DOCTORAL THESIS

Management and Orchestration of Network Slicing in Public-Private 5G Networks

University of Granada

Author:
Jose Antonio Ordóñez Lucena

Supervisors:
Dr. Pablo Ameigeiras Gutiérrez

A thesis submitted in fulfillment of the requirements
to obtain the Doctor degree as part of the
Doctoral Program in Information and Communication Technologies
in the

Wireless and Multimedia Networking Lab Research Group
Department of Signal Theory, Telematics and Communication

Granada, July 2022

Editor: Universidad de Granada. Tesis Doctorales
Autor: José Antonio Ordóñez Lucena
ISBN: 978-84-1117-566-1
URI: https://hdl.handle.net/10481/77685

“A man who cannot tolerate small misfortunes can never accomplish great things”

Chinese proverb

Acknowledgements

This thesis is the last step in an amazing journey that started six years ago. I want to thank all the people who have kept me on the right track and helped me bring this work to completion.

First, I would like to express my gratitude to my supervisor, Prof. Pablo Ameigeiras. Since I joined the Wireless Multimedia and Networking Lab (WiMuNet) group in 2015, with my first research grant, he provided me with support and advising during this research endeavor. But above all else, he has given me the opportunity to fulfill my wish of doing research. He has taught me practically everything I know about scientific research. And proof of this is the present dissertation, wherein the main findings are but the result of his valuable guidance. I could not have imagined having a better supervisor. This work is also yours.

I am also grateful to the rest of WiMuNet members for their comradeship as well. Pablo, Jorge and Juanjo, thanks for all these years with you. You have had an incredibly positive impact on me, not only on a research level, but also on a personal level. And how could I forget those moments with my lab mates? Óscar, Jonathan and Pilar are not only excellent researchers, but awesome friends. Thanks for having been there during the first steps in my research career, for encouraging me when things did not go well, for generously devoting much of your time to me, and for all the good moments we had together. I am fortunate to have them in my life, despite the miles that separate us. But that's nothing that telecommunications cannot solve, right? Finally, I would like to wish all the best to the young researchers who have joined the group over the past three years. Lorena, Natalia, Félix, Julia, Pablo… the future is yours.

Furthermore, I am in debt to all the colleagues I have had the opportunity to work with during these years in my professional life at Telefónica. Diego López, you are a well-respected authority in the field of network softwarization; indeed, the NFV MANO concept is your invention. You are an eminent researcher, and I am honored to be part of your team. Having the opportunity to learn from you every day is a privilege that not everyone has. And apart from being a great ham slicer and a "weirdo" in beer tastes, you're the perfect travel companion. Jesús Folgueira, thanks for having thought of me to join the company and bet on me to do great things. I recall our informal conversations every morning when arriving at the office, when we debated past episodes, present worries, and future plans. Your almost 30 years in Telefónica attest your proven experience, especially in what relates to day-to-day issues at operational telecom networks, which has served me a lot to give this dissertation a more operator-centric approach. And, for sure, thanks to Luis Miguel Contreras, whom I have the pleasure of co-authoring a lot of works and enjoying our ride back home in the metro. It's clear that the slicing topic has brought us closer; now, we need to bridge existing 3GPP-IETF gaps. Finally, I'd like to extend my gratitude to Jesús Martín, my team partner, and Juan Carlos García, my director. Jesús, thanks for your invaluable assistance, responsiveness and technical help since the first day I joined the company. You

are the real expert in standards, not only in the 3GPP technicalities, but also in the management of membership fees. Juan Carlos, you are a true leader. You have written glorious pages in the history of Telefónica, and I feel so fortunate to being now under your supervision. Thanks for giving me the opportunity to drive the slicing topic at Telefónica, and for your trust. This has enabled me to have confidence and feel there are no limits for me to keep growing within the company.

I have been also enriched by the interaction with so many people from other organizations (vendors, service providers, research institutions, universities, etc.) as result of the industrial and innovation activities carried out along these years. Part of my learnings from them are also captured in this work.

Lastly, I am obliged to Eva and my family, for all the time I took from them to conclude this work. My deepest and most loved gratitude to my brother, Pablo, for the particularly good moments I have shared with him during my life. And my biggest thanks go out to my parents, Jose and Leli, for instilling in me the values of hard work and perseverance, and for being there for me when I needed them, especially when I went through tough times. But overall, I am forever grateful for giving the gift of life, and for the sacrifices they made to give me the wonderful opportunities that I have always had. *Os quiero*.

José Antonio Ordóñez Lucena

Abstract

5G era is touted as the generation of mobile networks that will deliver an end-to-end ecosystem to enable a fully mobile and connected society. One of the most noticeable differences with respect previous generations is the irruption of the vertical industries (e.g., industry 4.0, transportation, energy, health) in the telco landscape. These *verticals* aspire to embrace 5G capabilities to shape and accelerate their digital transformation, in their mission to move towards a more modern business ecosystem, grounded upon service innovation and sustainability principles. The consequences of this paradigm shift are clear: in addition to supporting the evolution of the established prominent mobile broadband use cases, 5G will need to support countless vertical use cases with a high variety of applications and variability of their performance attributes, ranging from low bitrate high latency services to high bitrate low latency services, with many variants in between. These use cases will be also delivered across a wide range of devices (e.g., smartphone, IoT sensors, industrial equipment, vehicles), with quite different mobility patterns and energy consumption requirements each. Furthermore, 5G use cases will be provisioned using distributed infrastructures, including carrier networks but also 3rd party nodes that span beyond operators' footprint.

With this scene in mind, the challenge that lies ahead is how to accommodate all this casuistry in 5G, when most of these use cases will be active at the same time. The "one-size-fits-all" architectural approach that exists in today's networks is not feasible. The reason is that unlike 4G, which is user-centric, 5G will be industry-centric, so the design of one single physical network optimized to process mobile broadband traffic no longer makes sense. Additionally, the vertical use cases will have quite different (sometimes conflicting) service requirements, tied to Service Level Agreements (SLAs) that are much less flexible that those existing in traditional mass-market services. One solution could be to design dedicated networks per use case, following solutions such as those existing with DECOR. However, this solution is neither scalable (more and more use cases are defined every year) nor viable (too much upfront costs) for operators. In such a case, fresh solutions need to be explored. This is where the concept of network slicing fits in.

Network slicing is a solution whereby a physical network infrastructure is split into a set of logical network partitions, each tailored to satisfy the specific service requirements of a given vertical or use case. These partitions, referred to as network slices, are potentially operated isolated from each other but instantiated and running over the same physical network. ¡In 5G, the network will be a continuum spanning across different administrative domains. These domains typically correspond to infrastructures managed by the different mobile network operators, referred to as public land mobile networks (PLMNs); however, some of these domains can also correspond to private infrastructures managed by the verticals themselves, such as factories and transportation hubs (seaport, airport, etc.).

To make slicing happen, in thus needed to build an end-to-end (E2E) open infrastructure that integrates networking, computing and storage resources, together with technologies

for their segregation and programmability, to transform networks into a *flexible, reliable and secure well-orchestrated facility across multiple administrative domains*. Within this high-level objective, the goal of this dissertation is to design and validate solutions for network slicing management and orchestration in multi-domain environments, with applicability in public and private 5G network scenarios. To that end, this thesis is structured into three main workstreams.

The first one corresponds to the design of system architecture solution or multi-domain network slicing. This architecture will build upon network softwarization technologies (Software Defined Networking and Network Functions Virtualization) together with their orchestration-enabling artifacts, and integrate them into a robust, scalable, and standards-compliant system. The resulting system will provide all the necessary capabilities that shape network slicing concept (including isolation, customization, elasticity, programmability, and automation, among others), with multiple service-tailored logical networks running atop a distributed yet common physical network. Special focus is put on the aspects related to the lifecycle management of these network slices, from their design to their provisioning until their termination, with operation in between.

The second line of work will be focused on the implementation of the system architecture, and the prototype validation in 5G experimentation facility. The selected environment is 5G-VINNI. 5G-VINNI is a large-scale E2E infrastructure consisting of 5G nodes that are distributed across Europe, and that provides a testing and validation environment for vertical use case experimentation. For the system implementation, Open Source MANO (OSM) stack will be combined with Openslice. The result is an E2E management solution suite providing the network and service orchestration capabilities which are needed to partition infrastructure resources and allocate them to different slices, at both provisioning and operation time. This solution will incorporate federation capabilities, to facilitate operation in multi-domain environments. Finally, a Proof-of-Concept (PoC) will be set up, to showcase the behavior of the solution in different use cases: network slice design (creating a network slice descriptor and onboarding it to the catalog), network slice provisioning (commissioning a network slice upon a vertical-triggered service order) and scaling (increase the capacity of the slice, by allocating more resources). The PoC also exhibits also multi-domain aspects; to that end, Spanish and Greek nodes from 5G-VINNI facility have been selected.

The third and last line of work focuses on the study of the private 5G market, and the analysis of network slicing role in it. Unlike private Long-Term Evolution (LTE), based on the use of infrastructures totally separated from the public LTE network, the expected continuum in 5G will span nodes from public and private infrastructures. Actually, there are many verticals that require executing E2E services spanning these two types of infrastructures, with some workloads running on-premises (private infrastructure) and some others on the operator's footprint (public infrastructure). To break silos in this public-private network infrastructures and facilitate a seamless provisioning of E2E service across them, slicing is identified as "the solution". However, for this to happen, we need to first understand the private 5G ecosystem, why verticals prefer having standalone private 5G networks in the short term, and their motivations to partially migrate 5G public network in the medium and long term. Based on this understanding, which includes findings from business and technology viewpoints, we will outline a radar for network slicing roll-out in telco networks, to accompany the verticals in the transitions, managing their expectations of what capabilities will be available by when, and under which conditions.

Table of Contents

Declaration of authorship iii

Acknowledgements v

Abstract vii

Table of Contents ix

List of Abbreviations xvii

Thesis Details xxi

I Introduction 1

Setting the Scene 3

1 5G system and services 3
1.1 First steps of 5G 3
1.2 5G service categories 4
1.3 An adequate approach for 5G system design 5
2 Network slicing concept and principles 6
3 Network slicing: public or private networks? 8
4 Network slicing pillars 11
4.1 Network softwarization technologies 11
4.2 Edge computing 14
4.3 Management and Orchestration 15
4.4 Putting it all together 16
5 Network slicing landscape 16
5.1 Vertical Industry Organizations 17
5.2 Telco Industry Organizations 17
5.3 Standard Development Organizations 18
5.4 Open source and integration communities 19
5.5 Timing 19
6 Main challenges on network slicing 20
6.1 Technical challenges 20
6.2 Non-technical challenges 23
7 Scope and Objectives of the Thesis 25
8 Research Methodology 28
9 Impact Creation 29
10 Publications 31
11 Thesis Outline 31
References 32

II Multi-domain Network Slicing: System design 35

Literature Review and Problem Description 37

1 Background Context 37
1.1 Network slice concept 37
1.2 System architectures combining SDN and NFV capabilities. 38
1.3 Multi-domain orchestration 40
1.4 Network slice characterization 42
2 Ambition 43
2.1 Identification of key network slice capabilities 43
2.2 Standards-compliant network slicing architecture 44
2.3 Model-based network slice description 45
References 46

A Network Slicing for 5G with SDN/NFV: Concepts, Architectures and Challenges 49

Abstract 50
1 Introduction 50
2 Background on key concepts for Network Slicing 51
2.1 Resources 51
2.2 Virtualization 52
2.3 Orchestration 53
2.4 Isolation 54
3 ONF Network Slicing Architecture 54
4 NFV Reference Architectural Framework 56
5 Network Slicing use case with SDN-NFV Integration 58
6 Challenges and Research Directions 60
6.1 Performance issues in a shared infrastructure 60
6.2 Management and orchestration issues 61
6.3 Security and privacy 61
6.4 New business models 61
Acknowledgement 61
References 62

B The Creation Phase in Network Slicing: From a Service Order to an Operative Network Slice 63

Abstract 64
1 Introduction 64
2 NFV Network Services and Network Slices 65
3 Network Slicing Architecture 67
4 Network Slice Creation Phase 68
4.1 Service Ordering 69
4.2 Network Slice Resource Description 70
4.3 Admission Control 71
4.4 Optimization and Resource Reservation 72
4.5 Network Slice Preparation 72
5 Conclusions 73
Acknowledgement 74
References 74

III Multi-domain Network Slicing: Prototyping and Validation 75

Literature Review and Problem Description 77

1 Background Description 77
1.1 Progress on research and innovation projects 77
1.2 Progress on the standardization arena 79
1.3 5G experimentation facilities 85
1.4 Network slicing solutions and PoCs 86
2 Ambition 90
2.1 Network slicing orchestration in action: on-boarding 90
2.2 Network slicing orchestration in action: provisioning a multi-domain network slice 96
2.3 Network slicing orchestration in action: multi-domain slice auto-scaling..........97
References 97

C Modelling Network Slice as a Service in a multi-vendor 5G experimentation ecosystem 103

Abstract 104
1 Introduction 104
2 Related Work 106
2.1 Telecom Industry Organizations 107
2.2 Standard Development Organizations 107
2.3 Network Slicing in 5G-VINNI facility 108
3 Network Slices Modelling 108
3.1 The scope of VINNI-SB 109
3.2 The model of VINNI-SB 110
4 5G-VINNI Service Catalog 111
5 5G-VINNI Portal 112
5.1 Openslice framework 112
5.2 Service Deployment 113
6 Conclusions and Future Work 114
Acknowledgement 114
References 114

D Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility 117

Abstract 118
1 Introduction 118
2 Related Work 120
3 5G-VINNI Facility Architecture 121
3.1 Facility Site 122
3.2 Service Catalog 123
3.3 Portal 124
4 Cross-Domain Slice Operation 124
4.1 Federation Options 125
4.2 Solution for the Federation at the Service Orchestration Layer 126
5 Federation Use Case 127
5.1 Slice Ordering 127
5.2 Slice Fulfilment 128

5.3 Slice Operation 128
6 Conclusions and Future Work 128
Acknowledgement 129
References 129

E Automated Network Slice Scaling in Multi-site Environments: The ZSM PoC #2 report 131

1 PoC Project Details 132
1.1 PoC Project Review 132
1.2 PoC Team Members Review 132
1.3 PoC Project Scope Review 133
1.4 PoC Project Milestones Review 134
1.5 Confirmation of PoC Event Occurence 134
1.6 Other Dissemination Activities 135
2 ZSM PoC Technical Report 136
2.1 PoC Project Milestones Review 136
2.2 General 145
2.3 Gaps identified in ZSM standardization 145
2.4 PoC Suggested Action Items 146
2.5 Additional Messages to ZSM 147
2.6 Additional messages to Network Operators and Service Providers 147
References 147

IV Private 5G Networks and Network Slicing: The Art of Living Together 149

Literature Review and Problem Description 151

1 Background context 151
1.1 Public vs private mobile networks 152
1.2 Today’s private mobile network solutions 154
1.3 Private 5G: As capable as wires, but without the wires 155
1.4 Putting all together 158
2 Ambition 158
2.1 Profiling Private 5G in industry 4.0 158
2.2 Robust, scalable and future-proof solutions for private 5G in the vertical industries ecosystem 159
2.3 Provision of private 5G networks with network slicing 161
References 163

F The use of 5G Non-Public Networks to support Industry 4.0 scenarios 167

Abstract 168
1 Introduction 168
2 Overview of the 5G System Architecture 170
3 5G-enabled NPNs 172
3.1 Stand-alone NPN 172
3.2 Public Network Integrated NPN 173
4 Analysis of NPN Attributes 174
4.1 QoS Customization 175
4.2 Autonomy 175

4.3 Isolation 175
4.4 Security 175
4.5 Service Continuity 176
4.6 NPN management for verticals 176
4.7 Entry barriers for verticals 176
5 Conclusions and Future Outlook 177
Acknowledgement 178
References 178

G 5G Non-Public Networks: Standardization, Architectures and Challenges 181

Abstract 182
1 Introduction 182
2 Key Aspects for 5G NPNs 183
2.1 Spectrum 183
2.2 Interworking 183
2.3 Transport networks 184
2.4 Positioning 184
2.5 Hardware acceleration 185
2.6 Security and Privacy 185
3 3GPP Related Standardization 185
3.1 3GPP Support for Non-Public Networks 185
3.2 Network Sharing 186
4 Single-site NPN Architectures 187
4.1 Stand-alone NPN Architecture 187
4.2 PNI-NPN Architecture 188
4.3 On-premises RAN Sharing Scenario and Architecture 190
5 Mobility and Multi-site NPN Scenarios 192
5.1 Mobility in NPNs 192
5.2 Multi-site NPN 192
6 Challenges 194
6.1 Zero-touch Practices on NPN Management 194
6.2 Multi-WAT in 5G NPNs 194
6.3 Enabling and Validating 5G NPNs with E2E Deterministic QoS Support 195
6.4 Capability Exposure in PNI-NPN 195
Acknowledgement 196
References 196

H On the Rollout of Network Slicing in Carrier Networks: A Technology Radar 199

Abstract 200
1 Introduction 200
2 Network Slicing: Concept, System Architecture and Deployment 203
2.1 Network slice concept 203
2.2 Architectural framework for Network Slicing 205
2.3 Network Slice Description 207
2.4 From a Service Order to a Deployed Network Slice 209
3 Impact of Network Slicing 210
3.1 CN Slicing 210
3.2 RAN Slicing 212
3.3 TN Slicing 216

4 Network Slicing Technology Radar ... 219
5 CN Domain ... 221
5.1 Functionality ... 221
5.2 Add-on Features ... 225
6 RAN Domain ... 227
6.1 Functionality ... 228
6.2 Radio Resource Allocation ... 231
6.3 Penetration ... 235
7 TN Domain ... 236
7.1 Transport Technologies ... 236
7.2 Transport SDN Fabric ... 239
8 OSS Domain ... 243
8.1 OAM ... 243
8.2 Capability Exposure ... 246
9 Conclusions ... 249
Author Contributions ... 250
Funding ... 251
Institutional Review Board Statement ... 251
Informed Consent Statement ... 251
Data Availability Statement ... 251
Conflicts of Interest ... 251
Abbreviations ... 251
References ... 253

V Conclusions and Remarks ... 261

Conclusions ... 263

1 Main Findings ... 263
2 Future Work ... 268
3 References ... 270

VI Appendices ... 271

Annex A: Other merits related to research activities ... 273

1 Additional publications ... 273
1.1 Journals ... 273
1.2 Conferences ... 274
2 Standardization ... 274
2.1 3GPP SA5 ... 275
2.2 ETSI ZSM ... 276
2.3 IETF ... 276
2.4 GSMA NG ... 277
2.5 ETSI NFV ... 277
3 Talks ... 277
4 International research and innovation projects ... 278

Annex B: Resumen ... 281

1 Introducción a los sistemas 5G ... 281

1.1 Casos de uso 281
1.2 Una aproximación al diseño de sistemas 5G 282
2 Network slicing: concepto y principios 282
3 Network slicing: ¿redes públicas o privadas? 284
4 Pilares de network slicing 285
4.1 Tecnologías de *softwarización* de red 285
4.2 Computación en el borde 286
4.3 Gestión y orquestación 286
4.4 Resumen 287
5 Principales desafíos en network slicing 287
5.1 Desafíos técnicos 287
5.2 Desafíos no técnicos 290
6 Alcance y Objetivos de la Tesis 290
7 Conclusiones 292

List of Abbreviations

3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5G-ACIA	5G Alliance for Connected Industries and Automation
5G-MAG	5G Media Action Group
5G-PPP	5G Infrastructure Public Private Partnership
5GAA	5G Automotive Association
5GC	5G Core
5GMF	5G Mobile Forum
5QI	5G Quality Indicator
AI	Artificial Intelligence
AMBR	Aggregate Maximum Bit Rate
AMF	Access and Mobility management Function
API	Application Programming Interface
ASIC	Application Specific Integrated Circuit
ATCN	Abstraction and Control of Transport Networks
B2B	Business-to-Business
B2C	Business-to-Customer
BSS	Business Support System
CaaS	Container-as-a-Service
CN	Core Network
CNF	Cloud-Native Function
COTS	Commercial-off-the-Shelf
CSMF	Communication Service Management Function
CU	Centralized Unit
DN	Data Network
DRB	Dedicated Radio Bearer
DSCP	DiffServ Code Point
DU	Distributed Unit
DWDM	Dense WDM
E2E	end-to-end
EC	European Commission
eCPRI	Enhanced Common Public Radio Interface
EDSO	European Distribution System Operators
eMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
Flex-E	Flexible Ethernet
Flex-O	Flexible Optical Transport Network
FPGA	Field Programmable Gate Array

GBR	Guaranteed Bit Rate
GSMA	GSM Alliance
GST	Generic network Slice Template
IaaS	Infrastructure-as-a-Service
ICT	Information and Communication Technologies
IOC	Information Object Class
ISG	Industry Specification Group
IT	Information Technology
ITU	International Telecommunications Union
K8s	Kubernetes
L4S	Low Latency, Low Loss, Scalable throughput
LSP	Label Switched Path
MANO	Management and Orchestration
MBR	Maximum Bit Rate
MDAS	Management Data Analytics Service
MEC	Multi-access Edge Computing
MEF	Metro Ethernet Forum
mIoT	Massive Internet of Things
mMTC	Massive Machine-Type Communications
MPLS	Multiple Protocol Label Switching
MW	Microwave
NBI	NorthBound Interface
near-RT RIC	near-Real Time RIC
NEF	Network Exposure Function
NEST	Network Slice Type
NFV	Network Functions Virtualization
NFV	Network Functions Virtualization
NFVO	NFV Orchestrator
NG-RAN	Next Generation RAN
NGMN	Next Generation Mobile Network
non-RT RIC	Non-Real Time RIC
NPN	Non-Public Network
NR	New Radio
NRM	Network Resource Model
NSA	Non-Standalone
NSaaS	Network Slice as a Service
NSCAF	Network Slice Access Control Function
NSD	Network Service Descriptor
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NSSAI	Network Slice Selection Assistance Information
NSSI	Network Slice Subnet Instance
NSSMF	Network Slice Subnet Management Function
NSST	Network Slice Subnet Template
NST	Network Slice Template
NWDAF	Network Data Analytics Function
OAM	Operation, Administration and Maintenance
ODA	Open Digital Architecture
ONF	Open Networking Foundation
OS	Operation System

OSS	Operations Support System
OT	Operational Technology
PaaS	Platform-as-a-Service
PCE	Path Computation Element
PCF	Policy Control Function
PDU	Packet Data Unit
PLMN	Public Land Mobile Network
PNI-NPN	Public Network Integrated Non-Public Network
PNI-NPN	Public Network Integrated NPN
PoC	Proof of Concept
PoP	Point of Presence
PRB	Physical Radio Block
PSCE	Public Safety Communication Europe
RAN	Radio Access Network
RAT	Radio Access Tecnology
RIC	RAN Intelligent Controller
ROADM	Reconfigurable Optical Add-Drop Multiplexer
RRM	Radio Resource Management
RU	Radio Unit
S-NSSAI	Single Network Slice Selection Assistance Information
SA	Standalone
SBA	Service Based Architecture
SBI	SouthBound Interface
SBMA	Service Based Management Architecture
SD	Slice Differentiation
SD-WAN	Software Defined WAN
SDN	Software Defined Networking
SDN	Software Defined Networking
SDO	Standards Development Organization
SDO	Standards Development Organization
SLA	Service Level Agreement
SMF	Session Management Function
SNPN	Standalone Non-Public Network
SNPN	Standalone NPN
SR	Segment Routing
SST	Slice/Service Type
T-NSC	Transport Network Slice Controller
TAC	Tracking Area Code
TAI	Tracking Area Identifier
TE	Traffic Engineering
TEAS	Traffic Engineering Architecture and Signaling
TM Forum	Tele Management Forum
TN	Transport Network
TSN	Time Sensitive Network
UE	User Equipment
UPF	User Plane Function
uRLLC	Ultra-Reliable Low Latency Communications
URSP	UE Resource Selection Policy
VDU	Virtual Deployment Unit
VLL	Virtual Leased Line

VNF	Virtualized Network Function
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPRN	Virtual Private Routed Network
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
XR	Immersive Reality
YANG	Yet Another Next Generation
ZSM	Zero-touch network and Service Management

Thesis Details

Thesis Title: Management and Orchestration of Network Slicing in Public-Private 5G Networks
Ph. D. Student: José Antonio Ordóñez Lucena
Supervisors: Prof. Pablo Ameigeiras Gutiérrez, University of Granada

This PhD thesis is the outcome of years of research at the Wireless, Multimedia and Networking Lab (WiMuNet) group, from the Department of Signal Theory, Communications of the University of Granada, Spain. This research has been performed in close collaboration with Telefónica I+D, which is the research and innovation branch of Telefónica group.

This thesis is elaborated following the compendium format. The main body of this thesis consists of the following papers.

[A] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges", in *IEEE Communications Magazine,* vol. 55, no. 5, pp. 80-87, May 2017. DOI: 10.1108/MCOM.2017.1600935. Impact Factor = 9.27 (2/87 Q1; Category: TELECOMMUNICATIONS).

[B] J. Ordonez-Lucena, O. Adamuz-Hinojosa, P. Ameigeiras, P. Munoz, J. J. Ramos-Munoz, J. Folgueira and D. Lopez, "The Creation Phase in Network Slicing: From a Service Order to an Operative Network Slice", in *2018 European Conference on Networks and Communications (EuCNC),* Ljubljana, Slovenia, 2018, pp. 1-36. DOI: 10.1109/EuCNC.2018.8443255

[C] J. Ordonez-Lucena, C. Tranoris and J. Rodrigues, "Modeling Network Slice as a Service in a Multi-Vendor 5G Experimentation Ecosystem," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops),* 2020, pp. 1-6. DOI: 10.1109/ICCWorkshops49005.2020.9145225.

[D] J. Ordonez-Lucena, C. Tranoris, J. Rodrigues and L. M. Contreras, "Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility," in *2020 European Conference on Networks and Communications (EuCNC),* 2020, pp. 40-45. DOI: 10.1109/EuCNC48522.2020.9200940.

[E] J. Ordonez-Lucena, C. Tranoris and B. Nogales, "Automated Network Slice Scaling in Multi-site Environments: The ZSM PoC#2 report", 2021.

[F] J. Ordonez-Lucena, J. F. Chavarria, L. M. Contreras and A. Pastor, "The use of 5G Non-Public Networks to support Industry 4.0 scenarios", in *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2019, pp. 1-7. DOI: 10.1109/CSCN.2019.89313

[G] J. Prados-Garzon, P. Ameigeiras, J. Ordonez-Lucena, P. Muñoz, O. Adamuz-Hinojosa and D. Camps-Mur, "5G Non-Public Networks: Standardization, Architectures and Challenges," in IEEE Access, vol. 9, pp. 153893-153908, 2021. DOI: 10.1109/ACCESS.2021.3127482. Impact Factor = 3.476 (43/94 Q2; Category: TELECOMMUNICATIONS).

[H] J. Ordonez-Lucena, P. Ameigeiras, L. M. Contreras, J. Folgueira, D. R. López, "On the Rollout of Network Slicing in Carrier Networks: A Technology Radar", *Sensors*, 2021, 21, 8094. DOI: 10.3390/s21238094. Impact Factor = 3.847 (95/276 Q2; Category: ENGINEERING, ELECTRICAL & ELECTRONIC).

Part I

Introduction

Setting the Scene

1 5G system and services

5G communication systems are expected to enable a major societal transformation that will provide people, business, and governments with unprecedented capabilities to share information. Since the start of the exploratory phase in the early months of 2014, academy and research institutions actively have scouted technological innovations for 5G. Their aim was to transform future networks into secure, reliable, and flexible orchestration platforms able to satisfy the service demands of the 2020 decade.

1.1 First steps of 5G

Discussions on visions, requirements, and technologies for 5G mobile communication systems have already been gone by many organizations. For instance, the ITU-R Study Group 5 Working Party 5D (WP5D) issued a new recommendation named "IMT Vision-Framework and overall objectives of the future development of IMT for 2020 and beyond" [1], where some of these discussions were summarized in a consistent way. In addition, technical studies on 5G gained attention worldwide as evidenced by the acceleration of efforts that governmental entities and research bodies (from both academia and industry) have made. Indeed, several governments and groups of commercial companies and academic institutions set up projects and fora to study and promote 5G mobile technology. Examples of projects and initiatives with focus on 5G included the METIS project [2], the 5G Infrastructure Public Private Partnership (5G-PPP) [3], the Next Generation Mobile Networks (NGMN) Alliance [4] in Europe, the 5G Mobile Communications Promotion Forum (5GMF) [5] in Japan, and the 5G Forum [6] in Korea.

After several years of research by these projects and initiatives on future communication systems, there is a wide consensus on the 5G service landscape, particularly on the view that 5G will not only be a natural evolution of current mobile broadband networks. Besides having enhanced network capabilities (e.g., lower latencies, improved coverage, higher spectral efficiency, and higher peak throughputs), 5G systems may be programmable service enablement platforms, empowering business innovation. On the one hand, 5G will integrate resources into one unified programmable infrastructure. This programmability and unification may enable an optimized and more dynamic usage of resources, the convergence of fixed, mobile and broadcast services, and the accommodation of diverging use cases in an efficient and prompt manner. On the other hand, 5G will create a software-driven ecosystem, attracting fresh players apart from the traditional telco actors. In addition to chipset manufacturers, vendors and mobile (virtual) network operators, as well as solution integrators. All these stakeholders need to coexist in a "coopetitive" (i.e., cooperative and competitive) ecosystem which ultimately will unleash service and business innovation.

1.2 5G service categories

Communication services in 5G likely have a plethora of performance requirements, ranging from low bitrate high latency services to high bitrate low latency services, with variants in between. Indeed, 5G systems may support services requiring very low and very high bandwidths, very dense connectivity as well as coverage of areas where few connections are active at a given moment, centralization of functions to drive economy of scale, but also distribution of functions very close to the user to reduce latency and deliver content with local context. As seen, these requirements are not only quite different, but also incompatible with each other in some cases.

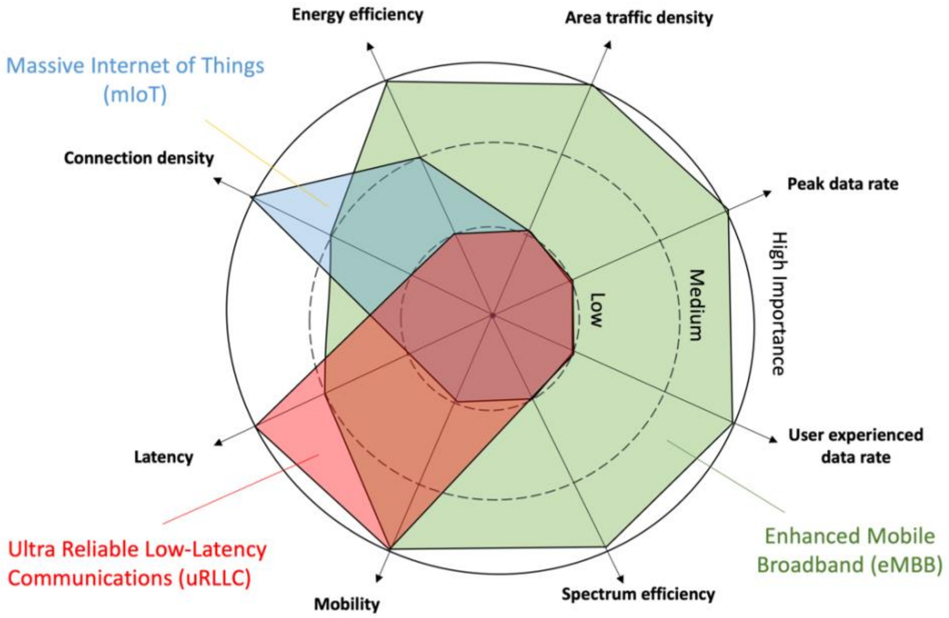


Figure I-1. Main 5G service categories.

To deal with this heterogeneity, industry agreed on the needed to define a classification where services with similar requirements could be grouped into the same category. After use case profiling and clustering, three different service categories were defined: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communication (uRLLC) and Massive Internet of Things (mIoT). This classification has become de facto for 5G, and assumes the following: depending on its performance requirements, any 5G communication service can fall into one of the following categories (see Figure I-1):

- **Enhanced Mobile Broadband (eMBB)**. This category includes communication services able to provide high data rates and supporting high traffic density values. eMBB service type is meant to extend traditional MBB scenarios, particularly those that require higher data rates, lower latency, and a more seamless user experience (reliable broadband access) over large coverage areas. eMBB distinguishes between two types of scenarios: hotspot scenarios and wide-area coverage scenarios. On the one hand, hotspot scenarios handle static users in indoor environments with high throughput requirements and with no (or very low) mobility. On the other hand, wide-area coverage scenarios support users in outdoor environments with medium to high mobility, providing them with uniform and seamless service experience.

- **Ultra-Reliable Low Latency Communications (uRLLC)**. This category focuses on providing an ultra-responsive connection with single-digit-millisecond latencies. The data rate is not expected to be very high, but offers high availability support. Potential uRLLC type services include industrial automation, mission-critical applications, or remote medical assistance, among others.
- **Massive Internet of Things (mIoT)**. Also referred to as massive Machine-Type Communications (mMTC), this category includes communication services offering improved network coverage with high connection density support. Most of these services involve the presence of a very large number of power constrained IoT devices with long operational lifetime, each transmitting small volumes of non-delay sensitive data to exchange information with other devices or apps hosted in remote cloud servers

As the work progressed, very particular use cases appeared, with requirements that could not be appropriately accommodated into one of these three categories, nor a combination of them. In response to them, additional 5G service categories needed to be defined, such as Vehicle-to-Everything (V2X, in late 2017) and High-Performance Machine-Type Communications (HMTC, in 2021).

1.3 An adequate approach for 5G system design

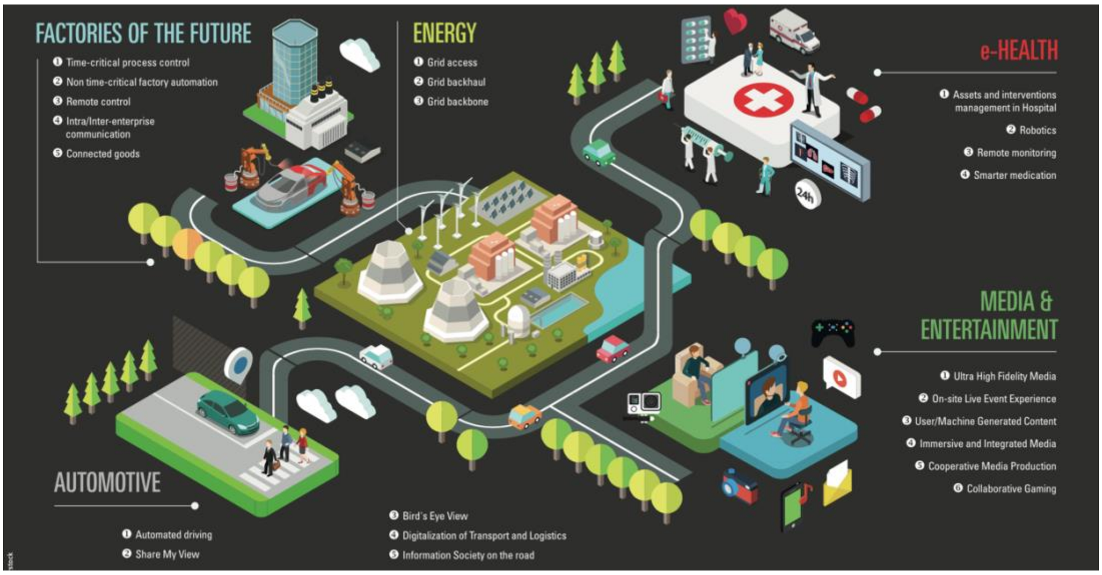

FACTORIES OF THE FUTURE
❶ Time-critical process control
❷ Non time-critical factory automation
❸ Remote control
❹ Intra/Inter-enterprise communication
❺ Connected goods
ENERGY
❶ Grid access
❷ Grid backhaul
❸ Grid backbone
e-HEALTH
❶ Assets and interventions management in Hospital
❷ Robotics
❸ Remote monitoring
❹ Smarter medication
24h
MEDIA & ENTERTAINMENT
❶ Ultra High Fidelity Media
❷ On-site Live Event Experience
❸ User/Machine Generated Content
❹ Immersive and Integrated Media
❺ Cooperative Media Production
❻ Collaborative Gaming
AUTOMOTIVE
❶ Automated driving
❷ Share My View
❸ Bird's Eye View
❹ Digitalization of Transport and Logistics
❺ Information Society on the road

Figure I-2. Examples of vertical industries. Source: [7]

One of the main key drivers of future 5G systems is the need to simultaneously accommodate a large variety of innovative use cases over a common network infrastructure in an efficient way. The use cases that 5G must support will not only include traditional data and internet services, but also services from enterprise customers that belong to non-telco industries, typically referred to as vertical industries (see Figure I-2). These *vertical customers* will tap into 5G to accelerate their digital transformation, in their mission to contribute to build a modern economy and a smart society. The aim is that 5G can provide an end-to-end (E2E) infrastructure capable of delivering a consistent Quality of Service (QoS) and Quality of Experience (QoE) across all these use cases. Achieving this goal would not be a problem if it were not for the fact that these use cases impose a much wider range of requirements than existing services do nowadays.

Existing network deployments are relatively static and monolithic. They are provisioned with a single set of resources and standard mobile network functions to support the MBB use cases, including high-data-rate mobile traffic from smartphones, applications from Over-the-Top (OTT) service providers, and some basic MTC devices for initial IoT. This “one-size-fits-all” architectural approach is not flexible and scalable enough to efficiently support the use cases that the 2020 timeframe will bring, particularly when having such a disparate set of service requirements. To efficiently accommodate future use cases along with increased demands for existing services over the same network infrastructure, it is assumed that 5G systems will require architectural enhancements with respect to current deployments.

One solution would be to design a network infrastructure able to simultaneously offer extremely high data rates, single-digit-millisecond latencies, and very high connection densities. However, designing such a network with a single set of resources and standard network functions as 4G networks do today, would be extremely complex (from a technical viewpoint) and prohibitive (from a cost viewpoint). For this reason, new solutions must be explored. This is where the concept of network slicing fits in. This concept is discussed in the next section.

2 Network slicing concept and principles

5G are called to support services targeting a wide variety of vertical customers, each having quite different (even conflicting) requirements. Additionally, some of these services need to be executed concurrently. This entails a great challenge, considering that a single 5G infrastructure should be able to satisfy diverging service requirements, potentially at the same time. To cope with this situation in a sustainable way, network slicing appears.

Network slicing is a cutting-edge solution that aims at splitting the infrastructure into a set of logical network partitions, each optimized (in terms of resources, topology, functions, configuration, and management) to satisfy a particular set of service requirements. A network slice is composed of a collection of 5G network functions and specific Radio Access Technology (RAT) settings that are combined to provide tailored control and user plane capabilities [8]. Not all slices contain the same functions, and some functions that today seem essential for a mobile network might even be missing in some of the slices. The intention of a network slice is to provide only the traffic treatment that is necessary for served use case(s), avoiding all other unnecessary functionality.

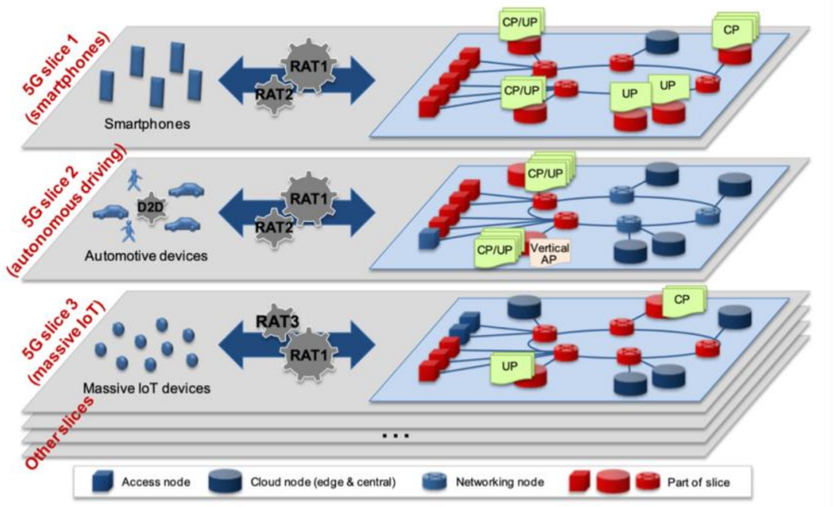


Figure I-3. Network slicing concept. Source: [8]

Figure I-3 illustrates an example of multiple slices concurrently operated on the same infrastructure. For example, a network slice for typical smartphone use can be realized by setting fully-fledged functions distributed across the network. For the network slice supporting automotive use case, security, reliability, and bounded latency will be critical; hence the need to implement redundancy and protection solutions and have user plane traffic processed as close as possible to vehicles. And for a network slice supporting sensors deployed at a large-scale, some basic control plane functions can be configured, omitting e.g., any mobility functions, with contention-based resources for access.

The survey in [9] reports a non-exhaustive list of main principles that shape network slicing concept and related capabilities. These include:

- **Isolation**. This property ensures the independent behavior and operation of individual slices, even though they all run on top of common infrastructure. This means that configuration, performance degradation or any security issue on a slice shall not have an impact on the rest of slices.
- **Customization**. This property ensures that each slice is provisioned only with the resources and configuration settings that are needed to satisfy the specific requirements of served use cases. This customization can be applied to different dimensions, including: i) slice capacity, allocating more or less compute and connectivity resources; ii) slice topology, with different variants in the number of nodes and the number of paths across them; iii) in-slice user plane traffic, by selecting different service-tailored network functions and configure them with appropriate forwarding policies; iv) in-slice control plane traffic, with different signaling protocols and mobility/connection/session management settings; and v) slice add-ons, with the possibility to include value-added applications (e.g., analytics, IoT servers, firewalls, traffic optimizers), as per service needs.
- **Elasticity**. This property refers to the ability of the slice to be re-sized, increasing and decreasing its capacity as needed. Elasticity allows a slice to always meet the SLA of hosted services, regardless of changing network conditions (e.g., traffic load surges, radio interface variability, faulty nodes), by allocating and deallocating resources accordingly, in a dynamic manner.
- **Programmability**. This property allows managing a slice as a software object. This means that i) slice resources and their properties can be modeled and captured in a machine-readable file; and ii) the allocation and configuration of these resources can be controlled with the use of Application Programming Interfaces (APIs).
- **End-to-end (E2E)**. The network slice provides a service-tailored connectivity pipe between two endpoints: a terminal connected to the radio node, and an application server hosted in a data network. In this path, the traffic traverse different domains, including radio access network (RAN), transport network (TN) and core network (CN) domains, some of them with different technologies (e.g., IP and optics in the transport network).
- **Hierarchical abstraction**. This feature has its roots on recursive virtualization, which allows repeating resource abstraction in a layered pattern. With this capability, the resources allocated for a slice can be further "sliced", with the goal of being traded to a 3rd party. A clear example would be a network operator delivering an eMBB slice to a mobile virtual network operator (MVNO), and the latter using the provided resource pool to define multiple slices for different vertical customers. A vertical customer with different services can in turn decide to decompose the allocated slice into fine-grained slices, one for each service.

- **Scalability**. This property refers to the specification of a system architecture able to cope with the operational needs of slicing, in terms of quantity and types of slices to be orchestrated. This specification is subjected to the operator criterium for network slice design, which can be coarse-grained (i.e., one network slice for each 5G service category) or fine grained (i.e., one network slice for use case), or any other variant in between. The first option allows working with a small number of network slice types, which brings significant advantages in terms of reusability (multiple slices can be deployed from the same slice type) and operation (the less the number of slice types, the easier their maintenance, and the lower scalability burdens). On the other hand, the second option allows more customization, at the cost of preventing reusability and making operator's system much more unstable; indeed, making much fine-grained offering of slices can provoke an unmanageable number of artifacts to be orchestrated.
- **Automation**. With different network slices running on the network, along with time-varying network conditions, it is impossible for human operators to manage the lifecycle of individual slices in a timely manner. Apart from being error-prone, human intervention cannot keep up with the timescales that slicing may bring. In this regard, it is needed to define self-management capabilities (e.g., self-configuration, self-monitoring, self-optimization, self-healing) for network slicing, minimizing the number of human "touches" on the network.

3 Network slicing: public or private networks?

For commercialization of 4G mobile services, operators have traditionally relied on two different business models:

- Business-to-Consumer (B2C), based on delivering voice and data services to carrier subscribers. These services are provided using of the public network infrastructure, also referred to Public Land Mobile Network (PLMN).
- Business-to-Business (B2B), based on delivering tailored connectivity and digital services to enterprises customers, including private companies and government institutions. The security requirements and purpose-specific functionalities of these services prevents their delivery using the public network; dedicated private LTE solutions completely separated from the PLMN are used instead.

According to the above rationale, public and private networks are seen as two competing, disjoint solutions, following 'either one or another' approach. However, this rationale is no longer valid in 5G service ecosystem. In the B2C market, some of user-centric services that 5G brings will require performance levels that cannot be met with best-effort capabilities offered by carrier PLMN. In the B2B market, delivering tailored services with the sole use of private network infrastructures is quite expensive; this constitutes a major entry for the new verticals, most of them unable to afford these costs.

In 5G, we need to find a solution which allows offering i) more performant B2C services, with costs similar to the carrier PLMN; and ii) B2B services with capabilities similar to those provided by using dedicated private infrastructures, but a much more reduced cost. The answer to these conditions is **network slicing**. As seen in Figure I-4, network slicing allows bridging the existing gap between private and public networks, by offering a solution that provides the benefits of both worlds. On the one hand, network slicing features traffic isolation capabilities, high degree of customization, and performance and security levels which are quite close to those of private networks. On the other hand, it provides the convenience and the coverage of a carrier PLMN, with all the benefits in terms of service continuity, mobility support, and reliable access to data and internet services. But unique to slicing is the flexibility and agility, with the ability to dynamically create, modify and tear down slices, and allocate/de-allocate resources at a much more reduced time scale.

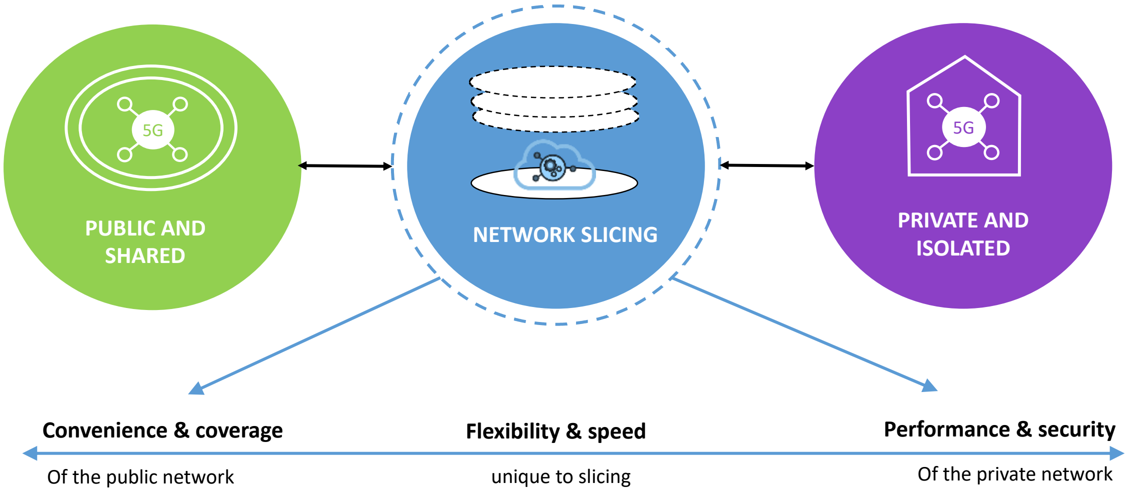
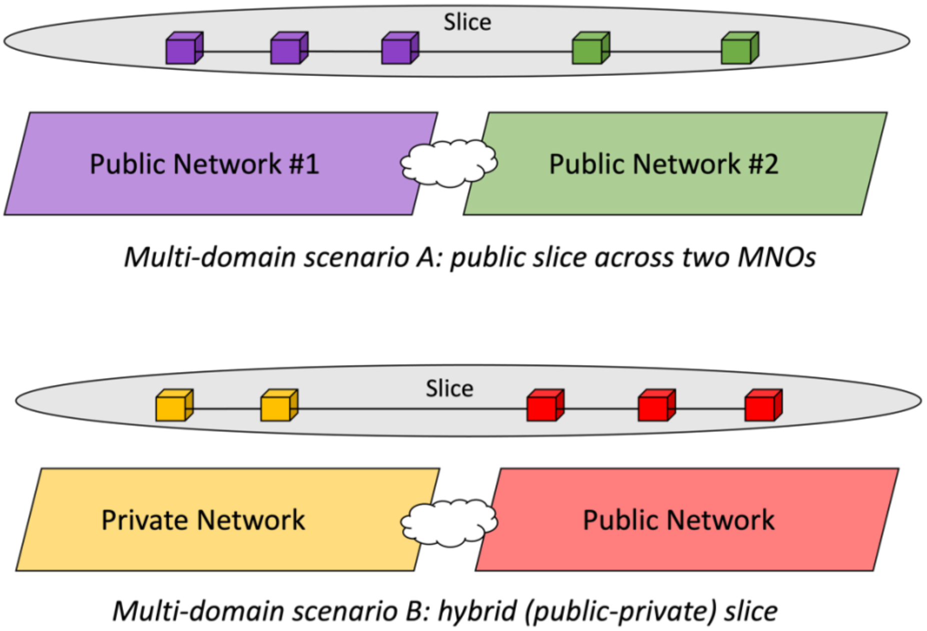


Figure I-4. On the relationship of network slicing with public and private networks. Figure adapted from [10].

For the commercialization of network slices, the operator may face three business models: the evolution of traditional ones, namely B2C and B2B, plus a new one: B2B2X. The description of these models is given below.

- **Business-to-Consumer (B2C) slicing**. In this model, the operator provisions network slice slices for top-tier mass-market services, such as cloud gaming or video streaming. These slices allow providing differentiated traffic treatment (e.g., higher throughput, bounded latency) for end-users who subscribe to them. An example could be a user who buys access to low latency mobile gaming slice on their smartphone. Analyst reports claim that the number of B2C slices in 5G is expected to be rather low. The reason is that there are not many mass-market services that end-users would like to pay extra for higher quality; therefore, it is difficult for the operator to monetize slice usage. However, this situation may change in the medium-long future, with the arrival of metaverse and user-centric applications around it.
- **Business-to-Business (B2B) slicing**. In this model, the operator sells a network slice to an enterprise customer, which uses it for internal operation. Apart from traditional companies and government institutions, in 5G these customers also include the verticals. For example, a manufacturer can request the operator to provision a dedicated slice to accommodate industry 4.0 applications (e.g., robotics, logistics, augmented reality) in a factory. B2B slices are typically fine-grained, particularized for the specific needs of the enterprise customer. Unlike B2C category, tens of B2B slices are expected to be deployed, one or more for ach enterprise customer.
- **Business-to-Business-to-X (B2B2X) slicing**. It is similar to the B2B flavor, with the exception that the slice buyer does not use the slice for internal operation, but to provide services to 3rd parties. Here, the hierarchical abstraction and recursion principles outlined in Section 2 apply. Examples of slice buyers conformant to B2B2C models are MVNOs and large-scale content service providers. Examples of

slice buyers conformant to B2B2B models can be global digital and cloud service providers, typically referred to as hyperscalers (e.g., Amazon Web Services, Microsoft Azure, Google Cloud), which act as main channels for small and medium-sized enterprises.



Multi-domain scenario A: public slice across two MNOs

Multi-domain scenario B: hybrid (public-private) slice

Figure I-5. Multi-domain network slicing. The scenario A corresponds to a B2C slice, while the scenario B represents a B2B slice.

In terms of infrastructure used, there exists differences between B2C slices and B2B/B2B2X slices. The first are entirely deployed using PLMN resources, while the second group makes (full or partial) use of private infrastructure. This infrastructure is typically located at customer premises, e.g., a factory (industry 4.0 vertical), a stadium (if media vertical) or a smart port (logistics vertical).

From the assumption made above, one can note that B2C slices can be profiled in the left half of the spectrum pictured in Figure, while B2B/B2B2C slices can be profiled in the right half of the spectrum.

Finally, it is worth noting that for all the models, it is not uncommon for a slice to cross multiple administrative domains.

For example, for the B2C model, slices host mass-market services which need to be available for subscribers in different geographical locations, even across countries. It might happen that the operator acting as network slice provider did not have coverage for certain locations, and therefore would need to rely on the public network of another operator partner. This situation is shown in the top-side picture of Figure I-5.

In the case of B2B and B2B2X models, this occurs when the slice is not entirely deployed in a private infrastructure. In this situation, the slice span across private and public network infrastructures, with certain resources and functions executing on-premises (customer-managed infrastructure) and the rest in the PLMN. The more resources and functions executing on-premises, the closer the slice to the right end of the spectrum, but higher the cost.

4 Network slicing pillars

This section provides a brief overview of the main enablers for operators to realize network slicing.

4.1 Network softwarization technologies

Network softwarization reflects the trend in which the networks are gradually being architected into systems that separate the software implementing network functions, protocols and services from the hardware running them. This transition is changing the way communication infrastructures are designed and operated, enabling dynamic provisioning and flexible configuration of services atop. Though several technologies fit in the network softwarization realm, two stand out: Software Defined Networking (SDN), scoping networking domain; and Network Functions Virtualization (NFV), with focus on the compute domain.

4.1.1 Software Defined Networking (SDN)

SDN departed from the idea of control user plane separation in network elements, based on moving their control capabilities to a centralized software controller. With this approach, network elements no longer need to make decisions on traffic processing (e.g., switching, routing, firewalling, filtering); instead, it is the controller which takes these decisions on behalf of them. This allows network elements to become inexpensive forwarding devices, whose only responsibility is to process incoming packets (user plane functionality) based on the configuration settings and instructions sent by the SDN controller (control plane functionality).

The radical proposition outlined above was the base of an industrial and research movement around the ideas of network programmability as mechanism for flexible and dynamic configurations of networks, as compared with traditional ways of configuration, either based on manual procedures or leveraging on per-vendor Network Management Systems (NMS). Figure I-6 shows a comparison on legacy vs SDN enabled networks.

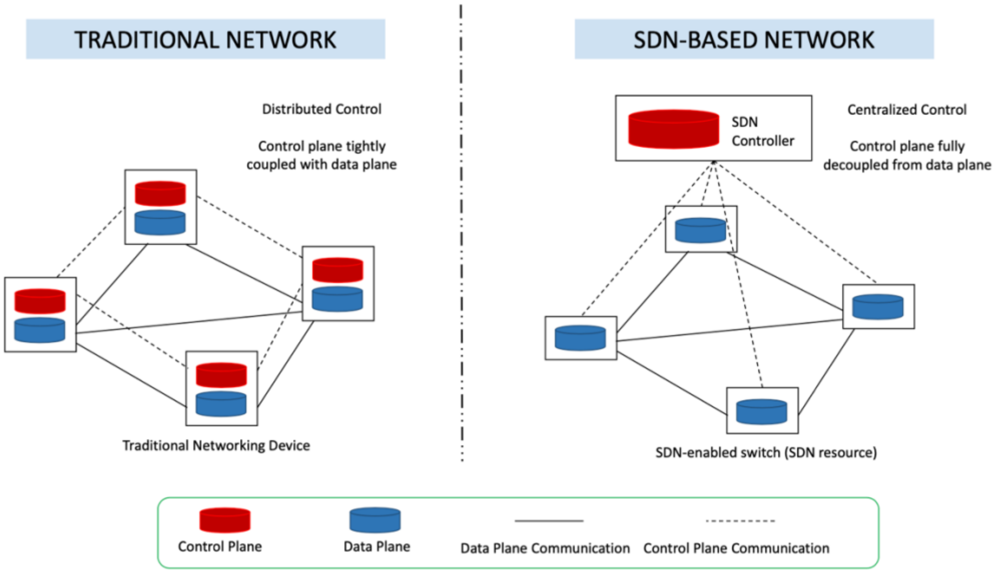

TRADITIONAL NETWORK
Distributed Control
Control plane tightly coupled with data plane
Traditional Networking Device
SDN-BASED NETWORK
SDN Controller
Centralized Control
Control plane fully decoupled from data plane
SDN-enabled switch (SDN resource)
Control Plane
Data Plane
Data Plane Communication
Control Plane Communication

Figure I-6. The impact of SDN.

SDN architectural principles have been promoted by industrial fora, being the Open

Networking Foundation (ONF) the first one producing prominent architectural designs, as in [11] and [12]. Figure I-7 pictures the original SDN architecture. As seen, it is articulated into three layers: i) the *infrastructure layer*, built out of a number of forwarding devices provisioned with pure user plane capabilities; ii) the *control layer*, which hosts the SDN controller and built-in network services, including topology management, path computation element engine, network virtualization protocol suite, among others; and iii) the *application layer*, grouping all the operator-internal assets (e.g., management and monitoring functions) and 3rd party add-ons (e.g., big data applications) that consume controller's network services and associated capabilities.

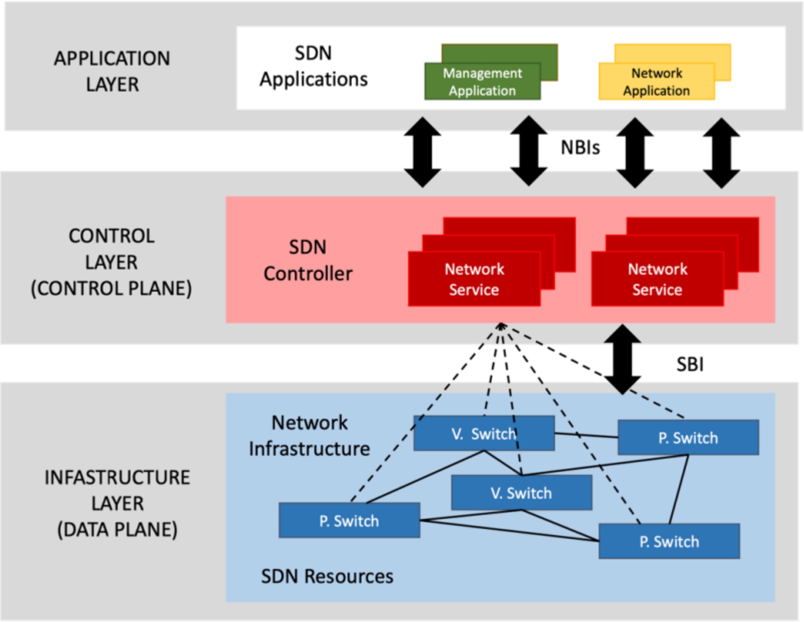


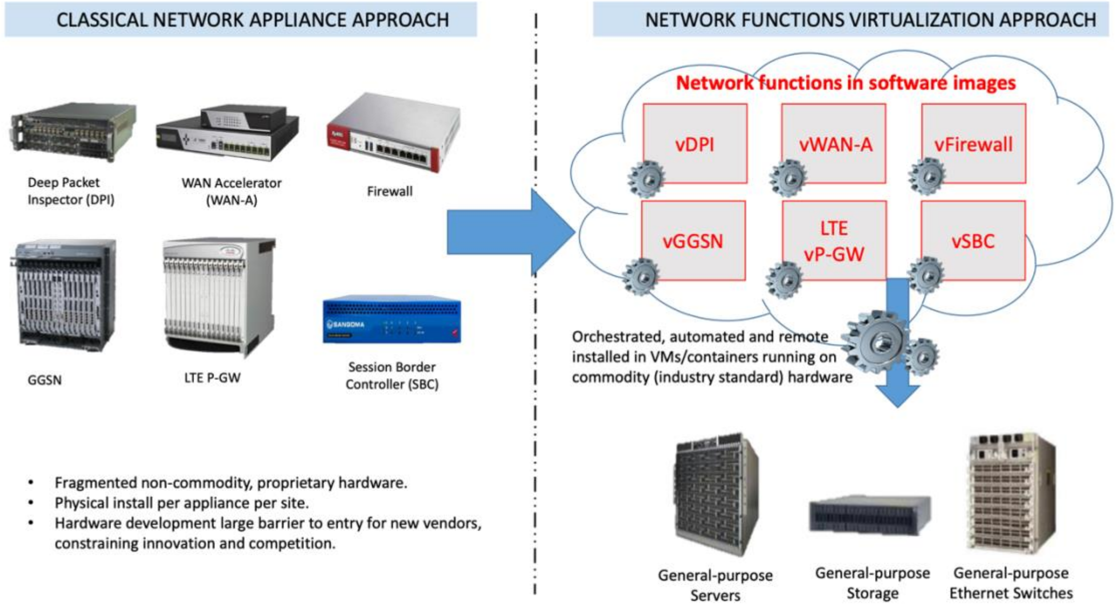
Figure I-7. SDN architecture

When looking at Figure I-7, one can notice that the SDN controller is the central piece of the entire architecture, interacting with the infrastructure and application layers using specific protocols and models at its South Bound Interface (SBI) and North Bound Interface (NBI), respectively. First implementations were based on using OpenFlow as open and standard protocol for the SBI, and REST-based solutions (e.g., RESTCONF, RESTful) for the NBI.

Though this reference ONF architecture was key to foster the evolution that telecom industry is facing now, particularly in what respects to the control user plane separation existing in 5G Core (5GC) [13], it is true that SDN-based solutions have been evolving along the time with different propositions nowadays, which are less radical than the original SDN approach. For example, the OpenFlow is no longer used as de-facto protocol for the SBI, at least in production networks, due to the scalability limitations it shows; instead, the SBI leverages on NETCONF/YANG. Additionally, not all control plane capabilities are moved out of network elements and migrated to the controller; instead, the trend is towards a programmatic interaction with the network rather than a full and complete programmability of the behavior of forwarding devices. Finally, it is worth noting that SDN solutions for data center networks (IT networks) and transport networks (WAN networks) are not equivalent nor comparable, due to the different network virtualization technologies in scope of both environments.

4.1.2 Network Functions Virtualization (NFV)

NFV departed from the idea of separating the software implementation of specific network functions from the dedicated nodes were traditionally those functions were running, in a monolithic and tightly integrated mode. This followed the successful approach experienced in the Information Technologies (IT) industry with the cloud computing model [14].



CLASSICAL NETWORK APPLIANCE APPROACH
Deep Packet Inspector (DPI)
WAN Accelerator (WAN-A)
Firewall
GGSN
LTE P-GW
Session Border Controller (SBC)
Fragmented non-commodity, proprietary hardware.
Physical install per appliance per site.
Hardware development large barrier to entry for new vendors, constraining innovation and competition.
NETWORK FUNCTIONS VIRTUALIZATION APPROACH
Network functions in software images
vDPI
vWAN-A
vFirewall
vGGSN
LTE vP-GW
vSBC
Orchestrated, automated and remote installed in VMs/containers running on commodity (industry standard) hardware
General-purpose Servers
General-purpose Storage
General-purpose Ethernet Switches

Figure I-8. The conceptualization of NFV technology.

The main goal of NFV is to decouple software from hardware in network function nodes, moving network function logic to a software image that can run on top of commercial-off-the-shelf (COTS) servers. This idea is captured in Figure I-8. In NFV jargon, this software image is referred to as a Virtualized Network Function (VNF). VNFs are deployed as virtualized deployment units (VDUs) over an execution environment, which is totally transparent to the actual hardware below. This execution environment could correspond to a hypervisor (Infrastructure-as-a-Service [IaaS] environment) or a container engine (Container-as-a-Service [CaaS] environment). In the first case, VNFs are deployed as Virtual Machines, while in the second case are implemented as containers.

According to the above rationale, one can notice that NFV allows moving away from the dependence of (a large variety of) purpose-built, vendor-specific nodes to the use of VMs/containers running on (a smaller number of) general-purpose, commodity hardware. This not only represents a gain in terms of costs and scalability, but also in terms of service provisioning; actually, NFV provides the operators with the ability to deploy VNFs where and when needed, combining them to form end-to-end network services. For further details on NFV benefits, see [15].

The concept and collaborative work on NFV were born in October 2012, when some of the world's leading telecom network operators jointly authored the first white paper [16]. To define a set of specifications that would facilitate the industrialization of NFV solutions, seven of these leading telecom operators (AT&T, BT, Deutsche Telekom, Orange, Telecom Italia, Telefonica, and Verizon) formed an Industry Specification Group (ISG) with open membership and selected the ETSI to be home of this ISG [17]. Figure I-9 pictures the reference architectural framework that ETSI ISG NFV defines. The core part is the Management and Orchestration (MANO) stack, responsible for the lifecycle management of individual VNFs (VNFM scope) and their composition into network services (NFVO scope). Depending on their capacity needs, these VNFs and network services may be provisioned with right-sized amount of IaaS/PaaS resources from the NFV Infrastructure (VIM scope). The allocation and release of NFVI resources to hosted VNFs and network services is performed dynamically, according to traffic fluctuations and in response to NFVO/VNFM triggered lifecycle management operations (e.g., scaling in/out).

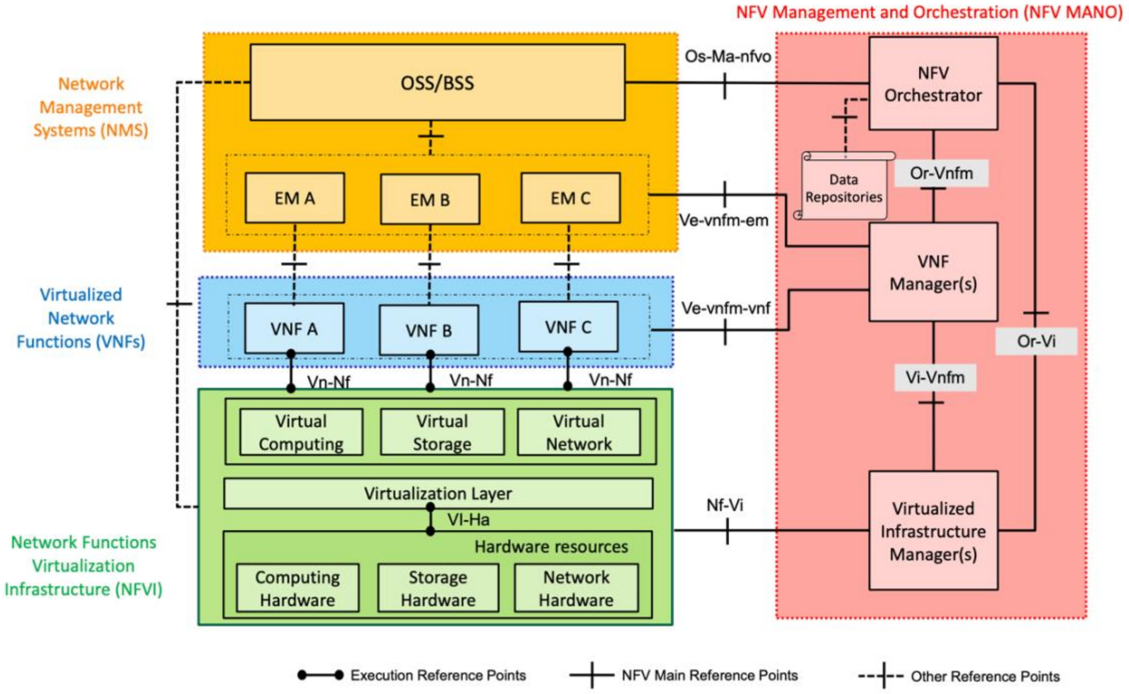


Figure I-9. ETSI NFV architectural framework. Source: [18].

Though defined more than one decade ago, the principles and architecture of ETSI ISG NFV still remain. The technology has matured since then, and the first commercial solutions are already available in production networks. Most of these solutions are based on hypervisors, which allow for the execution of VNFs as VMs. However, this is going to change. Telco industry is now pushing the need to start replacing VM-based VNFs to container-based VNFs, which are much more aligned with cloud-native principles that prevail in 5G, starting with 3GPP 5G Core (5GC) [13]. This migration from IaaS to CaaS is progressive, though it is expected to be concluded in the short term, especially considering the traction that Kubernetes suite (de-facto container orchestration solution) is gaining.

4.2 Edge computing

Cloud computing has historically been a major focus for many applications, wherein data is transferred to off-site servers for processing and analysis. These servers are typically located in centralized data centers, far from the source where this data is generated. With the arrival of 5G and use cases, hosting all workloads in such remote nodes is no longer feasible, especially in what relates to uRLLC and mIoT services. For example, in the case of a uRLLC service, terminating traffic far from data sources/consumers means exceeding latency budget. In the case of mIoT service, having the analytics server into a remote cloud node means all the traffic aggregated from millions of sensors need to be traverse the entire network, which pushes backbone capacity to its limits. To cope with these scenarios, the concept of edge computing was defined.

Edge computing represents a paradigm whereby data storage and processing takes place in a location as close as possible to the user, device or service that will consume the data. In a nutshell, it can be seen as the result of having a cloud continuum along the entire infrastructure, from end device up to the internet. Figure I-10 illustrates this continuum, where the concept of "telco edge" is present at different locations, typically within the boundary of operator networks (far edge, near edge), but also beyond it (on-premises edge). In the latter case, edge computing is placed at customer premises, for example within factory buildings or inside transportation hubs. The reasons for having on-premises workloads typically respond to security concerns and/or policy regulations (e.g., data residency). And for some very specific use cases, the reason can also be performance, e.g., achieve ms-level latency for industry 4.0 services such as "collaborative robots" and "zero-defect detection in smart manufacturing".

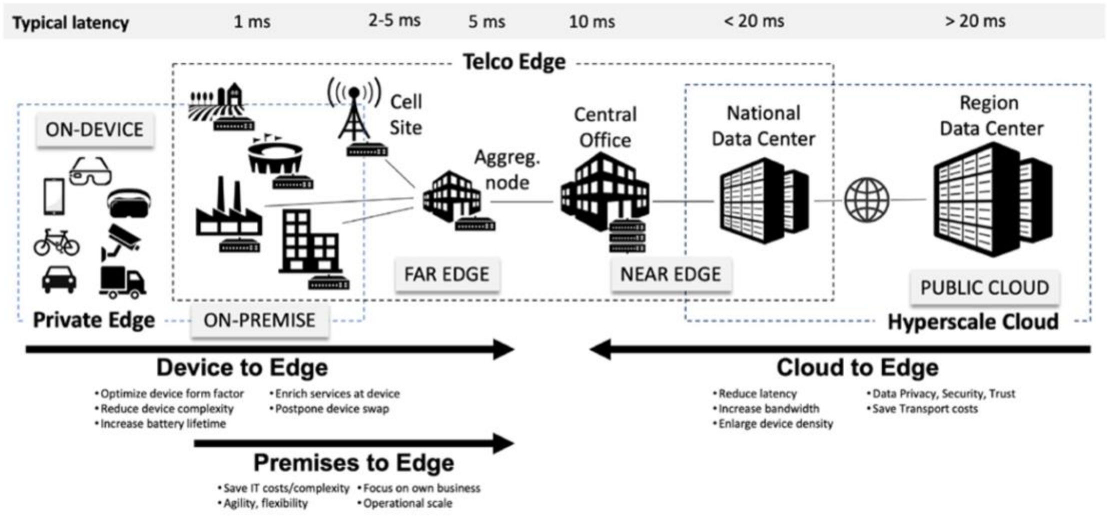


Figure I-10. Edge computing concept. Source: [23]

In relation to slicing, it is worth noting that the use of on-premises edge computing only applies to B2B/B2B2C slices, but not to B2C slices. An example can be the scenario pictured in Figure I-5, with the private network domain hosting on-premises edge computing, and the public network the far/near edge.

There exists extensive literature on the concept and usage of edge computing. For further information, see [19]-[22].

4.3 Management and Orchestration

Management and orchestration refer to the set of tools enabling the provisioning and operation of services in softwarized networks, including virtualized network services and network slices. The objective is to control these services and configure their behavior throughout their lifecycle. For the case of network slicing, the lifecycle is split into four phases [24], as pictured in Figure I-11:

- **Preparation phase**. In this phase, the network operator sets up the network environment, performing all the activities that are needed before provisioning the slice. This includes capacity planning tasks, design of models for network slice artifacts, and their onboarding into corresponding catalogs.
- **Commissioning phase**, consisting in the provision of a network slice instance. In this phase, different slice components are deployed with the necessary capacity

where needed, allocating and configuring infrastructure resources according to the slice specificities.

- **Operation phase**. This phase includes all the activities that are in scope while the network slice instance is up and running. These activities include: *activation*, i.e., make the network slice instance ready to process upstream/downstream traffic; *reporting*, i.e. capture all the metrics made available by the slice instance at run-time, including performance measurements and fault alarms; *supervision*, i.e. compare collected metrics against slice targeted behavior; *modification*, i.e. change slice capacity or topology, either partially or totally, as a response to SLA deviation as reported by the supervision stage; and *deactivation*, i.e., force slice instance to stop processing traffic.
- **Decommissioning process.** It consists in removing the network slice instance, releasing resources and configuring them back accordingly.

Apart from managing the lifecycle of individual network slice instances, management and orchestration is responsible to solve dependencies and conflicts that may exist across different slices.

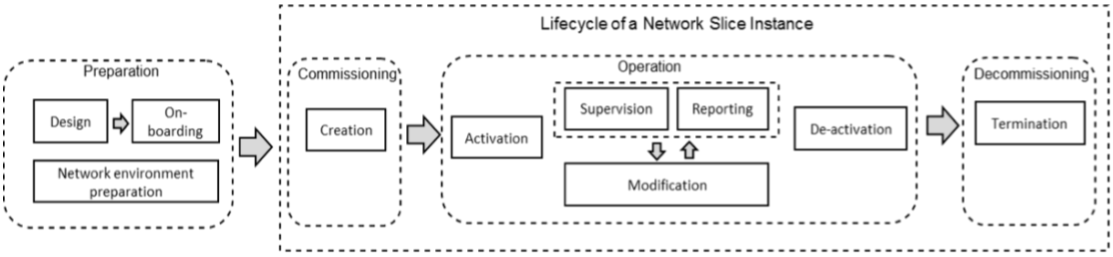


Figure I-11. Network slice lifecycle management. Source: [24]

4.4 Putting it all together

The concepts described here provide means for operators to deliver network slicing capabilities. Let's glue them together. On the one hand, **NFV** (Section 4.1.2) allows deploying the functions of every slice with necessary capacity where and when required. On the other hand, **SDN** (Section 4.1.1) allows operators to programmatically steer traffic within the slice, across the deployed functions. **Edge computing** (Section 4.2) provides the distributed computing substrate to host slice functions, allocating them closer or further from customers as needed, with the possibility of moving them along the substrate. Finally, the **management and orchestration** (Section 4.3) represents the overarching framework that hosts the intelligence and automation tools to instantiate, modify and remove slice instances, and make them coexist when running in parallel atop, guaranteeing their isolation and SLA-compliant behavior.

5 Network slicing landscape

This section provides an overview of the main organizations that contribute to shaping network slicing. They include vertical industry organizations (Section 5.1), telco industry organizations (Section 5.2), standard bodies (Section 5.3) and integration & open-source communities (Section 5.4). Figure I-12 illustrates their scope of work, and the relationship between them.

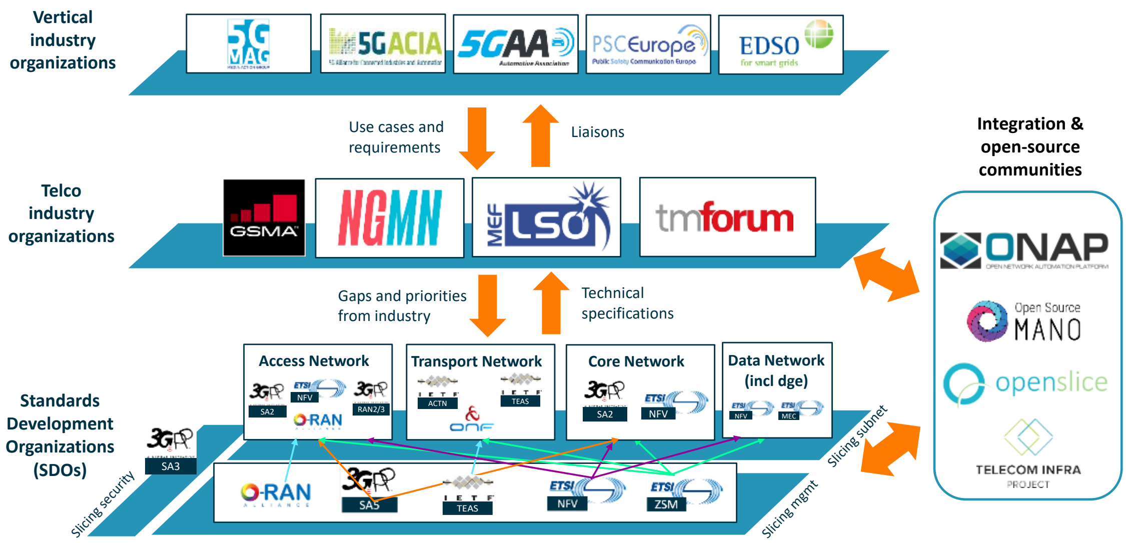


Figure I-12. Network slicing landscape.

5.1 Vertical Industry Organizations

Network slicing aims to support various vertical industries, hence the importance of having industry consortia representing the main drivers for 5G adoption in different vertical sectors. Examples of these consortia are: **5G Media Action Group (5G-MAG),** addressing media & entertainment; **5G Automotive Association (5GAA**), which covers automotive industry; **Public Safety Communication Europe (PSCE)**, representing public safety verticals; **European Distribution System Operators (EDSO**), scoping smart grid; and the **5G Alliance for Connected Industries and Automation (5G-ACIA)**, covering industry 4.0 market. The mission of these organizations is to ensure that the top-tier requirements of vertical customers are captured and prioritized in the ongoing 5G standardization and regulatory requirements, and that new developments in 5G are effectively communicated to and understood by these customers.

Some of the vertical industry organizations pictured in Figure I-12 have already created workstreams specific to slicing, which shows the usefulness of this solution for them.

5.2 Telco Industry Organizations

The mission of telco industry organizations is the same as vertical industry organizations, but now from the operators' point of view. They use the input captured from vertical industries to understand the needs of vertical customers, prioritizing lines of technical work accordingly, with the ultimate goal of influencing on the roadmap of solutions developed in the standard bodies. In relation to network slicing, activities have been started in these four organizations: GSM Alliance (GSMA), Next Generation Mobile Network (NGMN) Alliance, Metro Ethernet Forum (MEF) and Tele Management Forum (TM Forum).

The **GSMA** is a telecommunications industry association which represents the interest of mobile industry worldwide, uniting nearly 800 operators with almost 300 companies in the broader mobile ecosystem. GSMA has published a wide variety of white papers on network slicing, and leads the work of defining templates for slice description. Further details are shared in Part II and Part III.

The **NGMN Alliance** actively drives global alignment and convergence of technology standards and industry initiatives, with the objective to avoid fragmentation and to ensure 5G adoption. This telco industry organization releases white papers touching on hot topics, or in the eve of new mobile generations; indeed, NGMN Alliance was the first organization which coined the network slicing concept. For further information, see Part I.

The **MEF** has specified the Lifecycle Service Orchestration (LSO) framework [32], a high-level reference architecture to illustrate the relationships between service providers, customers and 3rd parties regarding service orchestration. The LSO has inspired different standard bodies to define technical solutions for orchestration in softwarized networks, and has been used to carry out different PoCs in network slicing.

Finally, there is **TM Forum**. It provides an open, collaborative environment and practical support which enables operators to rapidly transform and digitalize their business operations and IT systems, to capitalize on the opportunities presented by network softwatization. TM Forum's ZOOM project [33] started a workstream to analyse network slicing. Several user stories have been generated, and respective requirements have been derived and mapped to TM Forum assets, such as the Open Digital Architecture (ODA) [34].

5.3 Standard Development Organizations

The Standard Development Organizations (SDOs) are responsible for developing technical solutions which satisfy the operational and business requirements of telco and vertical industries. As pictured in Figure I-12, there are a number of SDOs that touch on network slicing, from different perspectives.

The 3rd Generation Partnership Project (3GPP) produces technical specifications and reports for mobile networks from 3G onwards. It is the standard leading the specification of 5G technology, from Release 15. 3GPP is also considered as the forefront ambassador for network slicing, with solutions developed in different working groups, including **RAN2/3** (5G radio access network), **SA2** (5G core network and service platform), **SA3** (security) and **SA5** (management, orchestration, and charging).

The European Telecommunications Standards Institute (ETSI) produces globally applicable standards for Information and Communication Technology (ICT) systems, including fixed, mobile, radio, converged and broadcast services. ETSI work on network slicing is spanned across different Industry Specification Group (ISGs), including Network Functions Virtualization (NFV), Multi-access Edge Computing (MEC) and Zero-touch network and Service Management (ZSM). **ETSI NFV** focuses on network slice virtualized resource management, leveraging use of NFV technology. **ETSI MEC** develops specifications for the use of edge computing in fixed-mobile converged environments. Finally, **ETSI ZSM** defines E2E network management and orchestration solutions for telco services, including network slicing, with focus on automation.

The Internet Engineering Task Force (IETF) is the standard for internet services. It covers the network domains which are out of scope of 3GPP, namely transport network and data network. Although late, it also started to work on network slicing, first on the **Abstraction and Control of Transport Network (ATCN)** working group, with focus on architecture, and later on the **Traffic Engineering Architecture and Signaling (TEAS)**, working on solutions in close collaboration with 3GPP. Further information is captured in Part IV.

Apart from these main three SDOs, other industry groups are worth mentioning, including ONF and O-RAN Alliance (O-RAN). **ONF** promoted the use of SDN technology (see Section 4.1.1) for a long time, before becoming an association in charge of developing de-facto prototype solutions in different technologies, including optical networks and private 5G. On the other hand, **O-RAN** was founded in 2018 by tier-1 operators to re-shape RAN industry towards more intelligent, virtualized and fully interoperable solutions, shifting from black-box to white-box approaches. This paradigm shift has made O-RAN become the reference standard for RAN advanced features, including wireless network slicing.

5.4 Open source and integration communities

These communities demonstrate and assess solutions conceptualized in telco industry organizations and developed in SDOs, with the mission to feed them back with lessons learnt. The validation environments typically build upon prototypes using open-source code. Examples include **Open Source MANO (OSM)**, an ETSI hosted project providing a reference NFV implementation for network and service orchestration; **Open Network Automation Platform (ONAP)**, similar to OSM but under the umbrella of Linux Foundation; and **Openslice**, a future-proof Operation Support System with built-in slicing management capabilities. Further details of these three solutions are given in Part III. In addition, other communities such as **Telecom Infra Project (TIP)** are relevant, though this one is more focused on private 5G networks.

5.5 Timing

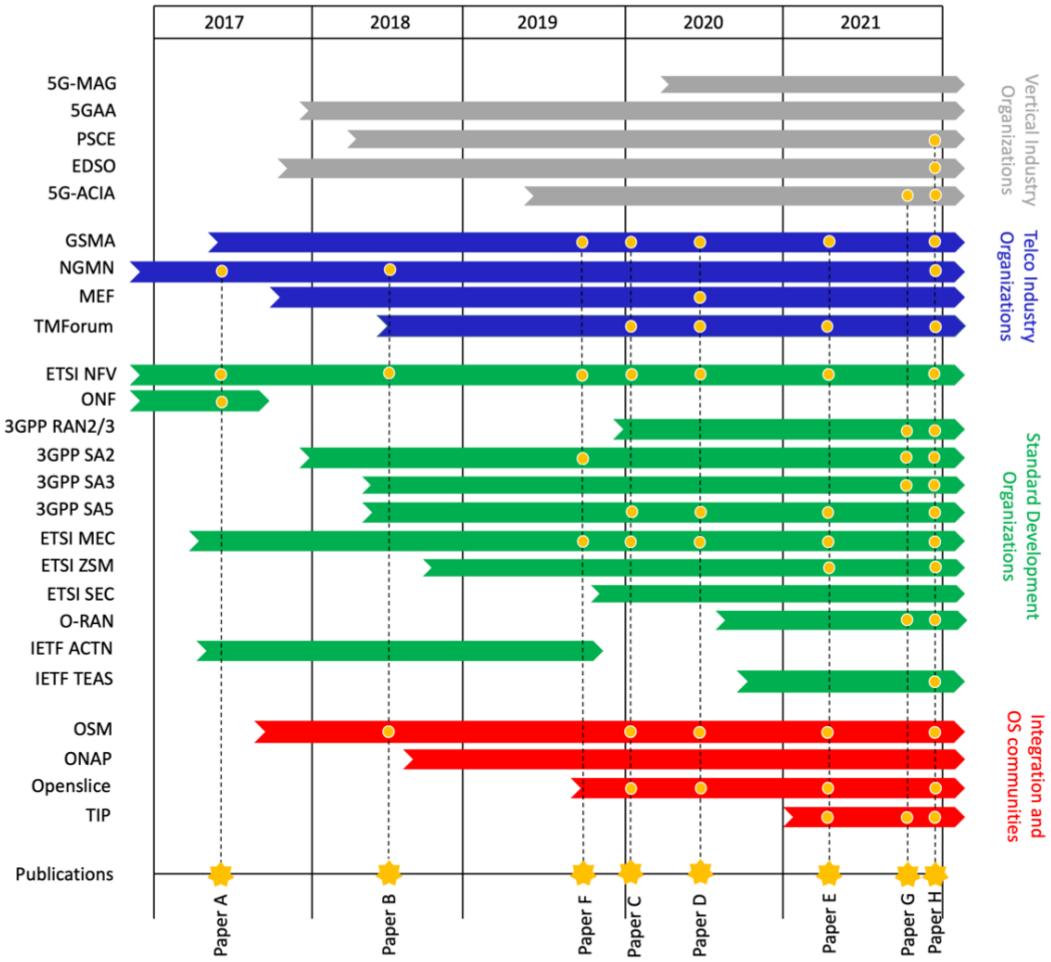


Figure I-13. Impact of standardization activities in the in-thesis publications.

The landscape described above and shown in Figure I-12 demonstrates the hyper-fragmentation that industry experiences around network slicing, with many initiatives working in different parts of the E2E problem, with different scopes that do not necessarily match with each other. This makes it difficult to have consistent solutions, across horizontal and vertical domains, resulting in many gaps that need to be addressed. However, the lack of coordination is not the problem, but also the timing; in fact, not all the organization started to discuss slicing at the same time. Figure I-13 illustrates this problem, explaining the evolution in the work captured in this PhD thesis, and why the inputs of specific SDOs were in some papers and not in others.

6 Main challenges on network slicing

This section lists the challenges associated to network slicing in 5G mobile systems.

6.1 Technical challenges

These challenges represent open issues that industry need to face when implementing system architectures, protocols and algorithms for network slicing, in such a manner that the design principles and capabilities listed in Section 2 are met. References [25]-[29] capture a large part of the technical challenges which were originally identified for network slicing. However, not all of them fall within the scope of the present thesis.

This section identifies and elaborates on the most noticeable technical challenges which are relevant for the work of this thesis.

Table I-1. In-scope network slicing technical challenges. These challenges result from the design principles and capabalities captured in Section 2.

	Design principles and capabilities							
Technical challenges	Isolation	Customization	Elasticity	Programmability	End-to-end (E2E)	Hierarchical abstraction	Scalability	Automation
Standards-compliant system architecture			X		X	X	X	
Resource segregation	X	X				X		
Resource allocation	X	X			X			
Translation of service requirements into infrastructure requirements	X				X			
Security	X			X				
Federation	X				X			

Capability exposure to slice customers	X	X		X		X		
Zero-touch commissioning								X
Zero-touch operation			X				X	X

- **Standards-compliant system architecture.** The challenge that lies ahead is how to design a system architecture that makes a combined use of the technology pillars, considering the different scope of these pillars and their standardization by different standard bodies. How to make all the pieces work together becomes more difficult as new functionalities appear in the standardization landscape (see Section 5) and multi-vendor solutions come into scene.
- **Resource segregation**. This allows splitting the infrastructure into a set of partitioned resources, referred to as resource chunks. This partitioning builds on the fact that multiple network slices running on top of the common infrastructure, hence the need to define different resource chunks for them. Resource segregation needs to be applied across all type of infrastructure resources, including wireless, compute and connectivity resources. Regarding *wireless resources*, the resource chunks are virtualized resource blocks. These result from the abstraction of physical resource blocks, according to N:1 mapping rules. Regarding *compute resources,* solutions can range from the dedication of compute nodes like the bare-metal approaches (full isolation) to the sharing of computing capabilities by means of hypervisors (medium isolation) or containers (low isolation) [14]. Finally, for connectivity resources, options span from the segregation per lambdas (hard isolation) to the definition of traffic-engineered shared pipes (soft isolation). As seen for all the cases, a lower level of partitioning means a lower number of resource chunks, but also a higher degree of isolation among them. At the end of the day, it means renouncing multiplexing gains at the cost of having more isolated slices. The challenge here is how to design solutions that carefully takes multiplexing gains and isolation into account, and that keep both dimensions as much balanced as possible. Multiple approaches can be used for these solutions, and the number of these approaches increase as new technological features appear.
- **Resource allocation**. This is the step that follows resource segregation. It consists in delivering the resource chunks to the different slices, according to their service needs. The challenge here lies on two fundamental questions: How to ensure consistency across all the network domains? What is the best time scale for making resource allocation decisions?
- **Translation of service requirements into infrastructure requirements**. To accommodate a service into a 5G sliced network, there is a need to translate customer expectations (e.g., guaranteed/peak UE throughput, number of UEs, availability, maximum packet delay budget, etc.) into operator internal decisions at the underlying infrastructure, related to enabling network features and resource capacity. Traditionally this has been resolved with pre-defined mapping policies, based on the assumption that i) the number of services with different performance profiles is relatively low, and ii) the infrastructure is built of network functions from a very limited number of vendors. These tenets are however no longer valid in 5G, which is now open to vertical services and flees from mono-vendor culture thanks to network softwarization. In this situation there are multiple open questions ahead: how to

formalize service requirements, in a format that both operator (telco experienced) and vertical (no telco background) can understand? How to map service requirements into network function level requirements, e.g., network function sharing, and activation/de-activation of certain features? How to map network function level requirements into resource segregation and allocation solutions? What are the operator criteria to decide if the service can be accommodate in an existing slice, or a new slice needs to be defined? How to design solutions such that the operator can make all these mappings and decisions in a few minutes, so feedback can be provided to the customer in a timely manner? In case the network cannot accommodate requested service requirements, should the operator need to suggest to the customer alternative requirements as part of the feedback?

- **Security.** The arrival of softwarization technologies in telco industry allows networks to be more flexible, elastic, and customizable. However, if not deployed carefully, these technologies may harm network security, increasing the attack surface. Actually, the resource abstraction that virtualization brings may difficult attack detection and defending. This is even worse in network slicing, where the sharing of resources may facilitate attackers cross network slices to misuse the networks for their desired purposes. A summary of the challenges and main directions regarding security in sliced networks can be found in [30].
- **Federation**. As outlined in Figure I-5, there exist situations where the network domains building out a slice span two (or more) administrative domains. To make slice behavior consistent in the E2E path, we need the management systems from the involved administrative domains to communicate with each other. This approach, referred to as federation, ensures coordination across the domains on decisions related to network slice lifecycle management, such as resource segregation and allocation, connectivity configuration, or scaling. Despite industry is a needed capability, there are a number of open issues that still prevent their adoption: How the interfaces, protocols and APIs for the federation should look like? Which body should be responsible for their standardization? Should federation be based on one-to-one interactions between administrative domains, or should a 3rd party act as broker to mediate these interactions? How to ensure trustworthiness when federating administrative domains that are owned by different stakeholders? What are the auditability and access control specificities when federating a private domain with a public domain?
- **Capability exposure to slice customers.** One of the most noticeable features that slicing enables is the ability to provide the customer with the perception of having a dedicated network. This means that the customer does not only get insight into slice metrics, but also can access the actual slice and programmatically configure it himself. This is especially relevant for B2B and B2B2X models. The challenge here lies on how to expose these operational capabilities to the customer, using APIs. In particular, the following open issues are found: which parameters these APIs shall convey, to ensure they are meaningful to the customer, considering he is no telco experienced? Should these APIs be operator-specific, or could we expect some standardization work here? If so, which standard should take the lead? How to apply access control to the customer? How to ensure actions triggered by a customer do not impact other slices? What are the mechanisms that are needed to trace request-response messages between operator and customers, and how these impact on the SLA violation and liability of the slice?
- **Zero-touch slice commissioning.** The process of translating service requirements

into infrastructure requirements, and proceed with resource segregation and allocation accordingly, needs to be automated as much as possible. The reason is the number of variants that exist in these decisions, and the time scales expected for them. However, going for this automation is a long journey which multiple questions that remain unanswered: How to design slice feasibility check algorithms to take these decisions? How the workflows related to catalogue and inventory management are impacted?

- **Zero-touch slice operation.** To make network slice assurance automated, gaps need to be bridged in the following dimensions.
    - Reporting**:** what are the slice Key Performance Indicators (KPIs) which need to be monitored? What are the performance measurements and fault alarms that I need to aggregate to compute those KPIs? What are the nodes that provides such metrics? How can I discover the reporting options of these nodes? How to correlate network function metrics with infrastructure metrics in virtualization environments?
    - Supervision and modification: what are the cross-slice and slice-specific policies that an operator shall configure based on received service requirements? How to ensure they are not incompatible among them? How to program slice scaling algorithms, and which are the input/output parameters that they should operate on? How scaling decisions are propagated towards SDN and NFV system, to enforce resource re-allocation accordingly? How to arbitrate when there are conflicting decisions? How can Artificial Intelligence (AI) assist in this arbitration?

6.2 Non-technical challenges

Apart from the complexities inherent to the actual technology (Section 6.1), there exist other challenges that operators may face prior to starting slicing commercialization. These fundamentally are issues related to monetization and regulation. These non-technical challenges arose when the telco industry started to discuss slicing from a market perspective, something which occurred two years after the PhD thesis kick-off.

In relation to monetization, the following challenges apply:

- **Pricing and sales processes**. Slicing enables customers to have flexibility, security, simplicity, and network performance tailored to specific and demanding use case requirements. To both articulate and capitalize on the value of network slicing, it is needed that the customers perceive the capabilities this solution brings to their business, avoiding for it to become a commodity such as other Internet-like connectivity services. This deep level of customization needs to be crystalized into concrete service and product offerings in the operator’s portfolio, all accompanied with new sales processes and new pricing strategies. The problem is that unlike IaaS services from top-tier cloud providers, offerings based on network slicing cannot be easily categorized in a few tiers or buckets, precisely because of the seamless infinite possibilities for customization. And using traditional commercial channels are not an option for these customers, who look for more ‘pick-and-choose’ approaches.
- **Go-to-market strategy.** Operators need to have a proper go-to-market plan in place for network slicing if they hope to gain experience and capture market share. A well-though-out strategy can make the biggest difference between an operator that succeeds and opens new revenue streams, to one that is just scratching the surface. The GSMA White Paper published in 2018 [31] provides some hints on how this

strategy could be. As seen in Figure I-12, these hints only offer guidelines, but not a magic recipe; at the end of the day, individual operators shall chart their own course by making decisions that, in most cases, are quite dependent on local market conditions. This means that for an operator with a footprint that spans across multiple countries, the go-to-market strategies of their operational business units can differ.

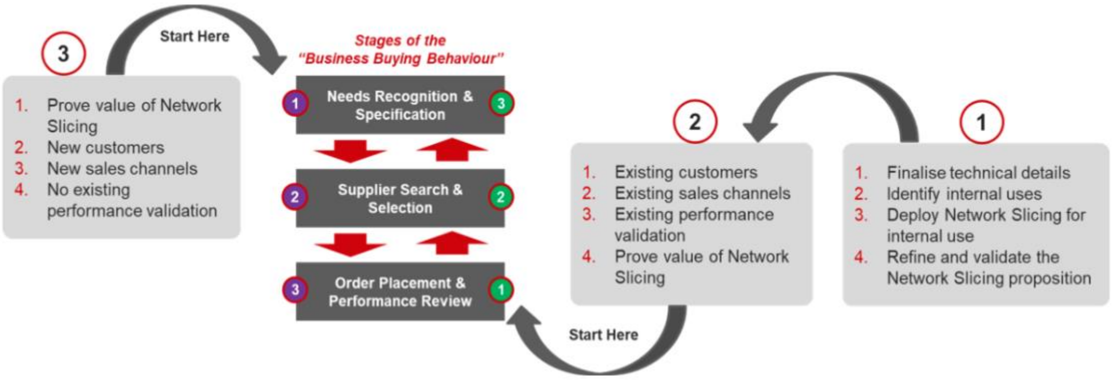

Start Here
3
1. Prove value of Network Slicing
2. New customers
3. New sales channels
4. No existing performance validation
Stages of the "Business Buying Behaviour"
Needs Recognition & Specification
Supplier Search & Selection
Order Placement & Performance Review
2
1. Existing customers
2. Existing sales channels
3. Existing performance validation
4. Prove value of Network Slicing
1
1. Finalise technical details
2. Identify internal uses
3. Deploy Network Slicing for internal use
4. Refine and validate the Network Slicing proposition
Start Here

Figure I-12: Stages of the network slicing go-to-market strategy. Source: [31].

Regulation touches on many aspects, including (though not limited to):

- **Net neutrality**. The differentiated service treatment that slicing promises to deliver can cause suspicion in policy makers. The reason is that isolating data traffic into logical partitions is a traffic discrimination on commercial grounds, which might contravene network neutrality rules. Telco industry position is that policy makers should support and open and non-discriminatory internet, which provides consumers with access to the content and applications they want, while promoting service differentiation; this means that operators shall have the flexibility to create network services that appropriately handle unique requirements. If not done carefully, network neutrality regulation might prevent operators from tapping network slicing and monetizing it accordingly. And this situation would hinder innovation in networks; actually, without these revenue streams, operators may not be able to continue investing in modernizing their infrastructure, risking the availability of a modern network for a smart society.
- **Cross-domain data transfers**. Cross-border transfers of personal data are now regulated by many instruments and laws intended to protect individuals' privacy, the local economy or national security. It is foreseen that network slices may be utilized to offer services outside of the home jurisdiction, potentially in a comparable manner to international roaming. If so, any transfer of data across borders may need to consider regulatory requirements [35], at regional, national, and international levels.
- **Spectrum**. For those cases where slicing is used to provision solutions behaving like private 5G, as commented in Section 3, spectrum issues arise. For example, in a multi-domain slice such as the one shown in Figure I-5, the on-premises domain of the slice will have local coverage (e.g., factory, seaport, etc.). For this coverage, some voices in the industry encourage regulators to define dedicated private 5G bands, while other voices advocate to go for leasing practices (national operators dedicating a portion of their acquired spectrum to provide indoor coverage for industry customers). The decision taken has an impact on the way network slicing shall be provided and monetized by the operator, and therefore on the interaction with the customer (and the ecosystem around it).

7 Scope and Objectives of the Thesis

With network slicing, the operator's network can be logically split into a set of programmable network partitions (i.e., network slices), each designed to satisfy a specific set of service requirements. The service-tailored logical networks resulting from this partitioning can be executed in parallel but need to be operated in isolation from each other. This means that despite running on a common (shared) network infrastructure, network slices require separate (independent) management, keeping up with the timescales of hosted services.

The industry has worked on network slicing during the last five years, but with different scopes and timing, resulting now in a hyper-fragmented landscape (Section 5). The natural consequence is a multitude of open issues and gaps that need to be bridged (Section 6), and that prevent ramping up network slicing in the market. Additionally, when reaching out to B2B market, "network slicing" and "standalone private 5G" seem to be opposing solutions, with industry obsessed with debating which one is better to use. However, the reality is that there are opportunities for them to coexist; actually, telco and vertical industries can benefit from synergies between private 5G and slicing, choosing the best of both worlds (Section 3).

The goal of the present thesis is **to design and validate solutions for network slicing management and orchestration in multi-domain environments, with applicability in public and private 5G network scenarios.** The pursuit of this goal is a long journey that requires going step by step, because of the ecosystem complexity and the challenges therein. In fact, there are multiple technologies and network capabilities in scope, and several aspects to consider when combining them. In this situation, it is important to start with the basics, validating solutions with them, to lay the ground for further progress.

Following the above recommendations, the work of this thesis has been structured into three objectives. The description of these objectives and their relationship with scoped technical challenges are detailed below.

Objective 1: Design of a system for the management and orchestration of *network slices* in multi-domain environments. This objective includes the following sub-objectives:

- **O1.1**: Design of a system architecture. This architecture will build upon the SDN and NFV reference frameworks, combining and extending them so that the resulting system satisfies the design principles and slicing capabilities listed in Section 2.
- **O1.2**: Design of a network slice descriptor, to allow for a model-driven slice deployment and operation. This descriptor will capture the information that the system needs to manage a network slice throughout its lifecycle, from commissioning to de-commissioning, including operation in between.

Timing: 2016Q2 – 2018Q2

The objective 1 aims at defining system level solutions that provide a solid foundation to discuss on the different technical challenges captured in Section 6.1. It is worth noting that the works associated to objective 1 put the focus only on compute and connectivity resources, which are the ones that are in scope of network softwarization technologies. This means that RAN slicing is not addressed in this objective.

Sub-objective O1.1 will define a standards-compliant system architecture, by integrating software SDN/NFV modules and build slicing awareness atop. Hints to address resource segregation, resource allocation, capability exposure and security challenges are captured in the context of this system architecture, exemplified with a slicing use case addressing different administrative domains. For resource **segregation**, it is specified the interactions between the different modules to define the different resource chunks. For **resource allocation**, it is explained how the multi-site orchestration capabilities that the NFV MANO defines are used to deliver resource chunks to different slices. For **security**, references to protection mechanisms are included. For **capability exposure**, it is specified how to provision the slice with a dedicated controller that allow the customer to access slice capabilities. This access is done through the consumption of APIs offered by this controller.

The sub-objective O1.2 will provide a solution design to address all the open questions inherent to the following technical challenge: **translation of service requirements into infrastructure requirements.** The solution also includes slice specific policies on that sets the basis for **zero-touch slice commission,** with mechanisms to perform slice admission control, allocation solution and resource reservation, and zero-touch slice operation, with information on metrics to be collected for reporting and input/output parameters for scaling algorithms. The proposed solution design leverages the model artifacts of NFV, i.e., network service descriptors (NSDs) and VNF Descriptors (VNFDs) [18], and extends them up to the slice level, with the definition of network slice descriptors.

Objective 2: System implementation and solution validation. This objective includes the following sub-objectives:

- **O2.1**: Prototyping a solution for the network slice descriptor
- **O2.2**: Prototyping a solution for the system architecture.
- **O2.3**: Validation of prototyped solutions during the slice preparation phase, with focus on onboarding operation.
- **O2.4:** Validation of prototyped solutions during the slice commissioning phase, with focus on instantiation operation.
- **O2.5**: Validation of prototyped solutions during the slice operation phase, with focus on auto-scaling operation.

Timing: 2018Q3 – 2021Q2

The goal of objective 2 is to validate the hypothesis and assumption which were made in Objective 1. To that end, objective 2 will prototype the designed system level solutions and validate them in relevant use cases.

The sub-objective O2.1 will prototype the network slice descriptor resulting from O1.2.

The sub-objective O2.2 will prototype the SDN/NFV powered architecture designed in O1.1, using OSM and Openslice frameworks (Section 5). The prototyped implementation addresses the **federation** challenge, with a solution that features one-to-one interactions across administrative domains, making use of TM Forum APIs. The solution has built-in auditability and logging capabilities that provide necessary trustworthiness when federating the orchestrators of the domains.

The sub-objective O2.4 demonstrates the automation of workflows from slice ordering to slice deployment and configuration, thus answering the last open question related to **zero-touch slice commissioning** challenge**.** Questions related to feasibility check are however left out of scope of this demonstration.

Finally, the sub-objective O2.5 allows addressing all open questions related to supervision and modification captured in the **zero-touch slice operation challenge**, except for those related to AI usage for conflict resolution. This open issue is out of scope of the demonstration.

Objective 3: Analysis of the private 5G ecosystem, and design of solutions for private 5G networks exploiting network slicing capabilities. This objective includes the following sub-objectives:

- **O3.1**: Study on the specificities of private 5G ecosystem, including actors, use cases, service requirements and in-scope technologies, marking differences with mass-market public 5G networks. Analysis will be focused on industry 4.0.
- **O3.2**: Design of solutions for private 5G networks, integrating available technologies and slicing capabilities to profile different deployment scenarios, ranging from stand alone to hybrid (public-private) networks.
- **O3.3:** Overview of network slicing solutions, profiling them in terms of i) *behavior*, with focus on isolation capabilities; ii) *timing*, capturing their availability in the short, medium and long term; and iii) *applicable network environments*, scoping both public and private networks. This overview is outlined in a technology radar. Apart from assisting operators in their strategy for network slicing roll-out, this radar will help industry understand the role of network slicing for the provisioning and operation of 5G services in public-private environments.

Timing: 2019Q2 – 2021Q4

The goal of objective 3 is the study of private 5G market, the analysis of network slicing role in it, and the design of solutions based on their combined solutions. The objective O3 leverages the outcomes of objectives 1 and 2, and complements them with the inclusion of i) the private 5G network concept, modelled as a new administrative domain; and ii) the RAN domain.

While sub-objectives O3.1 and O3.2 focus most on private 5G network segment, the sub-objective O3.3 is the one that addresses in detail E2E slicing, with solutions that illustrates its usage on public-private 5G infrastructures. The result of O3.3 is a survey of solutions that allows addressing the open questions captured in the following technical challenges:

- **Standards-compliant system architecture**. The original SDN/NFV system architecture for network slicing is further elaborated in O3.3. The updated architecture integrates new software components, according to the progress reflected in a Figure I-13, using open interfaces for them to ensure multi-vendor solutions.
- **Resource segregation.** Different resource partitioning solutions are defined are profiled, with isolation as main reference indicator.
- **Resource allocation.** Different solutions for RAN, TN and CN slicing are specified and compared, in terms of complexity, performance levels and applicable environment (public networks, private networks and public-private networks). In relation to the question "how to ensure consistency across all the network domains?", O3.3 does not showcase one specific solution, though provide technical directions on how to avoid conflicts/decompensation when deciding on resource chunks in the E2E path. In relation to the question "what is the best time scale for taking resource allocation decisions?", there are pros, there are pros and cons of going for shorter or longer time scales. On the one hand, short-term decisions allow operators to handle dynamic traffic fluctuations and variable network conditions, particularly acute in RAN side, thus ensuring Service Level Agreement (SLA) compliance of individual network slices. However, this would require developing algorithms able to integrate real-time data collection and processing engines, to cope with these decisions in such time scales; additionally, the system stability could be compromised. On the other hand, long-term decisions may make the system operation much easier, helping to make better resource planning. However, this is done at the cost of risking SLA fulfilment

during certain time intervals, which is unacceptable for mission-critical services.

- **Translation of service requirements into infrastructure resources,** with answers to all the open questions captured in Section 6.1.
- **Security.** O3.3 complements insights captured in O1.1, with solutions provided by specific functions in the public and private domain.
- **Federation.** O3.3 leverages the federation solution prototyped in objective 2.
- **Capability exposure to slice customers.** O3.3 specifies different capabilities suitable for exposure to the B2B/B2B2C customer, grouping them into different API families. However, it does not offer concrete answers to the following questions: "How to apply access control to the customer? How to ensure actions triggered by a customer do not impact other slices? What are the mechanisms that are needed to trace request-response messages between operator and customers, and how these impact on the SLA violation and liability of the slice?". These questions are left for further study.
- **Zero-touch commissioning**. O3.3 details input/output parameters that are needed for the feasibility check operation, and the model-driven workflows that are triggered for slice allocation and configuration, taking catalogs and inventories into account.
- **Zero-touch operation**. O3.3 provides references that allows answering the open questions posed for reporting. For the supervision and modification, O3.3 leverages the solutions prototyped in objective 2.

8 Research Methodology

In pursuit of these thesis's objectives, the PhD candidate has partially leveraged on

- his participation in different European Commission (EC) funded H2020 projects, including 5G-VINNI, 5GROWTH and 5G-CLARITY.
- his role as Telefónica's delegate in different telco industry organizations, SDOs and open-source communities. These include GSMA NG, 3GPP SA5, ETSI ZSM, Openslice and OSM, with occasional participation at IETF TEAS.

Part of the objectives achieved in this thesis have been supported with these two activities, in particular objectives 2 and 3.

As for the research methodology, the following approach was followed. *First*, literature was examined in detail, in order to find open problems and research questions. This literature did not only include research papers, but also industry papers, collected from the different organizations pictured in Figure I-12. *After* this state-of-the-art review, gaps were identified and prioritized, capturing them into one or more problem statements. The validity and relevance of these problem statements were then double checked with university group, to get early feedback. *Next*, solutions were proposed, socializing them with the partners of corresponding project. *Later*, when the solution validation was required, testbed for simulation/experimentation was setup, and test cases are executed. *Finally*, the main findings were compiled and presented in the form of a scientific paper.

There were findings which were relevant, in terms of time and scope, to the work done by certain industry bodies and open-source communities. For those situations, the PhD candidate also (re-)shaped these findings, turning them into contributions/PoCs that were submitted to these organizations.

9 Impact Creation

To ensure the effective uptake of thesis's results, an impact creation strategy has been carefully outlined. This strategy has been articulated into SMART (Specific, Measurable, Achievable, Relevant and Time bound) outreach measurements scoping dissemination, standardization, and exploitation. These measurements have been carried out continuously when there was an appropriate combination of availability of thesis results and appearance of an opportunity.

On the one hand, **dissemination** refers to the public disclosure of thesis' outcomes through a process of promotion and awareness-raising from the very beginning. It makes research results known to various stakeholder groups in a targeted way, to enable them to use the results in their own work. Table I-2 captures the main dissemination activities of thesis outcomes.

Table I-2. Thesis's results dissemination: outcomes and dissemination channels.

