5G smarT mObility, media and e-health for toURists and citizenS

Deliverable D3.4
Final architecture and deployment results: Achieved results in the final architecture, technologies and infrastructure deployment

Project Details

Call	H2020-ICT-19-2019
Type of Action	RIA
Project start date	01/06/2019
Duration	36 months
GA No	856950

Deliverable Details

Deliverable WP:	WP3
Deliverable Task:	Tasks T3.1-T3.5
Deliverable Identifier:	5G-TOURS_D3.4
Deliverable Title:	Final architecture and deployment results
Editor(s):	Linas Maknavicius (NOK-FR)
Author(s):	M.Gramaglia, A.Garcia-Martinez, G.Garcia-Aviles, J.Perez-Valero (UC3M), G.Sacco, M.Piccinino (ERI-IT), G.Calochira, P.Scalzo, A.Buldorini (TIM), S.Imadali (ORA-FR), D.Inkielman, A.Oziębło, I.Wojdan (ORA-PL), V.Gezerlis (OTE), L.Maknavicius, B.Sayadi, S.Betgé-Brezetz (NOK-FR), G.Mitropoulos (NOK-GR), I.Labrador (ATOS), I.Belikaidis, E.Giannopoulou, V.Kosmatos, C. Ntogkas, I. Chondroulis (WINGS), C.Thienot (EXP), X.Gilles (AMA), I.Patsouras (ACTA), E.Gatel (BCOM), A.Montilla Vicent, A.Fernandez Sierra (UPV)
Reviewer(s):	M.Gramaglia (UC3M), G.Vivier (SEQ), D.Desirello (RAI), B.Mouhouche (SRUK), S.Provvedi (ERI-IT), E.Giannopoulou (WINGS)
Contractual Date of Delivery:	31/12/2021
Submission Date:	30/12/2021
Dissemination Level:	PU
Status:	Final
Version:	1.0
File Name:	5G-TOURS_D3.4_Final architecture and deployment results_v1.0

Disclaimer
The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Deliverable History

Version	Date	Modification
V1.0	30/12/2021	Initial version, submitted to ECthrough SyGMa.

Table of Content

LIST OF ACRONYMS AND ABBREVIATIONS **6**

LIST OF FIGURES **9**

LIST OF TABLES **11**

EXECUTIVE SUMMARY **12**

1 INTRODUCTION **13**

2 5G-TOURS ARCHITECTURE **15**

2.1 VERTICAL TO NETWORK INTERACTION 15
2.1.1 Network setup through NEST templates *15*
2.1.2 Service interaction though Network Exposure *16*
2.2 OVERALL ARCHITECTURE DESCRIPTION 18
2.2.1 Architecture description *19*
2.2.2 Architecture Instantiation per trial site *21*
2.2.3 Technology integration with 5G EVE *24*
2.2.4 5G security-by-design for verticals *30*

3 5G-TOURS NETWORK INNOVATIONS **37**

3.1 INTRODUCTION 37
3.2 ENHANCED MANO 37
3.2.1 AI-Agent functionality *37*
3.2.2 Service level assurance in 5G RAN MANO enhancement *41*
3.3 AI ORCHESTRATION 43
3.3.1 Outlook on the integration of AI into the major standard architectures *43*
3.3.2 AI approach in resource forecasting in 5GC *45*
3.3.3 AI for Zero-Touch Network Slicing *51*
3.4 BROADCAST SUPPORT 53
3.4.1 LTE-based 5G Broadcast *53*
3.4.2 5GC Multicast *54*
3.5 SERVICE LAYER 57
3.5.1 AI-enhanced MANO *57*
3.5.2 AI-Agents in OSM *61*
3.5.3 Service layer for the vertical closed-loop integration *63*
3.5.4 Multicast/broadcast functionality *64*
3.5.5 Service Layer SDK *66*

4 NETWORK INFRASTRUCTURE & DEPLOYMENT **68**

4.1 TOURISTIC CITY DEPLOYMENT UPDATE 68
4.1.1 Deployment of Physical Infrastructure Phase 2 ..... *68*
4.1.2 Network Equipment *70*
4.2 SAFE CITY DEPLOYMENT UPDATE 72
4.2.1 Deployment of Physical Infrastructure *72*
4.2.2 Network Equipment *74*
4.3 MOBILITY-EFFICIENT CITY DEPLOYMENT UPDATE 77
4.3.1 Deployment of Physical Infrastructure *77*
4.4 5G EVE BLUEPRINTS SUMMARY 82
4.4.1 Turin blueprints *82*
4.4.2 Rennes blueprints *84*
4.4.3 Athens blueprints *84*

5 CONCLUSIONS **87**

ACKNOWLEDGMENT **88**

REFERENCES **89**

List of Acronyms and Abbreviations

3GPP	3rd Generation Partnership Project
5GPPP	5G Public-Private Partnership
4G	4th Generation mobile network
5G	5th Generation mobile network
5GC	5G Core
5G EVE	5G European Validation platform for Extensive trials
5GS	5G System
5G-Xcast	Broadcast and Multicast Communication Enablers for the Fifth Generation of Wireless Systems
5G-MoNArch	5G Mobile Network Architecture for diverse services, use cases, and applications in 5G and beyond
5GT	5G-Transformer
5GT-VS	5GT Vertical Slicer
ABR	Adaptive Bitrate
AAS	Advanced Antenna Systems
ACF	Autocorrelation function
AI	Artificial Intelligence
AF	Application Function
AKA, EAP-AKA	Authentication and Key Agreement / Extensible Authentication Protocol
AL-FEC	Application Layer Forward Error Correction
AMF	Access and Mobility Function
API	Application Programming Interface
APN	Access Point Name
AR	Augmented Reality
ARCEP	L'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (French telecom regulator)
BBU	Baseband Unit
BM-SC	Broadcast Multicast Service Centre
BSCC	Broadcast Service and Control Centre
BSS	Business Support Systems
CDR	Call Detail Record
CMAF	Common Media Application Format
CN	Core Network
CNF	Cloud Native Function
CNN	Convolutional Neural Network
COTS	Commercial Over The Shelf
CPU	Central Processing Unit
C-RAN	Cloud Radio Access Network
CRUD	Create, Read, Update, Delete operations
CUDB	Centralized User Database
DASH	Dynamic Adaptive Streaming over HTTP
DBSCAN	Density-based spatial clustering of applications with noise algorithm
DoS/DDoS	(Distributed) Denial of Service attack
DVB	Digital Video Broadcasting
E2E	End to end
EDGE	Enhanced Data rates for GSM Evolution
EE	Execution Environment
eMBMS	Enhanced Multimedia Broadcast Multicast Services
ENI	Experimental Network Intelligence
EPC	Evolved Packet Core
ESMLC	Evolved Serving Mobile Location Centre
FDT	File Description Table
FEC	Forward Error Correction
GAM	Galleria Civica d'Arte Moderna e Contemporanea di Torino
GMLC	Gateway Mobile Location Centre
GPRS	Generic Packet Radio Service
GSM	Global System for Mobile communications

GTP	GPRS Tunneling Protocol
GUI	Graphical User Interface
HLS	HTTP Live Streaming
HPHT	High Power High Tower
HSS	Home Subscriber Server
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
IGMP	Internet Group Management Protocol
IMSI	International Mobile Subscriber Identity
IoT	Internet of Things
IRP	Integration Reference Point
IWL	Interworking Layer
K8s	Kubernetes
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LCM	Life-Cycle Management
LPLT	Low Power Low Tower
LTE	Long Term Evolution
M-ABR	Multicast Adaptive Bitrate
MAE	Mean Absolute Error
MANO	Management and Orchestration
MBB	Mobile Broadband
MBS	Multicast Broadcast Services
MBSF	Multicast Broadcast Service Function
MB-SMF	Multicast Broadcast Bitrate Session Management Function
MBSTF	Multicast Broadcast Service Transport Function
MBSU	Multicast Broadcast Service User Plane
MB-UPF	Multicast Broadcast User Plane Function
MDAF	Management Data Analytics Function
MDAS	Management Data Analytics Service
MEC	Mobile Edge Computing
ML	Machine Learning
MME	Mobility Management Entity
mMTC	Massive MTC
MoD	Multicast on Demand
MSNO	Multi-Site Network Orchestrator
MTC	Machine Type Communication
MTRL	Market and Technology Readiness Level
NE	Network Element
NEF	Network Exposure Function
NEST	NEtwork Slice Type
NF	Network Function
NFV	Network Function Virtualisation
NFVI	NFV infrastructure
NFV-O	NFV Orchestrator
NMSE	Normalized Mean Square Error
NR	New Radio
NRF	Network Repository Function
NS	Network Service
NSA	Non-Standalone
NSaaS	Network Service as a Service
NSD	Network Service Descriptor
NSO	Network Service Orchestrator
NSSF	Network Slice Selection Function
NWDAF	Network Data Analytics Function
O&M	Operations and Management
O5C	Open5GCore
OAM	Operations, Administration and Maintenance
O-DU	O-RAN Distributed Unit
ONAP	Open Network Automation Platform
O-RAN	Operator Defined Next Generation RAN Architecture and Interfaces
O-RU	O-RAN Radio Unit
OSM	Open Source MANO
OSS	Operation Support System
OTT	Over The Top
OVS	OpenVSwitch
PACF	Partial autocorrelation function
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol

