level

In 5G-VINNI facility, the external facing interfaces at the NFV MANO level are defined between NFVOs. This means that the NFVO is the only functional block from the NFV MANO stack involved in the exchange of information and management services between different administrative domains (i.e., between different 5G-VINNI facility sites). The capabilities offered by 5G-VINNI external facing interfaces include both NS orchestration and resource orchestration functions, and can be arranged into the following interfaces:

- **Catalogue Management interface:** allows an NFVO to expose information and operations on the NSDs and VNF Packages under its management towards other NFVOs.
- **LCM interface:** allows an NFVO to invoke life cycle management actions towards NS instances managed by other NFVOs.
- **Monitoring interface:** allows a NFVO to perform monitoring actions over NS instances managed by other NFVOs.
- **Resource Management interface:** allows a NFVO to perform management actions over virtualized resources belonging to other facility sites, supporting NS instances managed by other NFVOs.

The high-level capabilities that an NFVO (from a given facility site) allows towards other NFVOs (from other different facility sites) via these interfaces are specified below. Some of these capabilities are recommended in each 5G facility site, so as to achieve a minimal level of interoperability for NFV NS orchestration across domains. However, others capabilities define advanced features that entail significant complexity. Although desirable, these features result in having cross-domain NFV operations that could be difficult to achieve in some cases. For this reason, their implementation in a 5G-VINNI facility site is optional, resulting in complementary interface elements. Whether mandatory or optional, it is assumed that these capabilities define operations involving NS instances that are deployed from NSDs (and VNF Packages) that the NFVO publicly exposes towards the other facility's NFVO, according to the cross-domain privacy policy agreed among all the 5G-VINNI facility sites.

4.7.1.1 Recommended interface elements

To enable cross-facility NFV management and orchestration, each NFVO SHALL at least allow other NFVOs to (i) collect information from NS instances running within its own administrative domain, and (ii) participate in the deployment and operation of those NS instances. Assuming that these NS instances are based on NSDs (and VNF Packages) shared across different administrative domains, the set of operations that an NFVO is able expose towards other NFVOs through the different 5G-VINNI external facing interfaces are the following:

- **Catalogue Management interface**: query NSD Info operation, query VNF Package Info operation, and subscription/notify operations to get informed about changes of NSDs and VNF Packages. **LCM interface**: instantiate/query/terminate NS operations n, and subscription/notify operations to keep track of the most remarkable lifecycle changes of operative NS instances (e.g. NS instantiation, termination) and to gather basic information of run-time NS modification (e.g. NS instance has been modified).
- **Monitoring interface**: subscription/notify operations to get basic performance data and fault alarms collected from operative NS instances.
- **Resource Management interface**: subscription/notify operation to get basic information (without any details on the constituent VNFs and virtual links) on the state of the resources supporting operative NS instances, and query facility site's topology operation to get abstracted topology information from a 5G-VINNI facility site.

Note these operations are quite useful for the cases when the nesting/composition mechanism is involved, as happened in the Or-Or reference point (see Section 3.6.1). Indeed, when a NS instance managed by a given NFVO has been nested into a composite NS instance managed by the other NFVO, the latter can take advantage of the capabilities offered by the former to operate the composite NS instances.

4.7.1.2 Complementary interface elements

Besides exposing the above-referred capabilities, an NFVO could allow other NFVOs to perform more advanced operations over the NS instances under its management, in search of a more fine-grained NFV management and orchestration across facility sites. However, this granularity comes at the cost of complexity that likely goes

beyond the cross-domain NFV features in some sites. For this end, the implementation of these operations is subjected to the NFVO's features of each facility site

- **Catalogue Management interface**: on-board/update/delete NSD operations, and on-board/update/delete VNF Package operations,
- **LCM interface**: Scale/Heal NS operations, and subscription/notify operations to get advanced information on run-time NS modifications (e.g. NS instance has been scaled in/out, healed, updated). . **Monitoring interface**: create/query/delete PM job operations, create/delete threshold operations, and subscribe/notify operations to get fine-grained information of peformance data and fault alarms.
- **Resource Management interface**: subscription/notify operations to get more-advanced information (including details on the constituent VNFs and virtual links) on the state of resources supporting he operative NS instances. query facility site's resource zones operation, query facility site's resource pool operation.

4.7.1.3 Interfaces at the Service Operations and Management Level

According to Figure 4.1, the Service Operations and Management is the functional block operating at the highest abstraction level: the application layer. Mapping this block into the Service Orchestration functionality defined in the E2E MANO from the MEF LSO framework (see Section 3.6.1), and taking as a reference the open APIs specified in that framework, up to three type of external facing interfaces can be defined.

4.7.2 Northbound interfaces

These correspond to the ALLEGRO (CUS:SOF) Management Interface Reference Point that allows Customer Application Coordinator supervision and control of dynamic service behaviour (see Section 8.2.3) of the LSO service capabilities under its purview through interactions with the Service Orchestration Functionality.

For the northbound interface the TMF Open API's will be used, as described in Table 4.3.

Table 4.3: TMF Open API's

TMF APIs	Comment
Service Ordering API	Customer Application Coordinator controls Service by requesting changes to dynamic parameters as permitted by service policies. Customer Application Coordinator requests change to administrative state or permitted attributes of a Service.
Activation and Configuration API	Customer Application Coordinator controls Service by requesting changes to dynamic parameters as permitted by service policies. Customer Application Coordinator requests change to administrative state or permitted attributes of a Service.
Service Qualification API	
Service Inventory API	Customer Application Coordinator queries operational state of the Service.
Service Catalogue API	

4.7.3 Southbound interfaces

These correspond to the PRESTO (SOF:ICM) Management Interface Reference Point needed to manage the network infrastructure, including network and topology view related management functions.

In the 5G-VINNI Reference Architecture the E2E MANO will have southbound interfaces to the NFV MANO and the three domain controllers – 5G-RAN, 5G-Core and Transport.

E2E MANO will interface to the NFV MANO across the Os-Ma-nfvo reference point. Interfaces supported will be the SOL005 interfaces.

- Network Service Descriptor (NSD) Management

- Network Service (NS) Lifecycle Management
- VNF Package Management

E2E MANO will interface to the transport-domain controller following the specifications provided by IETF (e.g. IETF-YANG-MODEL [49]). Transport networks are managed using proprietary interfaces to dedicated Element Management Systems (EMS)/NMS. However, Network providers need a common way to manage multi-vendor and multi-domain transport. Therefore it is important to ensure that deployment use cases and related functionalities are supported by all models to allow a seamless translation/mediation between systems using different models.

The interface of a transport controller towards a northbound client, MAY be characterized by the following six main functions identified in Table 4.4, taken from IETF-YANG-MODEL [49].

Table 4.4: functions of transport controller interface to northbound client

Functions	Description
Obtaining Access Point Info	Getting the necessary access points info
Obtaining Topology	Getting the topology info
Tunnel Operations	Tunnel Setup, Deletion Modification and Info Retrieval
Service Request	Requesting connectivity service and retrieval the list of service request
Path Computation	Path Computation pre service provisioning
Virtual Network Operations	Requesting a virtual network and control operations (e.g. update, deletion)

4.7.4 Eastbound/westbound interfaces

There will be two flavours of East/West interfaces depending upon the architecture of the site that is interfaced to.

The first corresponds to the INTERLUDE (SOF:SOF) Management Interface Reference Point that provides for the coordination of a portion of LSO services within the partner domain that are managed by a Service Provider's Service Orchestration Functionality within the bounds and policies defined for the service. Within the 5G-VINNI Reference Architecture the partner domain will be another 5G-VINNI site or another project facility from 5G-Eve or 5Genesis, or possibly a 3rd party test site. These are a strictly SOF:SOF interface whereby both E2E MANO expose and consume each others' services. This flavour will support the APIs in Table 4.5.

Table 4.5: APIs supported on SOF:SOF interface

TMF APIs	Comment
Service Ordering API	Service Provider controls aspects of the Service within the Partner domain (on behalf of the Customer) by requesting changes to dynamic parameters as permitted by service policies. Service Provider requests change to administrative state or permitted attributes of a Service. Service Provider request creation of connectivity between two Service Interfaces as permitted by established business arrangement.
Activation and Configuration API	Service Provider controls aspects of the Service within the Partner domain (on behalf of the Customer) by requesting changes to dynamic parameters as permitted by service policies. Service Provider requests change to administrative state or permitted attributes of a Service.

	Service Provider request creation of connectivity between two Service Interfaces as permitted by established business arrangement.
Service Inventory API	Service Provider queries operational state of the Service. Service Provider queries the Partner for detailed information related to Services provided by the Partner to the Service Provider.
Service Catalogue API	

The second flavour is the corresponds to the PRESTO (SOF:ICM) Management Interface Reference Point but is limited to integration between the SOF and NFV MANO. This east/west interface differs from the PRESTO reference point by virtue of the fact that the SOF and ICM reside within different operator's domains. The services supported by the API will be the SOL005 interfaces.

- Network Service Descriptor (NSD) Management
- Network Service (NS) Lifecycle Management
- VNF Package Management

4.8 Security

4.8.1 Adoption of 3GPP Security principles

A brief summary of 3GPP Security principles is provided in Section 3.1.3, and MAY be implemented by Facility Sites. However, detailed considerations on integral 5G security are out of the scope of 5G-VINNI. Therefore, in this document we consider security from the concrete angles related to infrastructure deployment and operation, related to what can be implemented at the different project facilities.

4.8.2 Zone Model and multi-tenancy

With the use of NFV and SDN in 5G, it becomes critical to secure the Cloud as a shared environment between tenants and slice instances. Security needs to be present in all sub-domains of the Cloud to address the current threat landscape with sophisticated actors, to adhere to privacy and security regulations in the different markets, and to ensure isolations of tenants and network slices.

Main security controls that SHOULD be established in the Cloud domain are;

- Infrastructure network zoning model that applies to all layers of the infrastructure,
- Solid Multi-tenancy capabilities through the infrastructure platform,
- Security monitoring through flow based network analytics,
- Centralized logging with analytics capabilities,
- Server integrity checking and configuration auditing.

Note that there are other important security functions, such as DDoS protection, Intrusion Detection and Prevention Systems (IPS/IDS), antivirus, antimalware etc. The aim is not to cover all security requirements and features in this deliverable

In the following we describe architectural principles for the zoning model and multi tenancy.

4.8.2.1 Zone Model

The *infrastructure network zoning model* is fundamental in addressing the shared technology challenge in a multi-tenant platform. The defined zone model regulates how the cloud infrastructure is built and how segmentation is achieved.