Dissemination activity	Mechanisms
Publications in top-tier scientific journals.	Publication of results in high impact-factor journals and magazines on communications /networking. Submissions can be made against regular series or special issues. *Dissemination channels: IEEE Communications Magazine, IEEE Access, MDPI's Sensors, IEEE Vehicular Technology Magazine, Journal of ICT Standardization (River Publishers)*
Publications and presentations in international conferences	Publication and presentation of high-quality results at reputable international conferences. *Dissemination channels: IEEE relevant conferences (i.e., ICC, CSCN, SDN-NFV), EuCNC and IFIP/IEEE NOMS.*
Contributions at major trade shows and events[1]	Presentations and participation in research-oriented workshops and panels, at both academic and industry events. *Dissemination channels: FOKUS FUSECO FORUM and 5GWeek.*
Participation in EC and 5G PPP clustering mechanisms[1]	White papers and case studies elaborated in conjunction with research projects from 5G-PPP programme. *Dissemination channels: 5G-PPP Architecture WG, 5G-PPP Architecture Technical Board, NetWorld2020*
Production of technical documentation from projects results[1]	Publication of deliverables, technical reports, posters and newsletters attained to project technical innovations. *Dissemination channels: Public websites and social media from 5G-VINNI, 5GROWTH and 5G-CLARITY projects.*
NOTE1: Details are captured in Annex A.	

On the other hand, the **standardization** framework in scope encompasses various SDOs, telco industry fora and open-source communities working on network slicing and related 5G applications. As captured in Section 8, these includes GSMA NG, 3GPP SA5, ETSI ZSM, Openslice and OSM and IETF TEAS. To maximize its impact on these groups, the PhD candidate has been pursued the following activities: i) setup a standardization roadmap, which has been regularly updated to reflect on new opportunities as they arise from the standardization landscape; ii) sync up the progress of thesis work with the progress of the different groups through constant monitoring of activities; and iii) seize opportunities to push technology contributions into ongoing specifications or recommendations, in some situations by re-shaping the main findings reported in published papers.

Finally, there is **exploitation**. As the innovation branch of a multinational telecommunication operator, Telefónica I+D (TID) seeks exploitation activities related to direct technology transfer to the relevant business units of Telefónica. These activities are aimed to i) improve Telefónica technology evaluation and testing facilities, focused on 5G infrastructure and innovative network services; and ii) gain experience from service validation practices, and structure feedback to Telefónica's business units, so that they can shape them into commercial solutions and products. In his role of TID's employee, the PhD candidate has been called upon to lead exploitation plan on 5G sliced services, leveraging his thesis's results around network slicing and its onto private 5G network environments, through the following activities:

- Testing and evaluation of particular slicing features on selected use cases. The use case validation has been done through the execution of PoCs at experimentation facilities such as 5TONIC lab, in collaboration with other partners, including academia, vendors, solution integrators and verticals.
- Participation on internal slicing projects, defined as part of the company's go-to-market strategy for this technology. The outcomes of thesis's projects are valuable for the first stage of this strategy, which according to GSMA [31] is to deploy network slicing for internal use, i.e., prove the validity of slicing by using it to serve internal customers within an operator or the operator sister companies. The PhD student has participated in the Request-For-Information (RFI) processes and in the definition of strategic field trials with shortlisted vendors. For the elaboration of requirements captured in the RFI, and their later evaluation against commercial solutions, lessons learnt from the thesis have been used, especially in the field of management and orchestration.

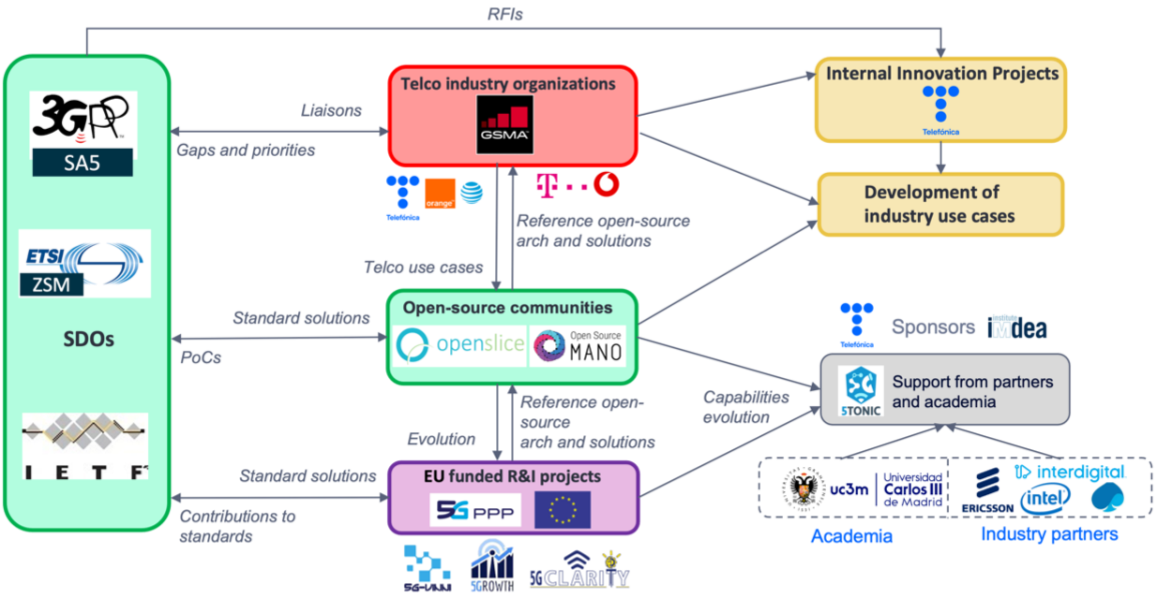


Figure I-15. Impact creation.

Figure I-15 summarizes the impact created with the work carried out in the present Thesis. Notice the workflows that govern the interaction between dissemination activities (scope research projects), standardization activities (scope telco industry fora, SDOs and open-source communities) and exploitation activities (scope Telefónica's internal innovation projects and development of industry use cases).

10 Publications

The study carried out in this dissertation and the proposed solutions have resulted in articles that have been published in top-tier magazines and renowned international conferences. These publications are:

[A] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges", *in IEEE Communications Magazine,* vol. 55, no. 5, pp. 80-87, May 2017. DOI: 10.1108/MCOM.2017.1600935. Impact Factor = 9.27 (2/87 Q1; Category: TELECOMMUNICATIONS).

[B] J. Ordonez-Lucena, O. Adamuz-Hinojosa, P. Ameigeiras, P. Munoz, J. J. Ramos-Munoz, J. Folgueira and D. Lopez, "The Creation Phase in Network Slicing: From a Service Order to an Operative Network Slice", *in 2018 European Conference on Networks and Communications (EuCNC),* Ljubljana, Slovenia, 2018, pp. 1-36. DOI: 10.1109/EuCNC.2018.8443255

[C] J. Ordonez-Lucena, C. Tranoris and J. Rodrigues, "Modeling Network Slice as a Service in a Multi-Vendor 5G Experimentation Ecosystem," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops),* 2020, pp. 1-6. DOI: 10.1109/ICCWorkshops49005.2020.9145225.

[D] J. Ordonez-Lucena, C. Tranoris, J. Rodrigues and L. M. Contreras, "Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility," in *2020 European Conference on Networks and Communications (EuCNC),* 2020, pp. 40-45. DOI: 10.1109/EuCNC48522.2020.9200940.

[E] J. Ordonez-Lucena, C. Tranoris and B. Nogales, "Automated Network Slice Scaling in Multi-site Environments: The ZSM PoC#2 report", 2021.

[F] J. Ordonez-Lucena, J. F. Chavarria, L. M. Contreras and A. Pastor, "The use of 5G Non-Public Networks to support Industry 4.0 scenarios," in *2019 IEEE Conference on Standards for Communications and Networking (CSCN),* 2019, pp. 1-7. DOI: 10.1109/CSCN.2019.89313

[G] J. Prados-Garzon, P. Ameigeiras, J. Ordonez-Lucena, P. Muñoz, O. Adamuz-Hinojosa and D. Camps-Mur, "5G Non-Public Networks: Standardization, Architectures and Challenges," in *IEEE Access*, vol. 9, pp. 153893-153908, 2021. DOI: 10.1109/ACCESS.2021.3127482. Impact Factor = 3.476 (43/94 Q2; Category: TELECOMMUNICATIONS).

[H] J. Ordonez-Lucena, P. Ameigeiras, L. M. Contreras, J. Folgueira, D. R. López, "On the Rollout of Network Slicing in Carrier Networks: A Technology Radar", *Sensors,* 2021, 21, 8094. DOI: 10.3390/s21238094. Impact Factor = 3.847 (95/276 Q2; Category: ENGINEERING, ELECTRICAL & ELECTRONIC).

In parallel to the study carried out in this dissertation, the PhD candidate has got other research merits, including the (co-)authoring of additional journal and conference papers, contributions to standards, talks at different venues, and other measurable outcomes. These merits are summarized in Section 9 (see impact creation in Table I-1) and further detailed in Annex A

11 Thesis Outline

This dissertation consists of a collection of papers, as opposed to a monograph. Thus, the contributions and findings of the study are presented in the included articles, which are collected in Parts II, III and IV of this document. Opening each of these parts we include two sections: 1) background context, which help reader to understand the precedents by providing a literature review; and 2) ambition, which identifies the main limitations of state-of-the-art, putting them in relation to the objectives of the thesis and the contributions in the different papers.

The outline of this thesis is as follows:

- **Part I.** It is introductory part, where we introduce and motivate the research carried out in this thesis. It also specifies the objectives in scope, the research methodology, the impact creation, and lists the publications conforming the compendium.
- **Part II.** This part addresses Objective 1 of this dissertation. Papers A and B compose this part, prefaced by the two canonical sections: background context and ambition.
- **Part III.** This part addresses Objective 2 of this dissertation. Papers C, D and E compose this part, prefaced by the two canonical sections: background context and ambition.
- **Paper IV.** This part addresses Objective 3 of this dissertation. Papers F, G and H compose this part, prefaced by the two canonical sections: background context and ambition.
- **Part V.** This part draws the main conclusions from this dissertation and outlines lines for future work.

The thesis makes use of numerous abbreviations which are spelled out in their first appearance for each chapter. We recommend that the reader use the List of Abbreviations included before Part I. A reference list is included at the end of each chapter. Note that references that are cited in different chapters may not be represented by the same number in all chapters.

Finally, though not part of the main body of this dissertation, this document includes two appendices. Annex A reports on the additional research activities performed by the author of this thesis, and which are not included in the compendium. Annex B includes a summary of the scope, objectives and main findings of the thesis, in Spanish.

References

[1] ITU-R Working Party WP SO: Draft New Recommendation, "IMT Vision-Framework and overall objectives of the future development of IMT for 2020 and beyond", Doc. RI2-SGOS-C-0199, June 2015.

[2] METIS project [Online]. Link

[3] 5G-PPP White Paper, ”5G Vision. The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services”, Mach. 2015 [Online]. Link

[4] The Next Generation Mobile Network (NGMN) Alliance [Online]. Link

[5] The Fifth Generation Mobile Communications Promotion Forum (5GMF) [Online]. Link

[6] The 5G Forum [Online]. Link

[7] 5G-PPP White Paper, ”5G Empowering vertical industries”, February 2016 [Online]. Link

[8] “5G White Paper”, NGMN Alliance, February 2015 [Online]. Link

[9] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini and H. Flinck, "Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429-2453, thirdquarter 2018. DOI: 10.1109/COMST.2018.2815638

[10] Ericsson and D. Little, "Network Slicing: A go-to-market guide to capture high revenue potential", Industrial Report 2021 [Online]. Link

[11] ONF TR-504, "SDN Architecture, Issue 1", November 2014 [Online]. Link

[12] ONF TR-521, "SDN Architecture, Issue 1.1", February 2016 [Online]. Link

[13] 3GPP TS 23.501, "5G; System Architecture for the 5G System (5GS)"

[14] Luis M. Contreras, "Progressive introduction of network softwarization in operational telecom networks: advanced at architectural, service and transport levels", July 2021 [Online]. Link

[15] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck and R. Boutaba, "Network Function Virtualization: State-of-the-Art and Research Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236-262, 2016. DOI: 10.1109/COMST.2015.2477041.

[16] ETSI White Paper, "Network Functions Virtualization: An Introduction, Benefits, Enablers, Challenges & Call for Action", October 2012 [Online]. Link

[17] European Telecommunications Standards Institute ISG NFV [Online]. Link.

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

[19] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637-646, Oct. 2016. DOI: 10.1109/JIOT.2016.2579198.

[20] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657-1681, 2017. DOI: 10.1109/COMST.2017.2705720.

[21] A. Filali, A. Abouaomar, S. Cherkaoui, A. Kobbane and M. Guizani, "Multi-Access Edge Computing: A Survey," in *IEEE Access*, vol. 8, pp. 197017-197046, 2020. DOI: 10.1109/ACCESS.2020.3034136.

[22] N. Hassan, K. A. Yau and C. Wu, "Edge Computing in 5G: A Review," in *IEEE Access*, vol. 7, pp. 127276-127289, 2019, DOI: 10.1109/ACCESS.2019.2938534.

[23] GSMA, "Telco Edge Cloud Value & Achievements Whitepaper", March 2022 [Online]. Link

[24] 3GPP TS 28.530, "Management and Orchestration; Concept, use cases and requirements".

[25] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," in *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[26] H. Zhang et al. , "Network Slicing Based 5G and Future Mobile Networks: Mobility, Re source Management, and Challenges," in *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[27] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega, D. Aziz, and H. Bakker, "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks," in *IEEE Communications Magazine*., vol. 55, no. 5, pp. 72–79, 2017.

[28] X. Li, M. Samaka, H. A. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, "Network Slicing for 5G: Challenges and Opportunities," in *IEEE Internet Computing*, vol. 21, no. 5, pp. 20–27, 2017.

[29] Davies and P. Thompson, "Challenges of Network Slicing", in *IEEE Network Softwarization*, January 2017 [Online]. Link

[30] V. Cunha et al., "Network slicing security: Challenges and directions", Wiley, vol. 2, Issue 5, 2019. DOI: 10.1002/itl2.125.

[31] GSMA, "Network Slicing Use Case Requirements", White Paper, April 2018. [Online]. Link

[32] Metro Ethernet Forum (MEF), "LSO (Lifecycle Service Orchestration)" [Online]. Link

[33] TM Forum, "Zero-touch Orchestration, Operations and Management (ZOOM) project" [Online]. Link

[34] TM Forum, "Open Digital Architecture (ODA)" [Online]. Link

[35] Chochliouros, I.P., Spiliopoulou, A.S., Lazaridis, P., Dardamanis, A., Zaharis, Z., Kostopoulos, A., "Dynamic Network Slicing: Challenges and Opportunities", in *Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) Artificial Intelligence Applications and Innovations*. AIAI 2020 IFIP WG 12.5 International Workshops. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 585. Springer, Cham. DOI: 10.1007/978-3-030-49190-1_5.

Part II

Multi-domain Network Slicing: System design

Literature Review and Problem Description

Part II addresses the Objective 1 of this dissertation, which is the design of a standards-compliant system architecture for multi-domain slicing. This objective is addressed in Papers A and B. Prefacing these publications, in this chapter we include two sections that help the reader have the full picture and understand the problem we want to address. Section 1 provides background context, capturing the precedents with a literature review. Section 2 identifies the main limitations of the state-of-the-art and puts them in relation with the contributions done in Papers A and B.

1 Background Context

1.1 Network slice concept

Network slicing concept in the context of 5G was first introduced by the NGMN Alliance in [1], where network slice is defined as a collection of 5G network functions and specific RAT settings that are combined together for the specific use case it is expected support. As a follow-up to this 5G White paper, the NGMN Alliance published in January 2016 a new document that further elaborated on the network slice concept [2]. This publication reports a multi-tier architecture for network slice implementation, consisting of three layers (see Figure), each contributing to the slice definition and provisioning with distinct tasks: i) service instance layer, ii) network slice instance and iii) resource layer. Building upon this network slicing concept outline, NGMN provides a characterization of a network slice instance, by listing the properties it shall be adhered to. Among these properties, three stand out: *isolation* (i.e., a network slice instance may be fully or partly, logically and/or physically, isolated from another network slice instance), *resource sharing* (i.e., a network slice instance may be composed of sub-network instances, which in turn might be shared by multiple instances) and *customizability* (i.e., instance-specific policies and configurations are required when creating a Network Slice Instance).

Leveraging NGMN work, other fora provide their own view on network slicing technology. For example, 3GPP defines network slicing as a technology that enables the operator to create networks, customized to provide optimized solutions for different market scenarios which demand diverse requirements, e.g., in terms of functionality, performance and isolation [3]. For ITU-T, network slicing is perceived as Logical Isolated Network Partitions (LINP) composed of multiple virtual resources, isolated and equipped with a programmable control and data plane [4]. Finally, GSMA claims that a network is a construction representing an independent end-to-end logical network that runs on a shared physical infrastructure, capable of providing negotiated service quality [5].

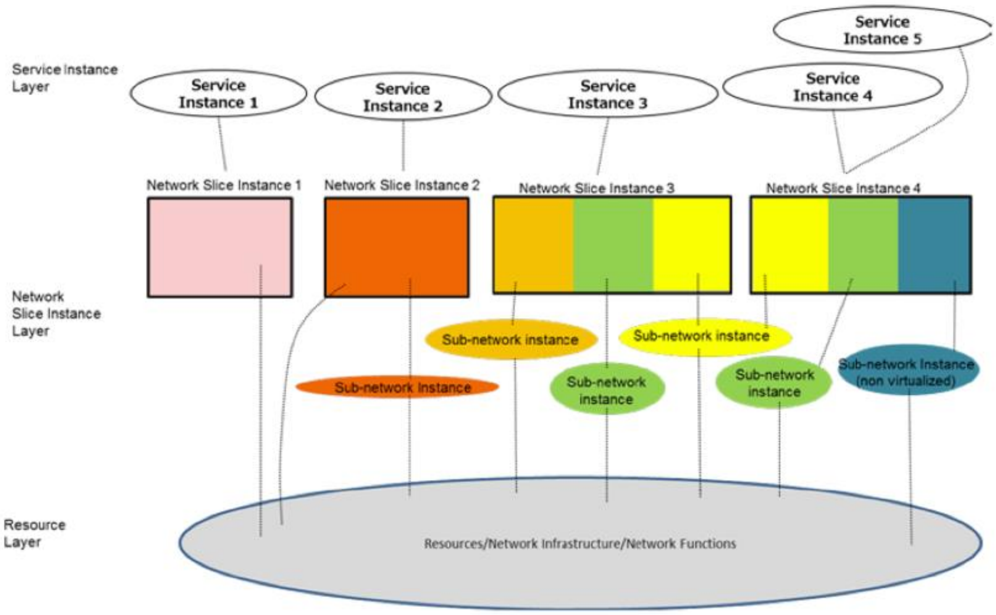


Figure II-1. Network slice conceptual outline. Source: NGMN Alliance [2].

1.2 System architectures combining SDN and NFV capabilities.

SDN and NFV are recognized to be the foundation of network slice realization. The software stacks associated to these two technologies need to be efficiently combined to provide necessary capabilities to create multiple logical network partitions atop a common infrastructure, ensuring their concurrent execution with service guarantees. As standard organizations and other industry fora kept working on solutions *propelling network softwarization* (e.g., ETSI, ONF, NGMN, GSMA) and *laying the groundwork for first 5G network solutions* (e.g., 3GPP), the European Commission launched the Horizon 2020 (H2020) programme, which allowed funding several research and innovation projects working on softwarized 5G networks. Most of them in scope the 5G-PPP initiative, the mission of these projects was to design, implement and validate network architectures that extend state-of-the-art solutions, with the goal of identifying gaps to guide progress and focus in the industry.

Though different projects bring different architectural proposals according to their targeted scope, it is fundamental to ensure consistency across them, in order to provide a consolidated view on the technical direction for the architecture design in the 5G era. In this regard, the 5G-PPP created the 5G Architecture Working Group. This group publishes regular white papers that present the architectural concepts developed in different 5G-PPP projects and other initiatives, identifying synergies, and proposing directions for future work, in terms of research and standardization. During the lifetime of the Part II in the present thesis, the 5G Architecture Working Group published two white papers, both scoping the latest findings from 5G-PPP Phase 1 projects (2015Q3-2017Q4).

The first white paper was released in July 2016 [6]. It captured novel trends and technology enablers for the realization of 5G architecture vision, putting them all together into a common framework. The architecture in Figure II-2 shows both mobile network functionality and management and orchestration functionality. This builds on ETSI NFV principles and MANO building blocks (i.e., NFVO, VNFM and VIM), extending them with the introduction of the E2E Service Management & Orchestration module as well as a programmable SDN controller. These two additional modules allow operators to flexibly configure, and control Virtualized and Physical network functions, according to the control/user plane requirements of the services executing atop.

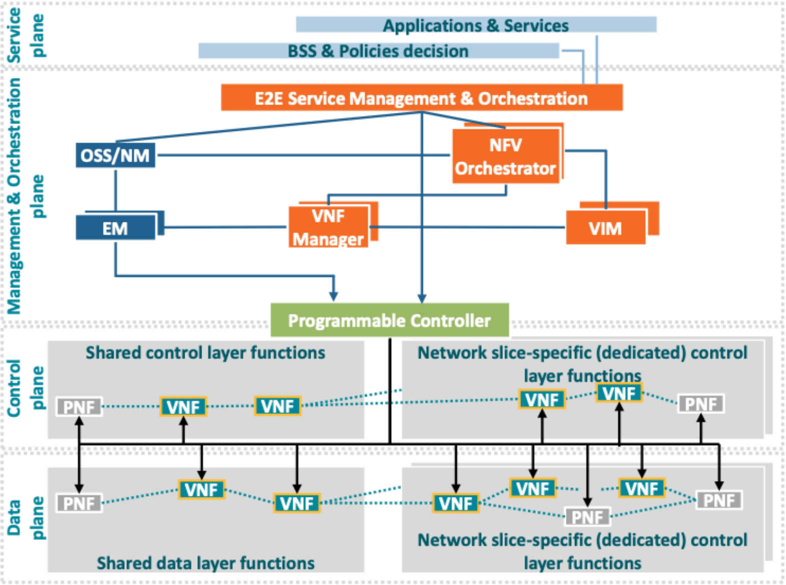

Service plane
Applications & Services
BSS & Policies decision
Management & Orchestration plane
E2E Service Management & Orchestration
OSS/NM
NFV Orchestrator
EM
VNF Manager
VIM
Programmable Controller
Control plane
Shared control layer functions
Network slice-specific (dedicated) control layer functions
PNF
VNF
VNF
VNF
VNF
PNF
Data plane
PNF
VNF
VNF
VNF
VNF
VNF
PNF
PNF
Shared data layer functions
Network slice-specific (dedicated) control layer functions

Figure II-2. SDN/NFV framework for control, management and orchestration of network functions. Source: 5G-PPP [6].

The second white paper, released in December 2017, presented the main findings and analyses of the different Phase I projects, along with the concept evaluations. The new document [7] put much more focus on the concepts of network slicing (and therefore multi-tenancy support), which are now considered an integral part of the system design. Figure II-3 pictures how the architectural framework and vision set by the original white paper is evolved to host slicing capabilities, with the integration of inter-slice resource broker [8] and the definition of common (shared) and dedicated (slice specific) network functions.

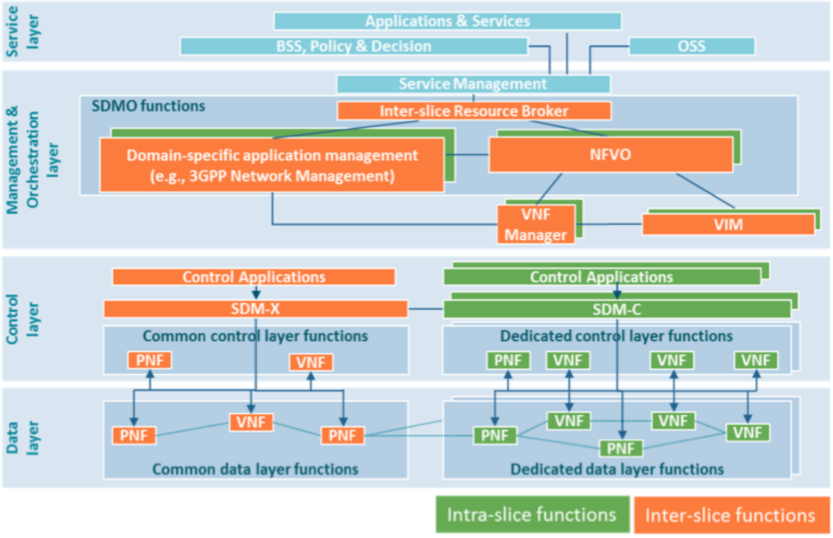


Figure II-3. 5G-PPP framework for the management and orchestration of network functions and slices. Source: 5G-PPP [7].

The abovementioned white papers capture the main insights of different 5G-PPP Phase I projects, but do not provide a deep dive on their specificities. Table II-1 tries to summarize this information, by presenting a non-exhaustive overview of all the relevant projects and their scope of work.

Table II-1. A summary of 5G-PPP Phase 1 projects working on SDN/NFV

Project Name	**Focus Area**		**Description**
	SDN	**NFV**	
5G-NORMA	Yes	Yes	Multi-service and context-aware adaptation of network functions to support a variety of services.
5G-Xhaul	Yes	Partial	Development of a scalable SDN control plane and mobility aware demand prediction models for optical/Wireless 5G networks
5G-CrossHaul	Partial	Yes	Design of 5G transport architectural solution that supports multi-domain orchestration among multiple network operators or service providers
5GEx	Yes	Yes	Specification of an SDN/NFV empowered framework that enables cross-domain orchestration of services over multiple administrative domains.
COGNET	Yes	Yes	Specification of an architectural framework that enables dynamic resource allocation to VNFs, such that SLA requirements can be met with optimal resource usage.
CHARISMA	Yes	Yes	Development of a software-defined converged fixed 5G mobile network architecture that offers both multi-technology and multi-operator features
COHERENT	Yes	No	Efficient radio resource modeling and management in programmable radio access networks.
METIS-II	Yes	Yes	Flagship project for a common evaluation of 5G radio access network concepts, preparing concerted action towards regulatory and standardization bodies
SELFNET	Yes	Yes	Development of efficient self-organizing network management framework for 5G through the combination of a virtualized and software-defined infrastructure with Artificial Intelligence technologies.
SESAME	Yes	Partial	Development of programmable 5G infrastructure that supports multi-tenancy and decreases network OPEX, whilst increasing the QoS and security.
Superfluidity	Partial	Yes	Specification of a converged 5G network architecture where network services with the possibility to deploy network services across the edge-to-cloud continuum.

1.3 Multi-domain orchestration

Multi-domain orchestration refers to the automated management of resources and services that span across different technologies, network segments and/or legal operational boundaries. The implementation of this feature in 5G is set to enable the interaction of multiple administrative domains at different levels, with different service and infrastructure providers. When applied to end-to-end slices, it shall ensure that allocation requests are mapped into corresponding (technology, network and administrative) domains while matching the service requirements of individual slices. Perez-Caparros et al. [9] was among the first to design the multi-domain orchestration use cases and its requirements. This was followed by several research works [10][12]-[12] that elaborate on how to map service requests on top of a federated environment. These works analyze existing VNF allocation embedding algorithms in the literature (e.g., [13]) and propose extensions to multi-domain environments.

In terms of terms of architecture work, one of the first ones was [14]. In this research-oriented paper, the authors plan explored a natural step ahead by considering the case of multiple administrative domains. They propose different solutions through different user stories and set out directions for future work. An initial analysis of multi-domain orchestration frameworks is given in [15]. Of particular interest is the architecture solution proposed by the 5GEx project, first presented in [16] and further elaborated in [17]. This solution laid the foundation for setting out a generic architectural framework at 5G-PPP community, with the idea of being i) applicable to all Phase 1 projects, and ii) used as a baseline solution in upcoming Phase 2 projects, e.g., 5G-TRANSFORMER [18].

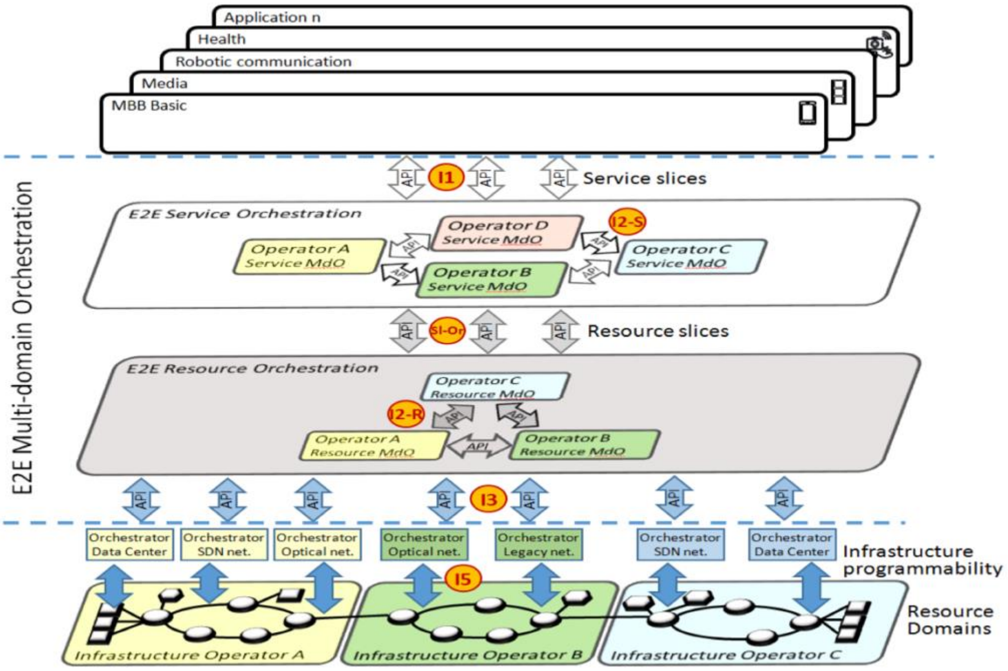


Figure II-4. 5G-PPP framework for multi-domain resource and service orchestration. Source: [6]

The abovementioned 5GPPP reference architecture is pictured in Figure II-4. The cornerstone is the Multi-domain Orchestrator (MdO) concept. A MdO coordinates resource and service orchestration activities at multi-technology level (i.e., network segments with different cloud and networking technologies) or multi-operator level (e.g., different administrative domains). The Resource MdO belonging to an infrastructure operator, for instance operator A, interacts with domain orchestrators, via **interface I3** APIs, to orchestrate resources within the same administrative domains. The MdO interacts with other MdOs via **interface I2-R (B2B)** APIs to request and orchestrate resources across administrative domains. Resources are exposed at the service orchestration level on interface **Sl-Or** to Service MdOs. **Interface I2-S (B2B**) is used by Service MdOs to orchestrate services across administrative domains. Finally, the Service MdOs expose, on **interface I1 (C2B)**, service specification APIs that allow business customers to specify their requirements for a service. The framework also considers MdO service providers, such as Operator D in Figure II-4, which do not own resource domains but operate a multi-domain orchestrator level to trade resources and services, acting as kind of MVNO or as a broker engine (3rd party enforcing federation).

1.4 Network slice characterization

To proceed with the provisioning of a network slice, it is important to have solutions that provide answers to the following questions:

- **Question 1 (Q1)**: How can an operator capture the capabilities that the slice shall provide to hosted services?
- **Question 2 (Q2):** How to flexibly map this slice description to the appropriate infrastructure elements and network functions?

The state-of-the-art solutions developed for these two questions are detailed below.

1.4.1 Solutions for Q1

The NGMN Alliance was the first to propose a way forward for Q1, with the introduction of network slice blueprint concept. As reported in [2], a network slice blueprint is a manifest that provides a complete description of the structure, configuration, and the plans/workflows for how to instantiate and control a network slice instance during its lifecycle. In other words, NGMN sees a network slice blueprint as an artifact that allows operator to create (and operate) instances of a particular network slice. This idea was later reinforced in [19], where 5GAmericas argued for the need to design blueprints following a similar structure as for the NFV descriptors, including NSDs and VNFDs. With this approach, the advantages inherent to model-based approaches (e.g., reusability, replicability, and provisioning automation) would be extended further, up to the slice level.

For the design of a network slice blueprint, there are two different yet complementary approaches. In one approach, the network slice blueprint is simply a template that allows describing slice capabilities, that can be matched against requirements of incoming service requests. The authors of [20] elaborates a solution following this top-down approach, with the definition of a slice manifest that captures information on the traffic characteristics, KPIs (e.g., throughput, latency, availability) and add-on services (e.g., localization service, monitoring application) that the slice can deliver to hosted services. In the second approach, the network slice blueprint is more detailed in the sense that it can identify functions or Radio Access Technologies (RATs) that are bundled together fitting the specific needs of a service. This bottom-up approach is for example presented in [21], where the slice blueprint allows defining a slice as a composition of building blocks, with these being network and application functions that are available in a marketplace.

One can notice that the main difference between the two approaches lies on the way that the network slice will be generated.

- In the top-down approach, the slice orchestrator will be assigned the more complex task of identifying the appropriate functions and technologies that offer the capabilities declared in the slice blueprint.
- In the bottom-down approach, slice-to-resource mapping gets simplified, as the building blocks of the slice are already specified in the slice blueprint. However, this simplicity comes at the cost of having less efficient solutions, as it leaves less flexibility to the slice orchestrator to tune the components of the slice.

1.4.2 Solutions for Q2

In relation to Q2, the issue is how to translate the information captured in the slice blueprint to deployment and configuration instructions of corresponding network components. According to [22], this translation entails two types of mapping: i) the functional/SLA mapping of the service requirements to network infrastructures; and ii) the mapping of network functions and infrastructure types to vendor implementations. The complexity of these mappings entirely depends on the approach chosen for Q1, as explained in section 1.4.1.

Apart from [20]-[21], there are other papers that provide alternatives to execute his mapping. For example, [23] follows the 5GAmericas recommendations, by developing a slice orchestrator that sits atop the NFVO, and connects network slice blueprints to NFV descriptors. This slice orchestrator consists of an internal module, referred to as "Slice2NS mapper", in charge of converting network slice blueprints into NSDs (and associated instantiation parameters). This builds on the fact that one network slice consists of one or more NFV network services [24]. The authors of [25] define tools and modelling primitives for correctly defining and on-boarding the information needed to create a network slice, bundling them into a Network Slice Design Studio. The tools available in this workspace guide the user to design the network slice blueprint and implement the workflows for network slice instantiation. Following the recommendations from ii), the Network Slice Design Studio considers the available network function models and infrastructure resource models, and associated network characteristics.

2 Ambition

2.1 Identification of key network slice capabilities

Section 1.1 summarizes the different definitions for network slicing when this concept was first defined. As one may notice from these definitions, all of them are incomplete. For example, NGMN Alliance [1] and 3GPP [3] focus on the definition of what a network slice contains, but not on what it is. ITU [4] and GSMA [5] bridge this gap, by clarifying that a network slice is an end-to-end logical network; however, the main drawback of this second round of definitions is that they omit the capabilities that differentiate network slices from traditional virtual network solutions (e.g., Virtual Private Networks or enterprise WAN networks). Some of these capabilities can be inferred from the list of properties that NGMN reports for a network slice instance in [2], including isolation, resource sharing and customizability. Even though these capabilities provide a good working basis, their description in the document is, by any reckoning, insufficient to have a solid specification of network slicing concept. For example, regarding *isolation*, it is reported that "a network slice instance may be fully or partly, logically and/or physically, isolated from another network slice". However, it is far from crystal clear what this physical and isolation means, when to apply one or another, their impact on the performance and security of the slice, or how can they be enforced at the infrastructure level. Additionally, how this isolation is related to the multi-tenancy feature that the slicing is presumed to support needs further clarification. Several questions also arise from the descriptions given for *resource sharing* and *customizability*.

Beyond the state-of-the-art: This thesis will overcome the limitations identified from the conducted literature review, by:

- providing a self-contained definition of network slice concept, with focus on the specificities that make network slicing an innovative approach compared to other carrier-grade virtual network solutions;
- identifying the key capabilities that govern the provisioning and operation of a network slice, elaborating on them to clearly specify how they provide means for multi-tenancy support. These capabilities extend the ones reported in [2], including orchestration.

Related objectives: O1.1

Means of verification: Paper A

2.2 Standards-compliant network slicing architecture

Section 1.2 overviews the SDN/NFV stacks which were provided by the different 5G-PPP Phase 1 projects. These stacks were combined to generate the reference 5G-PPP network slicing architecture, whose final design is pictured in Figure II-3. Though the system architecture is complete from a functional viewpoint, in the sense that it consists of an inter-slice resource broker (slicing capabilities) on top of SDM-C (SDN capabilities) and MANO (NFV capabilities), it presents two limitations that worth mentioning.

- There is not a well-defined list of slicing capabilities that the system shall fulfil, and therefore built upon. This might question the position of certain components in the architecture.
- The interfaces connecting these components are not specified. This prevents the disclosure of what specific information is exchanged across these components. For example, it is unclear how the inter-slice resource broker levers on NFVO offerings (e.g., SOL005) to policy resource allocation across different slices. What is more, there does not exist an interface between the MANO stack and the SDM-C, which prevents the combined use of NFV and SDN capabilities in this system.

In relation to the multi-domain orchestration system presented in Section 1.3, the main limitation is the lack of alignment with the solutions worked out in the different SDOs and industry fora. Actually, the design shown in Figure II-4 consists of a number of novel interfaces, such as I2-R, I2-S, I3, I1 and Sl-Or. However, it is not clear how some of these API based reference points leverage and extend standard NFV reference points. Additionally, there is no reference to any work in SDN and related orchestration activities, such as the ones promoted by ONF in [26] and [27]; actually, it is assumed that SDN has nothing to do with the support of slicing in multi-domain scenarios.

Beyond the state-of-the-art: This thesis will overcome the limitations identified from the conducted literature review, by defining a standards-compliant network slicing system architecture which is

- robust and scalable, so that the network slice properties on isolation, customizability, resource sharing and multi-tenancy support (among others) can be met;
- grounded on the use of SDN and NFV modules/interfaces that ONF and ETSI ISG NFV define in their specifications. The combination of these building blocks shall allow for slicing awareness in resource dispatching;
- applicable to single and multiple administrative domains, extending slice serviceability beyond the footprint of one single network provider.

Related objectives: O1.1

Means of verification: Paper A

2.3 Model-based network slice description

The literature review compiled in Section 1.4 reveals the problem statement on network slice characterization, with different solutions to address open issues captured in Q1 and Q2.

In relation to Q1, the pros and cons of going for a top-down approach (e.g., [20]) or bottom-up approach (e.g., [21]) have been remarked; however, there are no proposals in between.

The works described in [23] and [25] provide a sound starting point towards Q2. However, they present limitations that are worth mentioning. For example, though the ‘Slice2NS mapper’ reported in [23] promotes reusability and enables defining slicing constructions with minimal integration to existing MANO framework, this solution has one major drawback: for those cases that the slice requirements do not map to an already on-boarded NSD (i.e., available in the NFVO catalog), it requires automatic generation of a new NSD. The on-demand NSD creation out of a network slice blueprint is not easy task, and it involves several steps, including feasibility check (e.g., checking service requirements against network capabilities and state of resources in the inventories), packetization of existing VNFDs into a new NSD, and the integrity check of this NSD before onboarding it to the NFVO. The logic behind this workflow is complex, and requires it to be executed relatively fast, to cope with the dynamism presumed for network slicing. In relation to the Network Slice Design Studio reported in [25], the major limitation is that despite providing all the ingredients in scope of Q2, it does not specify the recipe to mix them up. Additionally, the solution is proposed in the context of a system architecture that is not standards-compliant, and thus it is not rooted to SDN and NFV capabilities captured in ONF and NFV. This is the same problem that was reported in the previous section.

Beyond the state-of-the-art: This thesis will overcome the limitations identified from the conducted literature review, by providing a solution for network slice characterization that

- it is contextualized in a standards-compliant system architecture. This architecture is the same as the one reported in paper A;
- proposes a novel approach to solve Q1, which is halfway between the top-down and bottom-up approaches, making the best of both worlds.
- specifies the steps to solve Q2, with a well-defined workflow that describes how to go from service ordering to a deployed network slice instance, solving the shortcomings reported from the literature review.

Related objectives: O1.2

Means of verification: Paper B

References

[1] “5G White Paper”, NGMN Alliance, February 2015 [Online]. Link.

[2] “Description of Network Slicing Concept”, NGMN Alliance, January 2016 [Online]. Link.

[3] 3GPP TR 22.864, “Feasibility Study on New Services and Markets Technology Enablers for Network Operation”, Stage 1, Release 14, v14.1.0, Sep. 2016.

[4] ITU-T Y.3011, “Framework of network virtualization for future networks”, June 2012 [Online]. Link.

[5] “An Introduction to Network Slicing”, GSMA White Paper, 2017 [Online]. Link.

[6] “View on 5G Architecture Version 1.0”, 5GPPP Architecture Working Group, July 2016 [Online]. Link.

[7] View on 5G Architecture Version 2.0”, 5GPPP Architecture Working Group, December 2017 [Online]. Link.

[8] K. Samdanis, X. Costa-Perez and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," in *IEEE Communications Magazine*, vol. 54, no. 7, pp. 32-39, July 2016. DOI: 10.1109/MCOM.2016.7514161.

[9] D. Perez-Caparros, I. Vaishnavi, S. Schmid and A. Khan, "An architecture for creating and managing virtual networks”, in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 2984-2988. DOI: 10.1109/PIMRC.2013.6666658.

[10] I. Houidi, W. Louati, W.B. Ameur, D. Zeghlache, “virtual network provisioning across multiple substrate networks”, *in Computer Networks*, 55 (4), 2013, pp. 1011–1023.

[11] I. Vaishnavi, R. Guerzoni and R. Trivisonno, "Recursive, hierarchical embedding of virtual infrastructure in multi-domain substrates”, in *Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft)*, 2015, pp. 1-9. DOI: 10.1109/NETSOFT.2015.7116141.

[12] J. Martín-Pérez and C. J. Bernardos, "Multi-Domain VNF Mapping Algorithms”, in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2018, pp. 1-6. DOI: 10.1109/BMSB.2018.8436765.

[13] M. Chowdhury, M. R. Rahman and R. Boutaba, "ViNEYard: Virtual Network Embedding Algorithms With Coordinated Node and Link Mapping", in *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 206-219, Feb. 2012. DOI: 10.1109/TNET.2011.2159308.

[14] R. V. Rosa, M. A. Silva Santos and C. E. Rothenberg, "MD2-NFV: The case for multi-domain distributed network functions virtualization," in *2015 International Conference and Workshops on Networked Systems (NetSys)*, 2015, pp. 1-5, doi: 10.1109/NetSys.2015.7089059.

[15] R. Guerzoni et al., “Analysis of end-to-end multi-domain management and orchestration frameworks for software defined infrastructures: an architectural survey”, in Transactions on Emerging Telecommunications Technologies, 28 (4), 2016, pp. 1–19. DOI: 10.1002/ett.3103

[16] A. Sgambelluri et al., "Orchestration of Network Services across multiple operators: The 5G Exchange prototype”, in *2017 European Conference on Networks and Communications (EuCNC)*, 2017, pp. 1-5. DOI: 10.1109/EuCNC.2017.7980666.

[17] 5GEx Deliverable 2.2, "5GEx Final System Requirements and Architecture", December 2017 [Online]. Link.

[18] 5G-TRANSFORMER Deliverable D1.2, "5G-TRANSFORMER Initial System Design", 2018 [Online]. Link.

[19] "Network Slicing for 5G Networks and Services", 5G Americas White Paper, November 2016 [Online]. Link.

[20] X. Zhou, R. Li, T. Chen and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," in *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146-153, July 2016. DOI: 10.1109/MCOM.2016.7509393.

[21] N. Nikaein et al., "Network Store: Exploring Slicing in Future 5G Networks", in *Proc. 10th ACM International Workshop Mobility in the Evolving Internet Architecture*", Sept. 2015.

[22] X. Foukas, G. Patounas, A. Elmokashfi and M. K. Marina, "Network Slicing in 5G: Survey and Challenges", in *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94-100, May 2017. DOI: 10.1109/MCOM.2017.1600951.

[23] B. Chatras, U. S. Tsang Kwong and N. Bihannic, "NFV enabling network slicing for 5G," in *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, 2017, pp. 219-225. DOI: 10.1109/ICIN.2017.7899415.

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

[25] A. Devlic, A. Hamidian, D. Liang, M. Eriksson, A. Consoli and J. Lundstedt, "NESMO: Network slicing management and orchestration framework," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 1202-1208. DOI: 10.1109/ICCW.2017.7962822.

[26] ONF TR-521, "SDN Architecture", Issue 1.1, 2016 [Online]. Link.

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

Paper A

Network Slicing for 5G with SDN/NFV: Concepts, Architectures and Challenges

Jose Ordonez-Lucena, Pablo Ameigeiras, Diego López, Juan J. Ramos-Munoz, Javier Lorca and Jesús Folgueira

Disclaimer:
This work has been published in the IEEE Communications Magazine.
Citation information: DOI 10.1109/MCOM.2017.1600935

Copyright:

Abstract

The fifth generation of mobile communications is anticipated to open up innovation opportunities for new industries such as vertical markets. However, these verticals originate myriad use cases with diverging requirements that future 5G networks have to efficiently support. Network slicing may be a natural solution to simultaneously accommodate over a common network infrastructure the wide range of services that vertical-specific use cases will demand. In this article, we present the network slicing concept, with a particular focus on its application to 5G systems. We start by summarizing the key aspects that enable the realization of so-called network slices. Then, we give a brief overview on the SDN architecture proposed by the ONF and show that it provides tools to support slicing. We argue that although such architecture paves the way for network slicing implementation, it lacks some essential capabilities that can be supplied by NFV. Hence, we analyze a proposal from the ETSI to incorporate the capabilities of SDN into the NFV architecture. Additionally, we present an example scenario that combines SDN and NFV technologies to address the realization of network slices. Finally, we summarize the open research issues with the purpose of motivating new advances in this field.

1 Introduction

5G systems are nowadays being investigated to satisfy the consumer, service and business demands of 2020 and beyond. One of the key drivers of 5G systems is the need to support a variety of vertical industries such as manufacturing, automotive, healthcare, energy, and media & entertainment [1]. Such verticals originate very different use cases, which impose a much wider range of requirements than existing services do nowadays. Today's networks, with their "one-size-fits-all" architectural approach, are unable to address the diverging performance requirements that verticals impose in terms of latency, scalability, availability and reliability. To efficiently accommodate vertical-specific use cases along with increased demands for existing services over the same network infrastructure, it is accepted that 5G systems will require architectural enhancements with respect to current deployments.

Network softwarization, an emerging trend which seeks to transform the networks using software-based solutions, can be a potential enabler for accomplishing this. Through technologies like Software-Defined Networking (SDN) and Network Function Virtualization (NFV), network softwarization can provide the programmability, flexibility, and modularity that is required to create multiple logical (virtual) networks, each tailored for a given use case, on top of a common network. These logical networks are referred to as network slices. The concept of separated virtual networks deployed over a single network is indeed not new (e.g., VPN), although there are specificities that make network slices a novel concept. We define network slices as end-to-end (E2E) logical networks running on a common underlying (physical or virtual) network, mutually isolated, with independent control and management, and which can be created on demand. Such self-contained networks must be flexible enough to simultaneously accommodate diverse business-driven use cases from multiple players on a common network infrastructure (see Figure A-1).

In this paper, we provide a comprehensive study of the architectural frameworks of both SDN and NFV as key enablers to achieve the realization of network slices. Although these two approaches are not yet commonplace in current networking practice, especially in public wide area networks (WANs), their integration offers promising possibilities to adequately meet the slicing requirements. Indeed, many 5G research and demonstration projects (such as 5GNORMA, 5GEx, 5GinFIRE, or 5G!Pagoda) are addressing the realization of 5G slicing through the combination of SDN and NFV. Thus, we present a deployment example that illustrates how NFV functional blocks, SDN controllers, and their interactions can fully realize the network slicing concept. Furthermore, we identify the main challenges arising from implementing network slicing for 5G systems.

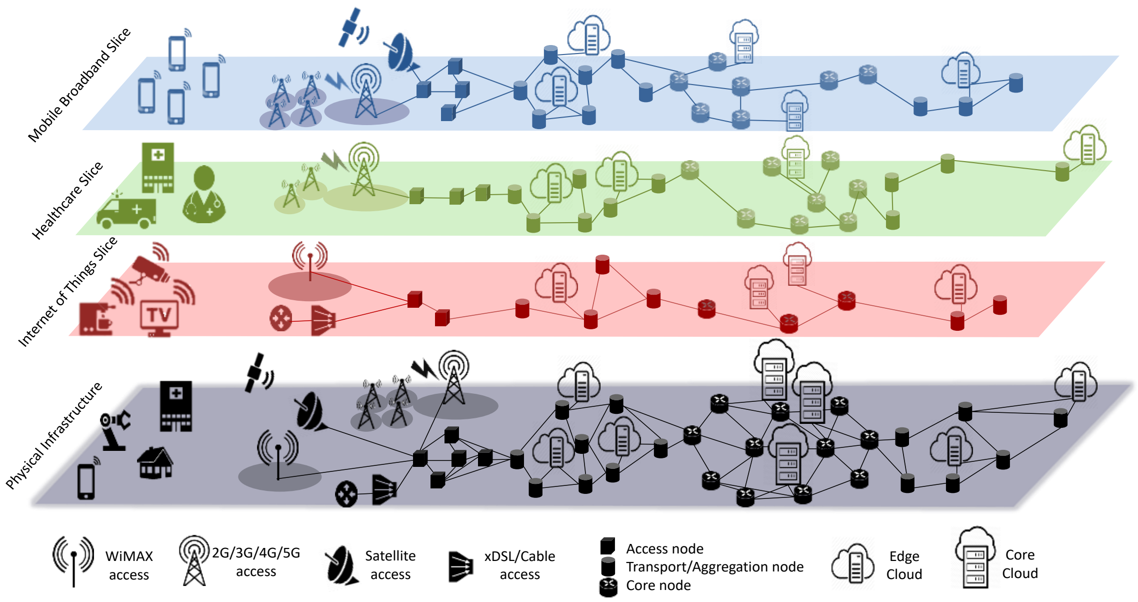


Figure A-1. 5G network slices running on a common underlying multi-vendor and multi-access network. Each slice is independently managed and addresses a particular use case.

The remainder of this paper is organized as follows: Section 2 provides a background on key concepts for network slicing. Sections 3 and 4 describe the SDN architecture from the ONF and the NFV architecture from the ETSI, respectively. Section 5 shows a network slicing use case with NFV and SDN integration, and Section 6 provides the main challenges and future research directions.

2 Background on key concepts for Network Slicing

In this section, we provide a background on key aspects that are necessary to realize the network slicing concept.

2.1 Resources

In its general sense, a resource is a manageable unit, defined by a set of attributes or capabilities that can be used to deliver a service. A network slice is composed of a collection of resources that, appropriately combined together, meet the service requirements of the use case that such slice supports. In network slicing, we consider two types of resources:

- Network Functions (NFs): functional blocks that provide specific network capabilities to support and realize the particular service(s) each use case demands. Generally implemented as software instances running on infrastructure resources, NFs can be physical (a combination of vendor-specific hardware and software, defining a traditional purpose-built physical appliance) and/or virtualized (network function software is decoupled from the hardware it runs on).

- Infrastructure Resources: Infrastructure Resources: heterogeneous hardware and necessary software for hosting and connecting NFs. They include computing hardware, storage capacity, networking resources (e.g., links and switching/routing devices enabling network connectivity) and physical assets for radio access. Suitable for being used in network slicing, the aforementioned resources and their attributes have to be abstracted and logically partitioned leveraging virtualization mechanisms, defining virtual resources that can be used in the same way as physical ones.

2.2 Virtualization

Virtualization is a key process for network slicing as it enables effective resource sharing among slices. Virtualization is the abstraction of resources using appropriate techniques. Resource abstraction is the representation of a resource in terms of attributes that match predefined selection criteria while hiding or ignoring aspects that are irrelevant to such criteria, in an attempt to simplify the use and management of that resource in some useful way. The resources to be virtualized can be physical or already virtualized, supporting a recursive pattern with different abstraction layers.

Just as server virtualization [2] makes virtual machines (VMs) independent of the underlying physical hardware, network virtualization [3] enables the creation of multiple isolated virtual networks that are completely decoupled from the underlying physical network, and can safely run on top of it.

The introduction of virtualization to the networking field enables new business models, with novel actors and distinct business roles. We consider a framework with three kinds of actors:

- Infrastructure Provider (InP): owns and manages a given physical network and its constituent resources. Such resources, in the form of WANs and/or data centers (DCs), are virtualized and then offered through programming interfaces to a single or multiple tenants.
- Tenant: leases virtual resources from one or more InPs in the form of a virtual network, where the tenant can realize, manage and provide network services to its users. A network service is a composition of NFs, and it is defined in terms of the individual NFs and the mechanism used to connect them.
- End user: consumes (part of) the services supplied by the tenant, without providing them to other business actors.

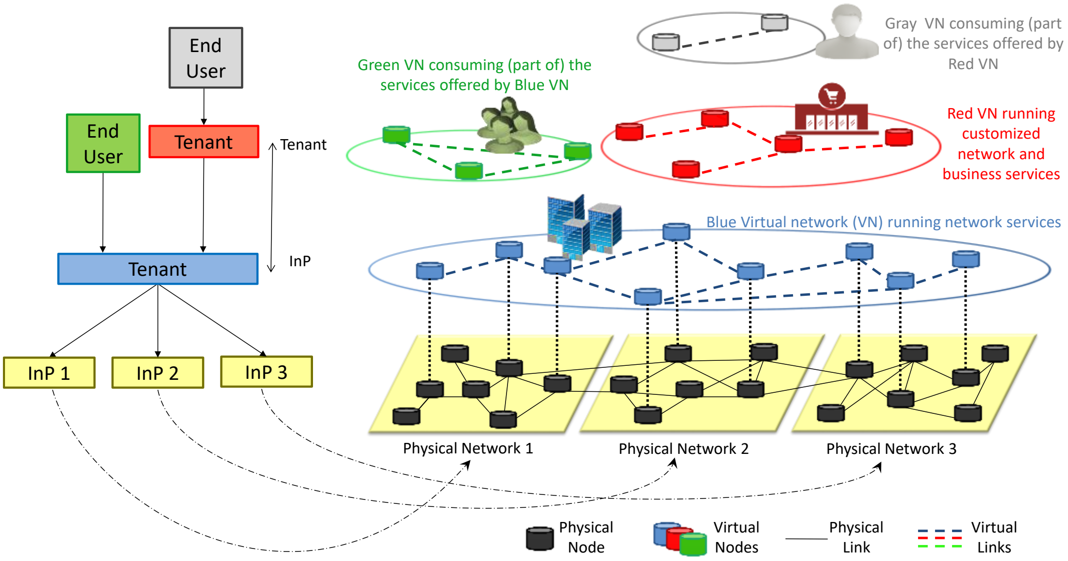

End User
End User
Tenant
Tenant
Tenant
InP
InP 1
InP 2
InP 3
Green VN consuming (part of) the services offered by Blue VN
Gray VN consuming (part of) the services offered by Red VN
Red VN running customized network and business services
Blue Virtual network (VN) running network services
Physical Network 1
Physical Network 2
Physical Network 3
Physical Node
Virtual Nodes
Physical Link
Virtual Links

Figure A-2. InPs and tenants as virtualization actors. These scenarios show the recursion principle, where these actors happen in a vertical multi-layered pattern.

As discussed above, virtualization is naturally recursive, and the first two actors can happen in a vertical multi-layered pattern, where a tenant at one layer acts as the InP at the layer immediately above. The recursion mentioned here implies that a tenant can provide network services to an end user, but also to another tenant (see Figure A-2). In such a case, this second tenant would provide more advanced network services to its own users.

2.3 Orchestration

Orchestration is also a key process for network slicing. In its general sense, orchestration can be defined as the art of both bringing together and coordinating disparate things into a coherent whole. In a slicing environment, where the players involved are so diverse, an orchestrator is needed to coordinate seemingly disparate network processes for creating, managing and delivering services.

A unified vision and scope of orchestration has not been agreed upon. According to the Open Network Foundation (ONF) [4], orchestration is defined as the continuing process of selecting resources to fulfill client service demands in an optimal manner. The idea of optimal refers to the optimization policy that governs orchestrator behavior, which is expected to meet all the specific policies and SLAs associated with clients (e.g., tenants or end users) that request services. The term continuing means that available resources, service demands, and optimization criteria may change in time. Interestingly, orchestration is also referred in [4] as the defining characteristic of an SDN controller. Note that client is a term used in SDN context.

The ONF states that the orchestrator functions include client-specific service demand validation, resource configuration, and event notification. For a more detailed description of these functions, see Section 6.2 in [5].

However, in network slicing orchestration cannot be performed by a single centralized entity, not only because of the complexity and broad scope or orchestration tasks, but also because it is necessary to preserve management independence and support the possibility of recursion. In our view, a framework in which each virtualization actor (see Section 2.2) has an entity performing orchestration functions seems more suitable to satisfy the above requirements. The entities should exchange information and delegate functionalities between them to ensure that the services delivered at a certain abstraction layer satisfy the required performance levels with optimal resource utilization.

2.4 Isolation

Strong isolation is a major requirement that must be satisfied to operate parallel slices on a common shared underlying substrate. The isolation must be understood in terms of:

- Performance: each slice is defined to meet particular service requirements, usually expressed in the form of KPIs. Performance isolation is an E2E issue and has to ensure that service-specific performance requirements are always met on each slice, regardless of the congestion and performance levels of other slices.
- Security and privacy: attacks or faults occurring in one slice must not have an impact on other slices. Moreover, each slice must have independent security functions that prevent unauthorized entities to have read or write access to slice-specific configuration/management/accounting information, and able to record any of these attempts, whether authorized or not.
- Management: each slice must be independently managed as a separate network.

To achieve isolation, a set of appropriate, consistent policies and mechanisms have to be defined at each virtualization level, following the ideas introduced in Section 2.3. The policies (what is to be done) contain lists of rules that describe how different manageable entities must be properly isolated, without delving into how this can be achieved. The mechanisms (how it is to be done) are the processes that are implemented to enforce the defined policies. From our point of view, to fully realize the required isolation level, the interplay of both virtualization and orchestration is needed.

3 ONF Network Slicing Architecture

The SDN architecture provided by the ONF comprises an intermediate control plane that dynamically configures and abstracts the underlying forwarding plane resources so as to deliver tailored services to clients located in the application plane (see SDN basic model in [5]). This is well aligned with the requirements of 5G network slicing, which needs to satisfy a wide range of service demands in an agile and cost-effective manner. Thus, the SDN architecture is an appropriate tool for supporting the key principles of slicing. The purpose of this section is to describe the SDN architecture and how it can be applied to enable slicing in 5G systems.

According to [4], the major SDN architectural components are resources and controllers. For SDN, a resource is anything that can be utilized to provide services in response to client requests. This includes infrastructure resources and NFs (see Section 2.1), but also network services, in application of the recursion principle described in Section 2. A controller is a logically centralized entity instantiated in the control plane which operates SDN resources to deliver services in an optimal way. Therefore, it mediates between clients and resources, acting simultaneously as server and client via client and server contexts, respectively. Both contexts are conceptual components of an SDN controller enabling the server-client relationships (see Figure A-3):

- Client context: represents all the information the controller needs to support and communicate with a given client. It comprises a Resource Group and a Client support function. The Resource Group contains an abstract, customized view of all the resources that the controller, through one of its northbound interfaces, offers to the client, in order to deliver on its service demands and facilitate its interaction with the controller. Client support contains all that is necessary to support client operations, including policies on what the client is allowed to see and do [4], and service-related information to map actions between the client and the controller.

- Server context: represents all the information the controller needs to interact with a set of underlying resources, assembled in a Resource Group, through one of its southbound interfaces.

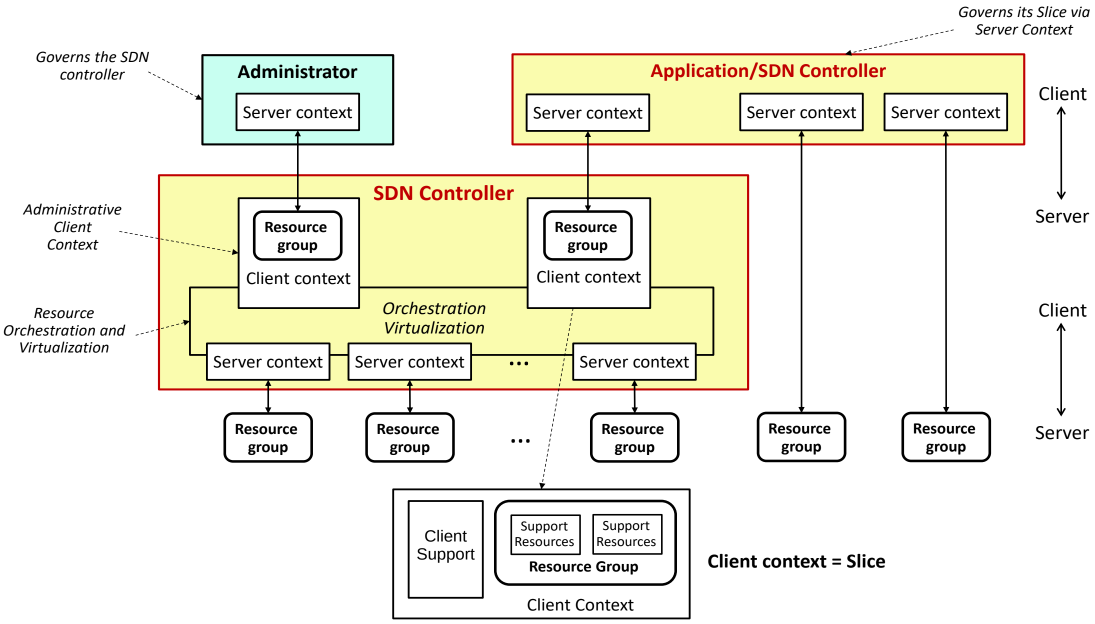


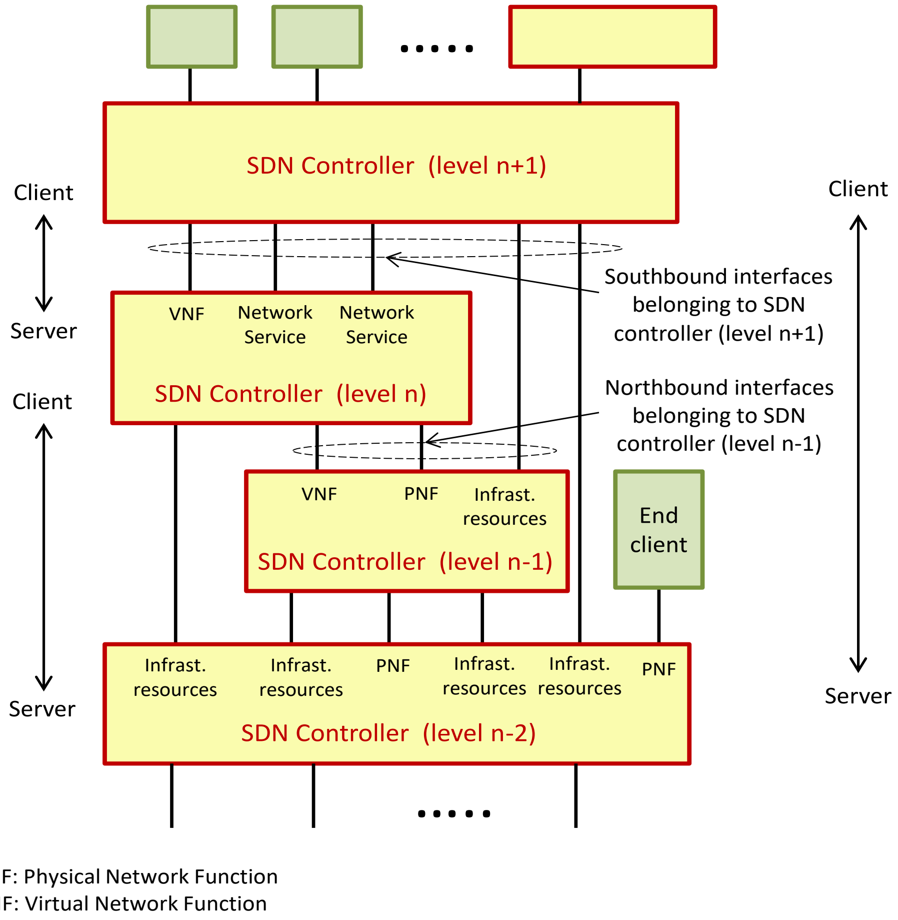
Figure A-3. ONF SDN Network Slicing architecture

The process of transforming the set of Resource groups accessed through server contexts to those defined in separate client contexts is not straightforward, and it requires the SDN controller to perform virtualization and orchestration functions.

When performing the virtualization function, the SDN controller carries out the abstraction and the aggregation/partitioning of the underlying resources. Thanks to virtualization, each client context provides a specific Resource Group that can be used by the client associated with that context to realize its service(s). Through the orchestration (see Section 2.3), the SDN controller optimally dispatches the selected resources to such separate Resource Groups. The interplay of both controller functions enables the fulfillment of the diverging service demands from all clients while preserving the isolation among them.

The SDN architecture also includes an administrator. Its tasks consist of instantiating and configuring the entire controller, including the creation of both server and client contexts and the installation of their associated policies.

According to the ONF vision, the SDN architecture naturally supports slicing [5], as the client context provides the complete abstract set of resources (as Resource Group) and supporting control logic that constitutes a slice, including the complete collection of related client service attributes.



SDN Controller (level n+1)
Client
Server
Client
Server
Client
Server
Southbound interfaces belonging to SDN controller (level n+1)
Northbound interfaces belonging to SDN controller (level n-1)
VNF
Network Service
Network Service
SDN Controller (level n)
VNF
PNF
Infrast. resources
SDN Controller (level n-1)
End client
Infrast. resources
Infrast. resources
PNF
Infrast. resources
Infrast. resources
PNF
SDN Controller (level n-2)
PNF: Physical Network Function
VNF: Virtual Network Function

Figure A-4. Complex client-server relationships enabled by the recursion in the SDN control plane, adapted from [5].

Another key functional aspect that makes SDN architecture ideal to embrace 5G slicing is recursion. Because of the different abstraction layers that the recursion principle enables, the SDN control plane can involve multiple hierarchically arranged controllers that extend the client-server relationships at several levels (see Figure A-4). According to these premises, it is evident that SDN can support a recursive composition of slices [5]. This implies that the resources (i.e., Resource Group) a given controller delivers to one of its clients in the form of a dedicated slice (i.e., client context) can, in turn, be virtualized and orchestrated by such client in case of being an SDN controller. This way, the new controller can utilize the resource(s) it accesses via its server context(s) to define, scale and deliver new resources (and hence new slices) to its own clients, which might also be SDN controllers.

4 NFV Reference Architectural Framework

Although the SDN architecture described in Section 3 gives a comprehensive view of the control plane functionalities enabling slicing, it lacks capabilities that are vital to efficiently manage the lifecycle of network slices and its constituent resources. In this respect, the NFV architecture [6] is ideal to play this role, as it manages the infrastructure resources and orchestrates the allocation of such resources needed to realize VNFs and network services.

To take benefit from the management and orchestration functionalities from NFV, appropriate cooperation between SDN and NFV is required. However, embracing SDN and NFV architectures into a common reference framework is not an easy task [7]-[8] .In this section, we briefly describe the tentative framework that ETSI presents in [8] to integrate SDN within the reference NFV architecture. This framework incorporates two SDN controllers, one logically placed at the tenant and another at the InP level. We commence providing a brief overview of the NFV architectural framework, and later describe the integration of the two SDN controllers (see Figure A-5).

The NFV architecture comprises the following entities:

- **Network Functions Virtualization Infrastructure (NFVI)**: a collection of resources used to host and connect the VNFs. While the broad scope of SDN makes resource a generic concept (see Section 3), the current resource definition in the NFV framework comprises only the infrastructure resources.
- **VNFs:** software-based implementation of NFs which run over the NFVI.
- **Management and Orchestration (MANO**): performs all the virtualization-specific management, coordination and automation tasks in the NFV architecture. The MANO framework [9] comprises three functional blocks:
  - *Virtualized Infrastructure Manager (VI*M): responsible for controlling and managing the NFVI resources.
  - *VNF Manager (VNF*M): performs configuration and lifecycle management of the VNF(s) on its domain.
  - *Orchestrat*or: according to ETSI, it has two sets of functions performed by Resource Orchestrator (RO) and Network Service Orchestrator (NSO) respectively. RO orchestrates the NFVI resources across (potentially different) VIMs. NSO performs the lifecycle management of network services, using the capabilities provided by the RO and the (potentially different) VNFMs.
- **Network Management System (NMS):** framework performing the general network management tasks. Although its functions are orthogonal to those defined in MANO, NMS is expected to interact with MANO entities by means of a clear separation of roles [9]. NMS comprises:
  - *Element Management (EM):* anchor point responsible for the FCAPS (Fault, Configuration, Accounting, Performance, and Security) of a VNF.
  - *Operation/Business Support System (OSS/*BSS): a collection of systems and management applications that network service providers use to provision and operate their network services. In terms of the roles, we consider in Section 2, tenants would run these applications.

ETSI proposal includes two SDN controllers in the architecture. Each controller centralizes the control plane functionalities and provides an abstract view of all the connectivity-related components it manages. These controllers are:

- **Infrastructure SDN controller (IC):** it sets up and manages the underlying networking resources to provide the required connectivity for communicating the VNFs (and its components [10]). Managed by the VIM, this controller may change infrastructure behavior on-demand according to VIM specifications, adapted from tenant requests.
- **Tenant SDN controller (TC**): instantiated in the tenant domain [11] as one of the VNFs or as part of the NMS, this second controller dynamically manages the pertinent VNFs used to realize the tenant's network service(s). These VNFs are the

underlying forwarding plane resources of the TC. The operation and management tasks that the TC carries out are triggered by the applications running on top of it, e.g., the OSS.

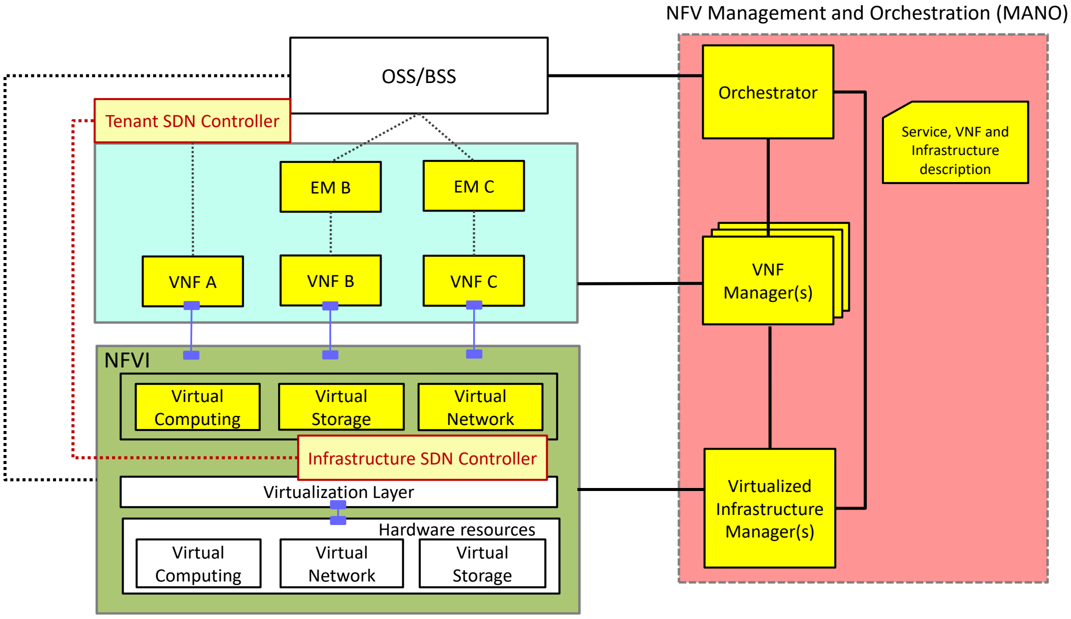
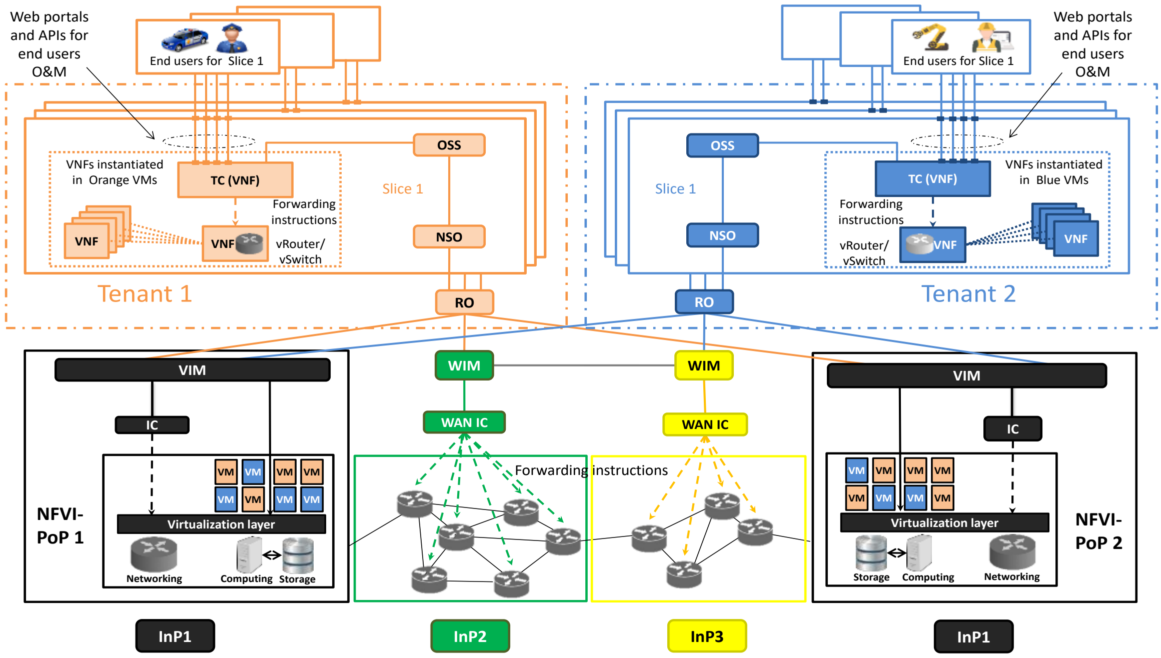


Figure A-5. Integrating SDN controllers into the reference NFV architectural framework at the two levels required to achieve slicing.

Both controllers manage and control their underlying resources via programmable southbound interfaces, implementing protocols like OpenFlow, NETCONF or I2RS. However, each controller provides a different level of abstraction. While the IC provides an underlay to support the deployment and connectivity of VNFs, the TC provides an overlay comprising tenant VNFs that, properly composed, define the network service(s) such tenant independently manages on its slice(s). These different resource views each controller offers through its interfaces have repercussions on the way they operate. On one side, the IC is not aware of the number of slices that utilize the VNFs it connects, nor the tenant(s) which operates such slices. On the other side, for the TC the network is abstracted in terms of VNFs, without notions of how those VNFs are physically deployed. Despite their different abstraction levels, both controllers have to coordinate and synchronize their actions [8]. Note that the service and tenant concept mentioned here can be extended to higher abstraction layers by simply applying the recursion principle, as shown in Figure A-2.

5 Network Slicing use case with SDN-NFV Integration

In this section, we describe an SDN-enabled NFV deployment example that illustrates the network slicing concept, with several slices running on a common NFVI (see Figure A-6). This deployment includes two tenants, each managing a particular set of slices. In the example, we only consider a single level of recursion, and thus the tenants directly serve the end users. Each slice consists of VNFs that are appropriately composed and chained to support and build up the network service(s) the slice (and thus the tenant) delivers to its users. Note that the deployment includes two distinct phases. First, a slice creation phase, in which an end user requests a slice from a network slice catalog, and then the tenant instantiates the slice. Next, a run-time phase, where the different functional blocks within each slice have already been created and are now operative. For simplicity, in Figure A-6 we only depict the run-time phase.



IC: Infrastructure SDN controller
TC: Tenant SDN Controller

Figure A-6. Network slicing deployment in a common framework, integrating both SDN and NFV

The example considers that the tenants access NFVI resources from three InPs. InP1 provides compute and networking resources, both deployed on two NFVI-Points of Presence (NFVI-PoPs) [12] in the form of DCs. InP2 and InP3 provide SDN-based WAN transport networks, used to communicate such NFVI-PoPs. The VMs and their underlying hardware, instantiated in the NFVI-PoPs and in charge of hosting VNFs (and their components), are directly managed by the VIMs. The networking resources, supporting VM (and hence VNF) connectivity at the infrastructure level, are programmatically managed by the ICs following the VIM and the WAN infrastructure manager (WIM) premises. Both VIMs and WIMs act as SDN applications, delegating the tasks related to the management of networking resources to their underlying ICs. Although in this example the ICs are deployed on the NFVI, it would be possible to integrate them into their corresponding VIMs, as [8] suggests.

On top of the InPs, the tenants independently manage a set of network slices. Each slice comprises an OSS, a TC, and an NSO. The OSS, an SDN application from the TC's perspective, instructs the controller to manage slice's constituent VNFs and logically compose them to efficiently realize the network service(s) the slice offers. The lifecycle of such network service(s) is managed by the NSO, which interacts with the TC via the OSS. The TC, deployed as a VNF, relies on the capabilities provided by virtual switches/routers (in the form of VNFs as well) to enable the VNF composition, forwarding pertinent instructions to such virtual switches/routers via its southbound interfaces. Through its northbound interfaces, the TC provides a means to securely expose selected network service capabilities to end users. Such interfaces allow end users to retrieve context information (e.g., real-time performance and fault information, user policies, etc.), operate, manage and make use of the slice's network service(s), always within the limits set by the tenant. The fact that each slice is provided with its own NSO, OSS and TC instances enables the required management isolation.

Each tenant must efficiently orchestrate their assigned resources to simultaneously satisfy the diverging requirements of the slices that are under its management. The RO is the functional block that performs such task on behalf of the tenant, providing each slice with the required resources via interfaces with each slice's NSO. The RO must perform the resource sharing among slices while fulfilling their required performance, following an adequate, effective resource management framework that must comply with both tenant and slice-specific policies. Such a framework is required so that the RO enables performance isolation among slices.

All the NFVI resources available for use by a tenant (i.e., those that RO orchestrates) are supplied by the different InPs. Each InP rents part of the virtual resources according to a business lease agreement that both InP and tenant had previously signed. To access, reserve and request such resources, the tenant's RO interacts with the VIM(s) /WIM(s) by means of interfaces that those functional blocks expose and that tenant's RO consumes. Indeed, we assume that VIMs and WIMs support multi-tenancy. We also assume that WIMs can communicate with each other according to predefined business agreements. In this respect, the interaction between a WIM and an RO might be achieved indirectly through another WIM.

As Figure A-6 suggests, the resource management must be performed at two levels: at the infrastructure level, where a slice-agnostic VIM/WIM provides the subscribed tenants with (virtualized) infrastructure resources, and at the tenant level, where the RO delivers its assigned resources to the corresponding slices. Both the VIM(s)/WIM(s) and the RO have to collect accurate resource usage information (each at its domain) and in turn to forecast resource availability in relatively short timescales to satisfy tenant and slice demands, respectively.

Please note that, with the exception of hardware resources, the functional blocks (e.g., VIM, RO, NSO, SDN controllers, etc.) are modeled as independent software components. The need for separate access, configuration and management suggests this modeling, wherein the software relationships are enabled with the help of the APIs that each component provides.

To preserve security and privacy isolation among slices, it is required to apply the compartmentalization principle at each virtualization level. In addition, each functional block and manageable resource (e.g., VNF) within a given slice must have its own security mechanisms, ensuring operation within expected parameters, and preventing access to unauthorized entities. This is intended to guarantee that faults or attacks occurred in one slice are confined to such slice, preventing their propagation across slice boundaries.

Additionally, although recursion has not been addressed in this example, it is readily applicable to this scenario by simply assuming some of the slice's users are tenants which in turn can deploy and operate their own slices.

6 Challenges and Research Directions

In this section, we identify the main challenges and future research arising from implementing slicing in 5G systems.

6.1 Performance issues in a shared infrastructure

When network slices are deployed over a common underlying substrate, the fulfillment of performance isolation requirement is not an easy task. If tenant's RO only assigns dedicated resources to network slices, their required performance levels are always met at the cost of preventing slices to share resources. This leads to over-provisioning, an undesired situation bearing in mind that the tenant has a finite set of assigned resources. One way to resolve this issue is to permit resource sharing (see e.g., [13]), although this means slices are not yet completely decoupled in terms of performance. Thus, it is required to design adequate resource management mechanisms that enable resource sharing among slices when necessary, without violating their required performance levels. To accomplish the sharing issue, the RO could use policies and strategies similar to those used in VIMs (such as the OpenStack Congress module, or Enhanced Platform Awareness attributes).

6.2 Management and orchestration issues

Given the dynamism and scalability that slicing brings, management and orchestration in multi-tenant scenarios are not straightforward. To flexibly assign resources on-the-fly to slices, the optimization policy that governs the RO must deal with situations where resource demands vary considerably in relatively short timescales. To accomplish this:

- An appropriate cooperation between slice-specific management functional blocks and RO is required.
- Policies need to be captured in a way that they can be automatically validated. This automation enables both the RO and slice-specific functional blocks to be authorized to perform the corresponding management and configuration actions in a timely manner.
- It is required to design computationally efficient resource allocation algorithms and conflict resolution mechanisms at each abstraction layer.

6.3 Security and privacy

The open interfaces that support the programmability of the network bring new potential attacks to softwarized networks. This calls for a consistent multi-level security framework composed of policies and mechanisms for software integrity, remote attestation, dynamic threat detection and mitigation, user authentication and accounting management. The security and privacy concerns arising from 5G slicing (see [14]) are today a major barrier to adopt multi-tenancy approaches.

6.4 New business models

The innovative partnerships between several players, each providing services at different positions of the value chain, and the integration of new tenants such as verticals, OTT service providers, and high-value enterprises, empowers promising business models. Given this business-oriented approach, new transition strategies must be broadly analyzed, allowing for a gradual evolution to future 5G networks and ensuring compatibility with past infrastructure investments. To accomplish this, a deep review of the telecom regulatory framework has to be made. Innovative ways of pricing, new grounds for cost sharing and standardized solutions, which provide the required support for interoperability in multi-vendor and multi-technology environments, must be studied as well.

Acknowledgement

This work is partially supported by the Spanish Ministry of Economy and Competitiveness, the European Regional Development Fund (Project TIN2013-46223-P), the Ministry of Education, Culture, and Sport of the Government of Spain, under grant Beca de Colaboración (2015-2016), and the University of Granada, under grant Beca de Iniciación a la Investigación (2016-2017).

References

[1] 5G-PPP, ERTICO, EFFRA, EUTC, NEM, CONTINUA and Networld2020 ETP, "5G empowering vertical industries", White Paper, Feb. 2016.

[2] M. Pearce, S. Zeadally, and R. Hunt, "Virtualization: Issues, security threats, and solutions", *ACM Computer Surveys (CSUR)*, vol.45, no.2, Feb. 2013, pp. 1-38.

[3] N. M. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization", Computer Networks, vol. 54, no.5, Apr. 2010, pp. 862–876

[4] ONF TR-521, "SDN Architecture", Feb. 2016.

[5] ONF TR-526, "Applying SDN Architecture to 5G Slicing", Apr. 2016

[6] ETSI GS NFV 002, "Network Functions Virtualization (NFV); Architectural Framework", V1.1.1, Dec. 2014.

[7] ONF TR-518, "Relationship of SDN and NFV", Oct. 2015

[8] ETSI GS NFV-EVE 005, "Network Functions Virtualisation (NFV); Ecosystem; Report on SDN Usage in NFV Architectural Framework", V1.1.1, Dec. 2015.

[9] ETSI GS NFV-MAN 001, "Network Functions Virtualisation (NFV); Management and Orchestration", V1.1.1, Dec. 2014.

[10] ETSI GS NFV-INF 001, "Network Functions Virtualisation (NFV); Infrastructure Overview", V1.1.1, Jan. 2015

[11] ETSI GS NFV-SEC 003, "Network Functions Virtualisation (NFV); NFV Security; Security and Trust Guidance", V1.1.1, Dec. 2014.

[12] ETSI GS NFV 003, "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV", V1.2.1, Dec. 2014

[13] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. Ramos-Munoz, and J. Lopez-Soler, "Virtualized MME Design for IoT Support in 5G Systems", Sensors, vol. 16, no. 8, Aug. 2016, pp. 1338-1362

[14] R. Harel, and S. Babbage, "5G security recommendations Package #2: Network Slicing", NGMN Alliance, Apr. 2016

Paper B

The Creation Phase in Network Slicing: From a Service Order to an Operative Network Slice

Jose Ordonez-Lucena, Oscar Adamuz-Hinojosa, Pablo Ameigeiras, Pablo Muñoz, Juan J. Ramos-Muñoz, Jesús Folgueira and Diego López

Disclaimer:
This work has been published in the European Conference on Networks and Communications (EuCNC), 2018.
Citation information: DOI 10.1109/EuCNC.2018.8443255.

Copyright:

Abstract

Network slicing is considered a key mechanism to serve the multitude of tenants (e.g., vertical industries) targeted by forthcoming 5G systems in a flexible and cost-efficient manner. In this paper, we present a SDN/NFV architecture with multi-tenancy support. This architecture enables a network slice provider to deploy network slice instances for multiple tenants on-the-fly, and simultaneously provision them with isolation guarantees. Following the Network Slice as-a-Service delivery model, a tenant may access a Service Catalog, selecting the slice that best fits its needs and ordering its deployment. This work provides a detailed view on the stages that a network slice provider must follow to deploy the ordered network slice instance, accommodating it into a multi-domain infrastructure, and putting it operative for tenant's consumption. These stages address critical issues identified in the literature, including (i) the mapping from high-level service requirements to network functions and infrastructure requirements, (ii) the admission control, and (iii) the specific information a network slice descriptor should have. With the proposed architecture and the recommended set of stages, network slice providers can deploy (and later operate) slice instances with great agility, flexibility, and full automation.

1 Introduction

The ongoing digital transformation is geared towards the integration of vertical industries into an ecosystem boosting technical and business innovation. This may bring a multitude of new vertical-driven use cases and application scenarios, with very distinct requirements. Current research efforts focus on finding ways to accommodate them on the same infrastructure in a flexible, agile, and cost-efficient manner. Network slicing will be key for this end. Leveraging network softwarization technologies such as Software Defined Networking (SDN) and Network Functions Virtualization (NFV), network slicing aims to logically split an infrastructure into a set of self-contained programmable network instances, each customized to only serve the particular needs of a given use case. The shared and multi-domain nature of the infrastructure on top of which these Network Slice (NSL) instances run makes isolation a capital requirement for network slicing.

Network slicing has brought the attention of the research community. Many standardization bodies and Fora have addressed this concept, including NGMN, IETF, ONF, and 3GPP. In [1], ETSI NFV provides an insight into the different views that some of these organizations have about slicing, analyzing how their visions match with the NFV constructs.

Network slicing is claimed to unlock new business opportunities, with flexible service delivery models. One of them is Network Slicing as-a-Service [2].This service delivery model enables an NSL provider (e.g., network operator) to deploy customized NSL instances for their clients (e.g., verticals) on request, and deliver them as a service. These clients, taking the role of NSL tenants, may in turn use the purchased NSL instances to deploy their business services for their own clients. This empowers recursive business models (e.g., Business-to-Business-to-X models), with multiple actors providing services at different positions in the value chain.

In our previous work [3], we proposed an SDN/NFV-based architecture enabling operation of NSL instances with recursiveness, multi-tenancy and multi-domain support. Although these issues have been addressed in architectural solutions proposed in different 5G-PPP projects (e.g., 5G-Crosshaul, 5GNORMA, 5GEx, etc.), none of them consider the isolation as the first criteria for architecture design. This has led to solutions that do not address all the isolation properties necessary in slicing: *performance, security, privacy, and management isolation.* Unlike those proposals, our solution satisfies each of these isolation properties while being compliant with ETSI NFV information model. For this end, two architectural enhancements are considered with respect to the NFV framework [4]: the decomposition of the NFV Orchestrator (NFVO) into resource and network service orchestration blocks, and the inclusion of a Tenant SDN Controller. The results derived from this work have contributed to ongoing standardization efforts, including those conducted by ETSI NFV [1] and IETF [5].

The vision given in [3] focused on the *run-time phase,* considering the NSL instances were operative and leased out to their tenants. However, the *creation phase* was omitted. In this phase, a tenant requests an NSL from a catalog, and orders its instantiation. The creation phase brings new challenges, including the translation of tenant-specific service requirements into network functions and infrastructure requirements, the specification of an NSL descriptor, and the admission control. These and other aspects have been identified in [6] as still open issues in the context of network slicing. Addressing them is thus essential to make a complete network slicing solution.

In this paper, we concentrate on the creation phase of network slicing, complementing the run-time phase addressed in our previous work. The main objective is to provide an insight into the procedures and mechanisms required to make the deployment of NSLs more flexible, agile, and automated from the perspective of both the NSL provider and the tenant. To incorporate these mechanisms and procedures, we extend our SDN/NFV-based architecture with two new functional blocks: the NSL Manager and the NSL Orchestrator. In the context of this architecture, we identify the stages the NSL provider shall follow for completing a catalog-driven NSL deployment. In each stage, we specify the input/output information, the steps involved, and the role that each functional block plays.

The remainder of this article is as follows. Section 2 shows how the concept of ETSI NFV network service is key to provide a resource-centric view of an NSL. Section 3 describes the slicing architecture, with focus on the new functional blocks. Section 4 provides a detailed view on the creation phase, on a step-by-step basis. Finally, Section 5 summarizes the main conclusions of this work.

2 NFV Network Services and Network Slices

The concept of Network Service (NS) introduced by ETSI NFV is key for network slicing. NSLs leverage the capabilities offered by NSs to satisfy the network requirements of the use cases they accommodate. From a resource-centric viewpoint, an NSL instance may be composed of one or more NS instances. Particularly, three scenarios can be considered:

a) The NSL instance consists of an instance of a simple NS.
b) The NSL instance consists of an instance of a composite NS.
c) The NSL instance consists of a concatenation of a simple and/or composite NS instances.

A simple NS includes one or more Virtualized Network Functions (VNFs), and virtual links providing connectivity between them. In search of modularity and recursiveness, the NFV framework provides the ability to include in the design of an NS one or more nested NSs. The result is a composite NS (see Figure B-1).

According to ETSI NFV, an NS instance is deployed from an NS descriptor. An NS descriptor is a deployment template used for creating and operating instances of an NS. The

NS descriptor provides a list of pointers to the VNF descriptors of the constituent VNFs, and additional information on connectivity between them. In case of a composite NS, the corresponding NS descriptor also references the NS descriptor(s) of the nested NS(s).

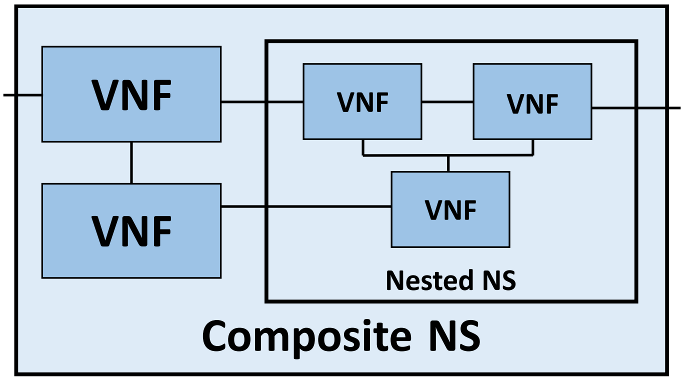


Figure B-1. An example of a composite NS. This NS consists of two VNFs and one simple NS.

A key mechanism in the NS descriptor is NS flavoring. NS flavoring enables customizing the deployment of an NS instance, in terms of *functionality* and *performance*. As stated in [7], an NS descriptor consists of one or more NS flavors, each specifying a different deployment configuration for the NS. Selecting an NS flavor within the NS descriptor enables selecting the VNFs and virtual links to be deployed as part of the NS, and hence the features to be activated for that NS.

A given NS flavor includes one or more NS Instantiation Levels (NS-ILs), each specifying a possible option of instantiating the NS using this flavor. An NS instance resulting from a NS-IL can only include instances of those VNFs and virtual links that have been declared in the flavor. The goal of a NS-IL is to describe how to deploy each constituent VNF and virtual link. To that end, an NSL-IL contains the following:

- For each VNF to be used for the NS instance, the NS-IL specifies the number of instances to be deployed, their resource levels (i.e., the level of resources to be allocated for each instance), and their applicable affinity/anti-affinity rules. Currently, the reliability requirements of a VNF (e.g., the subset of instances to serve as backup, if high availability hardware/software is required for any instance, etc.) are not part of the NS-IL, although their inclusion is expected for the NFV Release 3 [8]
- For each virtual link to be used for the NS instance, the NS-IL specifies transport reliability and the bitrate requirements.

According to the mentioned ideas, a triplet (NS descriptor ID, NS Flavor ID, NS-IL ID) provides a complete resource-centric description of an NS instance. The second term indicates the subset of VNFs and virtual links to be deployed for the NS, and hence the functionality selected for the NS. The third term specifies how instantiating each of those VNFs and links, thus setting the level of performance of the NS.

As seen, NS flavoring is key for slicing, as it enables selecting only the needed capabilities within an NS for a given NSL. To provide a complete resource-centric description of an NSL instance, it is required to specify which triplet is used to instantiate each constituent NS. For this end, we introduce the concept of NSL Instantiation Level (NSL-IL). The NSL-IL is an information element that provides a (list of) pointer(s) to the triplet(s) of the constituent NS instance(s). This means that if an NSL instance has “*M”* NS instances - see scenario c) -, then the NSL-IL will refer to the “*M* “triplets used for their instantiation.

3 Network Slicing Architecture

In this section, we describe an SDN/NFV based architecture for network slicing that extends our previous proposal [3].Note that this architecture focuses on the transport and core network domains, omitting the RAN domain for simplicity.

As Figure B-2 shows, this architecture enables an NSL provider to simultaneously operate multiple NSL instances. These instances run on top of a common infrastructure that spans across multiple administrative domains, each belonging to a different infrastructure provider. This infrastructure, consisting of geographically distributed Points of Presence (PoPs) and Wide Area Networks (WANs) connecting them, enables multi-site deployments. To manage the resources of the PoP(s) and/or WAN(s) within its administrative domain, an infrastructure provider leverages the capabilities of a Virtual Infrastructure Manager (VIM) and/or WAN Infrastructure Manager (WIM), respectively.

The NSL provider, taking the role of an infrastructure tenant, rents the infrastructure resources owned by the underlying infrastructure providers, and uses them to provision the NSL instances. For this end, the NSL provider has a resource orchestration functional block. The Resource Orchestrator uses the finite set of resources that are at its disposal (the resources supplied by the underlying VIMs/WIMs), and dispatches them to the NSL instances in an optimal way. This optimization means that all the NSL instances are simultaneously provided with the resources needed to satisfy their (potentially diverging) requirements, while preserving their performance isolation. The resource requirements of each NSL instance are stated by its NSL-IL (see Section 2).

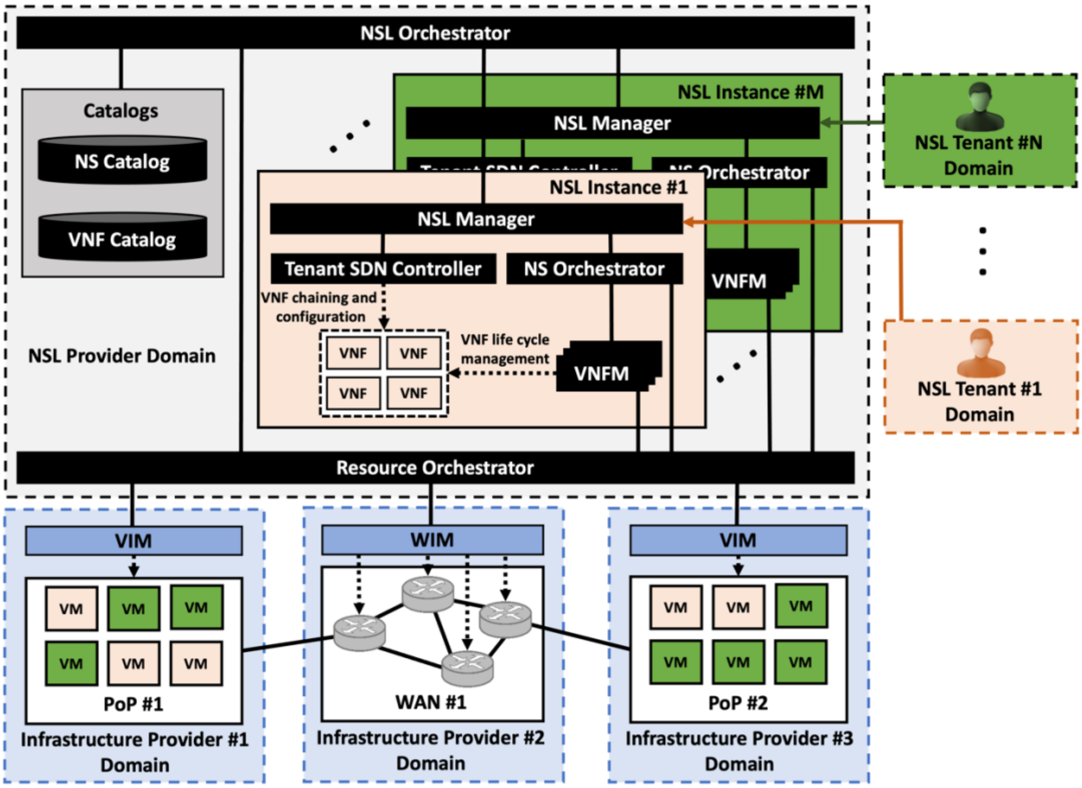


Figure B-2. SDN/NFV-based Network Slicing Architecture.

To preserve management isolation across NSL instances, each instance has its own management plane. This plane consists of four functional blocks: VNF Manager (VNFM), NS Orchestrator, Tenant SDN Controller, and NSL Manager.

The VNFM(s) and the NS Orchestrator perform the required life cycle operations (e.g., instantiation, scaling, termination, etc.) over the instances of the VNFs and NS(s), respectively. Since these operations involve modifying the amount of resources to be allocated for those instances, an interplay between these functional blocks and the Resource Orchestrator is required. The Tenant SDN Controller performs VNF configuration and chaining in a programmatic manner. On one hand, this SDN Controller configures the VNF instances at application level, taking the role of an Element Manager (EM) [4].On the other hand, it chains the VNF instances for NS construction, leveraging the forwarding capabilities provided by the data plane. Finally, the NSL Manager coordinates the operations and management data from both the Tenant SDN Controller and the NS Orchestrator, performing the fault, configuration, accounting, performance, and security management within the NSL instance. Additionally, it provides visibility and management capability exposure to external blocks. In this respect, note that the NSL Manager is of key importance for an NSL tenant. Each tenant consumes its NSL instance and operates it at its convenience (within the limits agreed with the NSL provider) through the NSL Manager. By way of example, the tenant could use an SDN application in the NSL manager to programmatically modify the VNF chaining rules on-the-fly, according to its needs.

Beyond the domain of an NSL instance, the NSL provider defines an NSL Orchestrator. This functional block plays a key role in the creation phase and the run-time phase. In the creation phase, it receives the order to deploy a NSL instance for a tenant, checks the feasibility of the order, and if feasible, triggers the instantiation of the NSL. For this end, it interacts with the Resource Orchestrator, and accesses the VNF and NS Catalogs. These catalogs contain VNF and NS descriptors, exposing the capabilities of all the VNFs and NSs that an NSL provider can select for the NSLs. At run-time, the NSL Orchestrator performs policy-based inter-slice operations. Particularly, it analyses the performance and fault management data received from the operative NSL instances to manage their Service Level Agreements. In case of Service Level Agreement violations, then the NSL Orchestrator decides which NSL instances need to be modified, and sends corrective management actions (e.g., scaling, healing, etc.) to their NSL Managers.

The interplay among the functional blocks described so far enables slicing. Abstraction is a key architectural principle for this end. Having different abstraction levels across functional blocks logically placed at different layers leads to a loosely coupled architecture. Each functional block is only responsible for a specific set of tasks, being they limited by the level of information the functional block understands.

In our architecture, the VIM/WIM, the Resource Orchestrator, and the NSL/NS Orchestrator operate at different layers, and hence provide different abstraction levels. The Resource Orchestrator maintains a PoP resource map derived from the information provided by VIM(s) and WIM(s), including data on geolocation, capabilities[1] and resource state. The Resource Orchestrator abstracts this information to the NSL/NS Orchestrator, providing a resource-agnostic view of the set of reachable PoPs. This view only includes high-level information on the locations and capabilities of those PoPs, without any information on their resources, nor the VIM(s) responsible for their management.

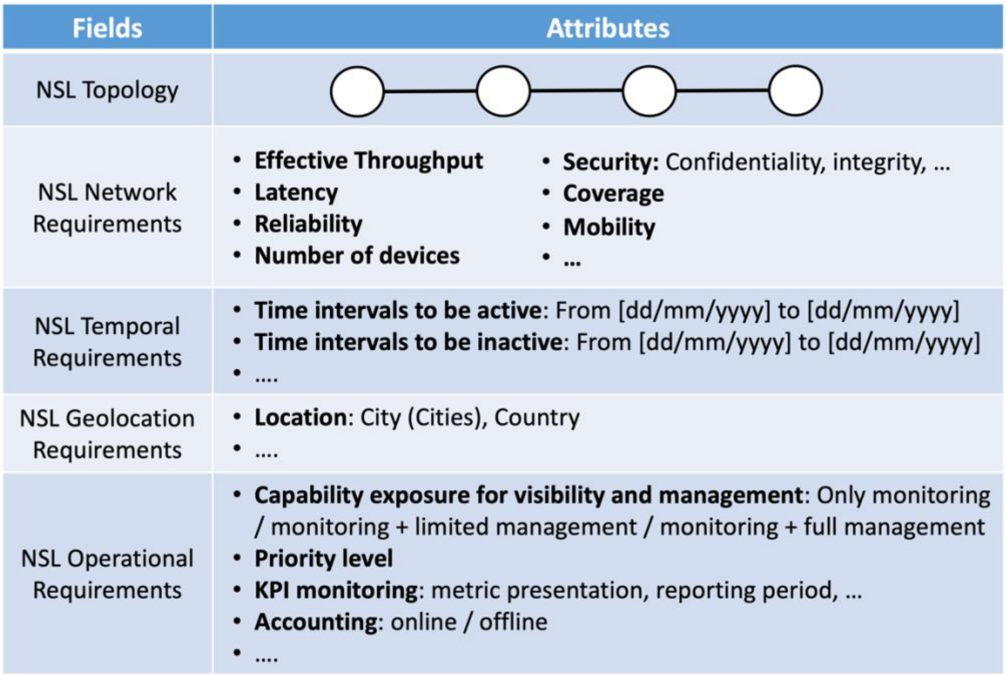
4 Network Slice Creation Phase

[1] The capabilities of a PoP depend on the PoP setup (e.g., setup for high availability and fault resiliency, setup for high I/O processing, etc.).

Section 3 focuses on the run-time phase of the network slicing concept, considering that the NSL instances are operative and being consumed by their tenants. However, prior to this phase, the creation phase occurs. This section concentrates on the creation phase, providing a detailed view on the steps the NSL provider must follow to instantiate a NSL according to the specificities gathered in a catalog-driven service order. For better understandability, these steps have been grouped into five well-defined stages. These stages are described below.

4.1 Service Ordering

The NSL provider defines a business-driven Service Catalog that contains a finite set of service templates, each describing a different service offering. These offerings include NSLs optimized to serve a multitude of usage scenarios, ranging from typical 5G services (e.g. eMBB, mMTC, and uRLLC) to vertical-specific applications (e.g. smart factory, remote surgery, connected cars, etc.). A service template is a readymade document that contains all the information that is required to drive the deployment of an NSL. In particular, it contains (1) the NSL topology, expressed as an ordered chain of technology-agnostic composable nodes, each providing specific functionality; (2) the NSL network requirements, including performance and functional requirements; (3) the NSL temporal requirements; (4) the NSL geolocation requirements; and (5) the NSL operational requirements. An example of a service template is shown in Figure B-2.



Fields	Attributes
NSL Topology	○—○—○—○
NSL Network Requirements	• **Effective Throughput** • **Latency** • **Reliability** • **Number of devices** • **Security:** Confidentiality, integrity, … • **Coverage** • **Mobility** • …
NSL Temporal Requirements	• **Time intervals to be active**: From [dd/mm/yyyy] to [dd/mm/yyyy] • **Time intervals to be inactive**: From [dd/mm/yyyy] to [dd/mm/yyyy] • ….
NSL Geolocation Requirements	• **Location**: City (Cities), Country • ….
NSL Operational Requirements	• **Capability exposure for visibility and management**: Only monitoring / monitoring + limited management / monitoring + full management • **Priority level** • **KPI monitoring**: metric presentation, reporting period, … • **Accounting**: online / offline • ….

Figure B-2. Service template structure. The nodes included in the topology depend on the use case the template is designed for (e.g., in an eMBB NSL, some nodes could be a cache, the EPC user plane, and the EPC control plane). Note that the value of some NSL requirements could be specified by the tenant, according to the NSL provider's policies. For typical values in different vertical-driven use cases, please see [8].

To facilitate the customization and automate the service definition, the NSL provider may suggest typical configurations of certain attributes, allowing tenants to focus on the key areas of the service template. The number and diversity of attributes that can be specified (including their allowed value ranges) by the tenant is up to the NSL provider's policies.

To order an NSL, the tenant makes use of the self-ordering APIs that the NSL provider exposes in a self-service Web Portal. With these APIs, the tenant gains access to the Service Catalog, from which it selects the service template that best matches its needs. Then, the tenant specifies the desired values for the attributes it can customize, according to the NSL provider's policies. The result is a catalog-driven NSL service order that the NSL Orchestrator must process. This order contains information mappable to RAN, transport, and core network domains. For simplicity, we focus on the latter two.

4.2 Network Slice Resource Description

The goal of this stage is to give a resource-centric view of the ordered NSL, expressed through an NSL-IL (see Section 2). This NSL-IL may be used to decide if the ordered NSL is feasible/infeasible from a resource viewpoint, and hence accepted/rejected for deployment (see subsection 4.3).

Upon receiving the service order, the NSL Orchestrator extracts the content that is relevant from a resource viewpoint: the NSL topology, and the NSL network requirements (i.e., performance and functional requirements). Using this information, the NSL Orchestrator constructs an NSL-IL for the NSL instance. For this end, it performs three steps.

In the first step, the NSL Orchestrator uses the NSL topology to identify which NS(s) need to be deployed for the NSL, retrieving the corresponding NS descriptor(s) from the NS Catalog. In the second step, the NSL Orchestrator selects within each descriptor the deployment option that best matches the features and the performance level required for the NSL. In other words, it selects the triplet (NS descriptor ID, NS Flavor ID, NS-IL ID) to be used to instantiate each NS. Finally, the NSL Orchestrator constructs the NSL-IL by referencing the selected triplet(s).

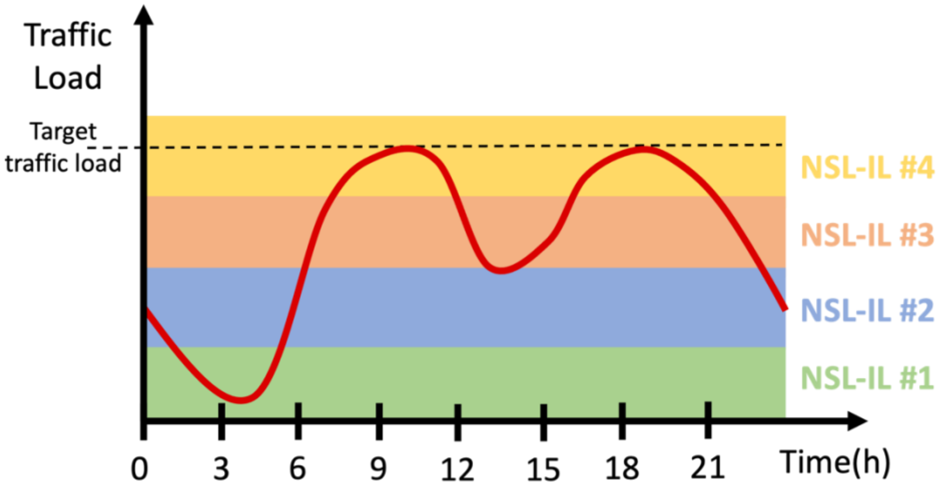


Figure B-4. Example of the traffic load expected for a given NSL instance during a typical day. NSL-IL #4 is the target NSL-IL, and the rest are the optional NSL-ILs. The entire set of NSL-ILs enables the NSL provider to adjust the level of resources within the NSL instance at run-time, in such a way it satisfies the desired performance, while making an efficient resource usage.

With the mentioned approach, the constructed NSL-IL meets the specified network requirements of the NSL instance, and hence is able to accommodate the target traffic load. From here on out, we will refer to this NSL-IL as the *target NSL-IL*. However, traffic fluctuations may occur throughout the lifetime of the NSL instance, resulting in periods of time where the traffic load is considerably lower than the target one. In this kind of situations, the triplet(s) used for the target NSL-IL may lead to a waste of resources. To solve this issue and take advantage of multiplexing gains, the NSL Orchestrator could make use of less resource-demanding triplets to accommodate lower traffic loads, and construct *optional NSL-ILs* with them (see Figure B-4). The number of optional NSL-ILs and the triplet(s) selected for each of them depend on the traffic fluctuations expected for the NSL instance. To estimate these fluctuations, the NSL Orchestrator may rely on traffic models that the NSL provider has inferred from historical data.