PFCP	Packet Forwarding Control Protocol
PLMN	Public Land Mobile Network
PM	Performance Monitoring
PNF	Physical Network Function
PoC	Proof of Concept
PSO	Particle Swarm Optimisation
QoE	Quality of Experience
QoS	Quality of Service
RAM	Random Access Memory
RAN	Radio Access Network
REST	REpresentational State Transfer
RLC	Radio Link Control
RMSE	Root Mean Square Error
RNIB	Radio Network Information Base
RNN	Recurrent Neural Network
RRH	Remote Radio Head
RTV	Real-Time Video
SA	Service Assurance
SA	Standalone
SA	System Architecture
SBA	Service Based Architecture
SDAP	Service Data Adaptation Protocol
SDN	Software Defined Networking
SDO	Standards Developing Organisation
SDR	Software Defined Radio
SEPP	Security Edge Protection Proxy
SFP	Small Form-factor Pluggable
SLA	Service Level Agreement
SMF	Session Management Function
SNMP	Simple Network Management Protocol
SOM	Self-Organizing Maps algorithm
SPGW	Serving Packet Data Network Gateway
SUCI	SUbscription Concealed Identifier
TCB	(5G EVE) Test Case Blueprint
TLS	Transport Layer Security
TMGI	Temporary Mobile Group Identity
TOI	Transport Object Identifier
TR	Technical report
TS	Technical Specification
TWAMP	Two-Way Active Monitoring Protocol
UC	Use Case
UDM	Unified Data Management
UHD	Ultra High Definition
UMTS	Universal Mobile Telecommunications System
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communications
VDU	Virtual Deployment Unit
vEPC	Virtual EPC
VIM	Virtualized Infrastructure Manager
VM	Virtual Machine
VNF	Virtual Network Function
VNFM	Virtual Network Function Manager
VR	Virtual Reality
VSB	(5G EVE) Vertical Service Blueprint
WEF	Wireless Edge Factory

List of Figures

Figure 1. Overall Network Ecosystem [5]. 19
Figure 2. The 5G-TOURS Network Architecture. 20
Figure 3. 5G-TOURS functional architecture instantiation for the Touristic City site. 22
Figure 4. 5G-TOURS overall functional architecture instantiation for the Safe City site 23
Figure 5. 5G-TOURS functional architecture instantiation for the Mobility-Efficient City site 24
Figure 6. Insertion Methodology. 24
Figure 7. Possible insertion points of 5G-TOURS on 5G EVE platform. 25
Figure 8. UC1 architecture instantiation 26
Figure 9. Integration of ONAP to 5G EVE interworking layer in the French Site. 27
Figure 10. AIA extension location of Athens site for use cases 10, 11, 12 and 13. 28
Figure 11. AIA extension location of Athens site for use cases 6 and 9. 29
Figure 12. Use cases 10, 11, 12 and 13 integration in 5G EVE Greek Site. 29
Figure 13. Use cases 6 and 9 integration in 5G EVE Greek Site. 30
Figure 14. 5G security drivers. 31
Figure 15. Example of multiple stakeholders involved in providing end-user 5G. 31
Figure 16. Security architecture model as defined in TS 33.501 (acronyms used are: ME=Mobile Equipment, SE=Serving Network, HE=Home environment) 33
Figure 17. Network scope for vulnerability analysis. 34
Figure 18. AI-Agents deployed on different VNFs in different Network Services managed from OSM. 38
Figure 19. AI-Agents connected to AI Model Servers. 39
Figure 20. Integrating AI Models. 40
Figure 21. AI-Agents solution deployment. 41
Figure 22. Example plotting the number of connected users in a given cell during a certain period of time. 42
Figure 23. Example of predicted (red) and real (violet) values regarding the number of connected users in a cell. 42
Figure 24. Example of a HeatMap depicting the NMSE monthly average of 19 cells for 14 months. 43
Figure 25. Traffic components in the Telecom Italia dataset. 46
Figure 26. Comparison of 10-minutes samples to 1H samples. 47
Figure 27. Time series decomposition using STL 47
Figure 28. ACF/PACF analysis. 48
Figure 29. SARIMAX and TBATS forecasts comparison. 48
Figure 30. Diagram showing the interaction between the API, Kubernetes cluster and the OS. 49
Figure 31. Diagram showing the application loop. 50
Figure 32. Diagram showing how many more request can be accepted with more resources. 51
Figure 33. The Orchestration Model. 52
Figure 34. Evaluation results. 53

Figure 35. Phase 1 trial's configuration. 54
Figure 36. UC4.b phase 2 trials architecture. 54
Figure 37. Reference point representation of the architecture for 5G multicast/broadcast services. 55
Figure 38. TR 26.802 proposed architecture. 56
Figure 39. 5G-TOURS adapted Use Case 4.c architecture. 56
Figure 40. 5G-TOURS Greek site architecture. 58
Figure 41. AI-enhanced MANO Graphical User Interface. 58
Figure 42. Workflow overview of the Performance Diagnosis tool. 60
Figure 43. AI-enhanced MANO migration operation of critical use case. 61
Figure 44 AI-enhanced MANO performance diagnosis operation. 61
Figure 45. Service Layer integrating AI Agents and AI-Models Server. 62
Figure 46. Closed loop architecture. 63
Figure 47. The integration of the PoC in the overall architecture. 64
Figure 48. Radio positioning at Palazzo Madama ground and first floor. 68
Figure 49. Installation in Palazzo Madama Phase 2. 69
Figure 50. Radio positioning at Palazzo Madama second floor. 69
Figure 51. Radio positioning at GAM. 70
Figure 52. DOT Radio installation simulation at GAM. 70
Figure 53. EVER building blocks [34]. 71
Figure 54. Overall Turin Network infrastructure. 72
Figure 55. 5G-TOURS 5G NR NSA wireless coverage at BCOM. 73
Figure 56. 5G TOURS 5G NR wireless coverage in the Wireless Operating Room at CHU. 73
Figure 57. Overall network architecture and physical deployment of network equipment and functions. 74
Figure 58. BBU, RRH 4G and RRH 5G under integration phase in BCOM labs. 75
Figure 59. 5G-TOURS integration with 5G EVE in Rennes 76
Figure 60. UPF deployment over OpenStack execution environment. 77
Figure 61. Final Athens node infrastructure for the needs of 5G-TOURS. 79
Figure 62. Outdoor Radio equipment installation at AIA 79
Figure 63. ACTA Probes and Server topology in Athens node. 80
Figure 64. Pictures of actual probes and ACTA's server (KMVaP) at OTE Labs. 80
Figure 65. High level view of 5G-TOURS RAN and CORE Network infrastructure at Greek Node. 81
Figure 66. RAN coverage of AIA for Smart Parking UC10 and Video enhanced airfield vehicles UC11 81
Figure 67. RAN coverage of AIA for Evacuation UC12 and AR/VR Bus Excursion UC13. 81
Figure 68. UC1 high-level infrastructure instantiation. 82
Figure 69. UC10 high level overview. 84
Figure 70. UC10 VSB overview. 84
Figure 71. UC10 VSB list of application metrics. 85

Figure 72. UC10 NSB overview. 85
Figure 73. UC10 TCB overview. 86

List of Tables

Table 1. The exposure to capability mapping..... 18
Table 2. List of UCs that will leverage on the “Connection” option..... 21
Table 3. 3GPP Security Features – 5G versus 4G..... 33
Table 4. Management Data Analytics Service. 44
Table 5. TIM Lab Core Capabilities..... 71
Table 6. Prerequisites for IaaS BCOM/UPF. 76

Executive Summary

This report provides the final update out of four planned for the 5G-TOURS project presenting work performed in WP3 "Network architecture and deployment". The scope of this deliverable is to report on the overall progress made in the 5G-TOURS network and system architecture and innovative technologies for the 5G network deployment in three trial sites (Turin, Rennes, and Athens).