An example zone model can be based on three levels of entities; (i) *Domains* that defines level of control. (ii) *Classes* that defines policies for exposure levels, data storage and data processing. (iii) *Zones* that isolates tenants, services and components within a class.

A security domain is an entity that defines the level of control and governance present; it does not itself have any policies, but exists to define where boundaries are present. A domain can consist of one or more security

classes. Domains defined in level of trust are; (a) un-controlled, (b) controlled, (c) restricted, (d) secure, and (e) externally controlled domain. Traffic is not allowed to flow directly between non-adjacent security domains, e.g. a device in (a) cannot directly communicate with device in (c), but have to pass through termination point in (b).

Security Zone Classes is a concept of aggregating network zones with the same exposure and security levels. Zones are built using infrastructure dedicated to the security zones and system components within that specific class. The entire operating environment SHOULD be taken into account when implementing the infrastructure according to the model (e.g. servers, networking, backup and storage, authentication infrastructure, virtualization).

An example zone model operating along a security architecture with six different Security Classes pre-defined is illustrated in Figure 4.11. Three Security Classes are for service production and three Security Classes for management.

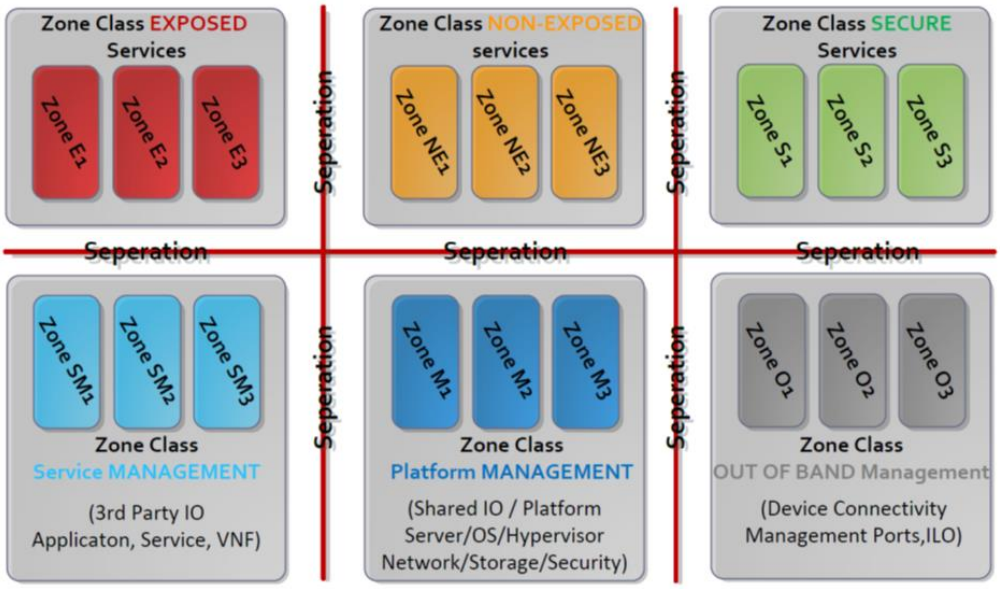


Figure 4.11: Example Security Classes and division into Zones for the Clo

Each of the security classes are further divided into *security zones* to segment services and tenants apart as illustrated in Figure 4.11. As opposed to security classes, a security zone by default use logical separation mechanisms to isolate tenants and services from other zones.

While the example zone model in Figure 4.11 is quite comprehensive, 5G-VINNI will use a scaled down version of it for which some of the guidelines are as follows. The Security Zone Model SHALL use strict separation between each Security Class (i.e. inter-security class traffic) either physical or logical. The separation needs to be reflected in all levels of the platform infrastructure components such as network, storage, compute, hypervisor and management systems.

Logical separation for intra-security class traffic SHOULD be used for most services (e.g. VxLAN based micro segmentation, virtual firewalls).

Physical firewalls SHOULD be used for inter-security class traffic. Hence, traffic from one security class to another will need to pass two different layers of security with one being logical or virtual and with the second layer being physical.

It SHOULD be possible to provision resources from a shared spare pool to the different security classes and tenants to maintain efficiency and automated management while maintaining the necessary segmentation. The required agility is achieved by having the physical storage and compute resources movable between classes and then creating dedicated logical software defined resources on top of them. The key enabler to achieving this is a shared network fabric and SDN.

4.8.2.2 Multi-Tenancy

Multi-tenancy enables separation between tenants running services in a shared environment. In a multi-tenant deployment, the resources controlled by one tenant are physically or logically separated and secured from other tenants. In addition to tenant isolation, reporting and capacity management individually per tenant is important. Multi-tenancy is a key requirement for services across both public and private cloud environments.

Tenants SHALL be represented by tenant-IDs that SHOULD be used in all configuration activities. No other tenant information SHALL be visible in the data plane. Each configuration activity within orchestration SHOULD be authenticated and verified before execution and then logged.

Tenants MAY be grouped into different classes, where typical groups requiring different access are; platform operators, generic application operators, regulated application operators, test and development, and business users. Strong authentication using multi-factor authentication SHOULD be used irrespective of which group a tenant belongs to.

4.9 Testing and Monitoring

Testing and Monitoring systems, as described below, SHALL be implemented in main facility sites. These are not mandatory for experimental sites.

The testing architecture has a purpose that is many-fold: first of all is establishing automation services that encompasses the CI/CD phase of the infrastructure deployment. This is nowadays considered a very important stage given the number of performance and interoperability variables at stake in modern virtualized systems. The second purpose of the testing infrastructure is to provide testing and experimentation services to the vertical service operators but also to the mobile operators themselves for network maintenance purposes. The final purpose is to provide an automated testing service to the orchestration layer. This is a fundamental step for ensuring reliability, resilience, and performance of the slice components and of the slice as a whole during the slice deployment phase. This is a due process before handing the slice over to the customer.

A second desirable component is the Monitoring architecture. Such system is interleaved with the slice components and it is used to monitor all the components, from the virtualised infrastructure to the Quality of Experience (QoE) of the traffic carried by the network. While its presence would be highly suggested, the complexity of virtualized networks make a full monitoring extremely heavy and resource-demanding. It will be here explained for the sake of completeness, but its implementation, even partial, mandated as optional.

4.9.1 Testing Architecture

The testing framework is centred round the Test Executor, that is in charge of commanding all the different components that play a role in a test, either testing tools, or System Under Test elements.

Web services provide an interface towards the customer, enabling the design and execution of tests. APIs enable the southbound communication towards the different elements in the test. The testing tools are stored as images or snapshots available to the Virtual Infrastructure Manager to be executed. Results are stored in a separate repository to be exposed to the Analytics components. A simplified Testing Architecture is displayed in Figure 4.12.

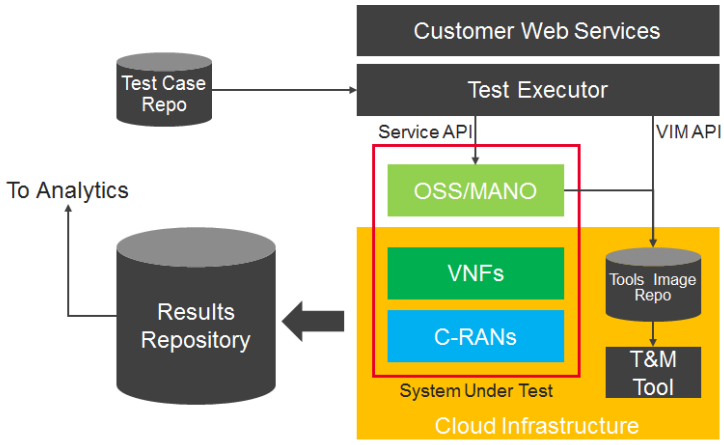


Figure 4.12: Testing Architecture

As displayed in Figure 4.12, the framework is complemented by a Test Cases Repository that enables quicker execution of testing life cycles.

4.9.2 Monitoring architecture

The foreseen Monitoring Service architecture is displayed in Figure 4.13.

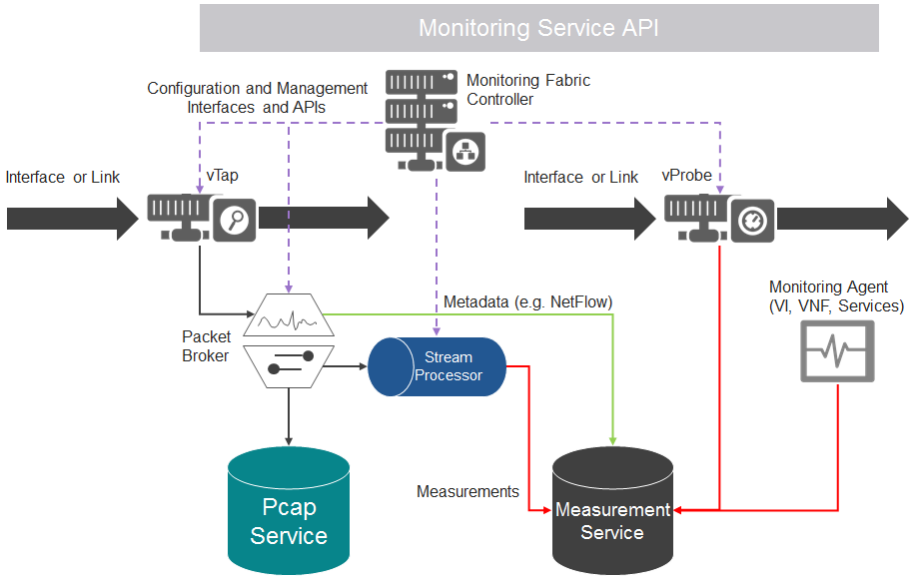


Figure 4.13: Monitoring Service Architecture

A Monitoring Service API is exposed for the management of the Monitoring Service Fabric whenever it will be deployed by the Service Orchestrator.

The Monitoring Services will be based on a combination of:

- vTaps: virtual network taps duplicating and forwarding packets or messages from network links or interfaces.
- vProbes: virtual network probes performing online measurements on network links or interfaces.
- Monitoring Agents: elements running as a part of the NFVI, or the VNFs, or Orchestration services providing data such as logs, events streams, infrastructure measurements (e.g. CPUs usage).
- Packet Brokers: virtual network element performing filtering and meta-analysis of the data flows coming from the vTaps.
- Pcap Service: packet capturing service storing the entire data exchange from a particular session.
- Stream Processors: online data stream processors performing measurements on aggregated data flows coming from the Packet Brokers (or vTaps).

- Measurement Service: heterogeneous data repository used as a foundation for providing input to Analytics services. A separate APIs will be used to expose the data to the Analytics services.