The target NSL-IL, along with the optional NSL-ILs, define the complete set of NSL-ILs among which the NSL instance can scale up/down during its entire life cycle.

4.3 Admission Control

The target NSL-IL specifies the resource requirements fitting the tenant's demands. Once derived, the NSL provider can perform the admission control. The admission control aims to check if the NSL provider can satisfy the resource, geolocation, and temporal requirements of the ordered NSL. For this end, the following information is needed:

1) The resource requirements of the NSL instance. This includes(a) the resources to be allocated for each VNF instance and virtual link, (b) the affinity/anti-affinity rules applicable between VNF instances, and (c) the reliability requirements for each VNF instance and virtual link.
2) The geographical region(s) where each VNF is needed.
3) The time intervals when the NSL instance needs to be active (operative).
4) Information of the PoPs (and the WAN network(s) connecting them) to which the NSL provider is subscribed.

The information shown in (1)-(3) is available to the NSL Orchestrator; indeed, (1) is part of the target NSL-IL, while (2)-(3) are derived from the geolocation and temporal requirements specified in the service order. The information specified in (4) is available to the Resource Orchestrator, and provided by the underlying VIM(s)/WIM(s). The fact that the NSL Orchestrator and the Resource Orchestrator operate at different abstraction levels means that they deal with different level of information, and hence none of them is able to perform the admission control at its own. The interplay of both functional blocks is needed. Following this idea, the admission control can be split into three steps. The NSL Orchestrator performs the first two steps, the latter being carried out by the Resource Orchestrator.

In the first step, the NSL Orchestrator calculates which PoP(s) is (are) candidate to host each VNF instance. A PoP is candidate for a VNF instance if the location and capabilities of the PoP satisfy the geolocation and reliability requirements of that instance. For this step, the NSL Orchestrator takes as input the information specified in (1c) and (2), and the resource-agnostic view provided by the Resource Orchestrator. As seen in Section 3, this view consists of high-level information of the location and capabilities of the reachable PoPs.

In the second step, the NSL Orchestrator sends two kinds of data to the Resource Orchestrator. On one hand, data concerning the NSL lifetime. For this end, the NSL Orchestrator takes the information shown in (3), and passes it down to the Resource Orchestrator. On the other hand, data concerning the target NSL-IL to be accommodated. For that, the NSL Orchestrator takes the resource requirements specified in (1), along with the candidate PoPs calculated in the first step, and passes then down to the Resource Orchestrator at VNF/virtual link level. For each VNF instance, the NSL Orchestrator communicates the candidate PoP(s), and the requirements shown in (1a) and (1b). For each virtual link, the NSL Orchestrator communicates the requirements specified in (1a) and (1c).

In the third step, the Resource Orchestrator seeks feasible solutions to deploy the target NSL-IL. A solution is feasible as long as each VNF instance can be allocated in a candidate PoP during the time interval(s) in which the NSL instance needs to be active, while satisfying the VNF affinity/anti-affinity rules and connectivity needs. For this step, the Resource Orchestrator takes the data received from the NSL Orchestrator, and compares it against the information specified in (4).

If there exists one feasible solution, the admission control is successful. In this case, the Service Level Agreement between the NSL provider and the tenant can be formalized; otherwise, these two parties shall re-negotiate the content of the service order.

4.4 Optimization and Resource Reservation

A successful admission control may derive multiple feasible solutions for the target NSL-IL (e.g. multiple PoPs can accommodate a given VNF instance). However, only one of them must be eventually selected for deployment. To solve this issue, the Resource Orchestrator may run an algorithm that calculates the optimal solution. Examples of optimality criteria that could be used for this algorithm include minimize resource usage, minimize energy consumption, etc.

Once the optimal solution is found, the Resource Orchestrator may proceed with resource reservation. The Resource Orchestrator sends resource reservation requests towards the underlying VIM(s)/WIM(s). The hard and soft nature of this reservation depends on the NSL provider's policies, as well as the nature of the use case the NSL instance will accommodate.

4.5 Network Slice Preparation

The NSL preparation is the last stage prior to put the NSL operative. It consists of setting up all that is required to manage the NSL instance throughout its entire life cycle, from commissioning (instantiation, configuration, and activation) to decommissioning (de-activation and termination) [9].This includes (1) preparing the network environment, and (2) designing and on-boarding the NSL descriptor.

In the network environment preparation, the NSL Orchestrator performs the following tasks:

- **It negotiates with the Resource Orchestrator a priority level for the NSL instance**. Having different priority levels allows the Resource Orchestrator to define a priority order between the NSL instances in case they compete for the same resources, or in case of resource scarcity. The geographical region(s) where each VNF is needed.
- **It prepares the management plane of the NSL instance**. First, the NSL Orchestrator instantiates the NSL Manager, the Tenant SDN Controller, the NS Orchestrator, and the VNFM(s). Then, it configures these functional blocks in an appropriate manner, making them ready for the run-time phase. By means of example, the NSL Orchestrator configures the NSL Manager in such a way it provides the tenant only with the visibility and management capabilities specified in

the service order.

In parallel to the network environment preparation, the NSL Orchestrator builds up the NSL descriptor. The NSL descriptor is a deployment template used by the NSL Manager to operate the NSL instance during its life cycle in an agile, automated fashion. This descriptor includes the following parts:

- **A set of policy-based workflows.** These workflows enable the NSL Manager to enforce the expected behavior of the NSL instance during its life cycle, in a timely manner. The NSL Manager translates the content of these workflows into appropriate NS and VNF management actions, and forwards them to the **NS Orchestrator and to the tenant SDN controller for their enforcement.**
- **The set of NSL-ILs available for use,** constructed in the Network Slice Resource Description phase (see subsection 4.2). The NS Orchestrator uses the triplets referenced by these NSL-ILs to scale the NS instance(s) at run-time, according to time-varying traffic demands.
- **VNF configuration primitives at application level, and VNF chaining management instructions.** Both are used by the Tenant SDN Controller to programmatically configure and chain the VNF instance(s).
- **Information about management data,** used for performance management (e.g., metrics to be monitored, metric presentation, reporting period) and fault management (e.g., alarms to be subscribed). Derived from the NSL operational requirements specified in the service order, the management data may be collected from the NS Orchestrator and the Tenant SDN Controller, and used for visibility/manageability purposes.

Note that the policy-based workflows contained in the NSL descriptor enables the NSL manager to automate all the life cycle operations that are manually triggered from the OSS in the ETSI NFV framework [10] making the NSL instance a self-contained entity. The remaining content of the NSL descriptor is used to feed these workflows (e.g., performance metrics may be taken as inputs for the workflows targeted at the NSL scaling operation). Manager to operate

5 Conclusions

An DN/NFV-based network slicing architecture has been presented in this work. This architecture addresses the two phases considered for network slicing: the creation phase and the run-time phase. This work focuses on the former.

We have provided detailed insight into the steps needed to successfully complete catalog-driven NSL deployments. These steps have been arranged into five stages: Service Ordering, Network Slice Resource Description, Admission Control, Optimization & Resource Reservation, and Network Slice Preparation. In each of these stages, the input/output information required, the steps involved, and the role of the participant functional block(s) have been specified.

With the architecture and the ordered set of stages proposed in this work, NSL providers are able to perform cost-efficient deployments, in an agile, flexible, and automated manner. The presence of a Service Catalog, with customizable service offerings that brings flexibility in service definition, and the interplay between the NSL Orchestrator and Resource Orchestrator are crucial for that end. Additionally, the correct design of a NSL descriptor in the creation phase is key for a successful operation in the run-time phase. This descriptor makes the NSL instance a self-contained entity, enabling the slice-specific management plane to operate the NSL instance in a customized way, with great agility, and full automation.

Acknowledgement

This work is partially supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (Project TEC2016-76795-C6-4-R), the Spanish Ministry of Education, Culture and Sport (FPU Grant 16/03354), and the University of Granada, Andalusian Regional Government and European Social Fund under Youth Employment Program.

References

[1] ETSI GS NFV-EVE 012, “Network Functions Virtualization (NFV); Evolution of Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework”, Dec. 2017

[2] X. Zhou et al., “Network slicing as a service: enabling enterprises’ own software-defined cellular networks”, *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146-153, 2016.

[3] J. Ordonez-Lucena et al., “Network Slicing for 5G with SDN/NFV; Concepts, Architectures and Challenges”, *IEEE Communications Magazine*, vol. 55, no. 55, pp. 80-87, 2017.

[4] ETSI GS NFV-MAN 001, “Network Functions Virtualization (NFV); Management and Orchestration”, Dec. 2014

[5] L. Geng et al., “Common Operation and Management on network Slices (COMS) Architecture”, in Internet Engineering Task Force (IETF), March 2018.

[6] X. Foukas et al., “Network Slicing in 5G: Survey and Challenges”, *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94-100, 2017.

[7] ETSI GS NFV-IFA 014, “Network Functions Virtualization (NFV); Management and Orchestration; Network Slice Templates Specification”, Aug. 2017

[8] Next Generation Mobile Networks (NGMN) Alliance, “Perspectives on Vertical Industries and Implications for 5G”, Sept. 2016

[9] 3GPP TS 28.801 V15.1.0, “Telecommunications management; Study on management and orchestration of network slicing for next generation network”, Jan. 2018

[10] ETSI GS NFV-IFA 013, “Network Functions Virtualization (NFV) Release 2; Management and Orchestration; Os-Ma-Nfvo reference point – Interface and Information Model Specification”, Aug. 2017

Part III

Multi-domain Network Slicing: Prototyping and Validation

Literature Review and Problem Description

Part III addresses the Objective 2 of this dissertation, which is the implementation and validation of solutions for network slice management and orchestration, in multi-domain environments. This objective is addressed in Papers C, D and E. Prefacing these publications, in this chapter we include two sections that help the reader have the full picture and understand the problem we want to address. Section 1 provides background context, capturing the precedents with a literature review. Section 2 identifies the main limitations of the state-of-the-art and puts them in relation with the contributions done in Papers C, D and E.

1 Background Description

1.1 Progress on research and innovation projects

When Part II of this thesis got started, the first results of 5G-PPP Phase 1 projects were made available. These results aimed to validate the architectural framework proposed therein, which focuses on particular functional aspects, such as multi-domain orchestration. For example, in relation to the 5GEx projects, the authors of [1] reported on the first MdO prototype; in particular, they demonstrated how it is possible to create and deploy network slices based on a multi-operator scenario. Vaishnavi et al. [2] provided an experimental implementation of multi-domain orchestration where multi-operator services can be deployed and monitor the service for SLA compliance over 5G networks. Finally, Draxler et al. [3] showcased how the 5G-Xhaul project's 5G Operating System (5GOS) can provide control and management for services running on top of a multi-domain 5G infrastructure. In 5GOS, the control and manipulation of resources scope different administrative and technological domains.

The main findings of Phase 1 projects were fed back to Phase 2 projects [4]. These projects used these outcomes as a basis to keep working towards the consolidation of 5G technology, developing novel future-proof solutions in the following areas: **A1**: 5G flexible RAN; **A2**: technology enablers for 5G RAN hardware and software platforms; **A3**: 5G fronthaul, backhaul and metro-haul; **A4**: 5G Autonomous Network Control and Management; **A5**: 5G Multi-Domains Multi-Tenants Plug & Play Control Plane and Slicing Control; **A6**: 5G Flexible and Agile Service Deployment; **A7**: E2E Orchestration across Optical, Packet, Wireless Virtualized Networks; **A8:** 5G Resilience and Availability; and **A9:** 5G Services Platforms and Programming Tools for NetApps.

It is worth noting that all Phase 2 projects built upon 3GPP Rel-15, which is the first release of 5G technology, when slicing became a normative feature; that is why this feature is present in all these projects. However, it is also true that these projects address the slicing concept from different perspectives and dimensions, depending on their scope of work. Table III-1 summarizes the work of the Phase 2 projects, providing a comparative analysis on their activity on slicing, in terms of i) targeted network domains, ii) orchestration solutions, and iii) multi-domain support. For further details on other non-slicing related takeaways, see the third version of the 5G-PPP Architecture White Paper [5], published in June 2019.

Table III-1. 5G-PPP Phase 2 projects: takeaways on slicing management and orchestration.

Project	**Main Areas**	**Network domain**			**Orchestration stack (slice specific modules underlined)**	**Multi-domain support**
		AN	**TN**	**CN**		
5GESSENCE	A4, A6	X			RAN Slicing Mgmt. + NFVO&VNFM + VIM[1]	No
5G-TRANSFORMER	A3, A7, A9	X	X	X	Vertical Slicer (VS)+ Service Orchestrator (SO) + Multi-Transport Platform (MTP)	Yes
5G-MONARCH	A1, A2, A7	X	X	X	Cross-domain M&O + Cross-slice M&O + NFVO&VNFM + Domain mgmt. + VIM[1]	Yes
BLUESPACE	A2, A4, A5	.X	X		Network Slice Manager[3] + NFVO&VNFM[4] + Transport SDN Controller + VIM[1]	Yes
METRO-HAUL	A3, A4, A5, A7		X		Slice Manager (thin layer) + NFVO&VNFM[2] + Transport SDN controller[5] + VIM + WIM[6]	Yes
SLICENET	A4, A5		X	X	Slice Service Orchestrator (SS-O) + Resource and multi network segment Orchestrator (NMR-O)	Yes
5G-MEDIA	A7, A8	X	X	X	NFVO&VNFM[2] + VIM[1] + WIM + Media Service MAPE	Yes
5GTANGO	A6 A8, A9			X	5Gtango Slice Manager[7] + Policy/SLA manager + NFVO&VNFM[2,8] + VIM[1] + WIM	No
MATILDA	A6, A8, A9		X	X	Computing Slicing Manager + NFVO&VNFM[2] + VIM + WIM	Yes
5G-PICTURE	A1, A2, A3, A7	X	X		Slicing Manager + NFVO&VNFM[2] + OSS + WIM	Yes
5GCITY	A5, A6		X	X	5GCity Slice Manager + NFVO&VNFM[2] + VIM	No

NOTE1: Implementation based on Openstack: https://www.openstack.org
NOTE2: Implementation based on Open Source MANO (OSM). https://osm.etsi.org
NOTE3: Implementation released under Apache2.0 license: https://github.com/nextworks-it/slicer
NOTE4: Implementation released under Apache2.0 license: https://github.com/nextworks-it/slicer
NOTE5: Implementation based on Open Network Operating System (ONOS): https://opennetworking.org/onos/
NOTE6: Add-on module incorporated to OSM stack: https://osm.etsi.org/docs/user-guide/06-osm-platform-configuration.html#wim-inter-vim-sdn-management
NOTE7: Add-on module incorporated to OSM stack: https://www.5gtango.eu/papers/2018/2018_EUCNC_WP5.pdf
NOTE8: Implementation based on SONATA: https://www.sonata-nfv.eu/content/agile-development-testing-and-orchestration-services-5g-virtualized-networks

In parallel to 5G-PPP Phase 2, there were projects from other initiatives that also worked on solutions that enable progress on network slicing. Examples of these projects were 5GPagoda! [6] and NECOS [7].

5G!Pagoda project aimed for the development of scalable network slice management and orchestration framework for multi-domain, distributed cloud network infrastructures. This project served as a catalyzer for a wide variety of works touching upon many important topics, from system architecture design to use-case driven slicing validation. On the architecture topic, the impact of multi-domain and resource federation features in network slicing were studied in works such as [8] and [9]. In relation to validation works, 5GPagoda! produced several results on network slice planning (e.g. [10][11]), in-slice traffic steering (e.g., [12][13]) and sliced content delivery networks (e.g., [14][15]).

The goal and outcomes of the NECOS project were similar to 5G!Pagoda, but much more focused on orchestration solutions scoping edge dominated infrastructure (e.g. [16]) and slice programmability, with SDN and cloud APIs (e.g., [17]). These APIs set the foundations for slice capability exposure, as they allow tenants to gain access to their slices and configure contained resources therein. This approach, coined by NECOS as slice-as-a-service, is elaborated on [18] and constituted the basis for the project's pilots that came afterwards. The showcasing was articulated into five demonstrators: Multi-Slice/Tenant/Service (MUST), Marketplace (MARK), Experiments with Large-Scale Lightweight Services Slices (ELSA), Machine-Learning based Orchestration of Slices (MLO) and Wireless Slicing Services (WISE). For further details on these demonstrators, see [19].

1.2 Progress on the standardization arena

Network slicing was appointed as a top-tier feature in 5G, upon a proposal from 3GPP community in late 2017. From that moment on, standards bodies and telco industry fora started to re-prioritize their activities, putting much more focus on the specification and recommendation of solutions enabling network slicing, horizontally (from the access network to the core network and internet) and vertically (from infrastructure layer to the service layer). In terms of management

From the point of management and orchestration, noteworthy progress was made at 3GPP SA5, GSMA and ETSI ZSM. This chapter summarizes the main findings in these organizations between 2018Q3 and 2021Q2, which is the period that covers the activities of Part III in the present thesis.

1.2.1 3GPP SA5

SA5 is the 3GPP working group responsible for developing solutions for the management, orchestration and charging of 3GPP networks and provided services. On the one hand, *management and orchestration* covers aspects such as operation, fulfilment, assurance and automation, including management entities with entities external to the network operator (e.g., digital service providers and verticals). On the other hand, *charging* covers aspects such as Quota Management and Charging Data Records (CDRs) generation, related to end-user and service-provider.

3GPP SA5 view on network slicing adheres to the three-layer scheme proposed by NGMN (Figure II-1), by defining constructions that map one-to-one with the NGMN original slicing concepts. The 3GPP constructions are three: i) *communication service,* corresponding to NGMN's service instance; ii) *network slice instance*, corresponding to the NGMN’s network slice instance; iii) *network slice subnet instance*, corresponding to the NGMN’s subnetwork instance. Figure III-1 illustrates an example of the relationship between these constructions.

- From a resource viewpoint, Network slice subnet instances (NSSIs) can be flexibly combined to form Network Slice Instances (NSIs). As seen, the NSSI concept is recursive in nature; actually, one top NSSI can be formed of domain-specific NSSIs, i.e., AN-NSSI and CN-NSSI.
- From a service viewpoint, a NSI can accommodate one or more communication services. This 1:N mapping relationship depends on the network slice criteria design. This design can be of quite different types, ranging from coarse-grained network slices, - i.e., one network slice for each 5G service category - to fine-grained slices - i.e., one network slice for each communication service.

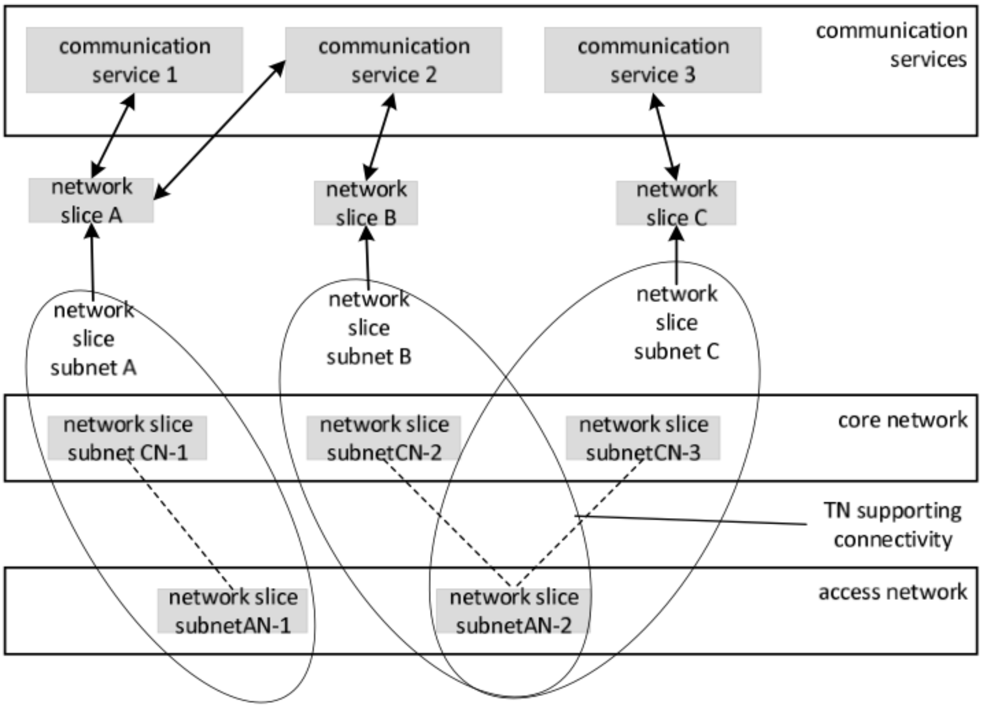


Figure III-1. A management view of 3GPP network slicing

For the provisioning and operation of the three 3GPP slicing constructions, TR 28.801 [20] introduced the following management functions: Communication Service Management Function (CSMF), Network Slice Management Function (NSMF) and Network Slice Subnet Management Functions (NSMF). Figure III-2 shows the relationships between them. In the event of a service order, the workflow will be as follows: upon capturing the service order, the CSMF translates the contained communication service-related requirements into network slice related requirements. These network slice related requirements are then forwarded to the NSMF, which decomposes them into network slice subnet related requirements for their forwarding to the NSSMF. Finally, the NSSMF processes these network slice subnet related requirements, translating them into ETSI NFV SOL-005 operations (deploying new network services or scaling existing ones) and modifications actions (configuring individual network functions to make their semantics aligned with slice expected behavior).

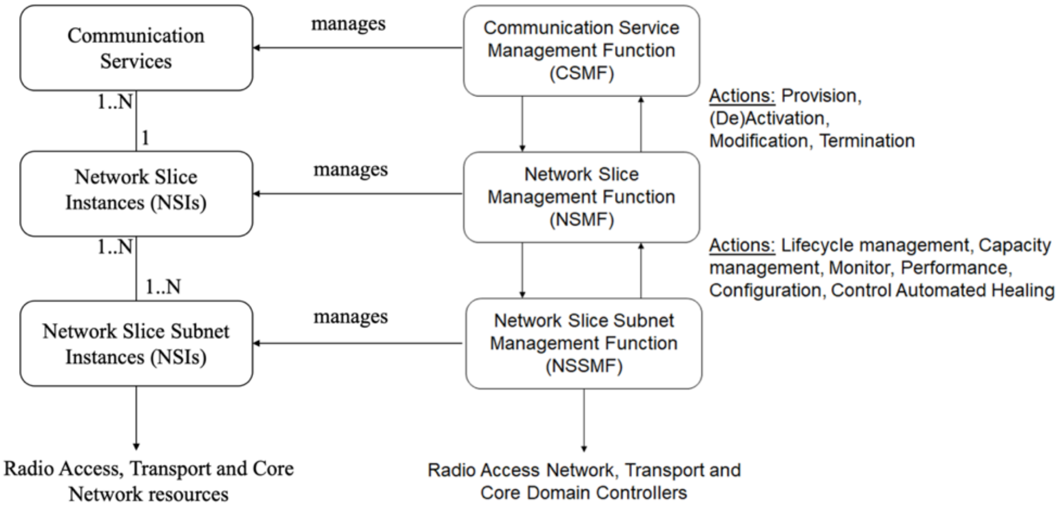

Communication Services
manages
Communication Service Management Function (CSMF)
Actions: Provision, (De)Activation, Modification, Termination
1..N
1
Network Slice Instances (NSIs)
manages
Network Slice Management Function (NSMF)
Actions: Lifecycle management, Capacity management, Monitor, Performance, Configuration, Control Automated Healing
1..N
1..N
Network Slice Subnet Instances (NSIs)
manages
Network Slice Subnet Management Function (NSSMF)
Radio Access, Transport and Core Network resources
Radio Access Network, Transport and Core Domain Controllers

Figure III-2. 3GPP management functions for network slicing support

Figure III-3 shows the complete map of 3GPP SA5 specifications, marking with colored balls those normative documents that contain solutions with direct impact on network slicing orchestration, and thus in scope of the present thesis. These documents can be clustered into three main groups:

- Network slicing use cases, requirements, and architecture. This cluster includes TS 28.530, TS 28.532 and TS 28.533. **TS 28.530** presents the business use cases and requirements of slicing. **TS 28.532** details the management services that are required for the provisioning, performance and fault management of 3GPP managed entities. Finally, **TS 28.533** explains the concepts and theory behind the construction of these management services, and their structure into a service-based management architecture (SBMA).
- Network slicing management capabilities. This cluster includes TS 28.531, TS 28.545, TS 28.550, TS 28.552, and TS 28.554. The first three specifications define the set of [request-response/subscribe-notify] operations that the vendors will later integrate into their CSMF, NSMF and NSSMF implementations. In particular, **TS 28.531** specifies management services and procedures for the provisioning of network slices (including the configuration and lifecycle management of constituent network slice subnets), while **TS 28.545** and **TS 28.550** are in charge of addressing performance and fault management aspects. The last two specifications list the performance measurements (**TS 28.552**) and KPIs (**TS 28.554**) that can be computed on a per slice level, leveraging S-NSSAI sub-counters. These metrics can be used to feed AI/ML engines, in order to make intelligent decisions that assist CSMF/NSMF/NSSMF in their orchestration activities.
- Network slicing modelling. This cluster includes TS 28.540 and TS 28.541. **TS 28.540** specifies the requirements for the definition of constructions that allows representing network slice resources in a Management Base Information (MIB), so that NSIs and constituent NSSIs can be catalogued and inventoried. **TS 28.541** builds on these requirements to define an information model (UML) for network slicing, with solutions for two data model languages: YAML and YANG. For further details on this model, see clause 6 of TS 28.541.

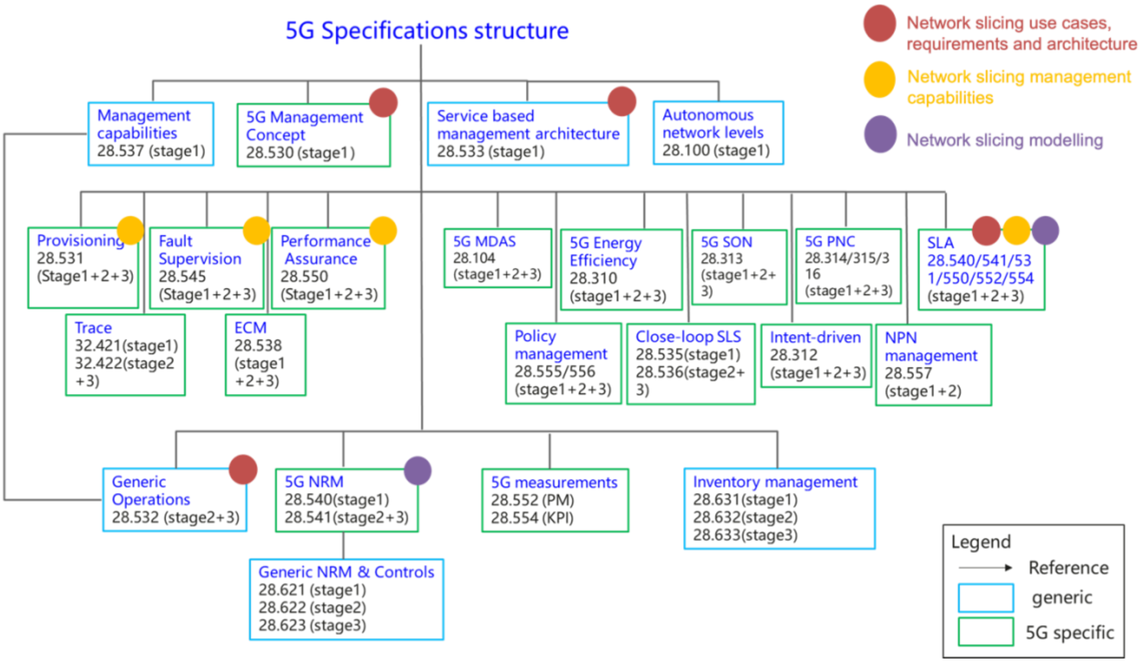

5G Specifications structure
Network slicing use cases, requirements and architecture
Network slicing management capabilities
Network slicing modelling
Management capabilities 28.537 (stage1)
5G Management Concept 28.530 (stage1)
Service based management architecture 28.533 (stage1)
Autonomous network levels 28.100 (stage1)
Provisioning 28.531 (Stage1+2+3)
Fault Supervision 28.545 (Stage1+2+3)
Performance Assurance 28.550 (Stage1+2+3)
5G MDAS 28.104 (stage1+2+3)
5G Energy Efficiency 28.310 (stage1+2+3)
5G SON 28.313 (stage1+2+3)
5G PNC 28.314/315/316 (stage1+2+3)
SLA 28.540/541/531/550/552/554 (stage1+2+3)
Trace 32.421(stage1) 32.422(stage2+3)
ECM 28.538 (stage1+2+3)
Policy management 28.555/556 (stage1+2+3)
Close-loop SLS 28.535(stage1) 28.536(stage2+3)
Intent-driven 28.312 (stage1+2+3)
NPN management 28.557 (stage1+2)
Generic Operations 28.532 (stage2+3)
5G NRM 28.540(stage1) 28.541(stage2+3)
5G measurements 28.552 (PM) 28.554 (KPI)
Inventory management 28.631(stage1) 28.632(stage2) 28.633(stage3)
Generic NRM & Controls 28.621 (stage1) 28.622 (stage2) 28.623 (stage3)
Legend
Reference
generic
5G specific

Figure III-3: 3GPP SA5 work on network slicing

1.2.2 GSMA

In 2018, GSMA published [21], providing a comprehensive overview about the service requirements on network slicing expressed by customers from different vertical industries, including AR/VR, automotive, energy, healthcare, manufacturing, public safety, and smart cities, among others. From the analysis conducted in this document, GSMA noted that service requirements on network slicing could be classified into performance, functional and control and management requirements; however, it concluded that there was no agreement on how vertical industries should express these requirements towards network operators. In this regard, GSMA agreed on the need to harmonize network slicing definition, identify network slice types with distinct characteristics and consolidate parameter and functionality requirements, from end-to-end perspective.

As per the recommendations above, GSMA Networks Group (NG) took action, with the development of a solution that would be able to offer verticals guidelines on how to issue service requirements on network slicing towards network operators, therefore addressing the existing gap between vertical and telco industries. This solution is called Generic network Slice Template (GST), which has been documented and maintained in GSMA PRD NG.116 [22]. The GST provides a universal description of a network slicing, containing all the potential attributes that can be used to characterize a network slice. It allows the network slice provider (e.g., network operator) and network slice customer (e.g., industry vertical) to agree on the Service Level Specification (SLS) for a network slice, by means of filling GST attributes with values based on service requirements. The result of this mapping defines a Network Slice Template (NEST), which in essence is a filled-in version of the GST that allows characterizing a network slice based on a service type. Different NEST's allow describing different types of network slices. On the one hand, for slices based on 3GPP 5G service categories (e.g., eMBB, mIoT, uRLLC), the operator may have a set of standardized NEST's (S-NEST). On the other hand, for slices addressing specific industry use cases (e.g., industry 4.0, logistics, eHealth), the operator can define additional private NEST's (P-NESTs). Both S-NEST's and P-NEST's are registered and published in the operator's service catalog. This can be seen in Figure III-4.

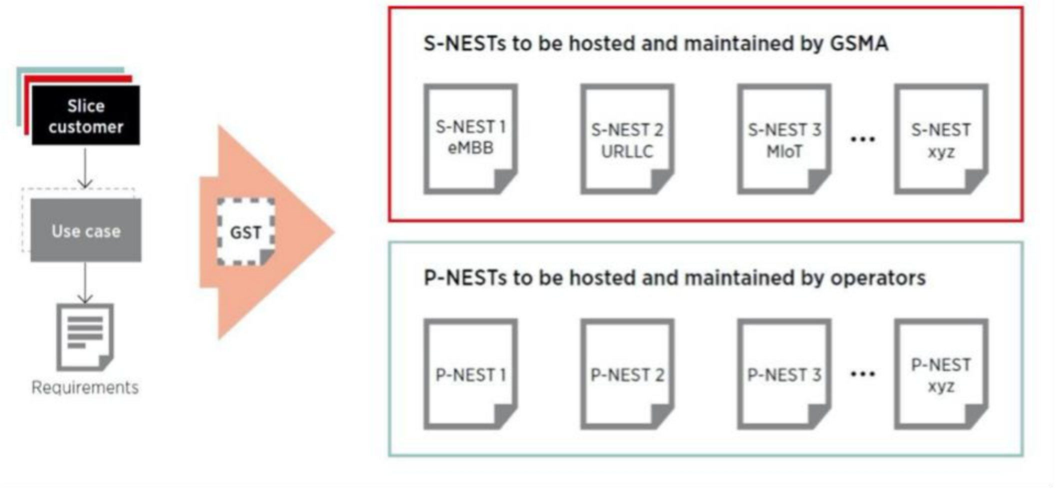


Figure III-4. 3GPP SA5 work on network slicing

It is worth mentioning that there exists a direct relationship between the GSMA work on GST/NEST and 3GPP SA5 work on slice modelling; actually, NEST attributes are considered as input for the ServiceProfile, which is the 3GPP slice construction that captures the network slice related requirements. As shown in Figure III-5, these requirements are further translated into network slice subnet requirements for RAN (i.e., NG-RAN SliceProfile) and CN (i.e., 5GC SliceProfile), and TN connectivity across them.

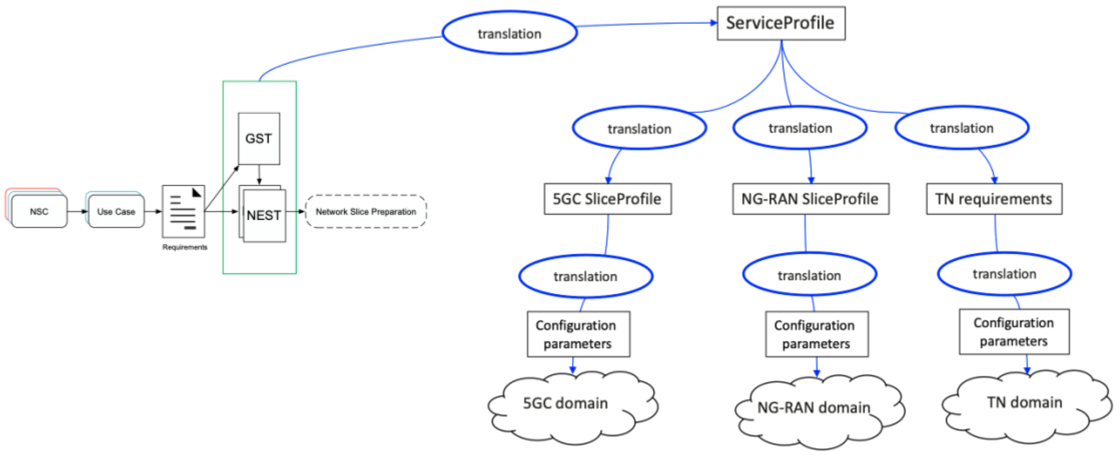
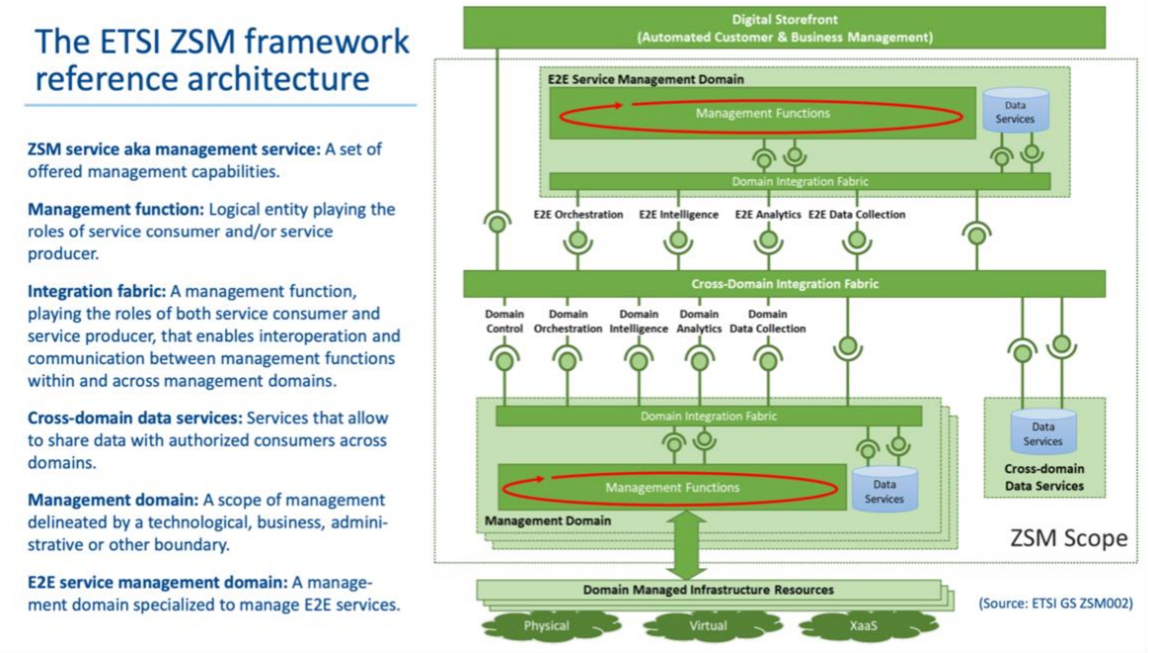


Figure III-5. The network slice journey – from GST to network slice configuration parameters

1.2.3 ETSI ZSM

The arrival of SDN, NFV and 5G are progressively turning networks into programmable, software-driven, and service-based infrastructures. This network transformation, together with the introduction of network slicing capabilities, has triggered the need to radically change the way resources, functions and services are managed and orchestrated. To face this challenge, ETSI defined a new ISG: Zero-touch network and Service Management (ZSM). ETSI ZSM has the mission of defining a robust, yet scalable OSS architecture based on cloud-native principles to accomplish zero-touch (fully automated) network and service operation. Figure III-6 pictures the as-is ZSM framework reference architecture and their building blocks.



The ETSI ZSM framework reference architecture

ZSM service aka management service: A set of offered management capabilities.

Management function: Logical entity playing the roles of service consumer and/or service producer.

Integration fabric: A management function, playing the roles of both service consumer and service producer, that enables interoperation and communication between management functions within and across management domains.

Cross-domain data services: Services that allow to share data with authorized consumers across domains.

Management domain: A scope of management delineated by a technological, business, administrative or other boundary.

E2E service management domain: A management domain specialized to manage E2E services.

(Source: ETSI GS ZSM002)

Figure III-6. ETSI ZSM architecture framework

The ZSM architecture [23] is composed of different management domains and one E2E service management domain, all interacting with each other through a cross-domain integration fabric. The scope of a management domain depends on the use case under consideration, and at the end of the day is a decision that only concerns the operator; actually, the operator can decide whether there exists a management domain per technology domain, vendor domain, network domain, administrative domain, or any combination in between. This gives the flexibility to adapt ZSM framework to different application scenarios, including both operator-internal and multi-operator services, as well as public-private network scenarios, where the slicing can play a key role, as outlined in Part I.

In 2018, ETSI ISG ZSM launched a work item entitled “End-to-End management and orchestration of network slicing”. This work item, which will result in the publication of ZSM003 [24] in late 2020, strives to specify requirements and solutions for the zero-touch operation of network slices, when deployed across multiple management domains. These requirements and management solutions are technology-agnostic, in the sense they can be applied on individual domains (e.g., access domain, transport domain, cloud domain), regardless of their specificities (e.g., mobile access vs fixed access in access domain, IP/MPLS network vs optical network for access transport, VM-based orchestration vs container-based orchestration). In pursuing this goal, the ETSI ISG ZSM:

- identifies relevant SDOs working on network slicing specifications, shedding light on their individual scope (e.g., 3GPP for mobile access and core network slicing, IETF for transport network slicing, ETSI NFV for virtualized network slice).
- leverages the on-going work in these SDOs, pointing out existing gaps/inconsistencies across them and providing necessary means for their addressment. These means, based either on plug-in-based adaptations and model translations among domains, are supplied by ZSM cross-domain integration fabric.

Apart from this network slicing activity, ZSM has launched other work items addressing features that can help operator to reduce “the number of touches”. One of them is Closed-Loop Automation (CLA). Documented in ZSM009 [25], this feature describes the ability to define and operate closed loops at different layers, to promote self-X capabilities (e.g., auto-scaling) based on a continuous "monitoring -> analysis -> decision -> execution" pipeline.

1.3 5G experimentation facilities

As 5G standards progressed and technical development of Phase 1 and 2 projects stabilized, the EC realized that there was a needed to accelerate the uptake of 5G in Europe, by providing an E2E facility that would allow lowering the entry barrier for vertical industries to pilot use cases. Such a European-wide 5G facility would require much more than interconnecting existing 5G labs (e.g., [26], [27]); indeed, it would require major European industry actors to sit in a common table and have them commit to establishing a pre-production facility with real 5G equipment, based on three tenets: i) the facility would need to run 24/7; ii) the facility would have to demonstrate that the key 5G network KPIs can be met; and iii) the facility would be accessed and used by vertical industries to set up research trials of innovative use cases, to further validate service KPIs in the context of concurrent usages by multiple users.

5G-PPP launched Phase 3 in July 2018 with the above goal in mind, by funding three projects: 5G-EVE [28], 5G-VINNI [29] and 5Genesis [30]. These projects constituted a big step forward, since it meant shifting from isolated piecemeals (Phase 1 and Phase 2 projects) to a large-scale 5G infrastructure supporting pre-commercial pilots. Table IIIII-2 summarizes the main capabilities of the three flagship projects.

Table III-2. Takeaways of three flagship projects: 5G-EVE, 5G-VINNI, and 5G-EVE

Supported Capabilities and Services			**Infrastructure projects**		
Area	**Features**	**Reference standards**	**5GEVE**	**5G-VINNI**	**5GENESIS**
ARCHITEC-TURE	5G NSA (NR+EPC)	3GPP TS 23.501	Rel-15&16	Rel-15&16	Rel-15&16
	5G SA (NR+5GC)	3GPP TS 23.501	Rel-15&16	Rel-15&16	Rel-15&16
	Sub-6 GHz NR	3GPP TS 38.401	Rel-15	Rel-15	Rel-15
	mmWave NR	3GPP TS 38.401	Rel-15	Rel-15	Rel-15
ACCESS AND SPECTRUM	Virtualized 5G RAN	Multiple SDOs	Rel-15	Rel-15	Rel-15
	NBIoT and LTE-M	3GPP TS 23.628 and 36.101	Rel-14	Rel-14	Rel-14
SLICING MGMT	Slicing for 5G NSA	3GPP: TR 28.801 ETSI NFV: IFA and SOL specs ITU-T: Y.3100	NFV Rel-2	NFV Rel-2	NFV Rel-2
	Slicing for 5G SA	3GPP: TS 28.530, 28.531, 28.533 ETSI NFV: IFA and SOL specs	Rel-15 NFV Ph-3	Rel-15 NFV Ph-2	Rel-15 NFV Ph-3
	Customized slicing	Multiple SDOs and open-source community	Rel-15&16	Rel-15&16	Rel-15&16
INTER-WORKING	Interworking with other projects	Builds upon ETSI NFV principles	Yes	Yes	Yes
EDGE COMPUTING	Multi-access Edge Computing	ETSI MEC specs	MEC Ph-1	MEC Ph-1	MEC Ph-1

BACKHAUL	mmWave for backhaul	IEEE 802.11ad or IEEE 802.11ay	No	Yes	Yes
	Satellite for backhaul	3GPP: TS 22.261, TR 22.819, 28.822, TR 38.811, 38.821 ETSI: TR 103 611 ITU-R: M.2460-0	No	Rel-15	Rel-15
VALIDATION	Automatic Testing Framework for KPI validation	ITU-R: M.2083-0, M.2410-0 ETSI: NFV-TST 5GPPP TMVWG	Yes	Yes	Yes

1.4 Network slicing solutions and PoCs

In this section, we will analyze state-of-the-art work on network slicing validation, considering solutions and pilots developed in the different research projects, as well as PoCs executed in the open-source communities such as ONAP and OSM.

1.4.1 Early lab solutions and pilots

Most of the demonstrations on network slicing come from the activity done in research and innovation projects. Apart from the works reported by 5GPagoda! (i.e., [10]-[15]) and NECOS (i.e., [18]-[19]) in Section 1.1, other projects from 5GPPP Phases 1 and 2 have also produced remarkable results. These include 5GEx [31], 5G-Essence [32][33], 5G-Transformer [35]-[39], 5G-MoNArch [40]-[44], Metro-Haul [45][46], SliceNet [47]-[49], 5GTango [50], 5GMATILDA [51][52], 5G-Picture [53] and 5GCity [54]. The literature referenced here includes both simulation work and small-scale pilots.

In relation to simulation work, there is a wide variety of solutions, with different scopes in terms of *network domains* (from domain-specific to end-to-end slicing) and *optimization goals* (e.g., maximizing the number of hosted services, or minimize the impact on running slices upon the arrival of unexpected traffic surges). To facilitate their discussion, the following taxonomy is proposed:

- **Multi-tenancy support**, addressed in [31]. In this work, the authors present a PoC demonstration of an SDN/NFV-based orchestrator that enables resource sharing among different tenants. The profit of an infrastructure provider is maximized by the proposed orchestrator using a dynamic slicing approach based on big data analytics.
- **RAN-only slicing,** addressed in [32]-[34] and [40]-[41]. The authors of [32] analyze different RAN slice configuration parameters at L2 and L3, and compare them through system-level simulations in a scenario with two slices: one providing eMBB service and another providing mission-critical services. The authors show how setting different parameters allows controlling the operation of 5GNR protocol stack to provide traffic differentiation and protection among RAN slices. The conclusions of this paper are used to elaborate solutions on cell planning and slice provisioning, which are validated and discussed in [33]. The authors of [34] design LACO, a RAN-specific network slice orchestrator that considers network slice request with strict latency request, and serves them with formal delay guarantees. In [40], Hans et al. present a novel online generic RAN slicing strategy optimizes to maximize the long-term network utility in wholesale offerings towards MVNOs. The solution is

evaluated through numerical simulations, exhibiting a satisfaction approximate to the global optimum, a fast convergence, a timely adapation to environment variation and a good scalability. Finally, [41] proposes a RAN framework built upon game theory, where each network slice unilaterally reacts to the settings of the others. The authors propose algorithms for admission control, weight allocation and user dropping, which jointly bring system to the Nash Equilibrium, making the solution an effective and implementable scheme for dynamically sharing resources across slices, both for elastic and for inelastic traffic.

- **TN-only slicing,** with mechanisms proposed and evaluated in [45] and [46]. The focus of these two works on the optical transport (L0-L1), without any references to the impact that the proposed solutions have on the management and orchestration plane.
- **End-to-end slicing,** covered in [35]-[37], [42] and [47]. The authors of [35] and [36] prove the feasibility and reliability of Overbooking Network Slices (OVNES) solution in a simulated Evolved Packet System (EPS) environment, consisting of an eNB, OpenEPC and some OpenFlow SDN controllers. This solution applies admission control policies based on a data-driven, ML-assisted revenue maximization strategy while, at the same time, assigns resources in terms of expected throughput (on the RAN), computational access (on the edge and core network) and latency constraints (on the transport network). Capitani et al. [37] demonstrate the deployment of a 5G mobile network slice through the 5G-Transformer architecture experimentally. In [42], Garcia Avilés et al. present PONSES, and open-source solution for practical end-to-end network slicing based on a slice-aware shared RAN. The authors design the require algorithms and protocols, and provide a full implementation based on Eurecom's OpenAirInterface, openLTE and srsLTE, which is used to validate the effectiveness of POSENS in achieving tenant isolation and network slice customization. Finally, the authors of [47] demonstrate how network slicing can guarantee a committed QoS performance through the end-to-end data plane (including RAN, MEC, CN and the wired connections among the different network segments), regardless of unexpected traffic surges that might lead to network congestion. The implementation is based on a fine-grain traffic adaptation supported by a three-layer hierarchical schema (scheduling, shaping and differentiating).

On the other hand, the pilots aim to assess developed solutions into real-world testbeds, to validate the lab hypothesis against vertical applications and services. The vertical industries in scope include media ([38][39][50][53]), logistics [43], tourism [44], eHealth [48], energy [49] and smart cities ([51][52][54]).

- **Media**. The authors of [38] showcase the end-to-end provisioning of high-definition streaming service using a dedicated network slice, with smart placement of virtual appliances at the edge to avoid bottlenecks in the core of the network and provide a low latency service. Measurements confirm that the coordinated operation of the 5G-TRANSFORME building blocks contributes to reducing service creation time to the order of minutes. Boubendir et al. illustrate in [39] the on-demand creation and deployment of network slices dynamically over multiple domains for live content services in a stadium. A network operator can achieve the federation of access and edge resources owned by private third-party actors through B2B relationships. [50] reports the use of 5TANGO stack to deploy a commercial operator real-time unified communication platform using network slices over a NFV infrastructure. Each instance of a collaboration system for real-time communications (e.g., multi-conference, screen sharing and whiteboard) is deployed on a different network slice, with different QoS requirements which are enforced through a Transport network, using WIM interfaces. Finally, [53] describes the 5G-PICTURE pilot showcased in Bristol, UK. The pilot consisted in a football stadium demonstration with ultra-high user density supporting entertainment services, with slices deployed to i) provide differentiated treatment of media applications, and ii) ensure service resilience in a multi-connectivity link scenario, based on combining LTE with Wi-Fi.

- **Logistics**. In [43], a full E2E slicing-enabled mobile macro network is provisioned in the Hamburg Port, with the 5G radio base station installed on the television tower close to the Hamburg trade fair and congress center. The slicing usage responds to the need to efficiently accommodate different services using a common infrastructure, guaranteeing their performance in terms of isolation. The demo showcased the provisioning of right-sized slices, each hosting applications from a different use case: traffic light control, mobile sensors on barges and improved port operations.
- **Tourism.** [44] reports on the “Turin Touristic City testbed”, which demonstrates the benefits of network slicing and edge computing to provide tourists with an interactive virtual reality visit of a representative room of Palazzo Madama in Turin. In this 5G-MoNArch pilot, the tourists (end users) interact through a virtual reality application that relies on the instantiation of two slice: one to manage a 360-degree video stream (eMBB slice with high throughput requirements), and the other to process haptic/voice communication (uRLLC slice with very low latency requirements).
- **eHealth.** The solution described in [47] is validated with an eHealth pilot. This pilot, reported in [48] demonstrates how to provide support to medical emergency first responders, by rapidly provisioning dedicated end-to-end broadband 5G slices to advance the emergency ambulance services through the design of better-connected, integrated and coordinate healthcare. The slice accommodates prioritized life-critical video-streaming from inside a high-speed moving ambulance, with the objective to offer reliable and dependable QoS with ‘zero-perceived’ downtime.
- **Energy.** In [49], it is demonstrated the use of a uRLLC slice to provide a fully decentralized high-speed self-healing solution for smart grid. These self-healing solutions rely on distributed automation and power system protection, and aim at increasing energy supply QoS by reducing the number of customer affected by power outages, as well as the frequency and duration of these outages.
- **Smart cities.** The authors of [51] display the use of a network slice to implement operational smart city intelligent lighting solution in Alba Luli, a middle-size city in Romania with about 70,000 inhabitants. This solution consists of the following application functions: i) IoT aggregator, collecting information from distributed lighting devices; ii) an IoT platform, in charge of provisioning, processing, visualization and device management role; iii) a dashboard, used by system administrators to manage lighting devices, iv) a storage system, responsible for data collection from IoT platform; v) the ticketing system, for tracking and resolving possible issues; and vi) the billing platform. The demonstration was conducted with 100 lighting devices, each with a bandwidth capacity of 100 Kbps. The results of the demonstration showed that by confining traffic to this slice, the following KPIs can be achieved: availability >99.9%, 100 Mbps throughput, latency <300 ms, jitter <100 us, and packet loss rate <0.1% [52].Finally, [54] showcases how the 5GCity neutral host platform enables an ICT infrastructure owner to slice and lease its infrastructure to 3rd parties, including MNOs and MVNOs, so that they can extend their service footprint in areas (e.g., stadia, shopping mall, railway) that are not economically

feasible for them to invest in CAPEX. The demos are conducted in Luca (Italy) and Barcelona (Spain).

1.4.2 Community-led PoCs

Research and industry projects are not the only ones which provide early solutions on network slicing; indeed, industry initiatives such as Open Network Automation Platform (ONAP) and Open Source MANO (OSM) have also used their ever-evolving software stacks to demonstrate slicing scenarios.

ONAP is an open-source project hosted by the Linux Foundation. It offers model-based management and orchestration services to network/cloud providers, offering them capabilities to deploy and operate network (function, service, function) instances in softwarized environments with great level of automation. 3GPP network slicing capabilities were first integrated in ONAP in late 2019, with the development of modules implementing the three management functions reported in TR 28.801 (i.e., CSMF, NSMF and NSSMF) together with some attributes from the slice model reported in TS 28.541 (i.e., ServiceProfile) [55]. These capabilities were made available in Frankfurt Release [56], and further enhanced in Guillin Release [57]. As can be seen from Figure III-7, the main difference between both releases is that Frankfurt (ONAP Release 6) only touches on RAN and CN slice subnets, while Guillin (ONAP Release 7) integrates slicing-awareness at TN domain. The presentation in [58] summarizes the main findings for the progress made on ONAP based slicing until December 2020. As the reader may notice from this documentation, despite the development of slicing features in the different ONAP modules [59], no formal PoC was conducted during the lifetime of the Part III of the present thesis; proof of this is that test cases listed in [56] and [57] have not yet been passed. The main (but not the only) reason is the complexity behind ONAP setup, as reported in [60].

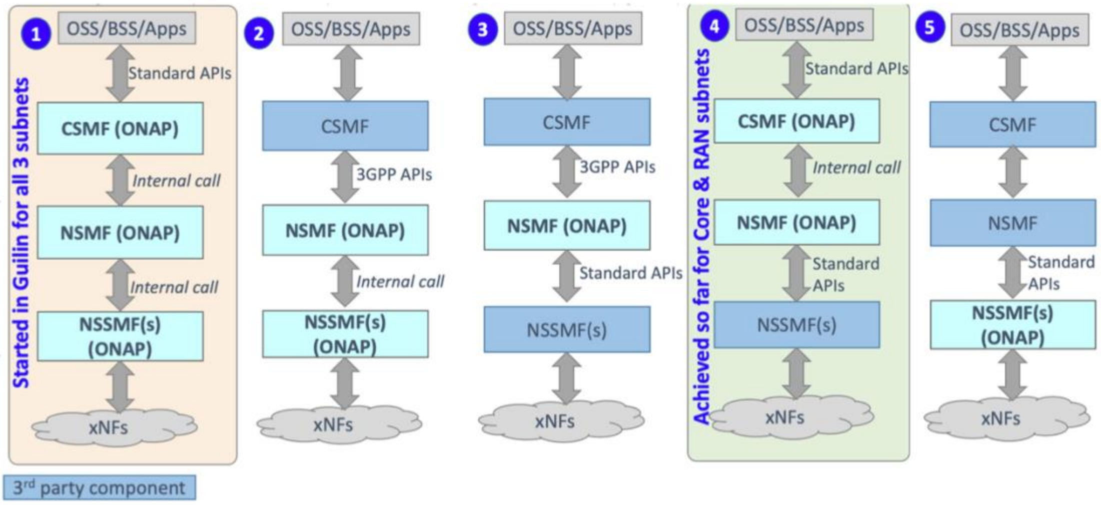
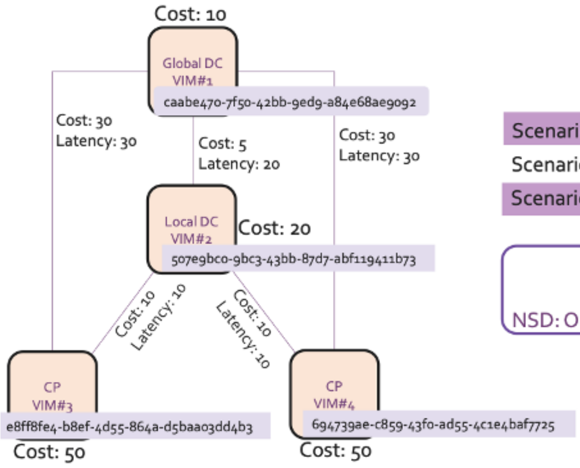
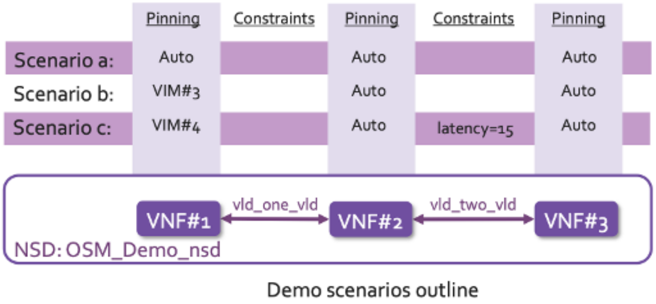


Figure III-7. ONAP slicing with Frankfurt Release (option 4) and Guilin Release (option 1).

OSM is an ETSI-hosted project which aims at developing an open-source MANO stack aligned with ETSI NFV specifications. The scope of OSM project covers both design-time and run-time aspects related to service delivery for telco operator environments. Since OSM Release FIVE (2019), OSM incorporates slicing features, thanks to the contributions done by 5GTANGO project (see NOTE7 in Table III-1). These contributions consisted in enriching YANG-based OSM information model with network slice (subnet) templates aligned with the NSST/NST constructions reported in TR 28.801. To make slice-to-network service association, these templates contain the pointers to the corresponding NSD together with SOL005 instantiation parameters [61]; as the reader may notice, this is a solution aligned with the one that PhD candidate had previously reported in Paper B. In OSM Release EIGHT (June 2020), two novel features were added for network slicing support: the definition of *Placement optimization module (PLA)* and the *quotas management functionality*. On the one hand, the PLA helps OSM user to find an optimal deployment of network slices, distributing the individual VNFs over the set of available VIMs based on user-provided models of i) computing and networking cost, and ii) latency and jitter metrics of inter-VIM connectivity. Figure III-8 shows an example on how PLA works. On the other hand, the quotas management functionality allows setting limits for the infrastructure, packages, and deployed instances (network service and slice instances) per OSM individual clients, therefore enabling multi-tenancy support. During the lifetime of the Part III, the following slicing PoCs have been conducted: “5G Network Slice Orchestration with OSM” [62], presented at MWC’19; and “provisioning of Magma EPC slices” [63], with Magma being a Facebook-led implementation of a production-ready vEPC.





	Pinning	Constraints	Pinning	Constraints	Pinning
Scenario a:	Auto		Auto		Auto
Scenario b:	VIM#3		Auto		Auto
Scenario c:	VIM#4		Auto	latency=15	Auto

Figure III-8: Example of the OSM’s PLA usage.

2 Ambition

The ambition is to validate the architectures and model-driven mechanisms which were conceptualized in Part II, in relation to the realization of network slicing in multi-domain environments. This validation is to be done using the infrastructure resources and orchestration stacks of real-world 5G experimentation facilities. Among the three facilities reported in Section 1.3, 5G-VINNI facility is selected. The reason is that the PhD candidate took an active participation in 5G-VINNI project, leading activities in relation to E2E network slice and service management. Therefore, it is natural to leverage the capabilities available in this facility in order to *assess the hypothesis and solutions* captured in Papers A and B, and *report main findings* in the papers conforming Part III.

2.1 Network slicing orchestration in action: on-boarding

The on-boarding is part of the preparation phase of a network slice lifecycle (see Part I, Section 4.2). It consists in uploading all the artifacts that are needed for the deployment of a slice, making sure that they are available for use by the time customers come to the system. These artifacts include Network Slice Descriptors (stored in the Network Slice catalogs), the NSDs and VNFDs (stored in the NFV catalogs) and other configuration files. As reported in Paper B, a Network Slice Descriptor is a model-based template that captures everything that is needed to manage a network slice throughout its lifecycle, from commissioning to de-commissioning. To that end, it was proposed that this descriptor included information on:

a) topology, specifying the components building up the slice, and the connectivity among them. The network slice components include a set of operator-provided network functions (most of them deployed in virtualized environment), which are arranged into one or more NFV network services. They might also include 3rd party components, typically functions providing value-added services (e.g., analytics) or vertical oriented applications (e.g., IoT server).

b) service requirements, which represent the set of KPIs and functionalities that the slice can fulfil.

c) temporal requirements, specifying the periods when the slice needs to be active during its lifetime; in fact, there are cases where there is no need for a slice to be 24/7.

d) geolocation requirements, specifying the geography where the slice shall provide coverage. In case this geography spans beyond the footprint of the operator, a slice across two or more administrative domains is required.

e) operational requirements, capturing the configuration parameters that allow operating instances of this slice (NSIs), when activated, at run-time.

In the following, we will describe the main findings of state-of-the-art solutions on network slicing modelling, and we eventually compare them against expectations reflected in Paper B.

As outlined in Section 1.1, it was not until 5G-PPP Phase 2 when the first solutions for network slicing were proposed (recall that the primer focus on 5G-PPP Phase 1 was instead architectural systems for multi-domain SDN/NFV infrastructures). Among all the 5G-PPP Phase 2 projects reported in Table III-1, three stand out in the topic of network slice descriptors: 5G-TRANSFORMER, SLICENET and 5GTANGO.

5G-TRANSFORMER proposes an orchestration stack that includes three main sub-systems: Mobile Transport and Computing Platform (MTP), Service Orchestrator (SO) and Vertical Slicer (VS). As seen in, MTP and SO focus on the deployment and operation of NSIs throughout a federated virtualized environment involving multiple administrative (and technology) domains, while the VS is the subsystem that handles the interaction of verticals with the operator system. The main findings of this solution are as follows:

- The VS is the sub-system that is conscious of the business needs of the verticals, their SLA requirements, and how they are satisfied by mapping them into given network slices
- The VS includes a catalogue of vertical service blueprints (VSBs), towards which verticals can issue service orders. A VSB is a baseline template that provides a vertical-oriented description of a service offering, allowing verticals to focus on the service logic and requirements, without caring on how they are eventually deployed at the infrastructure level. For this end, VSB is designed as a simple interconnection model that includes information on service graphs, vertical functions, traffic flows and connection points. The result of issuing a service order towards a given VSB is a Vertical Service Descriptor (VSD).
- The VS maps customer-facing requirements into resource-facing requirements, being the latter under the scope of SO and MTP. To this end, it translates VSD into an

extended Network Service Descriptor (NSD) that is used as deployment template for NSIs.

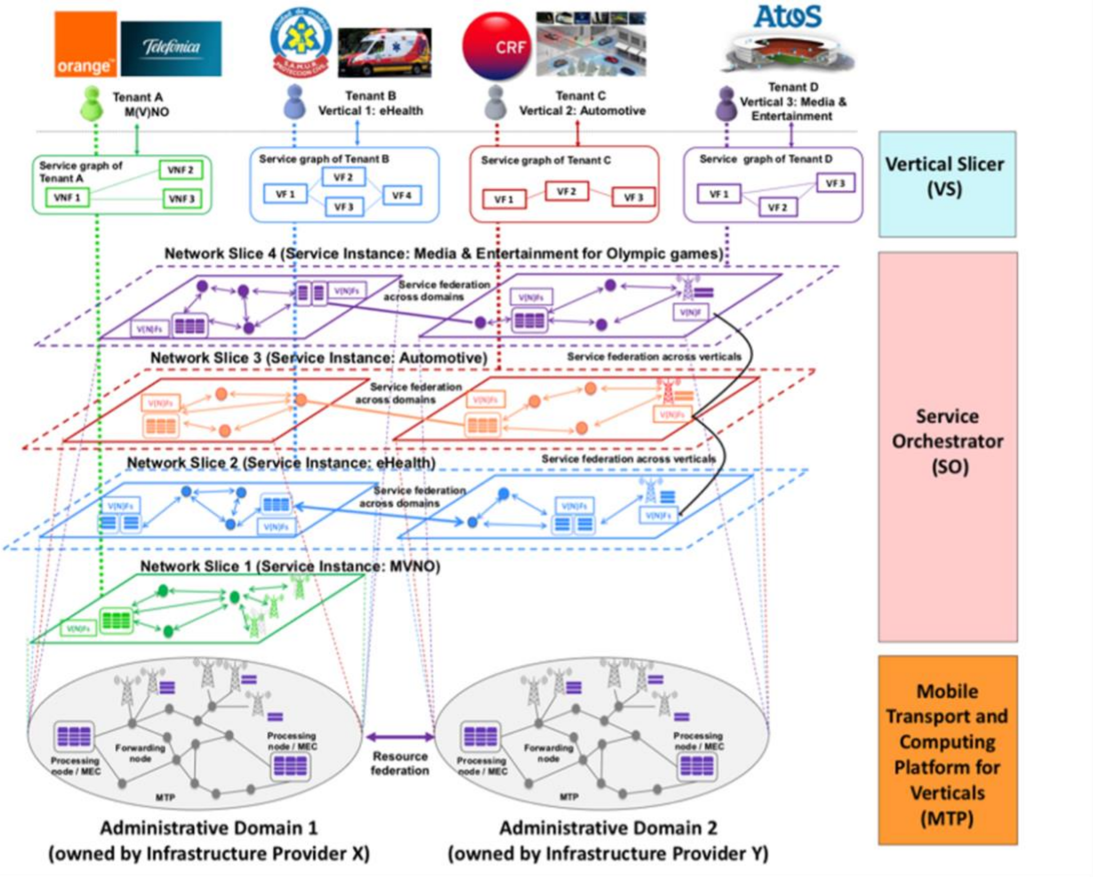

orange
Telefónica
CRF
AtoS
Tenant A
M(V)NO
Tenant B
Vertical 1: eHealth
Tenant C
Vertical 2: Automotive
Tenant D
Vertical 3: Media & Entertainment
Service graph of Tenant A
VNF 1
VNF 2
VNF 3
Service graph of Tenant B
VF 1
VF 2
VF 3
VF 4
Service graph of Tenant C
VF 1
VF 2
VF 3
Service graph of Tenant D
VF 1
VF 2
VF 3
Vertical Slicer (VS)
Network Slice 4 (Service Instance: Media & Entertainment for Olympic games)
Service federation across domains
V(N)Fs
V(N)F
Network Slice 3 (Service Instance: Automotive)
Service federation across verticals
Service federation across domains
Network Slice 2 (Service Instance: eHealth)
Service federation across verticals
Service federation across domains
Network Slice 1 (Service Instance: MVNO)
Service Orchestrator (SO)
Processing node / MEC
Forwarding node
Processing node / MEC
MTP
Resource federation
Administrative Domain 1
(owned by Infrastructure Provider X)
Administrative Domain 2
(owned by Infrastructure Provider Y)
Mobile Transport and Computing Platform for Verticals (MTP)

Figure III-9. Network slicing in 5G-TRANSFORMER. Source: [64].

SLICENET presents a service orchestrator which is equivalent to the 5GT-VS from 5G-TRANSFORMER. As pictured in Figure III-10, the Service Template (ST) and Service Descriptor (SD) in SLICENET are equivalent to VSB and VSD in 5G-TRANSFORMER.

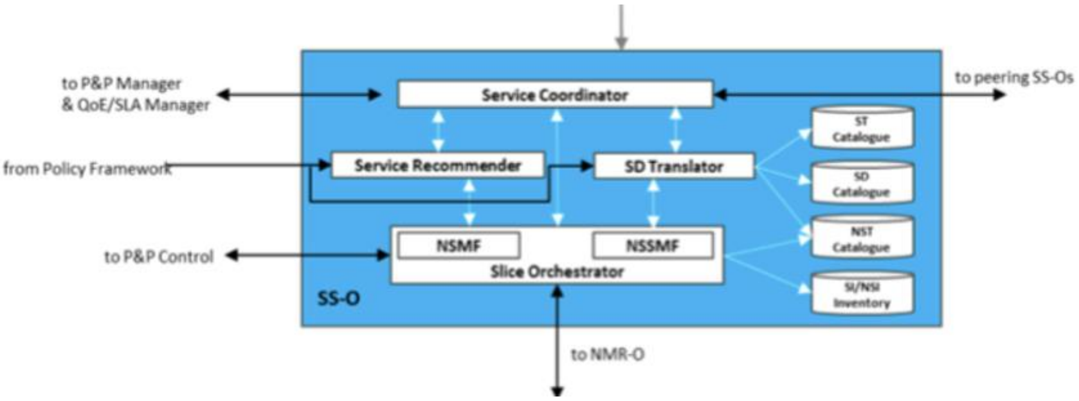


Figure III-10. SLICENET slicing service orchestrator. Source: [65]

Finally, **5TANGO** developed two artifacts for model-based network slice description, aligned with the approach propelled in 3GPP TR 28.801: Network Slice Template (NST) and Network Slice Subnet Template (NSST). Figure III-11shows the structure of a basic NST, composed of two NSSTs, each mapped to an NSD.

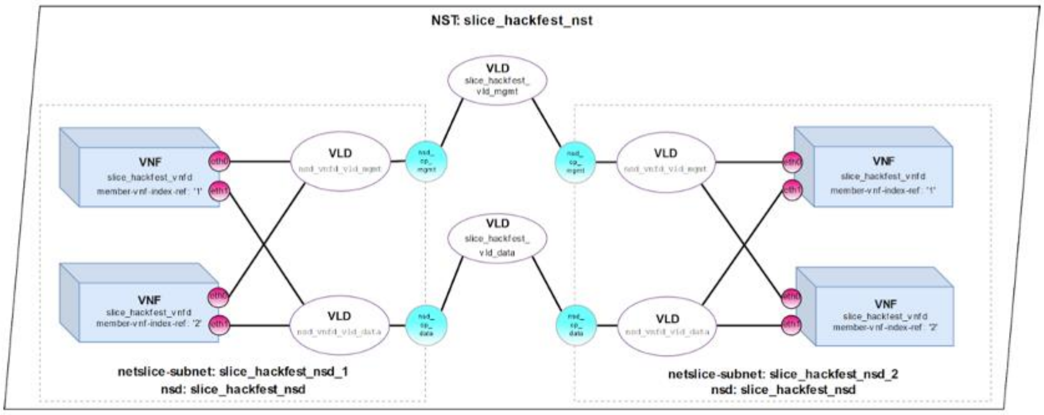


Figure III-11. An example of NST in OSM.

In addition to these three research projects (Section 1.1), OSM and ONAP (Section 1.4) also provide state-of-the-art solutions for network slice modelling.

On the one hand, **OSM** information model first incorporated network slicing artifacts in Release FIVE. Following the recommendations captured in ETSI NFV-EVE 012 [66], it was proposed to adopt the NST/NSST artifacts that 5GTANGO project had developed. Though it is true that original NST/NSSTs have evolved over the latest OSM releases, for example with the incorporation of attributes that allow associating NSDs with 3GPP slicing parameters (S-NSSAI for network slice signaling identification and 5QI for in-slice QoS treatment), their baseline structures remain the same.

On the other hand, **ONAP** provides four types of template, which are: i) Communication Service Template (CST), used to collect SLA requirements from the customer; ii) Service Descriptor (SD), used to record the requirements collected by the CST and convert them to the end-to-end network slice requirements; iii) NST, used to deploy of NSIs; and iv) NSSTs, used to describe slice subnet capability information, associating virtual resource (NSD) information to it. Figure III-12 pictures them and illustrates their relationship with reference standards. One can note that ONAP's CST and SD are equivalent to 5G-TRANSFORMER's VSB and VSD constructions, while ONAP's NST/NSST are equivalent to those available in OSM.

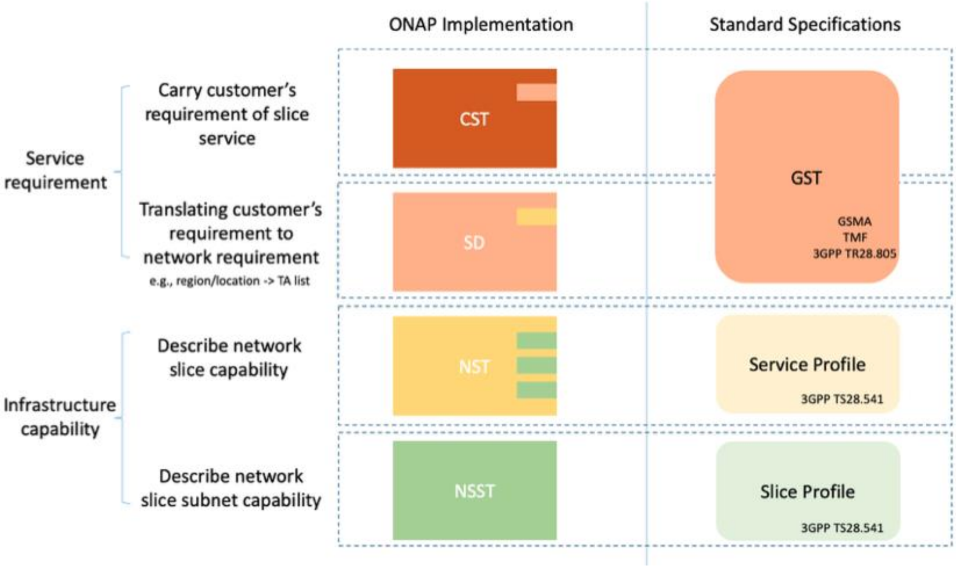

ONAP Implementation
Standard Specifications
Service requirement
Carry customer's requirement of slice service
Translating customer's requirement to network requirement
e.g., region/location -> TA list
Infrastructure capability
Describe network slice capability
Describe network slice subnet capability
CST
SD
NST
NSST
GST
GSMA
TMF
3GPP TR28.805
Service Profile
3GPP TS28.541
Slice Profile
3GPP TS28.541

Figure III-12. Network slice model in ONAP. Source: [67].

Table III-3. Comparative analysis of network slice modelling solutions

Topic	Means of verification	5GT	SLICENET	TANGO & OSM	ONAP
Topology	To specify the topology of operator-provider components, the network slice descriptor must include pointers to the NSD of individual network services.	Full support in VSD	Full support in SD	Full supported in NST/NSST	Full support in NST/NSST
	In relation to 3rd party components, the network slice descriptor must expose slice access points to the customers. The customer will use these points to attach external components into the slice.	Full support in VSB	Partial support in ST	Full support in NST/NSST	Not supported
Service requirements	The network slice descriptor must use GST parameters to capture this information	Partial support in VSB	Partial support in SD	Not supported	Full support in CST
	The filled-in parameters (NEST) are input to ServiceProfile	Partial support in VSD	Partial support in ST	Not supported	Full support in SD
Temporal requirements	The network slice descriptor must include scheduling for activation/deactivation, in case a network slice instance does not need to be up and running 24/7.	Not supported	Not supported	Not supported	Not supported
Geolocation requirements	The network slice descriptor must use GST coverage area parameter to specify this information.	Partial support in VSB	Not supported	Partial support in NST/NSST	Full support in SD
Operational requirements	The network slice descriptor must include information on NSI capability exposure towards customers, defining their visibility and control over the slice.	Not supported	Not supported	Not supported	Not supported
	The network slice descriptor must include allowed lifecycle management operations and policies in scope for the NSI	Partial support in VSD	Full support in SD	Full support in NSD policies and autoscaling rules	Partial support in NST/NSST
	The network slice descriptor must include information on NSI priority level, to handle network congestion scenarios.	Partial support in VSD	Not supported	Full support in NST/NST	Partial support in NST/NSST
	The network slice description must include information on monitoring, specifying the metrics and fault alarms that need to be collected at run time. This was not originally part of the GST.	Full support in VSD	Full support in SD	Full support in NSD policies	Full support in NST/NSST

Table III-3 provides a comparative analysis among these state-of-the-art solutions, compare them against the expectations for a network slice descriptor captured in Paper B. The main findings are summarized as follows:

- The NST/NSST artifacts in 5G-TANGO and OSM provide lightweight slicing awareness; this awareness is built upon the recommendation in NFV-EVE012, which clarifies the relationships existing between 3GPP constructions (network slice and subnets) and well-grounded NFV constructions (NSDs and VNFDs), including nested/composite patterns across them. However, most of the work developed in SA5 (in relation to Service/SliceProfile) and GSMA (in relation to GST and NEST) is not incorporated to the NST/NSST. In a nutshell, a priori they are the weaker solution in terms of capabilities. However, in terms of self-management, 5GTANGO & OSM constitute the best solution; the rich set of policies and auto-scaling rules available in NSDs allow automating slice operation, since NST/NSST reference the NSD constructions.
- 5G-TRANSFORMER and SLICENET present similar features, because of the similarities of their constructions. They exhibit reasonably good capabilities, and cover (though not fully) most of the expectations which were reported in Paper B. Their pain points are basically two. Firstly, their lack of alignment with GST/NEST and ServiceProfile parameters; instead, they use workarounds that are not standards-compliant, and therefore not openly reusable in other environments outside of the projects. Secondly, their inability to define the capabilities make available for consumption to the customer; this prevents the vertical to retain some control over the slice.
- ONAP constructions are the ones which are most aligned with GSMA and 3GPP SA5 work; indeed, most of SA5 companies also contribute to ONAP. However, they also exhibit weaknesses that are worth mentioning. First, there are four constructions to capture slicing information, which make their management quite unnecessarily complicated; optimization is needed here. Secondly, there is no possibility for a vertical customer to integrate external applications to the slice, making the slice as operator-only construction with no further extensibility. Finally, in what relates to information on slice lifecycle and priority management, there is room for improvement, lagging behind OSM models.

Beyond the state-of-the-art: This thesis will address the design, implementation and validation of a model-based network slice descriptor, according to the conceptual ideas and hypothesis which were captured in Paper B. More specifically,

- the design will overcome the limitations and pain points of existing modeling solutions. In pursuing this goal, special emphasis will be given to i) model simplicity, facing ONAP complexity; ii) full alignment with GST/NEST and ServiceProfile constructions, to make the solutions standards compliant; and iii) topology extensibility with 3rd party components, to facilitate customization by verticals.
- the implementation will be done using a data modelling language, to make the descriptor usable in an orchestration solution.
- the validation will demonstrate the process of creating the network slice descriptor, followed by its onboarding into the catalog. This will be showcased in a PoC, executed on the 5G-VINNI facility.

Related objectives: O2.1, O2.3.

Means of verification: Papers C and E

2.2 Network slicing orchestration in action: provisioning a multi-domain network slice

Provisioning is the set of activities that are needed to go from a service order to an operative network slice. Reported in Paper B, these activities include the translation of customer-facing requirements into operator-facing requirements, and the use of network slice descriptors (together with NSDs and VNFDs) to manage the workflows related to network slice instantiation, configuration, and activation.

As outlined in Section 1.4, different lab simulations and pilots have been generated using the slicing prototypes developed in the different 5G-PPP Phase 2 projects. Many of these outcomes precisely scope slice provisioning, in some cases assuming the need to allocate different slices for different verticals. Despite the notable results reported, there exist a number of limitations that are worth mentioning:

- Most of the prototypes are built out of tailored, non-replicable orchestration solutions. Although there is a consensus on using OSM + Openstack as reference implementation of NFV-MANO framework, the slicing orchestrator/manager running atop are project-specific. This can be observed in Table III-1. The problem of this approach is that the resulting network slice orchestrators/managers are typically non-standards-compliant, with particularities that made them only applicable to their own projects, preventing their usability in any other environment. And for sure, each slicing orchestrator/manager uses different slice models.
- Each pilot is executed in a small-scale testbed, with resources à *la carte,* that serves to demonstrate that the developed solution works for a given vertical. In addition, the orchestration stacks are purposely configured for individual pilots, with (almost) everything readymade beforehand. With this setup, there is no means to prove the validity of the solution in testbeds at a larger scale, with other verticals.
- All the vertical pilots scope one single administrative domain. There are no results that show how the developed slicing orchestration solutions behave in scenarios involving two (or more) administrative domains. As outlined in the Section 3 from Part I, these scenarios cannot be ruled out, given their importance for B2B market, either in the context of private-public networks (part of the slice deployed on the vertical's administrative domain) or in the context of transnational services (when the slice is allocated to a vertical that requires having worldwide coverage).

Beyond the state-of-the-art: This thesis will overcome the limitations identified from the conducted literature review, by validating the provisioning of a multi-site network slice, using the orchestration solution developed in 5G-VINNI facility. The orchestration solution

- makes use of the network slice descriptor, prototyped in O2.1.
- is open-source and standards-compliant, with focus on 3GPP SA5 solutions. This means that the solution is openly replicable in different environments.
- is much easier to use than ONAP, with a more loosely coupled and task-oriented design. It is an alternative to the integration issues that have been reported with ONAP in [60].

- sits atop of OSM, building on the robust NFV and automation capabilities that the latter provides, and complementing them with slice semantics. Examples of OSM capabilities that can be leveraged include the *PLA module* (see Figure III-8) and *quotas management* functionality.
- incorporates federation capabilities, using industry reference APIs to coordinate slice orchestration with other administrative domains. This allow having a slice deployed across multiple sites, when these sites are beyond the footprint of one single operator.

The ambition is to demonstrate how to deploy, configure and activate one network slice deployed across two facilities, managed by different operators.

Ambition: O2.1, O2.2, O2.4.

Means of verification: Papers D and E.

2.3 Network slicing orchestration in action: multi-domain slice auto-scaling

Once the network slice is up and running, it can be resized as needed, so that it can cope with the load surges while making an efficient resource usage. This lifecycle phase is the most difficult one, as it means keeping the slice compliant with SLA in variable environments, with other network slices running in parallel. The dynamism in traffic variations and the need to timely respond to them requires automation as well, from data collection (monitoring) up to corrective actions (e.g., scaling); human operators cannot deal with such short timescales.

Auto scaling is the features whereby a network slice orchestrator continuously compares data collected from running slice against the network status, in order to decide if resizing is needed. If so, it executes the action accordingly, either scaling-out (upsize the slice, allocating more resources) or scaling-in (downsize the slice, releasing allocated resources), automatically. Auto scaling is the third and last ambition of Part III, constituting the natural steps after slice onboarding and provisioning.

According to the literature review conducted in Section 1.4, neither Phase 2 projects nor OSM/ONAP communities have reported results in auto-scaling for a *slice deployed across different administrative domains*. This is the gap that the PhD candidate wants to bridge.

Beyond the state-of-the-art: This thesis will cover the gap identified from the conducted literature review, by demonstrating the auto-scaling feature in a multi-domain slice using the network slice orchestration prototyped in O2.2.

Ambition: O2.2, O2.5

Means of verification: Paper E.

References

[1] A. Sgambelluri et al., "Orchestration of Network Services across multiple operators: The 5G Exchange prototype”, in *2017 European Conference on Networks and Communications* (EuCNC), 2017, pp. 1-5. DOI: 10.1109/EuCNC.2017.7980666.

[2] I. Vaishnavi et al., "Realizing services and slices across multiple operator domains", in

NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1-7. DOI: 10.1109/NOMS.2018.8406168.

[3] S. Dräxler et al., "5G OS: Control and Orchestration of Services on Multi-Domain Heterogeneous 5G Infrastructures", in *2018 European Conference on Networks and Communications (EuCNC)*, 2018, pp. 1-9. DOI: 10.1109/EuCNC.2018.8443210.

[4] 5G-PPP Phase 2 projects [Online]. Link

[5] View on 5G Architecture Version 3.0", 5GPPP Architecture Working Group, June 2019 [Online]. Link.

[6] 5G!Pagoda project website [Online]. Link.

[7] NECOS project website [Online]. Link.

[8] S. Kukliński and L. Tomaszewski, "DASMO: A scalable approach to network slices management and orchestration", in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1-6. DOI: 10.1109/NOMS.2018.8406279.

[9] T. Taleb, I. Afolabi, K. Samdanis and F. Z. Yousaf, "On Multi-Domain Network Slicing Orchestration Architecture and Federated Resource Control" *in IEEE Network*, vol. 33, no. 5, pp. 242-252, Sept.-Oct. 2019. DOI: 10.1109/MNET.2018.1800267.

[10] A. Laghrissi, T. Taleb and M. Bagaa, "Canonical domains for optimal network slice planning", in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1-6. DOI: 10.1109/WCNC.2018.8377336.

[11] J. Prados-Garzon, A. Laghrissi, M. Bagaa, T. Taleb and J. M. Lopez-Soler, "A Complete LTE Mathematical Framework for the Network Slice Planning of the EPC" in *IEEE Transactions on Mobile Computing*, vol. 19, no. 1, pp. 1-14, 1 Jan. 2020. DOI: 10.1109/TMC.2018.2890235.

[12] H. Hantouti, N. Benamar and T. Taleb, "A Novel Compact Header for Traffic Steering in Service Function Chaining", in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1-6. DOI: 10.1109/ICC.2018.8422597.

[13] H. Hantouti, N. Benamar, T. Taleb and A. Laghrissi, "Traffic Steering for Service Function Chaining", in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 487-507, Firstquarter 2019. DOI: 10.1109/COMST.2018.2862404.

[14] I. Benkacem, T. Taleb, M. Bagaa and H. Flinck, "Performance benchmark of transcoding as a virtual network function in CDN as a service slicing", in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1-6. DOI: 10.1109/WCNC.2018.8377402.

[15] I. Benkacem, M. Bagaa, T. Taleb, Q. Nguyen, T. Toshitaka and T. Sato, "Integrated ICN and CDN Slice as a Service", in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1-7. DOI: 10.1109/GLOCOM.2018.8648051.

[16] I. Fotoglou et al., "Towards Cross-Slice Communication for Enhanced Service Delivery at the Network Edge", in *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, 2020, pp. 22-28. DOI: 10.1109/NetSoft48620.2020.9165442.

[17] K. B. Costa, F. S. Dantas Silva, L. M. Schneider, E. P. Neto, A. V. Neto and F. Esposito, "Enhancing Orchestration and Infrastructure Programmability in SDN With NOTORIETY", in *IEEE Access*, vol. 8, pp. 195487-195502, 2020. DOI: 10.1109/ACCESS.2020.3033486.

[18] S. Clayman et al., "The NECOS Approach to End-to-End Cloud-Network Slicing as a Service", in *IEEE Communications Magazine*, vol. 59, no. 3, pp. 91-97, March 2021. DOI: 10.1109/MCOM.001.2000702.

[19] NECOS project demonstrators [Online]. Link

[20] 3GPP TR 28.801, “Telecommunication management; Study on management and orchestration of network slicing for next generation network”, v15.1.0, April 2018.

[21] GSMA, “Network Slicing Use Case Requirements”, 2018 [Online]. Link

[22] GSMA, “Generic Network Slice Template v3.0”, July 2019 [Online]. Link

[23] ETSI GS ZSM 002, “Zero-touch network and Service Management (ZSM); Reference Architecture”, v1.1.1, August 2019 [Online]. Link

[24] ETSI GS ZSM 003, “Zero-touch network and Service Management (ZSM); End-to-end management and orchestration of network slicing”, v1.1.1, June 2021 [Online]. Link

[25] ETSI GS ZSM 009, “Zero-touch network and Service Management (ZSM); Closed-Loop Automation; Part 1: Enablers”, v1.1.1, June 2021 [Online]. Link

[26] 5TONIC lab [Online]. Link

[27] 5G lab Germany [Online]. Link

[28] H2020 5G-EVE project, “5G European Validation platform for Extensive trials” [Online]. Link

[29] H2020 5G-VINNI project, “5G Verticals Innovation Infrastructure” [Online]. Link

[30] H2020 5Genesis project, “5th Generation End-to-end Network, Experimentation, System Integration, and Showcasing” [Online]. Link

[31] M. R. Raza et al., "Demonstration of Resource Orchestration Using Big Data Analytics for Dynamic Slicing in 5G Networks," in *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1-3. DOI: 10.1109/ECOC.2018.8535466.

[32] J. Pérez-Romero, O. Sallent, R. Ferrús and R. Agustí, "On the configuration of radio resource management in a sliced RAN," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1-6. DOI: 10.1109/NOMS.2018.8406280.

[33] R. Ferrús, O. Sallent, J. Pérez-Romero and R. Agusti, "On the Automation of RAN Slicing Provisioning and Cell Planning in NG-RAN," in *2018 European Conference on Networks and Communications (EuCNC)*, 2018, pp. 37-42. DOI: 10.1109/EuCNC.2018.8442690.

[34] L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, H. D. Schotten and X. Costa-Pérez: "LACO: A Latency-Driven Network Slicing Orchestration in Beyond-5G Networks", in *IEEE Transactions on Wireless Communications (TWC)*, vol. 20, no. 1, pp. 667-682, Jan. 2021. DOI: 10.1109/TWC.2020.3027963.

[35] L. Zanzi, J. Xavier Salvat, V. Sciancalepore, A. Garcia-Saavedra and X. Costa-Perez, "Overbooking Network Slices End-to-End: Implementation and Demonstration," in *Proceedings of the ACM SIGCOMM 2018 Conference on Posters and Demos*, pp. 144-146. DOI: 10.1145/3234200.3234230.

[36] Salvat, J.X., Zanzi, L., García Saavedra, A., Sciancalepore, V. y Costa Pérez, X., “Overbooking Network Slices through Yield-driven End-to-End Orchestration”, in *CoNEXT '18: Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, 2018, pp. 353-365. DOI: 10.1145/3281411.3281435.

[37] A. Sgambelluri et al., "Experimental Demonstration of a 5G Network Slice Deployment Exploiting Edge or Cloud Data-Centers", in *2019 Optical Fiber Communications Conference and Exhibition* (OFC), 2019, pp. 1-3.

[38] 5G-TRANSFORMER demo, "Creating a media-oriented slice through the 5G-Transformer vertical slicer”, 2018. [Online]. Link

[39] A. Boubendir et al., "5G Edge Resource Federation: Dynamic and Cross-domain Network Slice Deployment", in *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, 2018, pp. 338-340. DOI: 10.1109/NETSOFT.2018.8460118.

[40] B. Han, J. Lianghai and H. D. Schotten, "Slice as an Evolutionary Service: Genetic Optimization for Inter-Slice Resource Management in 5G Networks", in *IEEE Access*, vol. 6, pp. 33137-33147, 2018. DOI: 10.1109/ACCESS.2018.2846543.

[41] P. Caballero, A. Banchs, G. de Veciana, X. Costa-Pérez and A. Azcorra, "Network Slicing for Guaranteed Rate Services: Admission Control and Resource Allocation Games", in *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6419-6432, Oct. 2018. DOI: 10.1109/TWC.2018.2859918.

[42] G. Garcia-Aviles, M. Gramaglia, P. Serrano and A. Banchs, "POSENS: A Practical Open Source Solution for End-to-End Network Slicing", in *IEEE Wireless Communications*, vol. 25, no. 5, pp. 30-37, October 2018. DOI: 10.1109/MWC.2018.1800050.

[43] “Smart Use Port Use Case”, 5G-MonArch pilot, 2019 [Online]. Link

[44] “Touristic City Use Case”, 5G-MonArch pilot, 2019 [Online]. Link

[45] Nadal et al., "Multi-vendor sliceable transceivers in partial disaggregated metro networks," in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1-4. DOI: 10.1049/cp.2019.0878.

[46] R. Casellas, A. Giorgetti, R. Morro, R. Martinez, R. Vilalta and R. Munoz, "Virtualization of disaggregated optical networks with open data models in support of network slicing", in *Journal of Optical Communications and Networking, vol. 12, no. 2, pp. A144-A154,* February 2020. DOI: 10.1364/JOCN.12.00A144.

[47] M. B. Weiss, A. Gavras, P. Salva-Garcia, J. M. Alcaraz-Calero and Q. Wang, "Network Management - Edge and Cloud Computing The SliceNet Case," *in 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC),* 2020, pp. 1-6. DOI: 10.1109/CCNC46108.2020.9045678.

[48] M. Roddy et al., "5G Network Slicing for Mission-critical use cases," in *2019 IEEE 2nd 5G World Forum (5GWF),* 2019, pp. 409-414. DOI: 10.1109/5GWF.2019.8911651.

[49] R. Ricart-Sanchez, A. C. Aleixo, Q. Wang and J. M. Alcaraz Calero, "Hardware-Based Network Slicing for Supporting Smart Grids Self-Healing over 5G Networks," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1-6. DOI: 10.1109/ICCWorkshops49005.2020.9145088.

[50] P. Alemany et al., "Experimental Validation of Network Slicing Management for Vertical Applications on Multimedia Real-Time Communications over a Packet/Optical Network," in 2019 21st International Conference on Transparent Optical Networks (ICTON), 2019, pp. 1-4. DOI:1109/ICTON.2019.8840187.

[51] B. Rusti et al., "5G Smart City Vertical Slice" in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2019, pp. 13-19.

[52] B. Rusti, H. Stefanescu, M. Iordache, J. Ghenta, C. Brezeanu and C. Patachia, "Deploying Smart City components for 5G network slicing," in *2019 European Conference on Networks and Communications (EuCNC),* 2019, pp. 149-154. DOI: 10.1109/EuCNC.2019.8802054.

[53] 5G Picture project Deliverable 6.3, "D6.3 Final Demo and Testbed experimentation results," May 2020 [Online]. Link

[54] 5GCity White Paper, “Enabling 5G Neutral Hosts: 5GCity Architecture and Business Models”, April 2020 [Online]. Link

[55] S. Seetharaman et. al. “ONAP Network Slicing: PoC proposal for Frankfurt”, Sept 2019 [Online]. Link

[56] ONAP, “E2E Network Slicing Use Case in R6 Frankfurt”, March 2020 [Online]. Link

[57] ONAP, “E2E Network Slicing Use case in R7 Guilin”, Sept 2020 [Online]. Link

[58] ONAP, “ONAP E2E Network Slicing Technical Overview – Providing End-to-end 5G Network Slicing Capability”, Dec 2020 [Online]. Link

[59] ONAP, “ONAP E2E Network Slicing Technical Overview – Providing End-to-end 5G Network Slicing Capability”, June 2020 [Online]. Link

[60] J. Collins, “ONAP is no Automation Panacea, quite the opposite by design”, March 2021 [Online]. Link

[61] ETSI OSM, “OSM Information Model” [Online]. Link

[62] ETSI OSM, “OSM PoC #6 – 5G Network Slice Orchestration with OSM”, 20190 [Online]. Link

[63] F. Díaz, “OSM#10 Hackfest – HD1.5 – Network Slicing”, 2021 [Online]. Link

[64] A. de la Oliva et al., “5G-TRANSFORMER: Slicing and orchestrating transport networks for industry verticals”, in *IEEE Communications Magazine*, vol. 56, no. 8, pp. 78-84, August 2018.

[65] SLICENET, “H2020 5G-PPP project SLICENET (End-to-End Cognitive Network Slicing and Slice Management Framework in Virtualised Multi-Domain, Multi-Tenant 5G Networks)”

[66] ETSI GR NFV-EVE 012, “Network Functions Virtualization (NFV) Release 3; Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework”, v3.1.1, December 2017 [Online]. Link

[67] ONAP, “Harmonizing Open Source and Standards: A case for 5G Slicing”, 2019 [Online]. Link

Paper C

Modelling Network Slice as a Service in a multi-vendor 5G experimentation ecosystem

Jose Ordonez-Lucena, Christos Tranoris and João Rodrigues

Disclaimer:
This work has been published in the IEEE International Conference on Communications (ICC), 2020.
Citation information: DOI 10.1109/ICCWorkshops49005.2020.9145225

Abstract

As the 5G standards mature and awareness of the capabilities of the technology increases, industry verticals are becoming more eager to test new services and develop them to the level of maturity required for market adoption. Consequently, there is a growing urgency to provide a realistic 5G test and experimentation platform that is open to verticals, and which supports rapid and agile testing of real-world use cases. The realism implies having a distributed multi-domain infrastructure, involving technologies and solutions from different vendors. This paper provides insights on how to allow advanced vertical experimentation in a trial end-to-end 5G facility with the above features, using Network Slice-as-a-Service for this end. Among other enablers, this includes the design of reusable, model-based service templates, their publication into a service catalog, and the exposure of open APIs towards the verticals supporting functions for creating, modifying and decommissioning network slices. All these mechanisms will be presented in this paper, based on the experience gained in the 5G-VINNI funded project.

1 Introduction

For over the last years, the telco industry has led standardization and technology exploration activities to accelerate the roll-out of fifth generation (5G) systems across the world. Unlike the past generation of mobile communications (4G), mainly focused on providing mobile broadband services to end users, 5G may offer programmable service platforms able to connect a wide variety of devices in a ubiquitous manner. These devices go beyond end user's smartphones, including vehicles, Internet-of-Things (IoT) appliances and industry 4.0 equipment (e.g., manufacturing robots, augmented reality-enabled tablets for remote worker operation). The irruption of industry verticals into the 5G ecosystem bring a new wave of use cases, with very different requirements in terms of performance (e.g., throughput, latency, reliability) and functionality (e.g., mobility, security, service continuity support), some of them very stringent [1].

To satisfy all the above service requirements in a cost-effective manner, network operators need to turn their networks into flexible, programmable multi-service platforms with the help of latest network softwarization technologies, including network slicing. To do that, and following current telco practices, each network operator may rely on the technology solutions developed by multiple vendors, typically providing operators with the resources (e.g., physical and virtualized network functions) and orchestration tools required to operate their networks. This approach will yield to the roll-out of multi-vendor 5G networks [2].

Before launching production-ready 5G networks with the above capabilities (e.g., interoperability across vendors and slicing support), testing and validation activities in large-scale testbeds involving vendors, network operators and industry verticals are required. For this end, and to accelerate the uptake of 5G in Europe, the 5G Private Public Partnership (5G-PPP) has financed three flagship projects: 5G-VINNI, 5G-EVE and 5GENESIS. These projects aim to provide a pan-Europe 5G validation network infrastructure able to demonstrate that the 5G Key Performance Indicators (KPIs) can be met, and that can be accessed and used by industry verticals to execute trials of innovative use cases, testing and validating specific applications that are dependent upon those KPIs. As shown in Fig. 1, the result of combining their facilities is a large-scale 5G service platform to set up a multiplicity of vertical use cases, covering about 20 EU sites and nodes on a pan-EU basis. In this work, we will focus on the

5G-VINNI project.

5G-VINNI is a large-scale, end-to-end facility providing advanced 5G capabilities that are made available to industry verticals for use case trialing [3]-[4]. This facility provides every vertical with an isolated service experimentation platform, deployed in the form of a slice. This means that each vertical will use the provided slice to set up one or more use cases, assessing their KPIs under different load conditions through the execution of a set of tests. According to Figure C-1, 5G-VINNI facility is composed of several interworking sites, each deployed at a different geographic location and defining a single administrative domain. Every 5G-VINNI facility site includes the following components:

- A Service Orchestrator, taking care of the lifecycle management of provided network slices at the application layer, i.e., slice semantics.
- A Network Orchestrator, which deploys and operates provided slices at the resource layer.
- Infrastructure resources, including access, transport, and core network functions. Some of these functions are physical, while some of them are executed on cloud environments.

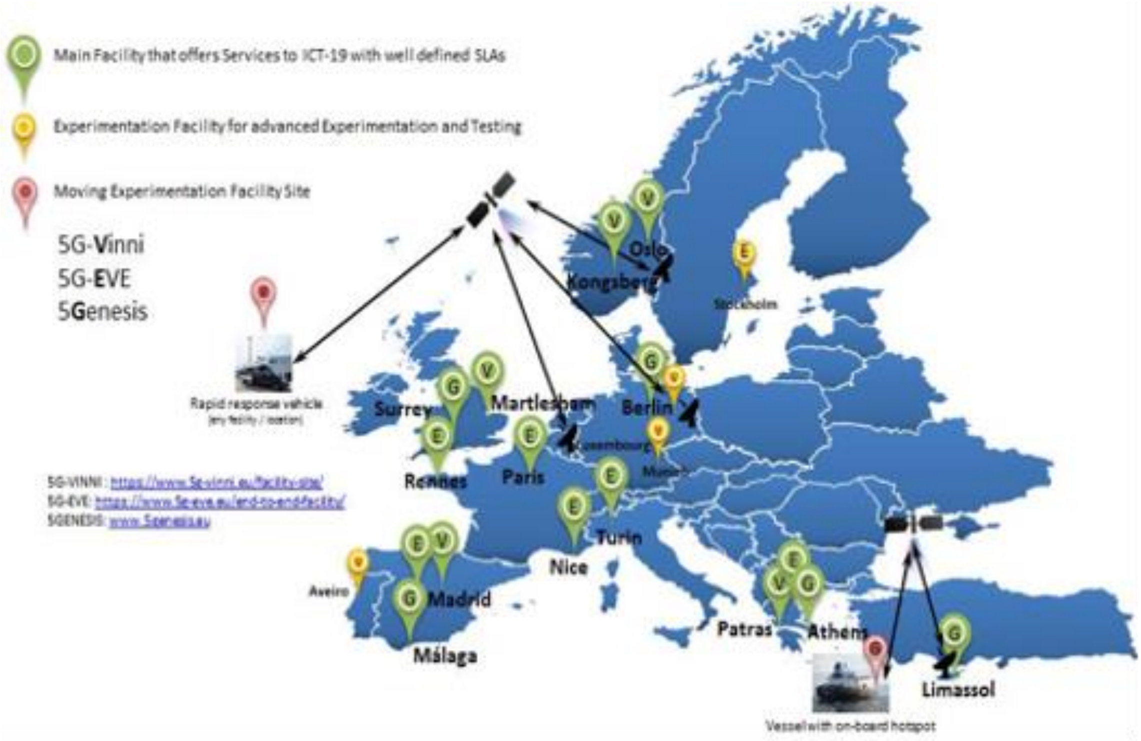

Main Facility that offers Services to ICT-19 with well defined SLAs
Experimentation Facility for advanced Experimentation and Testing
Moving Experimentation Facility Site
5G-Vinni
5G-EVE
5Genesis
Rapid response vehicle
(any facility / location)
5G-VINNI : https://www.5g-vinni.eu/facility-site/
5G-EVE: https://www.5g-eve.eu/end-to-end-facility/
5GENESIS: www.5genesis.eu
V
V
Oslo
Kongsberg
E
Stockholm
G
V
Berlin
G
V
Surrey
Martlesham
E
Rennes
E
Paris
Luxembourg
V
Munich
E
Turin
E
Nice
E
V
V
Aveiro
G
Madrid
Málaga
E
V
G
Patras
Athens
G
Limassol
Vessel with on-board hotspot

Figure C-1. 5G-PPP flagship infrastructure projects – Geography cartography.

All the components defined in a given site are from the same or different vendors, and all managed by a single operator. Despite this per-site description, the whole facility shall be viewed as a single platform from the verticals' side. This brings the need to implement adaptation layers, which unify the behavior and features offered in the different sites, abstracting underlying implementation details from each site. In this context, 5G-VINNI provides verticals with a single entry-point to the facility by means of a portal. As shown in Figure C-13, this portal allows any vertical to browse the service catalog and trigger corresponding service orders. On the one hand, the service catalog is used by 5G-VINNI to announce its service offerings, which are network slices to be delivered under the Network Slice as-a-Service (NSaaS) model [5]. On the other hand, service ordering is triggered by the vertical, and consists in selecting a service offering and issuing the request towards the 5G-

VINNI facility. In this request, the vertical can specify where he wants the slice deployed: in a single site or across different sites. In the latter, the portal should split (decompose) the service orders across requested sites.

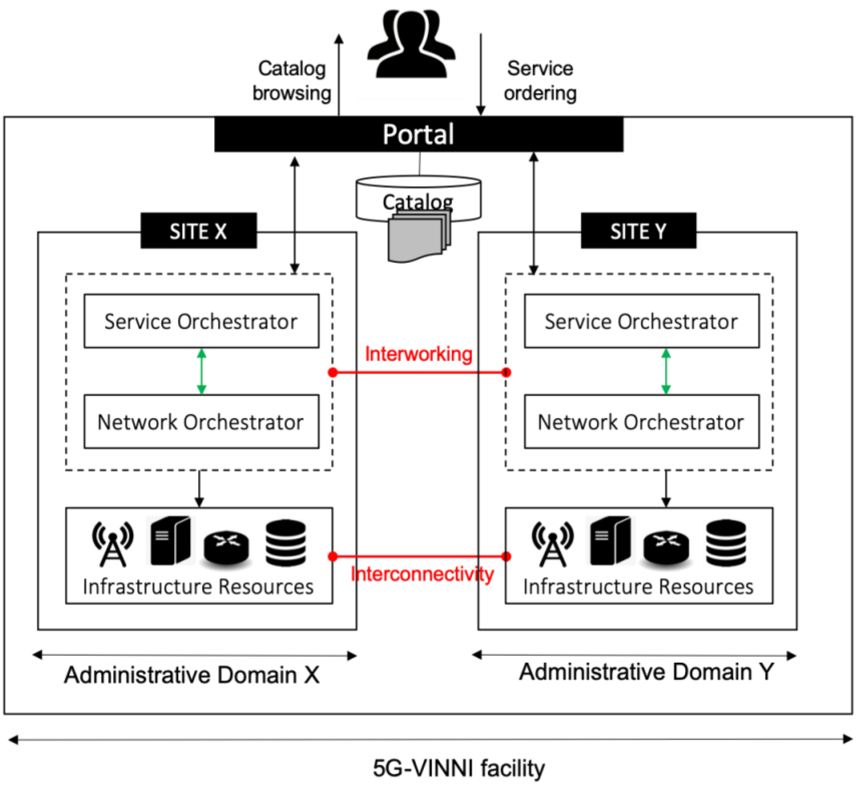


Figure C-13. 5G-VINNI facility baseline architecture

This work intends to describe the enablers and procedures to allow vertical experimentation in a 5G-ready facility like 5G-VINNI that provides NSaaS, and that span across multiple administrative domains, each being a multi-vendor environment under the management of a single operator.

This paper is organized as follows. Section 2 presents a status quo of slicing enablers, including on-going work from telecom industry and standards organizations, and discuss their applicability to 5G-VINNI. Section 3 describes the NSaaS-based models used to describe 5G-VINNI service offerings, including details on their implementation. Sections 4 and 5 focus on the service catalogue and service portal, respectively, describing the APIs offered to allow verticals to interact with the overall facility. Finally, Section 6 is used to summarize the conclusions of this work and provide insight on next steps.

2 Related Work

This section provides a summary of the relevant work in relation to network slicing, focusing on how this mechanism could be useful in vertical experimentation environments, i.e., turning deployed slices into service platforms where verticals can set up trials of different use cases. This will be discussed and analyzed in relation to the activities and outcomes from telecom industry organizations (Section 2.1), standards bodies (Section 2.2), and their applicability to 5G-VINNI project (Section 2.3).

2.1 Telecom Industry Organizations

The mission of these organizations is to collect information on service requirements from different vertical industry alliances (e.g., 5G-ACIA for Industry 4.0, 5GAA for automotive sector), identify potential technologies that can satisfy these requirements, and inform corresponding standards bodies, so that they can develop appropriate technology solutions. Relevant telecom industry organizations include Next Generation Mobile Networks (NGMN) Alliance, GSM Alliance (GSMA), Tele Management Forum (TM Forum) and Metro Ethernet Forum (MEF). In the context of the present paper, we will focus on GSMA and TM Forum.

On the one hand, GSMA is a trade body representing the interests of mobile operators worldwide. GSMA's work on network slicing focuses on how to translate service requirements from industry use cases into network slice requirements, addressing the existing gap between vertical and telco industries. To offer verticals some guidelines on how to issue their service requirements towards operators, the GSMA has defined the Generic Network Slice Template (GST) [6]. The GST is a set of attributes that can be used to characterize any slice in terms of performance, functionality, operation, and scalability. These attributes can be used by the operator and verticals to agree on an SLA. At this stage, 67 attributes have been defined in GST (see [6]).

On the other hand, TM Forum is actively working on the digital transformation and evolution of current Operations / Business Support Systems (OSS/BSS), seeking solutions that facilitate i) their consumption by verticals and ii) their integration into existing standards-defined architectural frameworks. In this respect, one of TM Forum's key contributions is the definition of the Open Digital Architecture [7], which is a multi-layer service platform that can be used by the operators to deliver XaaS, where X refers to the resource under consideration (e.g., infrastructure resource, network function, network service, etc.). To allow a given vertical to consume XaaS, the network operator provides him with corresponding resource management capabilities, offering them in the form of *Open APIs*. Defined in the TM Forum's Open API program, these APIs present two key features. First, they are vendor-agnostic, which means they are not tied to vendor-specific management solutions. Secondly, they can be flexibly defined across layers; indeed, Open APIs at a given layer result from the composition of open APIs from the layer immediately below, following recursive patterns. These two features make Open APIs a good candidate for fast integration of components in multi-vendor environments, guaranteeing reusability and interoperability in the heterogenous 5G ecosystem.

2.2 Standard Development Organizations

Although there exist several standards bodies in telco ecosystem, two of them are key in the context of network slicing: the Third Generation Partnership Project (3GPP), focused on developing solutions and mechanisms for the support of slicing in mobile networks; and the European Telecommunications Standards Institute (ETSI), dealing with aspects related to the deployment of slices in virtualized environments.

3GPP consists of different working groups. One of them is SA5, in charge of specifying requirements, mechanisms and procedures for the provisioning of network slices in 5G mobile networks. This provisioning is done with a service-based management architecture [8], which executes the lifecycle of multiple network slices based on a well-defined information model [9]. This information model describes the functional components of the slice (network functions arranged into one or more network slice subnets) and allows specifying the service requirements it shall support. For the latter, the attributes defined in

ServiceProfile struct are used. These attributes are translation of GST attributes into 3GPP domain.

ETSI is formed of a wide variety of industry specification groups, among which NFV plays a key role. This group defines a management and orchestration stack that allows the deployment and operation of network services, each defined as a composition of one or more Virtualized Network Functions (VNFs) that can be flexibly allocated in different cloud sites. The core component of this stack is the NFV Orchestration (NFVO), which handles the lifecycle management of network services and their VNFs based on the information retrieved from their descriptors, namely Network Service Descriptors (NSDs) [10] and VNF Descriptors (VNFDs) [11].