This report focuses on the following aspects:

- Finalization of the baseline network architecture which encompasses 5G EVE platform, relevant standards components and 5G-TOURS innovations, as well as architecture instantiation per trial site and interactions with vertical players.
- The technology evolutions conceived in the 5G-TOURS project, their implementation results, Service Layer components and theirs SDK descriptions.
- The final set of the network infrastructure deployments on each trial site.

In particular, several important outcomes from the architectural work, network innovations and trial deployments can be highlighted. Firstly, a comprehensive architectural setup was designed to suit vertical actors' needs, facing 5G-TOURS Service Layer (which integrates useful network exposure functions for the service "closed loop" e.g. KPI monitoring), incorporating 5G EVE interworking gear (for service onboarding through 5G EVE portal) and/or direct connection to local MANO functions (where exposure of specific functionalities is needed directly towards the Service layer). Then, specific 5G-TOURS network innovations were devised, such as Artificial Intelligence functions for network resource forecasting, dynamic slicing and orchestration support (showcased through successful Proof-of-Concept demos prepared by several partners), some others directly integrated within a selection use cases, e.g. state-of-the-art solution for the 5G broadcast support (in Turin site), Enhanced MANO extensions (largely deployed in Athens trial). Also, open-source initiatives were developed for the Service Layer to support some of these technologies, with SDK and API documentation, and release of full open-source software solutions and largely publicized, e.g., AI Agents for OSM. Finally, thorough design and deployment of complex network setup was achieved in three trial sites, integrating equipment, system solutions and application software from different suppliers towards in-the-field realization of a set of distinct use cases which improve citizen lives: media and touristic applications in a world heritage protected environment (Turin site), novel 5G mmWave radio solution and eHealth appliances' integration within restricted hospital environment (in Rennes site, one of the first experimental eHealth deployments in Europe), mobility-efficient city applications with full integration of diagnostic / network measurement extensions to achieve "closed loop" between Service layer and networking (in Athens site).

1 Introduction

This deliverable concludes the work of the 5G-TOURS network innovations carried out within WP3 "Network architecture and deployment". The 5G-TOURS project goal is to demonstrate the benefits of the 5G technology and integrated services in a pre-commercial environment for real users, tourists, citizens and patients and the respective vertical players, by implementing 13 representative use cases in 3 different types of cities:

- **Turin**, the touristic city (5 use cases),
- **Rennes**, the safe city (4 use cases) and
- **Athens**, the mobility-efficient city (4 use cases).

The objective of WP3 was to develop architectural vision and new enabling technologies of 5G-TOURS as well as extending the deployments of 5G EVE. Specifically, the key aims of WP3 were the following:

- development of an architecture based on innovations from 5G-TOURS that leverage and build on previous projects' innovations and standardization development,
- realisation of the necessary assessment and expansion of the 5G EVE network deployment for the three trial nodes in Turin, Rennes and Athens, and
- introduction of the developed technical innovations as part of the expanded infrastructure.

Essentially, WP3 is mostly concerned with research activities related to 5G network, system architecture and underlying technologies, through the lifecycle of scientific innovations: defining research challenges, prototyping, verifying hypothesis and algorithms, developing solutions, and then rooting them into architectural instantiations where feasible. Some academic papers are continuously published (e.g. IEEE INFOCOM, IEEE Network, IEEE JSAC, etc. [33], [36]) and standardization work produced (mainly at 3GPP [38-42]).

The first deliverable of WP3 D3.1 [1] described the initial design ideas for the baseline architecture, technologies and trial deployment objectives, alongside a section on 5G EVE [53] on-boarding and technological integration strategies.

The second deliverable of WP3 D3.2 [2] illustrated the initial progress on the architecture (as was designed to accommodate all the novel technologies developed during the project to support and enhance the proposed use cases) and physical deployment in 5G-TOURS. A description of the capabilities, including use case on-boarding and 5G EVE integration was also covered. The progress on the development of 5G-TOURS technologies were described. The status in standardization fora was covered as well as the implementation and integration of the subjacent NFs or algorithms into 5G-TOURS architecture.

The third deliverable D3.3 [3] covered the progress to date of WP3 ongoing work, as the project transitioned from Phase 1 to Phase 2 deployment (Phase 1 of the project covered a limited area in the cities, where pre-commercial network equipment and early application solutions were deployed, while Phase 2 concerned larger network scales, combining 5G EVE infrastructure and full system integration with 5G commercial equipment in trial sites). It illustrated the status of the network infrastructure deployment for 5G-TOURS and technological innovations, explaining advancement of their implementation.

This final deliverable D3.4 presents the final set of architectural and technological solutions, as well as ultimate deployment results in trial sites. This last stage was needed to support vertical requirements within the overall framework and finalize development of network innovation solutions.

The description of how the deliverable is structured follows. Section 2 introduces the requirements coming from the Use Case descriptions and their NEST definitions [4], discusses resulting vertical-to-network interactions (including network exposure and capabilities), provides the final update on the overall 5G-TOURS architecture (network and system), its high-level instantiation per trial site, integration aspects with 5G EVE platform, as well as considerations for 5G security-by-design for the verticals.

Section 3 presents the final set of the 5G-TOURS network-related innovations, specifically:

- MANO enhancements,
- AI orchestration solutions,
- 5G broadcast support,
- Service Layer, oriented towards verticals.

Each of those innovation areas encompasses two technological solutions, except for the Service Layer which counts five specific solutions. Some of those solutions enrich several 5G-TOURS use cases (e.g., 5G broadcast, also AI-enhanced MANO which is closely integrated with 5G EVE interworking layer and expands it in 5G-TOURS Greek trial site), or could possibly be applied to some of them (e.g., 5G core resource forecasting, Service assurance in 5G RAN MANO), others are stand-alone innovations resulting from the partners' research work (e.g., ETSI ENI PoC, AI-agents for OSM). Particular attention is given to the Service Layer SDK (Software Development Kit) descriptions linked to open-source initiatives in sub-chapter 3.5.5.

Finally, Section 4 describes the progress of the network deployments for 5G-TOURS. For each of the trial sites, it indicates how the 5G-TOURS architecture and innovations are applied to implement the targeted use cases, and the realized physical infrastructure.

2 5G-TOURS ARCHITECTURE

In this section, we describe the 5G-TOURS Network Architecture. This architecture has been designed with a set of requirements in mind, coming from the interaction with WP2 on the generation of the network slice blueprints through NEST templates, an activity that is also reflected in WP7. In the following, we first discuss the requirements coming from WP2 on the architectural design perspective, then we have an in-depth analysis of our designed architecture, structured along different axes.

2.1 VERTICAL TO NETWORK INTERACTION

In order to support the 5G-TOURS envisioned use cases, the WP3 included enhanced additional functionality that was developed on top of the existing 5G EVE platform. These development activities in some cases introduced additional side functionalities, like monitoring features, which were outside the scope of the 5G EVE project, or included some cornerstone network function, as for the case of broadcast support, AI-enhanced MANO, a functionality that is mandatory for the correct realization of the UCs. More specifically, the 5G-TOURS additions to the baseline 5G EVE ecosystem can be summarized as:

- More and diverse network infrastructure deployment, as thoroughly discussed in Section 4.
- Targeted vertical to network interaction, through the usage of the 5G-TOURS Service Layer. This includes the API for the usage of Artificial Intelligence, Broadcast, and enhanced monitoring (as discussed in Section 3).
- The specific algorithms implementing these functionalities (e.g., the AI agents).

The workflow took place as follows: the UC owners introduced requirements (by their initial description, but also through the NEST templates definition [4]) which were then fulfilled by WP3.

2.1.1 Network setup through NEST templates

Through the definition of NEST templates [4], UC owners defined their requirements on the network architecture, that is, novel features that are not directly related with hard KPIs such as maximum latency or minimum bandwidth, but rather soft KPIs related to the usability of the system.

2.1.1.1 **Network Exposure and Vertical Integration**

Vertical service providers need to have a more direct interaction with the network to increase the integration of their business logic developed with the network intelligence. In this context, an enhanced network exposure functionality from the operator, which has to make available fundamental capabilities such as the recorded network metrics to allow the service provider to tailor the operation of the service according to the changing environmental conditions, is needed. While we discuss in depth the enhanced exposure functionality in section 2.1.2, some of the key functionalities have been implemented in 5G-TOURS by the service layer, as discussed in Section 3.5. In particular, the service layer provides ways to foster the interaction between verticals and operators, in the following ways:

- Continuous KPI monitoring though network exposure functionality;
- Programmatic onboarding of network slices though the 5G EVE portal.

The selected KPI and the monitoring granularity are defined and required from WP2 through the NEST templates.