All the elements, except for the Monitoring Agents, will expose a set of Configuration and Management APIs towards the Monitoring Fabric Controller. This element can be considered an SDN/SDx controller, and as such, is in charge of connecting the various elements together and providing specific configurations such as the filtering rules for the packet brokers.

The Monitoring Agents are already considered part of the infrastructural elements, and for this reason it is best that their configuration is performed by NFVIM, VNFM, or SDNC, or any other infrastructure orchestration element of competence.

4.9.3 Interface elements

The interface elements exposed by the test execution framework are both north- and southbound, as displayed in Figure 4.14.

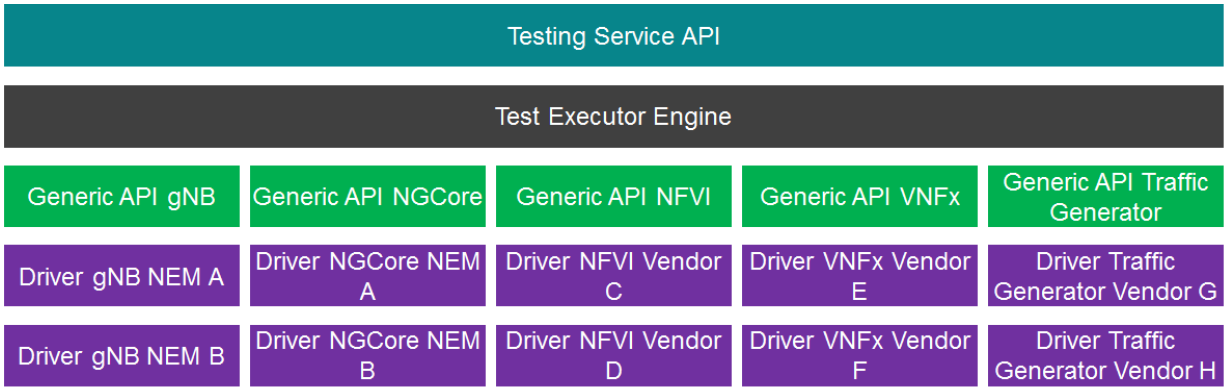


Figure 4.14: Testing Service Architecture

The Northbound interface is here called Testing Service API. It is used to request, configure, and manage the execution of the individual tests or test campaigns. It can be used by either web services exposed to the customer or the orchestration environments through a viable plugin.

The Southbound exposes a set of “Generic APIs”. Such APIs represent an abstraction layer for each individual component of the E2E system, ranging from the gNB to the virtualised infrastructure to VNFs. Such abstraction is used to de-couple specific configuration, management, and control implementations, ensuring a re-usability of test cases. A Driver plugin performs the translation between the vendor-specific implementation and the Generic API. The drivers can be either very simple due to existing implementation of the generic API in the equipment or more complicated implementing an API translation.

While being a critical part for guaranteeing the health of the 5G infrastructure, monitoring services come at a great burden of the infrastructure itself. They are therefore considered something that MAY be part of the 5G-VINNI facility whenever resources are available for their deployment.

Several types of monitoring deployments can be foreseen depending on how deep visibility is needed. Therefore, all the different management and configuration elements are considered optional.

5 Interconnection and Interoperability of 5G-VINNI sites

5.1 Interconnection between sites

This section explains the architectural requirements in terms of connectivity between sites. As a main principle, it is important that all sites have IP connectivity with each other, as represented in Figure 5.1.

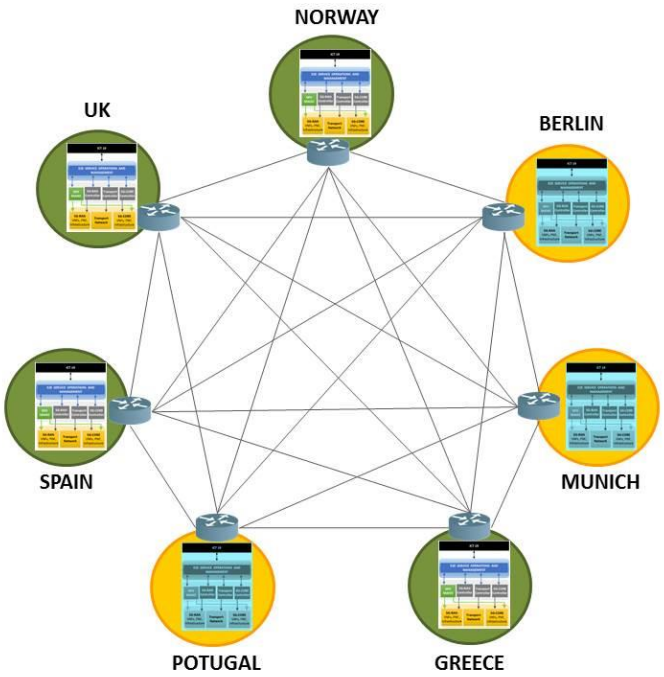


Figure 5.1: Interconnection of 5G-VINNI Sites

For such purpose, each site SHALL have one or more facility site gateways (FSGW) that fulfils the following requirements. In the long term solution, two FSGWs can be the preferred solutions (although not compulsory) for increased load balancing and resilience capabilities.

- Public IP.
- Routing protocols, especially exterior gateway protocols, such as Border Gateway Protocol (BGP).
- Generic Routing Encapsulation for the establishment of GRE Tunnels.
- Internet Protocol Security (IPsec) protocol suite.

For QoS Differentiation on the connection between sites the following options are under evaluation.

- Best Effort Internet (default for all facilities, according to basic service level).
- SD-WAN, where the traffic can be sent over a combination best-effort Internet paths and some way of using MPLS Core and/or using direct DiffServ enabled assured quality traffic exchange between the operator domains. This will enable the SD-WAN to support ASQ paths between facility sites.

This will allow various ways of enabling assured service quality (ASQ) interconnection paths between sites and value added connectivity (VAC) triggered and handled on-demand for end-point traffic flows whose traffic can be steered onto the ASQ paths. For more information on the multi-provider ASQ connectivity services, covering a hierarchy of various ASQ path and VAC services, and how they can be combined with NFVIaaS and VNFaaS, see 5GEx Deliverable D2.2 [39] and D2.3 [69]. The latter also includes suggested charging and business model principles.

For the support of use cases and scenarios beyond the basic service offers (based on best-effort Internet interconnection) the support for enterprise VPN will be considered. These services can be relevant for enterprise customers running their applications as tenants of several facility sites or connecting to partner enterprises that have their applications running as tenants of one or more different facility sites. The combination of enterprise VPN and Internet connectivity will also be considered.

- BGP based MPLS IP VPN, with reference to IETF RFC 4364 [74], provides specifications how BGP and MPLS can be used for realizing enterprise VPN service offered by one or more network service providers (NSP), using the IP/MPLS technical approach.
- The RFC 4362 specification also includes ways of realizing the VPN across multiple NSPs (aka. Inter-AS VPN, and option a, b or c).
- In addition, RFC 4364 also includes a hierarchical approach allowing a VPN to be used as a Carrier for other Carriers (NSPs) traffic, either when such a customer carrier needs to accommodate their ISP role or their customers' VPNs. This is aka. "Carrier's Carrier".

In the following paragraphs some characteristics of MPLS and SD-WAN are briefly presented.

Classic MPLS VPN with QoS: MPLS is a very well-known connectivity solution. With MPLS, the operator can make separate bandwidth reservations for different classes of traffic and give different forwarding behaviour based on the class. Figure 5.2 illustrates the concept of Different MPLS classes. Further information can be found in Zhang and Ionescu [50] and Davie and Rekhter [51] among many others.

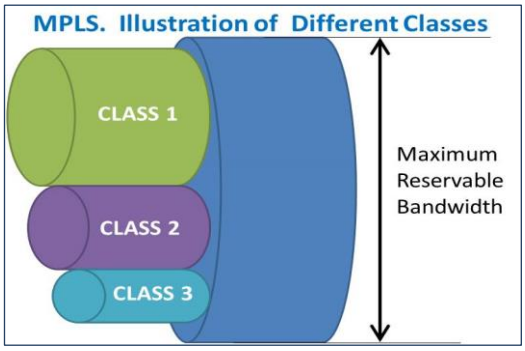


Figure 5.2: Illustration of MPLS with Different Classes

SD-WAN with MPLS Core: This concept uses software-defined networking in a wide area network (WAN), and is based on the traditional SDN principle of separating the CP from the data plane. One of the advantages that it offers is to enable the use of private WAN connections such as those based on MPLS, with conventional best effort internet access. Under this basis, the Operator has greater flexibility for the selection of channels that better fit different requirements. Figure 5.3 illustrates the concept of a SD-WAN deployment. Further information can be found in Michel and Keller [52] and Ali, Manel and Habib [53].

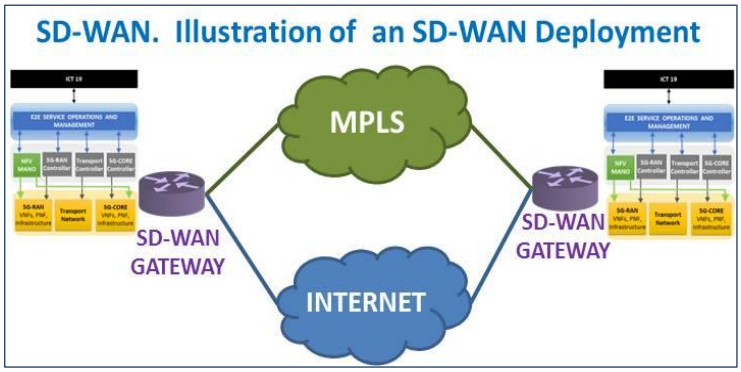


Figure 5.3: Illustration of a SD-WAN Deployment

In the initial phase, all sites will use Best Effort internet. However, the possibility that the connection between some specific pair of sites will be upgraded gradually to MPLS-VPN and/or SD-WAN with MPLS Core is under evaluation.

5.2 Interoperability

Once the different 5G-VINNI sites are interconnected as presented in the previous section, there is the need of providing interoperability in order that network services can be deployed and operated across sites. Note that while the connectivity part is drawn as a solid red line across the different technology domains as well as through the interconnection the individual end-user or end-device traffic flows will not be handled individually

in the transport domain or in the interconnection domain. In case of VAC on demand the individual connectivity services and traffic flow state information will be handled at the RAN and 5G Core level but not in the transport, backbone and interconnection segments which only handles the traffic at the aggregated levels. These traffic aggregates will most likely be handled at different hierarchical aggregate levels, according to NSP traffic engineering policies.