The relationship between 3GPP and NFV is as follows: the virtualized part of a slice can be deployed as a network service. This means mapping ServiceProfile into the corresponding NSD that NFVO can consume and understand. For communication between 3GPP and NFV, the NFVO offers a northbound interface [12] with multiple capabilities, including NSD management as well as network service lifecycle, performance and fault management.

2.3 Network Slicing in 5G-VINNI facility

5G-VINNI project adopts NSaaS as a service delivery model, whereby the entire facility provisions end-to-end network slices to verticals upon request. Every vertical makes use of the provided slice to test and validate their use cases. To ensure reproducibility (ability to replicate tests across different sites to assess KPIs under different load conditions) and interoperability (ability to ensure the interworking between the tools deployed on each site, typically from different vendors), 5G-VINNI facility architecture incorporates the guidelines form telecom industry organizations, and the normative specifications from standards bodies. This means that:

- The 5G-VINNI portal and service catalog offers TM Forum Open APIs, making them available for verticals, so they can invoke relevant NSaaS operations.
- The Service Orchestrator implements the 3GPP network slice management functionality, taking the roles of both Network Slice Management Function (NSMF) and Network Slice Subnet Management Function (NSSMF).
- The Network Orchestrator is built up with a NFVO, which orchestrates instances of VNFs and network services with the help of a Virtualized Infrastructure Manager (VIM). The VIM manages the virtual resources on top of which those instances are executed.
- The NFVO offers the northbound interface [12] towards the Service Orchestrator using ETSI SOL005 [13]. This normative specification defines the protocol and data model for the interface capabilities, in the form of RESTful APIs. These APIs have become de-facto solutions for most industry and open-source NFVOs.

3 Network Slices Modelling

To allow a vertical to specify the characteristics of the slice he wants, 5G-VINNI defines a baseline, model-service template called 5G-VINNI Service Blueprint (VINNI-SB). The VINNI-SB defines a common information model for the whole 5G-VINNI facility, in order to enable a site-agnostic design of network slices. This not only enables reproducibility (the ability to generate repeatable instances of a given slice at different sites), but also cross-site deployments.

In this section, we provide an insight on the VINNI-SB, giving details on its content (Section 3.1) and its implementation (Section 3.2). For more details on VINNI-SB, see [15].

3.1 The scope of VINNI-SB

Figure C-14 summarizes the structure of the VINNI-SB. As shown, VINNI-SB content is arranged into four main parts.

Slice Service Type (SST), which is a 3GPP parameter [14] that specifies the 5G service category the slice is meant to support. The following SST values apply: "1" (eMBB), "2" (uRLLC), "3" (mIoT) and "4" (V2X).

Slice topology, which provides information on how the slice is constructed from a logical viewpoint. This specifies i) what nodes the slice to be provided consists of, including information on their individual functionality; and ii) how these nodes are connected with each other along the entire topology, including information on their connectivity type. Each node taking part in the slice topology provides a well-defined functionality. Examples of these nodes include 3GPP components (i.e., gNodeB, 5G core control plane, user plane function), value-added service functions (i.e., firewalls, NAT) and edge applications. Depending on selected NFV criteria design, a node can be mapped to a network service or a VNF. The slice default topology can be flexibly extended, offering the verticals the opportunity to bring their own VNFs and applications (i.e., 3rd party VNFs) into slice definition, by simply attaching them to defined slice access points (red connection points in Figure C-14).

Slice attributes, which corresponds to GST attributes. The specification of values for these attributes allows the vertical to provide the service requirements the slice must satisfy, based on which SLA can be defined.

Slice testing and monitoring, relevant in experimentation environments. This part allows specifying the network capabilities the vertical needs to get from 5G-VINNI facility to execute use case trials at run-time. On the one hand, slice testing allows the vertical to specify the tools he needs from 5G-VINNI test framework to execute trialing activities within the slice. On the other hand, slice monitoring allows the vertical to specify telemetry information he wants to get from 5G-VINNI monitoring framework, including data sources and metrics collection method (threshold-based alarms or periodic notifications).

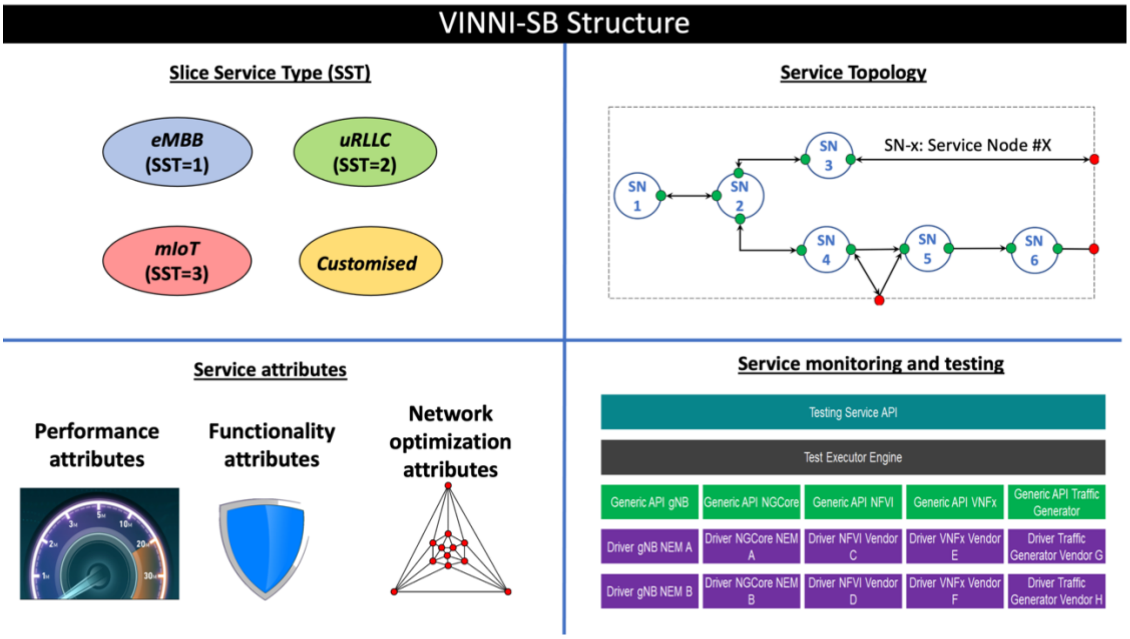


Figure C-14. An overview of VINNI-SB.

With the above-referred VINNI-SB content, it is possible to provide a customer-facing service (CFS) description of any network slice. This specification allows specifying the functionality and configuration settings of the ordered slice at the application layer, which is what the vertical typically is interested in (and understands). The next step is to map this CFS description into a resource-facing service (RFS) description of the slice, which provides details on how the slice is deployed at the resource layer. This means translating slice nodes and slice attributes into concrete NSD/VNFDs along with instantiation information, e.g., placement of network services/VNFs and their resource allocation. This will be done per facility site, based on the interactions between the SO and the NFVO.

3.2 The model of VINNI-SB

Despite the multiple approaches that can be used to define a model for the VINNI-SB, the well-known TM Forum's Information Framework (SID) [16] is the best positioned for this end. Following SID, the VINNI-SB is a top-level construction that can be modelled as a bundle of classes, each defining the invariant characteristics and behavior (attributes, methods, constraints and relationships) of a component taking part in the structure of that top-level construction. For VINNI-SB, the following classes have been defined: *sliceTopology*, *sliceAttributes*, *sliceTopology*, *slice3rdPartyVNFs*, *sliceTesting* and *sliceMonitoring*.

Figure C-15 provides a Unified Modelling Language (UML) representation of the SID-driven VINNI-SB model. As shown, the classes are of "ServiceSpec" type, meaning they represent CFS aspects. In addition to these classes, the VINNI-SB model also includes other types of classes, known as "LogicalResourceSpec" in SID terminology, and that represent RFS aspects of the slice. Relationships between both types of classes allow establishing mappings between CFS and RFS. For example, *slice3rdPartyVNF* class has 1:1 relationship (reference, pointer) with the VNFD class, which represents the VNF information model as stated in [11].

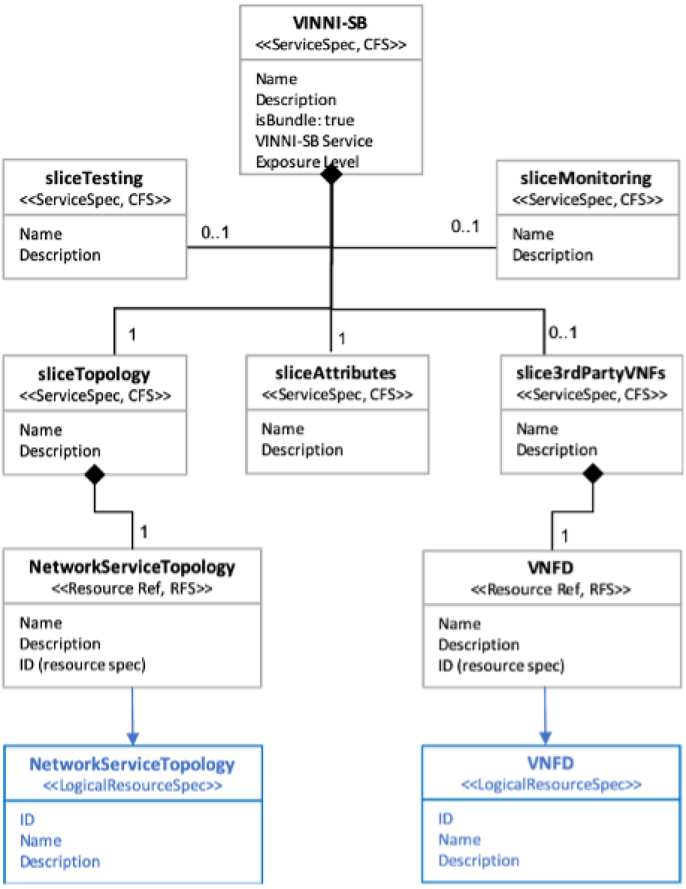


Figure C-15. UML diagram for VINNI-SB model.

4 5G-VINNI Service Catalog

Every facility site operator designs their VINNI-SBs, according to the features deployed on their administrative domains. For example, there are some sites that do not have edge computing nodes, and hence are unable to offer a uRLLC VINNI-SB. 5G-VINNI facility collects the VINNI-SBs from the different sites and registers them into a single service catalog. This approach allows providing verticals with a unified marketplace, informing them about available service offerings.

Any vertical is in position to browse the service catalog, select the VINNI-SB that best fits his needs, and issue a service order. This service order will be translated into a network slice instance, deployed across one or more facility sites.

To facilitate the handling of service catalogue, from VINNI-SB design to vertical-triggered service ordering, 5G-VINNI makes use of three TM Forum Open APIs:

- **Service Catalog API (TMF633),** which provides artifacts (e.g., models and dependencies) for the specification of VINNI-SBs, and with capabilities for their lifecycle management (e.g., registration, deletion, updating, etc.) in the service catalog.
- **Service Ordering API (TMF661),** which allows issuing a service order, which includes selected VINNI-SB and instantiation parameters.
- **Service Inventory API (TMF641),** which defines standardized mechanisms for CRUD operations over the records providing run-time information about deployed slice instances.

As described later, these TM Forum Open APIs may allow verticals to interact with the 5G-VINNI service catalog and consume the exposed capabilities.

5 5G-VINNI Portal

The portal provides a single entry-point for 5G-VINNI facility. Its mission is to expose the facility to verticals as a unified service platform rather than an interconnection of individual administrative domains, hiding the mechanisms, protocols and technologies adopted in every site for this end.

The portal supports the verticals, i.e., 5G-VINNI customers taking the role of experimenters, in obtaining access to the facility resources and available functionality (VINNI-SBs stored in the service catalog, testing and monitoring tools) for the purposes of use case trialing and KPI validation. The first prototype of this 5G-VINNI portal is Openslice [17], an open-source operation support system (OSS) designed for vertical experimentation. The following subsections provide information on Openslice framework, describing its main components and showing how the role he plays on the deployment of a network slice instance in 5G-VINNI facility.

5.1 Openslice framework

Openslice architecture is shown in Figure C-16. In general, the portal includes a set of loosely coupled modules exchanging messages via a message/routing service bus, following microservice architecture. Based on Apache Camel, this service bus allows communication across the different microservices either via message queues or via ‘publish/subscription’ model. These microservices help 5G-VINNI facility to manage (e.g., authorize, audit) the interactions between verticals and the service catalog, including catalog browsing, service ordering and service inventory related operations. Table C-1 provides a summary of the different microservices deployed in Openslice.

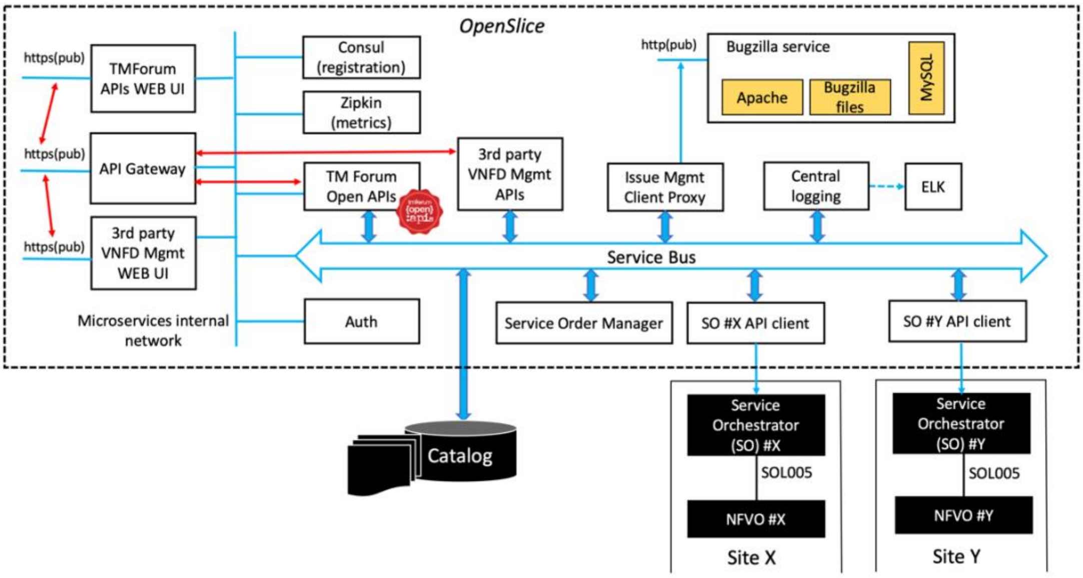


Figure C-16. 5G-VINNI portal – Openslice.

Table C-1. Openslice microservices

μService	Description
Auth	Server to authenticate/authorize verticals via OAuth 2.0
Zipkin	Distributed tracing system that is used to troubleshoot latency problems in mservice architectures.
Consul	One stop solution for typical procedures in μservice architectures, including service (self) registration, discovery, key-value store and load balancing.
Issue mgmt. client proxy	Offers interface to Bugzilla, which is a ticketing tool that allows issue tracking (fault alarms, service orders) and reporting (to verticals, facility operators, etc.) via tickets.
Central Logging	Logs all distributed actions into an ELK (Elasticsearch, Logstash and Kibana) stack.
TM Forum OpenAPIs	Offers TM Forum's OpenAPIs to allow consumption of service catalogue exposed capabilities. These open APIs include Service Catalog, Ordering and Inventory APIs.
3rd party VNFD Mgmt. APIs	Offers NFV APIs to manage 3rd party VNFDs (e.g., on-boarding, updating). These APIs allow verticals to bring their own VNFs to 5G-VINNI, to validate their KPIs.
Service Order Manager	Referred to as SOM, captures service orders triggered by verticals and propagates them to the corresponding SO. In case a service needs to be deployed across two (or more) sites, the SOM shall first decompose the received order.

To allow verticals to gain access to the framework, and hence to 5G-VINNI facility, Openslice offers a web frontend, with two UIs. This web frontend is implemented in Angular and interacts with the Openslice backend API using an API proxy.

5.2 Service Deployment

In this section, we describe the process of going from a service ordering to a deployed slice instance, made available to a vertical for use case testing and validation. In this process, three slice lifecycle phases are covered: slice preparation, slice commissioning (i.e., instantiation) and slice operation.

In the **slice preparation phase**, we consider that the VINNI-SBs describing 5G-VINNI facility service offerings are registered and published in the service catalogue. Under this assumption, the workflow of this phase is as follows. First, the vertical gains access to the 5G-VINNI facility through the portal, using Openslice web frontend UI. Once authenticated with Auth, he browses the 5G-VINNI service catalog using the Service Ordering API and selects the VINNI-SB that best fits his needs. In case the selected VINNI-SB does not include a VNF (e.g., application server) that the vertical needs to validate his use case, he can bring it to 5G-VINNI service catalog, on-boarding the corresponding VNFD. For this end, he relies on the *3rd party VNFD Mgmt. APIs*.

The **slice commissioning phase** begins when the vertical triggers a service order from the selected VINNI-SB. In this service order, the vertical provides a completed specification of the slice instance he wants, including information on slice topology (possibly extended with 3rd party VNFs) and slice attributes (filled in with values fitting use case requirements). To do that, the vertical makes use of the Service Ordering API. The issued service order is then captured by the SOM, which propagates it towards the corresponding site(s), communicating with the API client of the corresponding Service Orchestrator(s). Next, the service order is processed locally, at each site, as follows. This processing consists in translating the received service order (CFS, handled by the Service Orchestrator) into a set of resource requirements for the network slice to be instantiated (RFS, handled by the NFVO). To successfully achieve this translation, Service Orchestrator and NFVO exchange information relying on ETSI SOL005 capabilities. Once this translation is completed, the NFVO allocates the slice instance, instructing the VIM for that end.

In the **slice operation phase**, the slice is already instantiated, and can be made available for vertical experimentation. During this phase, the vertical keeps track of the status of the slice instance, making use of the *Service Inventory API*.

During the above-mentioned phases, all the operations are registered with the *central logging* service. In case there exists any issue, the *Bugzill*a service will come into play.

6 Conclusions and Future Work

As the 5G standards are completed and technical development of 5G products matures, the telco industry shall demonstrate that next-generation carrier networks are ready to satisfy KPIs required for upcoming vertical use cases. For this end, not only 5G capabilities need to be validated, but also interoperability. The latter is critical, considering that these use cases will be executed in heterogeneous environments, involving multiple technologies and actors. 5G-VINNI project provides a first approximation of these future scenarios. In this work we have explained how 5G-VINNI facility, despite being a multi-domain infrastructure involving multiple actors and technologies, is able to provide a unified service platform for vertical experimentation, using NSaaS as service delivery model. To achieve interoperability, common service models (VINNI-SB) and interworking mechanisms (enabled by Openslice) are used.

The first pilots with verticals are about to get started. For this end, 5G-VINNI relies on two orchestration solutions: one open source, which is Open Source MANO; and another proprietary, which is Nokia’s orchestration toolkit (i.e., FlowOne and CloudBand). The next steps are oriented to develop federation interfaces between these two solutions, needed for support use cases deployed across sites using different orchestration tools.

Acknowledgement

This work is supported by the H2020 European Project 5G-VINNI (grant agreement No. 815279).

References

[1] 5G-Americas, “5G Americas White Paper: Management, Orchestration and automation”, Nov 2019.

[2] NGMN Alliance, “5G White Paper”, 2016.

[3] 5G-VINNI project [Online]. Available: https://www.5g-vinni.eu/

[4] K. Mahmood, “Design of 5G End-to-End Facility for Performance Evaluation and Use Case Trials”, *in 2019 IEEE 2nd 5G World Forum (5GWF)*, 2019

[5] 3GPP TS 28.530, “Management and Orchestration; Concepts, use cases and

requirements”.

[6] GSMA PRD NG.116 v2.0, “Generic Network Slice Template”, 2019.

[7] TM Forum, “IG1167 Open Digital Architecture (ODA) Functional Architecture R19.0.1”, 2019.

[8] 3GPP TS 28.533, “Telecommunications management; Management of 5G networks and network slicing; Architecture framework”.

[9] 3GPP TS 28.541, “Management and orchestration; 5G Network resourcece Model (NRM); Stage 2 and 3”.

[10] ETSI GS NFV-IFA 014, “Management and Orchestration; Network Service Templates Specification”.

[11] ETSI GS NFV-IFA 011, “Management and Orchestration; VNF Descriptor and Packaging Specification”.

[12] ETSI GS NFV-IFA 013, “Management and Orchestration; Os-Ma-nfvo reference point – Interface and Information Model Specification”.

[13] ETSI GS NFV-SOL 005, “Protocols and Data Models; RESTful protocols specification for the Os-Ma-nfvo Reference Point”.

[14] 3GPP TS 23.501, “System architecture for the 5G System; Stage 2”

[15] 5G-VINNI Deliverable D3.1, “Specification of services delivered by each of the 5G-VINNI facilities”. [Online]. Available: https://zenodo.org/record/3345612#.XkFBTy2ZOL4

[16] TM Forum TM Forum “GB922 Configuration R14.5.1; Information Framework (SID)

[17] Openslice project [Online]. Available: http://openslice.io

Paper D

Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility

Jose Ordonez-Lucena, Christos Tranoris, João Rodrigues and Luis M. Contreras

Disclaimer:
This work has been published in the European Conference on Networks and Communications (EuCNC), 2020.
Citation information: DOI 10.1109/EuCNC48522.2020.9200940

Abstract

As 5G standards are completed and technical development of 5G products matures, the pressure for mobile operators to launch commercial networks with advanced capabilities (e.g. network slicing) increases. Despite that, industry forecasts suggest that adoption rates of next-generation networks will be slower compared to previous evolutions. Indeed, unlike 4G, in case of 5G there is i) a novel set of customers, i.e. the verticals, which may bring innovative use cases with unprecedented KPIs, and ii) a confluence of novel technologies, developed across different layers and provided by different vendors. Building, running and operating all these new innovations is extremely challenging because of the novelty and the lack of previous experience on integrating them altogether. 5G-VINNI project, based on the provisioning of end-to-end network slices for advanced vertical experimentation using a multi-domain 5G facility infrastructure, has been born to explore these novelties. This paper addresses the problem of cross-domain slice orchestration, proposing a federated-oriented, standards-based solution to allow transparent interoperability and interworking between different domains, each using a distinct orchestration solution.

1 Introduction

The fifth generation (5G) of mobile networks have been targeted to meet the requirements of a highly mobile and fully connected society. First 5G specifications are available with 3GPP Rel-15, where the focus has been primarily to satisfy traditional end-users needs in terms of extreme mobile broadband services, by providing more throughput and reduced latency. However, this is only the initial step, as further enhancements and optimizations are still needed to design a 5G system able to meet the challenging requirements from vertical industries (e.g., industry 4.0, agriculture, automotive, smart cities, energy), which are at the forefront of the ongoing digital transformation. Industry verticals, i.e. companies and organizations from these non-telco industries, will be the main beneficiaries of 5G, using provided capabilities to develop new services and application scenarios.

The result of integrating multiple industry verticals into the 5G system will be the coexistence of a wide variety of use cases on top of the same infrastructure, some of them with very different requirements in terms of performance and functionality. End-to-end slicing, service-based architecture, Software Defined Networking (SDN) and Network Functions Virtualization (NFV) are seen as the fundamental pillars to support the heterogeneous key performance indicators (KPIs) of these use cases in a cost-effective manner [1]. The use of these technologies will allow operators to transform today's monolithic networks into flexible, programmable service platforms with multi-tenancy features.

As these technologies reach a maturity level beyond pure laboratory experiments, operators and vendors work together in bilateral trials to accelerate their adoption in the market. These trials, typically carried out at local and limited environments, are not enough to validate and assess the readiness of these technologies before launching them in commercial 5G networks, taking into account the *end-to-end* and *multi-domain* nature of these networks. The end-to-end (E2E) feature means the services will potentially span across all network segments, including access, transport, core, edge and cloud segments. The multi-domain feature means the network will involve multiple administrative domains, each deploying a number of technologies based on both vendor-specific and open-source solutions. Every administrative domain will be operated by a single network operator, who may rely on the technology solutions supplied by multiple vendors. This will yield to the roll-out of heterogeneous 5G networks, with multiple actors involved.

Although some testbeds following the above premises have been already deployed in previous research projects (e.g., 5G-Exchange), the implications of having *different orchestration solutions on different domains* have not been studied. What is more, the engagement of industry verticals bringing real-world use cases has been rather scarce. Although it is true that some projects have verticals into their consortia (e.g., 5G-Transformer and 5G-Monarch), the *readiness of these verticals to deploy and orchestrate use cases when the operators provide them with different capability exposure levels* has not been assessed. This is crucial for a future-proof 5G ecosystem, which calls for a seamless integration between verticals (OT industry) and operators (ICT industry).

With the above reasoning, it is clear that before the roll-out of commercial 5G with the desired features (e.g., interoperability and slicing), intensive testing and validation activities in large-scale testbeds involving vendors, network operators and verticals are required. To this end, the 5G Private Public Partnership (5G-PPP) has funded three flagship projects: 5G-VINNI, 5G-EVE and 5GENESIS. These projects aim to provide a set of pan-Europe 5G validation facilities that permit experimentation of vertical services in close-to-production scenarios, ensuring that the 5G KPIs can be met. In this work, we focus on 5G-VINNI.

5G-VINNI is a large-scale E2E facility that provides 5G capabilities for advanced vertical experimentation in multi-domain environments [2]. These capabilities are made available for consumption using Network Slice as a Service (NSaaS). This service delivery model allows a vertical to use the provided network slice for validation activities, deploying one or more use cases and validating their KPIs under different load conditions.

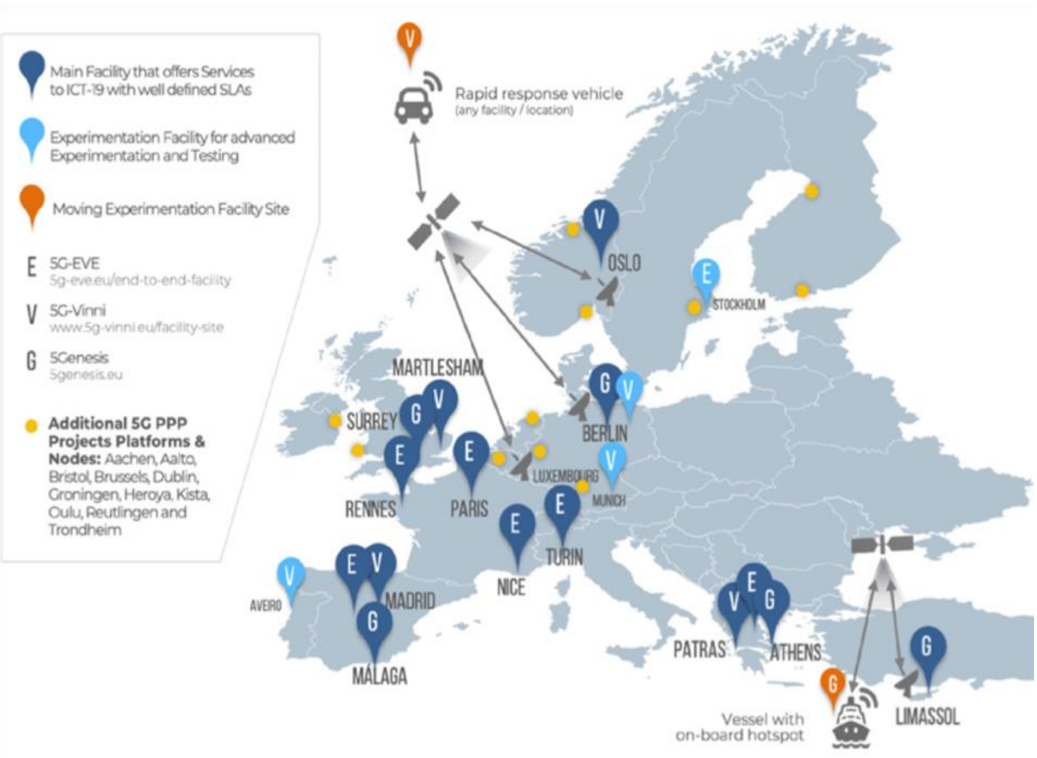

Main Facility that offers Services to ICT-19 with well defined SLAs
Experimentation Facility for advanced Experimentation and Testing
Moving Experimentation Facility Site
E 5G-EVE 5g-eve.eu/end-to-end-facility
V 5G-Vinni www.5g-vinni.eu/facility-site
G 5Genesis 5genesis.eu
Additional 5G PPP Projects Platforms & Nodes: Aachen, Aalto, Bristol, Brussels, Dublin, Groningen, Heroya, Kista, Oulu, Reutlingen and Trondheim
Rapid response vehicle (any facility / location)
V
OSLO
STOCKHOLM
MARTLESHAM
SURREY
BERLIN
LUXEMBOURG
MUNICH
RENNES
PARIS
NICE
TURIN
AVEIRO
MADRID
MÁLAGA
PATRAS
ATHENS
LIMASSOL
Vessel with on-board hotspot

Figure D-17. 5G-PPP flagship infrastructure projects – Geographic cartography.

Figure D-17 shows the 5G-VINNI facility. Despite of being formed through the combination of multiple smaller facilities (i.e., *facility sites*, each defining a single administrative domain), like in 5G commercial networks, it is necessary to orchestrate the entire 5G-VINNI facility as a single and unified entity independently on the particular placement of its components. This cross-domain orchestration procedure requires common standard interfaces and information models across those domains to enable the interoperability among the multi-vendor solutions adopted in each segment. This constitutes a real challenge, since most of the NFV, MEC and SDN solutions available today from vendors or open-source communities expose proprietary interfaces, which refer to non-standard information models. A well-founded discussion of this lack of convergence can be found in [3].

In this paper, we will address the above problem considering 5G-VINNI, which provides a 5G-ready E2E facility with multiple operators and technology solutions involved. We will focus on the problem of cross-domain slice orchestration, assuming the deployment and operation of slice instances across multiple facility sites, each making use of a different orchestration solution. Specifically, we will consider the use of two solutions: Open Source MANO (OSM) and Nokia's orchestration toolkit. In this environment, federation solutions to allow transparent interoperability between domains (i.e., without forcing the proprietary interfaces in all the different sites) will be proposed, discussed and developed.

The remainder of this article is as follows. Section 2 provides a state-of-the-art in federation approaches for multi-domain environments. Section 3 describes the 5G-VINNI facility architecture. Section 4 discuss the applicability and realization of federation in the 5G-VINNI facility, with a particular example shown in Section 5. Finally, Section 6 summarizes the main outcomes of this paper and informs of future work.

2 Related Work

Service provisioning in multi-domain environments brings new challenges in terms of interoperability, considering that different administrative domains may be equipped with different orchestration systems. Strategies, roadmaps, and solutions to address these challenges are within the scope of recent initiatives like ETSI Zero touch network & Service Management (ZSM).

To ensure unified orchestration of E2E slices in the abovementioned environments, as pursued in this work, the tools deployed on the different orchestration systems need to communicate with each other using standard reference points and common information models. These two assets allow the definition of federation interfaces, i.e. interfaces whereby two or more administrative domains can securely exchange data and operations under 'producer-consumer' role model.

Different solutions have been proposed in the standards bodies for the design of these interfaces. One of the precursors was ETSI NFV, the industry specification group (ISG) in charge of developing specifications for the deployment and operation of network services in virtualized environments. In [4], ETSI NFV reports on potential architecture options to support the offering of orchestration services across multiple domains. These options include both hierarchical and peer-to-peer approaches. Unlike the former, based on the reusability of carrier-grade interfaces like Or-Vi or Or-Vnfm, peer-to-peer approaches require the definition of novel (not yet mature) interfaces like Or-Or [5].

Metro Ethernet Forum (MEF) pretends to go a step beyond, scoping also non-virtualized environments and considering federation at multiple layers (e.g., business, services, resources), These features are illustrated in the Lifecycle Service Orchestration (LSO) reference architecture [6]. As seen in Figure D-2, this architecture considers three types of domains for the definition of federation interfaces: *service provider domain*, *partne*r (e.g., 3rd party service provider) *domain* and *customer* (e.g., vertical, end-user) *domain*. The LSO interface relevant for our work is INTERLUDE, which allows the service orchestration functionality (SOF) from a service provider to communicate with the SOF from a partner domain. The use of INTERLUDE for cross-domain slice provisioning is justified in any of these scenarios: i) part of a service provider's slice is hosted by the partner domain; and ii) part of a service provider's slice includes functions provided by the partner domain, e.g., 3rd party functions.

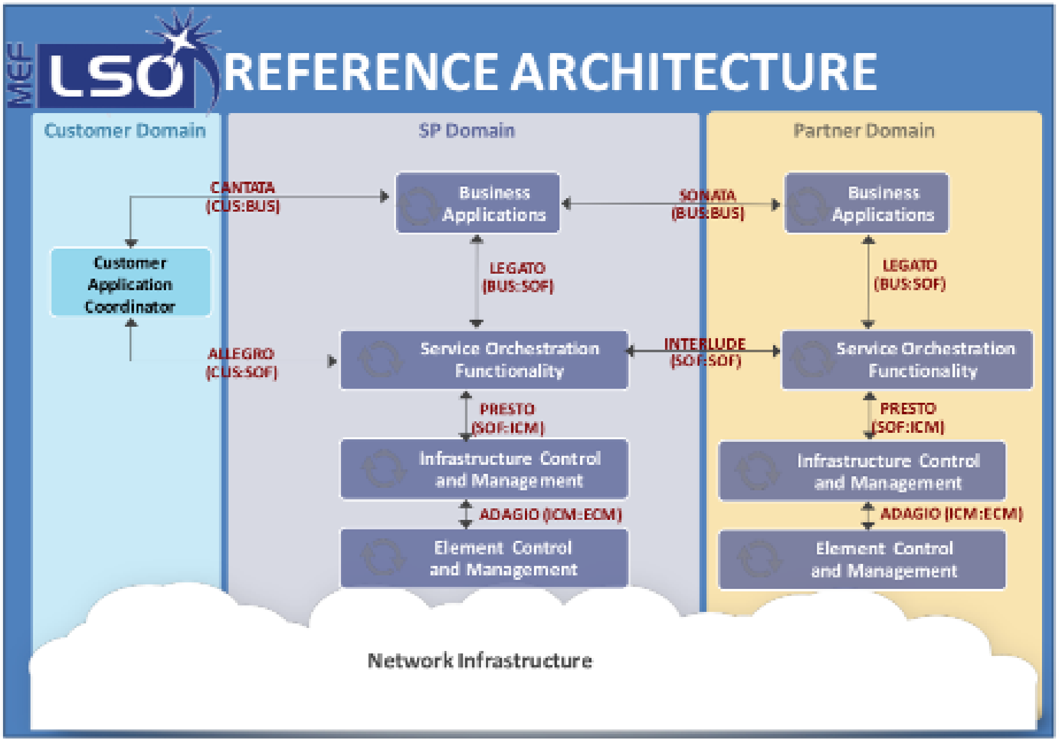


Figure D-2. MEF Lifecycle Service Orchestration (LSO) reference framework.

Leveraging the requirements and recommendations outlined by ETSI NFV and MEF, different research projects have also explored the federation problem, defining their own solutions at both service and resource levels. Some projects have proposed federation solutions based on peer-to-peer interactions between orchestrators, as it occurs in 5G-Exchange [7] and 5G-Transformer projects [8]. However, there are others like the 5G!Pagoda project [9], where a hierarchical approach is taken, with one parent orchestrator interacting with different domains.

3 5G-VINNI Facility Architecture

5G-VINNI leverages the latest 5G technologies to assemble a test and validation facility that provides industry verticals with isolated service experimentation platforms, in the form of E2E slices. These slices, accessed and used by verticals to set up innovative use case trials, are provisioned under the NSaaS model.

Figure D-3 shows the 5G-VINNI facility architecture. As seen, this facility consists of several interworking sites, each offering one or more slice types. The service offering from the different facility sites are registered into a single service catalog, which is made available to verticals through a common portal. In the next subsections, we provide more details on these components.

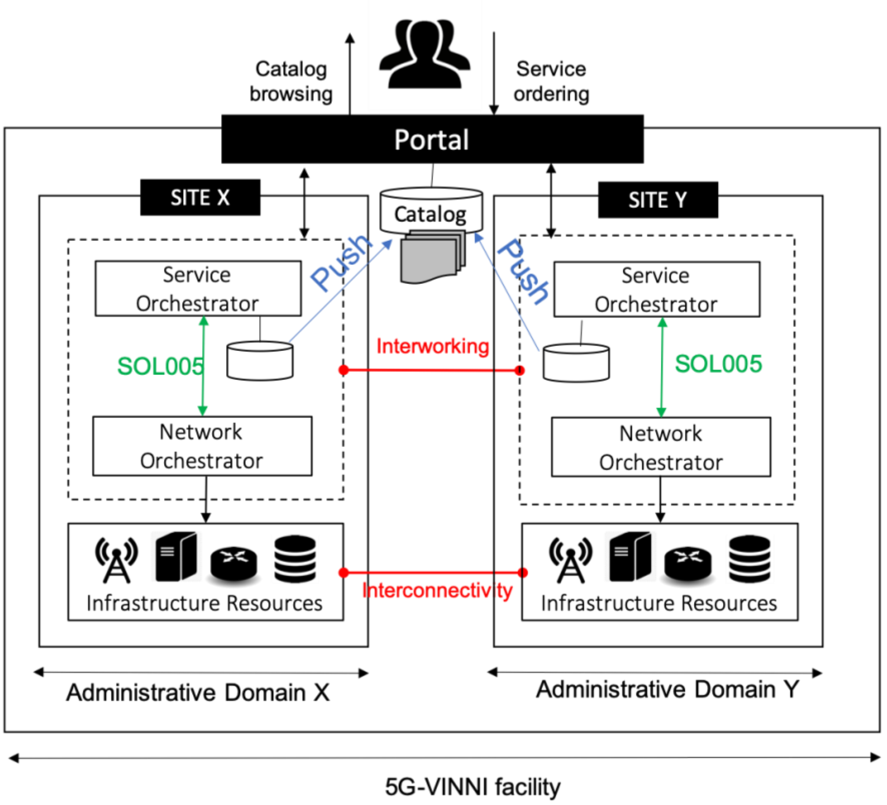


Figure D-3. 5G-VINNI facility architecture.

3.1 Facility Site

5G-VINNI facility consists of multiple sites, each deployed at a different geographic location (see Figure D-3) and defining a single administrative domain. Every 5G-VINNI facility site includes the following components: one Service Orchestrator, one Network Orchestrator and multiple infrastructure resources. All the components defined in a site can be from the same or different vendors, but all managed by a single network operator.

The *Service Orchestrator* (SO) takes care of the lifecycle management of the network slices at the application layer, i.e. network slice semantics. Taking the roles of Network Slice Management Function (NSMF) and Network Slice Subnet Management Function (NSSMF) as defined in 3GPP Rel-15 specifications, the SO implements the mechanisms and procedures for the provisioning of network slices in 5G mobile networks. This provisioning is done based on a well-defined information model, described in the *ServiceProfile* struct [10].

The *Network Orchestrator* handles the lifecycle of the network slices at the resource layer, i.e., network slice allocation. It is built with one NFV Orchestrator (NFVO), which orchestrates instances of VNFs and network services with the help of a Virtualized Infrastructure Manager (VIM). The relationship between the SO (3GPP scope) and the Network Orchestrator (NFV scope) is as follows: the virtualized part of a slice can be deployed as a network service. This means mapping the *ServiceProfil*e struct into one or more Network Service Descriptors (NSDs), each pointing to a number of VNF Descriptors (VNFDs). To allow this, the NFVO offers a northbound interface [11]. The protocol and data model for the interface capabilities are based on RESTful APIs, following the SOL005 normative specifications [12].

Finally, the *infrastructure resources* include access, transport, and core network functions. Some of these functions are physical, while some of them can be executed as VNFs on top of an NFV infrastructure. The cloud resources building this infrastructure are controlled and administrated by the VIM.

3.2 Service Catalog

Every facility site has a service catalog. This catalog allows the network operator to register and publish the network slices offered in his administrative domain. These service offerings are described using model-based service templates called 5G-VINNI Service Blueprints (VINNI-SBs). Designed under the principles of reusability and customizability, a VINNI-SB consists of a set of parameters that allow describing the capabilities that a vertical can get from a network slice.

Figure D-4 summarizes the structure of a VINNI-SB. As shown, the content of VINNI-SB is arranged into four main parts.

The **SST** is a 3GPP parameter [13] that specifies the 5G service category the slice is meant to support.

The **slice topology** provides information on how the slice is constructed from a logical viewpoint. This specifies i) the nodes the slice has, including information on their individual functionality; and ii) how these nodes are connected with each other along the entire topology, including information on their connectivity type. Each node taking part in the slice topology provides a well-defined functionality. Examples of these nodes include 3GPP components (e.g., gNodeB, 5G core control plane, user plane function) and value-added functionality (e.g., security, traffic shaping). Depending on the selected NFV criteria design, a node can be mapped to a network service or a VNF.

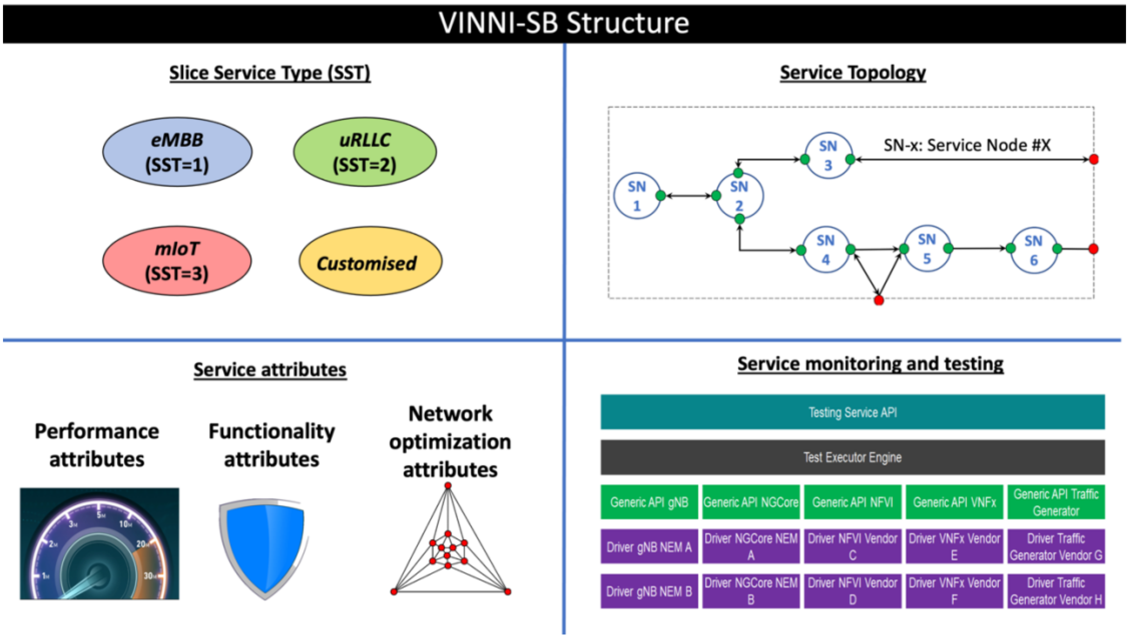


Figure D-4. Content of a 5G-VINNI Service Blueprint (VINNI-SB).

The **slice attributes** are based on those that GSMA has defined in the Generic Slice Template (GST) [14]. The specification of values for these attributes allows the vertical to provide the service requirements that the slice must satisfy.

Finally, there is the **slice testing and monitoring** part, relevant in experimentation environments. This last part allows specifying the testing and monitoring tools the verticals need to get from the slice to execute use case trials on top of it.

To allow site-agnostic orchestration of network slices, and eventually the deployment of slices across two or more sites, VINNI-SB defines a common information model for the whole 5G-VINNI facility. Taking advantage of this feature, 5G-VINNI facility collects the service offerings from the different sites, retrieving the VINNI-SBs from their catalogs and publishing them all into a single, centralized service catalog. This new catalog allows 5G-VINNI facility to provide a unified marketplace for verticals, helping them to browse the different service offerings and trigger corresponding service orders.

Note that the responsibility of keeping the above catalog updated is up to every site. Every time the catalog of a given site suffers from any modification (e.g., on-boarding of a new VINNI-SB, update of an existing VINNI-SB), the corresponding network operator shall notify this change to the centralized catalog, using pushing mechanisms to that end.

3.3 Portal

The portal provides a single entry-point for 5G-VINNI facility. Its mission is twofold. On the one hand, the portal support verticals, i.e., 5G-VINNI customers taking the role of experimenters, in obtaining access to facility resources and available functionality (VINNI-SBs, testing and monitoring tools) for the purposes of use case trialing and KPI validation. On the other hand, the portal allows exposing the facility to verticals as a unified platform rather than as an interconnection of individual administrative domains, hiding the mechanisms, protocols and technologies adopted in every site.

The first prototype of this 5G-VINNI portal is OpenSlice [15], an open-source operations support system born in the context of 5G-VINNI project. For more details on the OpenSlice architecture and capabilities, see [16].

4 Cross-Domain Slice Operation

A key enabler for vertical experimentation in 5G-VINNI facility is reproducibility, which can be defined as the ability to generate repeatable slice instances at multiple locations and at different time instants. Reproducibility allows any vertical to replicate experiments in controlled environments, assessing the variation of use case KPIs depending on selected capabilities. Different sites provide different 5G capabilities, not only in terms of resource capacity, but also in terms of functionality (e.g., edge support, telemetry/monitoring). To choose the capabilities that will support the use case execution, a vertical can decide where to deploy the slice: on one or another site, or across two or more sites. The latter is of particular interest for verticals, taking into account that many vertical services will span beyond the boundaries of a single administrative domain.

Cross-domain slice deployments brings several challenges in 5G-VINNI facility, as they do not only require data plane connectivity between the involved sites, but also interworking between their orchestration systems. For this interworking, two approaches can be followed: i) hierarchical orchestration; and ii) peer-to-peer orchestration. The first federation approach assumes the definition of a parent orchestrator, sitting on top of multiple child orchestrators, coordinating their workflows, and providing translation of their information / data models. This introduces significant burdens in management scalability, as the number of sites connected to this master orchestrator increases. Additionally, the scenario of having a network operator taking the broker role is unrealistic for upcoming operational networks, as it would raise concerns with the rest of operators in terms of privacy, auditability, and non-repudiation. For this reason, the peering approach is preferred for federating domains. In the following, we will discuss different federation options (Section 4.1) and provide a solution description for the option selected in 5G-VINNI facility (Section 4.2).

4.1 Federation Options

Considering the facility site components presented in Section 3.1, three options can be considered for federation:

- **Federation at Service Orchestration level (SO↔SO):** the SOs from different sites exchange information and expose their capabilities across them.
- **Federation at Network Orchestration level (NFVO↔NFVO):** the NFVOs from different sites exchange information and expose their capabilities across them.
- **Federation at different orchestration levels (SO↔NFVO):** the SO from one site communicates with the NFVO from another site.

All the above options are technically feasible when the federated sites rely on the same orchestration solution. In such a case, the use of proprietary interfaces is enough to enforce the required communication and capability exposure across domains. However, this scenario is rather unrealistic, especially looking ahead at upcoming commercial networks, where federation may involve multiple sites from different network operators, each making usage of a different orchestration solution. In the latter case, which is key for 5G-VINNI project, interoperability can only be achieved by means of standard interfaces. Table D-1 gives an insight into the three federation options, specifying their main features and the standard interfaces that can be used to fulfill these features. As seen, there exists at least one interface to implement every federation option. For example, SOL011 and SOL005, which define RESTful APIs for the implementation of Or-Or and Os-ma-nfvo interfaces, have become the de-facto solutions for the second and third federation options.

Table D-1. Federation Options

mService	Description	Interfaces
SO↔SO	*Information exchanged with external SO*: list of on-boarded VINNI-SBs, selected configuration of deployed slice (subnet) instances. *Operations exposed for external SO invocation*: slice (subnet) provisioning; slice (subnet) performance assurance; slice (subnet) fault supervision; network functions application layer conf & mgmt.	MEF LSO Interlude
NFVO↔NFVO	*Information exchanged with external NFVO:* list of on-boarded NSDs / VNFDs; records of deployed network service / VNF instances, with information on their resources. *Operations exposed for external SO invocation*: network service / VNF lifecycle mgmt.; network service / VNF monitoring; network service / VNF resources mgmt.	Or-Or [5]
SO↔NFVO	*Information exchanged with external SO:* the same as for NFVO « NFVO, but without information on instances resources. *Operations exposed for external SO invocation*: the same as for NFVO « NFVO, but without resources mgmt. *Information exchanged with external NFVO:* slice (subnet) – network service mapping.	Os-Ma-nfvo [11]

According to the above reasoning, from a technical viewpoint, the three options are equally valid and feasible for the intended federation. However, from an industry viewpoint, some options are less appealing than others, as happens with NFVO « NFVO and SO « NFVO. The main problem with these options is the difficulty of bringing them to the market, because of the reluctance of an operator to expose his NFVO beyond the boundaries of his administrative domains. Many reasons explain this reluctance. First, the need for the operator to expose on-boarded NSDs/VNFDs to other operators, especially considering that descriptor design is recognized as one of the main key enablers for revenue increase (and service differentiation) between operators. Secondly, the need for the operator to allow connection between his NFVO to an external NFVO/SO. In the case of SO « NFVO, this is even worse, since having two (or more) SOs connected to the same NFVO increases the risk of generating conflicting policies and inconsistencies in the status of that NFVO. Thirdly, the lack of in-built auditability in SOL011 and SOL005, which makes the corresponding NFVO exposed interfaces sensible points in terms of security and autonomy.

4.2 Solution for the Federation at the Service Orchestration Layer

MEF specifies in [6] the requirements and capabilities of the INTERLUDE interface (see Figure D-2). However, no data models or protocols have been defined for the interface implementation yet. Unlike, Or-Or and Os-Ma-nfvo interfaces, based on SOL005 and SOL011, no normative solution has been defined for INTERLUDE interface. In this context, Tele Management Forum (TM Forum) open APIs can be used. These APIs are not tied to vendor-specific orchestration solutions, allowing rapid integration and easy interoperability across domains.

As of today, a wide variety of Open APIs can be found in the TM Forum portfolio [18]. For the INTERLUDE interface implementation in 5G-VINNI, the following APIs apply:

- **Service Catalog Management API (TMF633),** providing artifacts for the registration and discovery of VINNI-SBs in the service catalog, as well as capabilities for their lifecycle management (e.g., registration, deletion, updating, etc.).
- **Service Ordering Management API (TMF641),** for issuing a service order. This order conveys the information required to deploy a slice instance: selected VINNI-SB and instantiation parameters. In some cases, this instance can be modelled as a network slice subnet instance.
- **Service Inventory Management API (TMF638),** which defines standardized mechanisms for CRUD operations over the records providing run-time information about the deployed slice (subnet) instances.
- **Service Configuration and Activation API (TMF640),** providing capabilities to allow the operation of a deployed slice (subnet) instance. This includes the ability to trigger lifecycle management actions (e.g., creation, modification, update, deletion) over that instance, and the ability to define rules to collect monitoring data from that instance (e.g., using threshold-based alarms or periodic notifications).

The SO of every 5G-VINNI facility site needs to offer these open APIs, so they can be consumed by the SOs from federated sites. The integration of open APIs in each site depends on the selected solution for the SO. In 5G-VINNI facility, two types of orchestration solutions exist:

- **Open Source MANO (OSM),** deployed in Spain site (Telefónica) and Greece site

(University of Patras). Although it was originally defined as a NFVO, OSM currently implements enhanced data models (based on SOL006) for 3GPP slicing support, thus taking the SO role as well.

- **Nokia's orchestration toolkit),** deployed in Norway site (Telenor) and UK site (British Telecom). This toolkit includes a SO (FlowOne) and a NFVO (CloudBand).

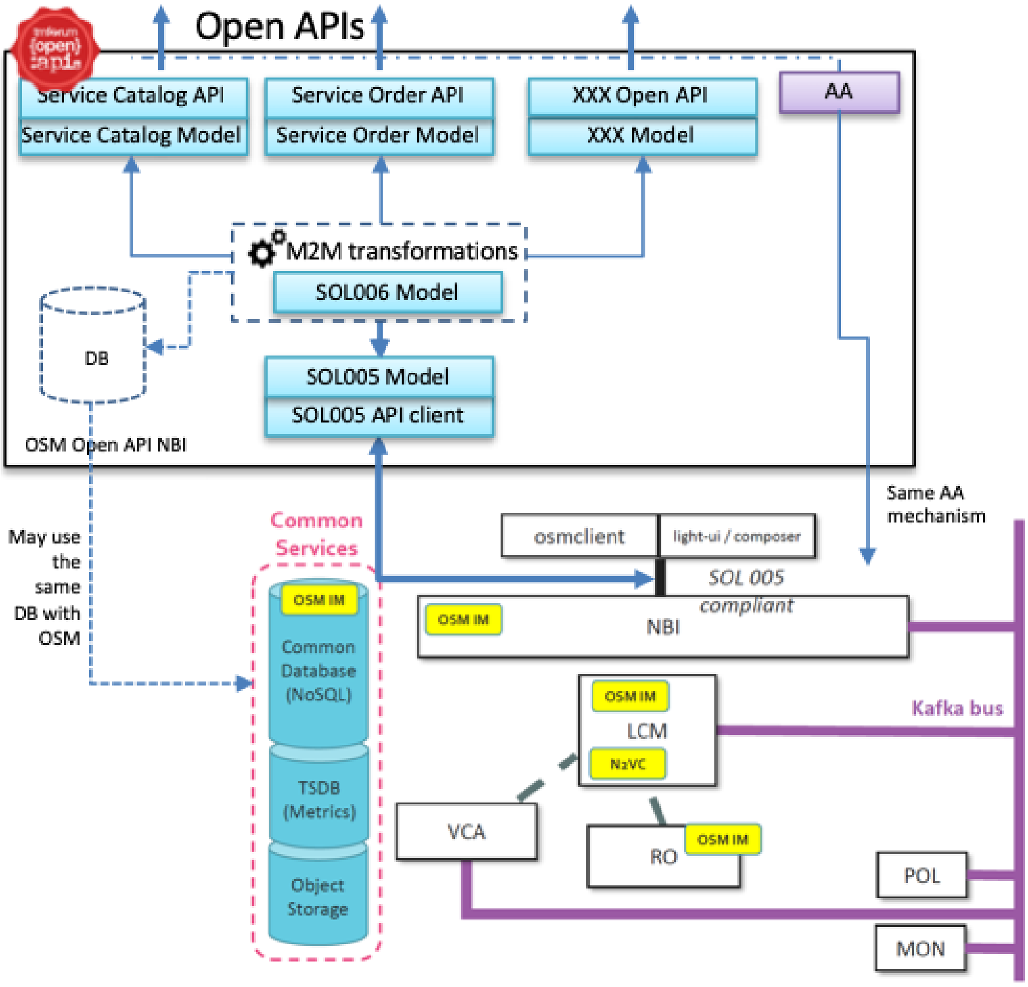


Figure D-5. TM Form Open APIs in OSM.

Figure D-5 shows how the integration of Open APIs is done in OSM.

5 Federation Use Case

We explain here how federation enables the deployment and operation of an E2E slice instance across two facility sites upon vertical request. In this process, three phases can be envisioned: *slice ordering*, *slice fulfillment* and *slice operation*.

5.1 Slice Ordering

In the first phase, the vertical gains access to the 5G-VINNI facility through the portal, browses the centralized service catalog, selects one VINNI-SB and issues the corresponding service order. In this service order, the vertical provides a completed specification of the slice instance he wants, including information on slice topology (possible extended with 3rd party VNFs), slice attributes (filled in with values fitting use case requirements) and slice location. We assume the following: i) the vertical wants the slice deployed across two facility sites, each having a different orchestration solution; and ii) the selected VINNI-SB was retrieved from the local catalog of one of these sites. By way of example, we consider that an industry vertical orders the provisioning of an eMBB slice instance across Norway and Spain. In this ordering, the vertical selects from the 5G-VINNI service catalog a VINNI-SB with SST=1, which was originally retrieved from Spain's OSM catalog.

The service order with the above setup is captured by the portal's order manager, which validates the order and send it to the Spain site. Then, the slice fulfillment phase begins.

5.2 Slice Fulfilment

In the second phase, upon receiving the service order, Spain site checks it, realizing that part of the ordered slice needs to be deployed at Norway site. This means that federation between the SOs of both sites (OSM and Nokia's FlowOne) is needed. From this point, the event workflow is as follows. First, OSM on-boards the VINNI-SB into FlowOne's catalog, using **TMF633**. Then, OSM decomposes the service order received from the portal, identifying the subnets which will be deployed on Spain and Norway sites. Finally, it triggers a service order towards FlowOne, using **TMF641**. With this order, OSM informs FlowOne about the topology and attributes of the slice (subnet) instance to be deployed on Norway site.

After the above actions, the slice can be commissioned. To this end, each SO first deploys the slice subnet at its site, providing day-0 and day-1 configuration on the different VNFs. Then, OSM and FlowOne exchange connectivity information of their slice subnets (e.g., IP addresses of VNF instances at the edge of each subnet) to set up a L2/L3 VPN connectivity service across these subnets, establishing an E2E data plane for the slice. The exchange of information is done **TMF638**, while the VPN connectivity service instantiation is done with **TMF640**.

5.3 Slice Operation

At the operation time, the cross-domain slice can be made available to the vertical for advanced experimentation activities. As part of these activities, advanced lifecycle management operations (e.g., scaling) can be issued. In this case, cooperation between SOs is needed by means of **TMF638** and **TMF640**

After the above actions, the slice can be commissioned. To this end, each SO first deploys the slice subnet at its site, providing day-0 and day-1 configuration on the different VNFs. Then, OSM and FlowOne exchange connectivity information of their slice subnets (e.g., IP addresses of VNF instances at the edge of each subnet) to set up a L2/L3 VPN connectivity service across these subnets, establishing an E2E data plane for the slice. The exchange of information is done **TMF638**, while the VPN connectivity service instantiation is done with **TMF640**.

6 Conclusions and Future Work

5G-VINNI project defines and develops an integrated, multi-domain E2E facility for advanced vertical experimentation in 5G environments. The facility provides a realistic reproduction of full 5G commercial networks, with multiple technologies and solutions in place, allowing ICT industry to explore and address some of the challenges these networks may bring in their rollouts.

In this paper, we have faced the problem of operating a vertical-oriented slice deployed across multiple administrative domains, each managed by a single network operator using a different orchestration solution. In 5G-VINNI facility, two different orchestration tools are deployed: OSM and Nokia's orchestration toolkit. An effective cross-domain slice operation requires the interoperation of both mentioned tools. To this end, 5G-VINNI implements a federation-oriented, standards-based solution across them, making use of TM Forum Open APIs.

At this stage, four APIs have been developed. To validate their usability, a set of conformance tests involving OSM and Nokia's FlowOne have also been carried out. However, validation is also needed with real-world vertical scenarios, assessing the readiness of these open APIs in terms of scalability. This work is planned to be performed in the forthcoming months of the 5G-VINNI project, once ICT-19 projects and their vertical use cases get on-boarded.

Acknowledgement

This work has been supported by the H2020 European Project 5G-VINNI (grant agreement No. 815279).

References

[1] 5G-PPP White Paper "View on 5G Architecture", v3.0, June 2019.

[2] 5G-VINNI project" [Online]. Available: https://www.5g-vinni.eu/

[3] L. M. Contreras, "Interworking of Softwarized Infrastructures for Enabling 5G Multi-Site Orchestration", *2019 IEEE Conference on Network Softwarization (NetSoft),* Paris, France, 2019, pp. 458-463.

[4] ETSI GS NFV-IFA 028, "Report on architecture options to support multiple administrative domains, 2019.

[5] ETSI GS NFV-IFA 030, "Management and Orchestration; Multiple Administrative Domain Aspect Interfaces", 2019.

[6] Metro Ethernet Forum, "MEF 55; Lifecycle Service Orchestration (LSO): Reference Architecture and Framework", 2019.

[7] 5GEx project D2.2, "5GEx Final System Requirements and Architecture", Dec. 2017.

[8] 5G-Transformer project D2.4, "Final design and implementation report on the MTP", Jan. 2019.

[9] 5G!Pagoda project D4.3, "End-to-end Network Slice", Jan. 2019.

[10] 3GPP TS 28.541, "Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and 3".

[11] ETSI GS NFV-IFA 013, "Management and Orchestration; Os-Ma-nfvo reference point – Interface and Information Model Specification".

[12] ETSI GS NFV-SOL 005, "Protocols and Data Models; RESTful protocols specification for the Os-Ma-nfvo Reference Point".

[13] 3GPP TS 23.501, "System architecture for the 5G System; Stage 2".

[14] GSMA PRD NG.116 v2.0, "Generic Network Slice Template", 2019.

[15] OpenSlice project [Online]. Available: http://openslice.io

[16] OpenSlice docs [Online]. Available: https://openslice.readthedocs.io/

[17] ETSI GS NFV-SOL 011, "Protocols and Data Models; RESTful protocols specification for the Or-Or Reference Point".

[18] TM Forum, "GB992 Open API Map R17.0.1", 2017.

Paper E

Automated Network Slice Scaling in Multi-site Environments: The ZSM PoC #2 report

Jose Ordonez-Lucena, Christos Tranoris, Borja Nogales Dorado

Disclaimer:
This document constitutes the report of a ETSI ZSM Proof of Concept (PoC)
The report is publicly available at the ZSM Wiki Portal.
Link: https://zsmwiki.etsi.org/images/8/84/ZSM_POC_2_Final_Report.pdf

1 PoC Project Details

1.1 PoC Project Review

PoC Number:	2
PoC Project Name:	Automated network slice scaling in multi-site environments
PoC Project Host:	Telefónica S.A.
Short Description:	This PoC has the aim of demonstrating the capability to automatically scale out a deployed network instance across multiple administrative domains. This will be achieved using the 5G assets of 5G-VINNI, which is a large-scale, end-to-end facility composed of several interworking sites, each deployed at a different geographic location and defining a single administrative domain. The management and orchestration capabilities of individual sites, and the enablers allowing for the interworking across them, are aligned with ZSM architectural design principles. The PoC fits the End-to-End (E2E) service management scenario category detailed in ZSM 001, considering the network slicing features specified in ZSM 003. The management and orchestration assets for this PoC, based on the combined use of Open Source MANO (OSM) and Openslice, are aligned with the ZSM architectural principles captured in ZSM 002 together with the fulfilment and assurance solutions specified in ZSM 008.
PoC Project Status: ***(Ongoing/Completed)***	Completed

1.2 PoC Team Members Review

	Organization Name	**ISG participant**	**Contact (email)**	**PoC Point of Contact (*)**	**Role (***)**	**PoC Components**
1	Telefónica S.A.	Yes	Jose Ordonez-Lucena joseantonio.ordonezlucena@telefonica.com Diego R. López diego.r.lopez@telefonica.com	X	Network/ service provider	Use case definition. Business model definition
2	Telenor ASA	Yes	Min Xie min.xie@telenor.com Pål Grønsund pal.gronsund@telenor.com Andres J. González andres.gonzalez@telenor.com		Network/ service provider	Use case definition. Business model definition.
3	Universidad Carlos III (UC3M)	No	Carmen Guerrero carmen.guerrero@uc3m.es Borja Nogales bdorado@pa.uc3m.es Iván Vidal ividal @it.uc3m.es Adrián Gallego adrgalle@pa.uc3m.es		University / Supplier	VNFs provider. Integrator

4	University of Patras (UoP)	No	Spyros Denazis sdena@upatras.gr Dimitris Giannopoulos dimit.giannopoulos@upnet.gr Panagiotis Papaioannou papajohn@upatras.gr Yiannis Chatzis ioannis.chatzis@upatras.gr		University / Supplier	VNFs provider. Integrator
5	Openslice	No	Christos Tranoris tranoris@ece.upatras.gr Kostis Trantzas ktrantzas@upnet.gr		Open source project	Openslice framework .
(*) Identify the PoC Point of Contact with an X. (**) The Role will be network operator/service provider, infrastructure provider, application provider or other						

All the PoC Team members listed above declare that the information in this report is conformant to their activities during the PoC Project.

1.3 PoC Project Scope Review

1.3.1 PoC Topics

Report the status of all the PoC Topics and Expected Contributions anticipated in the PoC proposal.

PoC Topic Code	PoC Topic Description	Related WI	Submitted Contribution link	Date	Status (*)

2	Automation in Multi-Stakeholder Ecosystem	ZSM 004(**)	ZSM(21)000162 "ZSM004 Add Openslice to Section 6" (***)
			ZSM(21)000163 "ZSM004 Openslice in ZSM architecture" (***)
				30/04/2021	Completed

(*) Planned, On-going, Completed, delayed (new target date), Abandoned					

(**) The planned contribution was made to ZSM 004, as it was not possible to contribute to the ZSM 001 and ZSM 003 (WIs officially linked to the PoC topic #2, see https://zsmwiki.etsi.org/index.php?title=Topic2_-_Automation_in_Multi-Stakeholder_Ecosystems):

- ZSM 001 -> this WI was dormant (no active revision) during the PoC project lifetime.
- ZSM 003 -> by the time when the PoC results were available for contribution, ETSI ZSM started the approval process for ZSM 003 final draft, meaning no further contributions were allowed.

(***) These contributions propose Openslice as a solution to automate network slice out operation, where different administrative domains (each managed by a different stakeholder) participate in the scaling out operation, as planned in the PoC proposal. The multi-domain and multi-stakeholder nature of Openslice, as well as the mapping of its services into ZSM framework, is captured in these contributions. Section 2 of the present report describes the solution in different scenarios and identifies some gaps in ZSM 008.

1.3.2 Other Topics in Scope

Report the status of all the PoC Topics and Expected Contributions anticipated in the PoC proposal.

PoC Topic Code	PoC Topic Description	Related WI	Submitted Contribution link	Date	Status (*)
(*) Planned, On-going, Completed, delayed (new target date), Abandoned					

"Automation in Multi-Stakholder Ecosystem" was the only PoC topic which was active during the PoC lifetime.

1.4 PoC Project Milestones Review

PoC Milestone	Milestone description	Target Date	Additional Info	Completion Date
P.P.1	PoC Presentation	02/12/2020	Presentation to ZSM NOC	15/11/2020
P.S	PoC Proposal submission	15/12/2020	Official PoC proposal submission	29/11/2020
P.P.2	PoC Public Announce	15/01/2021	Public Web announcement in 5G-VINNI media (web, twitter, etc.). *Once it is approved.	11/01/2020
P.PU	PoC user story detailed	22/01/2021	Detailing use case, specifying actors, pre-conditions & post-conditions, and exceptions.	03/02/2020
P.PT	PoC Test Plan	03/03/2021	Testbed setup and running	23/03/2021
P.D1	PoC Demo	17/03/2021	Demo for showcasing at ETSI endorsed Webinar	16/04/2021
P.C1	PoC Expected Contribution	17/03/2021	Propose contributions to several topics at ZSM meeting	30/04/2021 (to be presented in ZSM-14m Tech Call)
P.R	PoC Report	01/04/2021	PoC-Project-End Feedback	30/04/2021
P.E	PoC Project End	01/04/2021		30/04/2021

1.5 Confirmation of PoC Event Occurence

Due to COVID-19 restrictions, the PoC was showcased in a free-of-charge, ETSI endorsed webinar which took place on April 16th, 2021. More details of this webinar can be found below:

- **ETSI site for webinar registration**: https://www.etsi.org/events/1905-webinar-zsm-poc-2-showcase-automated-network-slice-scaling-in-multi-site-environments?jjj=1619193312791
- **Platform**: ETSI BRIGHTTALK CHANNEL
- **Webinar title**: "ZSM PoC#2 showcase: Automated network slice scaling in multi-site environments"
- **Webinar duration**: 105 min (15h00 – 16h45 CEST)
- **Webinar statistics**: 46 people attended live out of 90 pre-registered.

ETSI
The Standards People

ZSM PoC#2 showcase: Automated network slice scaling in multi-site environments

Presented by: Jose Ordonez-Lucena, Christos Tranoris, Borja Nogales

For: Webinar

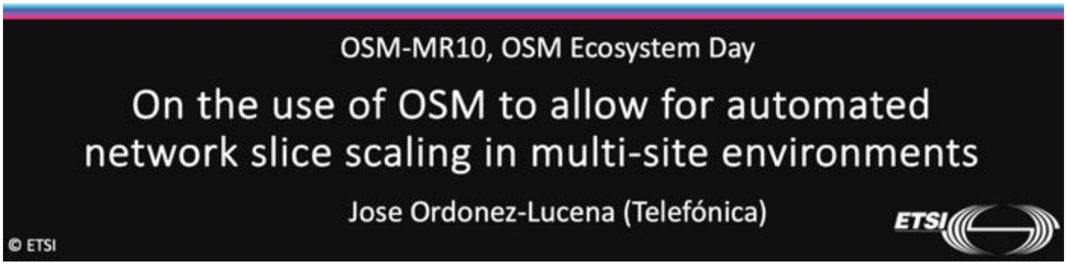
16/04/2021

© ETSI 2021 – All rights reserved

1.6 Other Dissemination Activities

In order to reach a wider audience, the PoC team participated in the OSM Ecosystem Day at OSM-MR10 (https://osm.etsi.org/wikipub/index.php/OSM-MR10_Hackfest), with a 25-min presentation that provided an overview of this PoC#2. This presentation is publicly available at the following OSM website: http://osm-download.etsi.org/ftp/osm-9.0-nine/OSM-MR10-hackfest/EcosystemDay/OSM-MR10%20ED1%20-%20Telefonica.pdf

Open Source MANO



OSM-MR10, OSM Ecosystem Day

On the use of OSM to allow for automated network slice scaling in multi-site environments

Jose Ordonez-Lucena (Telefónica)

© ETSI

ETSI

2 ZSM PoC Technical Report

2.1 PoC Project Milestones Review

2.1.1 PoC motivation

The PoC#2 focuses on the management of a network slice when deployed across multiple administrative domains. Specifically, this PoC aims at demonstrating how to automatically scale out a network slice instance in multi-site environments. The rationale of the demonstration is as follows:

- There is an existing (running) network slice instance. This instance is deployed across two different 5G-VINNI facility sites: Madrid (Spain) and Patras (Greece), each hosting a portion of the entire network slice.
    - From a functional viewpoint, the network slice consists of multiple NFV network services, each corresponding to a network slice subnet.
    - From an operational viewpoint, the network slice is deployed as a multi-site network slice instance.
    - From a network viewpoint, there exists L3 connectivity between Madrid and Patras, so that in-slice connectivity can be ensured along the entire data path.
- The behavior of existing (running) network slice instance is continuously monitored.
    - Policy-based performance management on individual facility sites.
    - There are pre-defined policy rules that allow triggering the need for scaling out operation based on collected metrics.
- When certain policy rules are met at Madrid facility site, a scaling out operation is triggered. This operation applies to the entire network slice instance.
    - This means that although the scaling out operationally is triggered at Madrid facility site, this operation needs to be propagated to Patras facility site accordingly.
    - Consistency is a must: increasing capacity of one network slice subnet on one

facility site requires modifying the capacity of the network slice subnet accordingly.

According to this rationale, the PoC#2 requires a use case that justifies (i) having a multi-site network slice instance; (ii) the triggering of scaling out operation at Madrid facility site, and (iii) the need to propagate the scaling out operation to the Patras facility site. The selected use case is based on vertical industry related (e.g. e-Health, PPDR) NetApps hackathon involving developers from Spain and Greece. For this short-lived event, a network slice instance is deployed.

The logic of the in-scope use case, illustrated in Figure E-18, is as follows:

- There is a NetApp submission service where developers continuously upload their solutions. The NetApps submission portal and the backend service broker are hosted in Madrid facility site.
- Due to EU defined General Data Protection Regulation (GDPR) policy, NetApps binaries and data must be hosted in the home country. Therefore, the services for managing the NetApps catalogue repositories need to be located at both Madrid and Patras facility sites.
- During the hackathon days, there is a sudden high demand of portal interaction, due to an unexpected prize to winner developers. The demand is first detected in Madrid, thus the backend hosts of NetApps catalogue repository will be automatically scaled there.
- The scaling out operation triggered in Madrid is propagated to the Patras facility site, since this sudden high demand of portal interaction is also expected at Greece side. Unlike Madrid, where the scaling out was a reactive corrective action, the scaling out operation triggered at Patras facility site is a pro-active corrective action (due to forecasting reasons).

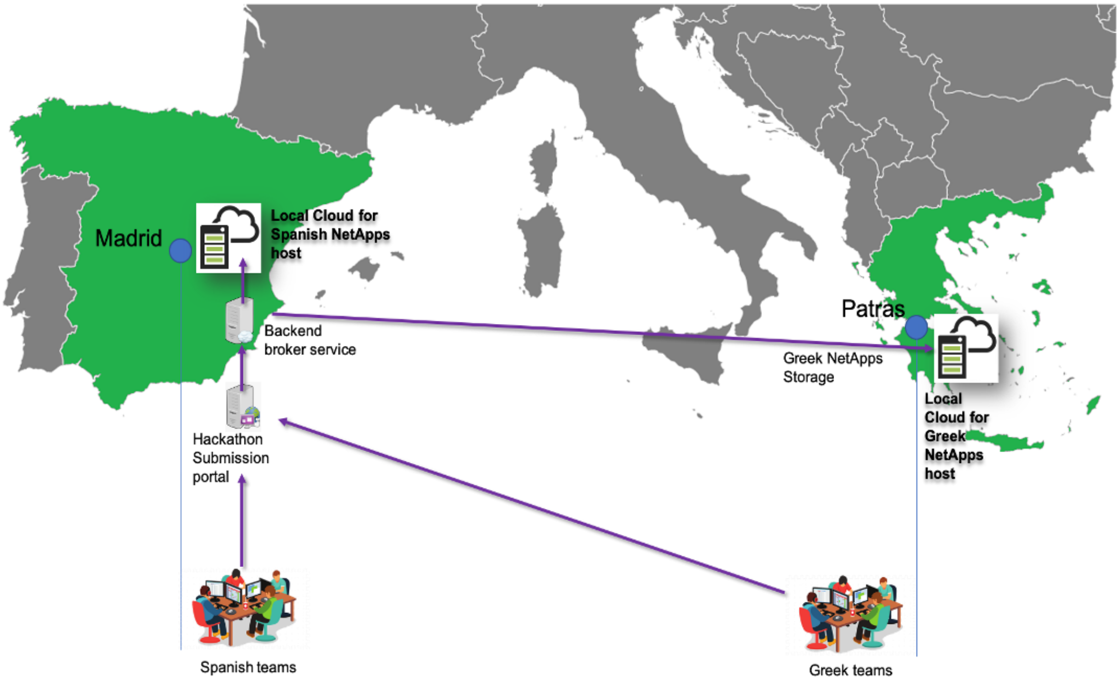

Madrid
Local Cloud for Spanish NetApps host
Backend broker service
Hackathon Submission portal
Spanish teams
Patras
Greek NetApps Storage
Local Cloud for Greek NetApps host
Greek teams

Figure E-18. PoC use case.

2.1.2 PoC architecture

The architecture of this PoC is illustrated in Figure E-19. As seen, the setup consists of two identical orchestration stacks, one for each 5G-VINNI facility site involved: Openslice (Service Orchestrator) + OSM (NFV Orchestrator) + Openstack (VIM). To allow for multi-site network slice orchestration, interworking between both stacks is a must. In the PoC, this interworking occurs at the Service Orchestration layer, with "Madrid-Openslice" and "Patras-Openslice" communicating using TMF Forum Open APIs.

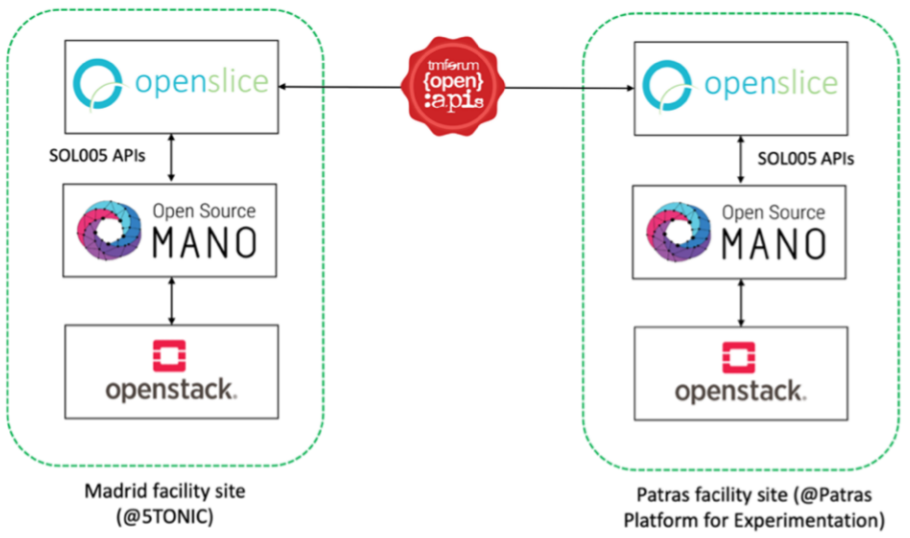


Figure E-19. PoC architecture

In ZSM architectural framework, OSM is mapped to a ZSM management domain (MD), while Openslice plays the role of E2E service MD. Figure E-20 and Figure E-21 provide a more detailed view of OSM and Openslice internal architectures, illustrating how their modules are related to ZSM grouping data and management services. More details of these relationships can be found in the first PoC#2 report [1].

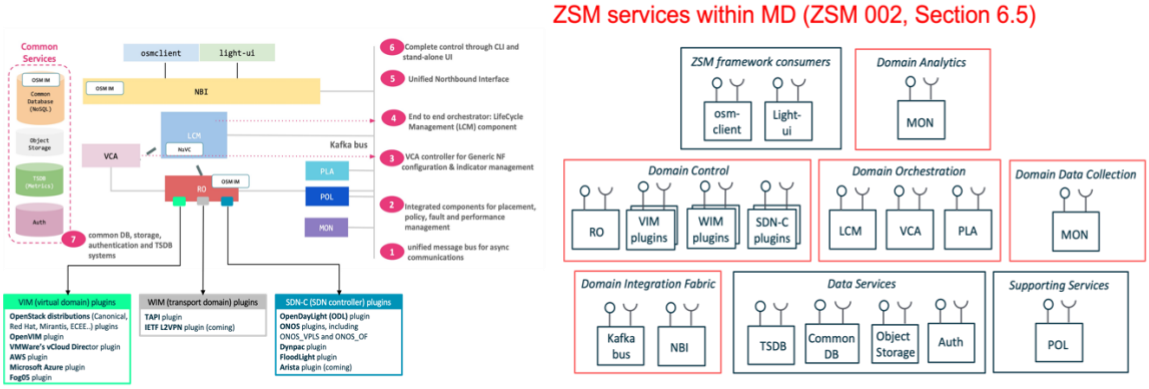

Common Services
OSM IM
Common Database (NoSQL)
Object Storage
TSDB (Metrics)
Auth
osmclient
light-ui
OSM IM
NBI
LCM
N2VC
VCA
Kafka bus
RO
OSM IM
PLA
POL
MON
6 Complete control through CLI and stand-alone UI
5 Unified Northbound Interface
4 End to end orchestrator: LifeCycle Management (LCM) component
3 VCA controller for Generic NF configuration & indicator management
2 Integrated components for placement, policy, fault and performance management
1 unified message bus for async communications
7 common DB, storage, authentication and TSDB systems
VIM (virtual domain) plugins
OpenStack distributions (Canonical, Red Hat, Mirantis, ECEE..) plugins
OpenVIM plugin
VMWare's vCloud Director plugin
AWS plugin
Microsoft Azure plugin
FogOS plugin
WIM (transport domain) plugins
TAPI plugin
IETF L2VPN plugin (coming)
SDN-C (SDN controller) plugins
OpenDayLight (ODL) plugin
ONOS plugins, including ONOS_VPLS and ONOS_OF
Dynpac plugin
FloodLight plugin
Arista plugin (coming)
ZSM services within MD (ZSM 002, Section 6.5)
ZSM framework consumers
osm-client
Light-ui
Domain Analytics
MON
Domain Control
RO
VIM plugins
WIM plugins
SDN-C plugins
Domain Orchestration
LCM
VCA
PLA
Domain Data Collection
MON
Domain Integration Fabric
Kafka bus
NBI
Data Services
TSDB
Common DB
Object Storage
Auth
Supporting Services
POL

Figure E-20. On the mapping of OSM with ZSM services.

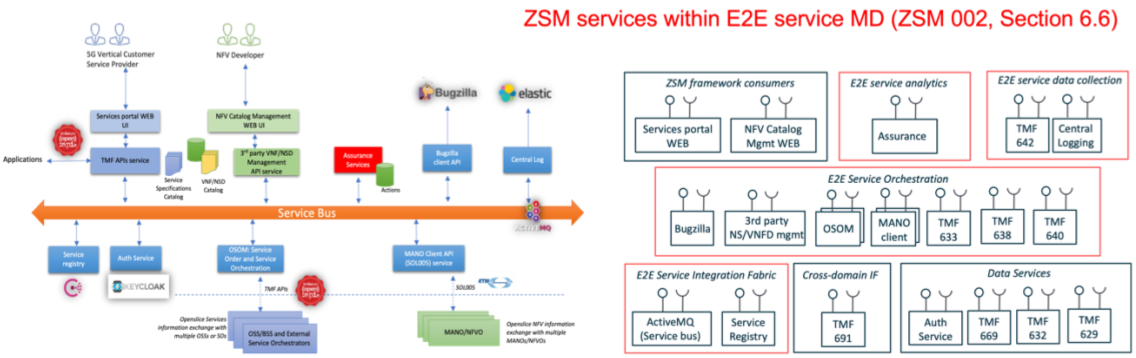

5G Vertical Customer Service Provider
NFV Developer
Bugzilla
elastic
Services portal WEB UI
NFV Catalog Management WEB UI
Applications
TMF APIs service
Service Specifications Catalog
VNF/NSD Catalog
3rd party VNF/NSD Management API service
Assurance Services
Actions
Bugzilla client API
Central Log
Service Bus
ACTIVEMQ
Service registry
Auth Service
KEYCLOAK
OSOM: Service Order and Service Orchestration
MANO Client API (SOL005) service
TMF APIs
SOL005
ETSI
Openslice Services information exchange with multiple OSSs or SOs
OSS/BSS and External Service Orchestrators
MANO/NFVO
Openslice NFV information exchange with multiple MANOs/NFVOs
ZSM services within E2E service MD (ZSM 002, Section 6.6)
ZSM framework consumers
Services portal WEB
NFV Catalog Mgmt WEB
E2E service analytics
Assurance
E2E service data collection
TMF 642
Central Logging
E2E Service Orchestration
Bugzilla
3rd party NS/VNFD mgmt
OSOM
MANO client
TMF 633
TMF 638
TMF 640
E2E Service Integration Fabric
ActiveMQ (Service bus)
Service Registry
Cross-domain IF
TMF 691
Data Services
Auth Service
TMF 669
TMF 632
TMF 629

Figure E-21. On the mapping of Openslice with ZSM services.

2.1.3 PoC user story

The NetApps hackathon event captured in Figure E-19 requires deploying a network slice across Madrid and Patras facility sites. This means that each facility site hosts a portion of the slice. The internal composition of this slice and the geographical distribution of their functional components is illustrated in Figure E-22. As seen, the network slice consists of three network slice subnets, each modelled as a separate Network Service Descriptor (NSD).

- **Network Slice Subnet A (NSS-A),** deployed in Madrid according to NSD$_F$. The NSD$_F$ is composed of four VNFs, including two Load Balancers (LB-1 and LB-2), one Web Server and one backend API brokering service.
- **Network Slice Subnet B (NSS-B),** deployed in Madrid according to NSD$_{SRV}$. The NSD$_{SRV}$, consists of three VNFs, including one Load Balancer (LB-3), one repository catalogue and one catalogue DB.
- **Network Slice Subnet C (NSS-C),** deployed in Greece according to NSD$_{SRV}$. Like NSS-B, NSS-C holds the repository catalogue and its supported DB, together with the LB-3 as an entry point of requests.

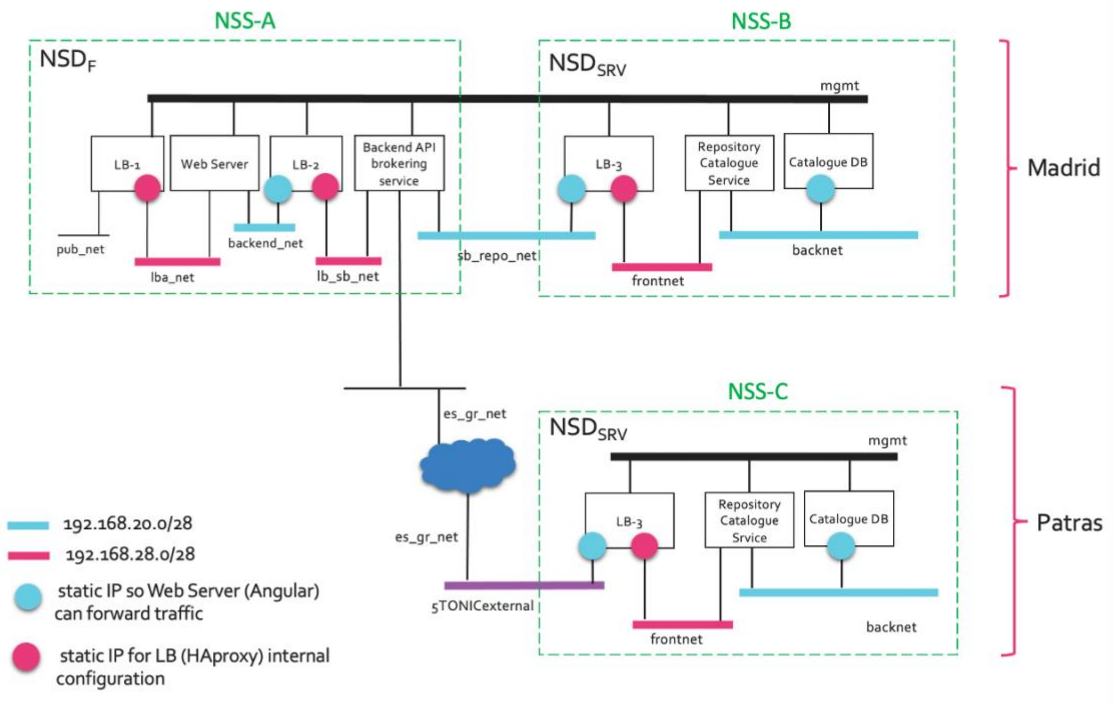


Figure E-22. PoC pre-conditions.

Madrid-Patras connectivity is based on a multi-site NFV and data communication pipe enabled through a VPN-based overlay solution. For more details of the specific solution used, see [2].

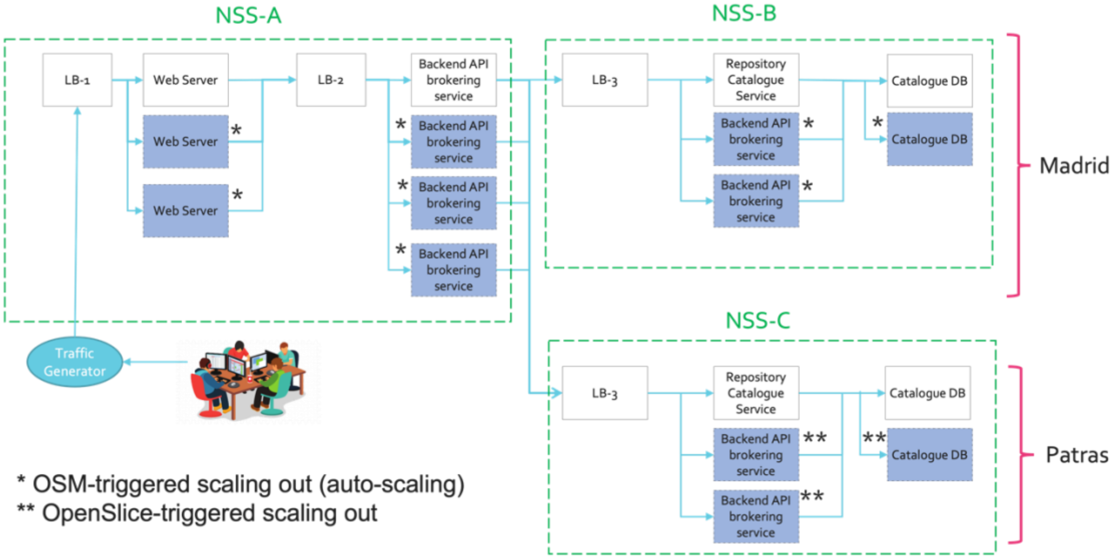
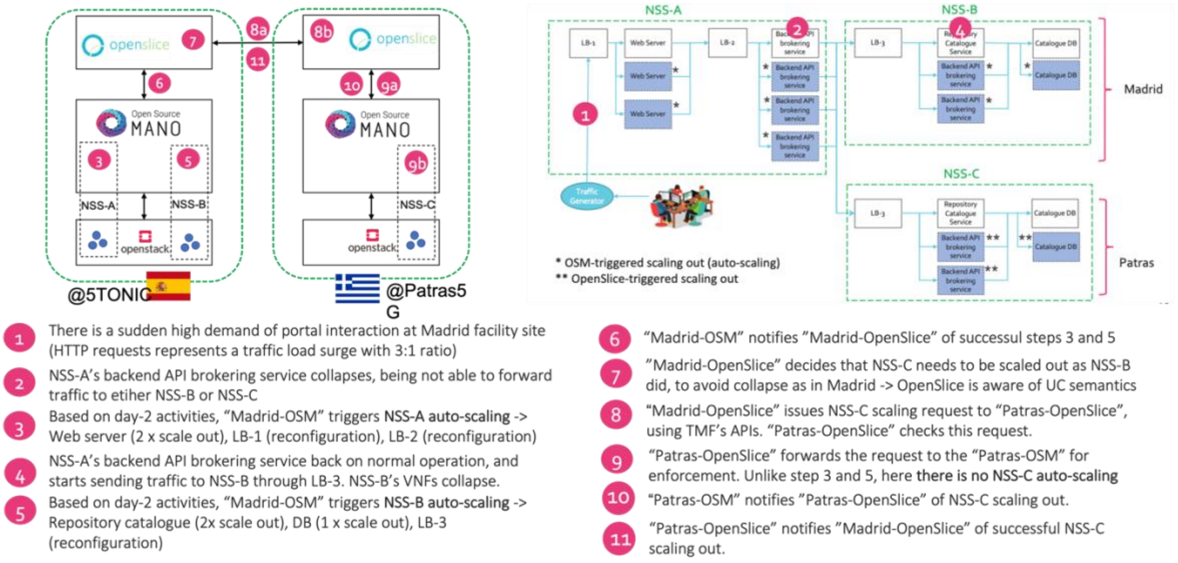

NSS-A
NSS-B
NSS-C
LB-1
Web Server
LB-2
Backend API brokering service
LB-3
Repository Catalogue Service
Catalogue DB
Madrid
Patras
Traffic Generator
* OSM-triggered scaling out (auto-scaling)
** OpenSlice-triggered scaling out

Figure E-23. PoC post-conditions

Figure E-23 illustrates the impact of the scaling operation over the running slice instance. The white-colored part of the figure captures the slice instance as originally deployed for the NetApps hackathon (PoC pre-conditions), while the full picture shows the state of the slice instance after being scaled out (PoC post-conditions). The user story that explains this transition is depicted in Figure E-24. For more details, see the first PoC#2 report in [1].



1 There is a sudden high demand of portal interaction at Madrid facility site (HTTP requests represents a traffic load surge with 3:1 ratio)

2 NSS-A's backend API brokering service collapses, being not able to forward traffic to etiher NSS-B or NSS-C

3 Based on day-2 activities, "Madrid-OSM" triggers **NSS-A auto-scaling** -> Web server (2 x scale out), LB-1 (reconfiguration), LB-2 (reconfiguration)

4 NSS-A's backend API brokering service back on normal operation, and starts sending traffic to NSS-B through LB-3. NSS-B's VNFs collapse.

5 Based on day-2 activities, "Madrid-OSM" triggers **NSS-B auto-scaling** -> Repository catalogue (2x scale out), DB (1 x scale out), LB-3 (reconfiguration)

6 "Madrid-OSM" notifies "Madrid-OpenSlice" of successul steps 3 and 5

7 "Madrid-OpenSlice" decides that NSS-C needs to be scaled out as NSS-B did, to avoid collapse as in Madrid -> OpenSlice is aware of UC semantics

8 "Madrid-OpenSlice" issues NSS-C scaling request to "Patras-OpenSlice", using TMF's APIs. "Patras-OpenSlice" checks this request.

9 "Patras-OpenSlice" forwards the request to the "Patras-OSM" for enforcement. Unlike step 3 and 5, here **there is no NSS-C auto-scaling**

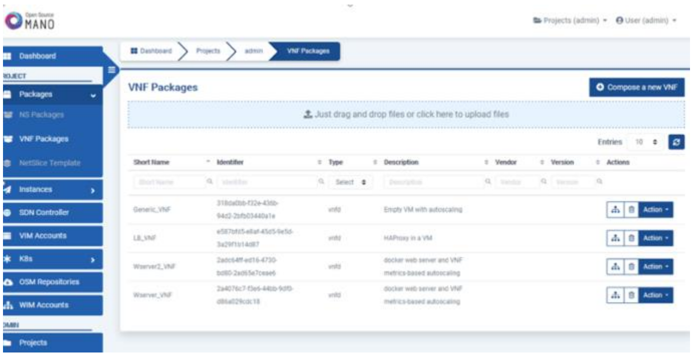
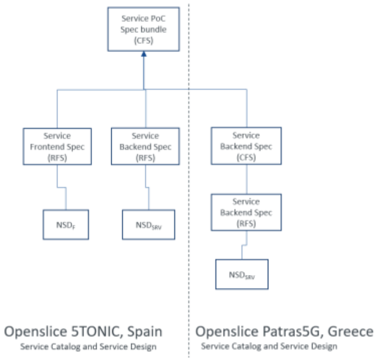
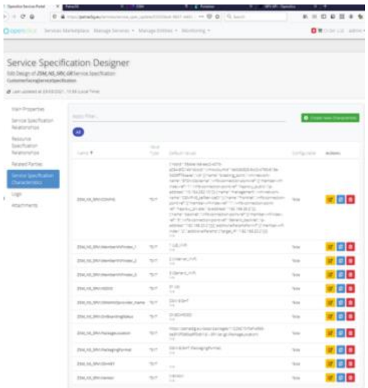
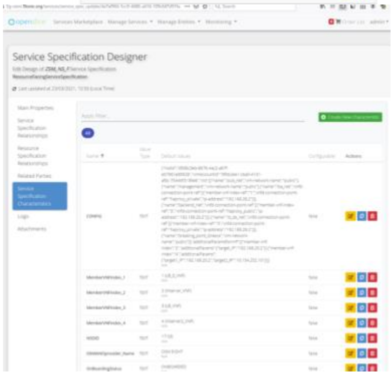
10 "Patras-OSM" notifies "Patras-OpenSlice" of NSS-C scaling out.

11 "Patras-OpenSlice" notifies "Madrid-OpenSlice" of successful NSS-C scaling out.

Figure E-24. PoC user story.

2.1.4 Scenarios

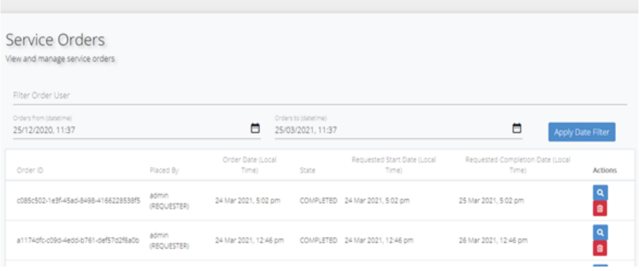
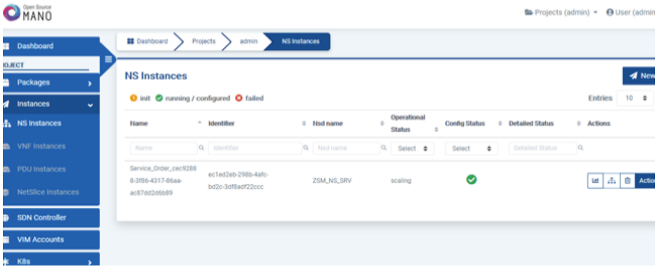
Table E-1. Scenario 1: Service on-boarding

Precondition	Each facility site includes their management and orchestration stack (Openslice+OSM+Openstack) in operation. Both Openslice instances are configured to communicate and exchange Service Catalog management and Service Ordering management related request-response/notify-subscription messages.
Verification	VNFDs and NSDs are onboarded to the OSM instance available in each facility. RFS and CFS specifications are onboarded to the Openslice instance available in each facility.
Sequence	
a. NSD_F (and constituent VNFDs) together with NSD_{SRV} (and corresponding VNFDs) are onboarded to “Madrid-OSM”. These descriptors specify the NFV resource requirements of NSS-A and NSS-B, respectively.	 Open Source MANO Projects (admin) User (admin) Dashboard PROJECT Packages NS Packages VNF Packages NetSlice Template Instances SDN Controller VIM Accounts K8s OSM Repositories WIM Accounts ADMIN Projects Dashboard > Projects > admin > VNF Packages Compose a new VNF Just drag and drop files or click here to upload files Entries 10 VNF Packages Short Name / Identifier / Type / Description / Vendor / Version / Actions Generic_VNF / 318da0bb-f32e-436b-94d2-2bfb03440a1e / vnfd / Empty VM with autoscaling / Action LB_VNF / e587bfd5-e8af-45d5-9e5d-3a29f1b14d87 / vnfd / HAProxy in a VM / Action Wserver2_VNF / 2adc64ff-ed16-4730-bd80-2ad65e7ceae6 / vnfd / docker web server and VNF metrics-based autoscaling / Action Wserver_VNF / 2a4076c7-f3e6-44bb-9df0-d86a029cdc18 / vnfd / docker web server and VNF metrics-based autoscaling / Action
b. NSD_{SRV} (and constituent VNFDs) is onboarded to “Patras-OSM”. This descriptor specifies the NFV resource requirements of NSS-C.	
c. The CFS and RFS specifications describing the slice subnets to be deployed for the PoC are designed in Openslice. The composition of these specifications and their relationships with the onboarded NSDs are captured in the right-side figure. This figure illustrates the diagram tree for the PoC slice specification.	 Service PoC Spec bundle (CFS) Service Frontend Spec (RFS) Service Backend Spec (RFS) Service Backend Spec (CFS) Service Backend Spec (RFS) NSD$_F$ NSD$_{SRV}$ NSD$_{SRV}$ Openslice 5TONIC, Spain Service Catalog and Service Design Openslice Patras5G, Greece Service Catalog and Service Design
d. The RFS specifications of those network slice subnets to be deployed at Spain facility site (NSS-A and NSS-B) are designed in “Madrid-Openslice”. These specifications include “Service Frontend Spec (RFS): NSS-A” and “Service Backend Spec (RFS): NSS-B”.	 Service Specification Designer  Service Specification Designer
e. The RFS specification of the network slice subnet to be deployed at Greece facility site (NSS-C) is designed in “Patras-Openslice”. This specification corresponds to “Service Frontend Spec (RFS): NSS-C”.	
f. The “Service Frontend Spec (RFS): NSS-C” needs to be available public in the “Patras-Openslice” service catalog, so that it can be exposed to the “Madrid-Openslice” service catalog. To that end, the RFS specification is turned into a CFS specification.	

g. The CFS specification of the PoC network slice is designed in “Madrid-Openslice” according to this bundle: “Service Frontend Spec (RFS): NSS-A” + “Service Backend Spec (RFS): NSS-B” + “Service Backend Spec (CFS): NSS-C”.	

Service Specification Designer

Table E-2. Scenario 2: Service deployment

Precondition	The CFS/RFS specifications and NSD/VNFDs are onboarded to Openslice and OSM at both facility sites.
Verification	The network slice instance is running. Metadata info (records) of deployed network slice subnet instances are stored in the service inventories of both facility sites. Action rules for service policy management are created.
Sequence	
a. The customer requests the allocation of a dedicated network slice. To that end, it issues a service order based on Service PoC bundle (CFS).	
b. The service order is captured in “Madrid-Openslice” OSOM.	
c. Following the diagram tree for the PoC slice specification, the network services (and VNFs) corresponding to individual network slice subnets are deployed on every facility site. Network service instances based on NSD$_F$ and NSD$_{SRV}$ are deployed in Spain facility site (via “Madrid-OSM”), while a network service instance based on NSD$_{SRV}$ is deployed in Greece facility site (via “Patras-OSM”).	
d. Day-0 + day-1 configuration of individual VNFs is performed via Juju	

Service Orders

View and manage service orders

Filter Order User

Orders from (datetime): 25/12/2020, 11:37

Orders to (datetime): 25/03/2021, 11:37

Order ID	Placed By	Order Date (Local Time)	State	Requested Start Date (Local Time)	Requested Completion Date (Local Time)	Actions
c085c502-1e3f-45ad-8498-41662285385f5	admin (REQUESTER)	24 Mar 2021, 5:02 pm	COMPLETED	24 Mar 2021, 5:02 pm	25 Mar 2021, 5:02 pm	
a1174dfc-c09d-4edd-b761-def57d2f6a0b	admin (REQUESTER)	24 Mar 2021, 12:46 pm	COMPLETED	24 Mar 2021, 12:46 pm	26 Mar 2021, 12:46 pm	

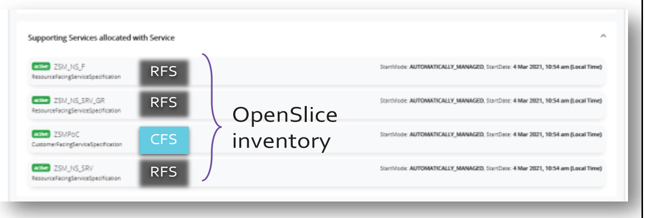
NS Instances

Name	Identifier	Nsd name	Operational Status	Config Status	Detailed Status	Actions
Service_Order_cec92888-3f86-4317-86aa-ac87dd2d6b89	ec1ed2eb-298b-4afc-bd2c-3df8adf22ccc	ZSM_NS_SRV	scaling	✓		

<table>
<tr>
<td>

charms. After this, the service order is successfully completed. The components building up the network slice instance are running.

</td>
<td>



Supporting Services allocated with Service

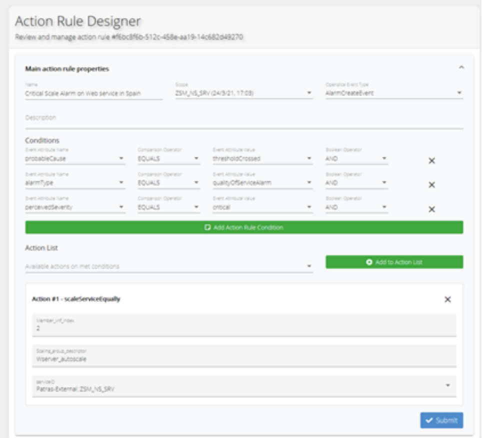
active ZSM_NS_F ResourceFacingServiceSpecification	RFS	StartMode: **AUTOMATICALLY_MANAGED**, StartDate: **4 Mar 2021, 10:54 am (Local Time)**
active ZSM_NS_SRV_GR ResourceFacingServiceSpecification	RFS	StartMode: **AUTOMATICALLY_MANAGED**, StartDate: **4 Mar 2021, 10:54 am (Local Time)**
active ZSMPoC CustomerFacingServiceSpecification	CFS	StartMode: **AUTOMATICALLY_MANAGED**, StartDate: **4 Mar 2021, 10:54 am (Local Time)**
active ZSM_NS_SRV ResourceFacingServiceSpecification	RFS	StartMode: **AUTOMATICALLY_MANAGED**, StartDate: **4 Mar 2021, 10:54 am (Local Time)**

</td>
</tr>
<tr>
<td>

e. An action rule is created on the scope of the running network slice. This action rule assists Openslice to perform service auto-scaling. The created alarm is as follows:

SCOPE affectedService="ZSM_NS_SRV"
ON AlarmCreateEvent
IF (probableCause = thresholdCrossed) & (severity = critical) & (alarmType = qualityOfServiceAlarm)
THEN actions = scaleServiceEqually(Patras-External::ZSM_NS_SRV, VNFIndex=2)

</td>
<td>



Action Rule Designer

Review and manage action rule #f6bc8f6b-512c-458e-aa19-14c682d49270

Main action rule properties

Name: Critical Scale Alarm on Web service in Spain

Scope: ZSM_NS_SRV (24/3/21, 17:03)

Openslice Event Type: AlarmCreateEvent

Description

Conditions

Event Attribute Name	Comparison Operator	Event Attribute Value	Boolean Operator
probableCause	EQUALS	thresholdCrossed	AND
alarmType	EQUALS	qualityOfServiceAlarm	AND
perceivedSeverity	EQUALS	critical	AND

Action List

Available actions on met conditions

Action #1 - scaleServiceEqually

Member_vnf_index: 2

Scaling_group_descriptor: Wserver_autoscale

serviceID: Patras-External::ZSM_NS_SRV

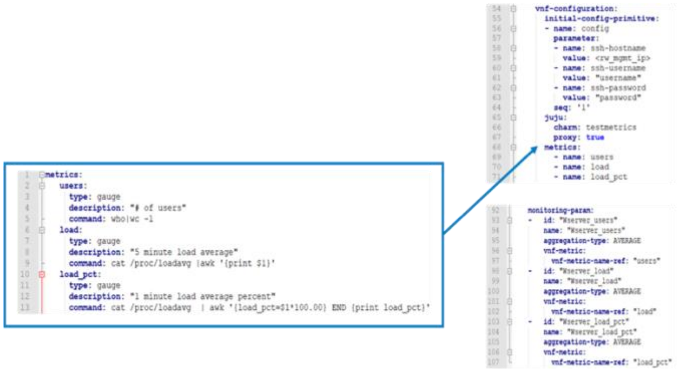
</td>
</tr>
</table>

Table E-2. Scenario 3a: Service auto-scaling (OSM)

<table>
<tr>
<td><b>Precondition</b></td>
<td>After scenario 2, the slice instance is completely deployed and configured across both facility sites.</td>
</tr>
<tr>
<td><b>Verification</b></td>
<td>The OSM component included in the orchestration stack of the Madrid facility site monitors the service performance offered by both subnets deployed in that facility site, and automatically executes the reactive scaling out operations in case of detecting a service performance degradation. NSS-A and NSS-B are successfully scaled out.</td>
</tr>
<tr>
<td colspan="2"><b>Sequence</b></td>
</tr>
<tr>
<td>

a. "OSM-Madrid" starts monitoring the performance offered by both subnets deployed in this facility site. For this purpose, the VNFDs include the metrics that are intended to be collected by the OSM monitoring framework, as well as the frequency for their collection.