2.1.1.2 **Service Layer / AI and Vertical Integration**

Artificial Intelligence(AI) is the most prominent tool used by network operators in 5G-TOURS to optimize the resource provisioning within the network. More specifically, operations such as intelligent resource assignment and service optimization are provided through AI, as discussed in Section 3.3. Specifically, in 5G-TOURS we target the network management through AI from the operational cost perspective, helping operators to optimize

that cost. Additionally, AI is also helpful to support the service layer operation bridging the intelligences developed in the network infrastructure and the ones developed by service provider. While this has not be implemented in one of the 5G-TOURS use cases it has been initially showcased in the 5G-TOURS supported PoC in the ETSI ENI ISG [52].

2.1.1.3 Broadcast and Vertical Integration

Use Case 4 sub use cases heavily rely on broadcast support. So 5G-TOURS partners fulfilled this functionality by providing an implementation of a 3GPP Rel. 16 High Power High Tower solution, and by the multicast-enabled 5G Core deployed in the UPV premises and connected to the 5G EVE Italian site as an additional branch. More information on this aspect can be found in Section 3.4.

2.1.1.4 Other required innovations

The very stringent requirements in terms of bandwidth and latency imposed by several UCs challenge even the most innovative hardware deployment available in 5G EVE, which has been planned just a couple of years ago. This is a clear indication of the pace at which the 5G deployment runs. Therefore, for some use cases 5G-TOURS partners deployed new hardware that either provides superior performance such as the 26 GHz network deployed in France or includes 5G functionality in challenging environments such as the museum in Turin (which is protected by cultural heritage regulations) or the Hospital in Rennes (where the network deployment is fully integrated with the existing medical infrastructure deployment). As discussed in Sections 4.1, 4.2, and 4.3, this required a very thorough engineering of the new deployment, which we consider as important innovations carried out by the project.

2.1.2 Service interaction though Network Exposure

A common distinctive aspect for the next generation of mobile networks with respect to legacy ones, is their tighter interaction between the service provider and the network operators. The final goal is achieving a continuum between the end users and the provided services through the network infrastructure, that is tailored for the specificity of the envisioned application. As a result, the network becomes a kind of "commodity" for the service provider, that has to be managed as many other infrastructure deployments.

This totally new paradigm, enabled by network softwarization and programmability, also allows to employ data-driven solutions for steering the operation of the network. However, as any other algorithm based on AI, the availability of input data (used, e.g., to train models) and the possibility of enforcing the decisions stemming from these models into the network becomes fundamental.

2.1.2.1 Network Domains

With the arrival of the network softwarization concept, different functionalities in the network adopted a software-driven approach. Among the most important examples, we can list the Service Based Architecture (SBA) in the network core, and the xApps defined in the O-RAN controller hierarchy. However, the flexible interaction guaranteed by the open interfaces devised by these approaches are traditionally confined into one specific domain (e.g., network functions), while the software landscape in the context of mobile networks is much broader, as we depict in Figure 1, we next discuss each of them in details:

The Network Functions Domain: The main functions and interactions of the functions within the RAN and between RAN and CN have already been specified by 3GPP. New Radio (NR) includes the Service Data Adaptation Protocol (SDAP) layer in the user plane, which enables the mapping of QoS flows to the radio bearers increasing the degree of freedom for QoS enforcement in RAN, and the F1 interface which enables the central unit-distributed unit (CU-DU) split between packet data convergence protocol (PDCP) and radio link control (RLC). The 5GC, instead, comprises also new NF/NEs supporting service based communication between 5GC CP NFs/NEs, such as the Network Repository Function (NRF) and Network Exposure Function (NEF), that are specified to enable the service registration and service discovery of the 5GC CP NFs/NEs in the same domain (e.g., 5GC of a network provider). Network Data Analytics Function (NWDAF) is another new NF introduced in 5GC. NWDAF represents operator managed network analytics logical function.

The Network Management Domain: Starting from 3GPP Release 15, the 3GPP Operations, Administration, and Maintenance (OAM) domain, in the following also referred to as management plane, has introduced the Service-Based Management Architecture (SBMA). In this framework, a management service offers management capabilities. The most essential management services include generic provisioning, fault supervision, and performance assurance management services, which are typically produced by the NF or a lower management layer.

The Orchestration Domain: While the traditional network domains (i.e., NFs and management plane) were already present in legacy networks (i.e., up to 4G/LTE), their Orchestration and the (advanced) configuration experienced an incredible boost, mainly due to the introduction of the SDN, NFV, and containerization technologies. Given its scope of software-driven, application-agnostic management of general-purpose cloud resources, orchestration and lifecycle management procedures have not been tackled by 3GPP. For instance, the orchestration of VNFs or containers is not being included in the 3GPP work. As a matter of fact, orchestration of network resources is currently achieved through vendor-specific and standard-compliant solutions.

The Service Provider Domain: The novel network softwarization paradigm introduced by 5G and beyond 5G networks allowed for a diverse and heterogeneous landscape of tenants. That is, different service providers such as industrial verticals, are now a fundamental piece involved in the network operation, in clear contrast with what has been traditionally happening up to the legacy 4G networks. The 4G network provisioning is characterized by a full Over The Top (OTT) service delivery model. Instead, the new model can provide a more integrated view of the system from the tenant and service provider perspective, who can, by leveraging on novel configuration primitives, act on the underlying network slices. From the standardization point of view, this concept has been increasingly attracting more attention. For instance, 3GPP currently defines two management models: Network Operator Internals (i.e., the legacy solution, in which the tenant has no visibility on the underlying system) and NSaaS in which the tenant can manage the network slice as manager via an exposed management interface, and optionally provide network slicing capabilities to other tenants. This case is also envisioned by 5G-TOURS with all the vertical service providers.

2.1.2.2 Network Capabilities

As previously discussed, although different elements of the state-of-the-art network architectures already provide data-driven functionality, their limited scope (very often bounded to one specific domain) may hinder the automated operation of the network that involves cross-domain activities such as reactive orchestration upon network triggers, or service-driven network re-configurations that are at the basis of the autonomous operation of the network.

Capability type 1: Monitoring and data collection. This capability is related to the provisioning of raw monitoring data from and to different Network Elements, and it is already implemented for some network domains. For instance, NFs can provide KPIs related to the cell performance (e.g., handover failure rates, cell load, etc.), user-centric KPIs (e.g., per user throughput, latency, etc.) and KPIs related to the end-to-end service performance. Also, monitoring data of the infrastructure (e.g., CPU and RAM utilization) falls into this category. Finally, this capability type includes the capability to define customized measurement jobs, trace collection configurations, or real-time performance measurements on the monitored element.

Capability type 2: Triggers, alarms, and fault supervision. While the data collection capability discussed before provides a way for collecting information from Network Elements at full granularity, there is the need for a more refined way of accessing it. That is, most of the elements in the network need to react according to well-defined state machines upon triggering events.

Capability type3: Actions, control, and configurations. Besides exposing data as well as alarms and events, Network Elements of different domains shall also expose configuration and control capabilities to other elements. Generically, this capability type comprises capabilities to act, i.e., to create, modify, delete objects as well as their parameters and configuration attributes.

Capability type 4: Network intelligence, and policy recommendation and enforcement. Future network operation will be intelligent. This means that most of the tasks that currently require human intervention to achieve optimality in the network will be handled automatically and with some kind of AI in the loop. Thus, Network Elements shall expose the capability of performing complex analytics on inputs coming from other

elements (i.e., the ones exposed by capability types 1-3 above), from which the operator can devise or enforce policies.

2.1.2.3 Capability exposure through the service layer

In Table 1 below we map the different network domains with the exposure capabilities, matching the kind of capabilities that shall be provided by the network. More specifically, we focus on the row D in the context of the service layer (when the service provider is a producer), while all the other metrics may be consumed by the service provider.

Table 1. The exposure to capability mapping.

Capabilities / Producing Network Domains	(1) Monitoring and data collection	(2) Triggers, Alarms and Fault Supervision	(3) Actions, control and configurations	(4) Network Intelligence and Policy Recommendations
(A) Network Functions	NW resource utilization UE traffic conditions UE counters 5GC counters	NW resource failure QoS unfulfillment Network Functions SW exceptions	NRM parameters Procedure (ICIC) Parameters Mobility Management	NWDAF RAN Analytics Long Term RRM
(B) Management	Cell Traces Network slice counters	Cell outages Slice-level SLA failures	SON Slice lifecycle management	MDAS
(C) Orchestration	NFVI monitoring WAN monitoring	NFVI alarms WAN links failures	VNF placement decisions VNF deployment flavor	AI as a service VNF placement algorithms Root Cause Analysis
(D) Service Providers	Manufacturing process monitoring Application Service Status	Manufacturing line failures Massive churn rates	Production cell layout reconfigurations Expected traffic patterns	Service domain analytics Business intelligence

In 5G-TOURS we implemented a few of such capabilities exposure through the service layer, namely the one related to the exposure of network intelligence towards the vertical (capability type 4), and monitoring and data collection (mostly KPIs, type 2), and Action, Control and Configuration of the Broadcast solutions (capability type 3).