Figure 5.4 illustrates a service across two different 5G-VINNI sites.

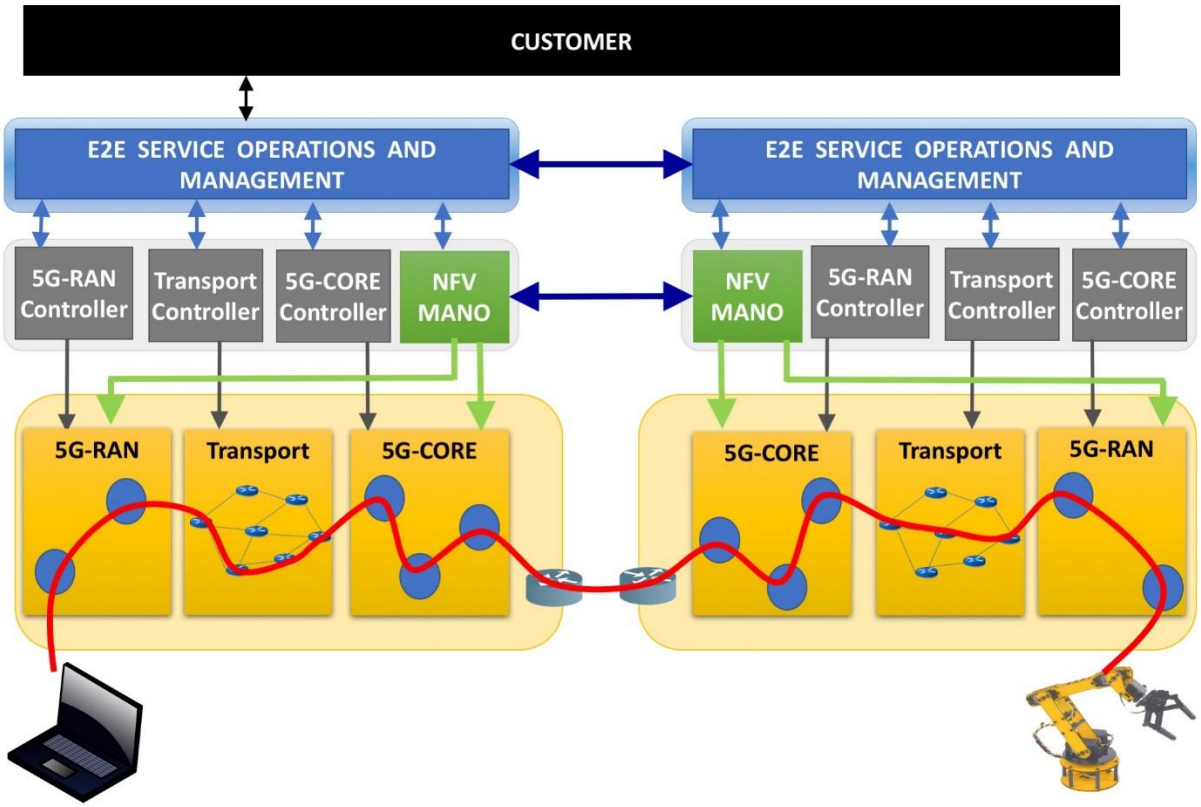


Figure 5.4: Illustration of Service Inter-Operation across 5G-VINNI Sites.

In order to enable such interoperability, the use of North-South Bound and East-West Bound interfaces are needed in order to guarantee the proper understanding of the service needs on each of the involved sites, as highlighted in Figure 5.5. The service and resource orchestration, control and management operations are introduced and discussed in the previous Section 3 and Section 4. While the arrows are here drawn directly to/from the NFV MANO in the neighbour domain this will typically be realized via inter-provider APIs (e.g. along the lines of 5GEx Project Ref [39]) and rather indirectly accessing these MANO capabilities, according to the specific APIs provided and the security policies in the provider domains. In this regard, it is important to note also that the specifications of these inter-provider APIs are still at an early stage and to be considered as a key activity and work in progress by the industry forums and specification defining organizations.

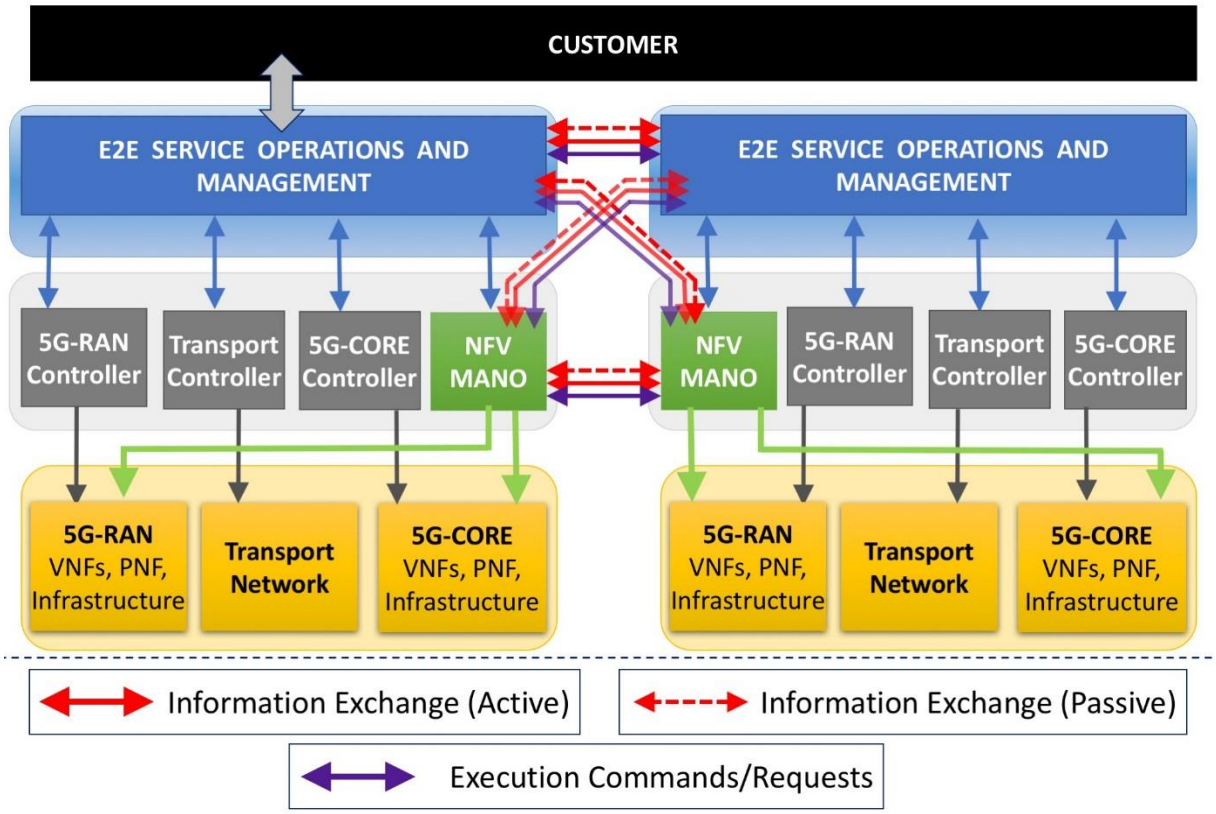


Figure 5.5: Interfaces needed for Sites Interoperability.

Based on the external interfaces presented on Section 4.7, here in Table 5.1, we propose a complementary classification of the functional aspects of such external interfaces.

Table 5.1: Levels and Functional features of 5G-VINNI site external interfaces

	Information Exchange (Passive)	**Information Exchange (Active)**	**Commands Execution**
Service Operations Level	Notify to a remote site changes (alarms, failures, updates, etc.) on local service operations	Request and access information on operational status of remote services	Request and execute operational changes on remote site services.
Virtualised Service Level	Notify to a remote site changes (alarms, failures, updates, etc) on local VNFs/NSs	Collect fault and performance alarms and info-status from remote sites VNFs/NSs	Authorize, invoke and execute LCM functions on VNFs/NSs on remote sites
Resources Level	Notify to a remote site changes (alarms, failures, updates, etc) on local resources allocating VNFs/NSs, and changes in topological information	Exchange of network topology and resource information between administrative domains	Authorize, invoke and execute changes over the resources that accommodate NS and VNF instances.

The specific type of external interfaces to be used and activated, as well as the interaction between management and orchestration components will depend on the respective use cases, and the respective progress on the mentioned inter-provider APIs work by industry forums and specification defining organizations. Therefore, the decision will be taken at further stages of the project in accordance with the specific use cases needs as specified in Table 5.2 below.

Table 5.2: Options available for the interoperability of 5G-VINNI Sites

	Information Exchange (Passive)	Information Exchange (Active)	Commands Execution
E2E Op. <-> E2E Op.	Yes / No	Yes / No	Yes / No
E2E Op. <-> NFVO	Yes / No	Yes / No	Yes / No
NFVO <-> NFVO	Yes / No	Yes / No	Yes / No

6 Future topics and Research

6.1 Edge Computing

6.1.1 Overview

In 2014, ETSI has created the Industry Specification Group (ISG) MEC to develop edge computing technology, in order to capture the broader scope. Edge computing is a key technology for 5G, in order to achieve ultra-low E2E latencies and increase bandwidth efficiency.

ETSI MEC architecture [71] is agnostic to 3GGP Generations. It can be applied to 4G networks; however, there are still many open issues, as 4G was not designed for MEC [72]. In 3GPP TS 23.501 [1], MEC is natively supported, further expanded on in the ETSI MEC paper on 5G [73], as it was designed from scratch with some edge computing enablers.

6.1.2 MEC on 4G Networks

On 4G, there are essentially two approaches to implement MEC: bump in the wire and distributed EPC, described in the ETSI MEC paper on 4G [72]. Figure 6.1 below depicts both approaches (bump in the wire, up; distributed EPC, down).

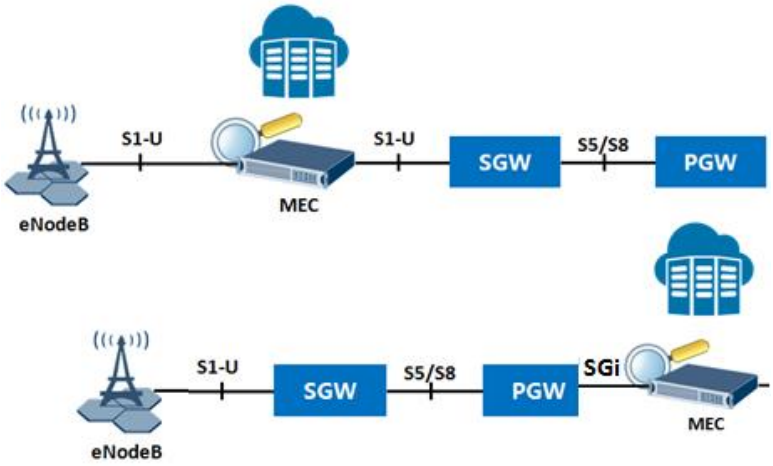


Figure 6.1: MEC in 4G networks: Bump in the wire and distributed EPC approaches.