</td>
<td>



```
metrics:
   users:
     type: gauge
     description: "# of users"
     command: who|wc -l
   load:
     type: gauge
     description: "5 minute load average"
     command: cat /proc/loadavg |awk '{print $1}'
   load_pct:
     type: gauge
     description: "1 minute load average percent"
     command: cat /proc/loadavg  | awk '{load_pct=$1*100.00} END {print load_pct}'
```

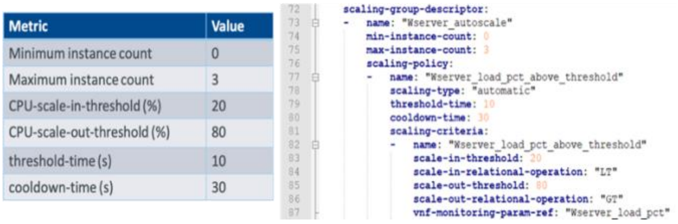
```
  vnf-configuration:
    initial-config-primitive:
    - name: config
      parameter:
      - name: ssh-hostname
        value: <rw_mgmt_ip>
      - name: ssh-username
        value: "username"
      - name: ssh-password
        value: "password"
      seq: '1'
    juju:
      charm: testmetrics
      proxy: true
    metrics:
      - name: users
      - name: load
      - name: load_pct
```

```
monitoring-param:
-   id: "Wserver_users"
    name: "Wserver_users"
    aggregation-type: AVERAGE
    vnf-metric:
      vnf-metric-name-ref: "users"
-   id: "Wserver_load"
    name: "Wserver_load"
    aggregation-type: AVERAGE
    vnf-metric:
      vnf-metric-name-ref: "load"
-   id: "Wserver_load_pct"
    name: "Wserver_load_pct"
    aggregation-type: AVERAGE
    vnf-metric:
      vnf-metric-name-ref: "load_pct"
```

</td>
</tr>
<tr>
<td>

b. Due to the high demand of user requests (cf. user story, step 1), the Web Server collapses in terms of CPU usage (cf. user story, step 2). "Madrid-OSM" detects a performance degradation on NSS-A, leveraging OSM performance management framework (with OSM's POL and MON modules involved), and decides that a corrective action needs to be

</td>
<td>



Metric	Value
Minimum instance count	0
Maximum instance count	3
CPU-scale-in-threshold (%)	20
CPU-scale-out-threshold (%)	80
threshold-time (s)	10
cooldown-time (s)	30

```
    scaling-group-descriptor:
    -   name: "Wserver_autoscale"
        min-instance-count: 0
        max-instance-count: 3
        scaling-policy:
        -   name: "Wserver_load_pct_above_threshold"
            scaling-type: "automatic"
            threshold-time: 10
            cooldown-time: 30
            scaling-criteria:
            -   name: "Wserver_load_pct_above_threshold"
                scale-in-threshold: 20
                scale-in-relational-operation: "LT"
                scale-out-threshold: 80
                scale-out-relational-operation: "GT"
                vnf-monitoring-param-ref: "Wserver_load_pct"
```

</td>
</tr>
</table>

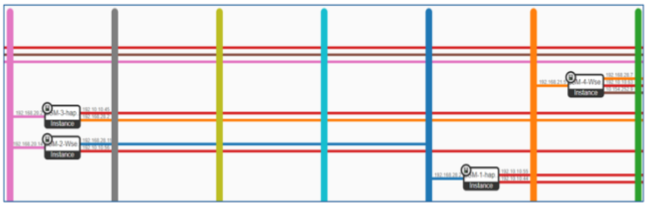
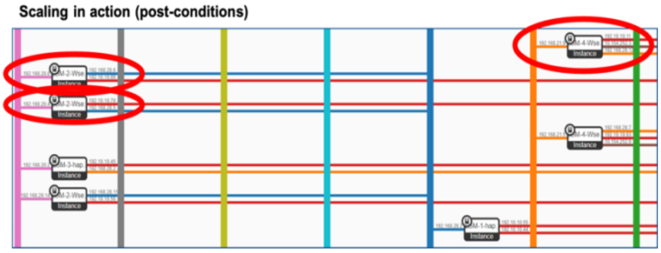
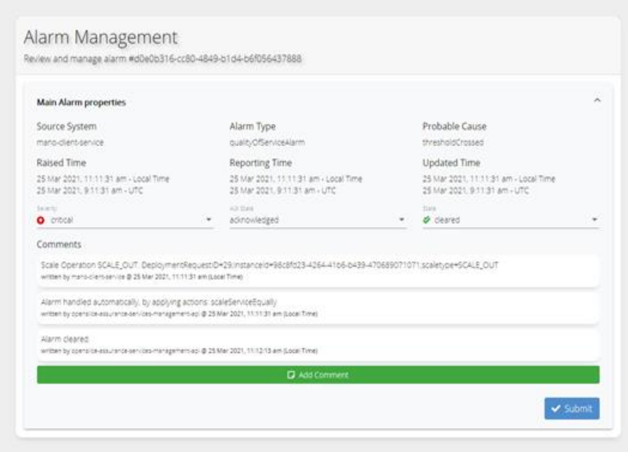
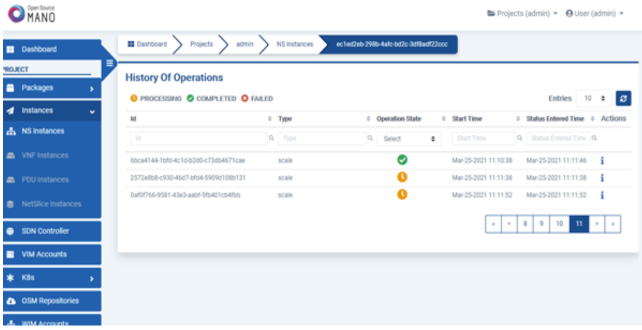
taken on this web server. To that end, the POL module checks the rules defined in the VNFD: trigger scaling out operation when CPU usage exceeds 80%.	
c. "Madrid-OSM" proceeds with the scaling operation on web server VNF.	
d. Once web server VNF has been scaled out, the rest of NSS-A components need also to be resized accordingly, by creating additional instances of backend API brokering service VNF, following the same procedure as steps 2 and steps 3.	Scaling in action (pre-conditions)  M-3-hap Instance M-2-Wse Instance M-4-Wse Instance M-1-hap Instance Scaling in action (post-conditions)  M-3-Wse Instance M-2-Wse Instance M-3-hap Instance M-2-Wse Instance M-4-Wse Instance M-4-Wse Instance M-1-hap Instance
e. The NSS-A components, including existing and newly created VNF instances, are configured accordingly (cf. user story, step 3). This allows capturing the results of scaling operation in the semantics of individual VNFs.	
f. Due to the high demand of user requests (cf. user story, step 1), the NSS-B VNFs collapse (cf. user story, step 4). "Madrid-OSM" detects a performance degradation on NSS-B leveraging OSM performance management framework, and decides that a corrective action needs to be taken on the "Repository Catalog service".	
g. The NSS-B is scaled out (cf. user story, step 5). To that end, steps c, d and e are similarly applied on NSS-B VNFs.	

Table E-3. Scenario 3b: Service auto-scaling (Openslice)

Precondition	NSS-A and NSS-B have been successfully scaled out at the Spain facility site.
Verification	NSS-C is scaled at Patras, following up the event captured at Madrid. "Madrid-Openslice" is notified about the successful NSS-C scaling out operation.
Sequence	
a. Upon NSS-A and NSS-B scaling out operation, "Madrid-OSM" sends a notification to "Madrid-Openslice" (cf. use story, step 6)	

b. “Madrid-Openslice” acknowledges the receipt of this notification by creating an alarm. The alarm matches the action rule originally designed in scenario 2 (service deployment), and therefore “Madrid-Openslice” handles it (cf. user story, step 7). **SCOPE** affectedService=“ZSM_NS_SRV ” **ON** AlarmCreateEvent **IF** (probableCause = thresholdCrossed) & (severity = critical) & (alarmType = qualityOfServiceAlarm) **THEN** actions = scaleServiceEqually(Patras-External::ZSM_NS_SRV, VNFIndex=2)	 Alarm Management Review and manage alarm #d0e0b316-cc80-4849-b1d4-b6f056437888 Main Alarm properties Source System: mano-client-service Alarm Type: qualityOfServiceAlarm Probable Cause: thresholdCrossed Raised Time: 25 Mar 2021, 11:11:31 am - Local Time, 25 Mar 2021, 9:11:31 am - UTC Reporting Time: 25 Mar 2021, 11:11:31 am - Local Time, 25 Mar 2021, 9:11:31 am - UTC Updated Time: 25 Mar 2021, 11:11:31 am - Local Time, 25 Mar 2021, 9:11:31 am - UTC Severity: critical Ack State: acknowledged State: cleared Comments Scale Operation SCALE_OUT. DeploymentRequestID=29,instanceId=98c8f023-4264-4166-b439-470689071071,scaletype=SCALE_OUT written by mano-client-service @ 25 Mar 2021, 11:11:31 am (Local Time) Alarm handled automatically, by applying actions: scaleServiceEqually written by openslice-assurance-services-management-api @ 25 Mar 2021, 11:11:31 am (Local Time) Alarm cleared written by openslice-assurance-services-management-api @ 25 Mar 2021, 11:12:13 am (Local Time) Add Comment Submit
c. Following up the directives from the action rule, **“**Madrid-Openslice” requests “Patras-Openslice” to scale NSS-C out, using TMF OpenAPIs (cf. user story, step 8).	 Open Source MANO Projects (admin) User (admin) Dashboard > Projects > admin > NS Instances > ec1ed2eb-298b-4afc-bd2c-3df8adf22ccc Dashboard PROJECT Packages Instances NS Instances VNF Instances PDU Instances NetSlice Instances SDN Controller VIM Accounts K8s OSM Repositories WIM Accounts History Of Operations PROCESSING COMPLETED FAILED Entries 10 Id | Type | Operation State | Start Time | Status Entered Time | Actions 6bca4144-1bfd-4c1d-b2d0-c73db4671cae | scale | ✓ | Mar-25-2021 11:10:38 | Mar-25-2021 11:11:46 2572e8b8-c930-46d7-bfd4-5909d108b131 | scale | ● | Mar-25-2021 11:11:38 | Mar-25-2021 11:11:38 0af0f766-9581-43e3-aabf-5fb401cb4fbb | scale | ● | Mar-25-2021 11:11:52 | Mar-25-2021 11:11:52 « ‹ 8 9 10 11 › »
d. NSS-C is scaled at Patras (cf. user story, step 9), following the step g from scenario 3a.	
e. Upon receiving notification from “Patras-OSM”, the **“**Patras-Openslice” notifies the “Madrid-Openslice” (cf. user story, steps 10 and 11).	

2.2 General

Contribution	WI/Document Ref	Comments	Meeting
ZSM(21)000162 “ZSM004 Add Openslice to Section 6”	ETSI GR ZSM 004 [3]	This contribution aims to include Openslice in the landscape of ZSM related open-source communities (ZSM 004, Section 6)	ZSM-14m Tech Call
ZSM(21)000163 “ZSM004 Openslice in ZSM architecture”	ETSI GR ZSM 004 [3]	This contribution is a proposal on how Openslice framework fits with the ZSM reference architecture, illustrating how Openslice components map with the ZSM grouping of management and data services.	ZSM-14m Tech Call

2.3 Gaps identified in ZSM standardization

Gap identified	Forum (ZSM ISG, Other)	WI/Document Ref	Gap details and Status

Policy-driven E2E service assurance	ZSM	ETSI GS ZSM 008 [4]	The PoC has demonstrated that automated scaling on a E2E service requires the definition of policies (action rules) in the service assurance set-up operation. However, no guidance on how a policy should be specified for this operation has been captured in ZSM 008 thus far. The PoC team recommends the ZSM 008 rapporteur (and contributors) to take action on this, providing some guidance for Communication Service Providers (CSPs)/vendors, so that they do not need to develop ad-hoc solutions every time they want to define a policy for a E2E service.
ZSM framework consumer	ZSM	ETSI GS ZSM 008 [4]	The PoC has demonstrated up to three different actors for the ZSM framework consumer: a ZSM management domain, a NFV developer and a vertical customer. The on-boarding, fulfilment and assurance operations detailed in ZSM 008 represents ZSM framework consumer in a generic, abstract manner, which are not always applicable to the different actors playing the role of ZSM framework consumer. The PoC team recommends the ZSM 008 rapporteur (and contributors) to take action on this, making it clear that not all the operations defined therein are applicable to every ZSM framework consumer. Examples captured in an informative annex could be valuable for outside readers.
WIs in scope of PoC topic #2	ZSM	ZSM PoC topic #2	The PoC topic #2 only includes ZSM 001 and ZSM 003 as concerned WIs. However, there are other recent WIs, such as ZSM 004 ("Landscape", w) and ZSM 008 ("Cross-domain E2E service lifecycle management"), that despite being related to the scope of PoC topic #2, they are not explicitely mentioned in the current PoC topic #2 description (see here). The PoC team recommends the PoC Management Team (PMT) to update the current PoC topic #2 description to include ZSM 004 and ZSM 008 as in-scope WIs.

2.4 PoC Suggested Action Items

The PoC#2 has leveraged the 5G-VINNI results and environment to demonstrate automation in multi-domain environments, with a focus on network slice lifecycle management, covering on-boarding, fulfilment, and assurance phases [4]. The PoC team has showcased how ZSM architectural framework is a key asset to achieve a zero-touch slice operation beyond the boundaries of one network operator (thanks to the definition of a SBMA that facilitate collaborative interactions among different stakeholder), bridging the gaps between a variety of standards with different focus (e.g., TM Forum, ETSI NFV) as well.

The PoC#2 exemplifies a complete technology evolution path, based on the triplet {research + experimentation + standardization} and with open-source communities (OSM and Openslice) along the entire path.

The PoC#2 sets the ground for future experimentation in the future, with:

- further integration of additional NFVO solutions. For future work, PoC team is exploring the use of both open-source and vendor-specific NFVOs, to assess the interworking of YANG and TOSCA models, in on-boarding and fulfilment phases. To achieve the required interoperability, it is important that selected NFVOs provide

SOL005 capabilities through their NBI.

- scenarios focused on ETSI GS ZSM 009 (Closed-Loop Automation) and ETSI GS ZSM 012 (AI enablers).

2.5 Additional Messages to ZSM

None in addition to the matters discussed above.

2.6 Additional messages to Network Operators and Service Providers

None.

References

[1] ZSM PoC#2 Report 1, "PoC#2 user story", Feb 2021. Available: https://zsmwiki.etsi.org/images/f/ff/ZSM_POC_2_User_Story.pdf

[2] B., Gonzalez, L. F., Vidal, I., Valera, F., Garcia-Reinoso, J., Lopez, D. R., Rodríguez, J., Gonzalez, N., Berberana, I., Azcorra, A. Integration of 5G Experimentation Infrastructures into a Multi-Site NFV Ecosystem. *J. Vis. Exp*. (168), e61946, doi:10.3791/61946 (2021).

[3] ETSI GR ZSM 004, "Zero-touch network and Service Management (ZSM); Landscape", v1.6.0, March 2021. ETSI GR ZSM 008, "Zero-touch network and Service Management (ZSM); Cross-domain E2E Service Lifecycle Management", v0.6.0, March 2021.

[4] ETSI GR ZSM 008, "Zero-touch network and Service Management (ZSM); Cross-domain E2E Service Lifecycle Management", v0.6.0, March 2021.

Part IV

Private 5G Networks and Network Slicing: The Art of Living Together

Literature Review and Problem Description

Part IV addresses the Objective 3 of this dissertation, which is the specification and analysis of solutions for private 5G networks exploiting network slicing capabilities. This objective is addressed in Papers F, G and H. Prefacing these publications, in this chapter we include two sections that help the reader have the full picture and understand the problem we want to address. Section 1 provides background context, capturing the precedents with a literature review. Section 2 identifies the main limitations of the state-of-the-art and puts them in relation with the contributions done in Papers F, G and H.

1 Background context

Private networks are not a new concept. Thousands of sites across the globe have deployed private networks in the form of Ethernet, enterprise Wi-Fi, TETRA (TErrestrial Trunked RAdio), DMR (Digital Mobile Radio) and MPT-1327 to support a variety of use cases, including emergency services, mining, as well as office communications. The focus of these traditional private networks is security and privacy protection for communication.

However, the enterprise digital and IoT transformation strategies call for new capabilities. The focus is now on connecting assets, including machines, sensors, and other objects, to address use cases such as monitoring, automation, and business analytics, with new services developed around these data streams. Private cellular (mobile) networks are best positioned to meet these needs, by offering top-tier security and privacy features while also enabling real-time communications for data produced by various networked elements. Over the past years industry and enterprises have set their sights on private cellular network solutions, with “private LTE” getting consolidated and “private 5G” now taking its first steps.

It is worth noting that private LTE/5G networks do not only provide secure communication and rich-featured network capabilities; it also delivers them via a single unified solution. This reduces the complexity for enterprises and large-scale public institutions to build a fully digitized environment, alleviating the integration efforts they faced in the past, while addressing the main limitations of legacy technologies. That’s why private LTE/5G solutions are called to be the main drivers to unleash digital transformation of industry verticals. Table IV-1 summarizes the main motivations for verticals to adopt private LTE/5G networks.

Table IV-1. Demand-side factors for private LTE/5G (source: Analysis Mason). Link.

Factor	Description
Operational efficiency	The demand for private LTE/5G networks is growing because large organisations’ digital transformation programmes are underway. Enterprises

	are in the process of digitizing their data and using it to drive processes and create new digital products and services
ICT and OT convergence	The convergence of Information & Communication Technology (ICT) and Operational Technology (OT) is also a key consideration. Ultimately, the need for high-bandwidth, low-latency networks to support increased automation will grow as enterprise data processing requirements increase.
Data privacy	Enterprises deploy private networks because data privacy is a key concern. They require more control and visibility of their data.
Cable substitution	Enterprises deploy private LTE/5G networks to support new applications as a more cost-effective alternative to extending their fixed networks.
Replacing legacy networks	Existing networks such as TETRA are reaching the end of their life and cellular technologies offer viable alternatives.
Wi-Fi limitations	Enterprises have used Wi-Fi successfully but have found that it has limitations in terms of supporting mobility and/or other factors such as reliability.

1.1 Public vs private mobile networks

Mobile network operators are today the main communication service providers (CSPs). Economies of scale permit national mobile operators to take on the heavy cost of building (CAPEX) and maintaining (OPEX) these networks to deliver nationwide coverage and mobile data services to a vast market of subscribers on domestic and business plans. However, when coming to industrial environments and mission-critical services such as public safety, government institutions, energy or banking, there exists a reluctance to (only) rely on mass-market public networks. The reasons behind are diverse, though they can be summarized in the following points:

- **Coverage limitations.** Mass-market networks do not provide ubiquitous coverage in every corner of the nation; actually, public coverage is determined by where the mobile network operator sees it worthwhile to provide. While the promise of ubiquitous connectivity might hold true in dense urban areas (where the investment in new radio nodes can be amortized over a short period of time), public coverage remains inconsistent or absent altogether in many remote industrial areas. Additionally, depending on the use case under consideration, indoor and underground coverage might also be in question. A clear example of these environments can be found in most logistics hubs, typically installed on the outskirts of urban centers. With this rationale, conditioning digital transformation of these hubs by the sole reliance on public coverage is not a good path.
- **Network longevity.** Network equipment and solutions for public mobile networks are typically designed for long-term use. Once installed, they remain there for years, or even decades without the need to be replaced. The reason is that regulators force mobile operators to offer essential connectivity capabilities worldwide, including voice and messaging services. This justifies why legacy equipment (2G, 3G) is still present in mass-market networks, and their need to coexist with the assets from newest technologies (4G, 5G). The result is a brownfield environment. The lifecycle of products used at industry and other mission-critical sectors is much shorter; indeed, the replacement of these products occurs at quicker pace, due to their deprecation in terms of features (their capabilities no longer satisfy the ever-evolving requirements of industry and mission-critical services) or security (they are not able to offer protection against latest cybersecurity attacks). This cadence is not compatible with

the CAPEX strategies of mobile network operators, whereby every upfront cost shall ensure revenue streams and/or cost savings sustained in the medium and long term.

- **Low customization.** Mass-market networks provide a similar service experience for different applications and customers; actually, there is a lack of differentiation in their traffic handling, largely imposed by net neutrality regulation. In addition, the functions and service platforms building up the public networks are typically enabled with a very low number of configuration options, either because of lack of features on vendor solution or due to operator's decision (e.g., disabling certain configuration settings greatly simplifies the day-to-day network operation). A clear example can be observed in the QoS. Although 3GPP defines a wide variety of values for QoS Configuration Indicators (QCI) and 5G QoS Indicators (5QI) parameters, only a small portion of these values is implemented in public networks, typically the three or four ones which fit the eMBB service requirements from B2C segment. The same occurs with other features such as mobility management and security. Vertical services exhibit requirements that cannot be met with the pre-defined yet limited settings existing in public networks; and it is unlikely that mobile operators would like to fine tune their large-scale networks to satisfy the differentiated requirements of individual customers.
- **Shared network usage and fault tolerance.** Mass-market networks are designed for best-effort usage, with all users sharing resources with everyone else. This means that when the network is overloaded, every mobile subscriber performance hits, including end users and business organizations. For those situations where the root of network misbehavior is other than the rush hour, for example a node malfunctioning or a hacking attack, the result can be a massive outage across public connectivity services. This is not tolerable for industry and mission-critical services, which call for a high reliability communication infrastructure. For undisrupted operations, many vertical industries strive for nearly 100% network uptime, which is often unrealistic for commercial carriers. This is a business case that justifies the use of private networks for critical infrastructure.
- **Security and data privacy.** Industry and mission-critical services process and manage quite sensitive information, including business-critical and personal data. The creation and movement of treasure troves of valuable data will draw the attention of cyber criminals, making robust security and data privacy a fundamental necessity, while compliance regulations will make it a responsibility. In this context, stakeholders of these services look for two things: i) to have end-to-end visibility of data flows, from source to destination; ii) to retain full and complete authority for cybersecurity decision-making. These two ambitions are incompatible with the use of public networks. Actually, in public networks data must always be routed through the operator's backend before reaching the end server or application platform, thus violating (i). In addition, carrier networks have built-in security and data protection mechanisms that do not always match the requirements of these stakeholders, some of them tied to the industry-specific regulation policies (e.g., ISO standards for industry 4.0), thus violating (ii). In such a situation, stakeholders prefer to go for on-premises solutions.
- **No control or maintenance.** Configuration and upgrades are performed according to schedules determined by the cellular provider, not its customers. This means that owners of industry and mission-critical services have no control on the Operation, Maintenance, and Administration (OAM) activities on the network. In this context, the possibility of verticals to innovate is quite limited. The reason is that their ability

to bring new services and applications is subjected to the activation of certain network capabilities, being the latter a decision that is entirely up to the operator; in fact, the operators decide when and how to activate every network capability in the mass-market networks, according to their individual feature and testing roadmap.

The reasons elaborated above motivate the ever-increasing adoption of private mobile networks for industry and mission-critical services. Private mobile networks provide stakeholders of these services (e.g., industry verticals, government, etc.) with total control of the network. They can define who can access the network, when they can access, which QoS they experience, what kind of coverage the network has and how much capacity is available. Likewise, they can freely implement preferential traffic treatment across services, since net neutrality and other regulation policies do not apply.

1.2 Today's private mobile network solutions

Private networks are not merely a theoretical construct; in fact, there already exists production-grade solutions in the market. Examples include Private LTE and Public Safety Mobile Broadband networks.

1.2.1 Private LTE

Private LTE networks are currently a commercial reality [1]. Private LTE is a miniature version of a public LTE network and functions much like its macro cousin, though with some differences: i) it covers a much smaller geography, typically a localized area such as a stadium, factory or mine; ii) it has built-in features can be enabled/disabled *à la carte*, allowing for customizability in terms of performance and functionality; and iii) it allows data to stay on premises, offering a potential advantage for many stakeholders requiring high levels of data security and privacy. In a nutshell, private LTE networks work the same as mass-market 4G networks but are *designed, deployed, operated and optimized* for their use by a private organization instead of a traditional mobile carrier.

The physical and operational separation from public commercial networks has motivated the use of private LTE in utility sectors such as energy, which needs to be built upon resilient and independent infrastructures, with fault tolerance even to natural disasters. Works in [2]-[7] provide some examples on how private LTE can be used in smart grid. Other market segments where this technology has gained traction over the last three years include factories [8], mines [9] and smart ports [10]. And the market share is expected to grow in the upcoming years, especially across small and medium enterprises (SMEs). The fact that private LTE can offer tailor-made 4G capabilities with minimal infrastructure build-out makes it attractive for enterprises and organizations seeking a cost-effective yet high-quality solution to have an on-premises private cellular network.

One of the major constraints of private LTE is the spectrum availability. In many countries, *licensed spectrum* is already congested. This means either the stakeholder needs to pay a premium to be granted an existing block of spectrum, or wait until someone else gives up their frequencies. Alternatives include using *unlicensed spectrum*, with technologies such as MultiFire [11], or *shared spectrum*, with models such as the Citizen Broadband Radio Service (CBRS) in the U.S. [12] and Licensed Shared Access (LSA) in Europe [13].

1.2.2 Public Safety Mobile Broadband

These networks are a specific type of private LTE network – one dedicated to public safety and first responders. They are not available to other types of critical communications users and are generally government-funded [14]. Examples include FirstNet in the U.S. [15], Emergency Service Network (ESN) in the United Kingdom [16], and SafeNet in South Korea [17]. The aim of such systems is to replace old DMR solutions, providing public safety agencies secure, nationwide, and interoperable communications with voice and broadband data.

1.2.3 Business Ecosystem

When landing private mobile networks, three main groups arise.

1. **Mobile network operators (MNO).** Apart from delivering broadband services with mass-market networks, carrier operators also have market share in private networks. For example, with private LTE, they can provide additional services to customers outside of the licensed spectrum they already have.
2. **Neutral hosts,** which own and manage infrastructures installed in public venues such as stadiums, museums, or transportation hubs (e.g., airports, underground stations). The venue owner invests on the infrastructure to provide internal services, but also to lease it to MNOs, for them to provide coverage to end-users in these venues. In the latter case, the venue owner plays the role of neutral host. A relevant example is a recent tender for the London Underground, where the contract secures a 20-year concession to provide a neutral host network in the tunnels, stations, and platforms [18].
3. **Industry verticals and government institutions,** which have their own private networks to provide connectivity to their factory, mine, campus, or utility service area. Though most of these organizations prefer to operate the private network infrastructures themselves, not all of them have skilled workforce; in other words, not all organizations have their own IT and networking teams. In such a case, these organizations hand over OAM activities to a managed service provider. Examples of managed service providers are i) the business units of mobile network operators, ii) the business units of vendors, or iii) solution integrators specialized in private networking.

1.3 Private 5G: As capable as wires, but without the wires

The previous section has surveyed the usage of private LTE in industrial applications from different sectors. However, industries have increasingly stringent performance requirements regarding throughput, latency, reliability, availability, security, and device density [19], which private LTE systems cannot meet. The reason is that the target KPIs and features are well beyond the capabilities of 4G technology; in this situation, 5G technology enters to scene.

The first phase of 5G deployments started during the second quarter of 2019. These early deployments are based on 3GPP Rel-15 in Non-Standalone (NSA) mode, typically using a carrier at 3.5 GHz, and will be destined to support enhanced Mobile Broadband (eMBB) use cases. Thus, the first phase of 5G deployments can be seen as a natural evolution of the 4G mobile broadband service, not yet delivering on the capabilities required by vertical services. According to Sylvain Fabre, senior research director at Gartner, "*In the short to medium term, organizations wanting to leverage 5G for use cases such as IoT communications, video, control and automation, fixed wireless access and high-performance edge analytics cannot fully rely on 5G public infrastructure for delivery*" [20]. Again, according to Mr. Fabre, "*most operators will only achieve a complete end-to-end*

5G infrastructure (including URLLC and edge computing) on their public networks during the 2025-to-2030 time frame".

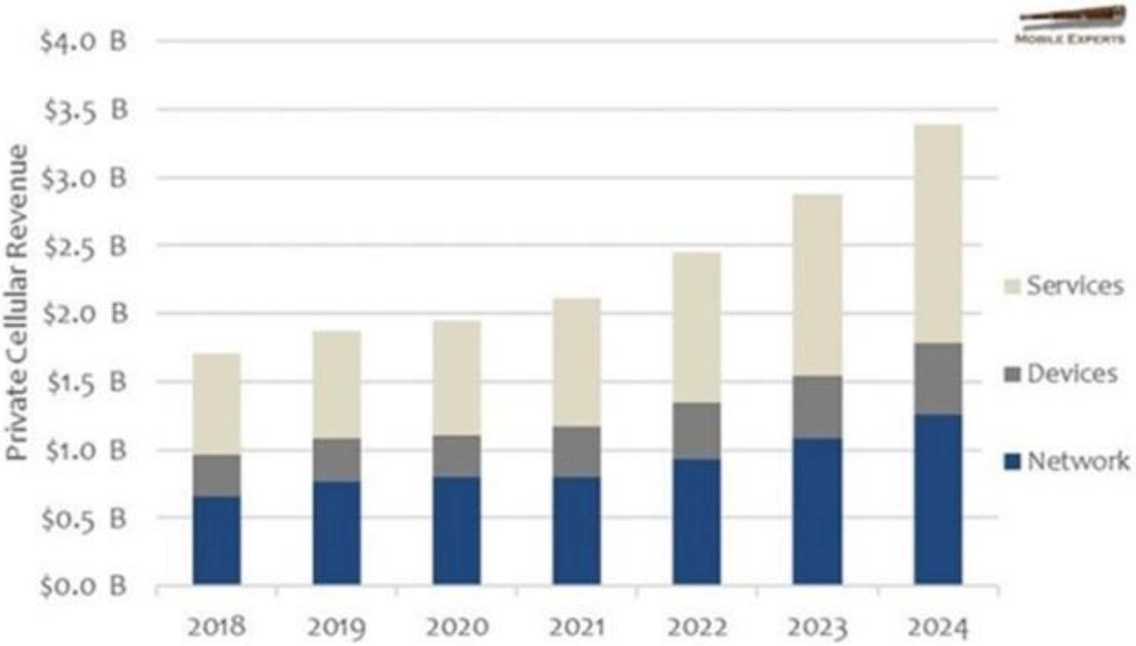
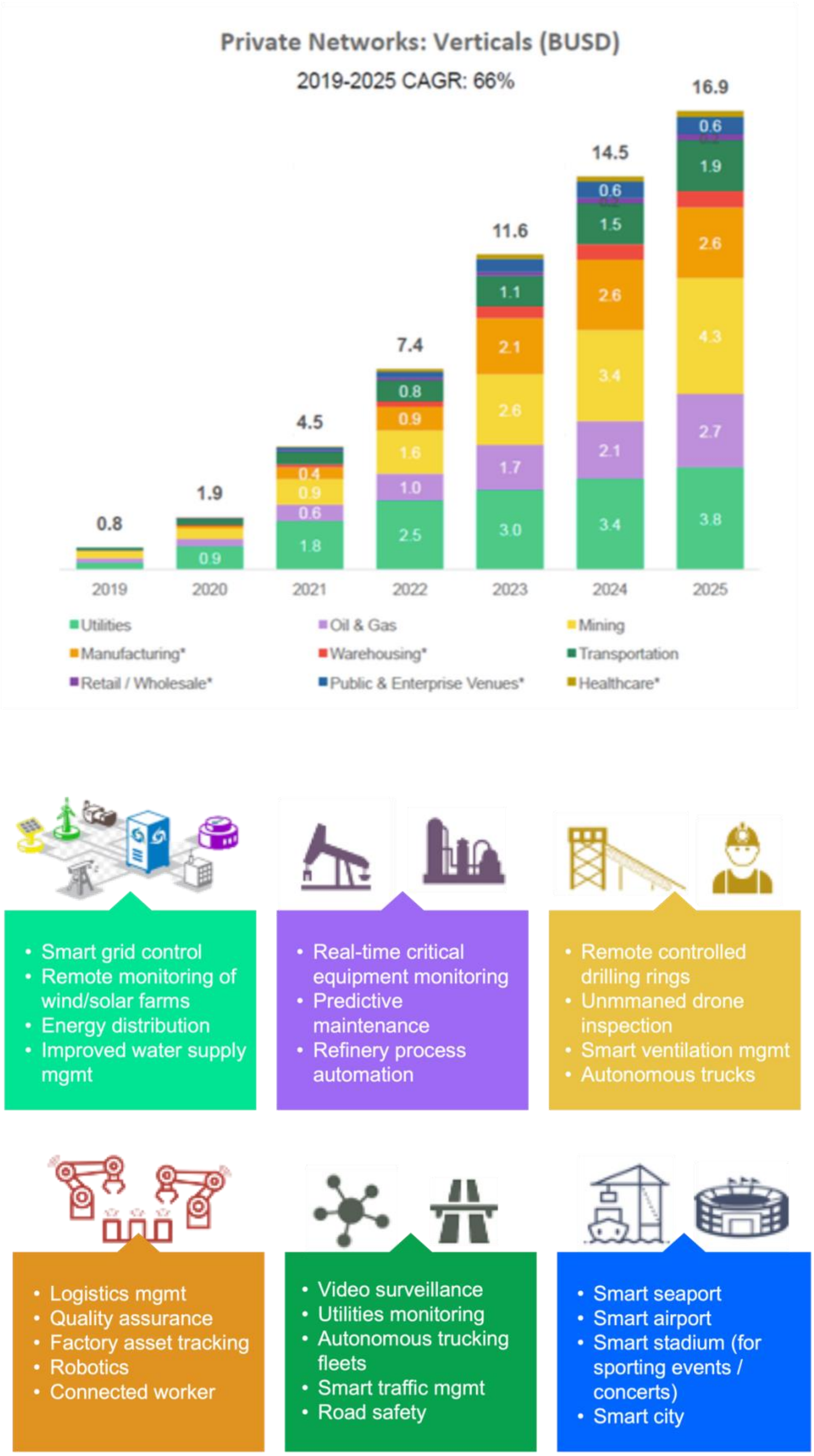


Figure IV-1. Market share for private 5G networks. Source: [21].

However, vertical use cases are demanding 5G critical communications early on. The answer to this market need is private 5G networks. A report by Mobile Experts [21] forecasts a growth of the private 5G markets at over 10 percent CAGR to $3.4 billion by 2024, with Figure IV-1 depicting the expected market share among network, devices, and services. Other market estimates for private 5G networks are even more optimistic and forecast a CAGR growth of approximately 30% between 2018 and 2021, eventually accounting for more than $5 billion by the end of 2021 [22].

From a technical viewpoint, the concept of private 5G network is known as non-public network (NPN) [23]. This term was coined by 3GPP in early 2019, with the release 16 kick-off, to refer to the use of 5G technology in a private mobile network environment. This sets a demarcation point with private LTE (Rel-13 and Rel-14) and the first generation of 5G networks (3GPP Rel-15), entirely focused on eMBB. Unlike these past releases, 3GPP Rel-16 focus is on uRLLC and mIoT capabilities, with low latency, high reliability and massive yet energy-efficient connectivity as core principles.

Figure IV-2 pictures the industries that would benefit most from private 5G, together with potential use cases. One can notice that the trend observed for private LTE remains, with utilities, mining, and industry 4.0 covering most of market share. The latter one is a critical sector ready to be disrupted by private 5G. Indeed, the 5G-PPP has identified Factories of the Future as a key vertical sector, identifying five main use case families where 5G is poised to disrupt manufacturing [24], including: i) control of time-critical processes inside the factory, ii) non-time-critical in-factory communications, iii) remote control, iv) intra/inter enterprise communications, and v) connected goods. Other efforts, like the 5G Alliance for Connected Industries and Automation (5G-ACIA), are gathering major European industrial and telecom players and are also working towards the vision of 5G factories [25].



- Smart grid control
- Remote monitoring of wind/solar farms
- Energy distribution
- Improved water supply mgmt

- Real-time critical equipment monitoring
- Predictive maintenance
- Refinery process automation

- Remote controlled drilling rings
- Unmanned drone inspection
- Smart ventilation mgmt
- Autonomous trucks

- Logistics mgmt
- Quality assurance
- Factory asset tracking
- Robotics
- Connected worker

- Video surveillance
- Utilities monitoring
- Autonomous trucking fleets
- Smart traffic mgmt
- Road safety

- Smart seaport
- Smart airport
- Smart stadium (for sporting events / concerts)
- Smart city

Figure IV-2. Vertical industries in private 5G. Adapted from Ericsson’s report.

Industry 4.0 requirements cannot be addressed with private LTE or 3GPP Rel-15, and thus factories have so far relied primarily on cable-based connectivity with very limited flexibility. However, the irruption of the 3GPP Rel-16 and beyond will provide an excellent opportunity for 5G to disrupt factory automation, replacing proprietary and costly Ethernet solutions with a rich-featured NPN. To make it happen, 3GPP is developing industry specific features for R16 and beyond such as uRLLC featured 5GNR design, support of Time Sensitive Networks (TSN) over 5G to wirelessly synchronize industrial machinery, and flexible usage of spectrum [26].

1.4 Putting all together

Market analysts agree that private 5G may be a big leap for industry digitalization, but also coincide that it is not the only technology that works. For uses and environments, private LTE will do simply fine, and we can expect companies to continue to build private mobile networks using it, as reported in Section 1.2.1. For example, Nokia has used private advanced LTE networks (4.9G) to automate one of its base station factories [27]. Another alternative that remains valid is Wi-Fi. Wi-Fi deployment is fast, easy, and cheap compared to private cellular networks, making it an attractive choice where speed and economy are a priority. Private Wi-Fi networks are already used in factories, typically for non-critical applications. New Wi-Fi standards, including Wi-Fi 6, are being launched, offering significant enhancements as reported in [28][29].

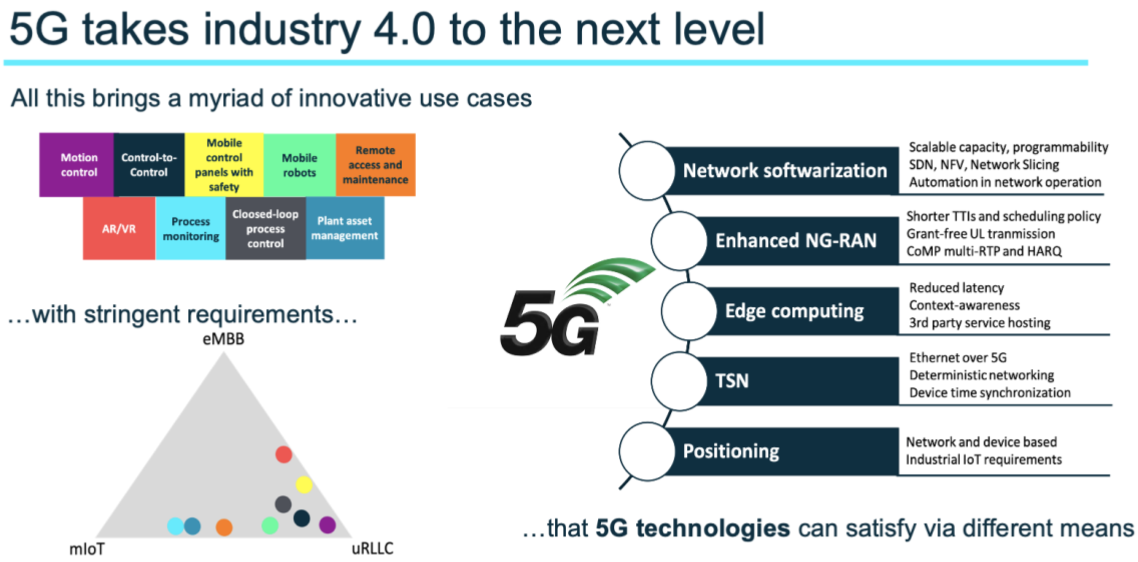
Many industry reports have focused over the last two years on making comparative analysis between industrial Wi-Fi, private LTE and private 5G, capturing their pros and cons in terms of performance and cost. Examples of these analysis can be found in [30]-[34]. The results show that it is not an either-or situation; at the end of the day, the reality is that private LTE/5G will continue to coexist with Wi-Fi and other legacies. Ultimately, enterprises should focus on understanding where and when each connectivity solution is best suited, find potential network synergies and take advantage of any opportunity for consolidation. In this regard, private 5G has considerable ground to make up. The reason is that Rel-16 5G networks exhibit inherent multi-connectivity features thanks to the introduction of functions such as Non-3GPP Interworking Function (N3IWF) and Trusted Non-3GPP Gateway Function (TNGF), which allows seamless integration of 5G with other technologies [35]. In other words, private 5G can provide not only 3GPP Rel-16 and beyond features, but also act as an aggregator/integrator of other connectivity solutions.

2 Ambition

2.1 Profiling Private 5G in industry 4.0

3GPP coined the term NPN in their specifications in early 2019. The NPN concept refers to a Rel-16 5G network that is intended for non-public use, i.e., for exclusive use of particular enterprise or organization. As captured in [23], a NPN can be either: i) a *Standalone NPN*, i.e., an NPN which does not rely on network functions provided by the public cellular network; or ii) a *Public Network Integrated NPN*, i.e., an NPN deployed with the support of the public cellular network. Within these two NPN categories, 3GPP recognizes that there can exist different types of deployment options, with the possibility of the deployed NPN to include both virtual and physical elements. However, the literature does not provide answers to the following questions:

- What are the dimensions/aspects that articulate the definition of these deployment options?
- What are the decision criteria for an enterprise customer to go for one or another deployment option?



5G takes industry 4.0 to the next level

All this brings a myriad of innovative use cases

…with stringent requirements…

…that **5G technologies** can satisfy via different means

Motion control
Control-to-Control
Mobile control panels with safety
Mobile robots
Remote access and maintenance
AR/VR
Process monitoring
Cloosed-loop process control
Plant asset management
eMBB
mIoT
uRLLC
5G
Network softwarization
Scalable capacity, programmability
SDN, NFV, Network Slicing
Automation in network operation
Enhanced NG-RAN
Shorter TTIs and scheduling policy
Grant-free UL tranmission
CoMP multi-RTP and HARQ
Edge computing
Reduced latency
Context-awareness
3rd party service hosting
TSN
Ethernet over 5G
Deterministic networking
Device time synchronization
Positioning
Network and device based
Industrial IoT requirements

Figure IV-3. Private 5G and Industry 4.0.

On the other hand, as commented in Section 1.3, industry 4.0 is one of the market segments that may benefit most from private 5G (see Figure IV-3). So this segment is ideal to elaborate on the two open questions above.

Beyond the state-of-the-art: This thesis will provide answers to the two open questions listed above, and use them for reporting on NPN solutions targeting industry 4.0.

- In relation to the first question, examples of dimensions/aspects that govern the deployment of a NPN include i) *NPN location*, i.e. where the individual NPN functions are deployed, either on-premises or on the public network; ii) *NPN management*, i.e. who are the stakeholders that leads the OAM activities for the NPN; and iii) NPN *spectrum nature*, i..e unlicensed, licensed but leased by the MNO, licensed but issued by the local regulator. Based on the identified dimensions/aspects, this thesis will identify the NPN flavors that makes more sense in the industry 4.0 ecosystem. Ecah flavor represents a different deployment option.
- In relation to the second questions, examples of decision criteria may include performance, security, integration efforts and cost figures, among others. Based on these decision criteria, this thesis will provide a comparative analysis of the different NPN flavors, assessing pros and cos.

Related objectives: O3.1.

Means of verification: Paper F

2.2 Robust, scalable and future-proof solutions for private 5G in the vertical industries ecosystem

The integration of NPN in 3GPP Rel-16 feature roadmap, along with the foundation of 5G-ACIA [25], contributed that private 5G started to gain momentum in industry and academia.

5G-ACIA has always been the main driver for the development of private 5G solutions. With a consortium formed of telco operators, industry stakeholders and regulators, the 5G-ACIA released a number of white papers that set the course for private 5G evolution from the very beginning. Captured in [36]-[41], these papers served as an inspiration for further work in other bodies, including:

- Telco industry organizations, embracing GSMA, 5G Americas, TM Forum and Global mobile Suppliers Association (GSA). In [42], GSMA provides insights on the use of 5G private and dedicated networks for manufacturing, production, and supply chains, with some examples on real-world customer success stories. This work was followed by [43], which reports pathways for the replacement of private LTE with private 5G in campus networks. 5G Americas generated two white papers on private 5G networks: one in 2020 [44] and another in 2021 [45]. TMForum provides in [46] a study that captures the main opportunities the mobile network operators can embrace with private 5G, when acting as managed service providers (see Section 1.2.3). Finally, GSA publishes yearly reports on private 5G market status, with the first one released in 2020 [47].
- Standardization development organizations, mainly 3GPP. 5G-ACIA has input use cases and service requirements to 3GPP SA1 (services), which sets the basis for the specification and development of solutions at 3GPP SA2 (architecture), 3GPP SA3 (security) and 3GPP SA5 (management, orchestration, and charging), in the Rel-16 and Rel-17 timeframe. SA2 work on NPN has been reported in TR 23.700-7 [48], with normative solutions captured in TS 23.501 [23]. SA3 has studied security in NPNs, documenting key issues, requirements, and potential solutions in TR 33.857 [49], which were compiled later in [50]. Finally, SA5 has reported the operational aspects related to NPNs in two documents: one informative, TR 28.807 [51], and other normative, TS 28.557 [52].

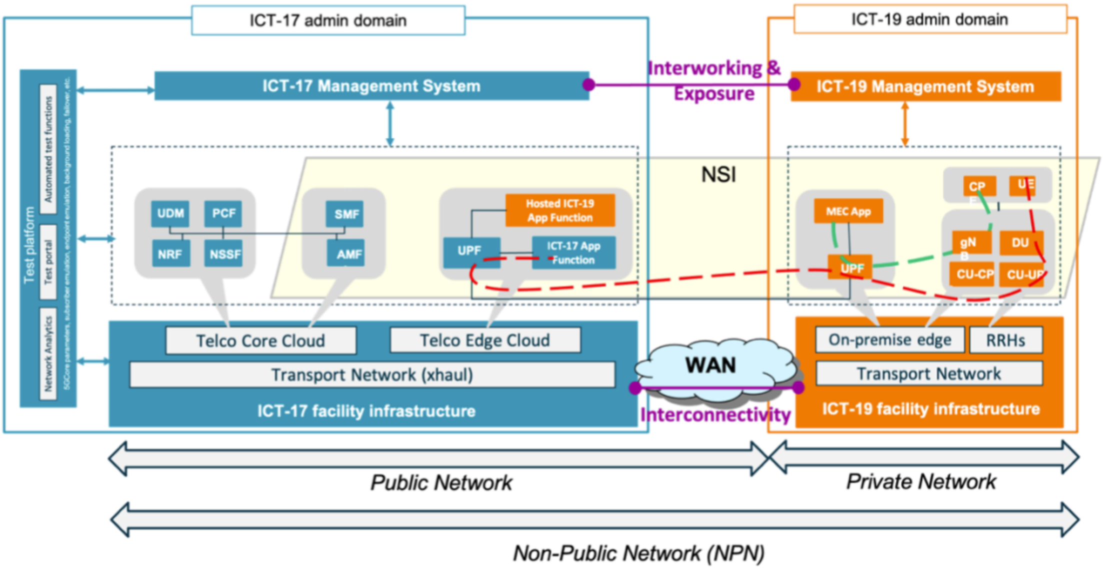


Figure IV-4. Combined use of ICT-19 and ICT-17 facilities to deliver NPN services.

Apart from this industry work, there exist several research and innovation activities around 5G private networks. These activities have crystallized in different academia papers [53]-[56] and demonstration activities in 5G-PPP Phase 3 projects. For example, ICT-19 projects such as 5Growth [57]and 5G-Tours [58] make use of ICT-17 facilities (i.e., 5G-VINNI, 5GEVE and 5Genesis) to work on solutions scoping public network integrated

NPN scenarios. Figure IV-4 illustrates how this is done, with ICT-19 facilities behaving as private network sites, and ICT-17 facilities acting as public network nodes. In addition, ICT-20 projects such as 5G-CLARITY [59], Affordable5G [60] and FUDGE-5G [61] are also working on advanced wireless solutions enabling private 5G networks.

The review conducted thus far demonstrates the ‘hype’ of private 5G. Nevertheless, there exist some gaps that are worth mentioning. The *first and main problem* is that the use cases, requirements, and technologies driving NPN solutions are quite scattered through the literature. For example, we have a plenty of capabilities such as TSN, edge computing, E2E orchestration, 5GC and RAN sharing, that need to be bundled together in NPNs; however, most of works address them separately, hence the difficulty in having the full picture. The *second problem* is that state-of-the-art literature puts the focus on industry 4.0 sector; however, there are other industry sectors and corporate enterprises that will bring additional service requirements, such as mobility support in NPNs spanning multiple sites. These requirements are not considered for the design of solutions in the literature. Finally, the *third problem* is that not all the works lay on a common ground for terminology and assumptions, which complicates comparative analysis and the search for synergies.

Beyond the state-of-the-art: This thesis will overcome the limitations identified from the conducted literature review, by:

- identifying the technology facilitators for NPNs, outlining their main capabilities, and putting them into a common framework.
- designing robust, scalable, and future-proof architectures for NPN, applicable to market sectors beyond industry 4.0. Different use case-driven architecture solutions will be outlined and characterized, with a special focus on mobility and multi-site use scenarios. The proposed architectures will be built out using standard solutions, in order to i) minimize integration efforts, when combing different technologies and capabilities altogether; and ii) implement common provisioning and operation patterns, which allows for much easier replicability when these private networks are to be deployed at large scale.
- reporting on the main challenges that industry may need to face for unleashing NPNs, and proposing way forward to address them, either via research or via standardization.

Related objectives: O3.1, O3.2.

Means of verification: Paper G

2.3 Provision of private 5G networks with network slicing

Though (most of) the ingredients enabling private 5G are already here, industry actors are still looking for the recipe to ensure success in the market. All actors taking part in the 5G value chain, from chipset and network equipment providers up to industry customers, need to face challenges within their scope of work. From the MNOs’ viewpoint, challenges ahead include:

- **Commercial related challenges.** Selling private 5G network is not easy for operators since most enterprise customers are not willing to assume upfront costs of getting a private 5G network up and running. In addition, the emergence of new actors in private mobile network landscape, such as solution integrators and hyperscalers [62] would put at risk the revenue streams for operators; in fact, these incumbents also want to uptake the role of private network and managed service providers, occupying the same market space as business units of operators. In this competitive landscape, operators' commercials need to move from (traditional) transactional to (modern) consultative approaches.

- **Product related challenges**. Solutions for private 5G networks (e.g., on-site 5GC) are lightweight versions of products originally designed for large-scale deployments (e.g., mass-market 5GC). The cost structure of these lightweight versions does not fit well with the scalability and customizability needs of private 5G networks, making them unattractive for early adopters.
- **Technological challenges**. These embrace the need for i) a tighter app-to-network integration, by making telco capabilities available to OT services; ii) legacy equipment/protocols to coexist with 5G and cloud capabilities, with much lower integration efforts; and iii) finding network provisioning and operation patterns that facilitate replicability, allowing for economies of scale (private 5G networks are today provisioned *à la carte*, with long cycles from their design to their set up).

Network slicing is a solution that can help operators to face these challenges. To make it happen, it is important for operators to deliver network slices offering performance and security comparable to those provided by today's standalone private 5G solutions (e.g., guaranteed SLA, traffic separation, controllable and configurable network), but a much more reduced cost. This, together with the flexibility and agility that network slicing will bring (see Figure I-4,) is what will drive vertical industries and mission-critical organizations to ask the operator for a dedicated network slice (i.e., PNI-NPN) rather than going for a full on-premises solution (i.e., SNPN). Actually, these ideas were already anticipated in the Section 3 from Part I.

Network slicing is an E2E solution, with many network and management domains involved. However, integrating all the capabilities that are needed to offer a full-blown network slice as a service (NSaaS) [63] may require years. In the meantime, operators need to look for workarounds to start commercializing and monetizing network slicing for the private network ecosystem in the short term, to gain traction from early adopters. Once first customers are captured and market share is ensured, operators can think about extending their network capabilities in the medium and long term to make slicing a more scalable and dynamic solution, becoming the preferred option (in terms of cost and flexibility) for the provision of private networks.

The path that the operators must take to offer network slicing solutions in the short, medium, and long term is key to ensure success in the market of private 5G networks. However, the population of this roadmap not an easy task, given the quite fragmented landscape in standards (see Figure I-12) and literature, with plenty of ad-hoc solutions that cover particular slicing aspects from different domains, and under different assumptions, both in time and scope. This calls for action.

Beyond the state-of-the-art: This thesis will overcome the problem stated above, by identifying the solutions that may help operator to provision private 5G networks with network slicing. These network slicing solutions will be captured in a radar, whose mission is articulated into three tenets:

- link identified solutions to different timelines (as-is, short-term, medium-term, and long-term), based on their technical viability and customer demands.
- specify the dimensions that have an impact on the usability (how and where) of these solutions, across all operator managed domains, including (i) radio access network, (ii) transport network; and (iii) core network; and (iv) orchestration.
- elaborate on the capabilities that individual solutions can provide to private networks.

Related objectives: O3.3
Means of verification: Paper H

References

[1] R. Ferrus and O. Sallent, "Extending the LTE/LTE-A Business Case: Mission- and Business-Critical Mobile Broadband Communications," *in IEEE Vehicular Technology Magazine*, vol. 9, no. 3, pp. 47-55, Sept. 2014. DOI: 10.1109/MVT.2014.2333695.

[2] Y. Sun, Y. Zhong and M. Wang, "Energy-efficient coordinated multi-point transmission for centralized power wireless private TD-LTE network", in *2016 China International Conference on Electricity Distribution (CICED)*, 2016, pp. 1-4. DOI: 10.1109/CICED.2016.7576124

[3] W. Miao et al., "Coverage and capacity analysis of LTE-based power wireless private network," in *2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID*), 2017, pp. 119-124. DOI: 10.1109/ICASID.2017.8285756.

[4] G. Li-Yuan, T. Xiao-Xu, M. Bao-Kun, C. Jin-Pei and Y. Xi-Ming, "Analysis of Service Load and Security of Power TD-LTE Wireless Private Network," in *2018 China International Conference on Electricity Distribution (CICED)*, 2018, pp. 387-392. DOI: 10.1109/CICED.2018.8592545.

[5] P. Ma et al., "Challenges and Solutions in Current Power Wireless Private Network," in *2019 4th International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, 2019, pp. 177-179. DOI: 10.1109/IGBSG.2019.8886311.

[6] Y. Liang et al., "Resource Allocation for Multi-class Businesses in LTE-A Uplink Communication for Smart Grid", in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019, pp. 1127-1132. DOI: 10.1109/IWCMC.2019.8766602

[7] Y. Guo, D. Zhu, L. Wei, H. Guo, S. Zhu and L. Feng, "Suitability Evaluation of LTE-based Wireless Private Network for Power Communication Business in Smart Grid," in *2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2019, pp. 704-709. DOI: 10.1109/SDPC.2019.00133.

[8] "Etteplan's private network as a service enables the smart factory of the future at Danfoss", Etteplan press release, 2020 [Online]. Link

[9] J. Blackman, "Three private LTE deployments in the mining industry", Enterprise IoT insights press release, 2019 [Online]. Link

[10] "Enabling a Digital port with private LTE at Rotterdam", Ericsson press release, 2019 [Online]. Link

[11] MulteFire Alliance [Online]. Link

[12] CBRS Alliance [Online]. Link

[13] L. Varukina, "Licensed Shared Access ", Nokia presentation [Online]. Link

[14] Tait communications, "Commercial Public LTE vs Private LTE vs Public Safety Mobile Broadband", 2020 [Online]. Link

[15] FirstNet [Online]. Link

[16] UK Government, "Emergency Services Network: overview" [Online]. Link

[17] Korean Ministry of the Interior and Safety, “Disaster and Safety Communications Network (Korea Safe-net)” [Online]. Link

[18] Computer Business Review, “London Underground Kicks Off 4G Tender Process”, 2018 [Online]. Link

[19] S. Doğan, A. Tusha and H. Arslan, "NOMA With Index Modulation for Uplink URLLC Through Grant-Free Access," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1249-1257, Oct. 2019. DOI: 10.1109/JSTSP.2019.2913981.

[20] Gartner, “Garnet Surveys Reveals Two-Thirds of Organizations Intend to Deploy 5G by 2020”, [Online]. Link

[21] Mobile Experts, “Private LTE and 5G”, 2019 [Online]. Link

[22] Prnswire, “Private LTE & 5G Network Ecosystem Market Grow More Than $5 Billion by 2021 and Forecast to 2030”, 2019 [Online]. Link

[23] 3GPP TS 23.501, “5G; System Architecture for the 5G System (5GS)”

[24] 5GPPP, “5G and the Factories of the Future”, 2015 [Online]. Link

[25] 5G-ACIA, “5G Alliance for Connected Industries and Automation”, 2019 [Online]. Link

[26] 3GPP List of Rel-16 Work Items, 2019 [Online]. Link

[27] Nokia, “Nokia’s digitalization of its 5G Oulu factory recognized by the World Economic Forum as an “Advanced 4th industrial revolution lighthouse,” Press release, July 3, 2019. Link

[28] D. Staley - IEEE 802.11 Working Group Chair, “IEEE 802.11 Standards: Wi-Fi 6 and Beyond”, *JPL Workshop*, October 2019 [Online]. Link

[29] E. Khorov, A.Kiryanov, A.Lyakhov, and G.Bianchi, “A tutorial on IEEE 802.11ax high efficiency WLANs,” in *IEEE Communications & Surveys Tutorial.*, vol. 21, no. 1, pp. 197–216, 2019Q1. DOI: doi: 10.1109/COMST.2018.2871099.

[30] Deloitte Insights, “Private 5G networks: Enterprise untethered - TMT Predictions 2020”, December 2019 [Online]. Link

[31] M. Bamforth, “Private 5G vs Wi-Fi vs Private LTE”, STL Partners, 2021 [Online]. Link

[32] Celona, “Private LTE vs Wi-Fi for Enterprise: Comparison & Use Cases”, 2021 [Online]. Link

[33] T. Crick, “Private LTE/5G Versus WiFi 6 for In-Building/Campus Wireless”, Astra Capital Management, August 2020 [Online]. Link

[34] M. Addicks, “Private 5G vs Wi-Fi: The New Wave of Wide-Area LAN”, September 2021 [Online]. Link

[35] 3GPP TS 24.502, “5G; Access to the 3GPP 5G Core Network (5GCN) via non-3GPP access networks”.

[36] 5G-ACIA, “5G for Connected Industries and Automation”, White Paper, Feb 2019 [Online]. Link

[37] 5G-ACIA, “5G for Automation in Industry”, White Paper, July 2019 [Online]. Link

[38] 5G-ACIA, “5G Non-Public Networks for Industrial Scenarios”, White Paper July 2019 [Online]. Link

[39] 5G-ACIA, “Integration of Industrial Ethernet Networks with 5G Networks”, White Paper, November 2019 [Online]. Link

[40] 5G-ACIA, “A 5G Traffic Model for Industrial Use Cases”, White Paper, November 2019 [Online]. Link

[41] 5G-ACIA, “Key Use Cases and Requirements”, White Paper, November 2019 [Online]. Link

[42] GSMA, “5G IoT Private & Dedicated Networks for Industry 4.0”, October 2020. [Online]. Link

[43] GSMA PRD NG.123, “5G Industry campus network deployment guideline”, version 1.0, November 2020 [Online]. Link

[44] 5G Americas, “5G Technologies in Private Networks”, White Paper, October 2020 [Online]. Link

[45] 5G Americas, “Private & Enterprise Networks”, White Paper, August 2021 [Online]. Link

[46] TM Forum, “Mobile private networks: Exploring the CSP opportunity”, June 2021 [Online]. Link

[47] GSA, “Private Mobile Networks: market status update”, October 2020 [Online]. Link

[48] 3GPP TR 23.700-07, “Study on enhanced support of Non-Public Networks (NPN)”.

[49] 3GPP TR 33.857, “Study on enhanced security support of Non-Public Networks (NPN)”.

[50] A. Jerichow et *al.*, “3GPP Non-public Network Security," *Journal of ICT Standardization*, vol. 8, no. 1, pp. 57–76, 2020.

[51] 3GPP TR 28.807, “Study on management of Non-Public Networks (NPN)”

[52] 3GPP TS 28.557, “Management and orchestration; Management of Non-Public Networks (NPN); Stage 1 and 2”.

[53] A. Rostami, "Private 5G Networks for Vertical Industries: Deployment and Operation Models", *in 2019 IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 433-439. DOI: 10.1109/5GWF.2019.8911687.

[54] I. Badmus, M. Matinmikko-Blue, J. S. Walia and T. Taleb, "Network Slice Instantiation for 5G Micro-Operator Deployment Scenarios," *in 2019 European Conference on Networks and Communications (EuCNC)*, 2019, pp. 133-138. DOI: 10.1109/EuCNC.2019.8802013.

[55] A. Aijaz, "Private 5G: The Future of Industrial Wireless," *in IEEE Industrial Electronics Magazine*, vol. 14, no. 4, pp. 136-145, Dec. 2020. DOI: 10.1109/MIE.2020.3004975.

[56] E. C. Strinati et al., "Beyond 5G Private Networks: the 5G CONNI Perspective," 2020 IEEE Globecomm Workshops (GC Wkshps, 2020, pp. 1-6. DOI: 10.1109/GCWkshps50303.2020.9367460.

[57] H2020 5Growth project [Online]. Link

[58] H2020 5G-Tours project [Online]. Link

[59] H2020 5G-CLARITY project [Online]. Link

[60] H2020 Affordable-5G project [Online]. Link

[61] H2020 FUDGE-5G project [Online]. Link

[62] AWS Private 5G Networks [Online]. Link

[63] 3GPP TS 28.530, “Management and Orchestration; Concept, use cases and requirements”.

Paper F

The use of 5G Non-Public Networks to support Industry 4.0 scenarios

Jose Ordonez-Lucena, Jesús. Folgueira, Luis M. Contreras and Antonio Pastor

Disclaimer:
This work has been published in the IEEE Conference on Service and Network Communications (CSCN), 2019.
Citation information: DOI 10.1109/CSCN.2019.8931324

Abstract

The on-going digital transformation is key to progress towards a new generation of more efficient, sustainable, and connected industrial systems allowing the so-called factories of the future. This new generation, commonly referred to as industry 4.0, will be accompanied by a new wave of use cases that will allow companies from logistics and manufacturing sectors to increase flexibility, productivity and usability in the industrial processes executed within their factory premises. Unlike typical use cases from other vertical sectors (e.g., energy, media, smart cities), industry 4.0 use cases will bring very stringent requirements in terms of latency, reliability and high-accuracy positioning. The combination of 5G technology with enterprise network solutions becomes crucial to satisfy these requirements in indoor, private environments. In this context, the concept of 5G non-public networks has emerged. In this article we provide an overview of 5G non-public networks, studying their applicability to the industry 4.0 ecosystem. On the basis of the work (being) developed in 3GPP Rel-16 specifications, we identify a number of deployment options relevant for non-public networks and discuss their integration with mobile network operators' public networks. Finally, we provide a comparative analysis of these options, assessing their feasibility according to different criteria, including technical, regulatory, and business aspects. The outcome of this analysis will help industry players interested in using non-public networks to decide which is the most appropriate deployment option for their use cases.

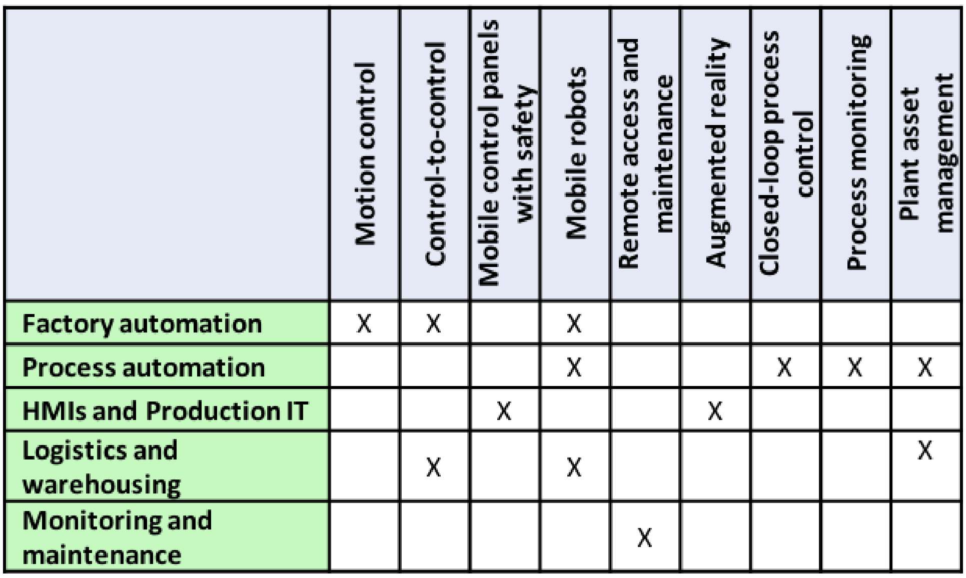
1 Introduction

Connectivity has become a pivotal driver to drive digitalization and product servitization in industrial environments. Industry 4.0 describes the "fourth industrial revolution", which aims at transforming today's factories into intelligently connected production information systems that operate well beyond the physical boundaries of the factory premises. Factories of the future leverage the smart integration of "cyber-physical-systems" and Internet of Things (IoT) solutions in industrial processes [1]. The Fifth Generation (5G) networks can play a key enabling role in this integration, offering programmable technology platforms able to connect a wide variety of devices in a ubiquitous manner [2].

The greatest beneficiaries of a 5G-enabled Industry 4.0 will be the non-telco players, typically operational technology (OT) companies from different vertical sectors such as manufacturing or logistics [3]. OT players, hereinafter referred to as industry verticals, may bring a large number of automation use cases (see Figure), most of them with stringent requirements in terms of availability, reliability, low latency, safety, integrity, and positioning with high accuracy. To meet these requirements in a cost-effective manner, industry verticals may leverage the capabilities provided by 5G technology [4].

The 3rd Generation Public Partnership (3GPP) leads the standardization activities in 5G. With the definition of the 5G system architecture in [6]. 3GPP provides a reference framework for the deployment and operation of upcoming 5G networks, ensuring global inter-operability and their compliance with IMT-2020 Key Performance Indicators (KPIs). Although the first generation of networks based on the 5G system architecture (3GPP Rel-15) were mainly conceived for public use, the possibility of having 5G networks also deployed for private use has recently raised a lot of interest in the industry community. As a result, their study has recently been included as part of the specifications related to the second phase of 5G networks

(3GPP Rel-16 and beyond). This has led to a new classification, whereby 3GPP states that, according to their intended use, networks can be classified into two big categories: Public Land Mobile Networks (PLMNs) and Non-Public Networks (NPNs). On one hand, a PLMN provides network services for public use within a given region, which typically scopes national coverage. A PLMN is operated by a Mobile Network Operator (MNO), who takes the role of PLMN operator. On the other hand, a NPN is intended for the sole use of a private organization, typically an industry vertical. The NPN provides coverage and private network services to devices that are within the vertical's defined premises (e.g. factory, campus). Examples of these devices include sensors, robots, auto-guided vehicles, and remote worker's AR-enabled tablets. From here on out, we refer to these devices as NPN devices.



	Motion control	Control-to-control	Mobile control panels with safety	Mobile robots	Remote access and maintenance	Augmented reality	Closed-loop process control	Process monitoring	Plant asset management
Factory automation	X	X		X					
Process automation				X			X	X	X
HMIs and Production IT			X			X			
Logistics and warehousing		X		X					X
Monitoring and maintenance					X				

Figure F-1. Application areas and use cases in the industry 4.0 ecosystem. Source: 3GPP TS 22.104 [5].

In the industry 4.0 ecosystem, the use of a NPN allows a vertical to have an end-to-end, in-premise 5G network, so that the private traffic can be confined within the boundaries of the defined premises, without the necessity to reach public domain. This is desirable for several reasons, including:

- Quality-Of-Service (QoS) requirements of mission-critical use cases, some of them demanding close-to-zero-ms latency and six nines reliability. The only way to satisfy these challenging requirements is to have dedicated 5G network within the factory, with 5G network functions and service applications as close as possible to the devices and making use of enhanced 3GPP reliability mechanisms, in some cases supported by technologies like Time Sensitive Networking (TSN) and DetNet [7].
- Very high security requirements, met by having strong security credentials and specific authorization mechanisms.
- Isolation from the public domain. This enables protecting the NPN against security attacks or malfunctions (e.g., service outage) in the PLMN.
- Independent network operation for the vertical, allowing him to manage the authentication and authorization of NPN devices, and keep track of their subscription data for accounting and auditing purposes.

However, despite the benefits mentioned above, making NPNs entirely independent of public networks is not always the best solution, either because of business reasons (verticals need to make an initial huge investment, followed by high operational expenditures) or technical reasons (when there is a need to provide NPN devices with connectivity when they are out of NPN coverage). For these cases, integration of the NPN with the PLMN is desirable, so that the MNO can provide device connectivity in out-of-coverage scenarios and reduce entry barriers to verticals. The integration brings open issues that have not been addressed yet in current 3GPP documentation. In view of this, 3GPP SA2 has proposed for Rel-17 a new study item called “Study on enhanced support of Non-Public Networks", which precisely aims to identify these issues and elaborate technical solutions to address them. At the time of writing, this work item has not started yet, although it is planned to begin in the second half of 2019.

In this article, we discuss the use of 5G-enabled NPNs as means to support industry 4.0 ecosystem. For this end, we will first provide a state-of-the-art of NPN in 3GPP specifications, identifying the work done so far. On the basis of this work, we will identify a number of network implementation options for NPNs that could be relevant for industry 4.0 ecosystem, ranging from NPNs completely separated from a PLMN, to NPNs that are entirely hosted by PLMNs, with some scenarios between these extremes. The selection of one or other option is up to the vertical, who can take this decision based on different criteria that include (i) service requirements of considered use cases, and (ii) business-related issues. To help vertical with this decision, we will provide a comparative analysis of the different options, discussing their pros and cons by means of different criteria, including QoS customization, autonomy, isolation, security, service continuity, NPN management and entry barriers for verticals.

The structure of this article is as follows. First, we will provide a overview of the 5G system architecture. Then, we will present the NPN concept in 3GPP ecosystem. Later, we will identify relevant deployment scenarios for NPN, and analyze them based on different criteria. Finally, we will provide some concluding remarks.

2 Overview of the 5G System Architecture

The 4G mobile network architecture was designed to meet requirements for conventional mobile broadband services. This architecture, consisting of a large number of coarse-grained network elements connected with point-to-point interfaces, is rather static and too complex to meet the flexibility, elasticity and scalability that are required to efficiently support the wide variety of vertical use cases that may arise in upcoming years. To meet the diversified requirements of these use cases with minimal complexity and costs, 3GPP has defined a completely new system architecture, shown in Figure . In this section, we provide a high-level description of the 5G system architecture. For more details, please see [6]-[8]

The key principles that explain the evolution from the 4G to the new 5G system architecture are the following:

- A converged core network, to support multiple access technologies. The new 5G Core (5GC) supports New Radio (NR), Evolved UTRAN (E-UTRAN) and non-3GPP access (e.g., Wi-Fi, Fixed). NR is the 3GPP air interface technology used in the new 5G radio access network (NG-RAN), consisting of one or more RAN nodes called next-generation NodeB's (gNBs).
- Control User Plane separation (CUPS). Following Software-Defined Networking (SDN) principles, control and user plane functions are separated for completely independent capacity scaling, decoupled technical evolution, and maximum topology flexibility.
- A unified User Plane Function (UPF), with modular forwarding and processing capabilities that can be flexibly programmed by the control plane.

- Compute and storage separation, allowing any network function to store data (e.g., UE and session context) in a centralized database (unstructured data storage function, UDSF), so that data can be shared across multiple instances of this network function. This supports multiple features such as scaling and 1:N resiliency models, making the 5G system more cloud native.
- Modularization of the architecture design, introducing a set of finer granularity network functions with looser implementation restrictions.
- Service-Based Architecture (SBA), whereby all control plane network functions are connected to a message bus, exposing their functionality to the rest of network functions over service-based interfaces. To allow every network function to discover the services offered by other network functions, the network function repository function (NRF) is defined.

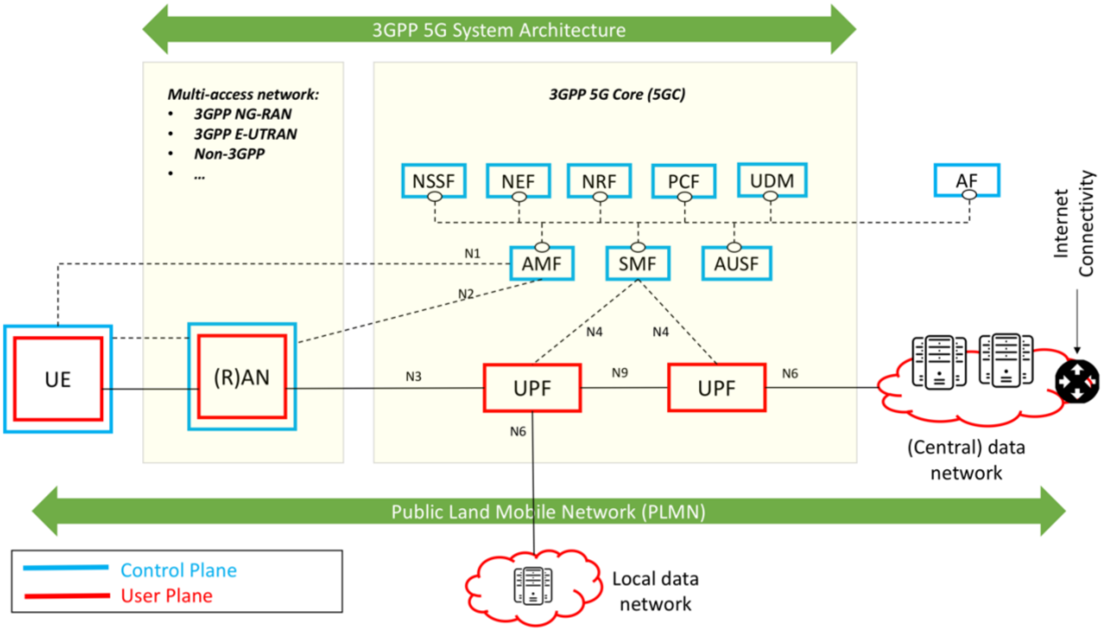


Figure F-2. 3GPP 5G System Architecture

Figure shows the 3GPP 5G system architecture in the context of the MNO's PLMN. As can be seen from the figure, 3GPP 5G system only includes Radio Access Network (RAN) and Core Network (CN) domains, but nothing beyond that. This means that data networks connected to the UPF via the N6 reference point are viewed by 3GPP as external network domains. Nonetheless, the role of these data networks is key to ensure effective support of 5G services in an end-to-end manner. In this paper, two types of data networks are considered:

- *Regional data network*. This data network is owned by the MNO, and thus formally belongs to the PLMN. It allows the MNO to provide UEs with (i) internet connectivity, and (ii) value-added network services, including IP Multimedia Subsystem (IMS) services and non-3GPP L4-L7 services (e.g., firewalling). To host these services, the regional data network consists of one or more high-volume servers with virtualization capabilities.
- *Local area data network*. Unlike the regional data network, a local area data network does not provide internet connectivity, and does not have high compute capacity; indeed, it consists of one or more edge nodes where paradigms like Multi-Access Edge Cloud (MEC) [9] can be applied. These nodes allow hosting delay-sensitive applications (e.g., for closed-loop robot motion control), so that they can be executed

as close as possible to the UE. The local data network can be owned by the MNO, or by under the administrative domain of an industry vertical. In the latter case, this data network can belong to an NPN.

In the following sections, we provide an overview of NPNs in 5G scenarios.

3 5G-enabled NPNs

The standardization work on the use of NPNs in 5G systems is still in its infancy. This is in part due to the lack of participation and influence of the OT players into the work progress of the relevant standards development organizations (e.g., 3GPP, ETSI, IETF and ITU). This has resulted in a misalignment between the service requirements in the industrial domain and the technical solutions delivered by the different standardization bodies. A first step to solve this has already been taken in 3GPP, with the definition of two Rel-16 study items: "Communication for Automation in Vertical domains" (3GPP TR 22.804) [10] and "LAN Support in 5G" (3GPP TR 22.821) [11]. In these study items, use cases from different vertical industries have been analyzed, with a special focus on those requiring the use of NPNs. Based on the requirements derived from this analysis, 3GPP has proposed an initial classification for NPNs, whereby NPNs can be divided into two main categories:

- *Standalone NPNs*, i.e., NPNs that do not rely on network functions provided by a MNO. A stand-alone NPN is an isolated private network that does not interact with a PLMN; indeed, the NPN and PLMN are deployed on separate network infrastructures.
- *Public Network Integrated NPNs,* i.e., NPNs deployed with the support of a PLMN. Unlike a stand-alone NPN, a public network integrated NPN is hosted (completely or in part) on PLMN infrastructure, relying on some MNO's network functions.

Despite having defined these two NPN categories, 3GPP documents do not provide further elaboration on them. This is the gap we cover in the following subsections, where these categories will be analyzed in detail, identifying some variants that could be found within them. Figure F-3 and Figure F-4 illustrate these two categories. For the sake of simplicity, we consider that the vertical's defined premises is a factory.

3.1 Stand-alone NPN

A stand-alone NPN is a private network based on the 3GPP 5G system architecture and completely separated from any PLMN. The independence between this NPN and a PLMN is manifested in the following terms: (i) the use of a unique identifier for the NPN, i.e., NPN ID, entirely independent of the PLMN ID; (ii) the assignment of private spectrum to the NPN; and (iii) the full deployment of a 5G system (including RAN and CN) within the logical perimeter of the factory. The fact that the NPN's CN is independent of the PLMN's CN means that subscription data, signaling traffic and user plane flows from NPN devices remain within the boundaries of the factory, and do not cross PLMN. For this reason, NPN devices are by definition non-public network subscribers.

In order to meet the stringent latency and reliability values required by some use cases, a licensed spectrum is highly preferred for the NPN. This licensed spectrum can be directly obtained from the regulator, or sub-leased from the MNO.

There are some situations where the NPN devices need to access public network services such as voice or internet, while within NPN coverage. In such scenarios, the establishment of a communication path between the NPN and the PLMN is required. As shown in Figure

F-3 a firewall can be used for this end. This firewall allows connecting the NPN data network with the PLMN data network. On the one hand, the NPN data network is within the factory, and usually consists of an edge node with MEC capabilities to run vertical-specific service applications. On the other hand, the PLMN data network consists of one or more regional cloud data centers hosting network services provided by the MNO. Note that the NPN and PLMN data networks illustrated in Figure F-3 corresponds to the local area and regional data networks from Figure . Also note that in this scenario, the firewall is a clearly and identifiable demarcation point that allows separation of responsibilities between the NPN operator (i.e., the vertical) and the PLMN operator (i.e., the MNO).

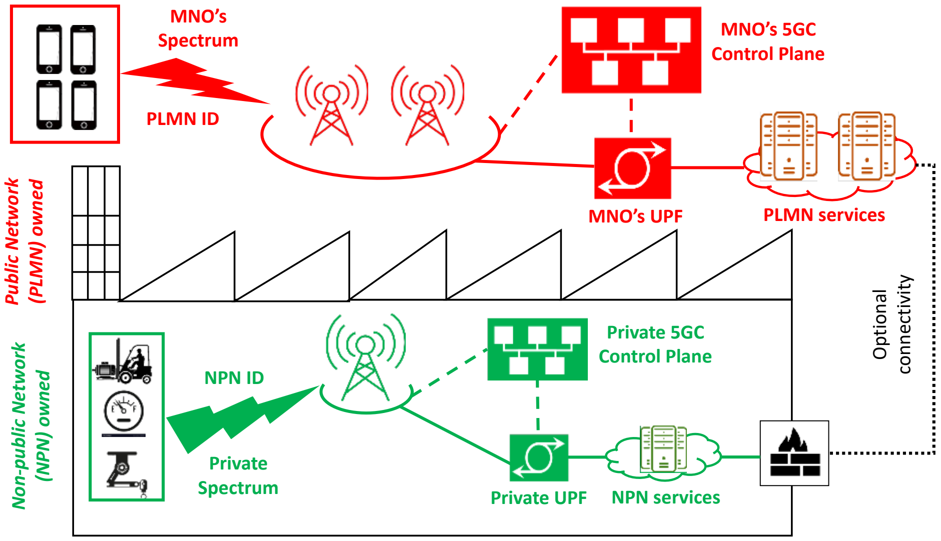


Figure F-3. Stand-alone NPN

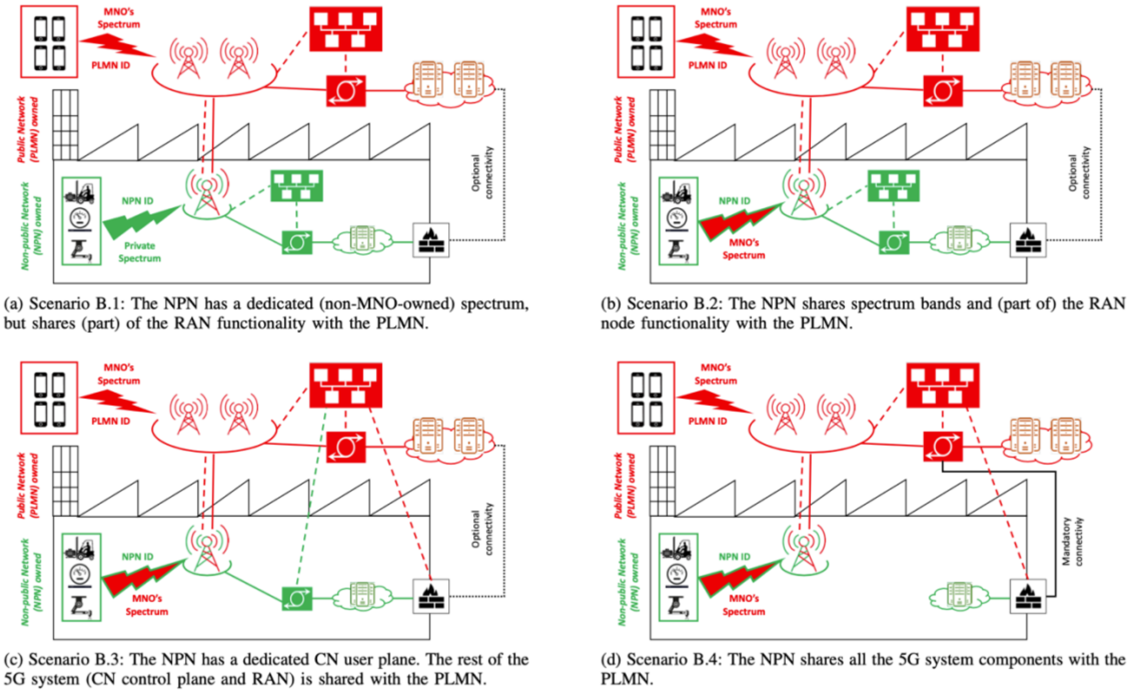
3.2 Public Network Integrated NPN

A public network integrated NPN is a private network based on the 3GPP 5G system architecture and deployed in conjunction with a PLMN. This category assumes the NPN consists of one public sub-network and one or more private sub-networks. On one hand, the public sub-network contains PLMN provided network functions. These functions are under the MNO's administrative domain, and usually deployed out of the factory. On the other hand, a private sub-network includes network functions that remain segregated from the PLMN, and that are allocated inside the factory. The deployment of public and private sub-networks in a public network integrated NPN can vary depending on the considered use case. In this paper, four deployment scenarios have been identified in this respect:

- Shared RAN, with MORAN [12] based approach (scenario B.1, Figure F-4.a): the NPN and PLMN have different IDs, segregated spectrum bands, and independent CNs. As seen, this scenario is quite similar to a stand-alone NPN, with NPN devices being non-public network subscribers. The novelty that this scenario brings is that the RAN segment of the NPN is partially shared with the PLMN. This means that some functions of the RAN nodes serving NPN devices within the factory can be provided by the PLMN. These functions are shared between the NPN and the PLMN, and thus define the public sub-network of the NPN. The rest of RAN functions remain segregated, and thus taken part in a private sub-network of the NPN.
- Shared RAN, with MOCN [13] based approach (scenario B.2, Figure F-4.b): this scenario is similar to B.1, with the exception that the NPN and PLMN also share the

spectrum. As it can be seen from the figure, this spectrum is public and owned by the MNO.

- Shared RAN and shared CN control plane (scenario B.3, Figure F-4.c): in this scenario, the only part of the NPN that remains entirely separate from the PLMN is the CN user plane. The CN control plane is provided by the PLMN, which means the (i) network control tasks in the NPN are performed in the MNO's administrative domain, and (ii) NPN devices are by definition public network subscribers. In this scenario, segregation of non-public and public traffic portions can be achieved by means of 3GPP-defined mechanisms, including network slicing.
- Shared RAN and CN (scenario B.4, see Figure F-4.d): the NPN is entirely hosted by the PLMN. This means that both public and non-public traffic portions are external to the factory, with all data flows routed towards the PLMN via the shared RAN node. However, to guarantee the separation and independence of both portions, these need to be treated as part of completely different networks. To enforce the needed segregation, slicing can also be used.



(a) Scenario B.1: The NPN has a dedicated (non-MNO-owned) spectrum, but shares (part) of the RAN functionality with the PLMN.

(b) Scenario B.2: The NPN shares spectrum bands and (part of) the RAN node functionality with the PLMN.

(c) Scenario B.3: The NPN has a dedicated CN user plane. The rest of the 5G system (CN control plane and RAN) is shared with the PLMN.

(d) Scenario B.4: The NPN shares all the 5G system components with the PLMN.

Figure F-4. Scenarios for the public network integrated NPN category

As in the case with the stand-alone NPN, a firewall installed in the outer edge of the factory allows connectivity between public and private domains. The presence of this firewall is optional for the scenarios B.1, B.2 and B.3, and it is only required when NPN devices want to consume public network services. It is however mandatory for the scenario B.4, since the firewall is the only way to allow these devices to access NPN services through the PLMN. For this end, the firewall connects PLMN provided UPF with the NPN data network.

4 Analysis of NPN Attributes

This section focuses on the attributes that are relevant for the 3GPP-defined NPNs, analyzing their implications for the different scenarios discussed in Section 3. The degree of compliance with these attributes should be considered by an industry vertical when assessing the suitability of an NPN deployment scenario for any planned industry 4.0 use case.

4.1 QoS Customization

It describes the ability to flexibly configure the parameters governing the behavior of a NPN, in such a way that the NPN can satisfy the specific requirements of targeted use cases. These requirements include coverage profiles (e.g., indoor), traffic patterns (e.g., uplink/downlink frame structures) and KPI values (e.g., throughput, latency and jitter, reliability values) that are dependent of the use case under consideration, and quite different from those typically considered in PLMNs. The more independent of a PLMN a NPN is, the more flexibility in NPN parameters setting is allowed, which naturally leads to a use case-tailored NPN configuration.

4.2 Autonomy

It is the ability to guarantee the normal operation of the NPN, regardless of any unexpected event (e.g., security failure, performance degradation) occurred in the PLMN.

4.3 Isolation

Isolation in NPN scenarios is the ability to make non-public traffic portion independent of any other traffic portion flowing in the PLMN infrastructure. This independence shall be assessed (i) in an end-to-end manner, from the device to the data network; and (ii) across the different networking planes, including user, control and management planes. In many deployment scenarios such as those considered in the public integrated NPN category, the NPN and the PLMN share (part of) the same infrastructure resources. It is therefore necessary to consider possible forms of isolation for those scenarios, according to their specificities. Despite their differences, all these forms of isolation need to converge into the following two principles:

- 3GPP network functions from the NPN and the PLMN shall be deployed separate from each other. This separation can be enforced not only at the physical level, but also at the logical level. The latter is particularly relevant for scenarios B.3 and B.4, with high levels of sharing between the two networks. In these scenarios, isolation can be guaranteed through the application of NFV paradigm (i.e., deploying 3GPP functions as virtual network functions) and the corresponding protection mechanisms. This protection shall be mostly focused on how resource sharing is applied, avoiding situations of resource starvation when a certain function sharing a virtualized infrastructure of any nature gets overloaded, depriving other functions of needed resources.
- Data of public and non-public network subscribers needs to be segregated and processed separately, in order to safeguard necessary privacy of the vertical and the MNO. To achieve this, it should be sought, to the extent possible, to avoid transmitting and storing private data outside the boundaries of the vertical's defined premises.

4.4 Security

Guaranteeing security in industrial scenarios requires that NPN communications provide full confidentiality and integrity, in particular when traversing PLMN paths shared with other traffic flows, which will be most, if not all, in practically any feasible scenario. This requires:

- The use of well-known network security techniques, to ensure the required confidentiality and integrity, poses the challenge of how cryptographic material is distributed in an acceptable way to the different deployed 3GPP network functions. Such an acceptable way implies it is trustworthy, so no element impersonation can happen, and verifiable, so identities can be securely verified by the communicating parties. For more details, see [14].
- Segregating the control plane and the management plane functions for the NPN and PLMN, to ensure that the vertical is only able to access network functions specific to the NPN (e.g., for configuration, accounting and/or auditing purposes), and unable to access other similar network functions specific to the MNO's PLMN.

4.5 Service Continuity

It is the ability to provide zero-time service interruption when the NPN devices move between the NPN and the PLMN, and vice versa. Service continuity assumes that PLMN is able to provide seamless connectivity to a device when leaving NPN coverage, either due to a temporal outage in the NPN, or simply because the device moves between two NPNs placed in different locations, although serving the same vertical, e.g., two factories administrated by the same vertical. To avoid service interruption in this type of situations, interworking mechanisms scoping signaling (e.g., automatic network selection) and security (e.g., certificates for device authentication and identification, and for access authorization) should be designed. Apart from well-studied roaming procedures, novel mechanisms based on the use of Non-3GPP Interworking Function (N3IWF) [6] are being explored in 3GPP specifications for this end. The N3WIF, deployed at the NPN (and the PLMN), performs a gateway-like functionality that allows handing over sessions from the NPN to the PLMN (and vice versa) when UE moves between both networks. This N3WIF-like gateway solution can be complemented with UE dual radio support mechanisms, as described in [14].

4.6 NPN management for verticals

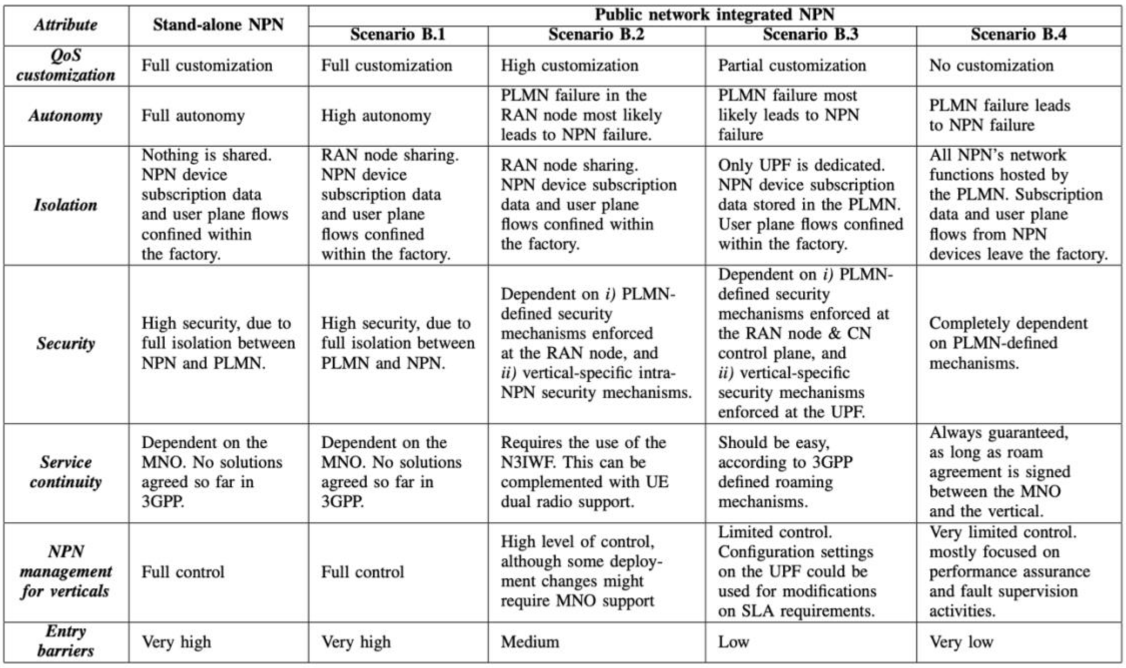
It refers to how much control the vertical can take to freely manage the NPN and its network functions. The more control the vertical company has, the better it can adapt the behavior of the NPN to the specific needs of the served use cases, in terms of performance, functionality and scalability. This control can be exercised through (i) the administration of specific policies; (ii) the execution of performance assurance and fault supervision activities; and (iii) the life cycle management of network functions and service applications, particularly relevant in NFV environments.

4.7 Entry barriers for verticals

A business KPI relevant for any industry vertical is the cost of having an NPN up and running. To estimate the amount of money a vertical shall invest for this purpose, a wide variety of cost sources should be assessed, including (i) spectrum acquisition; (ii) purchase/rental and maintenance of the compute, storage and networking hardware within the factory; (iii) purchase/rental and maintenance of software images executing CN functionality, if virtualized; and (iv) operational expenditure for a management and orchestration solution.

Table F-2 provides a comparative analysis of the different attributes and the implications these have for the different NPN deployment scenarios. As it can be seen, deployments close to stand-alone NPN and scenario B.1 make the NPN entirely independent in terms of performance, management and security; however, they may require significant investment from vertical industry side and might introduce some interworking issues with the PLMN, which hinders service continuity in mobility scenarios. These types of NPN deployment scenarios are ideal to support mission-critical, delay-sensitive industrial use cases demanding full isolation guarantees, and where participating devices are rather static. On the other hand, deployments close to scenarios B.3 and B.4 makes NPN more dependent on PLMN behavior. This facilitates the interaction between the two networks and lower entry barriers for verticals, at the costs of making NPNs less isolated in terms of performance and management. The selection of one or another deployment scenario depends on the cost-benefit ratio from the vertical's viewpoint, considering the requirements of the targeted use case.

Table F-2. Analysis of NPN features for different deployment scenarios



Attribute	**Stand-alone NPN**	**Public network integrated NPN**			
		Scenario B.1	**Scenario B.2**	**Scenario B.3**	**Scenario B.4**
QoS customization	Full customization	Full customization	High customization	Partial customization	No customization
Autonomy	Full autonomy	High autonomy	PLMN failure in the RAN node most likely leads to NPN failure.	PLMN failure most likely leads to NPN failure	PLMN failure leads to NPN failure
Isolation	Nothing is shared. NPN device subscription data and user plane flows confined within the factory.	RAN node sharing. NPN device subscription data and user plane flows confined within the factory.	RAN node sharing. NPN device subscription data and user plane flows confined within the factory.	Only UPF is dedicated. NPN device subscription data stored in the PLMN. User plane flows confined within the factory.	All NPN's network functions hosted by the PLMN. Subscription data and user plane flows from NPN devices leave the factory.
Security	High security, due to full isolation between NPN and PLMN.	High security, due to full isolation between PLMN and NPN.	Dependent on *i)* PLMN-defined security mechanisms enforced at the RAN node, and *ii)* vertical-specific intra-NPN security mechanisms.	Dependent on *i)* PLMN-defined security mechanisms enforced at the RAN node & CN control plane, and *ii)* vertical-specific security mechanisms enforced at the UPF.	Completely dependent on PLMN-defined mechanisms.
Service continuity	Dependent on the MNO. No solutions agreed so far in 3GPP.	Dependent on the MNO. No solutions agreed so far in 3GPP.	Requires the use of the N3IWF. This can be complemented with UE dual radio support.	Should be easy, according to 3GPP defined roaming mechanisms.	Always guaranteed, as long as roam agreement is signed between the MNO and the vertical.
NPN management for verticals	Full control	Full control	High level of control, although some deployment changes might require MNO support	Limited control. Configuration settings on the UPF could be used for modifications on SLA requirements.	Very limited control. mostly focused on performance assurance and fault supervision activities.
Entry barriers	Very high	Very high	Medium	Low	Very low

5 Conclusions and Future Outlook

In this paper we have described a number of deployment options for NPNs in the industry 4.0, based on 3GPP 5G specifications. These range from NPNs completely separated from a PLMN (stand-alone NPNs), to NPNs that are entirely hosted by the PLMN (scenario B.4). We also have provided a comparative analysis of the different options, based on different criteria. The outcome of this analysis may be useful for those verticals interested in NPNs, helping them to decide what is the best deployment option for them, according to their specific service needs and considering the effort they are willing to invest in designing, deploying and operating NPNs.

This paper provides guidelines that can be used as starting point for further progress in NPN standardization in 3GPP 5G systems. Much of the work that needs to be undertaken in the future includes the study on the applicability of network slicing in scenarios B.3 and

B.4, and the study on interfaces to enable interworking and seamless handovers between NPNs and PLMNs. Apart from these technical issues, other aspects should also be considered and explored, including.

- Regulatory aspects, most of them related to the spectrum.
- New business models. Indeed, the integration of NPN and PLMN enables a synergistic relation whereby industry verticals have incentives to invest in on-premises 5G evolved infrastructure, which can then be offered "as a service" to one or more MNOs to allow them to expand their service footprint at a reduced cost.

Acknowledgement

This work is supported by the H2020 European Projects 5G-VINNI (grant agreement No. 815279) and BOOST 4.0 (grant agreement No. 780732).

References

[1] M. Wollschlaeger, T. Sauter, and J. Jasperneite, “The Future of Industrial Communication: Automation Networks in the Era of IoT and Industry 4.0”, *IEEE Industrial Electronics Magazine*, pp. 17-27, 2017.

[2] S. K. Rao and R. Prasad, “Impact of 5G Technologies on Industry 4.0”, *Wireless Personal Communications*, vol. 100, no. 1, pp. 1-15, 2018.

[3] Industry 4.0 Market Research Report, “Industry 4.0 Technologies, Industry and Global Markets: 2019-2023”, 2019.

[4] M. A. Imran *et al*., “Enabling 5G Communication Systems to Support Vertical Industries”, 2019.

[5] 3GPP TS 22.104, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Service Requirements for cyber-physical control application in vertical domains (Release 17)," 2019.

[6] 3GPP TS 23.501, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System (Release 16)," 2019.

[7] A. Nasrallah *et al*., “Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 88–145, 2018.

[8] 3GPP TS 23.503, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and charging control for the 5G System (Release 16),” 2019.

[9] ETSI White Paper No.28, “MEC in 5G networks,” 2018.

[10] 3GPP TR 22.804, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on Communication for Automation in Vertical Domains (Release 16)”, Dec. 2018.

[11] 3GPP TR 22.821, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on LAN Support in 5G (Release 16)”, 2018.

[12] 3GPP TS 23.251, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Network Sharing; Architecture and functional

description (Release 15)”, 2018.

[13] 3GPP TR 33.819, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on security for 5GS enhanced support of Vertical and LAN Services (Release 16)”, 2018.

[14] 3GPP TR 23.734 “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on enhancement of 5G system (5GS) for Vertical and LAN Services (Release 16)”, 2018.

Paper G

5G Non-Public Networks: Standardization, Architectures and Challenges

Jonathan Prados-Garzon, Pablo Ameigeiras, Jose Ordonez-Lucena, Pablo Muñoz, Oscar Adamuz-Hinojosa and Daniel Camps-Mur

Disclaimer:
This work has been published in the IEEE Access.
Citation information: DOI 10.3390/ACCESS.2021.3127482.

Copyright:

Abstract

Fifth Generation (5G) is here to accelerate the digitization of economies and society and open up innovation opportunities for verticals. A myriad of 5G-enabled use cases has been identified across disparate sectors like tourism, retail industry, and manufacturing. Many of the networks of these use cases are expected to be private networks, that is, networks intended for the exclusive use of an enterprise customer. This article provides a comprehensive overview of the technical aspects for realizing private 5G networks while motivating them with illustrative examples. We first identify the key aspects for private 5G networks. Then, we follow an overview of the latest 3GPP specifications capabilities to support private 5G networks. Next, we address the realization of five worthwhile scenarios that cover single site, multi-site, radio access network (RAN) sharing, and mobility use cases in private 5G networks. Finally, we provide a summary of the key challenges for private 5G networks.

1 Introduction

Fifth Generation (5G) is here to accelerate the digitalization of economies and society. Over the last decade, the combined efforts from academy and industry have materialized in matured 5G standards that will bring services with data rates, latency, reliability, connection density, and security constraints never seen before, thus opening up innovation opportunities for verticals. Ericsson has identified more than 200 industry digitization use cases enabled or substantially enhanced by 5G technology [1]. Typical use cases can be found in disparate sectors such as agriculture, tourism (e.g., museums), transportation, healthcare, education (e.g., convention centers), retail industry (e.g., shopping malls), transport hubs (e.g., ports and airports), sport facilities (e.g., stadiums), energy industry, military bases, and manufacturing. In particular, 5G is acknowledged as a key enabler for Industry 4.0 [2]-[3].

Many of the networks of the abovementioned use cases, including the industrial sector, are private networks. A private 5G network, also termed Non-Public Network (NPN) by Third Generation Partnership Project (3GPP), is a 5G network deployed for non-public use. In contrast to Public Land Mo- bile Networks (PLMNs) that offer mobile network services to public subscribers, NPNs are intended for the exclusive use of an enterprise customer, such as an industry vertical or a state-owned company.

There are two basic options to deploy a 5G NPN: i) Stand- Alone NPN (SNPN), which does not rely on PLMN-provided network functions, and ii) Public Network Integrated NPN (PNI-NPN), whose deployment is supported by a PLMN. Whereas SNPNs enable the enterprise customer to retain full control of the NPN, PNI-NPNs represent a reduced entry barrier due to Capital Expenditure (CAPEX) and Operational Expenditure (OPEX) reduction. There are several works in the literature addressing 5G NPNs [2]-[5].

The 5G Alliance of Connected Industries and Automation (5G-ACIA) identified in [2] four deployment options for private 5G networks, namely, deployment as isolated network, deployment with shared Radio Access Network (RAN), deployment with shared RAN and control plane, and NPN hosted by the public network. The first two deployment options correspond to SNPNs, whereas the latter two are PNI-NPNs. These deployment options are intended to cover the necessities of different industrial scenarios. In [2], the 5G-ACIA provides a preliminary analysis of advantages and dis- advantages in terms of cost, control and security of these four deployment options. The authors in [4]-[5] provide a deeper analysis of the pros and cons of these deployment options. Additionally, in [3] the author motivates the suitability of 5G NPNs to unleash the potential of industry digitization. Also, the author in [3] revisits the primary use cases in industry 4.0 that can benefit from 5G NPNs, the deployment options proposed by 5G-ACIA, some related features included in 3GPP standards, and identifies some of the challenges for realizing 5G NPNs.

However, none of the above works provide a description on how to technically realize the proposed scenarios. In this work, we cover this gap by detailing the architectures to realize SNPNs and PNI-NPNs, which is essential to understand their technical feasibility and implications. Moreover, we provide simulation-based performance results for three PNI-NPN configurations in a campus network. Furthermore, we provide the description of the architecture for NPNs leveraging network sharing. Network sharing is also a key trend in 5G, as it enables notable cost reduction and may be a key lever to reduce the entry barrier for some enterprise customers interested on deploying 5G NPNs. In addition, it also fits the necessities of many private venues that cannot accommodate the deployment of several infrastructure net- works due to physical space limitations or aesthetics.

Besides single-site NPNs, there are other scenarios not covered in the literature that deserve attention. On the one hand, NPNs might spread across multiple sites, e.g., several enterprise branches. The support of a public network is needed to provide connectivity among the remote locations while ensuring the required performance and security levels. This article is the first addressing the multi-site 5G NPN scenarios by discussing their issues and identifying deployment alternatives. On the other hand, various private use cases involve devices that need to move out of the private venue without service interruption. In this vein, this article advances the state-of-the-art in the description of the technical realization of these scenarios when the UE does not support Dual Subscriber Identity Module.

Before providing the NPNs architectures covered in this work, we review the capabilities included in 3GPP specifications related to 5G NPNs (e.g., Local Area Data Network [LADN], Closed Access Group [CAG], Data Network Name [DNN]), Multi-Operator Core Network [MOCN], ...), which have not been described in the literature. Last, we identify additional challenges and research directions for realizing 5G NPNs to those proposed in the literature.

The remainder of the article is organized as follows. In Section 2, we review key aspects for 5G NPNs. Next, we provide an overview of the 3GPP specifications to support private 5G networks. In Sections 4 and 5 we address the realization of the 5G NPN scenarios covered in this work. Finally, in Section 6 we provide a summary of key challenges for 5G NPNs.

2 Key Aspects for 5G NPNs

This section reviews key aspects for 5G NPNs (see Figure G-1).

2.1 Spectrum

One of the key ingredients for the success of 5G NPNs is to make spectrum a handy resource. For this purpose, three regimes are available for spectrum access, namely, Licensed Spectrum (LS), Shared Spectrum (SS), Unlicensed Spectrum (ULS) [3]. Regarding LS, regional regulators are setting the spectrum aside (e.g., 3.7-3.8 GHz) for industrial private networks. Additionally, innovative SS solutions such as the Citizens Broadband Radio Service (CBRS) band in USA open new possibilities for NPN deployments.

2.2 Interworking

The integration of 5G with today's legacy private networks is essential to enable the gradual update of the network, thus lowering entry barriers for vertical customers, and specific use cases in which not all the devices (e.g., industrial controllers) are wirelessly connected. Many of the current private networks are based on isolated Ethernet environments and Wi-Fi deployments. Whereas 5G standards address the integration with Wi-Fi [6], proprietary solutions might be required for wired L2 networks like Industrial Ethernet [7].

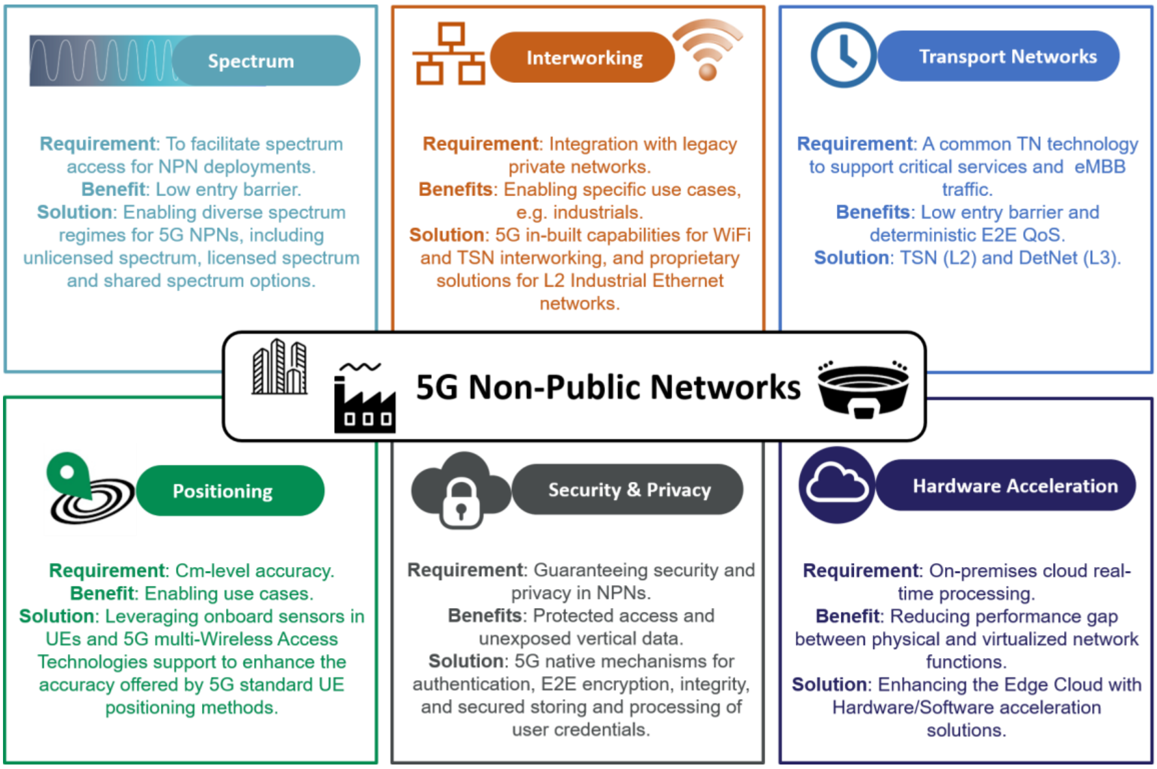

Spectrum
Requirement: To facilitate spectrum access for NPN deployments.
Benefit: Low entry barrier.
Solution: Enabling diverse spectrum regimes for 5G NPNs, including unlicensed spectrum, licensed spectrum and shared spectrum options.
Interworking
Requirement: Integration with legacy private networks.
Benefits: Enabling specific use cases, e.g. industrials.
Solution: 5G in-built capabilities for WiFi and TSN interworking, and proprietary solutions for L2 Industrial Ethernet networks.
Transport Networks
Requirement: A common TN technology to support critical services and eMBB traffic.
Benefits: Low entry barrier and deterministic E2E QoS.
Solution: TSN (L2) and DetNet (L3).
5G Non-Public Networks
Positioning
Requirement: Cm-level accuracy.
Benefit: Enabling use cases.
Solution: Leveraging onboard sensors in UEs and 5G multi-Wireless Access Technologies support to enhance the accuracy offered by 5G standard UE positioning methods.
Security & Privacy
Requirement: Guaranteeing security and privacy in NPNs.
Benefits: Protected access and unexposed vertical data.
Solution: 5G native mechanisms for authentication, E2E encryption, integrity, and secured storing and processing of user credentials.
Hardware Acceleration
Requirement: On-premises cloud real-time processing.
Benefit: Reducing performance gap between physical and virtualized network functions.
Solution: Enhancing the Edge Cloud with Hardware/Software acceleration solutions.

Figure G-1. Key aspects for 5G NPNs

2.3 Transport networks

The support of Ultra-Reliable Low-Latency Communication (URLLC) services requires all the network domains have the ability to provide deterministic Quality of Service (QoS) in terms of delay, jitter, frame loss and reliability, including the Transport Network (TN) [8]. Furthermore, the same TN infrastructure shall allow for the coexistence of 5G services heterogeneity to cheapen the costs. Time-Sensitive Networking (TSN) and Deterministic Networking (DetNet) meet these requisites and are, therefore, appealing solutions for providing layer 2 (L2) and layer 3 (L3) connectivity in NPNs, respectively [8]-[9].

2.4 Positioning

The positioning of User Equipment's (UEs) is essential to enable new use cases like Augmented Reality (AR)-assisted workers, motion control, and Automated Guided Vehicles (AGVs) in factories, which might require localization services with centimeter-level precision. Although standard UE positioning methods in 5G (based on radio signals measurements) do not offer such an accurate localization, 5G multi-Wireless Access Technology support and localization framework to collect measurements from onboard sensors in UE can be leveraged to provide localization solutions meeting the most stringent positioning requisites [10].

2.5 Hardware acceleration

Deploying network functions as Virtual Network Functions (VNFs) on commodity hardware significantly degrades their packet-processing performance compared to purpose-built hardware devices. This is a key aspect for critical services. To reduce the performance gap between traditional NFV and middleboxes, hardware acceleration solutions (e.g., Smart- NICs, PCIe cards, FPGA, GPUs, etc.) can be adopted [10].

2.6 Security and Privacy

From Release 16 on, 3GPP defines advanced security and privacy mechanisms for the support of NPNs [11]. These mechanisms provide solutions related to device-to-network communications, including device authentication (with the possibility of the enterprise customers to implement a second authentication in the local Data Network), end-to-end traffic integrity and encryption (at both user and control planes) and device credentials management. Additionally, other infrastructure-related solutions should be considered. Examples include remote attestation (ETSI NFV-SEC defined transitive mechanism ensuring trust and liability for the VNFs and underlying infrastructure) and proof-of-transit (allows for external verification in the compliance of traffic forwarding policies, ensuring packets traverses processing nodes as mandated) [12].

3 3GPP Related Standardization

In this section we provide an overview of the 3GPP Release 16 capabilities to support NPNs and network sharing.

3.1 3GPP Support for Non-Public Networks

According to 3GPP specifications [6], NPNs are categorized into SNPNs and PNI-NPNs.