2.2 OVERALL ARCHITECTURE DESCRIPTION

For the design of the overall 5G-TOURS architecture we started from the analysis of the different domains (besides the Network one, which has already been discussed in the previous section), analyzing how they build together into a compound network ecosystem.

To this extent it covered the 5G System (5GS) as a whole and discussed end-to-end (E2E) network slicing, service-based architecture, Software-Defined Networking (SDN), Network Functions Virtualisation (NFV), Management & orchestration, and E2E service operations & lifecycle management as the fundamental pillars to support the 5G Key Performance Indicators (KPIs). Given the new requirements coming from new stakeholders in the 5G ecosystem which have been described earlier, the recent advances in the softwarization of the mobile network ecosystem as well as the recent releases of the relevant standards for access, core, management and orchestration, we can draw architectural trends that are captured in the 5G-TOURS network architecture. A further trend that is newly introduced and that is quite intrinsic is the concept of Non-Public Networks (NPN). Sometimes called private network, an NPN provides 5G network services to a clearly defined user organisation or group of organisations and is deployed on the organisation's defined premises, such as a campus or a factory.

Owing to the architectural representation made by, e.g. the 5G PPP architecture whitepaper [5], we integrated the trends that form novel architectural aspects and which became very influential in the implementation of our project. The envisioned ecosystem is depicted in Figure 1 below, and comprises three main areas: the verticals, the network, and the infrastructure. These are then mapped into the 5G-TOURS architecture discussed in Section 2.2.1 below.

The Service Domain for Verticals includes all architectural innovations that help to include the business-related considerations to the offered services (among others, e-health, robotics, or enhanced video streaming services). Here, the key role is played by two innovations which have been considered in project, namely: the service layer and the concept of NetApps developed by UC owner. The service layer, which is described in Section 3.5, provides a common interface towards the management and the operation of the network, enabling the interaction

between the service intelligence and the underlying network. The concept of NetApps comprises all 5G network empowered applications that build a network service, through the usage of network slices. Slices are then used to provide such network services and encompass different network functions (including core and access functions), possibly orchestrated over different clouds.

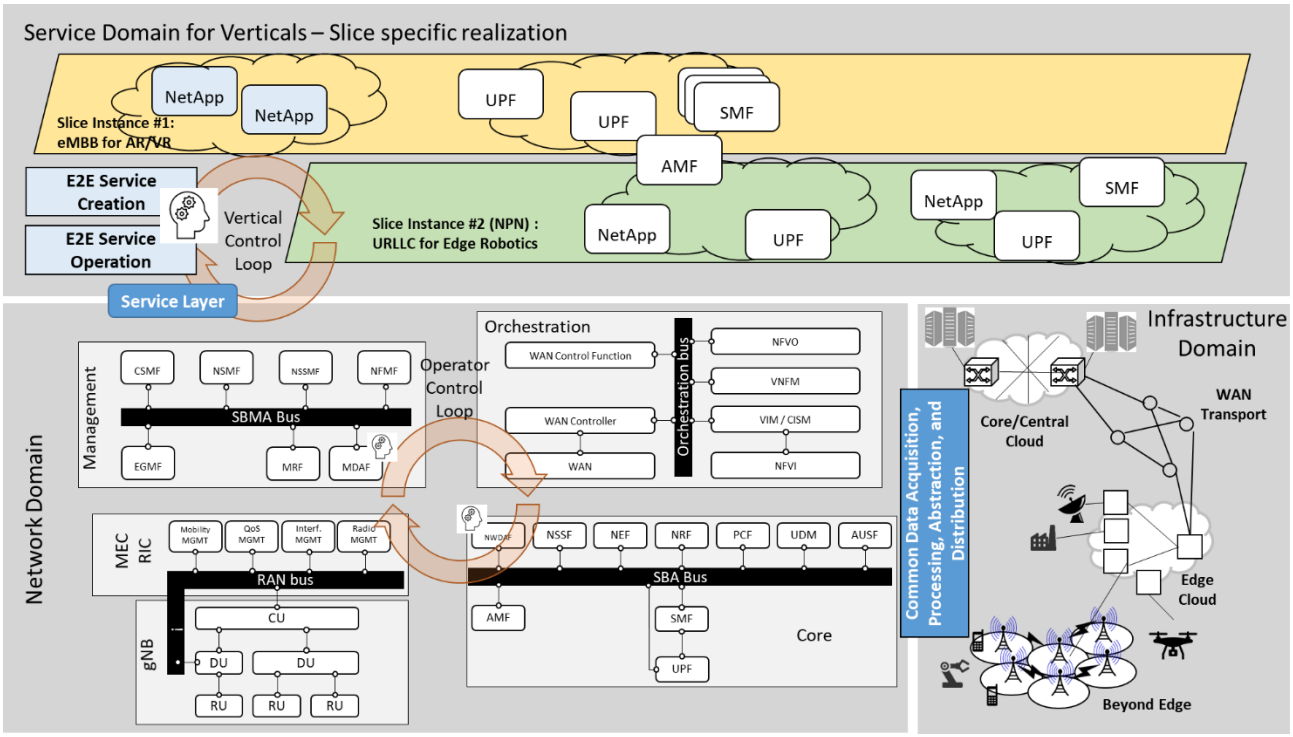


Figure 1. Overall Network Ecosystem [5].

The different functions are operated in the Network Domain, arranged in different slices according to the KPIs that they have to provide, as discussed in Section 2.1.2.1. Innovations in the Infrastructure domain are captured in the context of the specific instantiations, per trial site. The architecture natively supports the quest for network automation that is achieved through control loops and the usage of artificial intelligence algorithms. Specifically, we identified two main loops: the first loop enabled by the service layer that is leveraged by the service provider through the NetApps to steer the behaviour of the network, and the second loop that happens within the network domain, with specific modules such as the network data analytics function (NWDAF) or the other elements in the management that are specifically designed for this purpose.

2.2.1 Architecture description

The 5G-TOURS Network Architecture is depicted in Figure 2 below. It encompasses three different domains (Verticals, Network, and Infrastructure), in a layered fashion. As already discussed in D3.3 [3], the different layers are:

- The 5G-TOURS Verticals, that represent all the application ecosystem providing the UC related functionality. These VNF NetApps available in this layer, are then onboarded in the network using the Service Layer, either directly leveraging the 5G EVE portal, or the direct connection to the underlying MANO service available at each site/infrastructure deployment.
- The 5G-TOURS Service Layer implements all the facilities needed by NetApps to be fully integrated with the underlying network architecture. In 5G-TOURS this layer is provided by means of extensions to the 5G EVE portal and specific 5G-TOURS Service Layer solutions described in Section 3.5. However, as some UCs require a more direct interaction with the underlying MANO and Infrastructure to support specific operation, we devised an architectural option for this aspect, called “Connection”.
- The MANO, local to each site. This layer is mostly composed by 5G EVE configured elements, with the addition of 5G-TOURS specific elements, which are usually tightly integrated with the service layer,

using the connection. The MANO also includes AI specific modules. The full definition of the specific management and orchestration implementation (which are different for each site, as they are inherited from the 5G EVE) are described in the specific subsections of the section 4.

- The VNFs (Core and Access) available at each site. Again, these assets are partly inherited by 5G EVE, but they have been integrated with specific 5G-TOURS technology, especially in the access network side, but also for the specific multicast-enabled elements in the core.
- Finally, the infrastructure, mostly related to provide access network to the final trial sites (e.g., the museums in Turin, the hospital in Rennes, and the airport in Athens), but also related to virtualization deployed into each site, as discussed in Section 4.