The bump in the wire approach intercepts traffic at the S1-U interface (between eNB and S-GW). In the upstream direction, the GTP-U traffic is decapsulated and inspected in order to decide whether it is to be offloaded to an edge data center or forwarded towards a S-GW. A similar path is followed in the opposite direction. The offloading of some traffic at this point breaks the usual 3GPP model, as it skips tasks usually handled in the Core, like lawful interception, data retention, charging, or policy.

The distributed EPC approach has multiple flavours but basically intercepts the traffic in the SGi interface. At this point, the traffic is plain IP and has already crossed the Core, following the usual path. This way, it naturally solves the bump in the wire issues. However, traditional 4G deployments have centralized SGi interfaces, which can be several hundreds of Kilometres away from the customer. In this case, edge computing has a limited impact on latency and bandwidth efficiency as the edge is very far away from the customer, losing the expected edge benefits.

6.1.3 MEC on 5G Networks

5G natively supports MEC [73], as it has some enablers to divert traffic to the edge and it was designed to deploy gateways at the edge. With such mechanisms, 5G can be deployed close to customers, while it crosses gateway functions where actions like lawful interception, data retention, charging, or policy are handled.

In particular, the following features can be realised in the 5G-enabled edge computing capabilities:

- *The UPF is designed to be distributed.* This allows the UPF gateways to be deployed at the edge, unlike 4G typical deployments, which are usually on the Core. This allows the deployment of MEC platforms at the edge, while benefits from the gateway standard features, such as lawful interception, data retention, charging, policy, or others.
- *Uplink Classifiers on UPF can divert traffic locally. This* allows some traffic to be diverted to a local network, which can be an edge data centre. Those classifiers are 5G native and can be managed by the MEC platform.
- *PCF set policies to influence UPF routing*. This allows the Policy and Charging Function (PCF) to manage the UPF classifiers, via Session Management Function (SMF) and offload some traffic to the edge, where a MEC platform receive it.
- *AF via PCF (or NEF) can influence the traffic routing and steering.* This allows external entities, like MEC platforms, to act as AFs and interact with the PCF (in case MEC is a trusty entity for 5G), or NEF (if MEC not a trusty entity to 5G) to set rules to offload some traffic, offloading it to the edge data centre where MEC Apps are deployed to provide edge services.

Figure 6.2 depicts the interaction between MEC and 5G Networks (assuming that MEC is a non-trusted entity) [73].

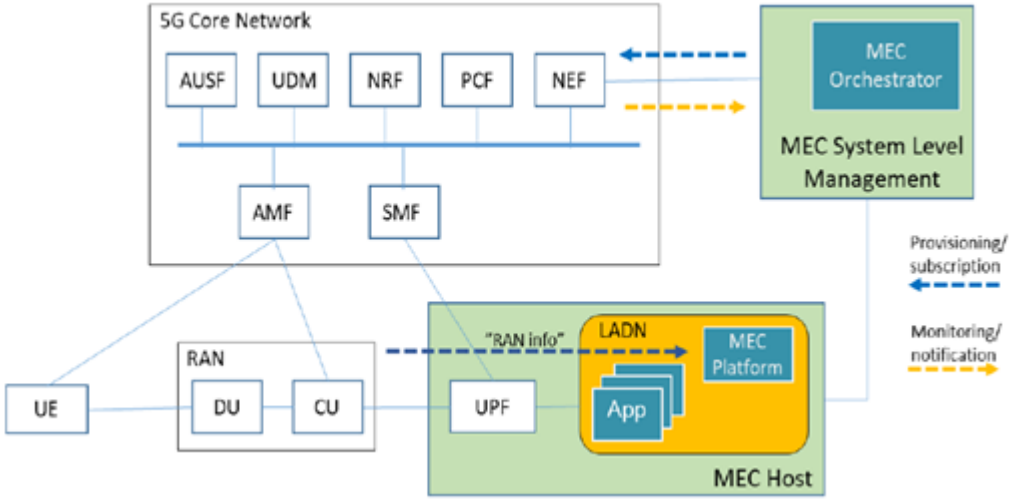


Figure 6.2: MEC in 5G networks: Integration [73].

6.1.4 MEC Implementation

A 4G-based MEC prototype implementation has been developed, which uses the bump in the wire approach. This implementation is able perform a full MEC operation; however, does not solve the issues related to lawful interception, data retention, policy or charging.

Regarding 5G, further research on this implementation intends to:

- Evolve the platform towards a 5G-based MEC, taking advantage of the 5G enablers;
- Integration with 5G networks, in particular with the PCF and NEF components;
- Tests and measure the performance of the integrated (MEC and 5G) deployment, in terms of latency and bandwidth efficiencies KPIs, with multiple vendors and in different locations.

6.1.5 Content Distribution Applications

Work has been conducted on the support of content distribution applications, at the edge of 4G (EPC) networks, supporting service orchestration operations related to cross-slice communication and routing optimization, in particular for edge caching.

Further development of this capability for MEC in 5G intends to:

- Support MEC integration within the 5G Core architecture
- Support MEC service orchestration primitives in the context of 5G-VINNI MANO architecture

- Develop and integrate Location Service primitives (ETSI GS MEC 013 [68]) with the purpose of supporting mobility management
- Investigate issues related to service continuity for streaming applications, including application context transfer and user (re-)direction
- Investigate and develop KPI measurement mechanisms tailored for the targeted edge computing environment e.g., orchestrating E2E performance measurements at the edge

6.2 Backhaul Automation

Software defined networking (SDN) is a promising technology for introducing the required programmability features into transport networks and it enables operators to efficiently share their transport network infrastructure resources among different services through network slicing, as described in Fiorani *et al* [54] and Mayoral *et al* [55]. SDN simplifies how transport networks deploy and operate, thus greatly reducing OpEx. Automation of configuration and service creation, traffic engineering and bandwidth control as well as common logical representation of network resources exposed through well- defined APIs, are key features of an SDN solution for the WAN.

Regarding the transport network automation, the ETSI NFV MANO framework defined the WAN infrastructure Manager (WIM), as a particular VIM. In this case, the VIM (e.g. OpenStack cloud controller) is responsible for controlling and managing the NFVI-PoP's resources, whilst the WIM is used to establish connectivity between NFVI-PoPs by communicating with transport SDN controllers. Figure 6.3 show a hybrid network environment example illustrating the goal of NFV to have fully programmatic open interfaces for service and resource orchestration within and across NFVI-PoPs.

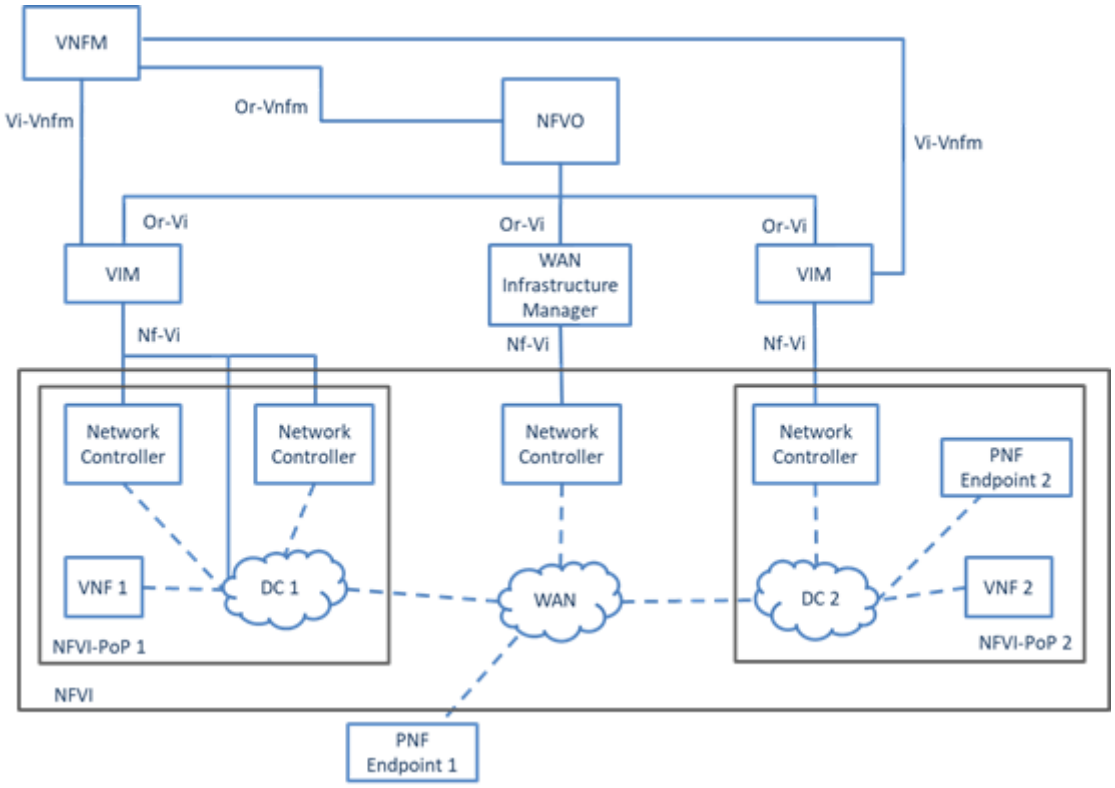


Figure 6.3: Network Controller example

The figure, taken from ETSI GS NFV-MAN 001 [16], shows:

- An example NFVI network that provides the necessary connectivity services to establish an E2E service between Endpoint 1 and Endpoint 2 through VNF 1 and VNF 2 as shown by the VNFFG representation.
- A model where the WAN between NFVI-PoPs provides Virtual Links based on network connectivity services between NFVI-PoPs, and where the NFVO may request Virtual Links based on network connectivity services through the Or-Vi reference point. This requires Resource Orchestration, an

NFVO function, to have an abstracted view of the network topology and interfaces to underlying VIMs and WAN Infrastructure Manager (WIM) to request connectivity services.