3.1.1 Stand-alone NPN

It is a NPN that operates without dependency on a PLMN, i.e., not relying on network functions provided by a PLMN. It requires a 5G System (5GS) separated from the PLMN, and NPN devices must have a subscription to the SNPN in order to access it. An SNPN is uniquely identified by the combination of a PLMN ID and a Network ID (NID). Thus, UE is configured with the tuple {PLMN ID, NID} to access an SNPN. The PLMN ID may be a private network ID (e.g., based on mobile country code [MCC] 999 as assigned by ITU for 3GPP), or the ID of a PLMN that is operating that SNPN. The NID could be self-assigned (i.e., chosen by SNPN at deployment time) or coordinated assigned (universally managed NID) [6].

3.1.2 Public Network Integrated NPN

It is a NPN deployed with the support of a PLMN. NPN devices must have a subscription to the PLMN in order to access the PNI-NPN. According to [6], a PNI-NPN may be provided by a PLMN by means of a dedicated DNN or by deploying network slices allocated for the NPN.

- Provision as DNN. In this the PNI-NPN is provided as a data network, which is used for hosting the NPN services and applications. The DNN identifies the data network, and whenever the subscriber executes the NPN application, the UE triggers the establishment of a PDU session to the NPN DNN. As typically NPNs provide services within a limited coverage area, the 3GPP has standardized the concept of LADN, which enables access to the DNN in a given area (e.g., stadium or museum), but not outside. The LADN service area is defined as one or several Tracking Areas (TAs). A TA is a group of cells where a user can move around without updating the Access and Mobility Management Function (AMF). When the UE is inside the LADN service area, it can request a PDU session establishment for the LADN DNN, and the network will grant such PDU session. The PLMN Operator (PLMN-Op) can use the UE Route Selection Policy (URSP) rules to control the PDU session request from the UE when this is inside (or outside) the LADN service area.

- Provision as a network slice. Network slicing is a technological solution consisting of providing isolated logical networks with diverging performance requirements over a common network infrastructure. A 5G network slice is composed of the 3GPP 5GS network functions (e.g., Next Generation NodeBs (gNBs), AMF, User Plane Function (UPF) UPF, SMF, etc.), it is identified by a Single Network Slice Selection Assistance Information (S-NSSAI), and it consumes a certain amount of radio resources in each cell. A PLMN-Op can use network slicing to provide public network services, or NPN services, i.e., a PNI-NPN. The PLMN-Op can deploy one or several dedicated network slices for the PNI-NPN, if NPN isolation or specific QoS treatment is desired. The customer can consume the received slice directly, or optionally extend it with additional features (e.g., device on-boarding, secondary authentication). Using network slicing for the PNI-NPN allows to control the access to the NPN because the subscriptions to the dedicated S-NSSAIs can be restricted to the NPN devices. In PNI-NPN, the UE needs to be pre-configured with the S-NSSAI to access the slice. The PLMN-Op can also use the URSP rules for this purpose.

A relevant requirement of a NPN is that it can control the access of NPN devices to the network in areas in which they are not permitted to. However, as in the case of LADN service area, network slices are set on a per TA basis [6]. That is, neither LADN nor network slicing allow the possibility to prevent UEs from automatically selecting and accessing specific cells within a TA. Closed Access Groups (CAG) may optionally be used in NPNs for this purpose. A CAG defines a list of subscribers who are allowed to access a CAG cell associated with it. A CAG cell is a cell that only UEs supporting CAG can access. Hence, CAG can be used in PNI-NPNs to prevent unauthorized UEs to access specific CAG cells inside a private venue (e.g., stadium or museum). Please note that CAGs are independent from any network slice.

3.2 Network Sharing

Network sharing is a key technical feature in 5G. 3GPP specifications for 5G provide support only for MOCN sharing architecture[6]. In the MOCN architecture the Next- Generation Radio Access Network (NG-RAN) segment (including RAN infrastructure, functionality, and spectrum carrier) is shared among multiple independent network opera- tors, while the 5G Cores are owned by each of them. The NG-RAN sharing functionality has been extended in Rel-16 to support MOCN scenarios involving NPNs. Specifically, the supported scenarios allow to share the NG-RAN among any combination of PLMNs, SNPNs, and PNI-NPNs (with CAGs).

In MOCN architecture, each cell of the shared NG-RAN must radiate the PLMN IDs and NIDs of the available PLMNs and SNPNs, respectively, through the Broadcast System Information (BSI) for selection by UE. Additionally, the PLMNs and/or SNPNs must be the same for all cells of a TA. The BSI also includes additional parameters per PLMN, such as cell ID, TAs, and CAG IDs. In the current version of 3GPP specifications a cell ID may only be associated with one of the following options: one or several SNPNs, one or several PNI-NPNs (with CAG), or one or several PLMNs [6].

4 Single-site NPN Architectures

This section presents the architectures for single-site NPNs.

4.1 Stand-alone NPN Architecture

The baseline SNPN consists of a private 5GS, comprising a NG-RAN and a lightweight 5G Core (5GC). The NG-RAN includes a set of gNBs providing indoor 5GNR coverage. The 5GC follows a Service Based Architecture (SBA) with control and user plane separation, i.e., it is designed with a 5G Core Control Plane (5GC-CP) decoupled from UPFs that build up the user data path. While the UPFs are always deployed on-premises, the 5GC-CP might be partially executed off-premises. The 5GC-CP can be hosted off-premises by 3rd party cloud providers, typically hyperscalers (e.g., AWS). Please note that some of these cloud providers also offer to bring their infrastructure and services on-premises (e.g., AWS Outposts), which could facilitate the complete SNPN deployment on-premises.

In SNPNs, the enterprise customer or a delegating company may take the role of NPN operator, thereby acting as a µOperator [13]. Alternatively, the enterprise customer may ask a PLMN-Op to take the NPN operator role.

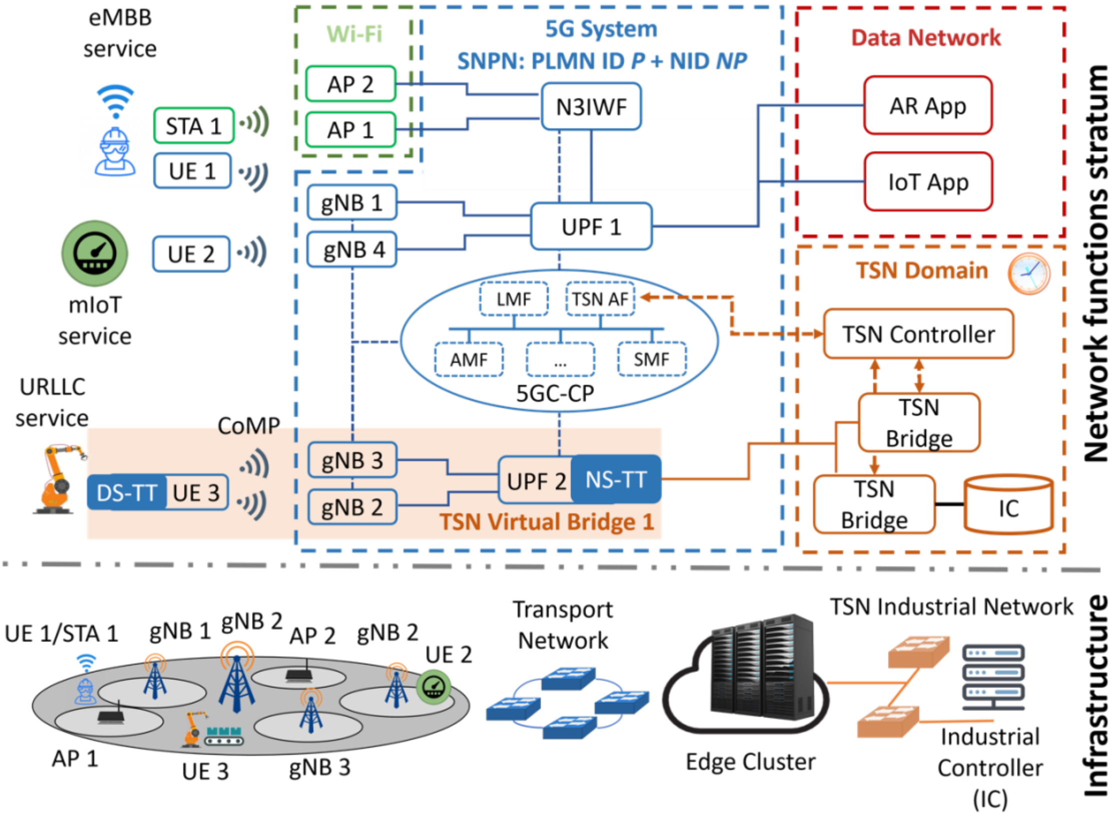


Figure G-2. SNPN architecture

One of the main use cases for an SNPN is a smart factory with industry 4.0 services that leverages 5G wireless connectivity capabilities. Figure G-2 captures an archetypal architecture of this SNPN. To better clarify the decoupling between functionality and infrastructure resources, the figure has been split into two separate strata: the infrastructure stratum and network function stratum (lower and upper figure side, respectively).

On the one hand, the infrastructure stratum represents the on-premises physical network substrate that hosts the SNPN. It comprises a set of wireless access nodes and a clustered NFV Infrastructure (NFVI), with a transport network providing TSN connectivity along the entire data path. The wireless access nodes include gNBs providing small cell 5GNR connectivity and Wi-Fi access points. Optionally, gNBs functional split could be considered if required. To that end, NFVI could be enhanced with hardware/software acceleration solutions for real-time processing of the virtualized gNB functions.

On the other hand, the network function stratum represents the different functional components building up the SNPN. Note that the SNPN includes four different network segments: 5GS (i.e., NG-RAN, UPF, 5GC-CP), Wi-Fi, TSN and the local data network. In the 5GS, UPFs and 5GC-CP are executed as VNFs on the edge cluster, while NG-RAN consists of gNBs deployed as physical network functions. The Wi-Fi segment, with technology features provided by underlay Wi-Fi access points, complements the 5GNR connectivity capabilities provided by gNBs. This segment allows increasing the reliability and throughput at access side leveraging on multi-access connectivity features (e.g., traffic offloading, bandwidth aggregation). The TSN segment allows providing deterministic QoS to the SNPN, which is key for typical URLLC-type industry 4.0 services where a wireless station (e.g., industrial robot) is operated by an industrial controller (IC) connected to the TSN industrial network. For these services, the 5GS behaves as a set of TSN bridges (one per UPF). The integration of 5GS and TSN requires the use of TSN translation modules (e.g., Device-Side TSN Translator (DS-TT) and Network-Side TSN Translator (NS-TT)) at the 5G entities interfacing with the TSN network, i.e., UE and UPF. The TSN controller transparently configures the 5GS as if it is a TSN bridge through the TSN AF. For more information of 5G-TSN interoperability, refer to [6] and [14]. Finally, the local data network allows hosting the applications (e.g., IoT app, AR app) providing the service logic.

Although not captured in the figure, it is worth noting that network slicing can be used in SNPN to differentiate traffic from different industry 4.0 services.

4.2 PNI-NPN Architecture

The PNI-NPNs represent a reduced OPEX/CAPEX deployment option compared to SNPNs as they may leverage the PLMN-Op’s infrastructure, spectrum and know-how. As described in Section 3.1.2, the PLMN-Op may provide the PNI-NPN by means of a DNN or a dedicated network slice.

The implementation of the PNI-NPN presents several issues:

- The on-premises 5GNR connectivity: the gNBs deployed in-house can be owned by the enterprise customer (e.g., purchased directly to the network equipment provider) or made available by the PLMN-Op.
- The ability to dedicate and customize the PNI-NPN: the PLMN-Op can configure the PNI-NPN in terms of functionality and capacity according to the enterprise customer’s needs, by provisioning network and application functions specifically dedicated and adjusted to the NPN requirements. For example, the PLMN-Op may deploy a customer-tailored, lightweight 5GC that includes only the network functions (UDM, AMF, SMF, NRF, UPF) and with the specific capacity as required by the

private services.

- The location of the PNI-NPN functions: some NPN scenarios require the network functions to be executed on the customer premises, either for performance (e.g., on-premises UPF, for low-latency support) or for privacy reasons (e.g., on-premises UDM, to keep subscription data locally stored.
- The UE access control: the PLMN-Op can enforce the access control by means of the CAG, LADN and network slicing mechanism as described in 3

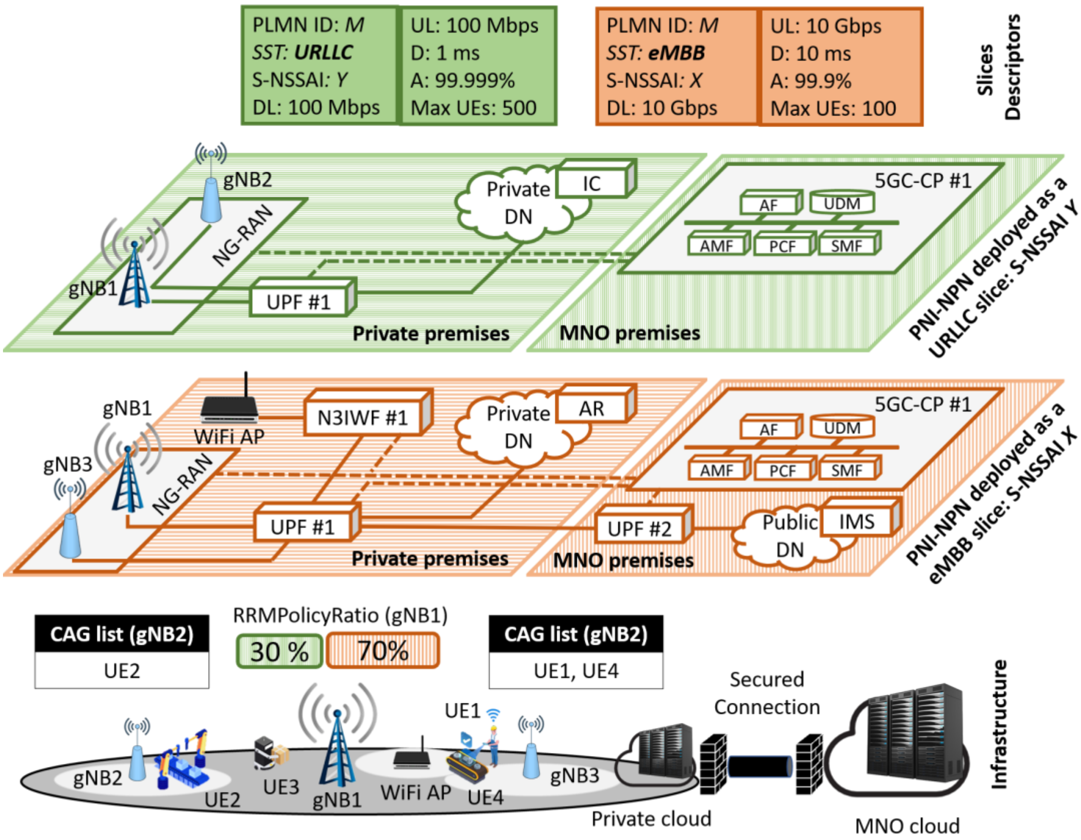
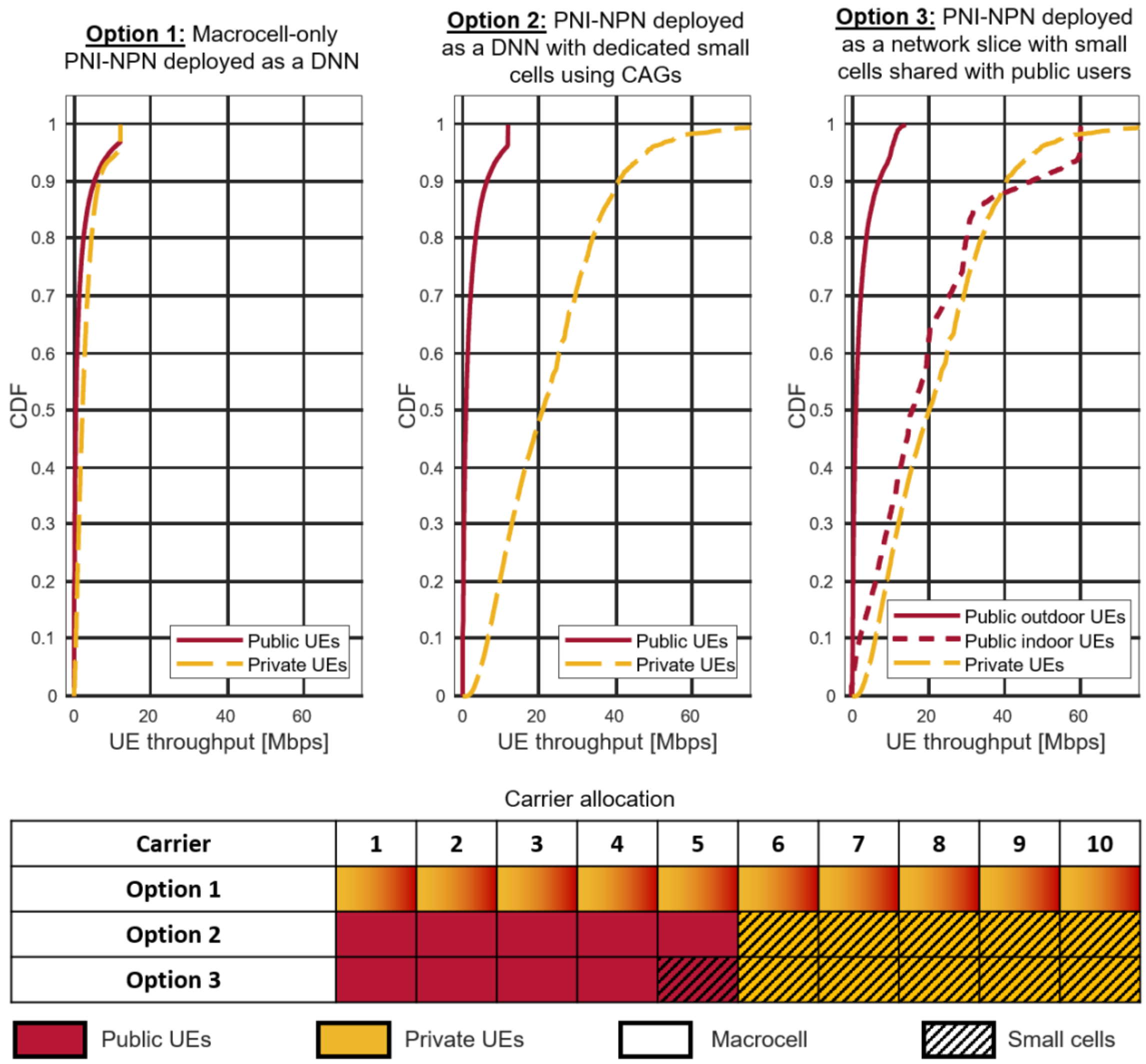


Figure G-3. PNI-NPN architecture

Figure G-3 captures an archetypal architecture for PNI-NPN scenarios. The figure illustrates two coexisting PNI-NPNs, both provisioned by the PLMN-Op as separate network slices. The gNBs broadcast the PLMN ID for individual PNI-NPNs. One of the slices, whose Slice/Service Type (SST) is URLLC, is destined to industrial critical applications. The second network slice provides access to enhanced Mobile Broadband (eMBB) services to the workers of the industry. This eMBB slice integrates Wi-Fi access through the N3IWF, which is also located on-premises. The PLMN- Op instantiates a UPF on-premises in the edge cluster and dedicates it to the URLLC slice. In this way, the critical traffic is kept in-house and its latency constraints can be met. On the other hand, the UPF for the eMBB slice and the 5GC-CP, which is shared by both slices, are hosted in the PLMN-Op's edge cloud.

For a DNN based implementation of a PNI-NPN please refer to [15].

Figure G-4 shows a comparison of the UE throughput in an industry campus for three deployment options. The setup considers 25 private users located inside three factory plants and 25 public users located inside and outside the factory plants. The deployment options are: 1) all users are served by a macrocell, and a PNI-NPN is deployed as a DNN for the private users, 2) public users are served by the macrocell whereas private users are served by small cells with CAGs located in the factory plants and the PNI-NPN is again deployed as a DNN, and 3) public outdoor users are served by the macrocell whereas indoor public and private users are served by the small cells and the PNI-NPN is deployed as a network slice. The system bandwidth is 100 MHz. It is split into ten carriers of 10 MHz each. Figure G-4 includes the carrier allocation among public and private users.



Carrier	1	2	3	4	5	6	7	8	9	10
Option 1	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥	🟨🟥
Option 2	🟥	🟥	🟥	🟥	🟥	🟨▨	🟨▨	🟨▨	🟨▨	🟨▨
Option 3	🟥	🟥	🟥	🟥	🟥▨	🟨▨	🟨▨	🟨▨	🟨▨	🟨▨

🟥 Public UEs 🟨 Private UEs □ Macrocell ▨ Small cells

Figure G-4. Throughput achieved by a PNI-NPN in an industry campus network for three deployment options.

As observed, for deployment option 1) the UE throughput is similar for both public and private users. For option 2) the throughput of private UEs significantly increases as they are served by the small cells with CAGs. For option 3), network slicing enables allocating one carrier for public use in the small cells, thus improving the throughput of the indoor public UEs.

4.3 On-premises RAN Sharing Scenario and Architecture

In a MOCN architecture for a private venue scenario, a Network Operator (NOP) deploys and operates an indoor small cells infrastructure, and opens this infrastructure and the spectrum to other NOPs for the provision of communication services. We will refer to the first NOP as the Master NOP (MNOP) and the remaining ones as Participating NOPs (PNOPs). Each PNOP, and possibly the MNOP, employs its own 5GC to deploy SNPNs or PNI-NPNs (with CAGs). Additionally, a PLMN-Op may also participate in the sharing with its own 5GC to merely extend the footprint of its public services inside the private venue. For the MOCN scenario, we identify the following possibilities:

- The MNOP is a PLMN-Op. In this case, the MNOP has primary access to a particular licensed spectrum which shares together with its NG-RAN infrastructure with the PNOPs.
- The MNOP is a µOperator. In this case, the venue owner or delegating company takes the role of MNOP and leases the NG-RAN to the PNOPs. The main difference with the previous case is that the µOperator does not have primary access to a particular licensed spectrum, and instead it requires a locally issued spectrum license. As mentioned in Section II, the µOperator has several alternatives to access spectrum in this situation.

The MOCN architecture is well suited for private venue scenarios as it enables multi-tenancy in the 5GS network, which makes it possible for various NOPs to provide communication services while sharing the NG-RAN. Some exemplary use case scenarios are a smart stadium, a shopping mall or a hospital, in which several NPNs could be deployed to provide various private localized services.

Figure G-5 depicts an architecture blueprint of such a MOCN deployment. The PNOPs act as tenants and interact with the MNOP to negotiate SLAs and request NG-RAN resources on demand. Under such requests, the MNOP has to allocate portions of network capacity to the PNOPs for a particular time period. Therefore, the NG-RAN network resources are to be sliced and delivered to each PNOP. Hereafter, we will refer to these resources as an infrastructure slice. This infrastructure slice is composed of all the set of wireless, virtualized compute and networking resources of the NG-RAN infrastructure, which are segregated and provided to a PNOP. It is worth noting that armed with a 5GC the PNOPs may advertise multiple 3GPP network slices, i.e., S-NSSAI, within their infrastructure slice, with each 3GPP slice having specific requirements in terms of network resources. Additionally, the NG-RAN architecture also has to expose the corresponding interfaces to PNOPs for networks resource requests, service monitoring and network management capabilities.

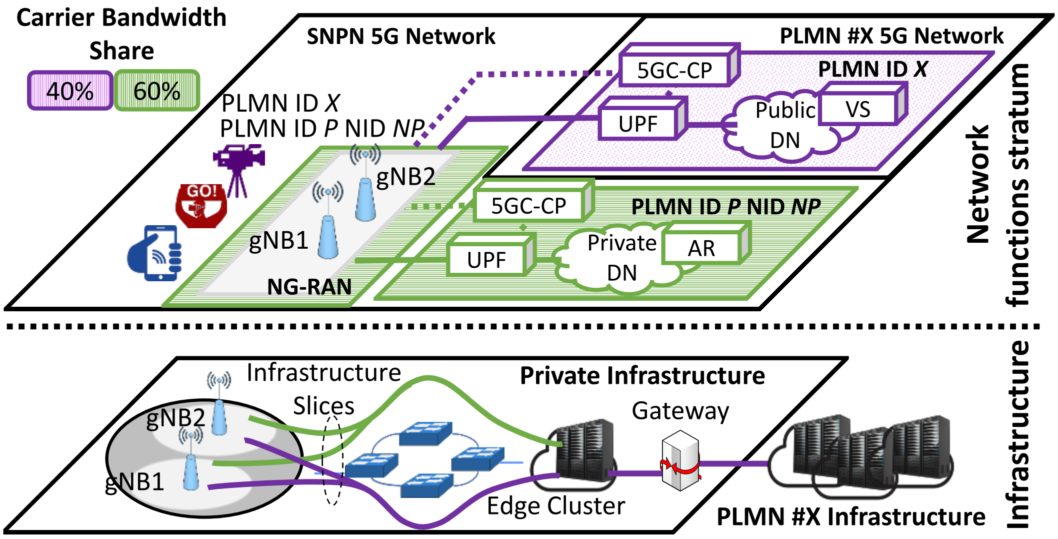


Figure G-5. On-premises NG-RAN sharing through MOCN architecture.

5 Mobility and Multi-site NPN Scenarios

This section describes, motivates and explores technical alternatives for mobility between NPNs and multi-site NPNs.

5.1 Mobility in NPNs

Several promising private applications require the devices move out of the private premises, such as an unmanned aerial vehicle fleet that needs to monitor the crop growth in agriculture or delivers a package in logistics, or even moves between private sites (e.g., factories). For example, real-time tracking of goods when they are moved between manufacturing, distribution, and retail centers, or even later incorporating those goods into the local factory inventory management system in an automated manner. These scenarios entail a PLMN that supplies wireless access out of premises and mechanisms to warrant Session and Service Continuity (SSC), i.e., to provide UEs with a seamless service experience, when devices leave or enter the NPN coverage area.

The specific solution to provide the services referred to above with SSC depends on the NPN deployment option: PNI-NPN or SNPN. For PNI-NPNs, the PLMN furnishes radio access both inside and outside the private premises. Then, ordinary intra-PLMN handover procedures are triggered when the UEs exit or enter the private venue's inner perimeter. These procedures ensure the SSC to the UEs when they cross the private venue borders. For SNPNs, typically, the SNPN only provides radio access on-premises, whereas a PLMN is needed to support 5G connectivity outside. Thus, either the UE has a Dual Subscriber Identity Module (SIM) and a subscription with the PLMN, or there is a roaming agreement between the SNPN and the PLMN are required to enable private UEs to maintain the connectivity out-of-premises. Here, we focus on the second option as many commercial mobile end devices, such as most of the Internet of Things (IoT) sensors, are equipped with a single SIM.

3GPP Release 16 allows for inter-PLMN mobility procedures with SSC assurance for both local breakout and home routed roaming scenarios. The same procedures might be used when the private UEs roam from the SNPN to the PLMN. Support of mobility between the SNPN and the PLMN imposes specific requisites on roaming agreement. For instance, there shall be direct communication between 5GC- CPs in the two networks. On the one hand, the public Unified Data Management (UDM)/Home Subscriber Server (HSS) needs to do an onboarding of the UE subscription data by requesting them to the private UDM/HSS. On the other hand, the public and private 5GC entities must interact to carry out the corresponding handover procedures.

5.2 Multi-site NPN

A multi-site NPN scenario represents a deployment use case whereby NPN provisioning aims at serving a given enterprise customer whose facility includes two or more sites, e.g., branch offices. Depending on the use case, a multi-site NPN scenario can represent (i) a connection of individual SNPNs, each deployed locally at every branch office; (ii) a single PNI-NPN (see Figure G-6).

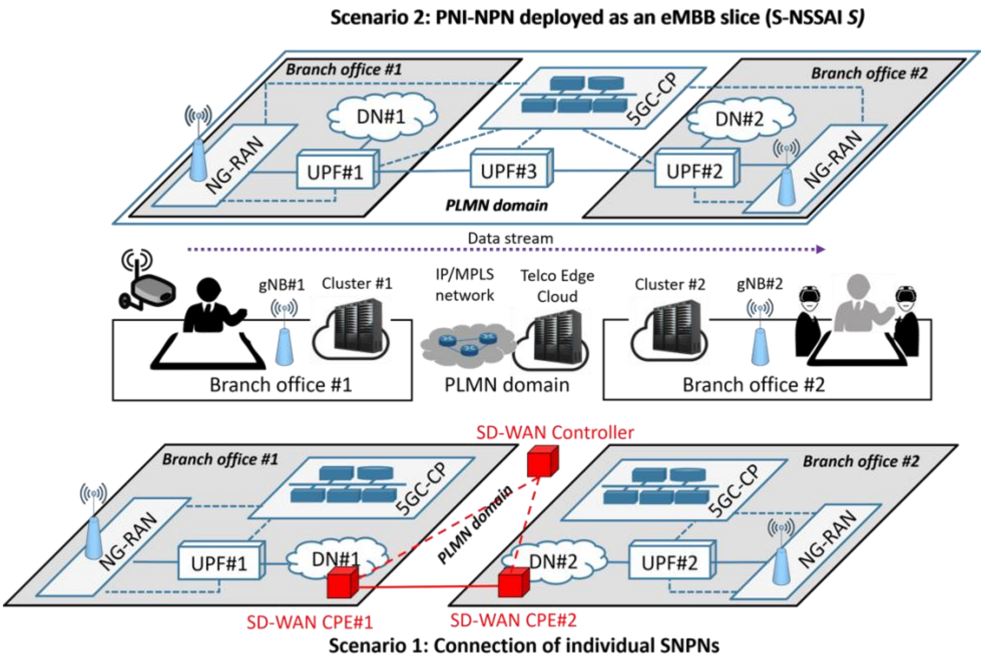


Figure G-6. Candidate multi-site NPN deployments.

The first category is typical for industry 4.0 enterprises, where independent 5G-enabled manufacturing tasks are executed at individual branch offices. The branch offices need to communicate between them to only exchange industry- specific data (e.g file exchange, database accessibility); this means that no signaling/data plane 5G traffic is exchanged among individual SNPNs. For this communication, a plausible solution is to set up a SD-WAN service (overlay) atop the PLMN-Op's IP/MPLS substrate (underlay).

In the second category, it is assumed there exists a single 5GS for the entire facility. Unlike the first category, the 5GS is now partially hosted by the PLMN. Typical layouts in this category consist of having lightweight branch offices, keeping user plane on premises and offloading 5GC-CP complexity towards PLMN-Op's edge node. The resulting deployment scenario is formed of a set of branch offices, each hosting a CP-less 5GS (i.e., RAN and UPF), and a PLMN- Op's edge node, which hosts 5GC (i.e., 5GC-CP and UPF).

The latter scenario may fit for customers requiring the use of eMBB capabilities among branch offices, for the delivery of 5G media services such as UHD video streaming (e.g., telepresence in council meetings) and XR video experience (e.g., AR assisted supervision on a remote factory). In both cases, the service consists in streaming video traffic from one branch office towards one or more remote offices, leveraging traffic casting (e.g., unicast/multicast/broadcast) mechanisms as needed. The on-premises UPF from source branch office, which performs UL Classifier (UL-CL) functionality, receives incoming IP packets corresponding to video service. Grouped in a PDU session, these IP packets are encapsulated in a GTP tunnel before their delivery to the PLMN hosted UPF. This UPF, deployed at PLMN-Op's edge node and performing the PDU Session Anchoring (PSA) functionality, receives the encapsulated sessions and applies necessary traffic casting policy to route them towards end branch offices, where local UPFs proceed with the GTP tunnel decapsulation, so IP packets can reach end users. In the overall process, participant UPFs are in charge of keeping 5QI-to- DSCP mapping (i.e., translation of 3GPP 5G QoS indicators into IP QoS indicators), so the QoS can be assured along the IP/MPLS substrate which connects the different branch offices.

6 Challenges

In this section, we identify some of the key challenges and future research directions arising from realizing 5G NPNs.

6.1 Zero-touch Practices on NPN Management

A simplified management of the NPN and a smooth integration in the IT infrastructure of the enterprise customer are key challenges for the success of 5G NPNs. To achieve those goals, NPNs have to embrace full automation in network and service orchestration and implement extensive zero-touch management approaches. The realization of this vision in SNPN leverages two principles: Artificial Intelligence (AI) and Intent-based interfaces. 3rd parties like µOperator can help enterprise customers to integrate these principles into their management stack solutions.

On the one hand, the introduction of AI principles allows for data-driven, self-X network and service management, minimizing the intervention of the NPN operator (i.e., the enterprise customer or the delegating company). Decisions that today traditionally takes slow human interactions, based on carrier-grade network characterization and optimization methods, should be autonomously performed by (ML) algorithms with a holistic view of the network, enabling software components to directly contribute into decision-making activities related with the SNPN management. Despite the general applicability of ML-based solutions, their practical application often relies on the possibility to access real-time data to perform analytics and diagnosis. To that end, further research work on data aggregation mechanisms (e.g., model-based streaming telemetry) needs to be made.

On the other hand, the design of intent-based language will allow the NPN operator to interact with the NPN resources, functions and services using business primitives, instead of low-level network configuration. With the use of an intent- based northbound interface, the NPN operator can operate the NPN in a user-friendly manner, by issuing expectations (intents) rather than specific network control/orchestration requests. Before getting this intent-based northbound ready for use, it is needed to understand how business intents are to be described and translated into enforceable goals and actions at resource, network and service layers. This requires further innovation and research work ahead on intent modelling, especially on intent decomposition, intent monitoring and intent assurance. Many of these aspects are still on early discussion, in both industry fora (e.g., TM Forum initiative on autonomous networks) and standardization bodies (e.g., 3GPP SA5 and ETSI ZSM).

6.2 Multi-WAT in 5G NPNs

Multi-Wireless Access Technology (WAT) is appealing to affordably improve the 5G NPNs performance. Many of the current private networks are based on Wi-Fi deployments for wireless connectivity. Thus, the integration of Wi-Fi with 5GNR in 5G NPNs reduces the entry barrier and enhances their QoS by leveraging the already deployed infrastructure. The integration of 5G with WiFi has been addressed in 3GPP Releases 15 and 16 by means of the Non-3GPP Interworking Function (N3IWF). This function abstracts the complexity of each Wi-Fi access point making it appear as a single gNB towards the UPF. Nonetheless, it is still required to devise and develop smart mechanisms that allow to easily combine 5GNR and Wi-Fi to provide advanced connectivity with improved reliability and throughput. For example, solutions to decide when switch, split or steer the eMBB traffic through the available WATs according to a given goal or SLA. Besides Wi-Fi, alternatives technologies like Light-Fidelity (Li-Fi) can also be integrated in multi-WAT 5G NPNs to further enhance their performance and increase the security of wireless communications.

6.3 Enabling and Validating 5G NPNs with E2E Deterministic QoS Support

One of the primary drivers behind Beyond 5G NPNs is the support for private critical services with stringent latency and reliability requirements such as connected robotics and autonomous systems. However, they are still open questions what is required, besides the URLLC capabilities included in recent 3GPP releases, to provide end-to-end (E2E) deterministic QoS support and which critical private services can be supported by 5G NPNs.

In addition to the data plane aspects, AI empowered management planes are also a requisite to cope with the complexity of configuring the different domains of the 5G NPNs and provide coherence among them, e.g., to ensure the end-to-end packet delay budget.

Last, deriving analytical performance bounds of the end-to-end QoS metrics (e.g., delay, jitter, packet loss, and reliability) and their experimental validation are essential to truly ensure that a given configuration of the 5G NPN meets the Service Level Agreement (SLA) of the private critical services. SLA violations might have a highly negative impact, e.g., long production downtimes in the factory or life-threatening in remote surgery. Therefore, the SLA violation probability has to be known and kept within the specific safety margins for the particular critical service.

6.4 Capability Exposure in PNI-NPN

The previous challenges mainly apply for SNPNs scenarios. However, as described above, PNI-NPNs represent a reduced entry barrier option to have an NPN for some enterprise customers such as SMEs or incumbent digital service providers. In PNI-NPNs, there are situations in which the customer enterprise wants to retain control and management of some specific parts of the network. In such a case, hybrid solutions can be defined, with PLMN-Ops taking the main control and management activities, while exposing needed capabilities to the enterprise customer. These capabilities can be of two types:

- Configuration related capabilities: this group of capabilities defines the ability of an enterprise customer to modify the parameters of certain network functions and infrastructure nodes. To that end, the PLMN-Op needs to characterize the permissions (i.e., *isReadable*, *isWritable*, *isInvariant*, *isNotifyable*) associated to these parameters accordingly.
- Assurance related capabilities: this second capability group defines the ability of an enterprise customer to subscribe to certain performance measurements and fault alarms, so that the customer can consume them in the format it sees more appropriate according to its business needs (e.g., for performance management, batches vs streaming).

To make capabilities available for consumption by the enterprise customer, the PLMN-Op shall have an BSS hosted integration fabric, in charge of mediating the request- response messages between the customer and PLMN-Op. It is important for the PLMN-Op to expose these capabilities in a controlled, secure and auditable way. To that end, the solution design for this integration fabric will require the implementation of an API gateway, together with mechanisms for token-based authentication and non-repudiation. However, how to build this solution is still unclear, and much work ahead is agreed in the telco industry community. On the one hand, it is still not clear for enterprise customers the capabilities they need to consume for their business processes; this is mostly due to their lack of knowledge/expertise with telco and networking issues. On the other hand, the PLMN- Ops need to think about the implementation of this BSS hosted integration fabric, with a particular focus on:

- the control, security and auditability implications of exposing these capabilities to the customer, especially considering multi-tenancy environments, where multiple customers will request the PLMN-Op to consume (potentially) different capabilities.
- the mapping of customer requests into network actions, and the API transformation behind this. In this regard, the PLMN-Op shall define a mechanism to map customer-facing, service APIs into low-level, internal network APIs.

Acknowledgement

This work has been partially funded by the H2020 project 5G-CLARITY (Grant No. 871428) and the Spanish national project TRUE-5G (PID2019-108713RB-C53).

References

[1] “5G for Business: A 2030 Market Compass -- Setting a Direction for 5G Powered B2B Opportunities", Ericsson White Paper, October 2019.

[2] “5G Non-Public Networks for Industrial Scenarios”, 5G-ACIA White Paper, July 2019.

[3] A. Aijaz, “Private 5G: The Future of Industrial Wireless”, *IEEE Industrial Electronics Magazine*, vol. 14, no. 4, pp. 136–145, 2020.

[4] A. Rostami, “Private 5G Networks for Vertical industries: Deployment and operation models", in *2019 IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 433–439.

[5] J. Ordonez-Lucena *et al.*, “The Use of 5G Non-Public Networks to support Industry 4.0 scenarios”, *Proc. of the 2019 IEEE Conf. on Standards for Commun. and Netw. (CSCN)*, 2019, pp. 1–7.

[6] 3GPP TS 23.501 V.16.5.0, “System architecture for the 5g system (5GS); Stage 2 (Release 16)”, July 2020.

[7] “Integration of industrial ethernet networks with 5G networks", 5G-ACIA White Paper, Nov. 2019.

[8] J. Prados-Garzon and T. Taleb, “Asynchronous time-sensitive networking for 5G backhauling”, *IEEE Network*, vol. 35, no. 2, pp. 144–151, 2021.

[9] A. Nasrallah *et al.*, “Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and related 5G ULL Research”, *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 88–145, Q1 2019.

[10] 5G-SMART Project, “Deliverable 5.1: First report on new technological features to be supported by 5G standardization and their implementation impact”, May 2020.

[11] A. Jerichow *et al.*, “3GPP Non-public Network Security”, in *Journal of ICT Standardization,* vol. 8, no. 1, pp. 57–76, 2020.

[12] INSPIRE-5Gplus project, “Deliverable 2.1: 5G security: Current status and future trends”, February 2021.

[13] M. Matinmikko-Blue and M. Latvaaho, “Micro Operators accelerating 5G deployment", *Proc. of the 2017 IEEE Int. Conf. on Ind. and Inf. Syst. (ICIIS*), 2017,

pp. 1–5.

[14] “Integration of 5G with Time-Sensitive Networking for Industrial Communications", 5G-ACIA White Paper, January 2021.

[15] “5G campus networks LTE and 5G-technology for local company networks", T-Systems International GmbH White Paper, 2020.

Paper H

On the Rollout of Network Slicing in Carrier Networks: A Technology Radar

Jose Ordonez-Lucena, Pablo Ameigeiras, Luis M. Contreras,
Jesús Folgueira and Diego R. López

Disclaimer:
This work has been published in the MDPI's Sensors journal.
Citation information: DOI 10.3390/s21238094.

Abstract

Network slicing is a powerful paradigm for network operators to support use cases with widely diverse requirements atop a common infrastructure. As 5G standards are completed, and commercial solutions mature, operators need to start thinking about how to integrate network slicing capabilities in their assets, so that customer-facing solutions can be made available in their portfolio. This integration is, however, not an easy task, due to the heterogeneity of assets that typically exist in carrier networks. In this regard, 5G commercial networks may consist of a number of domains, each with a different technological pace, and built out of products from multiple vendors, including legacy network devices and functions. These multi-technology, multi-vendor and brownfield features constitute a challenge for the operator, which is required to deploy and operate slices across all these domains in order to satisfy the end-to-end nature of the services hosted by these slices. In this context, the only realistic option for operators is to introduce slicing capabilities progressively, following a phased approach in their roll-out. The purpose of this paper is to precisely help design this kind of plan, by means of a technology radar. The radar identifies a set of solutions enabling network slicing on the individual domains, and classifies these solutions into four rings, each corresponding to a different timeline: (i) as-is ring, covering today's slicing solutions; (ii) deploy ring, corresponding to solutions available in the short term; (iii) test ring, considering medium-term solutions; and (iv) explore ring, with solutions expected in the long run. This classification is done based on the technical availability of the solutions, together with the foreseen market demands. The value of this radar lies in its ability to provide a complete view of the slicing landscape with one single snapshot, by linking solutions to information that operators may use for decision making in their individual go-to-market strategies.

1 Introduction

Over recent years, the telco industry has actively focused on the exploration of technologies to accelerate the roll-out of fifth generation (5G) systems worldwide. Unlike 4G, mainly focused on providing mobile broadband services to end users, 5G has been designed from its inception to help boost the digital transformation of vertical industries (i.e., industry sectors aiming at becoming fully digital, such as manufacturing, smart cities, transportation or agriculture [1]). The direct involvement of the so-called verticals and their specific needs within the 5G technology ecosystem implies the emergence of a new wave of use cases, with very different requirements in terms of performance (e.g., throughput, latency and reliability) and functionality (e.g., mobility, security, service continuity support), some of them very stringent.

To satisfy these different (and potentially conflicting) requirements in a cost-effective manner, operators need to turn their networks into programmable multi-service platforms, embracing the infrastructure and functional sharing mechanisms commonly referred to as network slicing. With network slicing, the operator's network can be logically split into a set of programmable network partitions (i.e., network slices), each designed to satisfy a particular set of service requirements. The service-tailored logical networks resulting from this partitioning can be executed in parallel but need to be operated in isolation from each other. This means that despite running on a common (shared) network infrastructure, network slices require separate (independent) management [2].

The technology foundations for slicing are already here. On the one hand, network functions virtualization (NFV) allows deploying the functions of every slice with necessary capacity where and when required. On the other hand, software defined networking (SDN) allows operators to programmatically steer traffic within the slice, across the deployed functions. In addition to these dynamic control means, NFV and SDN technologies also provide the ability to resize and move workloads at operation time, in such a way that the service requirements can be always met, regardless of network conditions, e.g., a faulty node or traffic load surges. However, to fully exploit the benefits that network slicing brings, it is important for the operator to apply the dynamic allocation and tailored partitioning of resources at all segments, from the radio access to the data network, including all the network domains in between. This means that the slice concept shall span the entire operator's managed network infrastructure, resulting in the provisioning and operation of end-to-end (E2E) network slices.

The E2E nature of slicing forces operators to keep consistency in the behavior of individual slices along the different domains. This may bring significant operational challenges in commercial networks, as outlined below:

- **Slicing readiness varies across the different domains**. In fact, the degree of penetration of slicing features in the different technology domains is not the same. For example, while the core network has incorporated network slicing support since the first 5G release (3GPP Release 15), the transport network does not support any native slicing feature yet, and first solutions have only recently been integrated into the radio access network. The main reason why the maturity level varies across technology domains (and their corresponding management domains) is mainly due to the existing fragmentation in the standardization arena, with a high number of participating Standard Development Organizations (SDOs). In the current landscape, each SDO addresses a portion of the E2E problem, developing slicing specifications for this portion under assumptions that do not necessarily match the assumptions made by other standard bodies, which typically address other portions. A clear example of this mismatching can be observed on the priorities that different SDOs set in relation to which slicing features need to be worked out in each release. In fact, these priorities are quite different across SDOs, both in time and scope.
- **Scalability burdens**. The higher the number of slices running in parallel, the heavier the burden on the operator's OSS (Operations Support System) in terms of scalability. In fact, having a high number of instantiated tiny network slices, each requiring separate control and management, may well imply a strong impact on OSS functions (orchestration, assurance, etc.). This requires the operator to find the right balance in the slice design and activation patterns, looking to minimize this impact while properly addressing service demands. The introduction of advanced configuration and automation capabilities in OSS assets is also a must, in order to reduce the number of touches, especially in the assurance phase.
- **Multi-provider solutions**. Upcoming 5G commercial networks are to be built out of solutions from multiple technology providers. The reason for this approach is essentially related to the dangerous effect of monoculture. Single-vendor dependency is a killer for innovation, as it restricts open collaboration from the broader 5G ecosystem of companies developing new technology, use cases, and services that the market expects. In this multi-vendor ecosystem, the challenge for operators will be in the appropriate combination of pieces from different providers and in ensuring they work together, within and across domains. The high integration efforts on the operator side to achieve multi-provider interoperability can be partially relieved by

selecting solutions which are standards-compliant, i.e., based on the use of open interfaces.

- **Brownfield environments**. Carrier networks are formed of already available equipment and functions (legacy is the common term for them), aimed at offering services from previous generations and even former releases of the current one. The need to keep this legacy up and running shall be combined with the introduction of the slicing functionality, avoiding the creation of silos. Unlike greenfield environments (e.g., private 5G networks), where network slicing can be easily launched as soon as commercial products are available, in carrier networks the operator needs to carefully upgrade its assets in such a way that the legacy and slicing features can coexist. This process needs to be conducted in a cost-efficient way, ensuring that the upfront CAPEX behind every required upgrade will be compensated with a large mass of customers willing to consume the added slicing features.

The above challenges outline the main issues that operators need to work out to fulfil the promise of E2E network slicing, in 5G and beyond. However, solving these issues towards this ultimate goal may require years, especially for the first issue (i.e., the different slicing readiness across the different domains). In the meantime, operators need to look for workarounds to start commercializing and monetizing network slicing, incorporating solutions in their portfolio according to the set of slicing capabilities available in their networks by then. The population of the portfolio is not an easy task, given the quite fragmented landscape in standards and literature, with plenty of ad-hoc solutions that cover particular slicing aspects from different domains and under different assumptions. Defining a network slicing rollout plan based on the current collection of solutions is a critical activity for operators to succeed in the market. This activity consists of two steps. First, identifying all relevant solutions and positioning them in a common space, with multiple dimensions that reflect the E2E conception of network slicing. Secondly, defining a go-to-market strategy [3], based on deciding which solutions will be made available, when, for which customers, and under which business models.

The purpose of this paper is to address the first step, outlining a technology radar to model this common space. This radar presents a phased-based vision for the introduction of network slicing capabilities in commercial networks, considering all the domains impacted in operator assets, including the main three technology domains and the OSS. In this vision, the radar identifies different solutions for network slicing and captures them into four rings, each corresponding to a different timeline: as-is ring (today's slicing), deploy ring (short-term slicing), test ring (medium-term slicing) and explore ring (long-term slicing). The position of each solution in the radar is done according to three different criteria: (i) the technology maturity of the solution, which is related to the readiness of the corresponding standards; (ii) the roadmap of commercial products, which specifies when the features associated with the solution will be available; and (iii) the relevance for the customers, which determines the prioritization of the solution over others.

To the best of our knowledge, this is the first work in literature that provides a radar for E2E network slicing, with a focus on the rollout of this technology in carrier networks. The radar captures a complete landscape of network slicing solutions, linking them to different timelines. In addition to this timing, the radar will also outline the dimensions impacting slice realization, from E2E viewpoint. These dimensions are to be analyzed in each of the operator managed domains, including Radio Access Network (RAN), Core Network (CN), Transport Network (TN) and OSS. In the *RAN domain*, network solutions are to be discussed based on three dimensions: functionality (e.g., disaggregation and O-RAN integration), radio resource allocation and penetration (in micro and macro cells). The *CN domain* will focus on how to use and combine core network functions for different slices, including baseline and value-added functions, depending on isolation and customer requirements. In the *TN domain*, slicing is to be discussed based on the availability of transport technologies and SDN-enabled capabilities, including programmability and automation. Finally, in the *OSS domain*, aspects related to network slice lifecycle management and capability exposure (i.e., to expose slicing capabilities to customers through service APIs) will be taken into account. These dimensions are used to characterize the different solutions captured in the radar, providing guidance on how and where to use them. This information, together with the timeline provided for these solutions, is the input material that enables an operator to define the plan for network slicing rollout. For further details on how to design and execute this plan, see recommendations reported in [4]-[6].

This article is structured as follows. Section 2 provides the technical background of E2E network slicing, with focus on the modelling, system architecture and deployment related aspects. Section 3 outlines the impact that the network slicing may introduce on the different technology domains. The understanding of these features will enable the reader to understand the radar, which is introduced in Section 4. The radar is the core contribution of this article, and hence deserves a detailed discussion, with a thorough analysis of all the solutions along the different dimensions: CN (Section 5), RAN (Section 6), TN (Section 7) and OSS (Section 8). Finally, Section 9 summarizes the main conclusions of this work.

2 Network Slicing: Concept, System Architecture and Deployment

This section provides a technical background of network slicing, outlining the main artifacts involved in their realization.

2.1 Network slice concept

A network slice provides a service-tailored connectivity pipe to one or more service applications hosted by the Data Network (DN). Examples of service applications include Immersive Reality (XR) streamers, IoT platforms or V2X backend servers. These applications can be associated with operator services (e.g., communication services) or with third party services. Devices subscribed to one service can establish communication with the service applications through the corresponding network slice, which will provide an enhanced connectivity profile in terms of functionality, performance and/or security [2].

The fact that makes network slicing an E2E concept is that the device-to-application connectivity pipe involves all the technical domains within the operator's managed network, including the RAN, CN and TN domains.

The **RAN domain** allows connecting the end devices to the operator's network using a wide variety of access technologies. In this paper, we focus on the Next Generation RAN (NG-RAN) [7][8]. The NG-RAN consists of multiple gNBs, which provide connectivity towards end devices using 5G New Radio (NR) technology. To take advantage of the benefits that RAN virtualization brings in terms of scalability and centralization, the standards have moved to a new architecture model where a gNB can be logically split into three entities denoted as radio unit (RU), distributed unit (DU) and centralized unit (CU). The NR protocol functions that correspond to each of these entities are determined by the so-called split options. Though there exist up to eight split options available for this gNB decomposition, after a thorough analysis the industry has opted for two: split 2, defined by the 3GPP and acting as a high layer split; and split 7-2x, defined by the O-RAN Alliance and acting as a low layer split [9]. The figure in Section 2.2 details the partitioning of NR protocol functions into RU, DU and CU, according to these two split options.

The **CN domain** allows end devices to send/receive mobile traffic to/from DN hosted applications or the Internet. In 5G, this functionality is provided by the 5G Core (5GC) [10]. Designed from its inception to be cloud-native, the 5GC follows a service-based architecture (SBA), with the definition of a disaggregated and modular, containerized control plane which is fully decoupled from User Plane Functions (UPFs).

Finally, the **TN domain** is in charge of providing infrastructure connectivity between the RU (the entry point to the network for the device) and the DN (where the service applications are hosted). To that end, it makes use of a wide variety of forwarding devices, which are founded on different technologies (e.g., IP/MPLS, optical/DWDM and microwave/backhaul) and connected forming different topologies (e.g., ring, mesh, hub-and-spoke), across different aggregation levels. The TN domain sets up the data path across the different RAN and CN functions, by mapping their interfaces into Wide Area Network (WAN) infrastructure resources. According to this mapping, different TN segments can be outlined:

- Fronthaul segment, scoping the data path between the RU and the DU. This data path implements the O-RAN fronthaul interface (split 7-2x). The control, data, management and synchronization planes of this interface are defined in [11][12].
- Midhaul segment, which sets up the data path between the DU and the CU. This data path implements the 3GPP F1 interface (split 2) [13].
- Backhaul segment, established between the CU and the UPF. It covers two 3GPP interfaces: N3 (CU-to-UPF) and N9 interface (UPF-to-UPF). When the UPF connected to the CU is the anchor UPF, then the N9 interface is not needed [10].
- DN segment, establishing connectivity between the (anchor) UPF [10] and the DN. This segment is the transport level realization of the 3GPP N6 interface [14].

To make slicing a reality, every technical domain is split into one or more logical network partitions, each referred to as a network slice subnet. The definition of multiple slice subnets on a single domain allows this segment to provide differentiated behaviors, in terms of functionality and/or performance. The stitching of slice subnets across the RAN, CN and TN results in the definition of network slices.

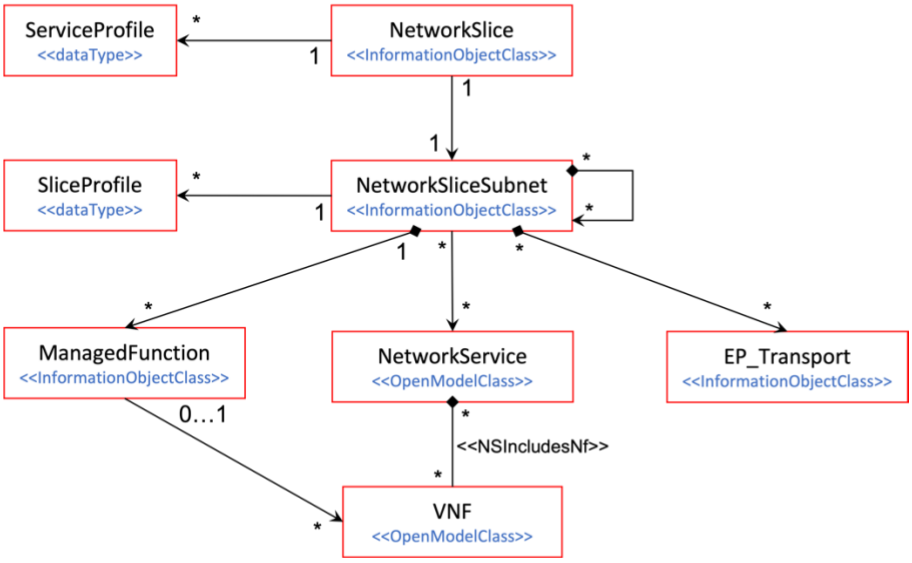


Figure H-1. 3GPP Information Model of a network slice: the Network Slice NRM fragment.

The rules for the definition of network slice subnets and their composition into network slices are detailed in the 5G Network Resource Model (NRM), specifically in the Network Slice NRM fragment [15]. This fragment captures the information model of 5G network slicing. As seen in Figure H-1, this model specifies the relationships across the manageable entities, each represented as a separate Information Object Class (IOC). An IOC captures the semantics and attributes of a manageable entity; in other words, it defines the class based on which instances (objects) from this entity can be created. In the model, we have four different IOCs: (i) *NetworkSlice IOC,* representing a network slice; (ii) *NetworkSliceSubnet IOC*, associated with a network slice subnet; (iii) *ManagedFunction IOC*, which represents a 5G network function; and (iv) *EP_Transport IOC*, which represents an interface associated with transport level information, e.g., transport address, reachability information, and QoS profiles. Note that for NetworkSlice and NetworkSliceSubnet IOCs, two additional constructions are defined:

- ServiceProfile: represents the requirements that the slice needs to support for a particular service. The 1:N relationship of this construction with the NetworkSlice IOC is because one network slice can host multiple services, as long as they do not impose conflicting requirements. These services can be from the same customer (the slice is dedicated for this customer) or different customers (the slice is used for serving multiple customers).
- SliceProfile: similar to the ServiceProfile, but applied to the slice subnet level.

Though multiple associations can be found across these IOCs, the most typical case consists in having one slice consisting of two slice subnets: one including NG-RAN functions (RAN slice subnet) and the other 5GC functions (CN slice subnet). Each network slice subnet can be deployed as an ETSI network service (via the *NetworkService* class), provided that one of the network functions is realized as a Virtualized Network Function (VNF) [16]. Finally, the EP_Transport IOC features the TN slicing behavior across the RAN and CN slice subnets, by mapping the QoS requirements associated to the different interfaces (e.g., F1, N3, N6, etc.) into appropriate WAN resources.

2.2 Architectural framework for Network Slicing

Figure H-2 illustrates the system architecture design for network slicing. This system is structured into two layers: the network layer, which provides the individual slices with the required user and control plane functionality, across all technical domains; and the OSS layer, which hosts all the assets that are used for the design, provisioning, and operation of network slices.

The network layer is formed of a collection of modular network functions that can be flexibly combined together to build up network slices. Figure H-2 shows an example with three different slices, one for each main 5G service category. The fact that every slice needs to be provisioned with a service-tailored user and control planes justifies the allocation of dedicated NR and 5GC network functions in their RAN and CN slice subnets. Which network functions are to be dedicated per slice and which ones can be shared with other slices needs to be analyzed case by case, as it depends (i) on the isolation requirements of the slice under consideration, and (ii) the type of customer that will consume this slice. Further discussion on this topic is captured in Section 5.1 and Section 6.1.

The OSS layer conveys all the Operation, Administration and Maintenance (OAM) tools that operators may use to manage the different slices across their entire lifetime [17]. These tools are classified into four main groups, depending on their scoped functionality: (i) design,

(ii) data management, (iii) assurance and (iv) orchestration. The most notable group is the orchestration, responsible for all the activities related to slice provisioning (i.e., going from a service order to a deployed network slice) and slice operation (i.e., keep the deployed slice at the desired state at run-time). This collection of activities shall be performed consistently across all the technical domains, with an E2E perspective. The specificities of these domains, each with a different pace of technological evolution and with legacy from multiple vendors, unveils non-negligible integration issues for operators. This is exacerbated as the number of slices running in parallel increases.

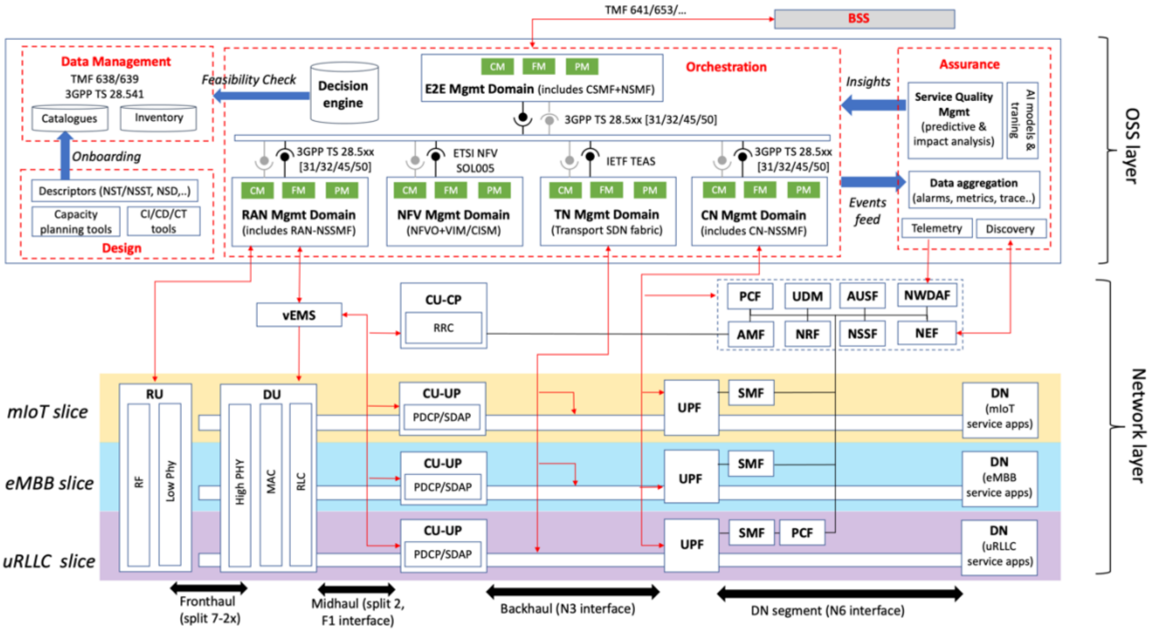


Figure H-2. Network slicing system architecture.

To cope with the above integration and scalability challenges, operators are required to adopt novel architecture approaches on the orchestration group. Service-based paradigm, which is about designing software architectures using Application Programming Interfaces (APIs) based on web-based technology, is considered as a potential facilitator in this respect. Originally conceived for 5GC, this architectural style can also be applied to the OSS layer, resulting in a Service-Based Management Architecture (SBMA). The SBMA consists of replacing traditional management entities (e.g., Network Managers) with a federated set of management functions that provide services to each other using REST APIs. The adoption of SBMA allows fleeing from point-to-point protocol interfaces (e.g., 3GPP Itf-N interfaces) to a service bus that interconnects all the management functions and polices the interactions across them. Different SDOs have already captured the benefits of having a SBMA in their architecture specifications. For example, 3GPP SA5 [18] and ETSI ISG ZSM [19] have defined their architectural frameworks based on SBMA. Even ETSI ISG NFV, which originally chose an interface-centric approach for the design of the Management and Orchestration (MANO) framework, has now decided to migrate towards a SBMA from NFV Release FOUR on wards [20].

As seen in Figure H-2, the management functions building up the OSS's orchestration group are arranged into five separate domains: RAN, NFV, TN, CN and E2E management domains. This design criterion represents a separation of concerns that is reasonable from the operator's viewpoint, and which relies on two principles:

- The independent management of network resources and functions from different technical domains. This facilitates a decoupled evolution of RAN, CN and TN, and allows the operator to select the technologies and vendor solutions they want for every technical domain.

- A clear separation between management (i.e., OAM activities on individual technical domains) and orchestration (i.e., coordination and conflict resolution activities across technical domains). In the proposed solutions, the RAN, CN and TN management domains are focused on management activities, while the NFV and E2E management domains are the ones responsible for orchestration.

The interactions across the different management domains are done with a service bus, which features the ZSM cross-domain integration fabric [19]. As seen, it is important for the different domains to make capabilities available for external consumption through standard APIs. Figure H-2 captures relevant references for these APIs.

2.3 Network Slice Description

One of the main business cases for network slicing is Network Slice as a Service (NSaaS) [17]. In this business model, an industry vertical (acting as the network slice customer) requests the network operator (acting as the network slice provider) to allocate a dedicated network slice satisfying a particular set of service requirements. With a large variety of emerging verticals in the market, it is fundamental for the operator to define a unified ability to interpret service requirements from different verticals, and to represent them in a common language. This unification will help the operator capture vertical-specific service requirements and translate them into appropriate network slice provisioning actions.

In this regard, the GSM Alliance (GSMA) has promoted the idea of having a universal slice blueprint providing a point of convergence between telco and vertical industries on network slicing understanding. This blueprint, known as the Generic network Slice Template (GST) [21], contains a set of attributes that allow the characterization of any network slice. The most representative GST attributes are included in Table H-1.

A Network Slice Type (NEST) is the result of filling GST attributes with values according to the service requirements. In essence, a NEST is a filled-in version of a GST, and can be used by an operator and a vertical customer to agree on the Service Level Agreement (SLA). Different NESTs allow the description of different network slices. For slices based on 3GPP 5G service categories, the operator may have a set of standardized NESTs (S-NESTs). For slices addressing specific industry use cases, the operator can define additional NESTs (P-NESTs) [22].

Table H-1 provides one example with three NESTs, one for each slice represented in Figure H-2. The table also qualifies the impact of the NEST attributes on every technical domain.

Table H-1. Examples of NESTs for the three slices represented in Figure H-2.

GST Attribute	Network domain			mIoT NEST	eMBB NEST	uRLLC NEST
	RAN	TN	CN			
Availability	X	X	X	99.9	99.99	99.9999
Session and Service Continuity (SSC) Support			X	SSC Mode 1: the IP address is preserved	SSC Mode 1: the IP address is preserved	SSC Mode 1: the IP address is preserved
Maximum DL (UL) throughput per UE	X		X	2 (4) Mbps	200 (200) Mbps	40 (40) Mbps
DL (UL) throughput per slice	X	X	X	Maximum: 30 (60) Gbps	Guaranteed: 300 (200) Gbps	Maximum: 20 (20) Gbps
Maximum number of PDU sessions			X	500,000	80,000	1,500
Slice Quality of Service (QoS)	X	X	X	3GPP 5QI: 9	3GPP 5QI: 1,2,5,8,7,8,0	3GPP 5QI: 82
Maximum supported packet size	X	X	X	300 bytes	1500 bytes	160 bytes
UE density (per km^2)	X		X	100,000	500	80
Simultaneous use of the slice	X	X	X	Can be used simultaneously with any slices with same SD value but different SST value	Can be used simultaneously with any slices with same SD value but different SST value	Cannot be used simultaneously with any another slice
Supported device velocity	X		X	10km/h - Pedestrian	500km/h – High speed vehicular	120km/h - Vehicular

NOTE 1: Slice QoS parameters attribute defines all the QoS relevant parameters supported by the network slice, including priority level, packet delay budget (i.e., maximum allowed latency), packet error rate, jitter, and maximum packet loss rate. These attributes are indexed using a 3GPP defined scalar called 5G Quality Indicator (5QI). For further details on 5QIs, see [10].

NOTE 2: The maximum number of PDU sessions attribute describes the maximum number of concurrent PDU sessions supported by the network slice. To enforce this quota on the slice, the 5GC network function called Network Slice Access Control Function (NSACF) is required. For further information of this function, see Section 5.2.

2.4 From a Service Order to a Deployed Network Slice

For enabling NSaaS, the operator registers in their portfolio a collection of service offerings, each representing a slice associated with an SLA. This SLA includes two main types of information: (i) technical information, which is captured in a NEST; and (ii) charging and pricing information. For the request of a network slice, the vertical customer browses the operator's portfolio, selects the service offering that best fits their needs, and issues the corresponding service order. From this point on, the following activities are triggered on the operator's side: The operator's BSS (Business Support System) captures the service order. It uses the charging and pricing information to configure the customer profile, and forwards the technical information (the NEST) to the E2E management domain using TM Forum Service Ordering API [23].

1. In the E2E management domain, the Communication Service Management Function (CSMF) translates the NEST parameter values into the ServiceProfile construction (see Figure H-1).
2. The CSMF requests the allocation of a network slice based on this ServiceProfile. The CSMF sends this request to the Network Slice Management Function (NSMF), using the *allocateNsi* operation (see clause 6.5.1 from 3GPP TS 28.531 [24]).
3. With the network slice allocation request, the NSMF is asked to deploy a network slice instance (NSI) on the operator's managed network infrastructure, in such a way that the service requirements captured in the ServiceProfile are fulfilled. Before beginning the deployment of network slice subnet instances (NSSIs) and the reservation of WAN resources across them, the NSMF shall make sure that the network slice allocation is feasible. To that end, it requests the Decision Engine (see Figure H-2) to perform a feasibility check procedure. The complete procedure execution can be separated into two parts. The first part checks for the qualitative network capabilities that the network slice instance requires, e.g., availability of a specific radio access technology or feasible network function configurations. This is expected to be completed rather quickly and can therefore provide a quick reply in the case of a negative ("network slice instance unfeasible") response. In case of a positive qualitative check, the second part quantitatively checks if there are enough infrastructure resources (including radio, WAN and computer resources) available for use. It also calculates confidence values if resource availability is associated with statistical uncertainty, e.g., due to statistical fluctuations in resource consumption of already deployed slice instances.
4. If feasible, the NSMF proceeds with the NSI allocation, based on the allocation of (i) the RAN NSSI, (ii) the CN NSSI, and (iii) the WAN resources providing end-to-end connectivity. In this process, the NSMF interacts with the Network Slice Subnet Management Functions (NSSMFs) from the RAN and CN management domains, and with the SDN fabric from the TN management domain. The NSSMFs may interact in turn with the NFV Orchestrator (NFVO) through SOL005 [25], for the cases where the NSSIs can be deployed as ETSI network services.

Figure H-5 illustrates the deployment view of the network slices shown in Figure H-2. This view shows these slices are allocated on the operator's managed infrastructure, in the form of NSIs. For this allocation, it is assumed that (i) the network slices have been ordered according to the NESTs specified in Table H-1, and (ii) the infrastructure consists of a RAN with cell sites attached via dedicated fibers to a three-tier TN. This capillarity in TN design allows the distribution of compute capacity, across Points of Presence (PoPs), which are physically deployed at three different aggregation levels:

- Central PoPs, which correspond to large-scale core cloud sites. They are typically built with commodity (x86 or ARM based) hardware and are ideal to host IT applications and delay-tolerant telco workloads.
- Regional PoPs, which represent Central Offices featuring the telco edge cloud [26]. The regional PoPs provide virtualization capabilities closer to service delivery endpoints in order to reduce the delay budget, making them ideal to host delay-critical telco workloads.
- Finally, access PoPs, which are associated with far edge sites. Much more distributed and closer to cell sites than regional PoPs, the access PoPs provide execution environments for hosting workloads with real-time requirements, e.g., virtualized DU instances (vDUs). In this regard, commodity hardware is no longer valid; they need to be equipped with advanced, rich-featured CPU architectures (e.g., Intel Xeon) and hardware acceleration solutions (e.g., FPGA, structured ASICs, etc.) instead [27].

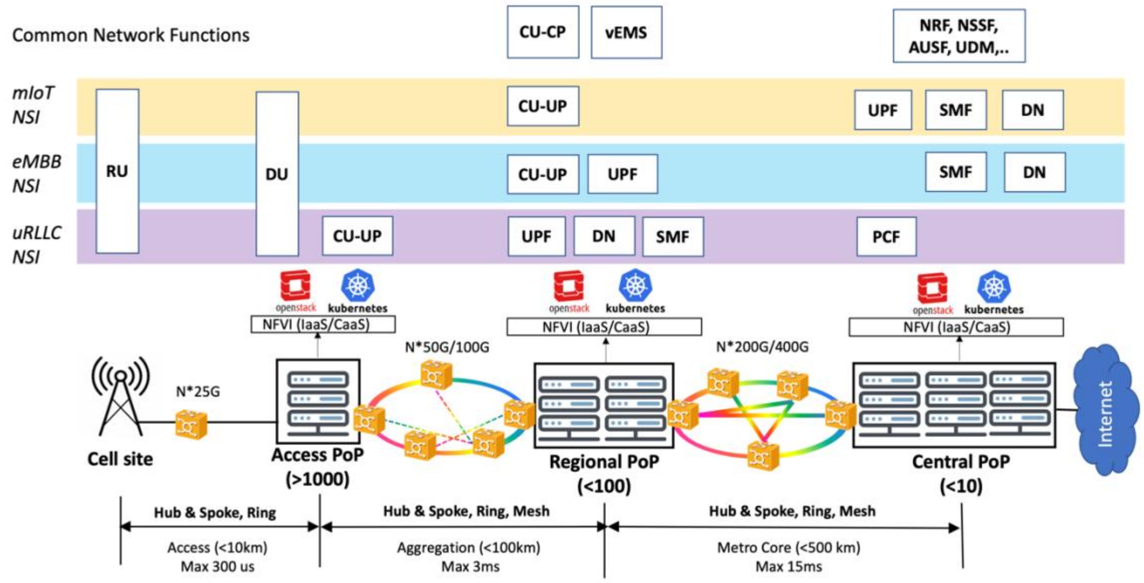


Figure H-5. Network slicing system architecture.

3 Impact of Network Slicing

The introduction of slicing will impact all the technical domains of the network. In this section, we review the features required on these domains to support slicing. These constitute the basis for the understanding of the different solutions that will be explained later, in Sections 5-8.

3.1 CN Slicing

The impact of network slicing in the CN domain can be summarized into three main topics: slice identity management, slice-aware device connectivity and the allocation of separate 5GC functions.

3.1.1 Slice Identity Management

The network slicing feature was first introduced in Release 15, with the ability of the 5GC to support multiple network slices and differentiate among them. This differentiation is done using two signaling identifiers: the Single Network Slice Selection Assistance Information (S-NSSAI) and the Network Slice Selection Assistance Information (NSSAI).

The S-NSSAI identifies a network slice across the UE, RAN and the 5GC. It is a 32-bit parameter comprised of two fields:

- A Slice/Service Type (SST): mandatory 8-bit field that refers to the expected network slice behavior in terms of features and supported services. The SST field may have standardized and operator-specific (non-standardized) values. The standardized SST range [10] includes values from 0 to 127, while values 128 to 255 belong to the operator specific range. For now, the following SST values have become normative: SST = 1 (enhanced Mobile Broadband), eMBB), SST = 2 (Ultra Reliable Low Latency Communication, uRLLC), SST = 3 (massive IoT, mIoT), SST = 4 (Vehicle to Everything, V2X) and SST=5 (High-Performance Machine-Type Communications, HMTC).
- A Slice Differentiator (SD): optional 24-bit field that allows the operator to differentiate among multiple network slices with the same SST. This differentiation can be in terms of slice features (e.g., mobile vs fixed-wireless access services, charging), customer information (tenancy) and slice priority.

An NSSAI is a collection of S-NSSAIs sent by the device to assist the network in selecting a particular network slice for this UE. Within the Public Land Mobile Network (PLMN), the NSSAI is managed at the Tracking Area level in the RAN, and at the Registration Area level in the 5GC. Different types of NSSAIs exist, including Configured NSSAI (NSSAI provisioned in the device), Subscribed NSSAI (NSSAI stored in the UDM), Requested NSSAI (provided by the UE to the serving PLMN during registration) and Allowed NSSAI (provided by the serving PLMN to the UE during registration) [10]. The 5GC uses the Requested NSSAI for slice selection and validation, and returns the Allowed NSSAI. The Allowed NSSAI indicates the S-NSSAI values that the UE can use in the serving PLMN for the current Registration Area.

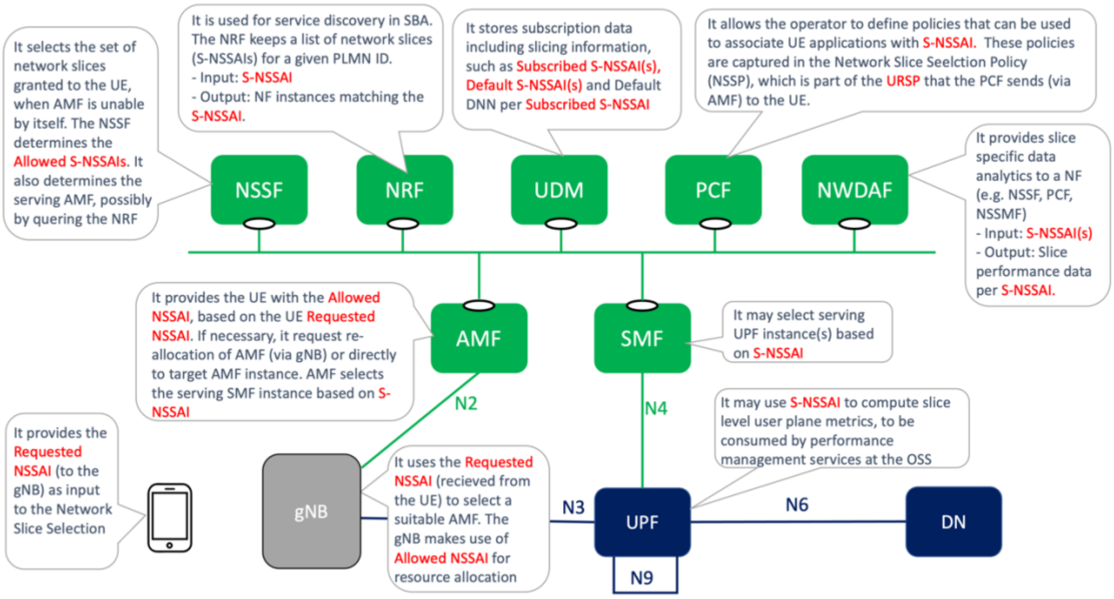

It selects the set of network slices granted to the UE, when AMF is unable by itself. The NSSF determines the Allowed S-NSSAIs. It also determines the serving AMF, possibly by quering the NRF
It is used for service discovery in SBA. The NRF keeps a list of network slices (S-NSSAIs) for a given PLMN ID.
- Input: S-NSSAI
- Output: NF instances matching the S-NSSAI.
It stores subscription data including slicing information, such as Subscribed S-NSSAI(s), Default S-NSSAI(s) and Default DNN per Subscribed S-NSSAI
It allows the operator to define policies that can be used to associate UE applications with S-NSSAI. These policies are captured in the Network Slice Seelction Policy (NSSP), which is part of the URSP that the PCF sends (via AMF) to the UE.
It provides slice specific data analytics to a NF (e.g. NSSF, PCF, NSSMF)
- Input: S-NSSAI(s)
- Output: Slice performance data per S-NSSAI.
NSSF
NRF
UDM
PCF
NWDAF
It provides the UE with the Allowed NSSAI, based on the UE Requested NSSAI. If necessary, it request re-allocation of AMF (via gNB) or directly to target AMF instance. AMF selects the serving SMF instance based on S-NSSAI
AMF
SMF
It may select serving UPF instance(s) based on S-NSSAI
N2
N4
It may use S-NSSAI to compute slice level user plane metrics, to be consumed by performance management services at the OSS
It provides the Requested NSSAI (to the gNB) as input to the Network Slice Selection
gNB
It uses the Requested NSSAI (recieved from the UE) to select a suitable AMF. The gNB makes use of Allowed NSSAI for resource allocation
N3
UPF
N6
DN
N9

Figure H-6. Impact of slicing in 5GC.

Figure H-6 summarizes the use of S-NSSAI and NSSAI artifacts in the 5G network.

3.1.2 Slice-Aware Device Connectivity

According to 3GPP specifications, the Allowed NSSAI can include a maximum of eight S-NSSAI values [10]. This means that a device can establish Packet Data Unit (PDU) sessions with up to eight slices at the same time.

The device can have different client applications (e.g., internet browsing applications, enterprise applications, XR applications), each requiring the connection to a different slice. To make this possible, the device needs to be made slicing aware, something that is achieved with the introduction of the UE Resource Selection Policy (URSP) [28]. The URSP is a network slicing feature enabled by the Policy Control Function (PCF), which informs the network slice status to the UE via the Access and Mobility management Function (AMF). It is composed of a number of URSP rules that map application information (e.g., client application ID, device Operation System ID, IP descriptors) with network slice information (e.g., S-NSSAI, Session and Service Continuity, Data Network Name). The device uses the URSP to determine which PDU session shall be chosen for a particular application based on URSP rules. For further information on the URSP and its use for slicing support at the device side, please see [29].

3.1.3 5GC Network Functions

As commented in Section 2.2, the allocation of dedicated network functions on a network slice allows it to be tailored to the specific needs of hosted service(s). Where there is more potential to make this customization is on the 5GC side.

It is not the goal of this subsection to discuss which 5GC functions are to be dedicated to a slice; indeed, as we will see in Section 5.1, this entirely depends on the business requirements of individual customers. The purpose of this section is instead to outline the importance of some 5GC functions when building up CN slice subnets. In this regard, the UPF is the most valuable network function to be dedicated, followed by control plane network functions (SBA).

The importance of having a dedicated UPF comes from two important reasons: (i) a tailored user plane QoS, and (ii) an improved availability and reliability. The first point refers to the ability to allocate an UPF with required resource capacity where needed, e.g., close to customer premises to ensure low latency. The second point means that having a dedicated UPF instance allows optimal redundancy level to be achieved, and the risk of service interruption for the slice to be reduced, ensuring that established sessions can survive for a period of time, even when the connection to the control plane functions is lost.

The control plane functions are the second-most valuable assets to dedicate. For example, with a dedicated Session Management Function (SMF) [10], it is possible to make changes to established sessions and establish new sessions for a period of time, even if the connection to UPFs is lost.

3.2 RAN Slicing

In the NG-RAN, the introduction of slicing has an impact on three main aspects: gNB configuration, mobility support and Radio Resource Management (RRM) procedures.

3.2.1 gNB Configuration

A gNB can be configured to support multiple slices. This configuration, done via the NSSMF, is based on the following principles:

a) Network slices are defined within a PLMN. In RAN sharing scenarios, where multiple PLMNs share the same cell, each operator needs to link S-NSSAIs with the PLMN ID.

b) The gNB serves a cell. The cell belongs to a tracking area, which is identified with two artifacts: Tracking Area Code (TAC), i.e., local identifier, and Tracking Area Identifier (TAI), i.e., universal identifier. The TAI is a {PLMN ID, TAC} tuple, and it is relevant in RAN sharing scenarios. To indicate the tracking area to which the cell belongs to, the gNB broadcasts one or more TAIs, i.e., one TAI per hosted PLMN ID [30].

c) A network slice is linked to a tracking area. This is because S-NSSAIs are managed per tracking area [31].

Based on the above principles, it can be noticed that all cells belonging to the same tracking areas must serve the same set of network slices. Once the gNB is set with supported slices (per TAI), its mission is to map traffic from individual PDU sessions into appropriate NG-RAN resources. This is done by associating the tuple {S-NSSAI, PLMN ID} with one Dedicated Radio Bearer (DRB). The profile of the DRB is configured with RRM parameters which are tailored to the service requirements of the slice.

3.2.2 Mobility Support

When the device moves from one cell to another, a handover procedure is triggered. The handover request (from the source gNB to the target gNB) includes the network slices assigned to the UE, specifying the tuple {S-NSSAI, PLMN ID} for each active PDU session. According to the principles listed earlier, it is clear that handover requests between gNBs from the same tracking area are always successful. However, in the case of mobility outside a tracking area, it might happen that one or more PDU sessions could not be transferred, because the associated S-NSSAI are not available in the target cell. In traditional radio control admission solutions, where handover acceptance is subject to the admission of all radio bearers, this scenario would result in an automatic handover rejection. To solve this all-or-nothing approach, partial admission control mechanisms are being developed. These mechanisms allow the admission of those PDU sessions whose associated S-NSSAIs are supported in the target gNB. The logic is as follows:

- The target gNB will send handover request ACK with Admitted PDU session and Not Admitted PDU session.
- If all the S-NSSAIs in the handover request are not admitted, the handover will be rejected.

Figure H-7 shows an example of mobility support using this partial admission control for handover management.

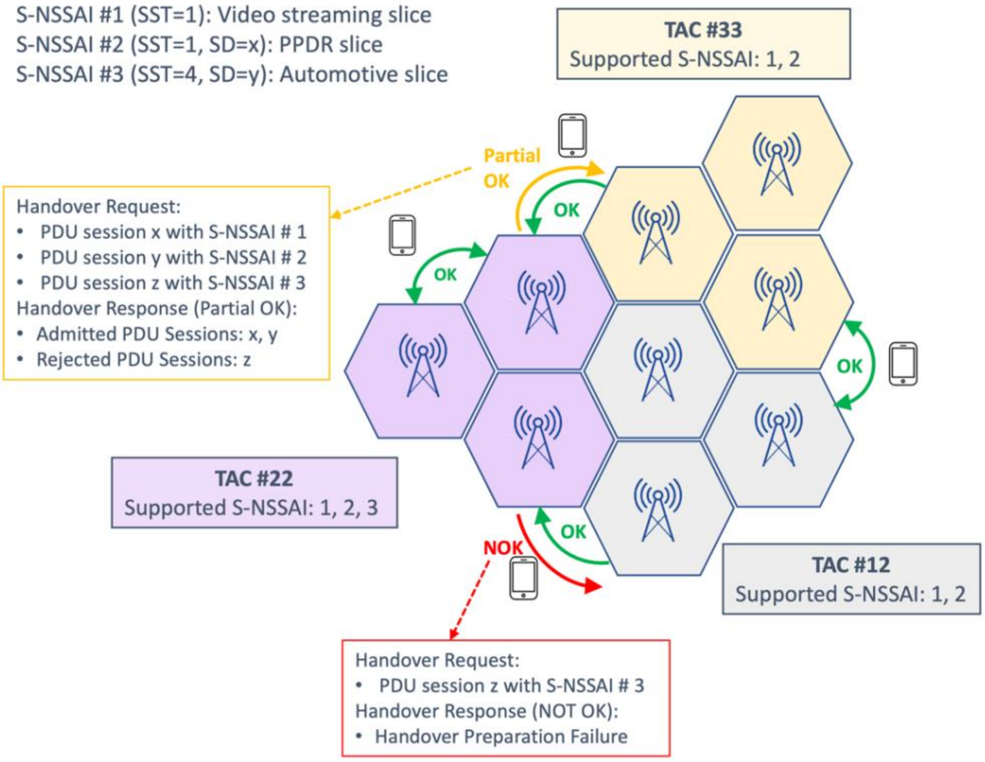

S-NSSAI #1 (SST=1): Video streaming slice
S-NSSAI #2 (SST=1, SD=x): PPDR slice
S-NSSAI #3 (SST=4, SD=y): Automotive slice
TAC #33
Supported S-NSSAI: 1, 2
Partial OK
OK
OK
OK
Handover Request:
• PDU session x with S-NSSAI # 1
• PDU session y with S-NSSAI # 2
• PDU session z with S-NSSAI # 3
Handover Response (Partial OK):
• Admitted PDU Sessions: x, y
• Rejected PDU Sessions: z
TAC #22
Supported S-NSSAI: 1, 2, 3
NOK
OK
TAC #12
Supported S-NSSAI: 1, 2
Handover Request:
• PDU session z with S-NSSAI # 3
Handover Response (NOT OK):
• Handover Preparation Failure

Figure H-7. Slice-aware mobility

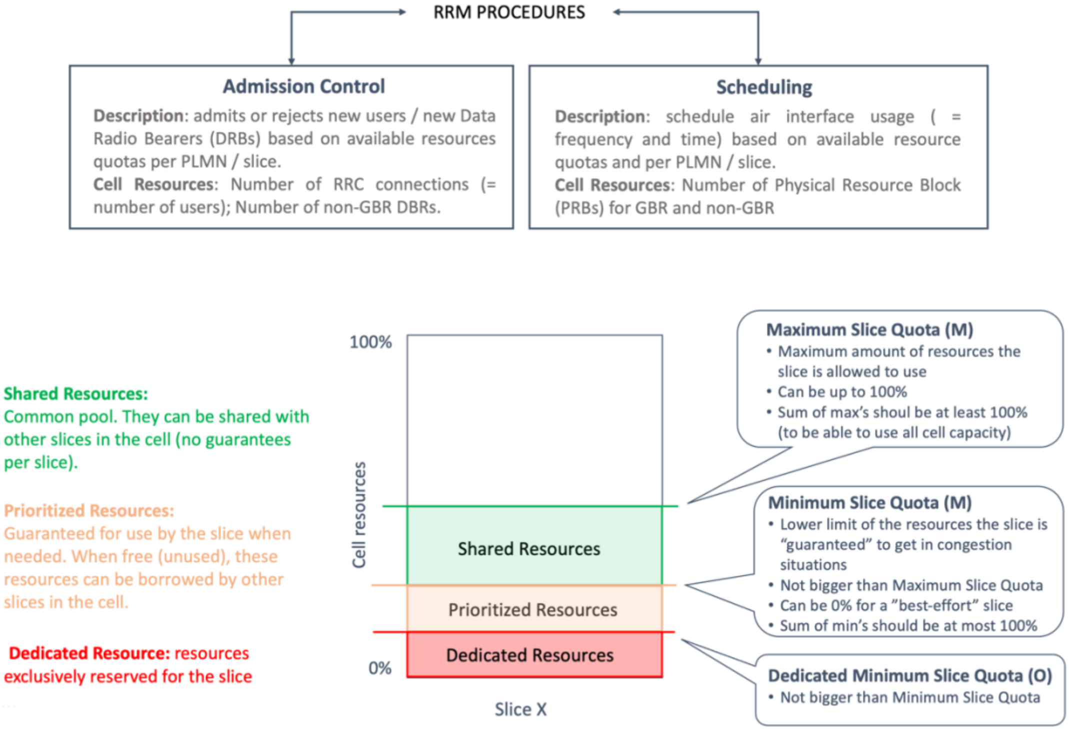
3.2.3 RRM Procedures

The gNB includes a set of RRM procedures that govern the allocation of NR cell resources across existing slices, in such a way that shortage of resources in one slice does not break the SLA of another slice. There are two fundamental RRM procedures: admission control and scheduling.

The task of admission control is to admit or reject the establishment requests for new radio bearers. Admission control can be based on number of users (RRC connections) or number of DRBs. The first option allows limiting the number of UEs accessing a specific slice based on SLA requirements. The second option is based on reserving enough DRBs for each slice, according to their estimated data volume.

The scheduling allows the gNB to dispatch available Physical Radio Blocks (PRBs), i.e., frequency-time resource grids, across the different slices, in such a way that the QoS requirements associated with their PDU sessions can be fulfilled. These requirements are expressed with the 3GPP 5G Quality Indicator (5QI) [10].

At a very high load, admission control provides the scheduler with sufficient resources to secure QoS of all the admitted users. To that end, it is very important to design an efficient admission control algorithm that can take into account the overall resource situation, the priorities of users based on their category level, the QoS of the in-progress request and the QoS requirements of the new radio bearer requests.



Admission Control — **Description**: admits or rejects new users / new Data Radio Bearers (DRBs) based on available resources quotas per PLMN / slice. **Cell Resources**: Number of RRC connections (= number of users); Number of non-GBR DBRs.

Scheduling — **Description**: schedule air interface usage (= frequency and time) based on available resource quotas and per PLMN / slice. **Cell Resources**: Number of Physical Resource Block (PRBs) for GBR and non-GBR

Shared Resources: Common pool. They can be shared with other slices in the cell (no guarantees per slice).

Prioritized Resources: Guaranteed for use by the slice when needed. When free (unused), these resources can be borrowed by other slices in the cell.

Dedicated Resource: resources exclusively reserved for the slice

Maximum Slice Quota (M)
- Maximum amount of resources the slice is allowed to use
- Can be up to 100%
- Sum of max's shoul be at least 100% (to be able to use all cell capacity)

Minimum Slice Quota (M)
- Lower limit of the resources the slice is "guaranteed" to get in congestion situations
- Not bigger than Maximum Slice Quota
- Can be 0% for a "best-effort" slice
- Sum of min's should be at most 100%

Dedicated Minimum Slice Quota (O)
- Not bigger than Minimum Slice Quota

Figure H-8. RRM procedures for slicing.

The operation in the admission control and scheduling procedures is, in both cases, based on the configuration of the following per slice quotas: dedicated minimum slice quota (optional), minimum slice quota (mandatory) and maximum slice quota (mandatory). Figure H-8 provides a summary of these three quotas. These quotas need to be specified for each RRM procedure, since the managed NR cell resources are different:

- For admission control, the NR cell resources correspond to either RRC connections (option 1) or DRBs (option 2). For option 1, it is the CU-CP which configures the quotas. For option 2, it is the CU-UP.
- When scheduling, the NR cell resources correspond to PRBs, based on which per slice quotas are defined.

In this work, we will focus on scheduling aspects. Table H-2 shows different examples on how to configure the slice quotas for scheduling. As seen, depending on the values set for these quotas, the slice can be profiled into different categories.

Table H-2. RAN slice characterization based on configured quotas.

Ded Min Slice Quota	Minimum Slice Quota	Maximum Slice Quota	RAN Slice Characterization
10%	10%	45%	Dedicated slice-profile 1 (dedicated + shared)
10%	20%	45%	Dedicated slice-profile 2 (dedicated + prioritized + shared)
-	20%	45%	Prioritized slice (prioritized + shared)
-	0%	45%	Best effort slice (can also use prioritized resources if spare left by another prioritized slice)

3.3 TN Slicing

Unlike the NG-RAN and 5GC, the TN domain is out of the scope of the 3GPP network slice concept. 3GPP provides slicing solutions for the RAN and CN domain, but not for the TN. However, to maintain consistency on the slice established between the device and the service application, there is a need to map 3GPP slice criteria into appropriate transport capabilities offered in the fronthaul, midhaul, backhaul and DN segments. This is not trivial.

On the one hand, there is the need to configure WAN resources in such a way that the requirements captured in the ServiceProfile and SliceProfile can be fulfilled in the TN substrate. This requires translating network function layer requirements associated with S-NSSAI information (e.g., maximum delay budget, data rates, availability, mobility speed, usage density) into transport network characteristics that include bandwidth, latency and criteria such as traffic prioritization, directionality, protection, and disjoint routes. This translation is done at provisioning time.

On the other hand, there exists a wide availability of transport technologies in carrier networks. These technologies provide multi-layer connectivity services using different topologies (e.g., hub-and-spoke, ring, point-to-point, point-to-multipoint). Though they do not support slicing natively, these technologies are able to mimic slicing behavior, if configured (and combined) properly.

In this section, we focus on the main enablers for these two open questions.

3.3.1 On the Mapping of 3GPP Slice Information into TN Nodes

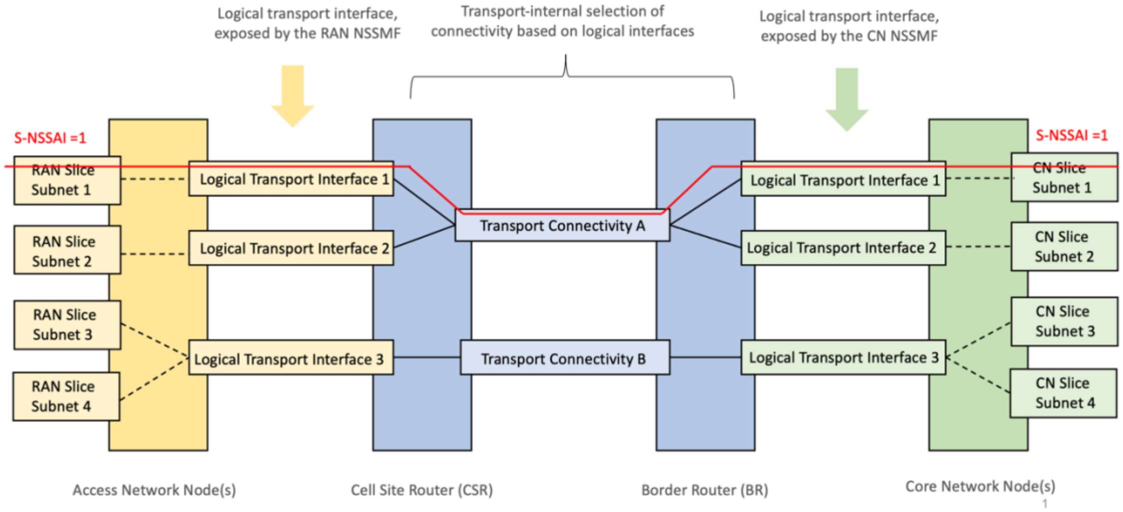
To configure the TN slicing behavior in the WAN resources, the TN management domain needs the following information:

- Network slice topology. The TN management domain needs to know the application endpoints of the slice to determine the needed WAN resources, which are either physical or virtual nodes. NSMF/NSSMFs provide the application endpoints [32] of 3GPP network functions taking part in the RAN and CN slice subnets and, if applicable, further information such as the next-hop router IP address configured in these network slice subnets. For example, the CU-UP application endpoints are the IP addresses/VLAN IDs associated with the F1 and N3 interfaces. The TN management domain correlates this information with the transport network topology and derives the (cell site or border) routers connecting to network function.
- Traffic segregation and mapping to S-NSSAIs. As 3GPP network functions can be shared by multiple network slices, it is necessary to segregate traffic belonging to specific slices on transport interfaces. One option for traffic segregation is to assign application endpoints to a specific set of S-NSSAI values. This solution is rather simple, as the TN can map packets to connectivity services based on application endpoints, provided that (i) the allocation of S-NSSAI to endpoints is known, and (ii) the application endpoints are visible on the transport layer. While this is the simplest solution in many cases, it is not a universal solution, as the application endpoint addresses are not always visible to the site router, e.g., when there is encryption using IPSec. An alternative solution is the concept of logical transport interfaces, as shown in Figure H-9.A. A logical transport interface is a virtual interface separated from application endpoints. It can be, for example, a specific IP address/VLAN combination corresponding to an IPSec termination point, or an identifier (e.g.,

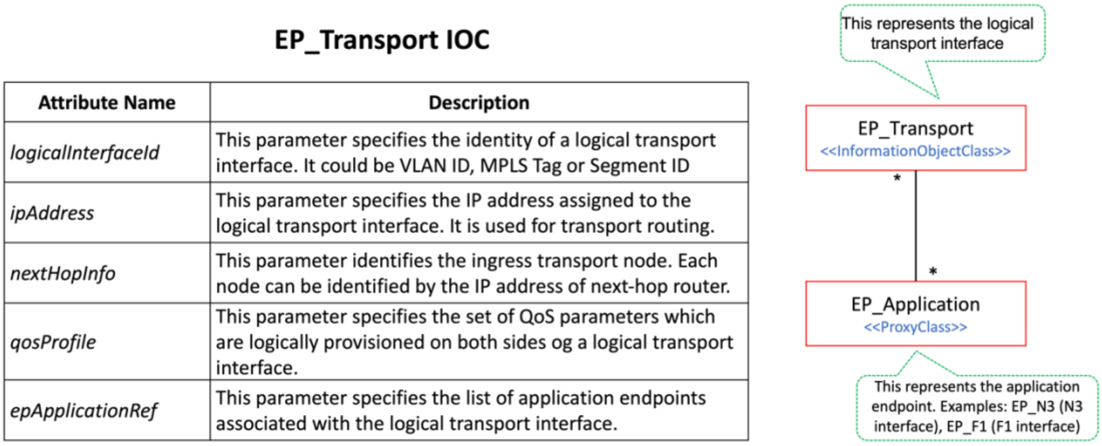
MPLS label, segment ID) that the TN recognizes, or it can be just a logical interface defined on top of a physical transport interface. As long as the interface identity can be derived from packet headers, the TN nodes can perform the mapping to transport connectivity services.

- Reachability information. Each logical transport interface carries the traffic associated with some application endpoints that may be using IP addresses separate from the transport interface. These IP addresses must be reachable; hence they need to be advertised to populate forwarding tables. A 3GPP network function can advertise such reachability information by running a dynamic routing protocol towards the next hop route.
- QoS requirements. To satisfy the service requirements captured in ServiceProfile and SliceProfile, each logical transport interface needs to be bound to a QoS profile that includes the applicability and use of DiffServ Code Points (DSCP) [33] and QoS related properties on that interface.

To allow the TN management domain to receive this information from the 3GPP management system, the EP_Transport IOC [15] is defined. Part of the Network Slice NRM fragment (see Figure H-1), this class allows the capture of the information that shall be exchanged between the 3GPP management system (E2E management domain, RAN management domain and CN management domain) and the TN management domain. This information is used to configure WAN resources in such a way that the requirements captured in ServiceProfile and SliceProfile can be fulfilled. Figure H-9.B shows the construction of the EP_Transport IOC, and how it maps the logical transport interface to application endpoints. Notice that one EP_Transport (representing a logical transport interface) can be associated with more than one multiple EP_Application (representing an application endpoint of a 3GPP network function), but also the other way around. While the first case captures the typical situation, the second case can be used for the sake of resilience or load balance in the TN. For example, in Figure H-9.A, instead of configuring multiple nextHops for one EP_Transport to allow multiple optional “links” between the gNB port and the cell site router, the solution adopted is as follows: to configure one nextHop for each EP_Transport, but have more than one EP_Transport for an EP_N3 to achieve similar load balance or resilience goal.


Logical transport interface, exposed by the RAN NSSMF
Transport-internal selection of connectivity based on logical interfaces
Logical transport interface, exposed by the CN NSSMF
S-NSSAI =1
RAN Slice Subnet 1
RAN Slice Subnet 2
RAN Slice Subnet 3
RAN Slice Subnet 4
Logical Transport Interface 1
Logical Transport Interface 2
Logical Transport Interface 3
Transport Connectivity A
Transport Connectivity B
Logical Transport Interface 1
Logical Transport Interface 2
Logical Transport Interface 3
S-NSSAI =1
CN Slice Subnet 1
CN Slice Subnet 2
CN Slice Subnet 3
CN Slice Subnet 4
Access Network Node(s)
Cell Site Router (CSR)
Border Router (BR)
Core Network Node(s)

(A)



EP_Transport IOC

Attribute Name	Description
logicalInterfaceId	This parameter specifies the identity of a logical transport interface. It could be VLAN ID, MPLS Tag or Segment ID
ipAddress	This parameter specifies the IP address assigned to the logical transport interface. It is used for transport routing.
nextHopInfo	This parameter identifies the ingress transport node. Each node can be identified by the IP address of next-hop router.
qosProfile	This parameter specifies the set of QoS parameters which are logically provisioned on both sides og a logical transport interface.
epApplicationRef	This parameter specifies the list of application endpoints associated with the logical transport interface.

This represents the logical transport interface
EP_Transport
<<InformationObjectClass>>
*
*
EP_Application
<<ProxyClass>>
This represents the application endpoint. Examples: EP_N3 (N3 interface), EP_F1 (F1 interface)

(B)

Figure H-9. On the use of logical transport interfaces for TN slicing support. (A) Traffic segregation and mapping to S-NSSAIs; the BR is also referred to as Provider Edge (PE) router. (B) Schema for the EP_Transport IOC.

3.3.2 Transport Technologies

To convey slice traffic in the transport network, multiple forwarding plane technologies can be used. Depending on their isolation capabilities, these technologies can be clustered into two main categories: soft slicing (the traffic loading on one slide may degrade the performance of other slices, because of the use of statistical multiplexing and service classes) and hard slicing (the traffic loading on one slice has no impact on the traffic from any other slice, including QoS effects).

The soft slicing category uses packet-based technologies to provide traffic-engineered and traffic-managed isolation of resources. This category encompasses Layer 2 and Layer 3 technologies, including tunnelling (e.g., VxLAN, MPLS) and virtualization (e.g., VPN, VLAN) based technologies. The mutual impact of QoS of slices sharing the same infrastructure resources may be mitigated by traffic engineering including, for example, limiting the statistical multiplexing ratio, or traffic policing on each network slice.

Hard slicing can be guaranteed through independent circuit switched connections (e.g., dedicated wavelength, dedicated TDM time slot) for the exclusive use of a single network slice. Unlike soft slicing, this category is thought of as being implemented at Layer 1, using techniques much more tightly coupled to the hardware itself, such as optical transport network switching [34] or novel Ethernet-based solutions like Flex-Ethernet [35][36].

Depending on the customer requirements, the operator may go for soft slicing or hard slicing, or a mixture of the two. Indeed, it is possible to combine them as shown in Figure H-10, with hard slicing ensuring a dedicated capacity chunk for the customer, and soft slicing providing traffic segregation among the services belonging to this customer. This approach preserves a cost-efficient solution to the customer that has multiple services, and wants them not to be impacted with traffic congestion or faults issued by services from other customers.

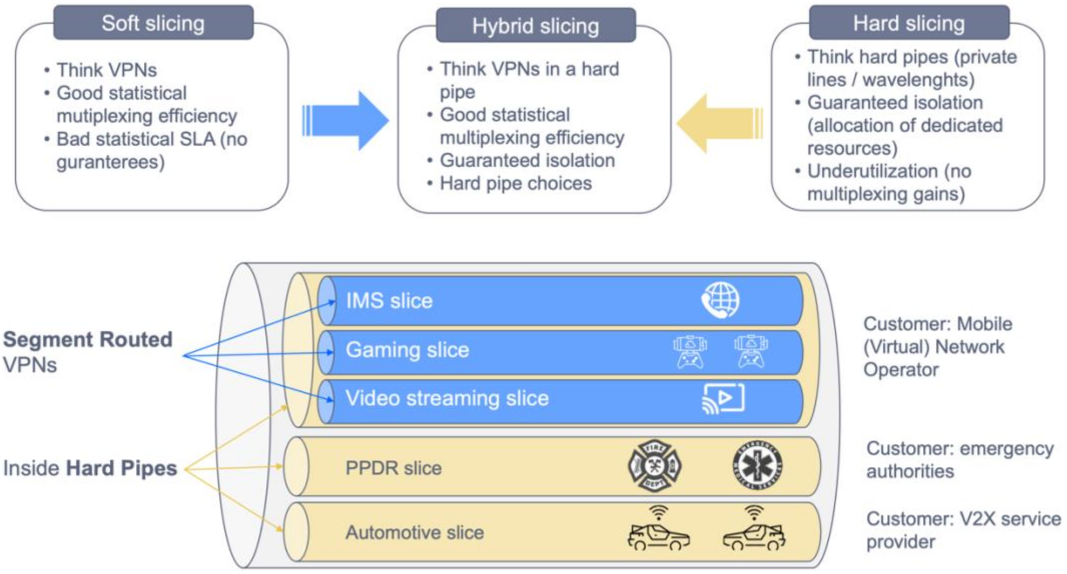

Soft slicing
• Think VPNs
• Good statistical mutiplexing efficiency
• Bad statistical SLA (no guranterees)
Hybrid slicing
• Think VPNs in a hard pipe
• Good statistical multiplexing efficiency
• Guaranteed isolation
• Hard pipe choices
Hard slicing
• Think hard pipes (private lines / wavelenghts)
• Guaranteed isolation (allocation of dedicated resources)
• Underutilization (no multiplexing gains)
Segment Routed VPNs
Inside Hard Pipes
IMS slice
Gaming slice
Video streaming slice
PPDR slice
Automotive slice
Customer: Mobile (Virtual) Network Operator
Customer: emergency authorities
Customer: V2X service provider

Figure H-10. Soft, hard and hybrid slicing in transport networks.

4 Network Slicing Technology Radar

Network slicing is an E2E solution, covering the three technology domains: RAN, CN and TN. The provisioning and operation of the different slices, and the lifecycle management of hosted services, is done with a cloud-native OSS stack composed of a number of management domains, including vertical domains (e.g., RAN, CN, TN management domains) and horizontal domains (e.g., MANO and E2E management domains). The maturity level of slicing varies across all these technology and management domains. This fact, together with the integration complexity of carrier networks (e.g., brownfield facilities, multi-vendor solutions), defines challenges that operators need to work out to enable the full promise of network slicing in a later phase.

Although standards and technologies do not currently support full slicing capabilities, it is important to get started with early-stage slicing and to use a vision of what is desired in the longer term to guide progress and focus. For operators, it is important to take a phased-based approach, establishing a process to incorporate learnings at each stage of the slicing journey.

The mission of this section (and upcoming ones) is to present a network slicing technology radar that can help operators to build their own journey towards the commercialization and monetization of slicing. As shown in Figure H-11, this radar captures a list of solutions for network slicing impacting all relevant operator’s sub-systems, including RAN, TN, CN and OSS. This is complemented by an assessment work, called ring assignment. In particular, we use four rings with the following semantics:

- **As-is ring**: represents solutions that are available in today’s carrier networks. These solutions are typically associated with technologies that operators have high confidence in, with low risk and recommended to be available across the entire service footprint. In terms of 5G roll-out strategy, this corresponds to 5G NSA (Non-Standalone) [37].
- **Deploy ring**: covers the slicing solutions that can be applied in early 5G SA (Standalone) networks, based on 3GPP Release 15 standards. Some operators have already started to activate their SA networks, while some others expect to get them

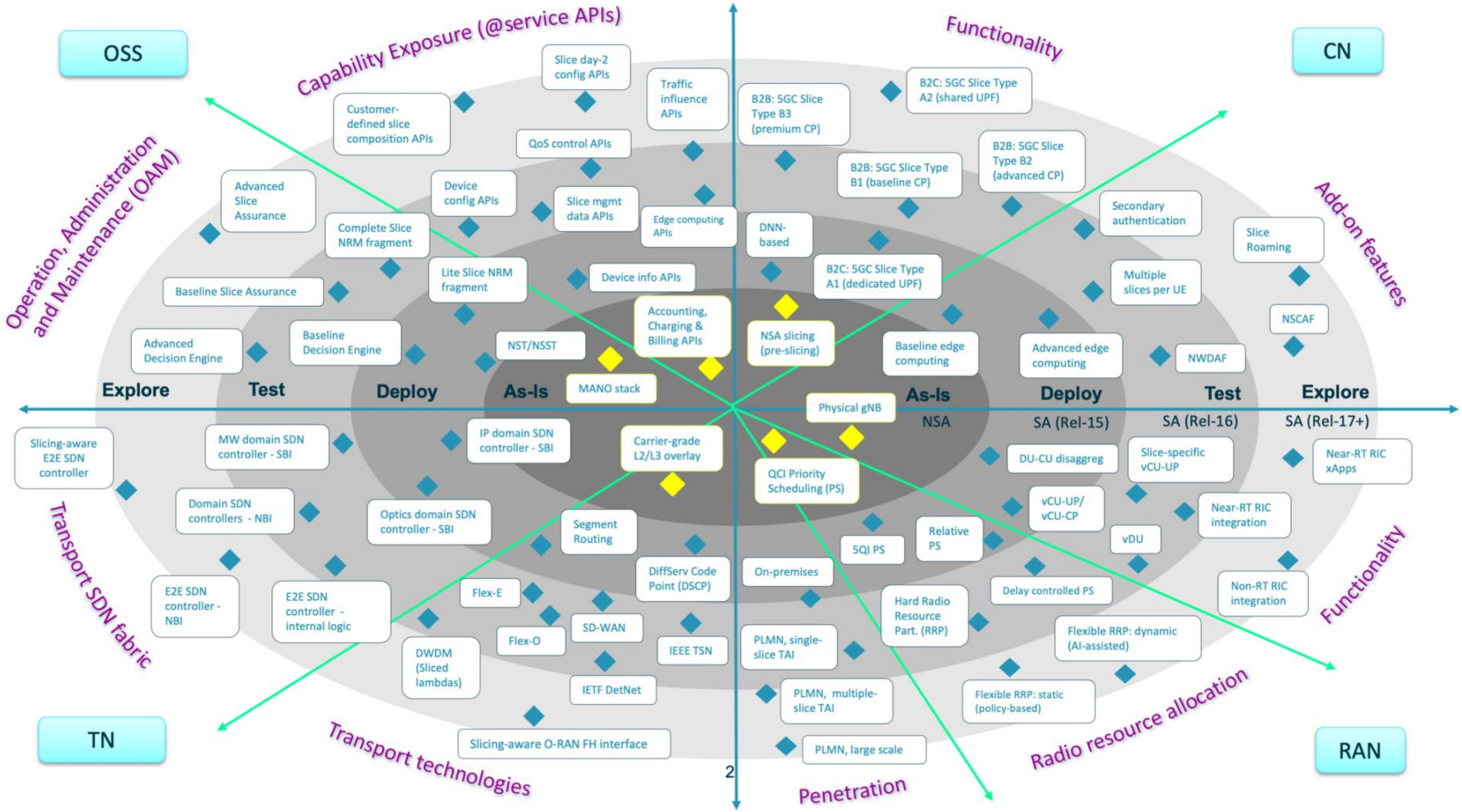


Figure H-11. Network slicing technology radar.

operationally ready within the next year. With this timing in mind, we can say that this ring captures proven slicing solutions that operators may integrate in the short-term.