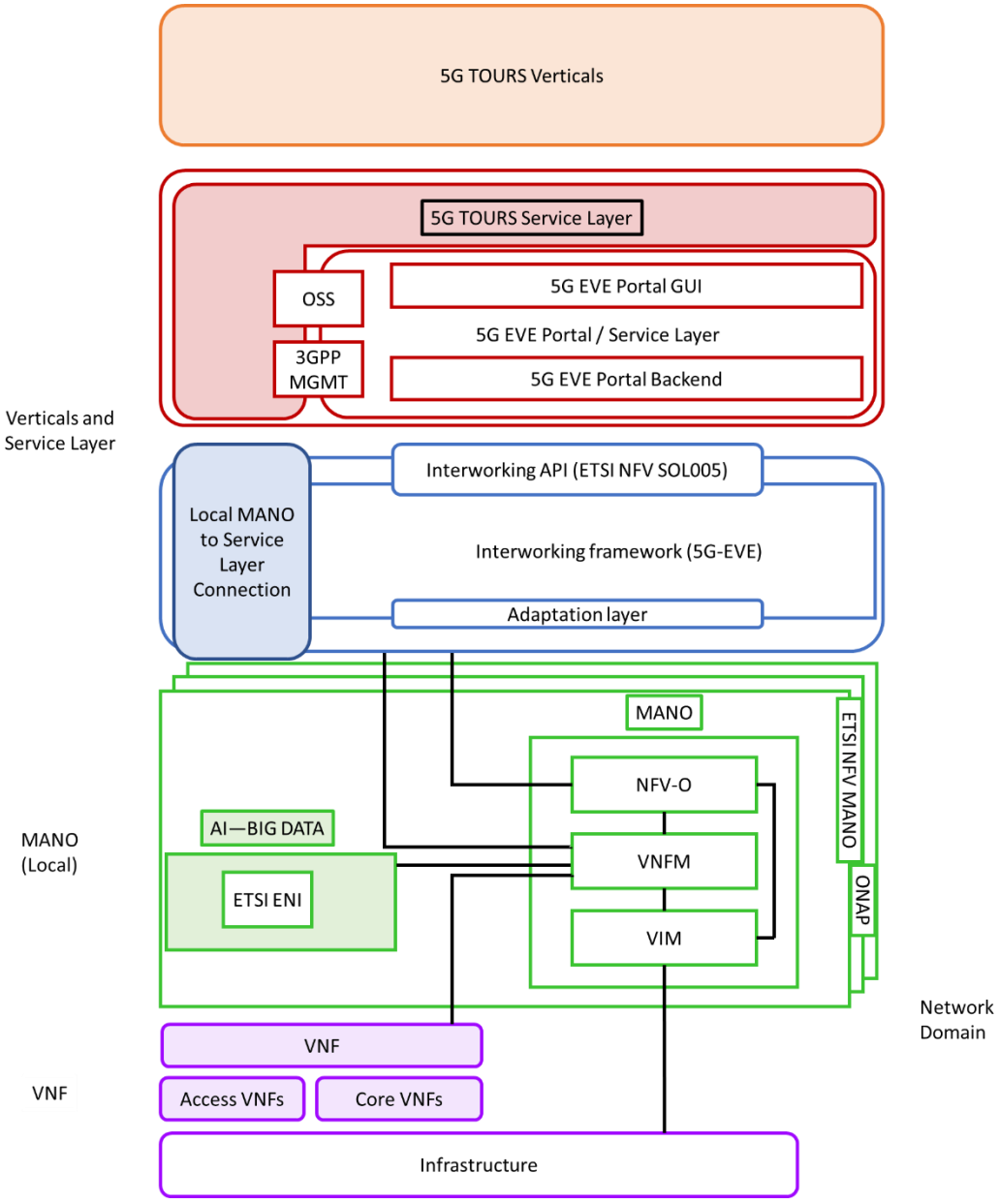


Figure 2. The 5G-TOURS Network Architecture.

The UCs are executed on enhanced 5G EVE platform, with possible integration of specific or hardware/software components. While their integration is discussed in section 2.2.3, we focus here on the UCs that use the direct connection between Service Layer and network operator's MANO.

The architecture as described above includes an architectural option, called "Connection" (in previous documents we called this option "By-pass") that enables the direct interaction between the vertical service provider

and the network operator. This functional requirement arose during the gap analysis for the 5G EVE platform we conducted and detailed in D3.2 [2], which was largely driven by the support of the vertical partners present in the 5G-TOURS consortium that gave feedback on the specific needs for the interface related to the specific use cases of 5G-TOURS. As it was designed and implemented, 5G EVE has an open loop between the network and the verticals, which can perform onboarding and termination of network slices through the 5G EVE portal but cannot access it at runtime. Instead, this is a very important functional requirement for the 5G-TOURS consortium verticals. Thus, 5G-TOURS tackled this issue with two activities, as follows:

- Integrate and extend as much as possible the 5G EVE platform to support our closed-loop needs, which are mostly related to the functionalities associated with the service layer. This is the case of the implementation of the service layer performed in the Greek site and discussed in Section 3.5.1, which will expose Diagnostic information directly to the verticals through the 5G EVE portal.
- Provide an architectural option to support the limited cases in which this integration will finally not be possible. We would like to remark that this option will be leveraged only to include the functionality that is strictly needed for the exposure of e.g. AI and Broadcast-related information through the Service Layer.

Due to the experimental activity performed in the former point, the full list the UCs that make use of the Local MANO to Service Layer Connection is detailed in Table 2.

Table 2. List of UCs that will leverage on the "Connection" option.

Use Case	Reason for using "Connection" option
UC2	The NetApp developed by the UC owner included proprietary software that could not be fully virtualized, requiring specific hardware not directly provided by 5G EVE and difficult to be integrated
UC3	The robot control app required a graphical user interface that could not be fully onboarded on the 5G EVE infrastructure
UC5	The video processing software required specific hardware which could not be natively integrated into the 5G EVE infrastructure
UC 7	The Connection is needed in order to support specific terminals that cannot be integrated with 5G EVE gear
UC 11, 13	Not all the components of the applications can be virtualized, (e.g., UHD video processing, VR/AR components)

Basically, 5G-TOURS researchers resorted to these direct connections only when the specific infrastructural support provided by 5G EVE was not enough to fulfill the usability characteristics of the NetApps. However, we remark that all the rest of the infrastructure is fully integrated with the 5G EVE ecosystem, as discussed next.

2.2.2 Architecture Instantiation per trial site

This section presents high-level "instantiations" of the architecture in the different trial sites, depicting technical innovations of the project as VNF/CNF/PNF functions. This highlights how the 5G-TOURS baseline architecture is applicable to various sites and more importantly that the results were achieved pooling different network equipment, CPEs, gateways and specific user appliances from various vendors. This proved that the consortium developed a generic architectural approach which can be reusable very quickly and so with high impact. The architecture instantiations are given for the phase 2 of the project, where the fully fledged network solutions are developed. The readers interested in transitions from Phase 1 to Phase 2 will confer to the 5G-TOURS D3.3 deliverable [3]. The infrastructure deployments are further detailed in Section 4.