- One VIM per NFVI-PoP but this is not mandated as there could be multiple VIMs in a single NFVI-PoP, each responsible for an Administrative Domain. Similarly, a single VIM could be responsible for multiple NFVI-PoPs. There could be multiple WIMs. Each VIM/WIM provides connectivity services within its domain and provides an abstraction of the underlying resources through a well-defined interface.

Current NFV MANO framework considers the network as a commodity that provides packet pipes with QoS, either pre-provisioned or dynamically provisioned by the WIM between the specified end-points. Since the introduction of the architecture of Figure 6.3 in [16], significant work has been conducted regarding the potential architectural options for the role and placement of WIM functional entity within the NFV-MANO architecture, as reflected in ETSI GR MFV-IFA 022 [56], and shown in Fig. 4.19. This architecture will have a lot of momentum in the 5G era, where the transport network will provide connectivity between virtualized 5G RAN and virtualized 5G Core.

Based on the previous architecture it is noted that a single SDN controller, comprising multiple network nodes featuring diverse technologies, provided by different vendors is not realistic. This is particularly because transport networks are fragmented into multiple vendor domains with its own control and MP technology. Apart from this reason, for modularity and administrative simplicity, in 5G-VINNI we follow a hierarchical approach, as laid out in Figure 6.4 where the WIM is the parent and the children are the SDN controllers which are dedicated to different transport network technologies. The WIM allows service orchestration by integrating the wireless, optical and satellite backhaul domains with the orchestration layer, thus allowing efficient cross-domain network slicing and service automation, as in Figure 6.4. WIM's role is twofold:

- Interaction with WAN Domain Controllers for slice stitching between the different transport network domains.
- Interaction with the orchestrator (NFVO) in order to stitch transport network slices with 5G RAN and Core slices.

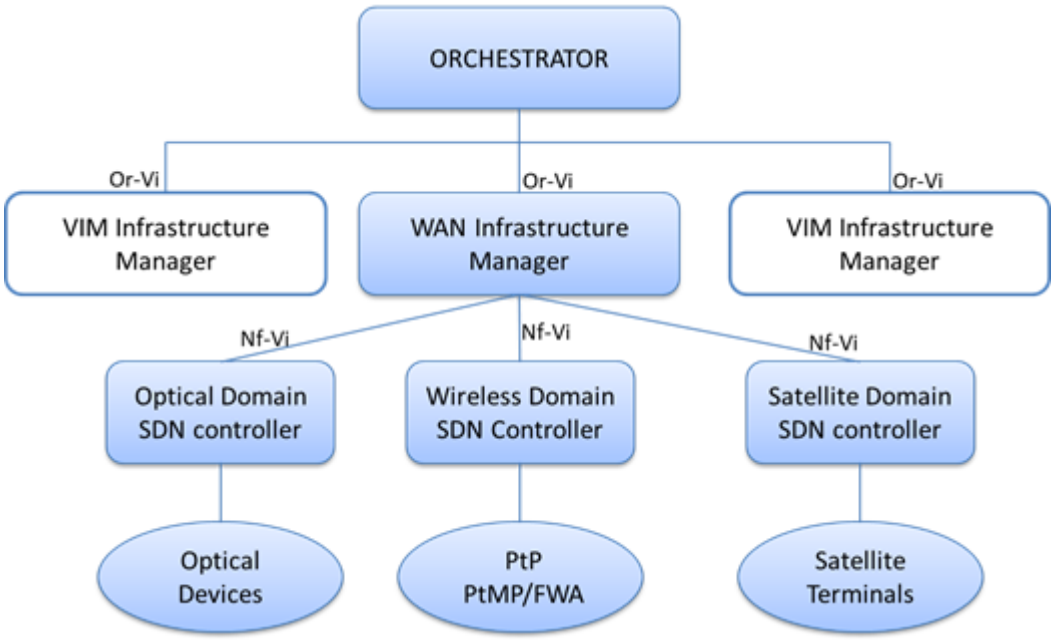


Figure 6.4: Hierarchical transport SDN controller architecture

6.3 Flexible Architecture for Verticals

Based on the analysis on on-going 3GPP SA1 study items shortly discussed in section 3 and on-going 3GPP technology focus on Core Network, in particular, the study items of 3GPP TR 23.791 [57] and 3GPP TR 23.742 [58], and the work item of 3GPP TS 23.725 [59], some high complexity issues relating to:

- the missing enabler for technologies integration required to address vertical requirements;

- the clarification of the relationship and scope of CP and MP;

can be identified as potential high interest research topics for future releases.

With the completion of Release 16, 3GPP system will already provide some effective but basic enablers for Vertical integration. In some aspects, however, Release 16 shows some notable limitations:

- Once deployed and operational, a (set of) network slice(s) in use by a 3rd parties cannot be further customized and optimized (e.g. resetting target KPIs);
- Vertical integration concepts relate solely to 3rd parties AFs.

In addition, the concept of SBA has been consolidated within 5G System, both in SA5 and SA2, and it has been spanning across both CP and MP. Notably, the formalization of the SBA framework(s) has been completed by specifying Producers and Consumers for services. The final result highlights similar services may be produced and consumed by both CP and MP, for example:

- MP PM will expose services to collect NFs data both to NWDAF and OAM;

This considerations provide evidence the scope separation between CP and MP requires further study, and that CP and MP may use mutually the services exposed by each other.

Possible items to study:

- Flexible architecture Framework including:
  - Plug and Play of vertical functions and micro services;
  - A “brain” to decide on the configuration of NFs/services/micro services according to vertical requirements;
  - Enhanced network management tools such as network monitoring, network configuration;
  - Flexible per vertical NF configurations;
  - CP and MP scope separation;
  - CP and MP exposed services and mutual interaction;

Figure 6.5 shows a high level flexible architecture framework.

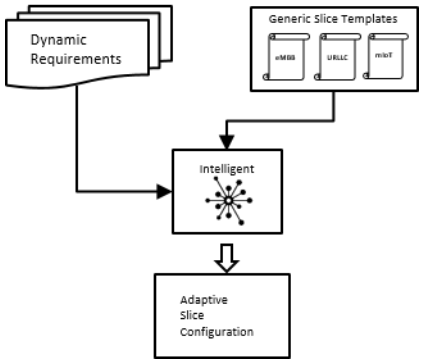


Figure 6.5: A high level view of flexible architecture framework

6.4 Analytics-driven service automation

The ultimate goal of 5G is to assure user experience, e.g., via SLA management, service assurance, service orchestration and management. Specifically, 5G targets at automating the process to meet dynamic and diverse requirements of various vertical services simultaneously. In order to meet this goal, intelligent analytics is expected.

With the completion of Release 16, the eNA study item (TR 23.791) will be addressing primarily on (i) interaction between CP and MP and (ii) data collection mechanism. However, Release 16 shows some notable limitations:

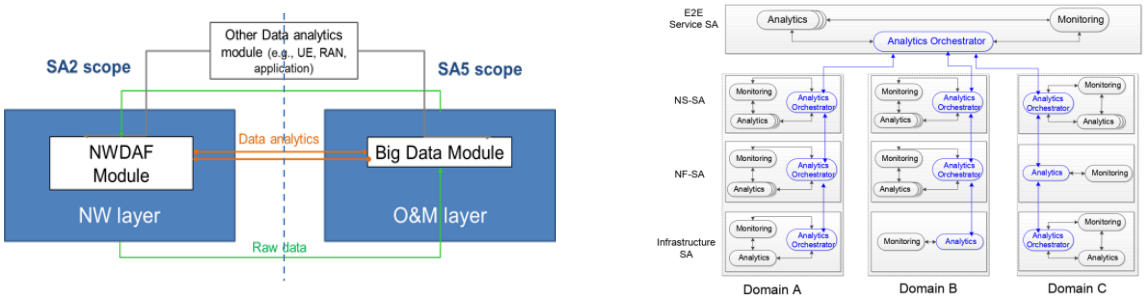
- No clear scope of 5G Core and OAM analytics data management, e.g., to support SLA management
- Unclear E2E service Assurance loop to support SLA management
- Dedicated data collection mechanisms between specific Producers and Consumers
- Monitoring service is solely managed by MP;

- No regulation of 5G System analytics exposure to 3rd parties, which may include a dedicated analytics system

A list of potential research topics include:

- Service assurance (SA) architecture to automate provisioning of network slice services
  - Fundamental functions
  - Role of Analytics in service assurance
  - Interaction of analytics with other Service Assurance functions to build agile, adaptive, and cost-efficient service assurance
  - Monitoring as a Service to support Analytics and SLA management
- Integrated Analytics framework:
  - An analytics framework with consideration of CP and MP as well as 3rd party functions
  - Applications and roles of the analytics framework in service automation;
  - Closed-loop between orchestration, control and assurance to automate service delivery

Figure 6.6 shows a high level integrated 5G Analytics framework and SA architecture.



(a) (b)

Figure 6.6: (a) The integrated 5G analytics framework and (b) Analytics supported SA architecture.

6.5 CoreRAN slicing and services

Most of the existing work on network slicing focuses on applying SDN and NFV principles to the Core network, but to maximize the efficiency of physical resource utilization and ensure proper isolation an E2E slicing paradigm should be used. Thus, RAN resources should be virtualized and future implementations should move in the direction of a service-oriented RAN architecture. Ideally, every tenant (such as OTTs and MVNOs) should have its own set of logical RAN resources (spectrum, base station functionality etc.), called slice, which in addition to the logical Core network resources, would create an E2E virtualized network instance with provisioning for the QoS requirements of the offered service. This approach also offers isolation among the tenants, who could possibly be competitors, and should only have a view of the network resources allocated to them.

6.5.1 RAN sharing in 4G LTE

There are many widely adopted approaches for RAN sharing in existing LTE networks. However, only two of them are standardized by 3GPP, documented in 3GPP TR 22.951 [60] and 3GPP TS 23.251 [61], and included in Alaez *et al* [62]. These two focus on sharing radio network elements (base stations) and radio resources (licensed spectrum) among up to six operators. In summary they are:

- **Multi-Operator Core Network (MOCN):** The operators share base stations, but each one has its own Core network (EPC).
- **Gateway Core Network (GWCN):** The operators share base stations as well as the MME element of the Core network. The rest of the Core network is isolated.

A UE may connect to a shared base station without any modification. Upon connection, the UE is informed of the PLMN-ids of the operators that are present. Then, the UE notifies the base station of its preferred PLMN-id.

The main goal of both is to reduce equipment, spectrum licensing and operational costs of the operators. RAN sharing schemes in LTE divide the resources among tenants, without offering isolation. Thus, one operator can view how the others are utilizing their resources.