- **Test ring**: captures slicing solutions that are much more focused on satisfying requirements from uRLLC and mIoT services. Associated with brand new Rel-16 features, these solutions have great potential but are unproven in production networks, hence it is worth operators investing in prototyping efforts in order to evaluate their performance and impact. This evaluation is typically done with commercial trials, either bilateral or multi-vendor, and different Proof of Concepts (PoCs). The upgrade towards Rel-16 is expected within the next 2–3 years; this means that test ring represents slicing solutions that might be available in the medium term.
- **Explore ring**: includes slicing solutions that are foreseen in the long run, starting in the next 4–5 years. These solutions, tied to features from 3GPP Rel-17 on wards, promise to provide great potential, though their impact and commercial availability is still far from crystal clear. The role of the operator is to keep track of their evolution through exploratory activities such as the ones done in research and innovation projects, e.g., 5G-VINNI [38], 5GROWTH [39] and 5G-CLARITY [40].

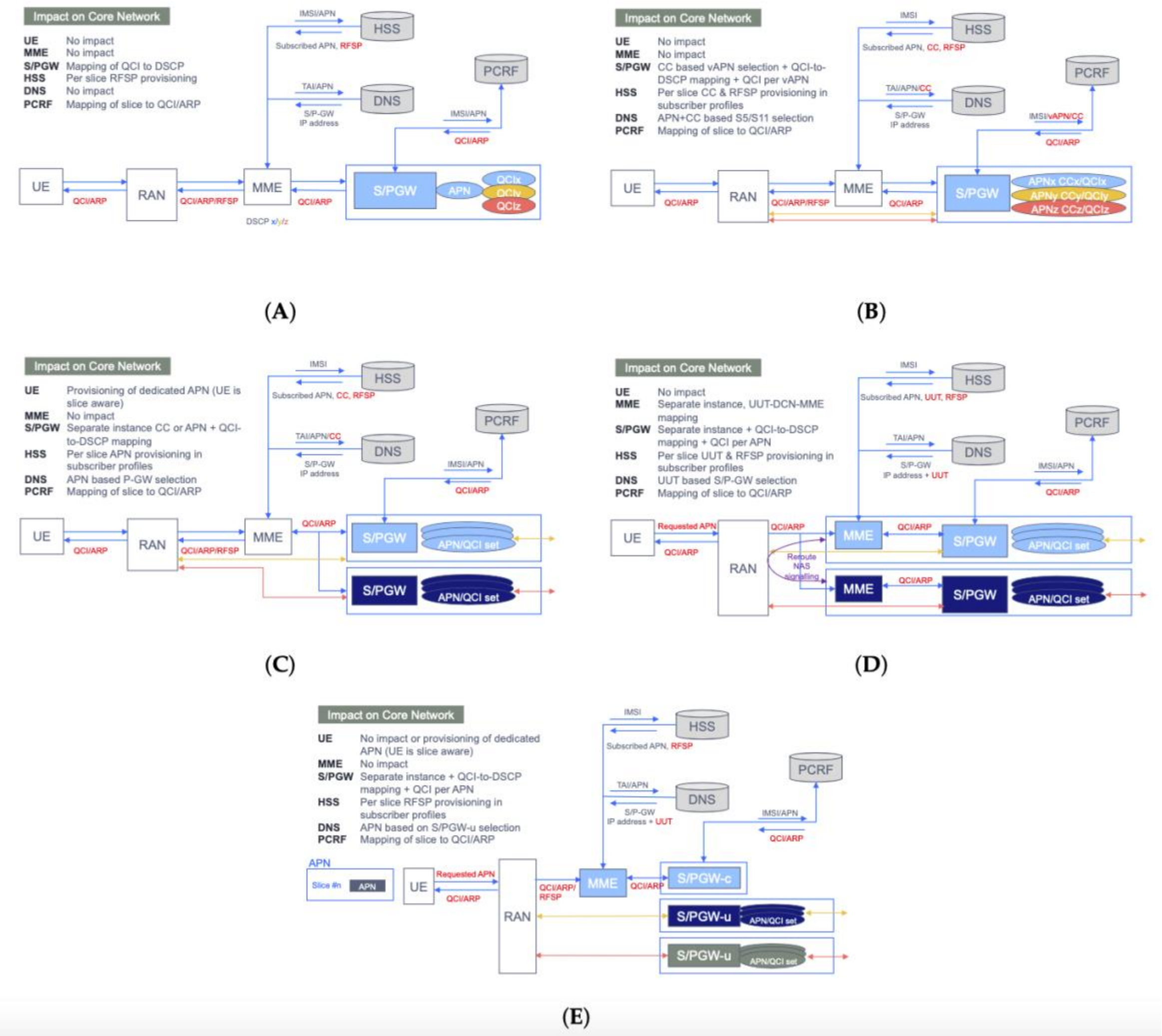
As outlined in Section 1, for the position of each solution into this radar, three criteria have been considered: (i) the technological maturity of the solution, which is subjected to the readiness of the standards; (ii) the roadmap of commercial products, which specifies when the features associated with the solution will be available; and (iii) the relevance for the customers, which determine the prioritization of the solution over others. The following sections provide details on these solutions, across the involved subsystems: CN (Section 5), RAN (Section 6), TN (Section 7) and OSS (Section 8).

5 CN Domain

In this domain, the technology radar captures information from two different dimensions: functionality and add-on features. The *functionality* dimension deals with the discussion on how to use CN functions for the construction of different network slices, scouting different deployment options for these slices depending on the isolation and business requirements of hosted services. On the other hand, the *add-on features* dimension refers to the set of value-added solutions that complement and extend baseline slice functionality. The network operator can optionally make use of these solutions to either (i) provision enriched services to the customer, i.e., new revenue streams; or (ii) streamline internal network operation, i.e., OPEX savings.

5.1 Functionality

The first 5G commercial networks available worldwide are based on NSA. The reason is that most communication service providers are looking to deliver mainly high-speed connectivity to consumers already with 5G-enabled devices today. For these providers, the NSA mode makes the most sense, because it allows them to leverage their existing packet core assets (EPC) rather than deploy a completely new 5G network. In this mode, where 5GC does not exist, there is no native slicing support, as outlined in Section 3.1. However, this lack of slicing enablers (e.g., no S-NSSAI support) does not mean the operators are unable to provide service differentiation and traffic segregation at the core side; in fact, there are a number of solutions that allow the EPC to enforce some level of traffic separation for overload mitigation when having multiple services. Figure H-12 depicts a functional description of some of these solutions, tagged with **NSA slicing (pre-slicing)** wording. As seen, the capabilities across these solutions are quite different, ranging from basic QoS differentiation (e.g., QCI based common APN-S/PGW) to a complete packet core separation (e.g., DECOR [41]), with a number of variants in between, some of them subjected to technology availability. For example, NFV technology is a must for the implementation of the control user plane separation solution (Figure H-12E).



Impact on Core Network

UE	No impact
MME	No impact
S/PGW	Mapping of QCI to DSCP
HSS	Per slice RFSP provisioning
DNS	No impact
PCRF	Mapping of slice to QCI/ARP

Impact on Core Network

UE	No impact
MME	No impact
S/PGW	CC based vAPN selection + QCI-to-DSCP mapping + QCI per vAPN
HSS	Per slice CC & RFSP provisioning in subscriber profiles
DNS	APN+CC based S5/S11 selection
PCRF	Mapping of slice to QCI/ARP

Impact on Core Network

UE	Provisioning of dedicated APN (UE is slice aware)
MME	No impact
S/PGW	Separate instance CC or APN + QCI-to-DSCP mapping
HSS	Per slice APN provisioning in subscriber profiles
DNS	APN based P-GW selection
PCRF	Mapping of slice to QCI/ARP

Impact on Core Network

UE	No impact
MME	Separate instance, UUT-DCN-MME mapping
S/PGW	Separate instance + QCI-to-DSCP mapping + QCI per APN
HSS	Per slice UUT & RFSP provisioning in subscriber profiles
DNS	UUT based S/P-GW selection
PCRF	Mapping of slice to QCI/ARP

Impact on Core Network

UE	No impact or provisioning of dedicated APN (UE is slice aware)
MME	No impact
S/PGW	Separate instance + QCI-to-DSCP mapping + QCI per APN
HSS	Per slice RFSP provisioning in subscriber profiles
DNS	APN based on S/PGW-u selection
PCRF	Mapping of slice to QCI/ARP

Figure H-12. Different deployment options for NSA slicing. (A) QCI based common APN-S/PGW. (B) Virtual APN/QCI based on Charging Characteristics (CC). (C) Dedicated S/PGW. (D) DECOR. (E) Control User Plane Separation.

As communication service providers set their sights on new revenue streams from groundbreaking 5G services (i.e., uRLLC, mIoT, V2X services), they realize the need to migrate to the SA mode, which represents the target 5G system architecture [42]. With the first commercial 5GC solution suites already available in the market, operators can bring native slicing functionality into their carrier-class facilities. Despite being Rel-15 complaint, these first solution suites typically provide very limited capabilities in relation to S-NSSAI support; in fact, many of them offer single-slice support, with their AMF/SMF implementations only able to deal with one S-NSSAI at a time. The fact that one single slice can be configured in the 5GC prevents the operator from using separate slices to achieve service/customer traffic segregation. In these circumstances, non-NSSAI assisted solutions shall be used instead. One example is the provisioning of separate Data Network Name (DNN) [10] for different services/customers. This **DNN-based solution** is equivalent to the pre-slicing solution shown in Figure H-12B, where virtual APN/QCI (EPC artifacts) are now replaced with DNN/5QI (5GC artifacts).

With the ever-increasing adoption of DevOps practices in the telco industry, 5GC software may be developed, delivered, tested, and brought into operation incrementally at a far higher cadence than it was before. The CI/CD pipeline [43] will allow vendors to reduce time-to-market and shorten release cycles in their product roadmap. Based on this rationale, it is expected that the first 5GC solution suites will be quickly upgraded with new features, including multi-slice support in AMF/SMF. This feature allows the configuration of two or more slices in the same 5GC, by fetching associated S-NSSAIs from the NSSF and injecting them into the corresponding AMF/SMF instances. The ability of having multiple 5GC slices (CN slice subnets) running in parallel may offer operators greater possibilities to tap new 5G use cases targeting public network users (B2C market) and industry companies (B2B market). These customers may have different service requirements in terms of performance and functionality, hence the need to define different 5GC slice types for them:

- Business-to-Customer (B2C) slice types, used for serving traffic from user-centric applications.
- Business-to-Business (B2B) slice types, used for the provisioning of non-public networks (NPNs) [44], in particular for public network integrated NPNs (PNI-NPNs) [45].

Figure 11 captures a representative number of these 5GC slice types.

On the one hand, there is the *B2C category* (Figure H-13.A-B). This category includes 5GC slices (i) designed for end user consumption, so there is no need to have in-slice dedicated control plane functions; and (ii) to be entirely built on the 5G public network infrastructure (PLMN). Within this category, two different slice types can be found:

- **5GC slice Type A1 (shared UPF)**: deployment flavor wherein the 5GC slice does not have a dedicated UPF; indeed, the in-slice UPF instance is also shared with other 5GC slices.
- **5GC slice Type A2 (dedicated UPF)**: deployment flavor wherein the 5GC slice is allocated with a separate UPF.

In the B2C category, the first 5GC slices to be launched may be of Type A2, with few slices hosting premium communication services that end users can subscribe to. With a commercial model based on offering VIP service experiences, the operator looks to keep existing users and attract new ones, generating moderate revenues from their subscriptions in the short term.

For Type A1 slices, the situation is rather different. Unlike Type A2 slices, where traffic isolation across them is preserved with the provision of dedicated UPFs, in Type A1 slices the UPF is shared among them. In this context, operators expect that one UPF can support multiple 5GC slices, which means that one UPF shall be able to manage user plane resources (e.g., UE IP addresses, GTP-U Fully Qualified TEIDs, CPU, memory, bandwidth, etc.) at the S-NSSAI granularity. Unfortunately, the ability for a UPF to perform resource management (e.g., resource separation, resource allocation, resource usage report) per 5GC slice is not yet available in the standards due to technology limitations inherent to UPF internals, though 3GPP have already started working on solutions to solve this [46]. Being a Rel-17+ feature, we are still years away from seeing

Type A1 slices running in production networks. However, their availability will mark a major turning point in operator B2C slicing strategies, with the ability to deploy 5GC slices at a much wider scale, offering performance levels similar to Type A2 slices, but using a much lower number of UPFs. Additionally, the operator can use Type A1 slices to aggregate users with similar performance profiles, all this in a transparent manner, in search of creating efficiency improvements. This use of slicing, which allows the operator to streamline network management operations, is referred to as ’Network Slicing for Network Operator internals’ in [17].

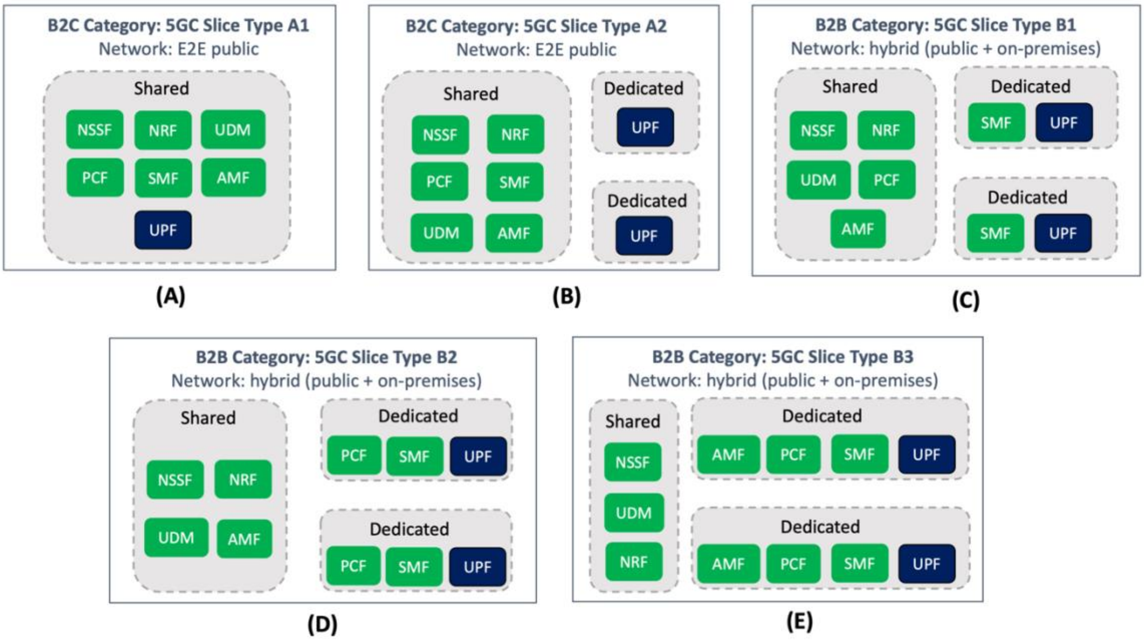


Figure H-13. 5GC slice types. (A)-(B) represent B2C slice types. (C)-(E) represent B2B slice types.

On the other hand, there is the *B2B category* (Figure H-13.C-E), which covers all the 5GC slices which are intended for industry customers. Examples of these customers include verticals and hyperscalers. Unlike the B2C category, the 5GC slices belonging to this new category will be used to host services for private use (i.e., only available for the customer’s subscribers), which typically span beyond the operator’s service footprint. This means that (i) every slice shall have dedicated control plane functions, so that the isolation of traffic management can be preserved across slices, and (ii) not all the in-slice functions will be hosted by PLMN; indeed, some of them might be deployed at customer facilities (e.g., UPF). Within this category, three different 5GC slice types are worth mentioning.

- **5GC slice Type B1 (baseline CP):** deployment flavor wherein 5GC slice is provided with dedicated UPF and a dedicated SMF. This ensures that in-slice traffic flows have an independent management and configuration, completely separated from other 5GC slices.
- **5GC slice Type B2 (advanced CP):** represents a 5GC slice Type B1 provisioned with dedicated PCF. Having a slice-specific PCF allows the customer to inject tailored QoS policies over in-slice traffic flows.
- **5GC slice Type B3 (premium CP)**: represents a 5GC slice Type B2 provisioned with dedicated AMF. Having a slice-specific AMF allows the customer to retain full control over mobility and connection management aspects regarding their subscribers.

The B2B category aims to exploit the real benefits that network slicing enables, which is the ability to provide separate network partitions with independent management for different industry customers. What these customers value most is to perceive allocated 5GC slices as dedicated, self-contained networks, under their own control. To that end, it is important for the operators to ensure that 5GC slices are delivered with network capabilities equivalent to those offered by private 5GC solutions (e.g., guaranteed SLA, traffic separation, controllable and configurable network), but at a much more reduced cost. This, together with the trust in the operator's proven know-how on OAM activities, is what will drive B2B customers to ask for a 5GC slice rather than purchasing a private 5GC from a 3rd party.

As evidenced from Figure H-13.C–E, the customer's perception of having a dedicated network requires the operator to provision 5GC slices with separate instances of some network functions. With the cloud-native design of 5GC and the consolidation of NFV practices into container-based environments, operators are conducting trials in this direction, assessing how the allocation of dedicated network functions impacts the number of 5GC slices that can be instantiated. This isolation vs scalability trade-off has demonstrated that the most optimal solution is to provision 5GC slices with dedicated instances of UPF and SMF. This flavor, which corresponds to 5GC slice type B1 (Figure H-13.C), will satisfy the service requirements of most industry customers [47][48]. However, there also exist specific customers whose business requirements may make them ask operators for more tailored 5GC slices, such as type B2 slices (Figure H-13.D) and type B3 slices (Figure H-13.E). Examples of these business requirements include the need for the customer to keep full control of QoS policies, or the need to get separate connection management of their subscribers.

Putting the B2B and B2C category solutions into the timeline reflected in Figure H-11, we can outline two things. First, in relation to the B2C category, we can see that type A2 slices may be commercially ready in the short term, while type A1 slices are expected in the long run. Secondly, in relation to the B2B category, we can see that all 5GC slice types may be available in the medium term, once Rel-16 features are integrated into the 5GC. Unlike the B2C category, where the difference between 5GC slice types is subjected to the availability of 3GPP solutions, in the B2B category the technology maturity of all the 5GC slice types is the same. The decision to go for one or another solution is entirely dependent on the customer-specific business requirements.

5.2 Add-on Features

The 5GC arena provides a lot of value-added capabilities that have a direct impact on the use of slicing, and that operators can progressively incorporate into their commercial networks.

In the short term (i.e., deploy ring in the radar), operators are focused on the introduction of edge computing. Edge computing is an evolution of cloud computing that allows moving workloads from centralized data centers (e.g., central PoPs) down to the telco edge nodes (e.g., regional PoPs), closer to consumers. In network slicing, edge computing is a must; in fact, bringing application hosting closer to the UE's access point of attachment allows achieving an efficient service delivery through the reduced end-to-end latency and load on the TN. The capabilities enabled by edge computing technology can be clustered into two solution sets:

- **Baseline edge computing**: provides support for application hosting and user-to-application connectivity. To that end, two network capabilities are needed. On the one hand, application placement capability, which allows for the optimized deployment of service applications at the target edge node, based on criteria such as resource availability, geographical areas, cost and latency requirements. On the other hand, edge node discovery capability, which represents the ability to identify an edge node capable of serving application clients (running on devices). In fact, when an application client wants to connect to an application, there is a need to discover the optimal edge node, which is the one that runs instances of the application, has the necessary resources (CPU, GPU, etc.) and provides the lowest network latency. For this discovery, there exists two solutions: DNS based (network layer solution, specified by 3GPP SA2) and device based (application layer solution, specified by 3GPP SA6). For further information on the pros and cons of these solutions, see [49].

- **Advanced edge computing**: provides mobility support in edge computing scenarios. As the user moves, it might happen that the current edge node is no longer valid, either because of SLA violation (e.g., the latency between the UE and serving node exceeds the maximum delay budget) or maintenance reasons (e.g., a node failure). This situation results in the user moving to a new edge node, a process that needs to be completed with the premise of keeping a seamless user service experience. This requires the availability of three main network capabilities: (i) service continuity capability; (ii) application re-location capability, i.e., to move the VM/container hosting the application instance from the source to the target edge node; (iii) context migration capability, i.e., to transfer the context from the stateful application towards the target edge node.

In the medium term (i.e., test ring in the radar), as long as standards and commercial products mature, operators are expected to enrich slicing functionality with the following solutions:

- **NWDAF**: the Network Data Analytics Function (NWDAF) [50] is a 3GPP Rel-16 function that provides network analysis information (upon request) about 5GC network entities. It provides S-NSSAI level analytics, and hence it may become the entry point to realize Artificial Intelligence (AI) in 5GC slices. NWDAF consumers can query for slice load levels and slice QoE measurements, or subscribe to slice-specific notifications that provide periodic updates or anomaly alerts. As shown in Figure H-6, examples of NWDAF consumers include: the Network Slice Selection Function (NSSF), which uses the S-NSSAI level analytics to add real-time intelligence to its slice selection algorithms; the PCF, which makes use of NWDAF info to optimize policy decisions on individual 5GC slices; and the NSSMF. 3GPP TS 23.288 [51] reports on the use of NWDAF to extract network analytics on a per network slice level.
- **Multiple slices per UE**: though 3GPP Rel-15 specifications allow a device to connect up to eight slices at the same time, thanks to the introduction of URSP (see Section 3.1) the reality is that most Rel-15 commercial solutions do not allow this feature. The existing limitations in commercial 5G SA handheld terminals prevent a UE from being connected to more than one slice at the same time. These limitations bet on the device's Operation System (OS) [29]. The device's OS mediates between the application clients and the device's modem, where the URSP is installed. Operators, vendors, device manufacturers and chipset providers are working together to find workarounds, with de facto solutions currently being assessed in different

PoCs. Among these solutions is the 5G slicing support in Android 12(S) devices, announced by Google in October 2021[2].

- **Secondary authentication**: Network Slice Specific Authentication and Authorization (NSSAA) attribute is defined in the GST [21] to specify whether, for a network slice, registered devices need to be authenticated by an external AAA server (Authentication, Authorization, Accounting server) using credentials different than the ones used for the primary authentication. This add-on feature, first introduced in 3GPP Rel-16 specifications, is intended for those industry customers that want to perform a second authentication over their subscribers. Operators are conducting bilateral trials with customers to help them understand the value of integrating NSSAA in B2B category 5GC slices, especially when used in the context of PNI-NPNs. In these trials, the operator-owned 5GC's NSSAA Function [52] contacts with the customer-owned AAA server via an AAA proxy. For further details on this interaction, see [10][30].

Finally, in the longer run (i.e., explore ring in the radar), the integration of Rel-17+ features into the 5GC will allow operators to unleash the full potential of 5GC slicing. In this scouting phase, operators have set their sights on these two featured solutions

- **Slice Roaming**: operators are expected to support roaming for network slicing, at least for network slices deployed from S-NESTs. However, this feature is still years away from being in commercial networks, as there are technical and commercial aspects that need to be agreed upon. The technical aspects are discussed in [53], a GSMA document where operators have captured their priorities on slice roaming so as to guide the specification of normative solutions in 3GPP. The commercial aspects include charging, billing and business models that are still under discussion. Unless all these aspects are agreed and reported, no multi-operator trials are expected shortly.
- **NSACF**: the Network Slice Access Control Function (NSACF) is a Rel-17 5GC function that monitors and controls (i) the number of registered UEs per network slice, and (ii) the number of PDU sessions per network slice [10]. With the NSACF, operators can enforce quotas on individual slices, making sure the signaling traffic and packet flows do not exceed the maximum slice load. NSCAF is still in stage 2 (functional definition), so no vendor solutions are yet available. In the meantime, operators are now trying to understand how to best apply this functionality to improve internal network operation, and how to link them with the admission control functionality at the NG-RAN side.

6 RAN Domain

In the RAN domain, the technology radar puts the focus on three different dimensions: functionality, radio resource allocation and penetration. The *functionality* dimension provides a deep dive on the applicability of open RAN principles on NR protocol stack functions to design and configure RAN slices, going from monolithic solutions towards more flexible, service-tailored composition patterns. The *radio resource allocation* dimension discusses the availability of solutions to segregate and dispatch cell radio resources to competing RAN slices, so that their targeted KPIs are met. Finally, the

[2] https://cloud.google.com/blog/topics/telecommunications/5g-network-slicing-with-google-android-enterprise-and-cloud, accessed on 20 October 2021

penetration dimension refers to the penetration of RAN slicing technology within the operator's footprint.

6.1 Functionality

As noted from Figure H-2, the target RAN slicing architecture lies on the possibility of (i) having a 3-tier NR protocol stack, distributed into RU, DU and CU modules; and (ii) provisioning a dedicated CU-UP instance to each slice, with tailored PDCP configuration settings, so that the delay and security requirements for a given S-NSSAI can be fulfilled. The achievement of these two milestones is mandatory for an operator to have a fully operable RAN slicing solution, as described later on. For the sake of network efficiency (i.e., OPEX savings) or further service innovation (i.e., new revenue streams), the operator might decide to enhance the baseline solution by integrating add-on features atop. One example is the integration of the RAN Intelligent Controller (RIC) [54], an optional AI-powered functionality originally defined in the O-RAN framework [55].

In the following, we describe the stepwise journey we foresee for a future-proof RAN slicing.

In current NSA scenarios (i.e., as-is ring in the radar), the predominant operator scenario is a **few physical gNBs** providing macro coverage to city and suburban areas. Installed in strategic geographic locations, these gNBs have inbuilt 4G/5G essential features, including massive MIMO, Dynamic Spectrum Sharing (DSS) [56], Narrow Band IoT (NB-IoT) and RAN sharing

As soon as the 5G coverage footprint needs to be extended, something that has already started with the rollout of first commercial SA networks, operators may migrate towards gNB cloudification, in search of CAPEX reduction. In fact, with this action, the operators are able to extend 5G coverage at large scale without the need to deploy costly physical gNBs everywhere. This bets on different, yet intertwined solutions:

- **DU-CU disaggregation**, whereby gNB is functionally split into one (centralized) CU instance and multiple (distributed) DU instances, conforming to a split 2 option. The result of this disaggregation is that CU can be entirely implemented in software, and therefore deployed as a VNF in any cloud environment. In fact, while individual DU instances remain colocated with RU at cell sites, the workload corresponding to the CU instance can be moved to the telco edge cloud.
- **Control User Plane Separation**, whereby the virtualized CU software is further decomposed into one CU-CP instance and multiple CU-UP instances. This requires a complete reshaping of CU software design, transforming a coarse-grained (VM-based) VNF into a number of modular (container-based) VNFs, each hosting a different instance.

These two solutions will be available in the short term; hence they are categorized in the deploy ring of the radar.

In the medium term, once the penetration rate of cloudified gNBs is significant, the operators may use deployed assets to have a fully operable RAN slicing environment. This environment, which shall be compliant with the two requirements captured in the beginning of this subsection, leverages on two solutions:

- **Slice-specific vCU-UP**, based on providing each RAN slice with a separate vCU-UP instance. This solution not only ensures user plane traffic isolation across different RAN slices [57], but also adapts/customizes the processing of DRB flows according to the slice specific needs.

- The implementation of **vDUs**, to allow for a 3-tier NR protocol stack. This solution is the result of a three-step journey, whereby the DU is first separated from the RU (introducing a fronthaul link between them, according to split option 7-2x), then designed in software (modelled as a VNF), and finally deployed into access PoPs (far edge sites). The ability to move DU workloads into a cloud environment is of particular interest for large-scale slices hosting distributed eMBB services or mIoT applications. However, this feature does not come like that alone; indeed, it needs to be accompanied with the provision of rich-featured CPU architectures and hardware acceleration solutions [27] in the access PoP, as outlined in Section 2.4. These assets are aimed at reducing the impact that virtualization overheads may introduce on DU packet-processing performance.

Industry stakeholders (including operators) have started to test these two solutions, validating their performance and assessing their techno-economic viability in typical scenarios where the vDU is shared across slices. In parallel, operators have launched a new activity on RIC, pushed by the O-RAN momentum. Unlike the first workstream, focused on searching for a baseline RAN slicing environment that can be easily scaled and replicated across the operator’s service footprint, this new workstream has the goal of checking how the RIC can further enhance RAN slicing features. As noted in [55], the RIC is a non-mandatory O-RAN component consisting of two modules: (i) near-RT RIC, which hosts ms-level RRM logic with embedded intelligence; and (ii) non-RT RIC, which provides delay-tolerant functionality including service and policy management, RAN analytics and AI/ML model training. Figure H-14 provides details on these two modules. In this testing phase, the focus is on the **near-RT RIC** (near-Real Time RIC), and in particular on the capabilities of the E2 interface [58].

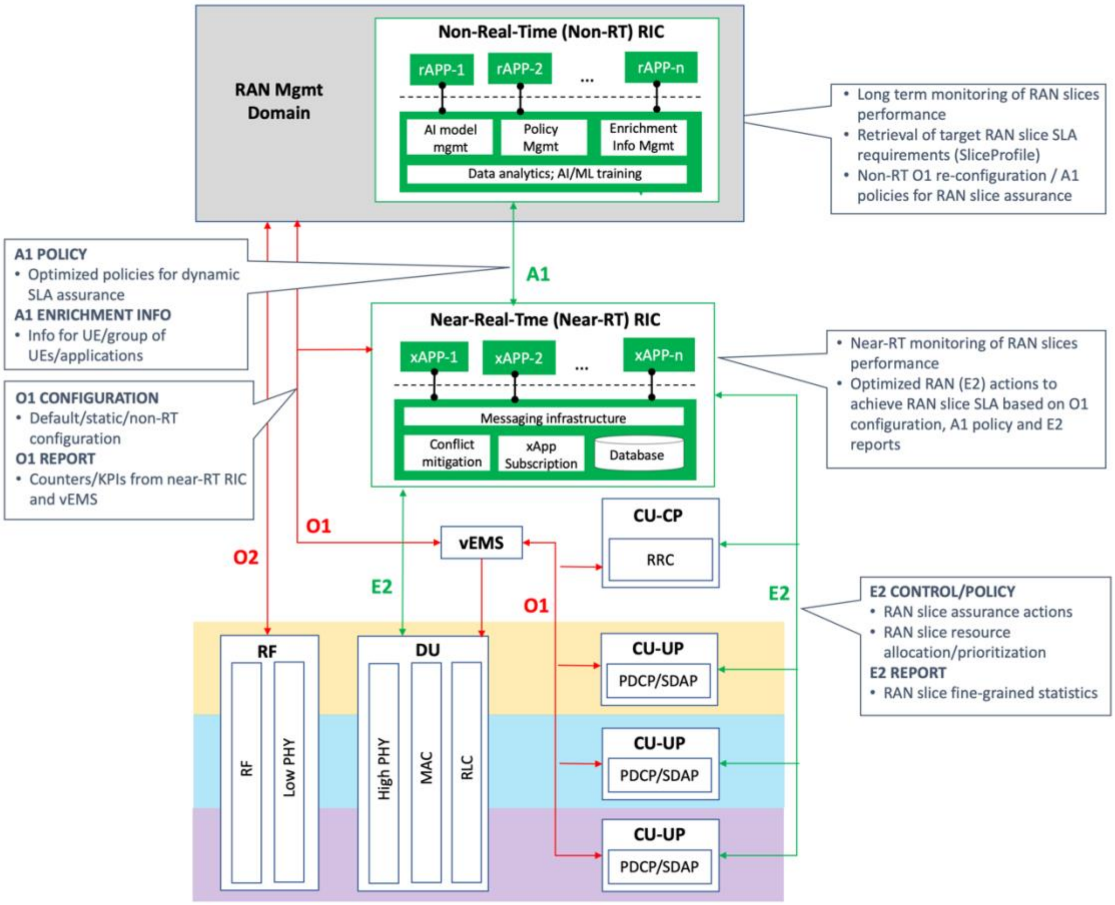

RAN Mgmt Domain
Non-Real-Time (Non-RT) RIC
rAPP-1
rAPP-2
...
rAPP-n
AI model mgmt
Policy Mgmt
Enrichment Info Mgmt
Data analytics; AI/ML training
• Long term monitoring of RAN slices performance
• Retrieval of target RAN slice SLA requirements (SliceProfile)
• Non-RT O1 re-configuration / A1 policies for RAN slice assurance
A1 POLICY
• Optimized policies for dynamic SLA assurance
A1 ENRICHMENT INFO
• Info for UE/group of UEs/applications
A1
Near-Real-Tme (Near-RT) RIC
xAPP-1
xAPP-2
...
xAPP-n
Messaging infrastructure
Conflict mitigation
xApp Subscription
Database
• Near-RT monitoring of RAN slices performance
• Optimized RAN (E2) actions to achieve RAN slice SLA based on O1 configuration, A1 policy and E2 reports
O1 CONFIGURATION
• Default/static/non-RT configuration
O1 REPORT
• Counters/KPIs from near-RT RIC and vEMS
O1
O2
vEMS
E2
O1
CU-CP
RRC
E2
E2 CONTROL/POLICY
• RAN slice assurance actions
• RAN slice resource allocation/prioritization
E2 REPORT
• RAN slice fine-grained statistics
RF
RF
Low PHY
DU
High PHY
MAC
RLC
CU-UP
PDCP/SDAP
CU-UP
PDCP/SDAP
CU-UP
PDCP/SDAP

Figure H-14. AI assisted RAN slice resource control.

In the long run, when the commercial O-RAN solutions are relatively stable, operators may focus on how to include the entire RIC in the day-to-day operation of RAN slices, making it more agile and intelligent [59]. In this regard, the following solutions are currently being explored:

- **xApps**. As seen from Figure H-14, the near-RT RIC delivers a robust, scalable, and secure platform for xApp hosting. These xApps are third party control applications that complement traditional RRM functionality, by bringing advanced algorithms applicable to real-time use cases, including QoS/QoE optimization, per-UE controlled load balancing, traffic steering and seamless handover control. Operators together with third party developers are exploring the possibility of implementing slice-specific RRM procedures through xApps, in order to decouple life cycles of slicing related innovation (e.g., 4/6-month release cycle) from CU-CP hosted RRM policies (e.g., 1/2-year release cycle).
- **Non-RT RIC**. The non-RT RIC (non-Real Time RIC) will be the main driver for AI-powered RAN slicing. This module will host and manage all AI/ML models that will later be pushed into the near-RT RIC down to the individual RAN slices. In the exploratory phase, where the operators are now, the main entry barrier they have encountered is its complex integration. In fact, the non-RT RIC needs to communicate with (i) the RU and the vEMS, using O1 interface [60]; (ii) the near-RT RIC, using A1 reference point [61]. In addition, it needs to interact with 3GPP

management system, something that today is still under discussion in O-RAN community.

Figure H-14 illustrates the integration of all the O-RAN components into the network slicing system architecture, including the reference points (E2, O1 and A1 interfaces) across them. To illustrate the information exchanged between these interfaces, we propose an example of the use of the RIC to make an AI assisted resource control for RAN slice SLA assurance. The logic of this resource control is hosted by xApps. Further details on this example can be found at the end of Section 6.2, once the details on radio resource allocation are explained.

6.2 Radio Resource Allocation

In this dimension, we encounter a number of solutions in the gNB scheduler enabling RAN slicing. These solutions are based on two different yet complementary mechanisms: priority scheduling and radio resource partitioning.

On the one hand, *priority scheduling* is a mechanism used for service differentiation between RAN slices when they all compete for the same PRBs. This corresponds to best effort slices, as captured in Table H-2. The service differentiation here is achieved with the definition of multiple DRB profiles, each featured with a different scheduling priority, and their allocation to existing RAN slices. This approach allows giving one or more RAN slices a preferential treatment with relative priority to others. The policy for DRB profile allocation is based on QoS criteria, though other business-wise criteria can be applied (e.g., based on platinum/gold/silver subscription).

On the other hand, *radio resource partitioning* is a mechanism based on PRB reservation. Unlike priority scheduling, radio resource partitioning allows segregating cell resources across different RAN slices. Here, resource segregation consists of allocating separate PRB chunks to the DRBs associated to these slices. For this PRB allocation, the gNB scheduler configures the resource quotas per slice. As shown in Figure H-8, the operator can configure up to three resource quotas in a slice, namely, maximum slice quota (mandatory), minimum slice quota (mandatory), and dedicated minimum slice quota (optional). For the definition of these per slice quotas, the operator makes use of RRMPolicy class defined in the 5G NRM, by applying the RRMPolicyRatio attributes (i.e., RRMPolicyMaxRatio, RRMPolicyMinRatio, RRMPolicyDedicatedRadio) to the {S-NSSAI, PLMN ID} tuple the RAN slice is associated with. For further information on this mapping, see clause 4.3.36 from 3GPP TS 28.541 [15].

Table H-3 provides a comparative analysis between priority scheduling and radio resource partitioning. While it may seem logical to define PRB reservation for each slice supported at the RAN, this is, in practice, suboptimal. There is a trade-off in performance between the gain from dedicating resources to specific slice services and the overhead in maintaining numerous resource partitions. The balance is to keep sufficient PRB chunks to guarantee resource isolation per RAN slice when needed, while not impacting radio performance due to excessive partitioning.

Table H-3. A comparative analysis of RAN slicing mechanisms.

Topic	Priority Scheduling	Radio Resource Partitioning
Solution	All slices share resources, and the slice SLA is guaranteed by increasing the scheduling priority of devices that have	Reserve dedicated resources and prioritized resources for specific slice(s) to ensure that devices in the

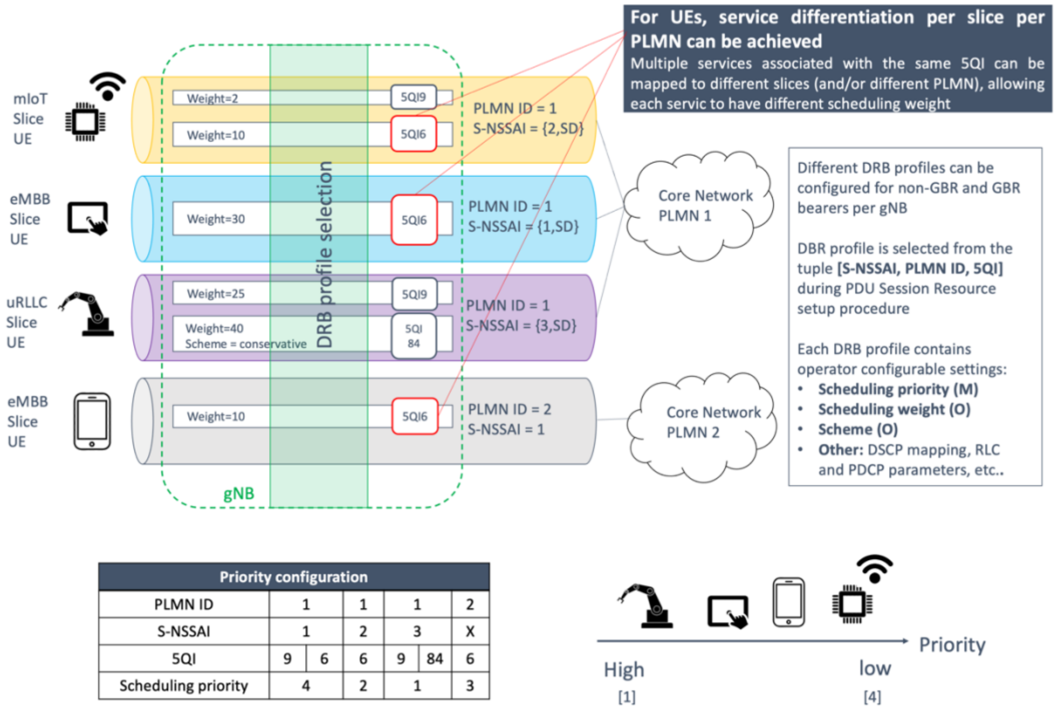
	not reached the minimum guaranteed rate	slices have sufficient available at any time
Guarantee	Best effort SLA guarantee: when the cell is congested, the SLA of some slices may not be guaranteed	Customers get accurate SLA guarantee
Isolation	Resource sharing between slices that can only provide limited soft isolation with different priorities	Hard isolation of resources between slices avoids mutual influence across slices and provides an isolation effect similar to that of dedicated frequency
Scenario	B2C market slicing	B2B market slicing

In the following, we describe the solutions that we foresee from these two RAN slicing mechanisms, and that are illustrated in the radar shown in Figure H-11.

In today's dominant NSA scenarios (as-is ring in the radar), the only possible solution for RAN slicing is **QCI Priority Scheduling**, based on giving preferential treatment to those DRBs associated with top-priority services. The priority of a service depends on the QoS parameters associated with the service flows, with a QCI as key parameter, complemented with other QoS related info (i.e., GBR + MBR values if service flows convey GBR traffic, AMBR values if service flows convey non-GBR traffic). Upon computation on the EPC side, the priorities of different services are sent to the gNB, which uses them to define corresponding DRB profiles. The number of DRB profiles available in current production networks is rather low; the reason is a coarse-grained design of DRBs, and that only few of them convey prioritized traffic.

As we move towards the rollout of the first 5G SA networks, new priority scheduling solutions are appearing:

- **5QI Priority Scheduling**. This solution is equivalent to the QCI Priority Scheduling, but using 5GC, which provides much more granular QoS control than the EPC. The fact that (i) the 3GPP 5G QoS framework allows one slice to convey multiple QoS flows, each featured with a specific 5QI, and (ii) the DRB profiles are computed based on 5QI, make it possible to have intra-slice service differentiation. Figure H-15 illustrates an example of how this solution works.
- **Relative Priority Scheduling**, which is an evolution of the previous solution. This evolution is based on enriching DRB profile characterization with additional configurable settings, including scheduling weight and scheme, among others. These settings, listed in Figure H-15 with the optional "(O)" tag, allow the gNB scheduler to resolve conflicting situations that are beyond the capabilities of 5QI priority scheduling. One example is when QoS flows from different slices have the same 5QI, but there is a need for preferential treatment of one of these slices. In such a case, scheduling weight can be used, as elaborated in the top-right corner of Figure H-15.



For UEs, service differentiation per slice per PLMN can be achieved
Multiple services associated with the same 5QI can be mapped to different slices (and/or different PLMN), allowing each servic to have different scheduling weight

Different DRB profiles can be configured for non-GBR and GBR bearers per gNB

DBR profile is selected from the tuple **[S-NSSAI, PLMN ID, 5QI]** during PDU Session Resource setup procedure

Each DRB profile contains operator configurable settings:
- **Scheduling priority (M)**
- **Scheduling weight (O)**
- **Scheme (O)**
- **Other:** DSCP mapping, RLC and PDCP parameters, etc..

Priority configuration						
PLMN ID	1		1	1		2
S-NSSAI	1		2	3		X
5QI	9	6	6	9	84	6
Scheduling priority	4		2	1		3

Figure H-15. Priority scheduling.

These two solutions build up the RAN slicing suite that operators have started to deploy, and may be operationally ready in the short term. However, with sights already set on evolving slicing features in the medium term, operators are also testing other solutions:

- **Delay controlled Priority Scheduling**. This solution consists of enriching gNB scheduling logic with the adaptive managed latency concept. It allows providing better experience to Rel-16 services, mostly interactive services that require high data rate and low latency communications. The adaptive managed latency concept represents the ability to provide bounded and steady low latency for these services, by coupling gNB scheduler and application in a feedback loop with dynamic rate adaptation signaled by the service application. This coupling can be enforced using either vendor-specific mechanisms or standard frameworks, such as Low Latency, Low Loss, or Scalable Throughput (L4S) [62]. Figure H-16 depicts how L4S congestion marking and feedback can be applied for gNB scheduling. Note that this Delay Controlled Priority Scheduling solution (application layer rate adaptation) is complementary to Relative Priority Scheduling (network layer service differentiation), and both can be used simultaneously.
- **Radio Resource Partitioning**. This is the first solution from the radio resource partitioning category. For the PRB allocation, this solution assumes that individual slices are only configured with dedicated resources, without prioritized resources. To achieve this, the operator sets the same value for the dedicated minimum slice quota (RRMPolicyDedicatedRatio) and the minimum slice quota (RRMPolicyMinRatio). From the set of flavors tabulated in Table H-2, one can note that the RAN slices resulting from this solution are compliant with the "dedicated slice-profile 1" settings. This flavor provides isolation and secure resources in high load conditions, at the cost of poor multiplexing gains. The reason is that dedicated resources of a slice cannot be used by others, even though the slice resource usage is below the dedicated minimum slice quota.

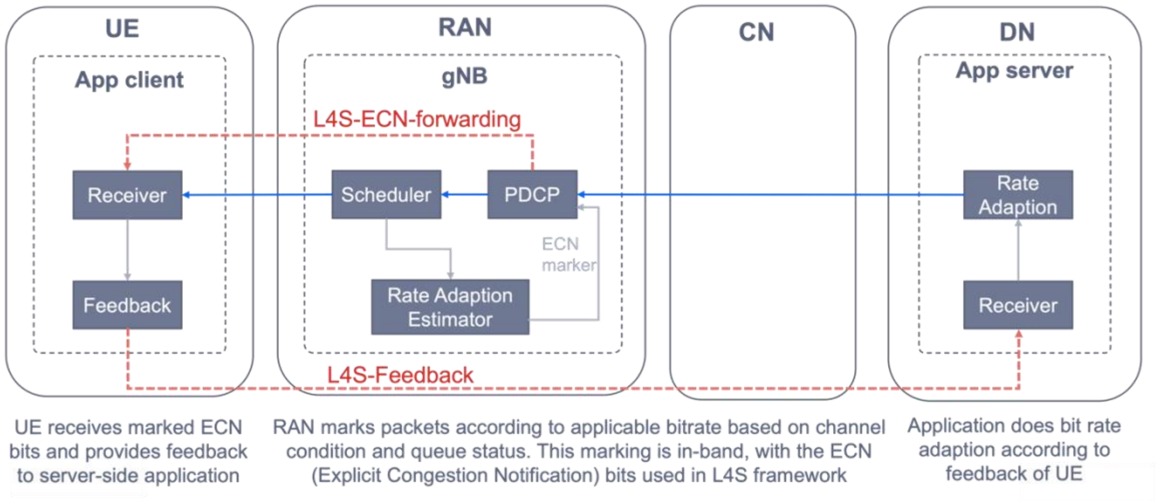

UE
App client
Receiver
Feedback
RAN
gNB
L4S-ECN-forwarding
Scheduler
PDCP
ECN marker
Rate Adaption Estimator
L4S-Feedback
CN
DN
App server
Rate Adaption
Receiver
UE receives marked ECN bits and provides feedback to server-side application
RAN marks packets according to applicable bitrate based on channel condition and queue status. This marking is in-band, with the ECN (Explicit Congestion Notification) bits used in L4S framework
Application does bit rate adaption according to feedback of UE

Figure H-16. Feedback loop for delay-controlled priority scheduling, with in-band L4S.

Finally, in the long run, we can find fully blown RAN slicing solutions. These solutions, listed below, are years away from being commercially available; in fact, they are still taking shape in the reference standards: 3GPP and O-RAN Alliance. Therefore, they are captured in the explore ring in the radar.

- **Static (Policy-based) Flexible Radio Resource Partitioning**. It allows for defining prioritized resources per slice, a feature which was disabled in the Hard Radio Resource Partitioning. With this new option, RAN slices can now be configured according to "dedicated slice-profile 2" and "prioritized slice" settings, as shown in Table H-2. The definition of prioritized resources per slice boosts resource efficiency, at the cost of making gNB scheduler logic much more complex, with a larger number of decision-making variables that need to be computed in real-time.
  It is important to note that gNB scheduler can work with both hard and flexible radio resource partitioning solutions, as depicted in Figure H-17. This example shows that the gNB is configured to serve four slices from two different PLMNs: PLMN#1, hosting mIoT, eMBB and uRLLC slices, and PLMN#2, hosting another eMBB slice. Looking at the slice specific quotas, one can note that the uRLLC slice is scheduled with hard approaches. For the remaining slices, flexible resource radio resource partitioning is applied, with the mIoT slice configured with "dedicated slice-profile 2" flavor and the eMBB ones configured with "prioritized slice" flavors.
- **Dynamic (AI-assisted) Flexible Radio Resource Partitioning**. This solution takes the previous solution to the next level, with the possibility of changing slice resource quotas over time, depending on collected performance metrics. To cope with this dynamism, operators need to take humans out of the loop, replacing them with novel AI-assisted artifacts enabling closed-loop automation. The xApps hosted by the near-RT RIC are perfect candidates for this role; indeed, they can provide agility and context-awareness in the decisions of changing resource quotas [63].

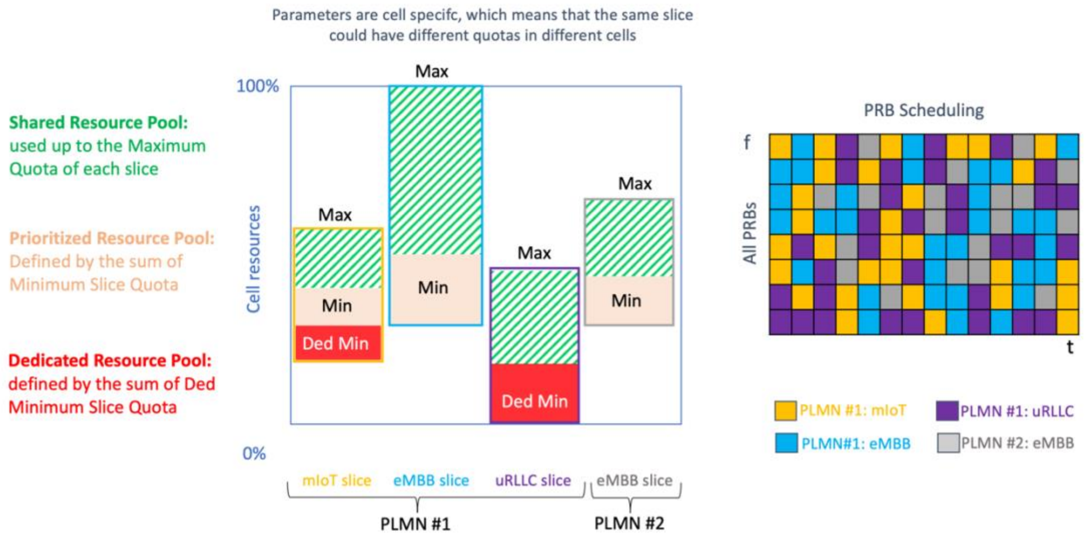

Parameters are cell specifc, which means that the same slice could have different quotas in different cells
Shared Resource Pool: used up to the Maximum Quota of each slice
Prioritized Resource Pool: Defined by the sum of Minimum Slice Quota
Dedicated Resource Pool: defined by the sum of Ded Minimum Slice Quota
100%
0%
Cell resources
Max
Min
Ded Min
Max
Min
Max
Ded Min
Max
Min
mIoT slice
eMBB slice
uRLLC slice
eMBB slice
PLMN #1
PLMN #2
PRB Scheduling
f
All PRBs
t
PLMN #1: mIoT
PLMN #1: uRLLC
PLMN#1: eMBB
PLMN #2: eMBB

Figure H-17. Radio Resource Partitioning.

Figure H-14 illustrates an example of the usability of near-RT RIC for real-time resource control of operative RAN slices. As seen, the near-RT-RIC makes use of E2 interface to interact with associated DU, CU-UP and CU-CP instances, for both monitoring (E2 REPORT) and configuration management (E2 CONTROL/POLICY) actions. The governance of these actions lies on xApps. The logic of these xApps can be assisted with AI/ML models, which are made available for consumption by the non-RT RIC using the A1 interface. By comparing collected metrics against AI/ML models, the xApps can make real-time decisions on quota changes, enforcing them in corresponding CU-UP instances. Though the ultimate invocation and execution of AI/ML models lies on xApps, this cannot be done without the Non-RT-RIC, which plays a key role in this workflow; indeed, non-RT RIC is responsible for the management of individual AI/ML model (e.g., AI/ML model deployment, configuration, performance evaluation, termination, validation, and testing).

6.3 Penetration

In this section, we provide a description of the penetration path we foresee for network slicing in cell sites. For this exercise, it is important to take into account the environment setup. In this regard, we consider three phases: early introduction (phase 1), full adoption on standalone private 5G networks (phase 2) and full adoption in public 5G networks (phase 3).

The *phase 1* covers the as-is and deploy rings of the radar. In this phase, it is assumed a baseline RAN slicing solution, consisting of applying 5QI priority scheduling (see radio resource allocation dimension) over cloud NG-RAN scenarios (see functionality dimensions), with no mobility support. This setup has minimal impact on production networks, without compromising backwards compatibility, and therefore constitutes a good candidate for early introduction in the PLMN. The integration of baseline RAN slicing in first SA networks ensures QoS-based service differentiation on RAN side, which is essential for operators to start commercializing 5GC Type A2 slices for B2C market (see Figure H-13).

However, for more elaborated setups leveraging open RAN principles, the above rationale is no longer valid. The major changes that these setups bring in RAN planning and configuration patterns prevent their large-scale introduction in today's commercial networks, which account for a large percentage of legacy RAN assets. Instead, operators may initially choose to start testing their RAN slicing on private networks (phase 2) before moving to macro cells (phase 3).

The *phase 2* covers the deploy, test and explore rings of the radar. In this phase, the operator may integrate RAN slicing solutions only in greenfield environments, **on-premises**, creating isolated islands of private 5G adoption. This setup constitutes a perfect niche for the operators to start commercializing their innovative RAN slicing solutions towards B2B customers, as the radio resource allocation mechanisms and O-RAN components become available. These solutions can be used either (i) to provide traffic isolation across on-premises customer services, or (ii) to provide the customer with a guaranteed SLA between the device and 5GC. In the second case, RAN slicing is used in conjunction with the B2B category 5GC slices to provide PNI-NPN services.

Finally, there is the *phase 3*, spanning across the test and explore rings of the radar. This phase 3 consists of applying the lessons learnt in phase 2 to macro cells. In fact, the trials in phase 2 will serve the operator as a playground before moving to the PLMN, where UE density and traffic patterns are radically different, and where mobility events now need to be taken into account. To facilitate this transition, a three-step journey is foreseen:

- **PLMN, single-slice TAI**. The operators will start to replicate the trials in specific tracking areas, with their cells configured with one single S-NSSAI.
- **PLMN, multiple-slice TAI**. The same solution as the previous one, but configuring all the cells from the same tracking area with two or more S-NSSAIs.
- **PLMN, large-scale**. The lessons learnt from the small-scale deployments allow improving RAN slicing before massive adoption in the PLMN, where more complex problems on capacity planning and mobility management may appear.

7 TN Domain

In the TN domain, the technology radar requires the analysis from two separate dimensions: transport technologies and transport SDN fabric. The *transport technologies* dimension captures the protocol encapsulation and data plane solutions across the different network segments. On the other hand, the *transport SDN fabric* dimension refers to the control and management plane aspects, discussing the application of programmability and automation through SDN.

These two dimensions are not coupled but complementary, in the sense they can evolve at a different pace, and they do not necessarily need to be used together. The transport technology dimension includes all the Layer 1/2/3 solutions that allow segregating connectivity resources and enforcing traffic separation at the WAN infrastructure, therefore enabling TN slicing realization. The sole use of these transport technologies is enough to have a sliced WAN infrastructure, but not to get it provisioned and operated in a dynamic way. The latter would require implementing programmability and automation capabilities atop these technologies, something that can only be brought by the SDN paradigm. The complementarity of the transport SDN fabric lies in the ability to operate with these technologies using a set of SDN controllers instead of traditional, siloed TN management systems (which typically lead to rather static, manual configurations). This requires equipping these SDN controllers with programmatic interfaces on the southbound side, towards the underlying technology devices. These interfaces shall be developed in such a manner that SDN controllers can inject configuration actions and retrieve collected data following a model-based approach.

7.1 Transport Technologies

The system architecture represented in Figure H-2 shows the entities on the path for an E2E slice, assuming a 3-tier deployment for the gNB: RU, DU and CU-UP. As seen, the user plane from the device to the application includes the following TN segments: fronthaul (O-RAN fronthaul interface), midhaul (3GPP F1-U interface), backhaul (3GPP N3 interface) and DN (3GPP N6 interface). The F1-U and N3 interfaces use GTP-U to transport packets (IPv4, IPv6, Ethernet or unstructured) in the PDU session established between the UE and the DN. The fronthaul interface carries the radio frames in the form of In-Phase (I) and Quadrature (Q) samples, using eCPRI [64] encapsulation over Ethernet or UDP over IP.

The mission of TN slicing is to implement slicing-featured capabilities across all these connectivity segments, ensuring that the SLA requirements are met not only at the 3GPP network function layer, but also at the transport underlay. Examples of these capabilities include guaranteed QoS (e.g., bandwidth reservation, upper latency bounds, controller delay variation), isolation, protection, and reliability (e.g., disjoint routes).

As outlined in Section 2.3 and shown in Figure H-10, there exist multiple networking solutions that can be used for the provision of these capabilities, ranging from soft slicing to hard slicing, with some trade-off solutions in between (i.e., hybrid slicing). Next, we present the journey we foresee towards the realization of an end-to-end TN slicing, based on a stepwise integration of different transport technologies into carrier networks.

Today’s scenarios are based on 5G NSA, where the slice data path only includes two segments: backhaul segment (between the physical gNB and S/P-GW) and DN segment (between S/P-GW and application). Typical implementations for these scenarios bet on the use of **traditional L2/L3 overlay** solutions, mainly VPN techniques. Layer 2 VPN techniques are, for example, Virtual Private LAN Service (VPLS), which sets up a Ethernet-based communication over MPLS tunnels, or Virtual Leased Line (VLL), which emulates a pipeline between two given endpoints. Layer 3 VPNs can be realized, for example, via Virtual Private Routed Networks (VPRN), featuring MPLS-based VPNs. Two main reasons explain the preference for these technologies. On the one hand, they do not have an impact in the underlay. Taking into account the brownfield design of today’s transport networks, it is natural for operators to reuse as much as possible existing overlay technologies, especially if they are carrier-grade and widely available along the operator footprint. On the other hand, they are cost-efficient, with a performance that is enough to satisfy the transport requirements of NSA slices.

In the short term, early B2C slicing offering will require evolving transport networks with different solutions such as the listed below.

- **DiffServ Code Point (DSCP)**. The integration of 5GC will lead to a change in the backhaul segment, now based in the N3 interface. Additionally, the DU-CU disaggregation will produce the F1 interface for the mid-haul segment. With this scenario, the setup is as follows: a GTP tunnel encapsulating user packets with slicing information captured in the {PLMN ID, S-NSSAI, 5QI} triplet transverses the backhaul and midhaul segments. Since the IP underlay across these segments is not able to interpret these 3GPP signaling identifiers, it is not possible for the border router (see Figure H-9.A) to apply the constraints represented by this tuple. In this regard, the UPF and RAN shall perform transport level packet marking in downlink and uplink, respectively, by setting the DSCP in the outer IP header [65]. This information is used by the corresponding border routers in the IP underlay to differentiate traffic from different slices [66][67].
- **Segment Routing (SR)**, and its use across backhaul and midhaul segments. Apart from the rich built-in features (e.g., highly scalable, responsive, programmability) [68], what makes SR [69] an ideal technology option for the IP underlay is that it can

be used with both the MPLS forwarding plane (SR-MPLS) and the IPv6 forwarding plane (SRv6). Examples of realization of slicing in SR networks can be found in [70][71]. The basis of this realization lies in the ability of SR to encapsulate additional information for discriminating traffic associated with different slices [72].

The solutions explained so far, belonging to the as-is and deploy rings in the technology radar, are part of the soft slicing category presented in Section 2.

In the medium term (the testing ring in the radar), as novel B2B offerings come into service portfolio and the slicing requirements get burdened, the capabilities offered by soft slicing solutions are not enough. In this regard, operators have started to evaluate new technologies, working on three main directions:

- **Hard slicing**. In this workstream, the operators are focused on trialing technologies like Flexible Ethernet (Flex-E), Flex-O and DWDM. These technologies are used to realize the concept of isolated traffic flows operating on common links that avoid negatively influencing the performance of each other in case of congestion. In particular, **Flex-E** [35] bets on the principle of calendar-based channelization, which consists of bundling or dividing physical Ethernet interfaces into multiple Ethernet hard pipes based on timeslot scheduling. The work in [73] gives further details on how Flex-E technology might be used to implement hard slicing. **Flex-O**, described in the ITU-T G.709.1/Y.1331.1 recommendation [74], provides Optical Transport Network (OTN) interfaces with comparable functionality to that of Flex-E based Ethernet interfaces. Finally, **DWDM** can be used for physical resource separation at wavelength level. Adaptive transponders over Wavelength Division Multiplexing (WDM), spectrum fragmentation and Optical Cross-Connect (OXC), and Reconfigurable Optical Add-Drop Multiplexer (ROADM) are optical network virtualization techniques that can be exploited for network slicing, so that wavelengths can be right-sized for the specific requirements of every slice.
- Deterministic techniques. Flex-O, Flex-E and DWDM are ideal for providing guarantees for throughput and delay, which will be a common pattern across most of the uRLLC scenarios. Apart from these hard slicing technologies, in order to enable deterministic real-time performance in transport networks, novel full-stack approaches are also proposed in the underlay, with techniques such as those outlined by the IEEE TSN Project Group (layer 2 aspects) and the IETF DetNet Working Group (layer 3 aspects). On the one hand, **TSN** is a set of IEEE 802.1 amendments that enable determinism of time-critical traffic flows, even in cases where traffic flows with different statistical characteristics are multiplexed. The use of slicing in TSN-based networks is discussed in [75]. On the other hand, **DetNet** defines a set of techniques to extend deterministic behavior to in-slice layer 3 paths. These techniques span from explicit routes, packet replication and elimination, to congestion protection with E2E synchronization [76].
- **SD-WAN**. It is not uncommon that B2B customers contracting 5GC slices will request QoS-assured connectivity across their enterprise sites. This setup may result in a multi-site B2B slice, typical in PNI-NPN scenarios. In this case, Software Defined WAN (SD-WAN) [77] is a good candidate solution for the DN segment. Some operators have started to execute PoCs in this direction, e.g., [78].

Finally, the focus in the longer term is on making **O-RAN fronthaul interface slicing-aware** [59]. Unlike the midhaul and backhaul segment (3GPP interfaces), where slicing information is injected with the EP_Transport IOC (see Figure H-9.B), the fronthaul segment (O-RAN interface) does not currently support slicing features. Figure H-18 shows the current protocol structure for the O-RAN fronthaul interface. As seen, the stack allows DU and RU to exchange (i) signaling information; (ii) user plane information, based on frequency-domain IQ samples; (iii) timing and synchronization information; and (iv) management plane information. However, no slicing related data is yet supported. The challenge that industry shall solve is how to encapsulate S-NSSAI (3GPP signaling information) in the O-RAN fronthaul interface, considering that this interface is out of the 3GPP specifications. O-RAN Working Group 4 is working towards this direction, though normative solutions (and therefore commercial solutions) are still far away.

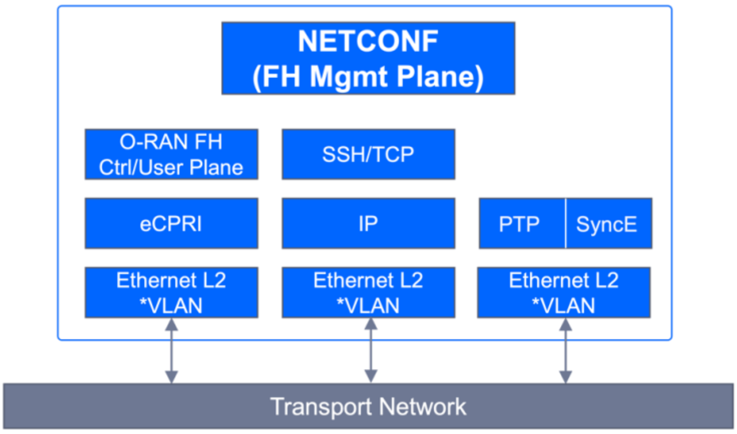


Figure H-18. O-RAN FH interface protocol structure.

7.2 Transport SDN Fabric

The use of a transport SDN fabric brings automatic network control and programmatic capabilities. These features are key for the operators in their need to cope with the agility and management complexity of slicing in their transport networks, with multiple vendor and technology solutions underneath.

As of today, the scope of SDN technology has been circumscribed inside the data center, using it to simplify the management of internal connectivity as the adoption of cloud solutions gets consolidated. However, the applicability of SDN into transport networks is a completely radical approach, with the use of new protocols and different model-driven operational practices quite tied to the specificities of the WAN technologies present in today’s carrier facilities. Making transport SDN a reality is not an easy task. In this regard, many efforts have been made over the past few years in this direction, with telco industry actors working out promising but misaligned solutions. Proof of this is the large number of architecture frameworks defined so far, e.g., [79]-[81].

It is now time to move forward and make progress. This requires the definition of a common strategy to reduce and select the most suitable standards to unify disparate SDN solutions into a single E2E, open transport SDN architecture. The preferred architecture option is shown in Figure H-19. Originally proposed by Telefónica with the iFUSION brand [82], this architecture follows a hierarchical model, with specific domain controllers per technology domain (IP/MPLS, optical/DWDM and microwave/backhaul) and an overarching Software-Defined Transport Network controller (SDTN controller). This two-layer control architecture has become the reference framework for operators worldwide, as echoed in the white paper published by Telecom Infra Project’s Open Optical and Packet Transport (OOPT) project group [83].

The benefits of the targeted SDN architecture lie in the system design, built upon three principles.

- Technology-domain controllers. The idea is to separate the control per technology concerns, then drive the different particularities of each technology with specific solutions suited to them. This separation of concerns also enables a higher scalability of the solution; in fact, in case a transport segment is divided among different administrative domains, multiple domain SDN controllers for the relevant transport segment might be included in the hierarchy, in a flexible way
- Technology-agnostic SDTN controller, abstracting the complexity below by offering a single-entry point for programmability of the overall transport network. This entry point is accessed by the TN management domain's consumers, which includes management functions from the rest of the management domains (see Figure H-2).
- Seamless integration of network slicing features. These capabilities are to be provided by the Transport Network Slice Controller (T-NSC), which is an add-on component of the SDTN controller, as shown in Figure H-19. This ensures that the SDTN controller can work with slicing and non-slicing services, by simply making use of or bypassing the T-NSC, respectively.

In the following, we describe the journey towards the goal scenario: having the E2E, open transport SDN architecture depicted in Figure H-19 up and running in commercial networks. We follow a bottom approach for describing the approach taken.

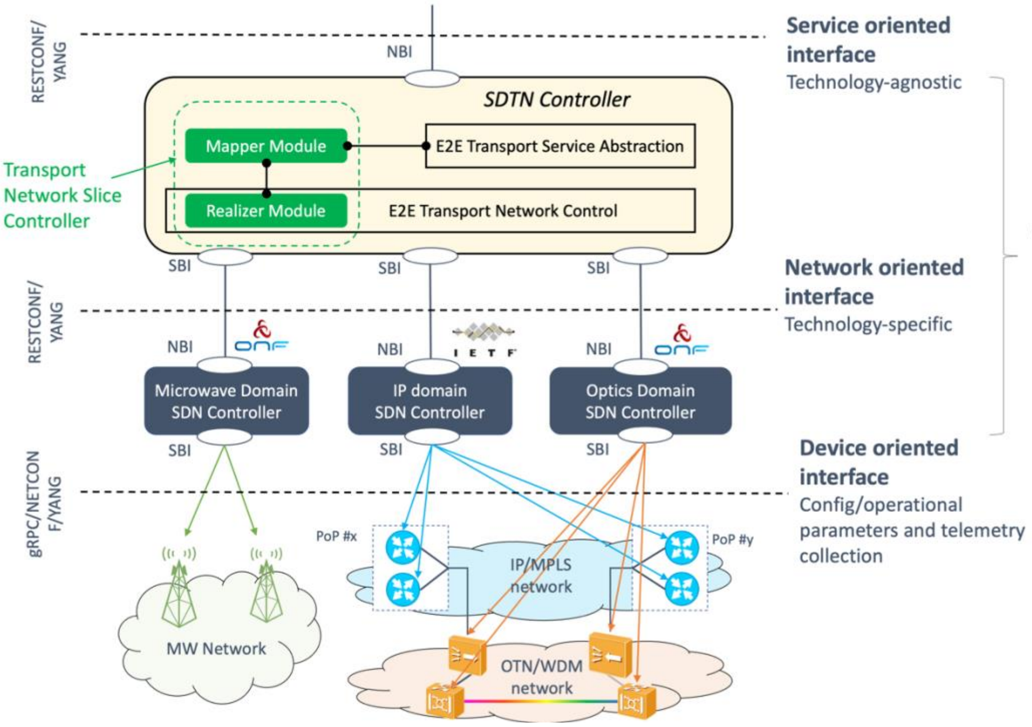

RESTCONF/ YANG
NBI
SDTN Controller
Mapper Module
E2E Transport Service Abstraction
Realizer Module
E2E Transport Network Control
Transport Network Slice Controller
Service oriented interface
Technology-agnostic
TN Management Domain
SBI
SBI
SBI
RESTCONF/ YANG
Network oriented interface
Technology-specific
NBI
ONF
NBI
IETF
NBI
ONF
Microwave Domain SDN Controller
IP domain SDN Controller
Optics Domain SDN Controller
SBI
SBI
SBI
gRPC/NETCONF/YANG
Device oriented interface
Config/operational parameters and telemetry collection
PoP #x
PoP #y
IP/MPLS network
MW Network
OTN/WDM network

Figure H-19. Transport SDN architecture.

In the lower part of the architecture, we see the device-oriented interface, which corresponds to the SouthBound Interface (SBIs) of the individual domain SDN controllers. As with any programmatic interface, the SBI is composed of selecting a protocol to transfer the data, and YANG data models to define how the message is formed. In all the domain SDN controllers, the SBI uses NETCONF [84] protocol, the main protocol for device management operations. Further details are provided below:

- **IP domain SDN controller: SBI**. For the IP/MPLS segment, the solution is based on a single multi-vendor IP domain controller, charged with configuring the Layer-2 and Layer-3 network elements. The management of this equipment is done through the SBI, which bets on declarative configuration and model-driven operations using device YANG models, such as those available in OpenConfig [85]. In addition to NETCONF/YANG, other protocols are considered in the SBI of an IP domain SDN controller, including (i) BGP-LS, to retrieve link-node topology of the IP/MPLS networks [86]; (ii) PCEP, to support Traffic Engineering [87]; and (iii) gRPC, to collect monitoring data [88].
- **Optics domain SDN controller: SBI**. For the optical segments, there is no way on having a ’one-size-fits-all’ SDN controller. The reason is that transport DWDM networks are highly implementation dependent, with no practical interoperability at optical level. Having a programmatic NETCONF/YANG based SBI would require the disaggregation of existing optical transceivers and line-side components.
- **Microwave (MW) domain SDN controller: SBI**. Though the number of hops from end links to the fiber aggregation point is progressively shorted (especially with the new Integrated Access Backhaul solutions [89]), it is still common for operators to have multi-vendor aggregation paths in their MW underlay. The operation and configuration of these paths is rather manual and static, using the proprietary interfaces from vendor-specific network management systems. The goal of this workstream is to avoid the integration complexity and scalability burdens of this approach (i.e., with OSS needed to maintain multiple interfaces) by introducing an SDN controller, which is a vendor-agnostic configurator of the MW network. To that end, standard communication protocol and YANG device models are being considered on the SBI, turning it into a programmatic interface.

Going up in the architecture, we arrive to the network-oriented interface, which permits the invocation of service fitting each technology underneath. This interface corresponds to the NorthBound Interface (NBIs) of the per-technology domain SDN controllers. **The Domain SDN controller: NBI** solution allows each controller to offer the following capabilities to the SDTN controller: (i) a vendor-agnostic provisioning interface, to request for the creation/deletion/modification of connectivity services; (ii) per-OSI layer topology and network inventory information; and (iii) active monitoring of network status, e.g., traffic statistics and event notifications. Unlike the SBIs, built with NETCONF operating over device YANG models, the NBIs bet on the use of RESTCONF [90] with network YANG models. On the one hand, the choice of RESTCONF (HTTP-based protocol) permits the reuse of all tooling around the REST interface, which is today the industry norm for non-device-oriented configuration management operations. On the other hand, the network YANG models provide abstract representation of relationships between multiple devices, including topology and connectivity services. Table H-4 captures the network YANG models that are for consideration at the domain controller NBIs.

Table H-4. Overview of network YANG models for the NBI of per-technology domains SDN controllers

Domain Controller	NBI models	Descriptioj
IP controller	L3SM [91] L2NM [92] L3NM [93] L2NM [94]	These models are used for the provisioning of L2/L3 connectivity services. They describe a VPN service from the customer (LxSM) or network operator (LxNM) viewpoint

	TE [95] TE Topology [96]	These models allow to manipulate Traffic Engineering (TE) tunnels. There exist extensions to work with a desired technology (e.g., MPLS RSVP-TE tunnels, Segment Routing paths).
Optics controller	T-API [97]	It is the model with higher market adoption. It includes technology-specific information from each transport layer, including DSR/Ethernet, OTN/ODU and photonics media.
Microwave controller	OpenBackhaul model set [98]	This repository captures the set of models for the microwave segment. They are extensions from those captured in ONF Core Information Model [99].

Finally, consuming these NBIs, there is the SDTN controller. This controller orchestrates the domain specific capabilities to provide real-time control of multi-layer and multi-domain transport network resources. The efforts on this workstream go in three directions:

- **E2E SDN controller: internal logic**. The focus here is on the implementation of the E2E Transport Network Control block of the SDTN controller. The scope of this module can be summarized into five functionalities: (i) E2E control across the different domains, by coordinating the disparate technologies through their corresponding SDN domain controllers; (ii) per-layer E2E visualization, i.e., per-layer topology composition; (iii) stateful control of provisioned network services; (iv) multi-layer Path Computation Engine (PCE), which has the role of computing paths across multiple technologies based on the per-layer topology composition; and (v) service binding to transport resources, which enables the controller to obtain the best Traffic Engineering (TE) connections for a given transport connectivity service. An example of the functionality described in (v) is as follows: for a VPN having certain bandwidth and latency constraints, compute the set of Label Switched Paths (LSPs).
- **E2E SDN controller: NBI**. The focus here is on the E2E transport network abstraction block, in charge of exposing an abstracted topology view of the network resources and the available set of network services to SDTN consumers through a unified NBI. There is no solution yet in the standards, although quite close cooperation between Telecom Infra Project's OOPT [83] and IETF's TEAS [100]. exists in this regard. Ongoing discussions reveal the idea to use LxSM models (see Table H-4) as a starting point for the NBI implementation.
- **Slicing-aware E2E SDN controller**. This consists of incorporating the T-NSC, which will have the awareness of slicing at the transport layer. Though there is not yet a common view of what T-NSC represents, the telco industry agrees on the need (i) to align the T-NSC concept with the IETF Slice Controller developed by the IETF TEAS's network slice design team [101], and (ii) to implement the T-NSC as an additional component of the SDTN controller. For the first point, the focus is to align the T-NSC with the use cases [102] and YANG network models [103] worked out for the IETF Slice Controller. For the second point, we foresee a solution similar to the one represented in Figure H-19. As seen, the T-NSC might be built out of two separate modules: the mapper and the realizer. The mapper module is responsible for collecting the customer-facing view of the TN slice for further processing the TN slice request. Thus, this module would integrate the customer-facing view on the provider view for triggering configuration, control and management actions. On the

other hand, there is the realizer module, which is in charge of coordinating different actions on a number of domain SDN controllers for effectively creating the TN slice, according to the original customer request. Integrated in the E2E Transport Network Control Abstraction block (see Figure H-19), this module would manage the workflows for the TN slice provision, as well as for its life cycle.

Putting all the discussed solutions in the technology radar shown in Figure H-11, we observe the following:

- The deploy ring includes **IP domain SDN controller: SBI** and **optics domain SDN controller: SBI** solutions. With the emergence of the first Rel-15 commercial networks, some operators have recently started the deployment of SDN technology in transport networks, although not at large scale yet. For now, it is limited to IP/MPLS and optical/DWDM segments, and highly coupled with specific use cases.
- The test ring covers **MW domain SDN controller: SBI**, **Domain SDN controller: NBI** and **E2E SDN controller: internal logic** solutions. The operators are currently conducting trials on features from these three solutions, which are expected to be available in the medium term.
- The explore ring captures the **E2E SDN controller: NBI** solution (for the integration of transport SDN fabric with the rest of OSS assets) and the **Slicing-aware E2E SDN controller** solution (to integrate the slice semantics in the transport SDN fabric).

According to the above rationale, it is clear that to get the targeted transport SDN architecture up and running in carrier networks, operators need to follow a staged approach where the full, E2E integration of slicing features constitutes the final stage. However, this does not mean we cannot have TN slicing meanwhile; indeed, as outlined in Section 7.1, there are multiple transport technologies available for slicing realization. What this really means is that the operator may not be able to have automation and programmability capabilities in the short and medium term when allocating and operating transport slices. The lack of SDN may force the operator to go for static provisioning and management operations, from the E2E management domain to technology-specific management systems, using traditional Command Line Interface (CLI) solutions with ad-hoc extensions to avoid vendor lock-in.

8 OSS Domain

Finally, for the OSS domain, the technology radar addresses two different dimensions: OAM and capability exposure. On the one hand, the *OAM* dimension refers to the set of activities related to network slice life cycle management. On the other hand, the *capability exposure* dimension touches on the need for providers to make network slice capabilities available for consumption to B2B customers, through easy-to-use service APIs.

8.1 OAM

The life cycle of a network slice is articulated into four different phases: preparation (slice design, slice on-boarding and network set-up), commissioning (slice instantiation and configuration), operation (slice activation, modification, reporting/supervision and de-activation) and decommissioning (slice termination) [17]. The operator's main role is to manage the life cycle of all slices running in the network. In this endeavor, the operator makes use of OAM tools available in the OSS layer, as illustrated in Figure H-2.

This section outlines the main solutions that enable the OSS to conduct network slice life cycle management. Their position in the radar (see Figure H-11) makes it clear that the trend is to evolve towards a fully digital OSS stack, wherein design, data management, assurance, and orchestration processes all need to be aligned and integrated across physical, virtual and cloud assets. This journey needs to be accompanied with new levels of automation, extending them further up the stack, into the E2E service and network management domain.

The first stop of the journey is the as-is ring of the radar. Today's OSS consists of isolated, monolithic systems passing through specialized applications that require complex integration. In addition, automation is quite limited (only present in specific parts of the OSS, and not covering all the management domains) and with many silos (part of the automation is often developed as standalone scripts used by separate domain teams, with each team creating and managing automation using its own environment). However, with the edge computing and 5G SA technologies just around the corner, most of the operators have already integrated a MANO stack in their OSS, using SOL005 [25]. Apart from the monetization of Infrastructure as a service (IaaS) services, the **MANO stack** allows operators to move carrier-grade functions and services to the cloud, leveraging NFV technology. This is critical for a cost-efficient delivery of some NSA slicing solutions, like the ones represented in Figure H-12.C–E, where there is a need to deploy multiple instances of the same network functions.

The second stop is the deploy ring of the radar. Coinciding with the early rollout of 5G SA networks and the commercialization of the first B2C slices, operators need to integrate slicing logic into their OSS. The short-term focus is on basic slice provisioning. In this regard, the following solutions are needed.

- **NST/NSST**. The Network Slice Template (NST) and Network Slice Subnet Template (NSST) are pre-configured service descriptors that help create a blueprint to ease replicability (i.e., design once, deploy everywhere) of offered network slices and slice subnets.
  For example, the NSST corresponding to a 5GC slice could include the following information: (i) the type of services that the 5GC slice supports, (ii) the 5GC slice topology, and (iii) the 5GC slice placement policy. Based on the fact that a network slice subnet can be deployed as an ETSI NFV network service (see Figure H-1), the information given in (ii) is a pointer to the corresponding Network Service Descriptor (NSD), while the information captured in (iii) is the specification of the PoP type where individual network service components can be deployed, and their affinity/anti-affinity rules.
  For the construction of the NST, the approach is similar, but including pointers to the NSSTs.
  The design, development, testing, and validation of NSTs/NSSTs/NSDs, and their subsequent onboarding to the catalogs, are activities that formally belong to the network slice preparation phase.
- **Slice NRM fragment**. As seen in Figure H-1, the 3GPP information model for network slicing is complex, with a high number of classes and different containment-naming relationships across them. Furthermore, this model is in continuous evolution, as long as the work in 3GPP SA2 and GSMA GST/NEST evolves. For this reason, it is preferable to have a lite version of the slice NRM fragment in the short term. This lite version may contain (i) all classes, except the EP_Transport IOC, which is intended for Rel-16; (ii) simple relationships across them, leveraging as much as possible on 1:1 mapping; and (iii) a limited number of attributes in the ServiceProfile and SliceProfile constructions. In these constructions, only the functionality and performance related attributes that are needed to provision 5GC type A2 slices (B2C market) will be implemented.

In case of selecting different vendors for the NSMF (in charge of interpreting ServiceProfile attributes) and NSSMF (in charge of interpreting SliceProfile attributes), the operator must ensure the compatibility of the lite slice NRM fragment with these NSMF and NSSMF solutions.

- **Baseline Decision Engine**. The Decision Engine is the OSS component in charge of performing the feasibility check procedure (see Figure H-2). In network slicing, this procedure makes use of three input data: (i) the service requirements that the slice must support, captured in the ServiceProfile; (ii) the NST, stored in the catalog; and (iii) the resource and network status, stored in the inventories. As detailed in Section 2.4, the feasibility check is a two-step operation. If the outcome of this operation is 'feasible', the decision engine uses internal policies and optimization algorithms to decide on the placement and resource allocation of the requested slice, and sends out the decision to the NSMF, which enforces it. The format of this decision is as follows: instantiate a new slice, configuring the NG-RAN slice with the radio resource allocation solution "X", and deploying the 5GC slice in this PoP "Y". For the instantiation of the 5GC slice, use the deployment flavor "Z" from the NSD referred in the NSST.

  In the short term, it is recommendable to use a baseline Decision Engine. This solution, captured in the deploy ring of the radar, assumes that the logic of the Decision engine is rather simple, built upon simple static rules or simply betting on trial-and-error approaches. The reason is that the variety of 5GC slice types is expected to be rather low, and all targeted for B2C market; therefore, there are no really a higher number of deployment options to choose from.

In the medium term, the upgrade to Rel-16 will allow operators to start commercializing network slicing for the B2B customers, with the 5GC slice types B1, B2 and B3. The shift from B2C to B2B market requires a much more rich-featured OSS than before, with better capabilities in terms of provisioning and scalability, and with the integration of first assurance mechanisms. On the one side, the enhancement in provisioning and scalability is due to the higher variety of slices, with tens of instances running in parallel. On the other side, the introduction of assurance features will allow operators to guarantee the QoS and contracted SLAs of each network slice, much more business-critical than those for B2C slices. The test ring of the radar captures the solutions that are needed to cope with the new OSS needs. These include

- **Complete Slice NRM fragment**. The new wave of slice offerings will be made available in the operator's service portfolio, with the publication of different NESTs. The wide variety of NEST parameters that can be configured by the customer prior to issuing service order, makes it necessary to evolve Slice NRM fragment. The main focus will be on the ServiceProfile and SliceProfile constructions, which need to extend their attributes to make them mappable to NEST parameters. Other minor enhancements are also expected, for example the addition of EP_Transport IOC and more flexible class relationships, as depicted in Figure H-1.
- **Advanced Decision Engine**. This solution is based on evolving the logic of the Decision Engine, with the integration of multi-objective policies and optimization algorithms. With many more slices running in parallel, each with completely different requirements, the operator's system becomes much more dynamic. If changes are too quick, the stability of the operator's network could be compromised. To avoid this, it is critical that the decision on slice placement and resource allocation (i) minimizes the probability of modifying the slice at operation time, and (ii) optimizes resource usage.

- **Baseline slice assurance**. The assurance represents the ability of the operator to retrieve management data from individual slices, using them as input for SLA verification. Examples of these data includes S-NSSAI level information on slice status (e.g., activated, de-activated), performance measurements and KPIs (e.g., UL/DL throughput, latency and packet loss rate) [104][105], notifications on fault events and alarms (e.g., threshold crossing) [106] and trace data.
  The operator may start with S-NSSAI level management data based on the aggregation of metrics collected from the 5GC (via NWDAF) and NFV MANO (via SOL005). This data will then be aggregated to check the health of the slice, verifying whether it meets the SLA requirements. In case of SLA violation, the operator will trigger corrective/remediation actions (e.g., capacity increase, re-configuration) on the corresponding slice. The number of actions available by then is expected to be limited, according to the experience the operators are getting from the trials.
  In this baseline assurance solution, it is assumed that the assurance group in the OSS (see Figure H-2) will include the 'data aggregation' and 'Service Quality Management' modules, but not the 'AI models and training' module.

Finally, in the longer run (explore ring in the radar), the operator focus will be on having an advanced slice assurance toolkit available in their OSS stack. To that end, operators are scouting solutions in two main directions. On the one hand, the incorporation of the RIC and the SDTN controller as new data sources, building on the two already in place (NWDAF and NFV MANO). To transform these quite different types of data into useful information for the Service Quality Management activities, the data aggregation framework may leverage on the Management Data Analytics Service (MDAS) functionality, reported by 3GPP in [107][108]. On the other hand, the full automation of the assurance pipeline (data collection and aggregation → insights → corrective /remediation actions), using zero-touch mechanisms such as those outlined in [109]. For the insights step, the 'AI models and training' module is to be used.

Once all the solutions in the radar have been discussed, one can wonder how the operators can assess the level of maturity of their OSS throughout this journey. This can be done using different OAM related KPIs, including scalability related KPIs (e.g., number of slice types, number of instances running in parallel), commissioning related KPIs (e.g., network slice instantiation time) and operation related KPIs (e.g., network scaling re-configuration time and scaling in/out time), among others.

8.2 Capability Exposure

Capability exposure represents the ability of a network slice provider to securely expose capabilities from its managed functions and services towards one or more authorized customers. These capabilities are made available for consumption through easy-to-use service APIs. When addressing customers with limited to no telco experience, it is important that the offered service APIs focus on the customer industry segment and hide the complexity of the underlying network. This requires the provider to abstract and combine low-level network APIs (e.g., 3GPP, ETSI and IETF stage 3 solutions) into customer-facing and intent-based APIs.

The importance of capability exposure in slicing environments has been already highlighted in [110][111]. When becoming a network slice provider, the operator can leverage this feature to offer NSaaS in multiple forms, from provider-managed slices (i.e., the provider is in charge of the slice operation, while the customer can merely use the network resources of the provider slice, without any further capability of managing or controlling it) to tenant-managed slices (i.e., the customer takes full control of the slice, and the provider just segregates the infrastructure as required for that purpose) [112], with some variants in between. By regulating the exposure, the operator can define the visibility and the degree of control the customer can take over the slice. Figure H-20 shows the logic behind the capability exposure concept.

In the following, we present the service APIs we foresee the operator may offer for capability exposure. For the sake of simplicity, these APIs have been clustered into API families.

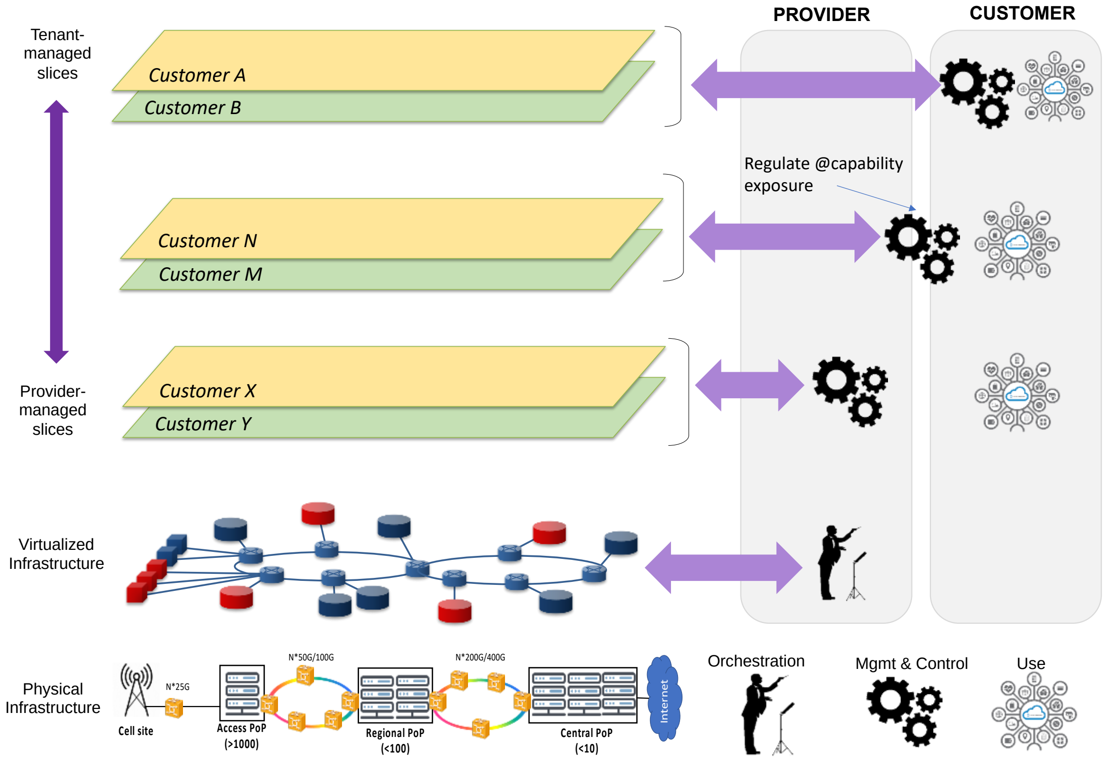


Figure H-20. Capability exposure in network slicing environments

Today's situation is as follows: only some operators have NSA slices, based on any of the technological solutions captured in Figure H-12. These slices are rather static, with everything configured beforehand; this means that no further customization is allowed from the customer side. In this context, the only actions that the customer can take are (i) to access the operator's portfolio, browse the service offering, select the NSA slices and issue a service order; (ii) to get high-level data on NSA slice status, for the purpose of SLA assurance; and (iii) to get information about the usage of the NSA slice and components that can be charged for. All these capabilities are provided with service APIs belonging to the **Accounting, Charging and Billing** API family. As seen, the possibilities with this as-is solution are quite limited, as it only permits provider-managed slices. However, this situation may radically change with the shift towards SA networks.

In the short term (deploy ring in the radar), the focus will be on premium B2C users, as noted in Figure H-13. For the provision of Type A2 slices, the operator will bet on internal NESTs. With the presence of NEF in the Rel-15 5GC solution suite, and the integration of the first Rel-15 features in their OSS systems, the operators are starting to launch a new API family: **Device Info**. It provides the customer with the ability to receive information related to one or more devices. This includes information on device subscription (e.g., subscribed

NSSAI), device location tracking (UE location and cell site), device mobility (handovers), device status (e.g., serving S-NSSAI, loss of connection, etc.) or CN type change (5GC to EPC coverage, and vice versa). This API family might offer the customer the possibility to explicitly query for this information (request-response mode) or be reported with notifications on subscribed events (subscribe-notify mode).

In the medium term (the test ring in the radar), the operators will publish the NESTs into their service portfolio, once 5GC capabilities necessary for building up B2B category slices are available. The use of NESTs provides an easy solution for the customer to request the allocation of a slice. Based on the service requirements captured in the NEST, the operator can then decide which B2B category slice is most appropriate, and provision it. In this time frame, it is also expected that customers gain experience with slicing technology, and hence want to retain more control of their slice. This requires the operators to increasingly expose further capabilities, with the definition of new API families. Below, there is a list of API families that some tier-1 operators have started to test and validate with some industry verticals and hyperscalers.

- **Edge computing**: these service APIs are available for those customers that want to extend their allocated slices with third party applications, with these applications hosted by the telco edge cloud. In a nutshell, this API family provides three main capabilities: (i) edge node discovery capability, which allows the customer to discover the set of edge nodes available in a certain region; (ii) edge node profile capability, whereby the customer can get information on capabilities and supported features of a given edge node, so as to check whether this node is valid for hosting the application; (iii) application resource allocation capability, that allows the customer to request the operator to deploy the application on a given edge node.
- **Device configuration**: provides the customer with the ability to register a device into the network slice, and to update device subscription. Based on the subscription information received by the customer, the operator updates the Unified Data Management (UDM) accordingly. This API family is quite useful for (industrial) IoT scenarios, where B2B customers want to retain control of the configuration of their devices. For further information on UDM functionality, see Figure H-6 and [10].
- **Network slice management data**: with this API family, the customers can subscribe to receive management data related to their allocated slices, at per S-NSSAI level. The individual customers might also choose how they want to consume these data, according to their preferences. For example, regarding performance management data, a customer could specify the KPIs to be informed, the batch format and the reporting period.
- **QoS control**: provides the customer with the ability to set and modify quality for a slice (e.g., maximum latency, guaranteed throughput, maximum admissible packet rate), on demand. The operator captures the customer-triggered request and routes it through the NEF down to the PCF, which will ultimately set/modify the 5QI associated with the in-slice PDU session(s).
- **Traffic influence**: provides the customer with the ability to modify the connection policies of UEs attached to the slice, in terms of how the traffic flows. For example, if the customer has deployed a 3rd party application in a specific edge node, with this API family the customer can then request the re-routing of the in-slice packet flows to this edge node.

Finally, in the longer run, the integration of Rel-17+ features into the network and OSS layers will allow operators to unleash the full potential of NSaaS. Examples of API families that operators foresee to make available in five-to-seven years time are captured below.

- **Slice day-2 configuration**: allows the customer to retain full control of slice day-to-day operation, which is the ultimate realization of tenant-managed slices. With this API family, the customer could request topology changes and resizing (e.g., scaling in/out) on the slice, based on the processing of alarms and KPIs received.
- **Customer-defined slice composition**: allows the customer to request a slice à la carte. To enable this feature, the operator shall define a marketplace with different functions and applications. The customer should be able to browse this marketplace, design an E2E slice by connecting selected functions/applications (following a 'plug-and-play' approach), and order its provisioning. This is a giant step forward, and requires thinking of novel ways of designing slices and capturing their service requirements, beyond today's NEST approach.

9 Conclusions

As 5G standards get completed and commercial products mature, operators shall demonstrate their ability to efficiently combine available technology solutions to deliver network slices for different customers, with a number of these slices being short-lived and provisioned on demand. Network slicing may allow network operators not only to achieve relevant CAPEX and OPEX reductions in their managed network infrastructure, but also to significantly enrich their portfolio with innovative service offerings, which is a key differentiator in the highly competitive environment the provision of network services has become today.

Network slicing is an E2E solution that has an impact on all subsystems, including the different technology domains (including RAN, CN and TN) and the operational systems (the OSS). The main problem is that the degree of maturity of slicing support is quite different in these subsystems. Indeed, while the CN is the technology domain that currently supports the most advanced slicing capabilities, the TN is still in an early phase, with RAN domain sitting in between. In the case of the OSS, in charge of the management of these three technology domains and of the orchestration activities across them, we can find different paces of technology evolution.

According to the above rationale, it is clear that it may take several years for operators to prepare their systems to fully harness the power of network slicing, and become able to support NSaaS. Awaiting this moment to start commercializing and monetizing network slicing is not an option, either for operators (who cannot offset the costs associated with the introduction of slicing capabilities into their assets) or for customers (which may not need the full set of capabilities for their use cases from the very beginning). In this context, what the industry demands is a phased-based rollout of slicing, starting with early-stage solutions and outlining a vision of what is desired in the longer run to guide progress and focus.

In this article, we have presented a radar with the mission to help industry in defining this rollout. This radar captures a complete landscape of network slicing solutions, linking these solutions to different timelines based on their technical viability and market demands. In addition to this timing, the radar has also outlined the dimensions that have an impact on the usability (how and where) of these solutions, across all operator managed domains. In the RAN domain, network slicing solutions have been discussed based on functional aspects (e.g., disaggregation and O-RAN integration), radio resource allocation and penetration in the operator's cell footprint. In the CN domain, discussions have been around the fulfilment of isolation and customer requirements of network slice customers, resulting in the use and combination of different 5GC functions, with different profiles. In the TN domain, solutions have been articulated around the technology capabilities available in the underlay WAN, complemented with the automation and programmability capabilities brought by SDN technology. Finally, in the OSS domain, aspects related to network slice OAM and capability exposure have been addressed, with a focus on provisioning and assurance activities.

The network slicing solutions captured in the radar have been analyzed based on their timeline, together with the above dimensions in Section 4, Section 5, Section 6, Section 7 and Section 8. In order to provide the reader with a common reference system to interpret the content of this radar, we started this work by providing the technical background of network slicing (Section 2) and its impact on individual technology domains (Section 3).

We are confident that the radar presented and discussed in this article will be a valuable reference for operators to outline their network slicing rollout plan within their service footprint. In fact, each operator can go through the radar and decide which specific solutions they want to activate in the different domains (CN, RAN, TN and OSS) and how to combine them, so that the resulting E2E capabilities can meet the service requirements of the targeted customers. This work has not entered into these decisions, which are entirely up to each operator, and subjected to certain market and business driven factors. In other words, this article provides the input material that an operator needs for the definition of a network slicing rollout plan (e.g., description, analysis, and profiling of network slicing solutions, in terms of timing and features), but with no details on how to execute this plan. Getting to these details would require going through two activities that are out of the scope of this article, and therefore not addressed in the discussion of this work:

- First, the definition of a well-defined methodology for the design of a robust rollout plan. The radar captures all available solutions and categorizes them into different dimensions and timelines; however, it does not provide clues on how an operator shall combine them in production networks. Examples of reference methodologies to this end can be found in [4]-[6]. These technology analyst reports provide high level recommendations and good practices for operators to design their individual rollout plans for network slicing.
- Secondly the assessment of proposed network slicing solutions, so that their main advantages and disadvantages can be outlined in advance. This needs to be done individually (i.e., per domain-specific solution) and E2E (i.e., based on selecting combinations of RAN+TN+CN solutions). One of the aspects we consider for future work is the evaluation of these solutions, using either simulation work or PoCs, depending on the maturity of each solution.

Besides operators, the outcomes of this work may be of relevance for other audiences, including vendors, verticals and researchers. On the one hand, the solutions captured in this radar can help the vendors to consolidate the feature roadmap in their products. On the other hand, the verticals can use this radar to understand what slicing capabilities may be available for consumption, and by when. This knowledge should be important for them, especially to manage expectations in terms of implementable use cases. Finally, this work provides a reference for researchers to keep working on the validation of E2E slicing solutions, using experimental setups that mimic the limitations and conditions of real-world carrier networks.

Author Contributions

Conceptualization, J.O.-L., P.A.; Investigation, J.O.-L., L.M.C.; Methodology: J.O.-L.,

P.A.; Writing—original draft, J.O.-L., P.A., L.M.C., J.F., D.R.L.; Supervision, P.A., J.F., D.R.L. All authors have read and agreed to the published version of the manuscript).

Funding

This work is partially supported by the H2020 European Projects 5G-VINNI (grant agreement No. 815279) and 5G-CLARITY (grant agreement No. 871428). It is also supported by the Spanish national project TRUE-5G (PID2019-108713RB-C53). This research reflects the consortia views, but neither the consortia nor the European Commission are liable for any use that may be done of the information contained therein.

Institutional Review Board Statement

Not applicable

Informed Consent Statement

Not applicable

Data Availability Statement

Not applicable

Conflicts of Interest

The authors declare no conflict of interest. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the companies and institutions they represent.

Abbreviations

The following abbreviations are used in this manuscript.

5G	Fifth Generation
5GC	5G Core
5QI	5G Quality Indicator
AI	Artificial Intelligence
AMBR	Aggregate Maximum Bit Rate
AMF	Access and Mobility management Function
API	Application Programming Interface
ASIC	Application Specific Integrated Circuit
B2B	Business-to-Business
B2C	Business-to-Customer
BSS	Business Support System
CN	Core Network
CSMF	Communication Service Management Function
CU	Centralized Unit

DN	Data Network
DRB	Dedicated Radio Bearer
DSCP	DiffServ Code Point
DU	Distributed Unit
DWDM	Dense WDM
E2E	end-to-end
eCPRI	Enhanced Common Public Radio Interface
EPC	Evolved Packet Core
Flex-E	Flexible Ethernet
Flex-O	Flexible Optical Transport Network
FPGA	Field Programmable Gate Array
GBR	Guaranteed Bit Rate
GSMA	GSM Alliance
GST	Generic network Slice Template
IaaS	Infrastructure as a Service
IOC	Information Object Class
L4S	Low Latency, Low Loss, Scalable throughput
LSP	Label Switched Path
MANO	Management and Orchestration
MBR	Maximum Bit Rate
MDAS	Management Data Analytics Service
MPLS	Multiple Protocol Label Switching
MW	Microwave
near-RT RIC	near-Real Time RIC
NEF	Network Exposure Function
NEST	Network Slice Type
NFV	Network Functions Virtualization
NFVO	NFV Orchestrator
NG-RAN	Next Generation RAN
non-RT RIC	Non-Real Time RIC
NPN	Non-Public Network
NR	New Radio
NRM	Network Resource Model
NSA	Non-Standalone
NSaaS	Network Slice as a Service
NSCAF	Network Slice Access Control Function
NSD	Network Service Descriptor
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NSSAI	Network Slice Selection Assistance Information
NSSI	Network Slice Subnet Instance
NSSMF	Network Slice Subnet Management Function
NSST	Network Slice Subnet Template
NST	Network Slice Template
NWDAF	Network Data Analytics Function
OAM	Operation, Administration and Maintenance
OS	Operation System
OSS	Operations Support System

PCE	Path Computation Element
PCF	Policy Control Function
PDU	Packet Data Unit
PLMN	Public Land Mobile Network
PNI-NPN	Public Network Integrated NPN
PoC	Proof of Concept
PoP	Point of Presence
PRB	Physical Radio Block
RAN	Radio Access Network
RIC	RAN Intelligent Controller
ROADM	Reconfigurable Optical Add-Drop Multiplexer
RRM	Radio Resource Management
RU	Radio Unit
S-NSSAI	Single Network Slice Selection Assistance Information
SA	Standalone
SBA	Service Based Architecture
SBMA	Service Based Management Architecture
SD	Slice Differentiation
SD-WAN	Software Defined WAN
SDN	Software Defined Networking
SDO	Standards Development Organization
SDTN controller	Software Defined Transport Network Controller
SLA	Service Level Agreement
SMF	Session Management Function
SNPN	Standalone NPN
SR	Segment Routing
SST	Slice/Service Type
T-NSC	Transport Network Slice Controller
TAC	Tracking Area Code
TAI	Tracking Area Identifier
TE	Traffic Engineering
TN	Transport Network
TSN	Time Sensitive Network
UE	User Equipment
UPF	User Plane Function
URSP	UE Resource Selection Policy
VLL	Virtual Leased Line
VNF	Virtualized Network Function
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPRN	Virtual Private Routed Network
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
XR	Immersive Reality
YANG	Yet Another Next Generation

References

[1] GSMA, “The 5G Guide: A Reference for Operators”, White Paper, 2019 [Online]. Link

[2] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges”, in *IEEE Communications Magazine*, no. 55, vol. 5 pp. 80–87, 2017.

[3] GSMA, “Network Slicing-Use Case Requirements”, White Paper, 2018 [Online]. Link

[4] S.W. De Gimarlo, “Create Value and Drive Revenue with 5G Network Slicing Phased Approach”, *Gartner*, 2021. [Online]. Link

[5] J. Crawshaw, “Network Slicing Management: A Key Use Case for Service Orchestration”, *Omdia*, 2021. [Online]. Link

[6] Ericsson and D. Little, “Network Slicing: A go-to-market guide to capture high revenue potential”, Industrial Report 2021 [Online]. Link

[7] 3GPP TS 38.401, “NG-RAN; Architecture Description”, v16.7.0, October 2021 [Online]. Link

[8] F. Launay, “NG-RAN Network—Functional Architecture”, in *NG-RAN and 5G-NR: 5G Radio Access Network and Radio Interface*; Wiley: Hoboken, NJ, USA, pp. 1–29, 2021.

[9] O-RAN Alliance, “O-RAN Use Cases and Deployment Scenarios: Towards Open and Smart RAN”, White Paper, 2020. [Online] Link

[10] 3GPP TS 23.501, “5G; System Architecture for the 5G System (5GS)”, v17.2.0, September 2021 [Online]. Link

[11] O-RAN Alliance, “Control, User and Synchronization Plane Specification”, O-RAN.WG4.CUS.0-v06, 2021.

[12] O-RAN Alliance, “Management Plane Specification”, O-RAN.WG4.MP.0-v06, 2021.

[13] 3GPP TS 38.470, “NG-RAN; F1 General Aspects and Principles”, v15.6.0, October 2021 [Online]. Link

[14] 3GPP TS 29.561, “5G System; Interworking between 5G Network and External Data Networks”, v17.3.1, September 2021 [Online]. Link

[15] 3GPP TS 28.541, “Management and Orchestration; 5G Network Resource Model (NRM); Stage 2 and 3 “, v17.4.0, September 2021 [Online]. Link

[16] ETSI GR NFV 003, “Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV”, v1.6.1, March 2021 [Online]. Link

[17] 3GPP TS 28.530, “Management and Orchestration; Use Cases and Requirements”, v17.1.0, September 2021 [Online]. Link

[18] 3GPP TS 28.533, “5G; Management and Orchestration; Architecture Framework”, v17.0.0, September 2021 [Online]. Link

[19] ETSI GS ZSM 002, “Zero-Touch Network and Service Management (ZSM); Reference Architecture”, v1.1.1, August 2019 [Online]. Link

[20] ETSI ISG NFV, “NFV Release 4 Definition (Release Documentation v0.3.0)”, 2021. [Online]. Link

[21] GSMA PRD NG.116 v5.0, “Generic Network Slice Template”, June 2021. [Online]. Link

[22] GSMA, “From Vertical Industry Requirements to Network Slice Characteristics”; White Paper, 2018 [Online]. Link

[23] TM Forum, “TMF641 Service Ordering API User Guide”, v4.1.0

[24] 3GPP TS 28.531, “5G; Management and Orchestration; Provisioning”, v17.1.0, September 2021 [Online]. Link

[25] ETSI GS NFV SOL-005, “Network Functions Virtualisation (NFV) Release 3; Protocols and Data Models; RESTful Protocols Specification for the Os-Ma-Nfvo Reference Point “, v3.5.1, October 2021 [Online]. Link

[26] GSMA, “Operator Platform Telco Edge Proposal Version 1.0”, White Paper, 2020 [Online]. Link

[27] J. F. Lorca, E. Serna, M. Aparicio, A. Chaissagne and J. L. Esplá, “Telefonica Views on the Design, Architecture and Technology of 4G/5G Open RAN Networks”, White Paper, 2021 [Online]. Link

[28] 3GPP TS 23.503, “5G; Policy and Charging Control Framework for the 5G System (5GS); Stage 2”, v17.2.0, September 2021 [Online]. Link

[29] NGMN Alliance, “5G Smart Devices Supporting Network Slicing Version 1.1”, White Paper, 2020 [Online]. Link

[30] 3GPP TS 23.502, “5G; Procedures for the 5G System (5GS); Stage 2”, v17.2.0, September 2021 [Online]. Link

[31] G. Fen, G. Feng, J. Zhou and S. Quin, “Mobility Management for Network Slicing Based 5G Networks”, in *Proceedings of the 2018 IEEE 18th International Conference on Communication Technology (ICCT)*, vol. 55, October 2018, p. 291-296.

[32] 3GPP TS 28.622, “Telecommunication Management; Generic Network Resource Model (NRM) Integration Reference Point (IRP); Information Service (IS)”, v16.9.0, March 2021 [Online]. Link

[33] K. Nichols, S. Blake, F. Baker and D. Black, “Definition of the Differentiated Services Filed (DS Field) in the IPv4 and IPv6 Headers”, RFC 2474 [Online]. Link

[34] Coriant, “The role of OTN Switching in 100G and Beyond Transport Networks”, White Paper, 2015 [Online]. Link

[35] IA #OIF FLEXE-02.1, “IA Flex Ethernet 2.1 – Implementation Agreement”, 2019 [Online]. Link

[36] Viavi Solutions, “OIF Flex-E 2.1 – Flexible Use of Ethernet”, 2020 [Online]. Link

[37] GSMA, “5G Implementation Guidelines”, White Paper, 2019 [Online]. Link

[38] H2020 5G-VINNI Project: “5G Verticals INNovation Infrastructure” [Online]. Link

[39] H2020 5GROWTH Project: “5G-Enabled Growth in Vertical Industries” [Online]. Link

[40] H2020 5G-CLARITY Project: “Beyond 5G Multi-Tenant Private Networks Integrating Cellular, Wi-Fi and LiFi, Powered by Artificial Intelligence and Intent Based Policy [Online]. Link

[41] 3GPP TS 23.707, “Architecture Enhancements for Dedicated Core Networks; Stage 2”, v13.0.0, December 2014 [Online]. Link

[42] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang and N. Li, "5G Deployment: Standalone vs. Non-Standalone from the Operator Perspective”, in *IEEE Communications Magazine*, vol. 58, no. 11, pp. 83-89, November 2020.