2.2.2.1 Italian site

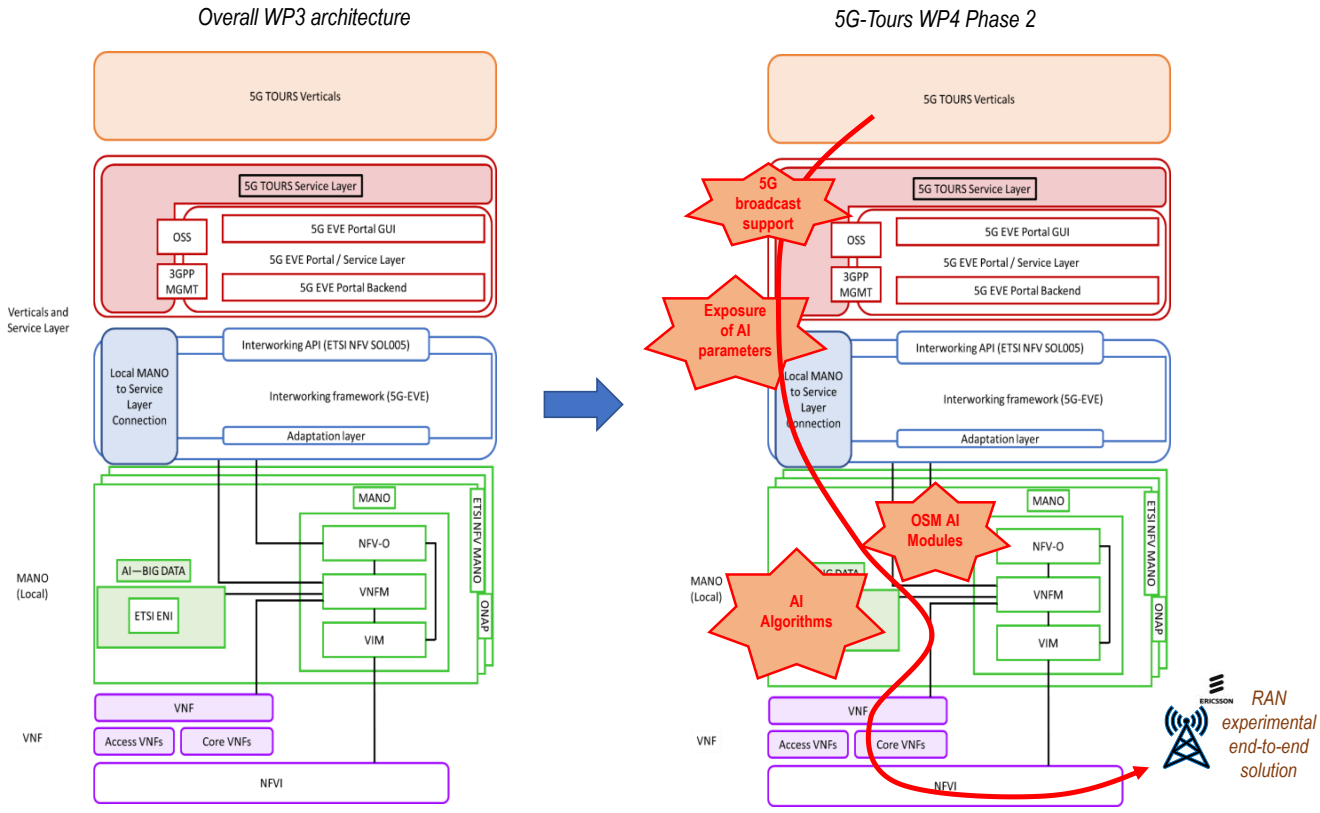


Figure 3. 5G-TOURS functional architecture instantiation for the Touristic City site.

Enhancements of the Italian site are integrated for phase 2 on the architecture solution as follows:

- The integration of the AI agents innovation into the latest release of the Open Source MANO installation in the 5G EVE site, this allows for a flexible orchestration of the system, especially for the part related to the application VNFs;
- The introduction of specific AI algorithms in the network such as the ones already demonstrated in the ETSI ENI PoC (see in Section 3.5.3.2), or further implementation of AI algorithms such as the ones described in Section 3.3.3;
- The exposure of some AI configuration parameters to the vertical. While the AI algorithms deployed in the network take care of the low level details of the configuration of VNFs, a 'knob' to configure them is available to the verticals through the service layer;
- Specific indoor end-to-end RAN deployment to respect the world heritage preserved environment.

2.2.2.2 French site

The functional setup in Rennes site for phase 2 is as follows:

- The 5G new-radio RAN is deployed over the 26 GHz band, to guarantee very high bitrates over a well-defined area (one of the first to be deployed in the context of 5G PPP projects);
- The core network (based on the BCOM WEF 2.2) is orchestrated using ONAP on the 5G EVE infra-structure, in a remote location (the Orange Labs in Chatillon). On the other hand, edge functions (RTV/AR servers) are located closer to the real applications;
- Use of CNF functions (cloud native using containers and microservices) on the control plane instead of VNF, which has become a new telecom virtualization trend in the last few years. This solves the prob-lem of optimal resource allocation (key value to support correct SLA and deliver to proper QoS) and reduces the cost of virtual infrastructure (relevant when AMF server handles IoT traffic such as in UCs

6,9,10). The architecture of the ONAP integration with 5G EVE portal is presented in section 2.2.3 in French Site. A PoC with AI/ML model empowering CNF rescaling is described in Section 3.3.2.

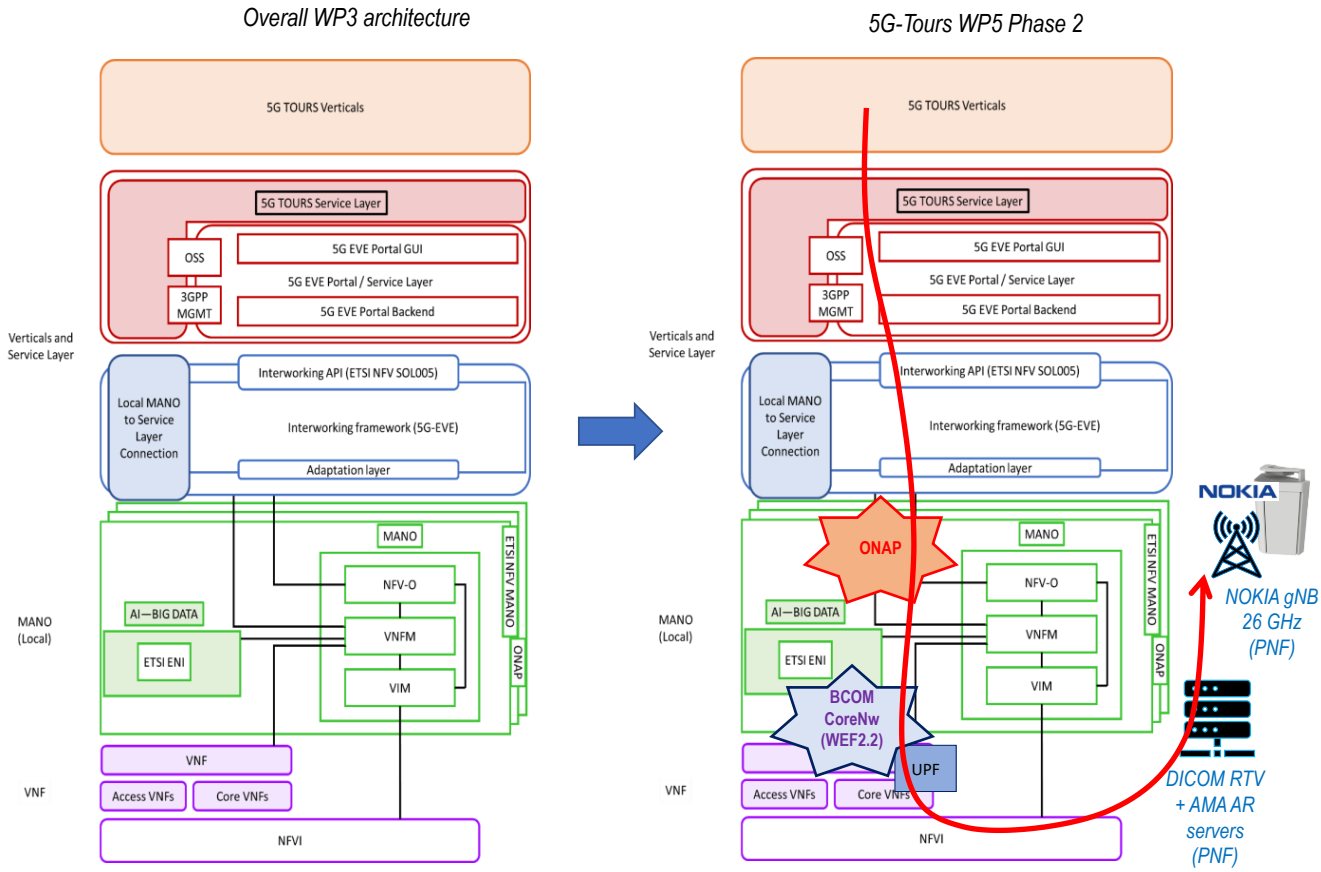


Figure 4. 5G-TOURS overall functional architecture instantiation for the Safe City site

2.2.2.3 Greek site

Phase 2 in Athens site is providing the following enhancements over original architecture:

- A 5G new-radio SA solutions, relying on the 5G Core functions implemented by Nokia (more details can be found in Section 4.3).
- The exposure, through the service layer, of specific network metrics gathered through the probes that are deployed in the network, as well as AI-enhanced MANO gear (more details in Section 3.5.1).
- A VPN interconnection with Rennes site has been established for the the need of running a Multisite UC8 (assistance of a doctor from Athens during a simulated intervention in the Wireless operating room in Rennes).
- Installation and configuration of TWAMP/VIAVI software in BBUS and Probe server for Latency, Throughput, Jitter, packet loss measurements of different segments of the E2E Network.
- Installation of new back-end VMs for the needs of the AR/VR content of UC13.