5G RAN sharing is further expected to aid in better delivery of services respecting QoS guarantees per service and improve the overall resource utilization of the base station by dynamically allocating resources in an isolated manner.

6.5.2 State of the art on RAN slicing proposals for 5G

The proposals for RAN sharing on 5G vary on the number of elements that are shared (see as examples Garcia *et al* [63], Foukas, Marina *et al* [64], Foukas, Patounas *et al* [65] and Rost *et al* [66]). By increasing the number of shared elements, the overall resource utilization increases, at the cost of increased complexity. On one extreme is "slice-specific RAN", where every operator owns all the components up to the upper PHY layer and only the transmission point is shared. On the other extreme is "slice-aware shared RAN", where all the RAN components are shared among the participating operators, allowing for maximum efficiency. The overall trend is towards including in the pool of shared resources all the network components and NFs. This is made possible by the increasing use of commodity hardware (as opposed to specialized equipment) to implement NFs and advances in virtualization.

6.5.3 Challenges

Resource isolation, while respecting QoS guarantees and maximum utilization of the physical resources are orthogonal targets. Ideally, a hypervisor would be monitoring the utilization of each slice and dynamically make a better distribution of resources.

To guarantee isolation, all the RAN functionality, such as scheduling, should be virtualized, so that each slice can view only its own resources, as if it is the sole owner of a physical network. Thus, each functionality should be replicated as many times as the number of slice owners and a mapping between the physical resources and the virtual ones should be created. Every slice should be able to handle all the mobile network operations of a given mobile network architecture (eg. handovers, roaming).

This flexibility inevitably makes the architecture more complex, so potential impacts on performance should be studied.

5G networks will support multiple RATs (including emerging technologies like 5G new radio and NB-IoT), which implies that RAN slicing should be able to accommodate multiple RATs. This presents a new challenge where a single user's traffic may be multiplexed over several elements that are responsible for different RATs. Such virtualization requires tight coordination between multiple RAN elements, which can be very challenging given that the RAN operations generally come with a tight time budget.

Since the slices are expected to be provisioned dynamically with their service requirements not known a priori, the RAN slicing needs to be flexible enough to provide support for various RATs and protocol stacks on-the-fly, following a RAN as a Service (RaaS) paradigm. The RaaS requires going beyond virtualising the access point to providing virtual RAN control functions that can be stitched on the fly together.

Finally, these changes in the network architecture should require minimal modifications on the UE side to reduce the added complexity as much as possible. If the tenants are OTT service providers, it is possible that the same UE might belong simultaneously to different slices, which can give conflicting control commands to the UE. For example, each OTT service might keep a different Radio Resource Control (RRC) state. Such cases should be handled gracefully. Also, special care should be taken to forward UE related information, such as signal quality, only to the slices associated with it.

6.6 IoT Slicing

6.6.1 Motivation

Emerging 5G networking technologies can foster the IoT vision of seamless integration between the real world of Things and the Information Technology domain, both for business *e.g.*, Industry 4.0, and for everyday life scenarios *e.g.*, smart cities. The expected support for mMTC, promises the simplification of IoT infrastructure deployment and operation, via scalable cellular coverage. This in turn supports a convergence between the IoT platform and the telecommunication domains, allowing for the integration of IoT platform capabilities within the operational scope of 5G networks. It follows that the support of IoT-enabled vertical domains goes through

the envisioned 5G service orchestration mechanisms, allowing the flexible and programmable deployment and management of the corresponding services.

In this context, it is of particular interest to view the relation between network slicing capabilities, inherent in 5G network architectures, and the integrated IoT resources. The extension of the network slicing concept to the IoT domain, termed here as *IoT Slicing*, targets at enabling the sharing of the IoT infrastructure between multiple isolated tenants (slice owners), as a means towards CapEx/OpEx reduction for IoT applications. In an analogy to the sharing of computing resources in Platform as a Service (PaaS) environments, the objective is to present each tenant with a view seemingly identical to a non-virtualized (i.e., non-"sliced") environment of IoT resources, and their operational environment, that allows for the retrieval or sensor data, the potential support of actuation, as well as the support thin device configuration primitives.

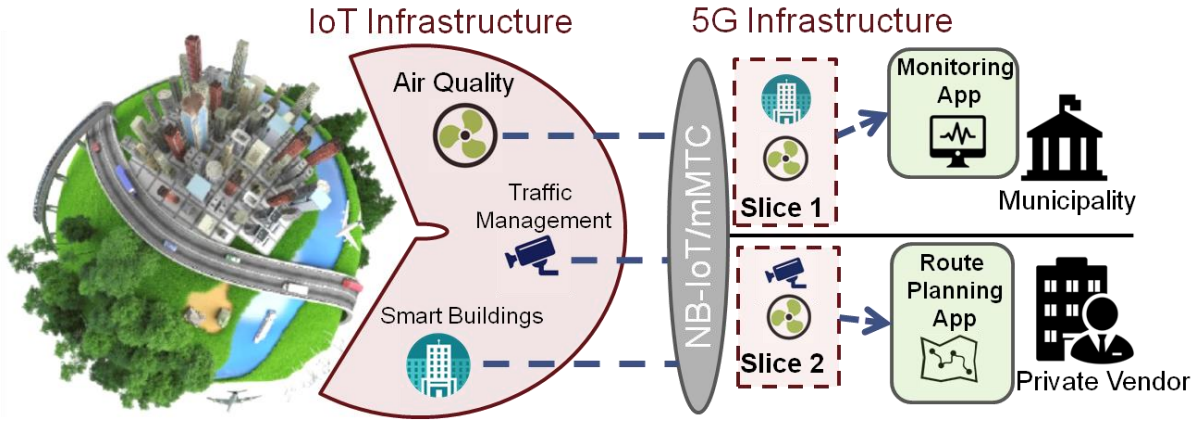


Figure 6.7: A high level view of the IoT Slicing concept

This approach facilitates the distinction of roles in an IoT ecosystem, guiding the various responsibilities and offloading them to different entities. Entities in this ecosystem can be the following:

- *Telecom Operators*—responsible for managing the physical & virtual IoT *infrastructure*, including operation and maintenance of physical/virtual resources as well as Slicing & Virtualization, interacting with actual IoT devices.
- *IoT (Platform) Slice owners*—responsible for developing & deploying their IoT services. Provided with mediated ownership of virtualized underlying devices and platforms, as well as configuration and management primitives within the limits of their slices.
- *IoT device providers*—responsible for physical infrastructure, i.e., smart sensors/actuators and edge equipment. They can be Telecom Operators, IoT Platform operators, civilians and essentially any actor in the IoT ecosystem able to register/connect their infrastructural resources to a designated IoT Gateway via open, standardized APIs.

In this ecosystem, IoT slicing aims to promote the efficient reuse of existing IoT infrastructure (IoT devices, gateways, southbound mechanisms/configurations, management), lowering CapEx/OpEx costs and removing entry barriers for IoT application owners. In addition, it allows the "Plug & Play" deployment of applications built for legacy IoT systems, since slice owners are presented with a virtualization-agnostic view of the underlying hardware, obsoleting the use of wrapper layers and glue code for interoperability.

As an example in context of the smart city domain, we showcase in Figure 6.7 a high level view of how IoT Infrastructural resources (smart devices, networking, control/management functionality) can be shared between isolated tenants—in this case the Municipality and a private Service Provider providing environmental friendly routes—who are each provided with a sandboxed environment (i.e., slice). In this environment, each tenant is presented with a private view over underlying HW resources and hence, they can interact with IoT resources and roll out their own smart city applications. The overall envisioned functionality is based on the support of IoT virtualization capabilities as explained in the following.

6.6.2 Approach

During the 5G-VINNI Project, research activities on IoT Slicing, focusing on the application of virtualization and orchestration technologies (NFV MANO) for the support of IoT Slicing Management, Operation and Automation mechanisms will be undertaken. In order to fully exploit separation of concerns and streamlining of IoT services on-boarding and deployment, the approach of slicing on the IoT Gateway (GW) level, following a PaaS sharing model can be employed. This sharing model targets IoT platform services including the exposure of IoT device, their capabilities and configuration interfaces; Rule Management; User Management, as well as traditional FCAPS tasks.

To this end, the envisioned approach foresees the realization of virtualized IoT GW instances dedicated to the support of the corresponding IoT slices, in the form of VNFs , termed (Figure 6.8). A north-bound interface (NBI) is expected to provide application support, in an IoT Slicing-orthogonal manner i.e., exposing existing IoT GW north bound interfaces, for instance MQTT based data delivery. A south-bound interface (SBI) handles the interconnection of IoT GW VNFs with an underlying IoT GW operated by the telecom operator. The latter supports the physical interface to the IoT infrastructure in an IoT Slicing-orthogonal manner. At the same time it realizes the IoT GW virtualization primitives by mediating access of IoT GW VNFs to the underlying infrastructure e.g., delivering sub-streams of data, realizing conflict resolution for actuators. This corresponds to the realization of a Virtualized IoT Infrastructure Management (VIoTIM) system, in an analogy to VIM for compute and storage resources, however bearing the PaaS flavor.

The envisioned solution will focus on the design and development of the mediating functionality as well as the orchestrated establishment of the corresponding interfaces.