[43] Spirent, “Building the New Telecom Innovation Pipeline”, Spirent EBook Series, Spirent: Crawley, UK, 2021 [Online]. Link

[44] J. Ordonez-Lucena, J. F. Chavarria, L. M. Contreras and A. Pastor, "The use of 5G Non-Public Networks to support Industry 4.0 scenarios" in *2019 IEEE Conference*

on Standards for Communications and Networking (CSCN), 2019, pp. 1-7.

[45] W. Y. Poe, J. Ordonez-Lucena and K. Mahmood, "Provisioning Private 5G Networks by Means of Network Slicing: Architectures and Challenges" in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1-6.

[46] 3GPP Liaison Statement, "LS on UPF Support for Multiple Network Slices (S2-2105240 / C4-212560), February 2021 [Online]. Link

[47] S. Balasubramanian (Nokia), "End-to-End Network Slicing in 5G System: 3GPP Standards Perspective", 2016. [Online]. Link

[48] Y.-H. Chai and F.J. Lin, "Evaluating Dedicated Slices of Different Configurations in 5G Core", in Journal on Computing Communications, vol.9, no. 7, pp. 83-89, July 2021.

[49] ETSI, "Harmonizing Standards for Edge Computing – A Synergized Architecture Leveraging ETSI ISG MEC and 3GPP Specifications", White Paper, vol. 36, 2020 [Online]. Link

[50] 3GPP TS 29.520, "5G; 5G System; Network Data Analytics Services; Stage 3", v17.4.0, September 2021 [Online]. Link

[51] 3GPP TS 23.288, "Architecture Enhancements for 5G System (5GS) to Support Network Data Analytics Services", v17.2.0, September 2021 [Online]. Link.

[52] 3GPP TS 29.526, "5G; 5G System; Network Slice Specification Authentication and Authorization (NSSAA Services); Stage 3", v17.2.0, September 2021 [Online]. Link

[53] GSMA PRD NG.113 v4.0, "5GS Roaming Guidelines", May 2021 [Online]. Link

[54] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena and S. Yoon, "Intelligent O-RAN for beyond 5G and 6G wireless networks", 2020, arXiv:2005.08374.

[55] O-RAN Alliance, "O-RAN Architecture Description", O-RAN.WG1.O-v03", 2020.

[56] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang and B. Ottersten, "Dynamic Spectrum Sharing in 5G Wireless Networks With Full-Duplex Technology: Recent Advances and Research Challenges", in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 674-707, 2018.

[57] C. -L. I, S. Kuklinskí and T. Chen, "A Perspective of O-RAN Integration with MEC, SON, and Network Slicing in the 5G Era", in *IEEE Network*, vol. 34, no. 6, pp. 3-4, November/December 2020

[58] O-RAN Alliance, "O-RAN Near-Real-Time RAN Intelligent Controller Architecture and General Aspects and Principles", O-RAN.WG3.E2GAP-v01.01", 2020.

[59] O-RAN Alliance, "Slicing Architecture", O-RAN.WG1.Slicing-Architecture-v05", 2021.

[60] O-RAN Alliance, "O-RAN O1 Interface Specification for O-DU", O-RAN.WG5.MP.0-v01.00", 2020.

[61] O-RAN Alliance, "O-RAN A1 Interface: Type Definitions 2.0", O-RAN.WG2.A1TD-v02", 2020.

[62] B. Briscoe, K. De Schepper, M. Bagnulo and G. White, "Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service Architecture", draft-ietf-tsvwg-l4s-arch-10 (work in progress) [Online]. Link

[63] KDDI and Samsung, "E2E Network Slicing PoC with RIC Radio Resource Management", O-RAN Alliance Plugfest [Online]. Link

[64] CPRI, "Common Public Radio Interface: eCPRI Interface Specification V2.0", 2019 [Online]. Link

[65] J. Herny, T. Sziget and L. M. Contreras, "DiffServ to QCI Mapping", draft-henry-tsvwg-diffserv-to-qci-04 (work in progress) [Online]. Link

[66] J. Kaippallimalil, Y. Lee, T. Saboorian, M. Shalash and U. Kozat, "Traffic Engineered Transport for 5G Networks", in *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2019, pp. 1-6

[67] T. Saad, V. Beeram, B. Wen, D. Ceccarelli, J. Halpern, S. Peng, R. Chen et al. "Realizing Network Slices in IP/MPLS Networks", draft-bestbar-teas-ns-packet-03 (work in progress) [Online] Link

[68] D. Jaksic, "Segment Routing in Service Provider networks", in *Proceedings of the Cisco Connect Event*, Rovinj, Croatia, 19-21 March 2021.

[69] Filsfils, "Segment Routing Architecture", RFC 8402 [Online]. Link

[70] D. Borsatti, G. Davoli, W. Cerroni and F. Callegati, "Service Function Chaining Leveraging Segment Routing for 5G Network Slicing" in *2019 15th International Conference on Network and Service Management (CNSM)*, 2019, pp. 1-6.

[71] M. Gramaglia, V. Sciancalepore, F. J. Fernandez-Maestro, R. Perez, P. Serrano and A. Banchs, "Experimenting with SRv6: a Tunneling Protocol supporting Network Slicing in 5G and beyond", in *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2020, pp. 1-6.

[72] Filsfils, "Stateless and Scalable Network Slice Identification for SRv6", draft-filsfils-spring-srv6-stateless-slice-id-04 (work in progress) [Online]. Link

[73] K. Katsalis, L. Gatzikis and K. Samdanis, "Towards Slicing for Transport Networks: The Case of Flex-Ethernet in 5G", in *2018 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2018, pp. 1-7.

[74] ITU-T G.708.1/Y.1331.1, "Flexible OTN Short-Reach Interfaces", April 2021 [Online]. Link

[75] S. Bhattacharjee et al., "Network Slicing for TSN-Based Transport Networks," in *IEEE Access*, vol. 9, pp. 62788-62809, 2021.

[76] IETF Deterministic Networking (DetNet) Working Group [Online]. Link

[77] MEF 70.1 Draft Release 1, "SD-WAN Service Attributes and Service Framework", [Online]. Link

[78] MEC 3.0 Proof of Concept (PoC) 133, "SD-WAN and 5G with Network Slicing" [Online]. Link

[79] ONF TR-522, "SDN Architecture for Transport Networks", 2016 [Online]. Link

[80] D. Ceccarelli, Y. Lee, "Framework for Abstraction and Control of Traffic Engineered Networks (ATCN)", RFC 8453 [Online]. Link

[81] Q. Mu, M. Boucadair, D. Lopez, C. Xie and L. Geng, "A Framework for Automating Service and Network Management with YANG", RFC 8969 [Online]. Link

[82] L. M. Contreras, Ó. González, V. López, J. P. Fernández-Palacios and J. Folgueira, "iFUSION: Standards-based SDN Architecture for Carrier Transport Network", in *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2019, pp. 1-7.

[83] Telefónica, Vodafone, MTN Group, Orange, Telia Company and Deustche Telekom, "Open Transport SDN Architecture", Telecom Infra Project Whitepaper [Online].

Link

[84] R. Enns, M. Bjorklund, J. Schoenwaelder and A. Bierman., "Network Configuration Protocol (NETCONF)", RFC 6241 [Online]. Link

[85] Openconfig, "Data Models and APIs" [Online]. Link

[86] H. Gredler, J. Medved, S. Previdi, A. Farrel and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) using BGP", RFC 7752 [Online]. Link

[87] J. P. Vasseur and J. L. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440 [Online]. Link

[88] gPRC, "A High Performance, Open Source Universal Remote Procedure Call (RPC) Framework" [Online]. Link

[89] C. Madapatha et al., "On Integrated Access and Backhaul Networks: Current Status and Potentials," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1374-1389, 2020.

[90] A. Bierman, M. Bjorjund, K. Watsen, "RESTCONF Protocol", RFC 8040 [Online]. Link

[91] S. Litkwoski, L. Tomotaki and K. Ogaki, "YANG Data Model for L3VPN Service Delivery", RFC 8049 [Online]. Link

[92] B. Wen, G. Fioccola, C. Xie and J. Jalil, "A YANG Data Model for Layer 2 Virtual Private Network (L2VPN) Service Delivery", RFC 8466. [Online]. Link

[93] S. Barguil, O.G. de Dios, M. Boucadair, L. Munoz and A. Aguado, "A Layer 3 VPN Network YANG Model", draft-ietf-opsawg-l3sm-l3nm-05 (work in progress) [Online]. Link

[94] S. Barguil, O.G. de Dios, M. Boucadair, L. Munoz, L. Jalik and J. Ma, "A Layer 2 VPN Network YANG Model", draft-ietf-opsawg-l2sm-l2nm-02 (work in progress) [Online]. Link

[95] T. Saad, R. Gandhi, X. Liu, V. Beeram and I. Bryskin, "A YANG Data Model for Traffic Engineering Tunnels, Label Switched Paths and Interfaces", draft-ietf-teas-yang-te-25 (work in progress) [Online]. Link

[96] X. Liu, I. Bryskin, V. Beeram, T. Saad, H. Shah and O.G. de Dios, "YANG Data Model for Traffic Engineering (TE) Topologies", RFC 8795 [Online]. Link

[97] ONF Transport API (T-API) Project [Online]. Link

[98] Openbackhaul Project [Online]. Link

[99] ONF TR-512 v1.4, "Core Information Model (CoreModel)", v1.4.1, 2018 [Online]. Link

[100] IETF Traffic Engineering Architecture and Signaling (TEAS) Working Group [Online]. Link

[101] A. Farrel, "Framework for IETF Network Slices", draft-ietf-teas-network-slices-04 (work in progress) [Online]. Link

[102] L. M. Contreras, "IETF Network Slice Use Cases and Attributes for Northbound Interface of IETF Network Slice Controllers", draft-contreras-teas-slice-nbi-05 (work in progress) [Online]. Link

[103] X. Liu, "IETF Network Slice Service YANG Data Model", draft-liu-teas-transport-network-slice-nbi-yang-05 (work in progress) [Online]. Link

[104] 3GPP TS 28.552, "5G; Management and Orchestration; 5G Performance Measurements", v17.4.0, September 2021 [Online]. Link

[105] 3GPP TS 28.554, "5G; Management and Orchestration; 5G end to end Key Performance Indicators (KPIs)", v17.4.0, September 2021 [Online]. Link

[106] 3GPP TS 28.545, "5G; Management and Orchestration; Fault Supervision (FS)", v17.0.0, September 2021 [Online]. Link

[107] 3GPP TR 28.809, "5G; Management and Orchestration; Study on Enhancement of Management Data Analytics (MDA)", v17.0.0, September 2021 [Online]. Link

[108] 3GPP TS 28.104, "5G; Management and Orchestration; Management Data Analytics (MDA)", v0.2.0. September 2021 [Online]. Link

[109] ETSI GS ZSM 009-2, "Zero-Touch Network and Service Management (ZSM); Closed-Loop Automation; Part 2: Solutions, v0.7.0 draft, October 2021. [Online]. Link

[110] GSMA, "An Introduction to Network Slicing", White Paper, 2017 [Online]. Link

[111] NGMN Alliance, "Security Aspects of Network Capabilities Exposure in 5G", v1.0, White Paper, 2018 [Online]. Link

[112] L. M. Contreras and D. López, "A Network Service Provider Perspective on Network Slicing" [Online]. Link

Part V

Conclusions and Remarks

Conclusions

This thesis has focused on the design and validation of solutions for network slicing management and orchestration in multi-domain environments, with applicability in public and private 5G network scenarios. In this last chapter, the main findings of this thesis are summarized. Furthermore, it also provides avenues for future research arising from the presented conclusions.

1 Main Findings

The work in this dissertation has been articulated into three objectives.

The first objective was to design a standards-compliant system architecture for multi-domain network slicing. The most remarkable conclusions derived from this objective are:

- Isolation in network slicing is a multi-faceted problem, articulated into three separate dimensions: i) performance, ensuring that the SLA is always met on each network slice instance, regardless of workloads or faults from other running instances; ii) security, ensuring that any type of intentional attack occurring in one slice instance has no impact on any other running instance; iii) management, ensuring that each slice instance can be operated as a separate network partition, with an independent lifecycle management. These dimensions impose requirements on the operator side, when acting as network slice provider, as follows. As for the performance dimension, the operator shall perform resource segregation, splitting the infrastructure into a set of logically partitioned resources (resource chunks), and then allocating them to different slices. In relation to the security dimension, the operator shall provide means to guarantee the management data is securely stored, and accessible only for the authorized customer. Depending on the criticality of the data, the operator can decide on different security management solutions for data protection, including integrity, confidentiality, and privacy protection solutions. Finally, regarding the management dimension, the operator shall provide means to allow for multi-tenancy support (controllability separation in the network), based on the definition of separate yet tailored management spaces for different customers. Each management space should be provisioned with only the configuration and monitoring capabilities that the customer needs to consume from their network slice instances.
- This thesis has proposed a system architecture for network slicing management and orchestration. The solution design leverages the standard frameworks for SDN and NFV technologies, as defined by ONF and ETSI ISG NFV, extending them so that their combined use provides expected slicing capabilities in multi-domain environments, especially in what refers to "isolation in shared multi-provider infrastructures". The system architecture covers the different phases of lifecycle management, including preparation and commissioning phase (Paper B), as well as

operation phase (Paper A).

- In the preparation phase, the network slice instance does not yet exist. It mostly covers design-time work, with the elaboration of the different descriptors and their onboarding to the catalogs, to allow for catalog-driven service orderings later on. This thesis has proposed a solution design for network slice descriptor. Similar to NSDs and VNFDs in NFV environments, a network slice descriptor is a machine-readable template that provides a full characterization of the slice, with information on expected behavior, so that system architecture can know how to orchestrate it through its lifecycle. The proposed solution considers the following fields for the descriptor: i) slice topology, expressed as an ordered chain of technology-agnostic composable nodes, each providing a particular network functionality; ii) slice network requirements, including performance and functional requirements; iii) slice temporal requirements, with information on start and end dates of the slice instance, with activation/de-activation scheduling in between, if needed; iv) slice geolocation requirements, specifying the service footprint; and v) slice operational requirements, including slice priority level, metrics to be reported, accounting data, and information to be made available to the customer (for visibility and monitoring purposes). The fields i), ii) and iv) assist the system architecture at provisioning time, when commissioning the slice instance. The fields ii), iii) and v) are relevant at operation time, to keep the slice up and running, conformant to SLA specifications.
- In the commissioning phase, the network slice instance is deployed. In the case of Network slice as a Service (NSaaS), the provisioning workflow is triggered upon a catalog-driven service order. This thesis has provided detailed insight into this workflow, arranging the steps into various stages. The first stage is service ordering, whereby vertical browses the catalog, selects one slice descriptor and requests for the allocation of a slice. Upon capturing this service order, the operator executes the network slice resource description, consisting in translating network slice descriptor parameters into resource requirements, mapping then NSDs/VNFD parameters and other infrastructure configurations. Then, the operator executes the admission control stage (feasibility check), followed by the optimization & resource reservation, where it decides the actual provision solution and reserve resources. This information is finally reported to the management plane, which proceeds with the actual execution of network slice provisioning. For each of the stages, the input/output information, the steps involved, as well as the role of the participant system architecture components have been specified.
- In the operation phase, the network slice instance is up and running. The slice instance executes on top of an infrastructure, which is shared with other slice instances. Additionally, during this phase, one or more slices can be resized (e.g., scaled in/out, see Figure I-11) to cope with time-varying traffic demands, modifying their capacity accordingly. Keeping slice instances isolated from each other in such an environment is not easy. If the operator only assigns dedicated resources to network slices, their required performance levels are always met at the costs of preventing slices to share resources. This leads to over-provisioning, an undesired situation bearing in mind that the operator has a finite set of assigned resources. One way to resolve this issue is to permit resource sharing, although this means slices are not yet completely decoupled in terms of performance. Thus, it is required to design adequate resource management mechanisms that enable resource sharing among slices, when necessary, without violating their required performance levels. For wireless resources, this can be done with appropriate intra-slice and inter-slice strategies. For compute resources,

the operator can accomplish the sharing issue using policies similar to those used in cloud computing solutions, together with Enhanced Platform Awareness solutions.

The second objective aimed to prototype solutions related to the designed system architecture, and validate them in relevant scenarios. The most remarkable conclusions from this objective are:

- For the network slice descriptor proposed in the first objective, a solution has been prototyped. The prototype implements all the fields the original design. In particular, it leverages GST/NEST parameters to specify the *network, temporal and geolocation requirements*. As for the *slice topology*, the prototype references corresponding NSDs, and exposes service access points to the customer, for those cases where the customer wants to attach 3rd party applications. Finally, for the *operational requirements*, the prototype includes monitoring information, specifying data sources and metrics collection method (threshold-based alarms or periodic reports).
  The prototype described above addresses the pain points of state-of-the-art solutions. It overcomes the main limitations and operational complexities existing in 5G-PPP projects (5G-TRANSFORMER, SLICENET and 5GTANGO) and open-source communities (OSM and ONAP), with a solution that builds upon progress in GSMA with respect slicing profiling, and that is programmed using simple yet effective modelling practices from TM Forum's Information Framework (SID).
- For the system architecture proposed in the first objective, a solution has been prototyped. The prototype implements an orchestration stack based on the combined use of OSM and Openslice. Both are open-source orchestration frameworks, with different scope. On the one hand, OSM is a reference implementation of NFV MANO, and is responsible for the lifecycle management of network slices at the virtualized infrastructure layer. On the other hand, Openslice features a Service Orchestrator, taking care of the network slice lifecycle management at the application layer. Openslice hosts slice semantics, and consumes OSM's NBI (SOL005) to instruct the deployment and operation of virtualized network slice components (i.e., VNFs) when and where needed.
- Unlike OSM, Openslice is an orchestration suite resulting from the research of this PhD thesis. Openslice follows a disaggregated and service-based architecture, formed of modular components that produce and consume APIs through a ActiveMQ service bus, policed by a service registry based on Consul. In a nutshell, Openslice provides three types of capabilities: customer-facing capabilities, resource-facing capabilities, and federation capabilities. Regarding the customer-facing capabilities, Openslice offers APIs to 3rd parties, for them to gain access to the system. These 3rd parties include: i) application developers, who onboard applications and VNFs into the marketplace, so that operators can use them to design slice offerings; and ii) vertical customers, who browse the catalog, selects among available network slice descriptors and issue corresponding service orders. In relation to resource-facing capabilities, Openslice include modules scoping provisioning (order management, inventory & catalog management, instantiation, and configuration plus activation) and assurance (e.g., performance management, fault management and policy management). These modules are responsible for the lifecycle management of network slices, interacting with OSM when necessary. Finally, the federation capabilities refer to Openslice's ability to interact with orchestration stacks from other administrative domains (either public or private), so as to allow multi-domain slicing. To make it happen, Openslice exchanges information with the counterpart stack using the following TM Forum Open APIs: Catalog Management API (TMF633), Service Ordering Management

API (TMF641), Service Inventory Management API (TMF638) and Service Configuration & Activation API (TMF640).

- This thesis has shed light on the 'ways of working' of developed prototyped, across all lifecycle phases, by specifying different user stories and detailing their workflows. This includes i) network slice descriptor design and onboarding, ii) slice instance provisioning, upon service ordering; iii) slice operation, with focus on scaling.
- The hypothesis anticipated in the above-mentioned user stories have been validated in an 5G experimentation infrastructure: 5G-VINNI. 5G-VINNI is a large-scale, pan-European network that leverages 5G technologies to assemble a pre-commercial facility for vertical use case experimentation. It integrates multiple nodes across Europe, including Spain, Greece, UK, Norway, among others. As for this thesis, a Proof of Concept (PoC) involving Spanish and Greek nodes was executed, to validate hypothesis in multi-domain environment. The results of this PoC were submitted to ETSI ZSM, to demonstrate the alignment of prototyped solutions with the zero-touch principles that this SDO promotes.

The third and last objective focused on the specification and analysis of operator provided solutions for private 5G networks, exploiting network slicing capabilities. The most relevant conclusions from this objective are:

- Vertical customers will bring use cases and services with stringent performance requirements regarding throughput, latency, reliability, availability, security and device density, which private LTE cannot meet. The reason is that the target KPIs and features are well beyond the capabilities of 4G technology. In this situation, 5G technology enters to scene. The first generation of 5G (3GPP Release 15) will not provide features for uRLLC and mIoT support. These features will come with 3GPP Release 16, and it may take time for operators to integrate this release into public network; however, vertical sectors are demanding 5G critical communications early on. In such a situation, where private LTE cannot be used (because of limited capabilities) and Rel-16 featured public 5G will not be available in the short term, the only solution is the use of private 5G. From a technical viewpoint, the concept of private 5G network is known as non-public network (NPN). This term was coined by 3GPP in early 2019, with the Rel-16 kick-off, to refer to the use of 5G technology in a private mobile network environment.
- Industry 4.0 is a critical sector ready to be disrupted by private 5G; indeed, analyst reports point out that industry 4.0, utilities and mining will be covering most of market share in the short and medium run. This thesis has profiled the role of private 5G in industry 4.0, by specifying different flavors. These flavors capture options for a NPN to be deployed, with variants on the ownership and location of the constituent network functions, from radio access network to data network. It also has provided a comparative analysis across these flavors, in order to assess the usability in different industry 4.0 applications based on a number of decision criteria that include performance, security, service continuity support, integration efforts and cost figures, among others. This comparative analysis provides a solid foundation to understand the pros and cons of each solution, and identify existing gaps, so industry and research can start working to bridge them.
- This thesis has surveyed the use of private 5G in 3GPP Rel-16 and beyond networks, with insights that are applicable to a wide variety of vertical sectors, well beyond industry 4.0. This survey captures the technology facilitators for private 5G, including spectrum (different licensing options), interworking (i.e., integration of 5G

with legacy industrial technologies, such as Ethernet and industrial Wi-Fi), deterministic networking (with technologies such as TSN and DetNet), positioning (for advanced, cm-level localization support in indoor environments), hardware acceleration (to bridge the performance gap that virtualization brings) and security & privacy (authentication, authorization and trustworthy access to sensitive data). It also overviews the impact that private 5G brings to 3GPP system, outlining which capabilities makes a private 5G mobile network (NPN) different from a public 5G mobile network (PLMN). This impact includes novel features for RAN sharing, and the possibility of a NPN to be deployed either stand-alone (SNPN) or assisted by the public network (PNI-NPN). Network slicing scopes this latter scenario.

- Laying on the above groundwork, robust and future-proof system architectures have been proposed for different scenarios. These architectures integrate technology facilitators and capabilities featuring RAN sharing, SNPN and PNI-NPN. For PNI-NPN realization, the public network can interact with the private network in different forms, each with its pros and cons. To validate the hypothesis anticipated in this analysis, three deployment options have been proposed for an industry campus network. The setup considers 25 private users located inside three factory plants and 25 public users located inside and outside the factory plants. The deployment options are: 1) all users are served by a macrocell, and a PNI-NPN is deployed as a DNN for the private users; 2) public users are served by the macrocell whereas private users are served by small cells with CAGs located in the factory plants and the PNI-NPN is again deployed as a DNN; and 3) public outdoor users are served by the macrocell whereas indoor public and private users are served by the small cells and the PNI-NPN is deployed as a network slice. The simulation results prompt the following observations: for deployment option #1, the UE throughput is similar for both public and private users. For deployment option #2, the throughput of private UEs significantly increases, as they are served by the small cells. And for deployment #3, slicing enables allocating one carrier for public use in the small cells, thus improving the throughput of the indoor public UEs.
- Most acclaimed technology analysts report that SNPNs will be the preferred option for private 5G commercialization in the short-term. The reason is that 5G Rel-16 innovation is relatively easy here, since there are almost no legacies. However, having greenfield environments does not mean low cost, but just the opposite. The fact that customized 5G networks need to be set up for individual customers, à la carte, makes CAPEX and OPEX quite high, making this option only affordable to large-sized companies. However, as the 5G technology matures, the transition from isolated private networks (SNPN) to hybrid models (PNI-NPNs) will be reality, empowered with network slicing. Network slicing not only features traffic isolation, high degree of customization, security and performance levels comparable to those offered by private networks at a more reduced cost; it also provides flexibility and speed, with the ability to create, modify and tear down logical networks in a matter of minutes, allocating/de-allocating resources when and where needed, all following easily replicable provisioning and operation patterns. With these capabilities, network slicing will allow dramatically reducing upfront costs for customers, reducing entry barriers, and making private 5G accessible to a large portion of the B2B market.
- It may take several years for operators to prepare their systems to fully harness the power of network slicing, which is the ability of operators to offer PNI-NPNs to vertical customers using self-service capabilities. Waiting for the availability of this

NSaaS model is not an option, either operators (who cannot offset the costs associated with the introduction of slicing capabilities into their assets) or for customers (which may not need the full set of capabilities for their use cases from the very beginning). In this context, what the industry demands is a phased-based rollout of slicing, starting with early-stage solutions and outlining a vision of what is desired in the longer run to guide progress and focus. This thesis presents a radar with the mission to help industry in defining this rollout. This radar captures a complete landscape of network slicing solutions, linking these solutions to different timelines (as-is, short-term, medium-term, long-term) based on their technical viability and market demands. In addition to this timing, the radar has also outlined the dimensions that have an impact on the usability (how and where) of these solutions, across all operator managed domains. In the RAN domain, network slicing solutions have been discussed based on functional aspects (e.g., disaggregation and O-RAN integration), radio resource allocation and penetration in the operator's footprint. In the CN domain, discussions have been around the fulfilment of isolation and customer requirements of network slice customers, resulting in the use and combination of different 5GC functions, with different profiles. In the TN domain, solutions have been articulated around the technology capabilities available in the underlay WAN, complemented with the automation and programmability capabilities brought by SDN technology. Finally, in the OSS domain, aspects related to network slice orchestration and capability exposure have been addressed, with a focus on provisioning and assurance activities.

- The radar demonstrates that 5G slicing requires having 5G SA deployed. A Rel-15 5GC will allow operators to start with a few B2C slices; nevertheless, the full explosion is expected from Rel-16 onwards, with verticals progressively shifting from SNPN (costly to maintain and difficult to scale) to PNI-NPNs (cheaper, extensible, and enabling simultaneous access to public and private data services). The position of the different solutions in the radar also shows that the network domain most advanced in slicing support is 5GC, followed by the RAN and then the TN. The reason why the TN lags behind 3GPP domains is twofold. On the other hand, it requires major renovations in the technology substrate, not only on the data plane, but also on the SDN controller plane. On the other hand, the progress in standards is rather low, with many gaps existing between 3GPP SA5 and IETF models.

2 Future Work

The work presented herein opens many possibilities for future research on the use of network slicing for advanced 5G services in public-private scenarios. Some of them are a direct consequence of the findings stated above, others are aspects that could not be addressed due to the limited time resources of the PhD study and, finally, there are points which fall outside the scope of the thesis.

In relation to the management and orchestration of network slicing, much of the work to be done is on Openslice, by:

- Refining models used for the network slice descriptors, updating them according to the progress made in GSMA's GST and 3GPP Rel-17 ServiceProfile. For now, GST parameters correspond to those available in GST version 3.0, and attributes in ServiceProfile scopes the ones that were available by mid 2020.
- Developing advanced feasibility check engine, for Openslice to decide on the best provisioning solution for a slice request. As for today, upon capturing a service order,

Openslice looks for a feasible instantiation solution using a “trial-and-error” strategy, deploying the first one it finds. Through simpler and faster, this approach is not optimum, in the sense that there can be the multiple feasible solutions, as reported in Paper B. The work to be done lies on the need to develop an engine that identifies all feasible options first, and among them, selects the optimum one. The optimization criteria can include multiple dimensions (e.g., minimize resource consumption, minimize number of workload migrations in mobile scenarios, etc.), appropriately weighed according to the specific needs of the use case under consideration, and the overall status of operator’s system.

- Improving data collection mechanisms in Openslice’s assurance framework. Current solutions are based on collecting every bit of telemetry information from every data point using polling strategies, and dispatch this information to giant data lakes. Apart from being inefficient (most of data stored is not used later on, either because it is raw data or irrelevant for the analytics and AI/ML model training purposes) and resource consuming (when moving data from sources to the data lake), these approaches are reaching their limits, especially as the number of nodes increase. In this situation, it is needed to define model-based telemetry solutions with pushing strategies, subscribing only to required data, in the required format as per consumer needs. First steps in this direction have been already taken in [A4], co-authored by the PhD candidate.
- Evaluating system scalability, assessing how Openslice behaves as the number of slices running in parallel increases. This analysis will also serve to identify how going for a fine-grained vs coarse-grained slice design impacts on the behavior of individual slices (how SLA compliant they are) as well as the stability system (how much frequent orchestration actions are triggered over a period of time).
- Validating the Openslice + OSM stack over a commercial 5GC stack. The latest trials have been conducted over 5G NSA (the results have not been reported in this dissertation).
- Reproducing the PoC results in a cloud-native environment, using Kubernetes (k8s) instead of Openstack. This also would mean re-shaping the VNFs (VM-based) into CNFs (container-based).

Going down to the network layer, the network slicing technology radar has demonstrated that several challenges lie ahead before having full E2E slicing, even within a single administrative domain. In particular, further research work is needed in the following issues:

- The role of O-RAN’s RAN Intelligent Controller (RIC) in network slicing assurance. It is clear that the policy and AI capabilities provided by A1 interface, together with xApps, can make slicing operation more intelligent and agile. However, the synergies of these capabilities with other 3GPP RAN features such as Self-Organizing Networks (SON) needs further investigation.
- TN slicing, with solutions bridging these gaps: i) no means for dynamic mapping between 5QI and DSCPs; ii) no slicing support in the fronthaul; and iii) characterization of isolation. For the third point, first results have been made available in [B6], wherein the PhD candidate is co-author.

Finally, in relation to the combined use of private 5G networks and network slicing, many challenges lie ahead, some of them reported at the end of Paper G:

- At the infrastructure layer, the need to make a combined use of multiple access technologies, including 3GPP and non-3GPP technologies, to enhance reliability and increase throughput channel. There are concepts that need to be re-thought, and very few validations.

- At the network layer, the need to keep assessing which network functions can be delivered to the private segment and which ones make sense to remain in the public network domain, considering aspects such as security, data privacy. Hyperscaler solutions are also gaining momentum, so it is time to see how their solutions can fit into the private-public networks.
- Finally, at the management and orchestration layer, there are two main lines of work. On the one hand, the introduction of self-management capabilities, in order to minimize the number of "touches" on the network. To that end, work on closed loop automation and intent-based is on-going, with PhD candidate addressing these aspects in both standard bodies and other research projects. On the other hand, capability exposure, so that customer can gain access to operator capabilities using open and user-friendly APIs. This exposure allows the customer to retain (partial) control on the slice and PNI-NPN. PhD candidate has also started working in this topic, with participation in the CAMARA initiative [1][2].

3 References

[1] Linux Foundation press release, February 2022 [Online]. Link

[2] CAMARA GitHub [Online]. Link

Appendices

Annex A: Other merits related to research activities

This Annex reports additional research activities performed by the author of this Thesis. First, additional (co-)authored publications are listed, being these publications not included in the compendium of this Thesis, though related to the topics referenced within. Second, results from the participation in standardization activities are provided, just reporting those including explicit evidence of contribution or authorship. Third, a list of diverse talks (keynotes, invited talks, etc.) at different scientific and industrial venues is provided. Finally, it is described the role taken in several international research and innovation projects where the author has been involved.

1 Additional publications

The following is a list of additional co-authored publications not included as part of the dissertation. They are listed in chronological order.

1.1 Journals

[A1] 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. DOI: 10.1109/MCOM.2018.1701336.

[A2] O. Adamuz-Hinojosa, P. Munoz, J. Ordonez-Lucena, J. J. Ramos-Munoz and J. M. Lopez-Soler, "Harmonizing 3GPP and NFV Description Models: Providing Customized RAN Slices in 5G Networks," in *IEEE Vehicular Technology Magazine*, vol. 14, no. 4, pp. 64-75, Dec. 2019. DOI: 10.1109/MVT.2019.2936168.

[A3] X. Li, A. Garcia-Saavedra, X. Costa-Perez, C. J. Bernardos, C. Guimaraes, K. Antevski, J. Manges-Bafalluy, J. Baranda, E. Zeydan, D. Corujo, P. Ioavanna, G. Landi, J. Alonso, P. Pixao, H. Martins, M. Lorenzo, J. Ordonez-Lucena and D. R. López, "5Growth: An End-to-End Service Platform for Automated Deployment and Management of Vertical Services over 5G Networks," in *IEEE Communications Magazin*e, vol. 59, no. 3, pp. 84-90, March 2021. DOI: 10.1109/MCOM.001.2000730.

[A4] A. Pastor, D. Lopez, J. Ordonez-Lucena, S. Fernandez, J. Folgueira, “A Model-based Approach to Multi-domain Monitoring Data Aggregation", in *Journal of ICT Standardization*, vol 9, issue 2, August 2021. DOI: 10.13052/jicts2245-800X.9210

[A5] X. Li, C. Guimaraes, G. Landi, J. Brenes, J. Mangues-Bafalluy, J. Baranda, D.

Corujo, V. Cunha, J. Fonseca, J. Alegria, A. Zabala, J. Ordonez-Lucena, P. Iovanna, C. J. Bernardos, A. Mourad and X. Costa, "Multi-Domain Solutions for the Deployment of Private 5G Networks," in *IEEE Access*, vol. 9, pp. 106865-106884, August 2021. DOI: 10.1109/ACCESS.2021.3100120.

[A6] T. Cogalan, D. Camps-Mur, J. Gutiérrez, S. Videv, V. Sark, J. Prados-Garzon, J. Ordonez-Lucena, H. Khalili, F. Cañellas, A. Fernández-Fernández, M. Goodarzi, A. Yesilkaya, R. Bian, S. Raju, M. Goraishi, H. Haas, O. Adamuz-Hinojosa, A. Garcia, C. Colman-Meixner, A. Mourad and E. Aumayr, "5G-CLARITY: 5G-Advanced Private Networks Integrating 5GNR, WiFi, and LiFi," in *IEEE Communications Magazine*, vol. 60, no. 2, pp. 73-79, February 2022. DOI: 10.1109/MCOM.001.2100615.

1.2 Conferences

[B1] P. Martinez-Julia, J. Ordonez-Lucena, V. P. Kafle, H. Asaeda and D. Lopez, "Exploiting Case Based Reasoning to Automate Management of Network Slices" in *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1-6. DOI: 10.1109/NOMS47738.2020.9110423.

[B2] W. Y. Poe, J. Ordonez-Lucena and K. Mahmood, "Provisioning Private 5G Networks by Means of Network Slicing: Architectures and Challenges," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1-6, DOI: 10.1109/ICCWorkshops49005.2020.9145055.

[B3] A. Gonzalez, J. Ordonez-Lucena, B. E. Helvik, G. Nocioni, M. Xie, D. Lopez and P. Gronsund, "The Isolation Concept in the 5G Network Slicing," in *2020 European Conference on Networks and Communications (EuCNC)*, 2020, pp. 12-16. DOI: 10.1109/EuCNC48522.2020.9200939.

[B4] D. Camps-Mur, M. Ghoraishi, J. Gutierrez, J. Ordonez-Lucena, T. Cogalan, H. Haas, A. Garcia, V. Sark, E. Aumayr, S. Meer, S. Yan, A. Mourad, O. Adamuz-Hinojosa, J. Perez-Romero, M. Granda AND R. Bian, "5G-CLARITY: Integrating 5GNR, WiFi and LiFi in Private Net- works with Slicing Support," in *European Conference on Networks and Communications (EuCNC)*, Dubrovnik, Croatia, June 2020.

[B5] M. Xie, P. H. Gomes, J. Harmatos and J. Ordonez-Lucena, "Collaborated Closed Loops for Autonomous End-to-End Service Management in 5G," in *2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2020, pp. 64-70. DOI: 10.1109/NFV-SDN50289.2020.9289902.

[B6] L. M. Contreras and J. Ordonez-Lucena, "On Slice Isolation Options in the Transport Network and Associated Feasibility Indicators," in *2021 IEEE 7th International Conference on Network Softwarization (NetSoft)*, 2021, pp. 201-205. DOI: 10.1109/NetSoft51509.2021.9492546

2 Standardization

The following is the list of contributions that the doctoral candidate has drafted in the standardization arena. This list only includes the contributions which have been accepted, and which are directly linked to the research work of the thesis. This means that all the documents mentioned in this section address network slicing and non-public networks concepts, or any other system feature intrinsically tied to them (e.g., capability exposure).

2.1 3GPP SA5

The candidate is Telefónica delegate at 3GPP SA5. The following contributions have been authored during the thesis.

[C1] "pCR 28.807 Roles related to NPN management". *Doc identifier*: S5-196713. *Work item*: Study on Non-Public Networks Management (Rel-16). *Plenary meeting*: SA5#127, France, October 2019.

[C2] "pCR 28.807 Use case deployment of a stand- alone Non-Public Network (SNPN) with 3GPP and non-3GPP segments". *Doc identifier*: S5-196734. *Work item*: Study on Non-Public Networks Management (Rel-16). *Plenary meeting*: SA5#127, France, October 2019.

[C3] "pCR 28.807 Network Slice as a Service in the management of Public Network Integrated Non-Public Network (PNI-NPN)". *Doc identifier*: S5-201596. *Work item*: Study on Non-Public Networks Management (Rel-16). *Plenary meeting*: SA5#129e, Virtual, February 2020.

[C4] "pCR 28.807 Add requirements for management of SNPN and PNI-NPN". *Doc identifier*: S5-202215. *Work item*: Study on Non-Public Networks Management (Rel-16). *Plenary meeting*: SA5#130e, Virtual, April 2020.

[C5] "Rel-16 CR 28.533 Add clarifications to description of tenant concept". *Doc identifier*: S5-202217. *Work item*: Study on multi-tenancy support (Rel-16). *Plenary meeting*: SA5#130e, Virtual, April 2020.

[C6] "pCR 28.557 Roles related to NPN management". *Doc identifier*: S5-204463. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#130e, Virtual, April 2020.

[C7] "pCR 28.557 Add use case on NPN provisioning by network slice of PLMN". *Doc identifier*: S5-205400. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#133e, Virtual, October 2020.

[C8] "pCR 28.557 Add generic management aspects". *Doc identifier*: S5-205402. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#133e, Virtual, October 2020.

[C9] "pCR 28.557 Add use case on NPN provisioning". *Doc identifier*: S5-205403. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#133e, Virtual, October 2020.

[C10] "pCR 28.557 Add CAG management". *Doc identifier*: S5-211479. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#135e, Virtual, January 2021.

[C11] "pCR 28.557 Add 5GLAN group management in UE related management aspects". *Doc identifier*: S5-212360. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#136e, Virtual, March 2021.

[C12] "pCR 28.557 Applicability of management modes considering the deployment options of individual NPN functions". *Doc identifier*: S5-212361. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#136e, Virtual, May 2021.

[C13] "pCR 28.557 Add NG-RAN related management requirements". *Doc identifier*: S5-212483. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#137e, Virtual, May 2021.

[C14] "CR 28.531 CR clarify misleading information on network slicing use cases". *Doc* *identifier*: S5-212461 + S5-212462. *Work item*: Maintenance (Rel-16 + Rel-17). *Plenary meeting*: SA5#137e, Virtual, May 2021.

[C15] "28.811 use case – support isolation in network slice subnet". *Doc identifier*: S5-213506. *Work item*: Study on Network Slice Management Enhancements (Rel-17). *Plenary meeting*: SA5#137e, Virtual, May 2021.

[C16] "S5-215526 Key issues relative to network slice management capabilities exposure". *Doc identifier*: S5-215526. *Work item*: Study on Network Slice Capability Exposure (Rel-17). *Plenary meeting*: SA5#139e, Virtual, October 2021.

[C17] "S5-221526 pCR 28.557 Add new NPN management related requirements". *Doc identifier*: S5-221526. *Work item*: Management of non-public networks (Rel-17). *Plenary meeting*: SA5#141e, Virtual, January 2022.

[C18] "S5-221531 Rel-17 CR 28.541 Update RANSliceSubnetProfile attributes". *Doc identifier*: S5-221531. *Work item*: Enhanced Management of 5G Service Level Agreements (Rel-17). *Plenary meeting*: SA5#141e, January 2022.

2.2 ETSI ZSM

The candidate is Telefónica delegate at ETSI ZSM. The following contributions have been authored during the thesis.

[C19] "ZSM003 Use of GSMA Generic Template in NSaaS scenarios". *Doc identifier*: ZSM(19)000488r6. *Work item*: Network Slicing (ZSM003). *Plenary meeting*: ZSM#8, Bonn, October 2019.

[C20] "On ZSM relationship with OSM". *Doc identifier*: ZSM(20)000044. *Work item*: Landscape (ZSM004). *Plenary meeting*: ZSM#10, Luxembourg, February 2020.

[C21] "ZSM004 OSM in ZSM Architecture". *Doc identifier*: ZSM(20)000320. *Work item*: Landscape (ZSM004). *Plenary meeting*: ZSM#12e, Virtual, September 2020.

[C22] "ZSM004 Update on GSMA description". *Doc identifier*: ZSM(20)000321. *Work item*: Landscape (ZSM004). *Plenary meeting*: ZSM#12e, Virtual, September 2020.

[C23] "ZSM004 Closed Loop Automation in OSM relevant to ISG ZSM". *Doc identifier*: ZSM(20)000322. *Work item*: Landscape (ZSM004). *Plenary meeting*: ZSM#12e, Virtual, September 2020.

[C24] "ZSM003 Management exposure to vertical customers". *Doc identifier*: ZSM(20)000322r3. *Work item*: Network Slicing (ZSM003). *Plenary meeting*: ZSM#12e, Virtual, September 2020.

[C25] "ZSM003 Network slice management and SLAs". *Doc identifier*: ZSM(20)000327r1. *Work item*: Network Slicing (ZSM003). *Plenary meeting*: ZSM#12e, Virtual, September 2020.

[C26] "ZSM003 Clarify network slice as a service". *Doc identifier*: ZSM(21)000023r2. *Work item*: Network Slicing (ZSM003). *Plenary meeting*: ZSM #14a Tech Call.

[C27] "ZSM004 Updates on GSMA and OSM sections". *Doc identifier*: ZSM(21)000256r1. *Work item*: Landscape (ZSM004). *Plenary meeting*: ZSM Interim #09e.

[C28] ETSI ZSM; PoC#2 Report 1 "PoC#2 user story", February 2021. Link

2.3 IETF

The candidate occasionally co-authors IETF individual drafts, his contributions being limited to assessing the impact of network slicing in the TN segment.

[C29] L. Geng, L. Qiang, J. Ordonez-Lucena, O. Adamuz-Hinojosa, P. Ameigeiras, D. Lopez and L. Contreras, "COMS Architecture," draft-geng-coms-architecture, March 2018.

[C30] S. Homma, H. Nishihara, T. Miyasaka, A. Galis, V. Ram OV, D. Lopez, L. Contreras- Murillo, J. Ordonez-Lucena, P. Martinez-Julia, L. Qiang, R. Rokui, L. Ciavaglia and X. De Foy, "Network Slice Provision Models," draft-homma-slice-provision-models, March 2019.

[C31] S R. Rokui, S. Homma, D. Lopez, X. de Foy, L. Contreras-Murillo, J. Ordonez-Lucena, P. Martinez-Julia, M. Boudacair, P. Eardley, K. Makhijani and H. Flinck, "5G Transport Slice", draft-rokui-5g-transport-slice, July 2019.

[C32] Luis M. Contreras, S. Homma, J. Ordonez-Lucena, "Considerations for defining a Transport Slice NBI", draft-contreras-teas-slice-nbi, November 2019.

[C33] Luis M. Contreras, I. Bykov, J. Ordonez-Lucena, "Connecting 3GPP slices through IETF Network Slice services", dradt-contreras-teas-3gpp-ietf-slice-mapping, March 2022.

2.4 GSMA NG

The candidate has contributed to the following GSMA white paper on E2E slicing. He has drafted section 3.3 (isolation), section 4 (network slicing architecture), section 5.4 (O&M), section 5.4.2 (IETF), section 5.4.6 (ETSI OSM), section 6 (On a phased-based rollout for E2E network slicing) and Annex A.2 (O&M, IETF).

[C34] GSMA, "E2E Network Slicing Architecture" White Paper, version 1.0, June 2021 [Online]. Link

2.5 ETSI NFV

The main findings from Paper A have been published in the following ETSI technical report.

[C35] ETSI GS NFV-EVE 012, "Network Functions Virtualisation (NFV); Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework", V3.1.1, December 2017 [Online]. Link

3 Talks

[D1] J. Ordonez-Lucena "Slice and Expose – Enhancing Network Slicing Applications by Means of Capability Exposure", invited talk at the *FOKUS FUSECO FORUM*, Berlin, Germany, November 2021.

[D2] J. Ordonez-Lucena, "Slicing with Non-Public Networks – And other orchestration challenges for the next decade", invited talk at the *NetWorld 2020 Visions on Future Communications Summit*, Lisbon, Portugal, November 2021.

[D3] J. Ordonez-Lucena, "On the Operation of Non-Public Networks: The Operator's Perspective", special session at the *5G-PPP Technical Board* (going virtual), December 2020.

[D4] J. Ordonez-Lucena, "5G-CLARITY Architecture, Innovation and Use Cases", special session at the *5G-PPP Architecture Working Group* (going virtual), January 2021.

[D5] J. Ordonez-Lucena, “On the use of OSM to allow for automated network slicing in multi-site environment”, especial session at the *OSM-MR10 Hackfest* (going virtual), March 2021.

[D6] J. Ordonez-Lucena, “MNOs and Private 5G Networks – A Perfect Catalyst for Industry Digitization”, special session at the *5GWeek*, London, UK (going virtual), June 2021.

[D7] J. Ordonez-Lucena, “Outlook for Operator Adoption for Private 5G Networks”, special session at the *2020 European Conference on Networks and Communications (EuCNC),* Duvronik, Croatia (going virtual), June 2021.

[D8] J. Ordonez-Lucena, “ZSM NOC – Emerging Use Cases”, invited talk *at the Layer123 Reunion,* Madrid, Spain, April 2022.

[D9] J. Ordonez-Lucena, “ZSM PoC#2: Automated network slice scaling in multi-site environments”, invited talk *at the Layer123 Reunion*, Madrid, Spain, April 2022.

4 International research and innovation projects

The following is a list of funded projects where the PhD candidate has actively participated.

<u>5G-VINNI (ICT-815279): 5G Verticals INNovation Infrastructure.</u>

- **Project duration**: July 2018 – December 2021.
- **Project objective**: 5G-VINNI intends to accelerate the uptake of 5G in Europe by providing an end-to-end (E2E) facility that validates the performance of new 5G technologies by operating trials of advanced vertical sector services. The project is committed to demonstrate the achievement of 5G KPIs across a range of combinations and permutations of new 5G access technologies and end-user equipment types interconnected by the most advanced 5G core network technologies available, applying Network Function Virtualization, Network Slicing and a rigorous automated testing campaign to validate the 5G KPIs under various combinations of technologies and network loads.
- **Main role in the project**: leader of Task 3.1 on the definition of network slice descriptor, at both the NFVO and Service Orchestration layer, for the specification of network slice service offerings. These offerings include single-domain and multi-domain (cross-site) slice services.
- **Thesis’s objectives in scope**: Objective 2

<u>5GROWTH (ICT-856709): 5G-enabled Growth in Vertical Industries</u>

- **Project duration**: June 2019 – February 2022.
- **Project objective**: 5GROWTH has as main objective the technical and business validation of 5G technologies from the vertical points of view, following a field-trial-based approach on vertical sites (TRL 6-7). Multiple use cases of vertical industries will be field-trialed on four vertical-owned sites in close collaboration with vendors and operators (Altice, Telecom Italia, Telefonica). The project leverages on the results of 5G-PPP Phase 2 projects where slicing, virtualization and multi-domain solutions for the creation and provisioning of vertical services are being developed and validated.

- **Main role in the project**: leading specification of PNI-NPN scenarios for 5Growth use cases, at both network and orchestration layers. For the orchestration layer, the PhD candidate designs solutions for controllable exposure of 5Growth capabilities.
- **Thesis's objectives in scope**: Objective 3

5G-CLARITY (ICT-871428): Beyond 5G Multi-Tenant Private Networks Integrating Cellular, Wi-Fi, and LiFi, Powered by Artificial Intelligence and Intent Based Policy.

- **Project duration**: October 2019 – December 2022.
- **Project objective**: 5G-CLARITY puts forward the design of a beyond-5G system that addresses the wide variety of challenges identified today in private network environments, including spectrum flexibility, delivery of critical services, integration with public network infrastructures, and automated network management with built-in slicing. It integrates multiple capabilities, at the infrastructure layer (5GNR+WiFi+LiFi and on-prem edge computing), network layer (ATSSS and O-RAN RIC) and intelligence layer (intent and AI/ML engines), that all together measurable enhancements with respect to the eMBB and URLLC services defined by 3GPP in Release 16, in terms of low latency, area capacity, reliability, and accurate positioning and synchronization features.
- **Main role in the project**: Leader of WP2 (Scenario Description, Architecture and Requirements). Leading Task 2.2 on the system architecture definition, and Task 4.2 on the specification and validation of E2E service orchestration solutions based on the combined use of public and private 5G infrastructures (PNI-NPN).
- **Thesis's objectives in scope**: Objective 3

Annex B: Resumen

Este anexo incluye un breve resumen de los puntos principales de la tesis, con el objetivo de cumplir con la normativa de la Escuela de Posgrado de la Universidad de Granada referente a la redacción de tesis doctorales cuando éstas son escritas en inglés.

1 Introducción a los sistemas 5G

La introducción de la tecnología digital en los procesos económicos e industriales puede jugar un papel esencial para hacer frente a los desafíos de esta nueva década. Los sistemas de quinta generación (5G) serán un activo clave para impulsar la digitalización de estos procesos, y construir con ellos una sociedad completamente móvil y conectada.

1.1 Casos de uso

Más allá de redes de comunicación con tecnologías de acceso radio vanguardistas, los sistemas 5G serán plataformas digitales capaces de dar soporte a las industrias verticales como las fábricas del futuro (también referidas como industria 4.0), la e-Salud, la automoción, o la energía. Estas industrias quieren aprovechar las capacidades que 5G brindará para acelerar en su transformación hacia un ecosistema empresarial más moderno y sostenible, construido sobre los principios de eficiencia (optimización de costes) e innovación (acelerando el desarrollo de nuevos servicios). La irrupción de los *verticales* en el sector telco trae como resultado nuevos casos de uso, muy variopintos en cuanto a prestaciones y características. En los primeros estudios de la 3GPP y NGMN Alliance, se identificaron más de 70 casos de uso nuevos. Atendiendo a sus requisitos, se clasificaron estos casos de uso en tres categorías de servicio:

- **Banda ancha móvil mejorada (*enhanced Mobile Broadband*, eMBB)**. Esta categoría agrupa todos aquellos casos de uso destinados a ofrecer una mejor experiencia de usuario en áreas de cobertura tradicionales, mejorando las prestaciones de las redes 4G en cuatro ejes fundamentales: tasa de transferencia de datos (mayores velocidades de carga/descarga), capacidad (mayor densidad de conexión), retardos (latencias más bajas) y movilidad (continuidad del servicio en los sistemas de transporte público de alta velocidad, como trenes). Estas prestaciones se dirigen a aplicaciones de tipo media, como el *streaming* de vídeos de alta resolución, *cloud gaming*, realidad aumentada, y realidad virtual.
- **Comunicaciones ultra fiables y de baja latencia (*ultra-Reliable and Low Latency Communications*, uRLLC)**. Esta categoría agrupa todos aquellos casos de uso que impone requisitos muy estrictos en cuanto a fiabilidad (tolerancia *quasi*-cero a fallas) y latencia (del orden del ms, con variaciones mínimas y controladas). Ejemplos de

caso de uso incluye robots conectados en entornos de fabricación distribuidos, control remoto de dispositivos en tiempo real, sistemas de seguridad y servicios de emergencia, cirugía remota, así como la gestión inteligente del transporte y de los recursos energéticos.

- **Internet de las Cosas masivo (*massive IoT,* mIoT)**. También conocida como comunicaciones masivas de tipo máquina (*massive Massive-Type Communications* mMTC), esta categoría recoge aquellos casos de uso que implica conectar a millones de dispositivos, sin intervención humana, a gran escala. La capacidad de gestionar grandes densidades de conexión tiene el potencial de revolucionar los procesos y aplicaciones de la industria moderna e inclusive de la agricultura y de las ciudades inteligentes.

Acomodar los distintos casos de uso sobre una misma infraestructura supone uno de los grandes desafíos de 5G, pues se han de satisfacer simultáneamente los requisitos específicos de cada caso de uso (muy distintos entre sí), y a la vez garantizar su completo aislamiento (la provisión de un caso de uso a un vertical no debe afectar al resto de verticales). Para lograr esto, se requieren diseñar nuevas soluciones de red flexibles, adaptables y escalables.

1.2 Una aproximación al diseño de sistemas 5G

Las redes móviles actuales son sistemas monolíticos e inflexibles, basados en hardware de propósito específico, y optimizados para dar soporte únicamente a servicios de banda ancha (es decir, servicios de voz y datos). Los sistemas 5G no sólo requerirán mejorar las prestaciones de estos servicios, sino también dar soporte a los verticales. El resultado es un gran conjunto de casos de uso que deben ser acomodados simultánea y eficientemente sobre una misma infraestructura de red, manteniendo su correcto aislamiento. Cada uso de uso impone un conjunto de requisitos exigentes; sin embargo, estos requisitos no involucran simultáneamente todas las prestaciones de 5G. Por ejemplo, un caso de uso de la categoría uRLLC exigirá muy baja latencia y muy alta fiabilidad; sin embargo, la velocidad de transferencia de datos y la densidad de conexión no son críticos. Del mismo modo, un caso de uso eMBB exigirá altas prestaciones en cuanto a velocidad y soporte de movilidad, pero puede tolerar ciertas variaciones del retardo y caída del servicio en momentos puntuales.

Atendiendo a esto último, el concepto de *network slicing* puede ser una aproximación a adoptar para dar solución al problema planteado. *Network slicing* propone la creación de multiples redes lógicas (virtualizadas) denominadas *slices*, cada una diseñada y adaptada para satisfacer un conjunto específico de requisitos, pero todas compartiendo una misma infraestructura. La ejecución concurrente de varias *slices* en una infraestructura común debe hacerse garantizando el correcto aislamiento entre las mismas.

2 Network slicing: concepto y principios

Network Slicing es una solución tiene como objetivo dividir la infraestructura en un conjunto de particiones de red lógicas, cada una optimizada (en términos de recursos, topología, funciones, configuración y administración) para satisfacer unos requisitos muy específicos. La intención es que cada *slice* se provisione únicamente con aquellas capacidades que son indispensables para procesar el tráfico de los servicios/casos de uso acomodados, evitando todas las funcionalidades innecesarias. Por ejemplo, una *slice* para un servicio de tipo mIoT será diseñado para soportar altas densidades de conexión y proporcionar una alta eficiencia energética de los dispositivos, sin preocuparse del soporte de movilidad ni de proporcionar bajas latencias. Esto quiere decir que la *slice* se provisionará con algunas funciones básicas del plano de control, y con una configuración en la radio basada en protocolos específicos de acceso al medio compartido.

La realización de *network slicing* se apalanca en una serie de principios, entre los cuales se incluyen:

- **Aislamiento**. Esta propiedad garantiza la independencia de las *slices*, aunque todos se ejecuten sobre una infraestructura común. Esto significa que cualquier congestión, falla, brecha de seguridad o configuración aplicada en un *slice* no tendrá impacto en el resto de *slices*.
- **Personalización.** Esta propiedad garantiza que cada *slice* se provisiona únicamente con lo necesario para satisfacer los requisitos específicos de los casos de uso que sirve. Esta personalización se puede aplicar en varias dimensiones, incluyendo i) capacidad, asignando más o menos recursos; ii) topología, con distintas variantes en el número de nodos y caminos entre ellos; iii) plano de usuario, configurando las funciones de procesamiento y reenvío del tráfico con las políticas adecuadas; iv) plano de control, activando únicamente las funcionalidades necesarias; y v) añadiendo aplicaciones de servicio, como firewalls, servidores IoT, servicios de analíticas y *big data*, adaptados a las necesidades de los casos de uso en cuestión.
- **Elasticidad**. Esta propiedad hace referencia a la posibilidad de ajustar la capacidad del *slice* de forma dinámica. Este principio permite que una *slice* siempre cumpla con los requisitos recogidos en el acuerdo con el cliente (*Service Level Agreement,* SLA) independientemente de las condiciones variables de la red, sin más que modificar asignar / desasignar recursos de acuerdo con las fluctuaciones del tráfico.
- **Programabilidad**. Esta propiedad permite manejar la *slice* como un artefacto *software*. Esto quiere decir que i) los recursos y las propiedades de una *slice* se pueden modelar y capturar en un fichero, que se puede procesar por sistemas automatizados; y ii) que la asignación y configuración de esos recursos se puede controlar con el uso de Interfaces de Programación de Aplicaciones (*Application Programming Interfaces,* APIs).
- **Despliegue extremo-a-extremo**. La *slice* es una construcción lógica que proporciona una conectividad customizada entre dos puntos: los consumidores del servicio (los dispositivos, incluyendo smartphone, sensor, vehículo), conectado a la red de acceso; y el productor del servicio (p.e., servidor de aplicación), desplegado en la red de datos. Para conectar los consumidores con el productor del servicio, la *slice* debe atravesar varios dominios de red, incluyendo los dominios de acceso (RAN), transporte (TN) y núcleo (CN), cada uno integrando distintas tecnologías. Es imprescindible asegurar un comportamiento consistente de la *slice* a lo largo de este camino, asegurando que las prestaciones y funcionalidades en un dominio sean coherente con las implementadas en el resto de los dominios. Un ejemplo de incongruencia sería asignar mucha capacidad en el plano de usuario del acceso y el núcleo, pero tener poco ancho de banda en la red de transporte.
- **Abstracción jerárquica**. Esta propiedad tiene sus raíces en el principio de recursión, que permite replicar patrones de virtualización de forma recurrente, en distintos niveles. Esto quiere decir lo siguiente: un recurso virtualizado en un nivel L, que resulte de la abstracción de un recurso en el nivel L-1, puede volver a virtualizarse, generando recursos más abstractos en el nivel L+1. Aplicando este mismo razonamiento al *slicing,* esto quiere que una *slice* se puede volver a abstraer (virtualizar) en otras *slices,* con el propósito de aplicar modelos de negocio mayoristas, y de ventas a 3ros. Por ejemplo, un operador de red puede entregar una

slice eMBB a un operador móvil virtual, y este a su vez particionarlo en nuevas *slices*, comercializándolos a 3ros. Del mismo modo, un operador de red puede vender una *slice* a un vertical (p.e., empresa de manufacturación), y el vertical segregar esta partición en *slices* más específicos, para servir distintos casos de uso (p.e., robots conectados, IoT industrial, control remoto de dispositivos).

- **Escalabilidad**. Esta propiedad hace referencia a la especificación de una arquitectura capaz de lidiar con las necesidades operaciones de *slicing,* en términos del número y tipo de *slices* que se requieren orquestar sobre una infraestructura común. Esta especificación está sujeta al criterio de diseño de las *slices*. Si tenemos diez casos de uso, el operador podría decidir diseñar tres *slices* (una *slice* por categoría de servicio), diez *slices* (una *slice* por caso de uso), o cualquier otra variante. Cuanto más nos acerquemos al número diez, significa que tenemos *slices* más granulares; esto permite una mayor personalización, a costa de disparar el número de slices ejecutándose en paralela, lo que puede en poner en compromiso la estabilidad de la arquitectura. En cambio, cuanto más nos acerquemos al número tres, tendremos un efecto contrario (entorno más estable y gestionable, a costa de tener *slices* más genéricas). El operador ha de buscar una solución de compromiso en el criterio de diseño de las *slices,* que no siempre es fácil de hacer teniendo en cuenta el número de variables en juego.
- **Automatización**. Con diferentes *slices* ejecutándose simultáneamente sobre una misma infraestructura, y unas condiciones de tráfico bastante variables y dinámicas, es imposible operar la red de forma manual. En este sentido, es necesario definir mecanismos de automatización en la gestión de ciclo de vida de las *slices*, que supervisen su comportamiento y lancen acciones correctivas oportunas, escalando el problema a los operadores humanos sólo en aquellos casos que sea estrictamente necesario.

3 Network slicing: ¿redes públicas o privadas?

Tradicionalmente, las redes públicas y privadas se han visto como dos planteamientos radicalmente distintos, dos opciones que compiten entre sí por ser la ‘elegida’ para dar solución a un caso de uso. Sin embargo, *network slicing* permite acercar ambos planteamientos, ofreciendo una solución que aún a los beneficios de los dos mundos. Por un lado, el *slicing* permite unos niveles de aislamiento, personalización, prestaciones y seguridad muy similares a los que ofrecen las redes privadas. Por otro lado, brinda niveles de cobertura similares a los que ofrecen las redes públicas, con capacidades de soporte de movilidad, y acceso confiable a servicios de datos e Internet. Y, además, ofrece flexibilidad y agilidad, con la capacidad de provisionar, modificar y terminar *slices* en cuestión de minutos, desplegando las funciones y aplicaciones donde (y cuando) sea necesario.

Un operador de red comercializa tres tipos de *slices*, dependiendo del segmento del mercado al que van dirigido:

- **Business-to-Customer (B2C)**: cualquier *slice* diseñada para cursar tráfico de usuarios públicos. Un ejemplo puede *slic*e de cloud gaming.
- **Business-to-Business (B2B)**. cualquier *slice* optimizada para un vertical o institución gubernamental, acomodando uno o varios casos de uso. Por ejemplo, una *slice* para una empresa de industria 4.0.
- **Business-to-Business-to-X (B2B2X)**: es un sabor similar al B2B, pero aplicando el principio de recursión, con relaciones cliente-proveedor a varios niveles. Un ejemplo

puede ser un *slice* para operador móvil virtual, que éstte revende a sus usuarios (B2B2C). Otro ejemplo puede ser una slice para un proveedor de nube pública (Amazon Web Services, Microsoft Azure, Google Cloud), que éste revende a clientes empresariales (B2B2B).

Para todos esos casos, no es raro que una *slice* atraviese dos o más dominios administrativos. Por ejemplo, los *slices* B2C suelen alojar servicios globales como el *cloud gaming* y distribución de contenidos, con una cobertura que va más allá de la huella de un solo operador. En este caso, se requiere que la *slice* se despliegue sobre infraestructura gestionadas por distintos operadores, cada uno definiendo un dominio administrativo diferente. Del mismo modo, para el caso B2B, es habitual que parte de las funciones y aplicaciones se desplieguen en las premisas del vertical (p.e., fábrica, campus, etc.) y el resto en la red pública. Aquí nuevamente tenemos dos dominios distintos: el gestionado por el vertical, y el gestionado por el operador.

4 Pilares de network slicing

Esta sección proporciona una breve descripción general de las capacidades que habilitan la realización de *network slicing*.

4.1 Tecnologías de *softwarización* de red

La *softwarización* de red representa la transformación de las redes en sistemas construidos sobre el principio de abstracción y desacoplo. Este principio se basa en separar el *software* que implementa funciones, protocolos y servicios de red del *hardware* que los ejecuta. Esta transformación está cambiando la forma en la que las infraestructuras de comunicación se diseñan y operan, siendo la frontera con el mundo IT cada vez más difusa. Dentro del paradigma de softwarización de red, hay dos tecnologías que destacan por encima del resto: la de las redes definidas por software (*Software Defined Networking*, SDN) y la de virtualización de funciones de red (*Network Functions Virtualization*, NFV).

4.1.1 Redes Definidas por Software (SDN)

SDN es un nuevo paradigma de red caracterizado por desacoplar los planos de datos y control. Con SDN, los dispositivos de red tradicionales se transforman en dispositivos programables muy sencillos, dedicados exclusivamente al reenvío de paquetes (plano de datos). La inteligencia de la red (plano de control) se encuentra lógicamente centralizada en un controlador software que, a través de una interfaz estandarizada, permite configurar en tiempo real el comportamiento de los dispositivos de red. Esto habilita la programabilidad de la red, y una automatización en su gestión.

4.1.2 Virtualización de Funciones de Red (NFV)

Por otra parte, NFV hace uso de los principios de virtualización, popularizados por el paradigma de *cloud computing*, y los aplica a las redes telco. NFV permite que las funciones de red (hasta ahora ejecutadas en hardware dedicado, propietario y no escalable) sean virtualizadas (desacopladas del hardware). Esto da lugar a las denominadas Virtualized Network Functions (VNFs), implementadas como software ejecutándose en hardware de propósito general (abierto, no propietario y barato).

ETSI ha propuesto un marco de referencia para NFV, basado en el despliegue y operación de ñas VNFs sobre una infraestructura de red virtualizada. La composición de

VNFs permite definir de forma flexible y dinámica servicios de red virtualizados, particularizados a las necesidades específicas de cada caso de uso. Para la gestión y orquestación de todos los artefactos en NFV, el marco de referencia de ETSI define una arquitectura llamada NFV-MANO. Esta arquitectura se compone de tres bloques funcionales: el *Virtualized Infrastructure Manager (*VIM*)*, responsable del control de la infraestructura virtualizada; *el VNF Manager* (VNFM*)*, que gestiona el ciclo de vida de cada una de las VNFs; y el *NFV Orchestrator (*NFVO*)*, que gestiona el ciclo de vida de los servicios de red, y supervisa la ejecución de las tareas del VIM y el VNFM.

4.2 Computación en el borde

La computación en el borde (*edge computing*) representa un paradigma consistente en acercar las capacidades *cloud* al perímetro de la red, trasladándolas desde los centros de datos a nodos más cercanos al acceso radio. La idea es que el almacenamiento y procesamiento de los datos se lleve a cabo en una ubicación lo más cercana posible al usuario, dispositivo o servicio que consumirá esos datos. Esto permite al operador i) reducir cuellos de botella en la red del núcleo de red, y ii) mover las funciones del plano de datos lo más cerca de la red de acceso, con el objetivo de minimizar los retardos y cumplir con los requisitos de baja latencia de los casos de uso correspondientes.

El objetivo último de esta tecnología es tener un continuo de *cloud, d*esde el usuario/dispositivo final hasta los centros de datos en Internet. Sin embargo, este es un proceso que lleva su tiempo, y que requiere dotar de capacidades *cloud* a nodos que actualmente no las tienen.

En la red pública, los primeros pasos se están dando ya, con los operadores reacondicionando sus centrales regionales para transformarlas en entornos de virtualización. Esto permitirá tener a corto plazo el *near edge*. A medio-largo plazo, este mismo reacondicionamiento se pretende replicar en las centrales de la última milla, lo que se conoce como *far edge*. En

En el contexto de redes privadas, los verticales también han empezado a migrar sus servicios y aplicaciones, moviéndolas desde la nube pública hasta sus premisas, donde cuentan con entornos de *cloud* a pequeña escala. Esto es lo que se conoce como *on-premises edge computing*, generalmente aplicable a servicios B2B/B2B2C. Las razones de esta migración no sólo responden a rendimiento, sino también a cuestiones de seguridad, o cuestiones legales (residencia de los datos).

Un *slice* podría hacer uso de los distintos tipos de *edge*, dependiendo del caso de uso. Por ejemplo, imaginemos un *slice* B2B multidominio, con parte del *slice* ejecutándose en las premisas del cliente, y la otra parte en la red del operador. En este caso, las cargas del primero segmento se desplegarían en el *on-premises edge*, mientras que las cargas del segundo segmento podrían desplegarse en el *near edge*.

4.3 Gestión y orquestación

La gestión y la orquestación hace referente a la lógica que gobierna la provisión y operación de servicios en redes *softwarizadas*, incluido las *slices*. En objetivo es controlar estos servicios, configurando el comportamiento de cada uno de ellos a lo largo de su ciclo de vida. Para el caso de una *slice*, el ciclo de vida se articula en cuatro fases:

- **Preparación**. En esta fase, el operador hace la “puesta a punto”, configurando el entorno de la red antes de iniciar el despliegue de la *slice*. Esta puesta a punto incluye tareas de planificación de capacidad, el diseño de descriptores, y su incorporación a los catálogos correspondientes.

- **Provisión**. Esta fase se corresponde al despliegue de la *slice* en la infraestructura de red. Dependiendo de los requisitos de servicio a soportar, la *slice* deberá ofrecer unas prestaciones u otras, y por tanto las funciones que la componen deberán desplegarse y configurarse de forma diferente. En esta fase la *slice* aún no procesa tráfico.
- **Operación**. Esta fase comprende todas las actividades que van desde la activación (la *slice* empieza a procesar tráfico) hasta la desactivación (la *slice* deja de procesa tráfico). Estas actividades incluyen i) monitorización, consistente en recoger métricas y alarmas sobre el estado actual de una *slice*; ii) supervisión, comprobando que las prestaciones que la *slice* ofrece en cada momento se ajustan a los requisitos decretados en el SLA; y iii) modificación, cambiando las prestaciones de la *slice* de acuerdo con las variaciones del tráfico o las condiciones de servicios. El escalado (incrementar la capacidad de la *slice,* asignándole más recursos) es un ejemplo claro de en qué consiste la actividad de modificación.
- **Terminación**. En esta fase, la *slice* se elimina. Los recursos que se asignaron se liberan, quedando disponibles para su utilización en otras *slices*.

Aparte de para administrar el ciclo de vida de cada *slice*, la lógica que representa el concepto de "gestión y orquestación" es también responsable de resolver dependencias y conflictos entre *slices* (por ejemplo, gestionando prioridades entre ellas en situaciones de escasez de recursos).

4.4 Resumen

Si ponemos en relación los conceptos vistos en esta sección, podemos ver que la **computación en el borde** (Sección 4.2) proporciona un sustrato de computación distribuida, un continuo de cloud que se desde el dispositivo hasta los centros de datos. La tecnología de **NFV** (Sección 4.1.2) permite instanciar las funciones de la *slice* dentro de ese continuo, desplegadas en los nodos adecuados con la capacidad necesaria. La tecnología de **SDN** (Sección 4.1.1) permite a los operadores controlar de forma programática los flujos de tráfico dentro de la *slice*, a lo largo de las funciones que NFV desplegó. Finalmente, la **gestión y orquestación** (Sección 4.3) es el marco responsable de gestionar el ciclo de vida de las distintas *slices*, desde su provisión hasta su terminación, garantizando el correcto aislamiento de las mismas cuando se ejecutan sobre una infraestructura compartida.

5 Principales desafíos en network slicing

En esta sección se resumen brevemente los principales desafíos inherentes a la tecnología de *network slicing*. Esto incluye desafíos técnicos (Sección 5.1), aunque también otros no necesariamente asociados a la tecnología (Sección 5.2).

5.1 Desafíos técnicos

- **Diseño de una arquitectura ajustada al estándar**. Es preciso hacer un uso combinado de los distintos pilares de slicing (p.e., SDN, NFV, computación en la nube), integrando los componentes en una arquitectura software. El desafío aquí reside en que la funcionalidad de estos componentes es muy distinta, y que su estandarización corre a cargo de distintos organismos. Cómo hacer que todos los componentes software encajen se hace aún más difícil a medida que el ecosistema de estándares se hace más grande y fragmentado, y la necesidad de tener soluciones multi-proveedor. En este contexto, el uso de interfaces normativas y abiertas se hace más necesario que nunca.

- **Segregación de recursos.** Consiste en particionar la infraestructura en un conjunto de recursos lógicos, abstraídos de los recursos físicos, y segregados entre sí. Esta partición se basa en el hecho de que varias *slices* se ejecutan sobre una misma infraestructura común, de ahí la necesidad de proveerlas con estos recursos lógicos. La partición se debe aplicar extremo a extremo, sobre recursos inalámbricos, de conectividad y de cómputo. El rendimiento del sistema dependerá del nivel de abstracción al cualo se hace la partición; cuanto más cerca del nivel físico, más aislados estarán los recursos lógicos entre sí, a costa de renunciar a tener un menor número de recursos a asignar entre *slices*, y por tanto a las ganancias de multiplexación. El desafío aquí es cómo diseñar soluciones que tengan en cuenta las ganancias de multiplexación y el asilamiento, y que mantengan ambas dimensiones lo más equilibradas posibles. Se pueden utilizar distintos enfoques para el diseño de estas soluciones, un número que aumenta a medida que aparecen nuevas capacidades tecnológicas.
- **Asignación de recursos.** Es el paso que sigue a la segregación de recursos. Consiste en entregar los recursos lógicos (resultantes de la partición de la infraestructura) a las distintas *slices*, de acuerdo a sus necesidades. El desafío aquí se articula en dos preguntas: ¿Cómo asegurar un reparto de recursos coordinado extremo a extremo, de tal forma que los dominios RAN, CN y TN de una slice exhiben unas prestaciones consistentes? ¿Con qué frecuencia se deben re-asignar (orquestar) recursos entre slices?
- **Traducción de requisitos de servicio a requisitos de infraestructura**. Consiste en mapear los objetivos del servicio (p.e., tasa de transferencia garantizada/máxima por usuario, números de usuarios, retardo máximo) en capacidades de red y número de recursos. Tradicionalmente, los operadores han resuelto el problema con el uso de reglas de mapeo predefinidas, algo fácil teniendo en cuenta i) los servicios a acomodar tenían perfiles de rendimiento similares, y ii) que la infraestructura estaba construida con equipos de pocos suministradores. Sin embargo, la irrupción de los verticales y las tecnologías de *softwarización* plantean ahora un escenario totalmente distinto, con muchas más variables en i) y ii). En este nuevo contexto, hay muchas cuestiones aún por resolver: ¿Cómo formalizar los requisitos del servicio, en un lenguaje que tanto el operador (con experiencia en telecomunicaciones) como el vertical (sin experiencia en telecomunicaciones) puedan entender? ¿Cómo traducir requisitos de servicio a requisitos de función de red, p.e., compartir/dedicar funciones de red y la activación/desactivación de ciertas funcionalidades? ¿Cómo mapear los requisitos de función de red en soluciones de segregación y asignación de recursos? ¿Cuáles son los criterios del operador para decidir si el servicio se puede acomodar en una *slice* existente o si es necesario definir una nueva *slice*? ¿Cómo diseñar soluciones de manera que el operador pueda tomar todas estas decisiones en unos pocos minutos, de modo que se pueda brindar feedback al cliente de forma rápida? En caso de que la red no pueda acomodar los requisitos de servicio solicitados, ¿debería el operador sugerir al cliente requisitos alternativos como parte de este feedback?
- **Seguridad**. La llegada de las tecnologías de *softwarización* permite que las redes sean más flexibles, elásticas y personalizables. Sin embargo, los principios de virtualización y programabilidad también traen consigo nuevos vectores de ataque y

brechas de seguridad, algo que entornos de recursos compartidos como el *slicing* requiere de especial atención.

- **Federación**. Aplicar el *slicing* de manera consistente a lo largo de los distintos dominios de red es por sí un desafío. Pero si a esto le añadimos la componente de los dominios administrativos, como la presentada al final de la Sección 3, el problema es aún más complejo. La razón es que ahora requerimos federar capacidades de distintas organizaciones, coordinando decisiones como i) segregación y asignación de recursos, ii) configuración de conectividad), iii) escalado. A pesar de que la industria coincide en la necesidad de aplicar la federación para garantizar esta coordinación entre dominios administrativos distintos, hay aún un montón de cuestiones abiertas: ¿Qué interfaces, protocolos y API deben usarse para la federación? ¿Qué estándar debería ser responsable de su especificación? ¿Debería la federación basarse en interacciones a pares entre dominios administrativos, o debería un tercero actuar como intermediario (bróker) para mediar en estas interacciones? ¿Cómo garantizar la confiabilidad al federar dominios gestionados por organizaciones distintas? ¿Cuáles son las especificidades de control de acceso y auditabilidad al federar un dominio privado con un dominio público?
- **Exposición de capacidades al cliente.** Una de las características más notables que permite el *slicing* es la capacidad de proporcionar al cliente la percepción de que tiene una red dedicada a su disposición. Esto significa que el cliente no sólo se limita a monitorizar las métricas de la *slice* (operación pasiva), sino que también puede acceder a la *slice* y configurarla de forma programática (operación proactiva). Esto es especialmente relevante para los modelos B2B y B2B2X, donde los clientes de la *slice* son verticales, operadores virtuales, o proveedores de nube pública. El desafío aquí radica en cómo exponer estas capacidades operativas al cliente, utilizando las API. En particular, se encuentran los siguientes problemas abiertos: ¿Qué parámetros deben contener estas APIs, para asegurar que son amigables de cara al cliente? ¿Deberían estas API ser específicas del operador o podríamos esperar algún trabajo de estandarización aquí? Si es así, ¿qué organismo debería liderar esta estandarización? ¿Cómo aplicar el control de acceso al cliente? ¿Cómo garantizar que las acciones ordenadas por un cliente tengan un impacto mínimo en otras *slices*? ¿Cuáles son los mecanismos que se necesitan para trazar los mensajes intercambiados entre el operador y el clientes en relación a la operación de *slice*, y cómo impacta en la auditabilidad de la SLA?
- **Provisión y operación automatizadas en una *slice*.** El objetivo final es eliminar la intervención humana, haciendo que las slices sean entidades autónomas, con capacidades de autogestión. Esto requeriría automatizar el ciclo del ciclo de vida de las distintas *slices*, programando los flujos de orquestación y las políticas de decisión asociadas. Sin embargo, conseguir este objetivo en el corto-medio plazo es poco realista; no sólo se requiere una evolución de las capacidades de gestión y orquestación sino experiencia operativa. En otras palabras, se necesita entender primero la operativa del *slicing* en las redes en producción, y a medida que se vayan ganando aprendizajes, se podrá pensar en automatizar los procesos manuales.
    - En la parte de provisión, algunas de las cuestiones abiertas son las siguientes: ¿Cómo diseñar algoritmos para la verificación de factibilidad en el despliegue de una *slice*?¿Cuáles deberían ser los parámetros de entrada y salida de estos algoritmos?¿Cuál es el impacto en los flujos relacionados con la gestión de catálogos e inventarios?
    - En la parte de operación, tenemos cuestiones referentes para el reporting

(¿cuáles son los KPIs que deben monitorizarse para la *slice*?¿qué métricas necesito para computar estos KPIs? ¿de qué nodos los puedo obtener? ¿cómo correlar métricas de función de red con métricas de infraestructura en entornos virtualizados?) y para la supervisión y modificación (¿cuáles son las políticas *cross-slice* e *intra-slice* que se configurar a partir de los requisitos de servicio recibidos por cada cliente?¿cómo garantizar que estas políticas no son incompatibles entre sí?¿cómo programar algoritmos de escalado de *slices*, y cuáles son los parámetros de entrada y salida requeridos?¿cómo propagar las decisiones de escalado hacia los sistemas SDN y NFV, para ejecutar estas decisiones en la infraestructura?¿cómo arbitrar situaciones en las que las decisiones son incompatibles entre sí?¿cómo puede ayudar la inteligencia artificial en este arbitraje?).

5.2 Desafíos no técnicos

Aparte de los desafíos técnicos, existen otros desafíos que los operadores deben abordar antes de empezar a comerciar servicios con *slicing*. Estos desafíos son fundamentalmente de dos tipos.

- **Monetización.** La monetización del *slicing* es uno de los grandes retos a los que se enfrentan los operadores en el corto plazo. Existen aún muchas preguntas sin respuesta: ¿cómo cristalizar las capacidades del *slicing* (aislamiento, personalización, conectividad bajo demanda) en productos de valor para el cliente? ¿Cómo catalogar estos productos en el portafolio del operador? ¿Se pueden definir varios sabores de un mismo producto, cada uno asociado a una tarificación diferente? ¿Cuál son los criterios para definir estos sabores? ¿Cómo gestionar la relación con el cliente? ¿En qué casos siguen válidos son los canales de venta existentes hoy en día, y en qué casos se tienen que migrar hacia portales de autogestión?
  Otro punto a abordar es la estrategia de comercialización, que responde a cómo debe el operador gestionar la salida al mercado de su solución de *slicing*:¿con qué proveedores y socios? ¿con qué cartera de clientes y servicios? Si un operador tiene presencia en varios países, estas decisiones se deben tomar en cada una de sus operaciones locales, ya que las condiciones del mercado suelen ser distintas entre ellas.
- **Regulación.** El éxito del *slicing* como solución dependerá en gran medida de las decisiones que los reguladores tomen en torno a cuestiones como la neutralidad de red, la residencia de datos, y gestión del espectro. La primera cuestión aplica sobre todo a *slicing* B2C, mientras que las dos últimas cuestiones tienen impacto en redes público-privadas 5G, y por ende en *slicing* B2B/B2B2X..

6 Alcance y Objetivos de la Tesis

Con el *network slicing*, la red del operador se puede dividir lógicamente en un conjunto de particiones de red programables, cada uno diseñada para satisfacer un conjunto concreto de requisitos de servicio. Las *slices* resultantes de esta partición operarán de forma simultánea, y deberán estar debidamente aislados entre sí. Esto significa que a pesar de estar desplegados sobre una infraestructura de red común (compartida), las *slices* requieren una administración separada (independiente), acorde con los tiempos y necesidades de los servicios que acomodan.

La industria ya ha empezado a trabajar en *slicing*, aunque de forma poco coordinada. Existente muchos organismos de estandarización y foros, cada uno con su propia temporización y prioridades. El resultado es un ecosistema muy fragmentado, con muchas capacidades que no terminan de casar, y que imposibilitan tener soluciones extremo a extremo. Como se ha recogido en la Sección 5, existen una gran variedad de cuestiones aún por resolver. Además, al llegar al mercado B2B, la industria se enfrasca en la eterna discusión "red privada 5G" vs "slicing", asumiendo una dicotomía entre ellas que realmente no existe. La realidad es todo lo contrario; existe una relación simbiótica entre ambas soluciones, tal y como se puso de manifiesto en la Sección 3.

La ambición de la presente tesis es **el diseño y validación de soluciones para la gestión y orquestación de *slicing* en entornos multi-dominio, con foco en infraestructuras 5G público-privadas.** Para la consecución de esta ambición, el trabajo se ha articulado en tres objetivos.

El primer objetivo se centra en el diseño de soluciones a nivel de sistema para *slicing*. El objetivo es tener un marco de referencia que permita discutir sobre los enfoques a adoptar para abordar los desafíos planteados en la Sección 5.1. El trabajo llevado a cabo en el objetivo 1 pone el foco en los recursos de cómputo y conectividad, que son los que están dentro del alcance de las tecnologías SDN y NFV. Esto significa que los aspectos inherentes a RAN *slicing* quedan fuera del alcance de este objetivo.

Objetivo 1: Diseño de un sistema de gestión y orquestación para *slicing* en entornos multi-dominio. Este objetivo se compone de los siguientes subobjetivos:

- **O1.1**: Diseño de una arquitectura para el sistema, utilizando como palancas los marcos de referencia SDN/NFV definidos en los organismos de estandarización correspondientes. La arquitectura debe dotar al sistema de capacidades de gestión y orquestación de *slicing* en infraestructuras multi-dominio.
- **O1.2**: Diseño de un descriptor de *slice*. Este descriptor capturará la información que el sistema necesita para gestionar instancias de *slices* a lo largo de su ciclo de vida, desde la provisión hasta la terminación, incluyendo la fase de operación.

Calendario: 2016Q2 – 2018Q2

El segundo objetivo se centra el validar las hipótesis y asunciones que se hicieron en el objetivo 1. Para tal fin, el objetivo 2 creará un prototipo de las soluciones a nivel de sistema diseñadas, y las validará en casos de uso relevantes.

Objetivo 2: Implementación y validación de un prototipo de sistema. Este objetivo incluye los siguientes subobjetivos:

- **O2.1**: Implementar un prototipo del descriptor de slice.
- **O2.2**: Implementar un prototipo de la arquitectura del sistema.
- **O2.3**: Validación de los prototipos desarrollados, evaluando sus capacidades durante la fase de preparación del *slice*. El foco se pondrá en la operación de *on-boarding*.
- **O2.4:** Validación de los prototipos desarrollados, evaluando sus capacidades durante la fase de provisión del *slice*. El foco se pondrá en la operación de instanciación.
- **O2.5**: Validación de los prototipos desarrollados, evaluando sus capacidades durante la fase de operación del *slice*. El foco se pondrá en la funcionalidad de auto escalado.

Calendario: 2018Q3 – 2021Q2

Finalmente, el tercer objetivo se centra en el estudio del rol de 5G en entornos de red privada, el análisis del papel que juega aquí *network slicing*, y en la identificación y caracterización de soluciones que ilustran las sinergias entre *slicing* y redes 5G privadas. El objetivo 3 aprovecha los resultados reportados en los objetivos 1 y 2, y los complementa con la inclusión de i) el concepto de red privada 5G, modelada como un nuevo administrativo federable, y ii) el dominio RAN.

Objetivo 3: Análisis del ecosistema de 5G privado, y diseño de soluciones para redes 5G privadas, explotando capacidades de *slicing*. Este objetivo incluye los siguientes subobjetivos:

- **O3.1**: Estudio sobre las especificidades del 5G en entornos de red privada, incluyendo actores, los casos de uso, requisitos y tecnologías involucradas, identificando principales diferencias con respecto al uso de 5G en redes públicas. El análisis se centrará en la industria 4.0.
- **O3.2**: Diseño de soluciones para redes 5G privadas, integrando tecnologías habilitantes y capacidades de *slicing*. Dependiendo del escenario de aplicación, se tendrán distintas opciones de despliegue, con sabores que van desde redes aisladas (infraestructura 100% privada) hasta redes híbridas (infraestructura público-privada).
- **O3.3:** Overview de soluciones de *slicing*, caracterizándolas en términos de i) prestaciones, con foco en el aislamiento; ii) temporización, explicando su disponibilidad en el corto, medio y largo plazo; y iii) entornos de red aplicables, incluyendo redes públicas y privadas. Este overview se proporcionará a través de un radar tecnológico. Aparte de ayudar a los operadores a definir su estrategia de despliegue y comercialización de slicing, este radar ayuda a la industria a entender la utilización de *slicin*g en distintos escenarios, con foco en la provisión de soluciones de red privada 5G para el mercado B2B.

Calendario: 2019Q2 – 2021Q4

7 Conclusiones

El trabajo de esta tesis se ha articulado en torno a tres objetivos fundamentales.

El primer objetivo versa sobre el diseño de una arquitectura de red para el soporte de soluciones de *slicing* en entornos multidominio. Las conclusiones más relevantes en relación con este objetivo se resumen en los siguientes puntos:

- Garantizar el aislamiento entre *slices* no es una cuestión baladí, teniendo en cuenta que todas estas redes lógicas están desplegadas sobre una misma infraestructura y, por tanto, sujetas a la compartición de recursos. El aislamiento entre *slices* es un problema con múltiples aristas, y que debe abordarse desde tres perspectivas fundamentales: i) rendimiento, ii) seguridad y iii) gestión. En primer lugar, se debe asegurar que las prestaciones de una *slice* no se vean impactadas/deterioradas por la ejecución de otras *slices* en paralelo. Para ello, el operador de red debe tener en cuenta el SLA al que cada *slice* está sujeto a la hora de aplicar los mecanismos de orquestación de recursos (consistente en segregar la infraestructura en un pool de recursos virtualizados, para luego balancearlos entre las distintas *slice*s). En segundo lugar, se debe asegurar que cualquier ciberataque o falla en una *slice* no se propague al resto, y que el acceso y consumo de una *slice* esté limitado restringido al cliente correspondiente. Para ello, el operador debe implementar mecanismos de control de acceso que sean robustos, y que no estén limitados a la autenticación/autorización del cliente, sino también a una adecuada gestión de su acceso a los datos, incluyendo soluciones que protejan la integridad, confidencialidad y la privacidad de los mismos. Finalmente, en tercer lugar, se debe asegurar que cada *slice* tenga una gestión de ciclo de vida independiente, separada del resto. Esto garantiza la customización de cada *slice,* tanto en lo referente a su configuración (programar sus parámetros con unos valores u otros) como en su operación (puede ser más o menos dinámica, dependiendo de los patrones de tráfico del servicio o servicios que acomoda, o de las necesidades del cliente). Y no sólo esto; también habilita a que el cliente puede monitorizar la *slice*, incluso asumir un control parcial sobre él, sin riesgo de que esta monitorización y control impacte en el comportamiento de otras *slices*.

- En esta tesis se ha especificado una arquitectura para el sistema de gestión y orquestación de *slices* en entornos multi-dominio. La arquitectura definida se apalanca en los marcos de referencia propuestos por la ONF y ETSI ISG NFV, extendiéndolos de tal manera que un uso combinado de los mismos proporcione las capacidades de *slicing*, especialmente en lo que se refiere a “garantizar el aislamiento en infraestructuras compartidas, y que se extiende por la huella de varios operadores. El sistema diseñado permite gestionar el ciclo de vida completo de cada *slice*, incluida las fases de preparación y puesta en marcha (publicación B), así como la fase de operación (publicación A).
- La fase de preparación hace referencia a todo lo que es preciso poner en marcha en los sistemas del operador, previo al despliegue de una *slice*. Abarca principalmente actividades de planificación de red, la elaboración de los descriptores de las *slices*, y su posterior registro en el catálogo. Un descriptor de *slice* es un fichero estructurado, conforme a un modelo de datos bien definido, que proporciona una caracterización completa de una *slice*. Permite que los sistemas gestión y orquestación sepan cómo desplegarlo y operarlo, conforme a un SLA. Sin embargo, la estructura de este descriptor, qué parámetros debe recoger, y cómo éstos impactan en la gestión de ciclo de vida de una *slice*, son cuestiones aún por resolver; y la única referencia que tenemos similar, la de los descriptores en el dominio NFV (NSDs y VNFDs), es a todas luces insuficiente. Para abordar este problema, en esta tesis se ha propuesto una solución de diseño de un descriptor de *slice*. La propuesta considera un descriptor estructurado en torno a cinco grandes campos: i) topología de la *slic*e, con las distintas funciones que la integran, y sus relaciones de conectividad; ii) requisitos de servicio, incluyendo las funcionalidades y prestaciones que la *slice* debe ofrecer; iii) tiempo de vida, especificando la fecha de inicio y final de la *slice*, y los periodos de activación/desactivación de la misma, en caso de que los hubiere; iv) cobertura, especificando la geografía donde la *slice* debe estar operativa; y v) requisitos operacionales, que recogen información relevante en tiempo de ejecución, por ejemplo, las métricas que se deben recoger de la slice (relevante para computar KPIs y detectar fallas), la priorización de la *slice* (relevante para gestionar situaciones de escasez de recursos) o los datos que se ha de exponer al cliente (con fines de visibilidad y monitorización). Los campos i), ii) y iv) asisten al sistema de gestión y orquestación en la provisión de la *slice*. Los campos ii), iii) y v) asisten al sistema de gestión y orquestación en la operación de la *slice*, asegurando que a lo largo de su ciclo de vida se comporta conforme a las especificaciones del SLA.
- La fase de provisión hace referencia al despliegue de la *slice*. En esta tesis se han detallado los flujos de orquestación correspondiente a esta fase de provisión, y se han agrupado en distintas etapas. La primera etapa es la correspondiente a emitir la de orden de servicio; el cliente accede al catálogo, navega por los distintos descriptores, selecciona el que mejor se ajusta a sus expectativas, cumplimenta los parámetros

configurables, y ordena instanciar una *slice* en base a esa descripción. En cuanto recibe esa orden, el operador traduce la caracterización de la *slice* a recursos de red, mapeando los parámetros del descriptor a configuraciones NFV y a capacidades de infraestructura. Luego, el operador procede con control de admisión (verifica si la infraestructura tiene recursos suficientes para servir la nueva *slice*), seguida de una etapa de optimización y reserva de recursos, donde se selecciona la realización final del *slice* (en caso de que haya más de una forma posible de desplegarlo) y se reservan los recursos necesarios (para evitar que otro *slice* pueda hacer uso de ellos). Para cada una de las etapas subrayadas, se han especificado los flujos de trabajo, detallando input/output y los componentes del sistema que intervienen en los mismos.

- En la fase de operación, la *slice* está activa, procesando tráfico, desplegado sobre una infraestructura en la que también hay otras *slices* ejecutándose en paralelo. En esta fase del ciclo de vida, el sistema de gestión y orquestación debe asegurar el aislamiento entre todos las *slices*, teniendo en cuenta: i) que todos ellos corren sobre una infraestructura común; ii) que el conjunto de recursos disponibles en la infraestructura es finito, y por ende, la flexibilidad a la hora de orquestar recursos disminuye a medida que aumenta el número de *slices*; y iii) que es posible que algunas *slices* tengan que escalar, para lidiar con las variaciones de tráfico de los servicios asociados. Gestionar el aislamiento en este tipo de entornos, dinámicos y basados en la compartición de recursos, no es sencillo. Y por supuesto, asignar recursos dedicados a las distintas *slices* no es una opción, ya que esto rompe con las ganancias de multiplexación que el *slicing* precisamente brinda. Se requiere por tanto el diseño de mecanismos de orquestación de recursos que permitan compartir recursos cuando sea posible, pero siempre asegurando el cumplimiento de los SLAs de las distintas *slices*. En el caso de los recursos inalámbricos, esto se puede conseguir con distintas configuraciones en la pila del protocolo de 5GNR, tanto *intra-slice* como *inter-slice*. En el caso de recursos de cómputo, el operador puede implementar políticas de orquestación similares a las que se usan en el *cloud computing*, complementadas con soluciones de aceleración software, específicas de entornos NFV.

El segundo objetivo consiste en la implementación y validación de soluciones para el diseño de arquitectura. Las conclusiones más relevantes se resumen en los siguientes puntos:

- Se ha hecho una implementación del descriptor de slice. El prototipo hace uso de los parámetros del GST/NEST para especificar los requisitos de servicio, del tiempo de vida y de cobertura. Para describir la topología del slice, la solución hace referencia a los NSDs/VNFDs, y expone aquellos puntos en los que el cliente puede extender la topología, incluyendo aplicaciones adicionales. Finalmente, en relación a los requisitos operacionales, el prototipo incluye información para ejecutar operaciones de reporte y supervisión, por ejemplo, qué métricas hay que recoger. Para ello, se especifican las fuentes de datos, qué datos recoger de ellas, cómo consumirlas (alarmas basadas en umbrales o informes periódicos), y cómo procesadas (por ejemplo, para computar KPIs).

  El prototipo planteado va más allá del estado del estado del arte, abordando las limitaciones y complejidades de las implementaciones existentes en distintos proyectos de la 5G-PPP (5G-TRANSFORMER, SLICENET y 5GTANGO) y comunidades de código abierto (OSM y ONAP). En particular, el prototipo plantea una solución alineada con el progreso de los estándares, por ejemplo, la incorporación de GST/NEST para el perfilado del slice, e implementada de forma sencilla, siguiendo las recomendaciones y buenas prácticas recogidas en el marco de SID del TM Forum.

- Se ha hecho una implementación del sistema de gestión y orquestación. El prototipo es un desarrollo software que combina las soluciones de OSM y Openslice. Ambas son pilas de orquestación basadas en código abierto, aunque con distinto alcance. Por un lado, OSM es una implementación de referencia del MANO definido por ETSI NFV, responsable de la gestión del ciclo de vida de las *slices* en la capa de infraestructura virtualizada; esto es, orquesta los recursos de las distintas VNFs que integran la *slice*. Por otro lado, Openslice hace las veces de orquestador de servicios, ocupándose de la gestión del ciclo de vida de los *slices* en la capa de aplicación; esto es, gestiona la semántica de la *slice*, las configuraciones a aplicar, etc. La interacción entre ambas pilas se hace a través de la interfaz de norte de OSM, basada en SOL005. Openslice consume las APIs de SOL005 para informar a OSM de cómo y dónde tiene que desplegar las distintas VNFs de la *slice*, y de cualquier otra modificación que sea necesaria en tiempo de operación (por ejemplo, escalado). En términos sencillos, se puede afirmar que Openslice es quien toma las decisiones de la *slice* (porque conoce su semántica, incluyendo configuraciones, SLA asociado, etc.) y OSM es quien las ejecuta (porque es quien tiene el control de los recursos virtualizados sobre los que corre el slice).
- A diferencia de OSM, Openslice es una pila de orquestación que sí se implementó a raíz la investigación de esta tesis doctoral. Openslice sigue una arquitectura desagregada y basada en servicios, con módulos que producen y consumen APIs a través de dos artefactos: un bus (ActiveMQ) y un registro de servicios (Consul). Openslice proporciona tres tipos de capacidades: i) interacción con 3ros, ii) operación interna, y iii) federación. Para la interacción con 3ros, Openslice ofrece APIs a través de su interfaz norte. Estas APIs son consumidas por dos tipos de 3ros: i) los desarrolladores de aplicaciones, que traen sus VNFs y sus aplicaciones de servicio, y las cuelgan en el catálogo, a modo de *Marketplace*, para que el operador pueda hacer uso de ellas a la hora de diseñar distintas *slices*; y ii) los verticales, que son los que solicitan la provisión de la *slice*, para acomodar sus casos de uso. Para cuestiones relacionadas con la operación interna, Openslice incluye módulos que ofrecen capacidades de provisión (gestión de órdenes de servicio, gestión de inventarios y catálogos, creación de instancias de slices, configuración y activación de las mismas, etc.) y aseguramiento (recolección de métricas, detección de fallos, supervisión, gestión de políticas, etc.). Estos módulos son la columna vertebral del sistema, quienes toman decisiones, y quienes articulan los flujos que gobiernan el ciclo de vida de una *slic*e. Finalmente, las capacidades de federación son aquellas que permiten a Openslice interactuar con pilas de orquestación en otros dominios administrativos (públicos o privados). Esta interacción permite hacer una *slice* que se expanda más allá de la huella de un operador, involucrando infraestructuras públicas multi-operador, o infraestructuras público-privadas. Para esta interacción, Openslice expone un conjunto de APIs abiertas, definidas por el TM Forum, que incluyen las siguientes: TMF633, TMF638, TMF641 y TMF640.
- Esta tesis ilustra la usabilidad de los prototipos desarrollados, describiendo los flujos de orquestación en distintas fases del ciclo de vida: i) preparación, con el diseño del descriptor de *slice* y su posterior catalogación; ii) provisión, con la instanciación y configuración de una *slice,* en respuesta a la orden de servicio emitida por el vertical; y iii) operación, con foco en la operación de escalado.
- Esta tesis ha validado las hipótesis planteadas. Para ello, ha desplegado los prototipos

desarrollados en 5G-VINNI, y ha reproducido los flujos de orquestación aquí. 5G-VINNI es una infraestructura de experimentación 5G a gran escala, con distintos nodos desplegados por la geografía europea, incluyendo España, Grecia, Reino Unido y Noruega, entre otros. En esta tesis, se ejecutó una prueba de concepto (PdC) con los nodos españoles y griegos, para validar las hipótesis en un entorno multi-dominio. Los resultados de la PdC se reportaron a ETSI ZSM, para demostrar el alineamiento de los prototipos desarrollados con los principios de automatización promovido por este organismo de estandarización.

El tercer y último objetivo se centró en la especificación y análisis de soluciones de operador para el 5G privado, utilizando soluciones de slicing como palanca para la provisión de estas soluciones. Las conclusiones más relevantes de este objetivo son:

- Los verticales traerán casos de uso con requisitos muy estrictos en cuanto a ancho de banda, latencia, confiabilidad, disponibilidad, seguridad y densidad de conexiones. Estos requisitos exceden las prestaciones que la tecnología 4G puede ofrecer, ni siquiera con su versión industrial (private LTE). Ante esta situación, entra en escena la tecnología 5G. La primera fase del 5G, la denominada 3GPP Release 15, está centrada en la mejora de las capacidades eMBB, pero nada más. Esto significa que lo referente a funcionalidades para el soporte de servicios tipo uRLLC y mIoT queda para Release 16. La integración de esta Release en las redes comerciales llevará algún tiempo; el problema es que los verticales demandarán estas capacidades mucho antes. Ante tal situación, en la que no se puede usar LTE privado (sus prestaciones no son suficientes) ni tampoco la red pública 5G (las características de Release 16 no estarán disponibles a corto plaza), la única solución es el uso de redes 5G privadas. Desde un punto de vista técnico, a estas redes se les denomina redes no públicas (NPN). Este término fue acuñado por 3GPP a principios de 2019, y se refiere al uso de la tecnología 5G en un entorno privado, con carácter industrial.
- La industria 4.0 es un sector clave para la aplicación de NPN's. Esta tesis ha caracterizado el uso del 5G en la industria 4.0, especificando distintos sabores de NPN's. Cada sabor representa una opción de despliegue y configuración del 5G distinto. Se ha hecho un análisis comparativo entre estos sabores, para evaluar su aplicabilidad en distintos casos de uso de la industria 4.0; los criterios utilizados para este ejercicio incluyen figuras de rendimiento, funcionalidades como soporte de la continuidad de servicio, cuestiones referentes a la seguridad, así como aspectos operativos (esfuerzos de integración), económicos (costos) y de negocio (barreras de entrada para nuevos clientes). Este análisis comparativo proporciona una base sólida para entender las ventajas/inconvenientes de cada sabor. También permite identificar cuestiones abiertas (problemas sin resolver), abriendo nuevas líneas de trabajo tanto para la industria como la academia, que han de trabajar conjuntamente para dar solución a esas cuestiones.
- Esta tesis ha examinado las especificidades del 5G privado en entornos que van más allá de la industria 4.0, ampliando el alcance de las NPN's a otros sectores verticales. En primer lugar, se han identificado y discutido los facilitadores del 5G privado: espectro (diferentes opciones de licenciamiento), integración con tecnologías industriales legadas (p.e., Ethernet, Wi-Fi), conectividad determinista (con latencias aseguradas y sincronismo, como las reportadas en las tecnologías TSN y DetNet), posicionamiento (permite el soporte de localización con gran precisión, a nivel de cm), aceleración hardware (para contrarrestar el *overhead* y, con él, la degradación del rendimiento que a veces conlleva el uso de la virtualización), y seguridad y privacidad (control de acceso, y protección de la integridad y confidencialidad de los

datos, todos ellos contiene información sensible). En segundo lugar, se ha analizado el impacto que el 5G privado trae en el diseño de redes, destacando aquellas capacidades que hacen que una red móvil 5G privada (NPN) sea diferente a de una red móvil 5G pública (comercial). Este impacto incluye funcionalidades avanzadas para compartir la red de acceso, y la posibilidad de implementar una NPN como una SNPN o una PNI-NPN. En el primer caso, la NPN es una isla aislada, totalmente separada de la red pública. En el segundo caso, la NPN es un despliegue híbrido, con parte de las funciones desplegadas en las premisas del cliente y el resto en la red del operador. Este modelo público-privado se puede provisionar con soluciones de *slicing.*

- Los facilitadores y las capacidades discutidas en el punto anterior se han integrado en un sistema de red, diseñado con arquitecturas robustas, extensibles y escalables. Para la realización de PNI-NPN's, el dominio público (red del operador) puede interaccionar con el dominio privado (red desplegada en las premisas del cliente) de distintas formas, cada una con sus ventajas e inconvenientes. Para validar las hipótesis planteadas en este análisis, se ha hecho una simulación de una red privada, consistente en tres factorías conectadas. El escenario de simulación considera la existencia de 25 usuarios privados ubicados dentro de las factorías, y 25 usuarios públicos con cobertura de interiores y exteriores. Con esta configuración, se plantea tres sabores. El primero (#1) una macrocelda sirve a todos los usuarios, y la PNI-NPN se implementa como una red de datos dedicada, usando la solución DNN de 3GPP. En el segundo (#2) macrocelda sólo sirve a los usuarios públicos, mientras que para los usuarios públicos se sirven usando microceldas indoor (con características de CAG para control de acceso); nuevamente, la PNI-NPN se implementa usando DNN. Finalmente, en el tercer sabor (#3), la macrocelda sirve a los usuarios públicos en exteriores, utilizando para el resto las microceldas indoor; además, LA PNI-NPN se implementa ahora como una *network slice*. La simulación de los tres sabores arroja los siguientes resultados: para el escenario #1, las prestaciones de los usuarios privados y públicos es similar. Para el escenario #2, los usuarios privados tienen mejores prestaciones, al ser servidos por microceldas indoor; además, la funcionalidad CAG impide que los usuarios públicos que estén en cobertura de interiores se puedan conectar a estas microcelda. Finalmente, para el escenario #3, la utilización de *network slicing* permite mejorar las prestaciones de los usuarios públicos, sin apenas degradar el rendimiento de los usuarios privados.
- Analistas reputados en el sector telco afirman que las SNPN serán la opción preferente para la comercialización a corto plazo del 5G privado. La razón es que la innovación es relativamente sencilla aquí, al no haber tecnologías 3G/4G con las que coexistir, como sucede en la red del operador. A estos entornos "limpios" se les denomina *greenfield*. Sin embargo, contar con este tipo de entornos no significa que los costes sean menores, sino todo lo contrario. Las redes hay que diseñarlas, dimensionarlas y configurarlas *a la carta*, atendiendo a los requisitos específicos del vertical; esto se traduce en unos gastos de CAPEX y OPEX bastante altos, que son únicamente asumibles por las grandes empresas (p.e., compañías globales, con grandes cifras de facturación). Sin embargo, a medida que la tecnología 5G vaya madurando, se espera una transición desde la SNPN (uso único de infraestructura privada, en las premisas del cliente) a PNI-NPN (uso de infraestructuras público-privadas, combinando recursos de casa del cliente con las de la red del operador), gracias al *slicing*. Las soluciones de *slicing* no sólo brindan prestaciones y capacidades de aislamiento, seguridad, y customización similares a las que ofrecen las infraestructuras privadas a un coste más reducido; también brinda capacidades de

flexibilidad y agilidad, permitiendo crear, modificar y eliminar *slices* en cuestión de minutos, asignándoles recursos de forma dinámica, y todo ello siguiendo patrones de provisión y operación fácilmente replicables, es decir, reusables y aplicables para clientes diferentes, en entornos distintos. Esto posibilita crear una economía de escala que reducirá drásticamente los costes, haciéndolos ahora mucho más atractivos para clientes, y haciendo que el 5G privado sea accesible para todos los verticales, incluyendo medianas y pequeñas empresas (PYMES).

- El objetivo último del *slicing* es permitir a los operadores ofrecer Network Slicing as a Service (NSaaS). NSaaS es un modelo de servicio que permite provisionar PNI-NPNs de forma dinámica, proporcionando a los clientes la capacidad de solicitar la provisión de *slices* bajo demanda, usando portales de autogestión. Incorporar este modelo de servicio no es baladí, y los operadores tardarán varios años en preparar sus redes y sistema comerciales para tal fin. Esperar todo ese tiempo no es una opción, ni para el operador (que no puede amortizar las inversiones realizadas con nuevos ingresos) ni para los verticales (que en el corto-medio plazo no van a requerir del uso completo de NSaaS para sus casos de uso). En este contexto, la única opción factible es introducir el *slicing* de forma progresiva, de tal forma que las redes comerciales puedan ofrecer capacidades básicas desde el principio, y sobre esta base, seguir construyendo para ofrecer funcionalidades más avanzadas a futuro.

  Esta tesis presenta un radar tecnológico, con la misión de ayudar a la industria (tanto a los operadores como a su entorno de proveedores y clientes) a definir esta estrategia de lanzamiento de *slicing*, basada en fases. Este radar recoge un conjunto completo de soluciones de *slicing*, y asocia estas soluciones a distintas temporizaciones (actual, corto plazo, medio plazo y largo plazo) en función de su viabilidad técnica y las demandas del mercado. Además, este radar identifica las dimensiones que caracterizan la disponibilidad (dónde y cómo) de estas soluciones, en todos los dominios, incluyendo acceso radio (RAN), núcleo de red (CN), transporte (TN) y sistemas (OSS). En el dominio RAN, las soluciones se han discutido en base a distintos aspectos funcionales (p. ej., desagregación e integración de capacidades O-RAN), a estrategias de asignación recursos de radio (priorización vs partición) y a la penetración en la huella del operador (empezando con coberturas en interiores antes de un despliegue a gran escala). En el dominio CN, las soluciones se basan en combinar las funciones del 5GC en base a criterios de aislamiento, con distintos sabores para distintos tipos de cliente. En el dominio TN, las soluciones se apalancan en la introducción de nuevas tecnologías en el plano de datos, complementadas con las capacidades de automatización y programabilidad que brinda la tecnología SDN en el plano de control. Finalmente, en el dominio OSS, se han abordado aspectos relacionados con la gestión y orquestación, y con la exposición de capacidades a 3ros, cubriendo todas las fases del ciclo de vida de la *slice*.

- El radar pone de manifiesto que sólo tendremos un *slicing* verdadero cuando tengamos 5G SA (5GNR + 5GC). Un 5G SA de Release 15 permitirá a los operadores ofrecer *slices* básicas para el segmento B2C; sin embargo, la verdadera exposición llegará a partir de Rel-16, con verticales migrando desde modelos SNPN (costosos de mantener y difíciles de escalar) a modelos PNI-NPN (más baratos, extensibles y que permite el acceso simultáneo a servicios de datos públicos y privados). La posición de las diferentes soluciones en el radar también muestra que el dominio más avanzado en el soporte de corte es el CN, seguido de la RAN y por último el TN. Hay dos razones que justifican el por qué TN es el dominio menos maduro en *slicing*. Por un lado, se requiere introducir nuevas tecnologías, tanto en el plano de datos como en el de control (SDN); esto llevará tiempo. Por otro lado, el progreso en los estándares en relación al TN *slicing* es lento, existiendo aún muchas brechas entre los modelos 3GPP SA5 e IETF.