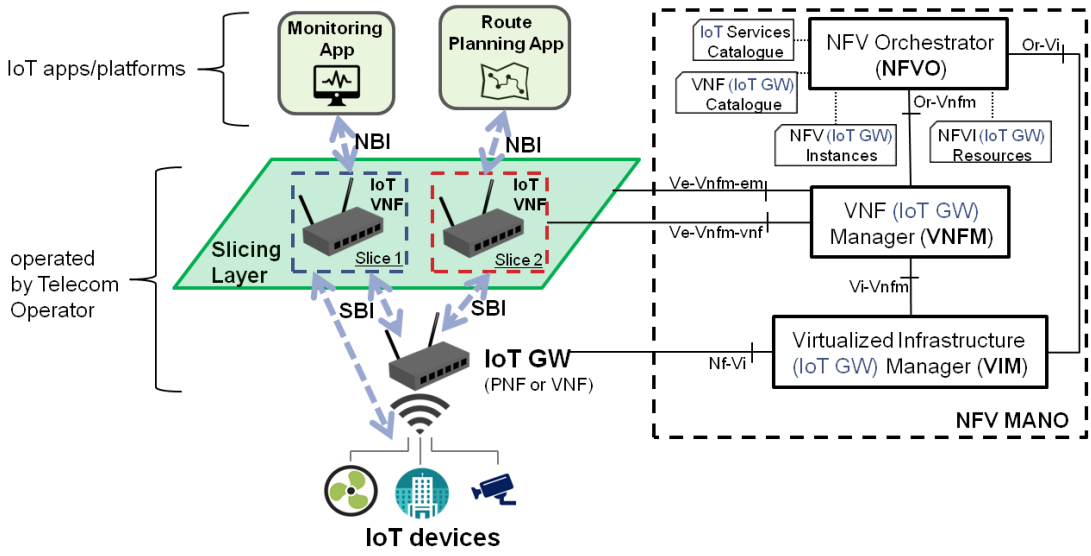


Figure 6.8: Slicing of the IoT Gateway and relation to NFV MANO

Research challenges of IoT Slicing, a subset of which will be investigated during the 5G-VINNI Project, include the following:

- Resource isolation and non-interference among tenants, especially with respect to scalability of computing/networking resources
- Conflict resolution/hierarchical control in case of multi-tenant actuation over the same physical devices. Policy based resolution mechanisms will be investigated.
- Integration of examined slicing mechanisms with MEC and 5G components
- Security, privacy and safety

References

[1] 3GPP TS 23.501: System Architecture for the 5G System
[2] 3GPP TR 22.891: Study on New Services and Markets Technology Enablers
[3] 3GPP TS 23.502: Procedures for the 5G System
[4] 3GPP TS 23.503: Policy and Charging Control Framework for the 5G System; Stage 2
[5] 3GPP TR 22.804: Study on Communication and Automation in Vertical domains (CAV)
[6] 3GPP TR 22.821: Feasibility Study on LAN support in 5G
[7] 3GPP TR 22.886: Study on enhancement of 3GPP support for 5G V2X services
[8] 3GPP TR 22.904: Study on User centric identifiers and authentication
[9] 3GPP TS 28.530: Management and orchestration; Concepts, use cases and requirements
[10] 3GPP TS 28.533: Management and orchestration; Architecture framework
[11] 3GPP TS 28.532: Management and orchestration; Generic management services
[12] 3GPP TS 28.554: Management and orchestration; 5G end to end Key Performance Indicators (KPI)
[13] 3GPP TS 23.222: Common API Framework for 3GPP Northbound APIs
[14] 5G-IA Architecture WG: View on 5G Architecture, Version 2.0
[15] O. Adamuz-Hinojosa, J. Ordonez-Lucena, P. Ameigeiras, J. J. Ramos-Munoz, D. Lopez and J. Folgueira, “Automated Network Service Scaling in NFV: Concepts, Mechanisms and Scaling Workflow”, in IEEE Communications Magazine, vol. 56, no. 7, pp. 162-169, July 2018.
[16] ETSI GS NFV-MAN 001: Network Functions Virtualisation (NFV); Management and Orchestration
[17] ETSI GS NFV-IFA 006: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Vi-Vnfm reference point - Interface and Information Model Specification
[18] A. Gonzalez, G. Nencioni, A. Kaminsiń, B.E. Helvik and P. E. Heegaard, “Dependability of the NFV Orchestrator: State of the Art and Research Challenges”, in IEEE Communications Surveys & Tutorials.
[19] ETSI GS NFV-IFA 007, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Or-Vnfm reference point - Interface and Information Model Specification”, V1.1.1, August 2018.
[20] ETSI GS NFV-IFA 005, “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Or-Vi reference point - Interface and Information Model Specification”, V1.1.1, August 2018.
[21] ETSI GS NFV-IFA 008: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Ve-Vnfm-em reference point - Interface and Information Model Specification
[22] ETSI GS NFV-IFA 013: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Os-ma-nfvo reference point - Interface and Information Model Specification
[23] ETSI GS ZSM 002:
[24] SaT5G Consortium, “SaT5G: Satellite and Terrestrial Network for 5G”, July 2017. Available: http://sat5g-project.eu.
[25] B. Tiomela Jou, O. Vidal, F. Arnal, J.-M. Houssin, K. Liolis, J. Cahill, H. Khalili, P. Sayyad Khodashenas, M. Boutin, D.-K. Chau, S. Sendra Diaz, “Architecture Options for Satellite Integration into 5G Networks”, 27th European Conference on Networks and Communications (EuCNC 2018), Ljubljana, Slovenia, June 2018.
[26] SATis5 Consortium, “SATis5 - Demonstrator for Satellite-Terrestrial Integration in the 5G Context”, ESA ARTES, October 2017. [Online]. Available: https://artes.esa.int/projects/satis5.
[27] M. Corici, K. Liolis, et al., “SATis5 Solution: A Comprehensive Practical Validation of the Satellite Use Cases in 5G”, 24th Ka and Broadband Communications Conference and 36th International Communications Satellite Systems Conference (ICSSC), Niagara Falls, Ontario, Canada, 2018.

[28] ETSI TR 103 611: Satellite Earth Stations and Systems (SES); Seamless integration of satellite and/or HAPS (High Altitude Platform Station) systems into 5G system and related architecture options
[29] 3GPP TR 23.737: Study on architecture aspects for using satellite access in 5G
[30] 3GPP TS 23.401: General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access
[31] 3GPP TS 23.402: Architecture enhancements for non-3GPP accesses
[32] 3GPP TS 23.203: Policy and charging control architecture
[33] 3GPP TR 38.801: Study on new radio access technology: Radio access architecture and interfaces
[34] 3GPP TR 28.801: Telecommunications management; Study on management and orchestration of network slicing for next generation networks
[35] ETSI TS 132 450: Key Performance Indicators (KPI) for Evolved Universal Terrestrial Ratio Access Network (E-UTRAN)
[36] IETF RFC 7426: Software-defined networking (SDN): Layers and architecture terminology.
[37] MEF Lifecycle Service Orchestration (LSO): Reference Architecture and Framework (2016)
[38] ETSI GS NFV-IFA 030: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Multiple Administrative Domain Aspect Interfaces Specification
[39] 5GEx Deliverable D2.2: 5GEx Final System Requirements and Architecture
[40] 5G-TRANSFORMER D1.2: 5G-TRANSFORMER Initial System Design
[41] Service Operations Specification MEF 55: Lifecycle Service Orchestration (LSO) : Reference Architecture and Framework
[42] TM Forum TR258: Mapping MEF Lifecycle Service Orchestration Reference Architecture to TMF APIs
[43] TM Forum Specification TMF641: Service Ordering Management API REST Specification
[44] TM Forum Specification TMF640: Activation and Configuration API REST Specification
[45] TM Forum Specification TMF645: Service Qualification API REST Specification
[46] TM Forum Specification TMF638: Service Inventory API REST Specification
[47] TM Forum Specification TMF633: Service Catalog Management API REST Specification
[48] ETSI GS NFV-SOL 005: Network Functions Virtualisation (NFV) Release 2; Protocols and Data Models; RESTful protocols specification for the Os-Ma-nfvo Reference Point
[49] IETF YANG Model - Cummings, Tara & Belotti, Sergio & Sharma, Anurag & Zhang, Xian & Xu, Yunbin & Ryoo, Jeong-dong & Shi, Yan & Jing, Ruiquan & King, Daniel – YANG Models for the Northbound Interface of a Transport Network Controller: Requirements and Gap Analysis. https://tools.ietf.org/id/draft-zhang-ccamp-transport-yang-gap-analysis-01.html
[50] D. Zhang and D. Ionescu, “QoS Performance Analysis in Deployment of DiffServ-aware MPLS Traffic Engineering”, Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007), pp. 963-967, Qingdao, 2007.
[51] Bruce S. Davie and Yakov Rekhter. 2000. MPLS: Technology and Applications (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA
[52] O. Michel and E. Keller, “SDN in wide-area networks: A survey”, 2017 Fourth International Conference on Software Defined Systems (SDS), pp. 37-42, Valencia, 2017.
[53] E. K. Ali, M. Manel and Y. Habib, “An Efficient MPLS-Based Source Routing Scheme in Software-Defined Wide Area Networks (SD-WAN)”, 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, pp. 1205-1211, 2017.
[54] M. Fiorani, et al., “On the Design of 5G Transport Networks”, Springer PNET, 2015.
[55] A. Mayoral, et al., “Need for a Transport API in 5G for Global Orchestration of Cloud and Networks through a Virtualized Infrastructure Manager and Planner”, JOCN, 2017
[56] ETSI GR NFV-IFA 022: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on Management and Connectivity for Multi-Site Service

[57] 3GPP TR 23.791: Study of enablers for Network Automation of 5G
[58] 3GPP TR 23.742: Study on Enhancements to the Service-Based Architecture
[59] 3GPP TS 23.725: Study on Enhancement of Ultra Reliable Low Latency Communication (URLLC) support in the 5G Core network (5GC)
[60] 3GPP TR 22.951: Service aspects and requirements for network sharing
[61] 3GPP TS 23.251: Network Sharing; Architecture and functional description
[62] Alaez, R. M., Calero, J. M. A., Belqasmi, F., El-Barachi, M., Badra, M., & Alfandi, O., "Towards an open source architecture for multi-operator LTE Core networks", Journal of Network and Computer Applications, 75, pp101-109, 2016.
[63] García, G., Gramaglia, M., Serrano, P., & Banchs, A., "POSENS: a practical open-source solution for end-to-end network slicing", IEEE Wireless Communications Magazine (Special Issue: 5G Testing and Field Trials), 2018.
[64] Foukas, X., Marina, M. K., & Kontovasilis, K., "Orion: RAN Slicing for a Flexible and Cost-Effective Multi-Service Mobile Network Architecture", Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (pp 127-140), October 2008.
[65] Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K., "Network slicing in 5G: Survey and challenges", IEEE Communications Magazine, 55(5), pp94-100, 2017.
[66] Rost, P., Mannweiler, C., Michalopoulos, D. S., Sartori, C., Sciancalepore, V., Sastry, N., ... & Aziz, D., "Network slicing to enable scalability and flexibility in 5G mobile networks", IEEE Communications magazine, 55(5), pp72-79, 2017.
[67] 3GPP TR 38.806: Study of separation of NR Control Plane (CP) and User Plane (UP) for split Option 2.
[68] ETSI GS MEC 013: Multi-Access Edge Computing; Location API
[69] 5GEx Deliverable D2.3: 5GEx Business and Economic Layer
[70] SliceNet Deliverable D2.2: Overall Architecture and Interfaces Definition
[71] ETSI GS MEC 003: Framework and Reference Architecture
[72] ETSI MEC Paper, MEC Deployments in 4G and Evolution Towards 5G
[73] ETSI MEC Paper, MEC in 5G Networks
[74] IETF RFC 4364: BGP/MPLS IP Virtual Private Networks (VPNs)

[end of document]