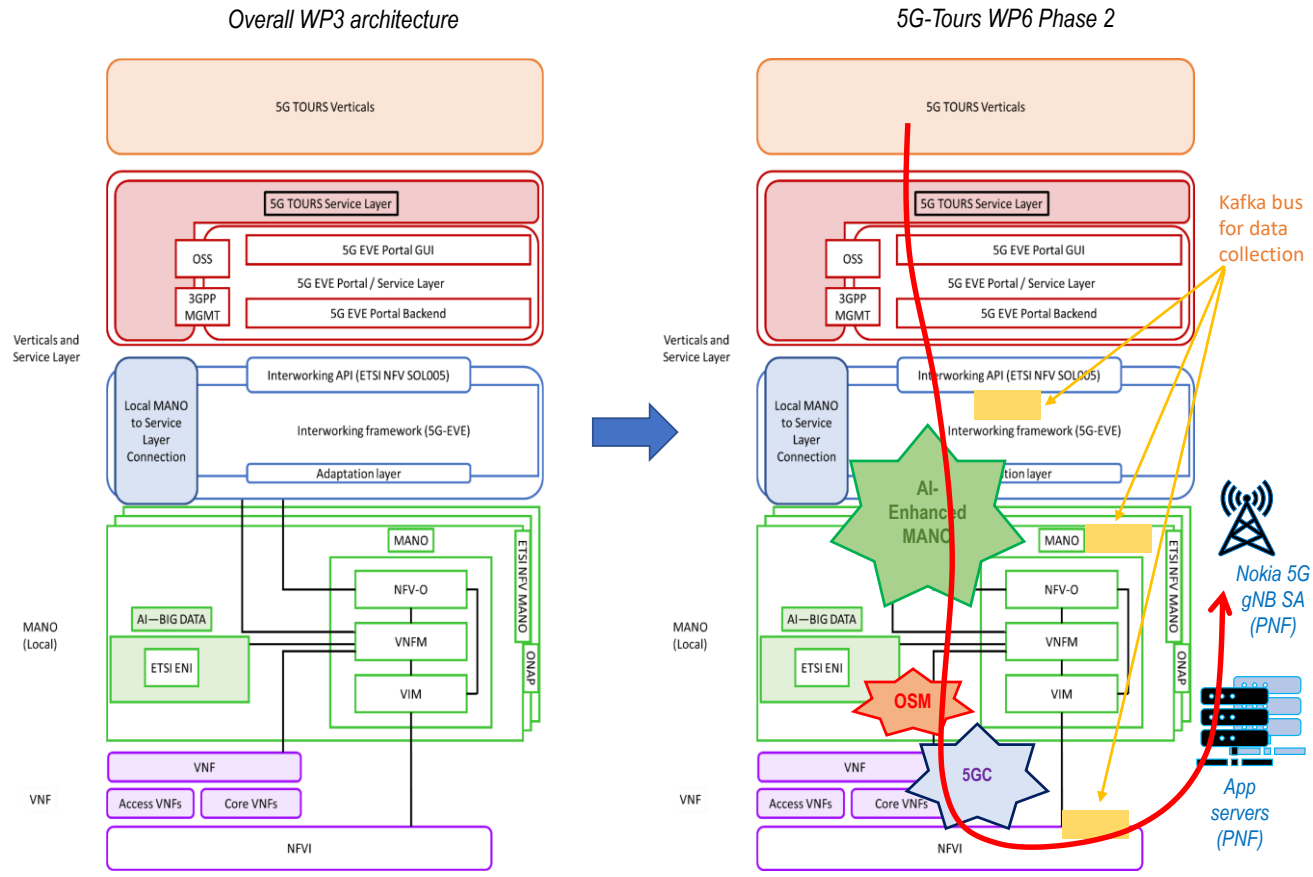


Figure 5. 5G-TOURS functional architecture instantiation for the Mobility-Efficient City site.

2.2.3 Technology integration with 5G EVE

In D3.2 [2], the strategy of onboarding of 5G-Tours vertical use-cases is exposed. It is summarized in the following Figure 6:

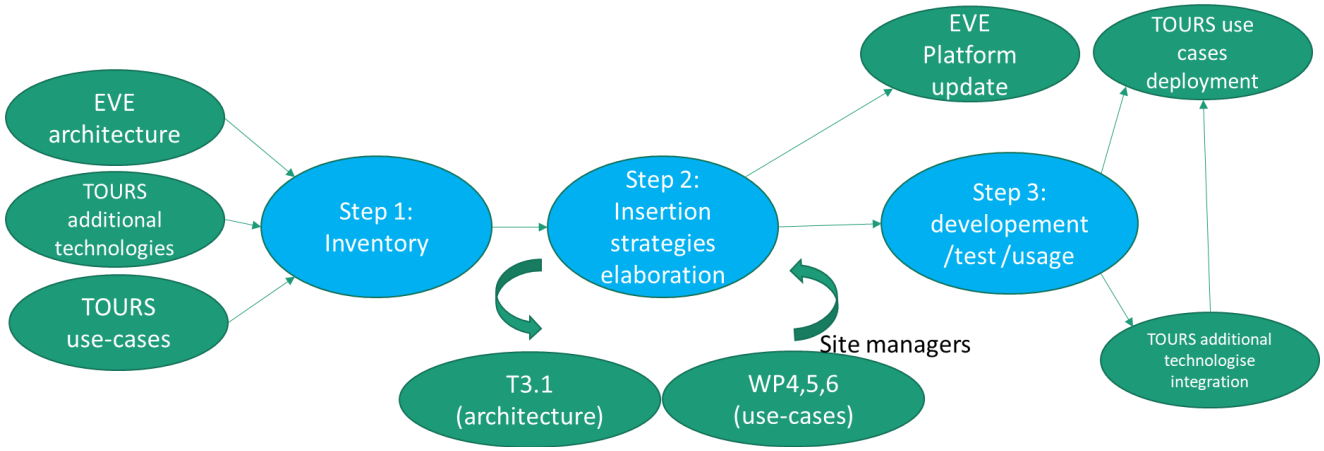


Figure 6. Insertion Methodology.

It is basically a three steps strategy: inventory, insertion strategies and development.

We already identified the possible extension or entry points to integrate 5G-TOURS in the 5G EVE platform. They are depicted in Figure 7 and highlighted by the red pictogram. We can:

- inject new experiment blueprint in the 5G EVE Portal;
- upload new VNF to the catalogue to build the experiment blueprint in the 5G EVE Portal;

- enhance some Locations using a different frequency band;
- extend the radio coverage by adding new locations.

For more clarity, 5G-TOURS has three sites: Rennes, Athens and Turin. Each Site has different locations like airport, hospital, ambulance, different museums etc.

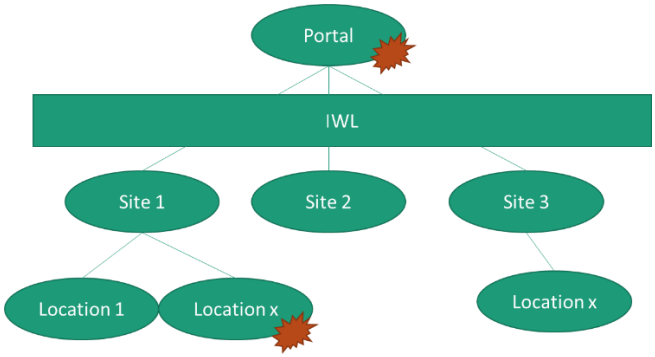


Figure 7. Possible insertion points of 5G-TOURS on 5G EVE platform.

Each step took a different execution time considering the multi-actor and cross-project interactions. At this date, the project finalized the step 3.

At the technology level, 5G-TOURS is aiming to build new 5G networking ideas and technologies around four main areas: enhanced MANO, broadcast, Service Layer for verticals and AI for network orchestration.

Most of the use-cases deployment conducted in 5G-TOURS is VM or container based. For that, the 5G EVE platform already supporting VMs has been enhanced in 5G-Tours via the container support. The OSM and ONAP, both selected orchestrators in Turin and Rennes have been upgraded using the latest release to support Kubernetes and Docker cluster management, while the OSM orchestrator in Athens is realised based on the OpenStack.

Another angle exploited by 5G-TOURS is how we can leverage the monitoring data that is continuously collected by the 5G EVE platform. An AI-based solution has been developed to influence and optimise placement decisions made by the Virtualized Infrastructure Manager (VIM), while ensuring that resources allocation and SLAs are adhered to. Moreover, by using the monitoring information, we can further optimise resource utilisation by:

- enabling higher density for a given set of workloads under the associated SLA;
- anticipating and reacting to changing loads in different slices and assisting the VIM in avoiding resource conflicts, and/or;
- timely triggering of up/down scaling or in/out scaling of associated resources.

2.2.3.1 Turin site

The 5G-TOURS use cases required a coverage extension of 5G EVE in two Locations (Palazzo Madama and GAM). The details on the current and foreseen architecture solution are reported in Section 4.1.1.

The UC implementation has been verified in the field relying on network solutions that provide indoor and outdoor radio coverage, and which are connected to a commercial core network. The architecture instantiation for those verification activities is conducted using the network infrastructure described in Section 4.1.2.

5G-TOURS evolutions will exploit the end-to-end 5G NSA network solution provided by 5G EVE at the TIM laboratory. We use the vanilla 5G EVE regarding the NFV orchestration, we added the capability of using Dockers on top of it. Figure 8 shows 5G-TOURS functionalities for UC1.

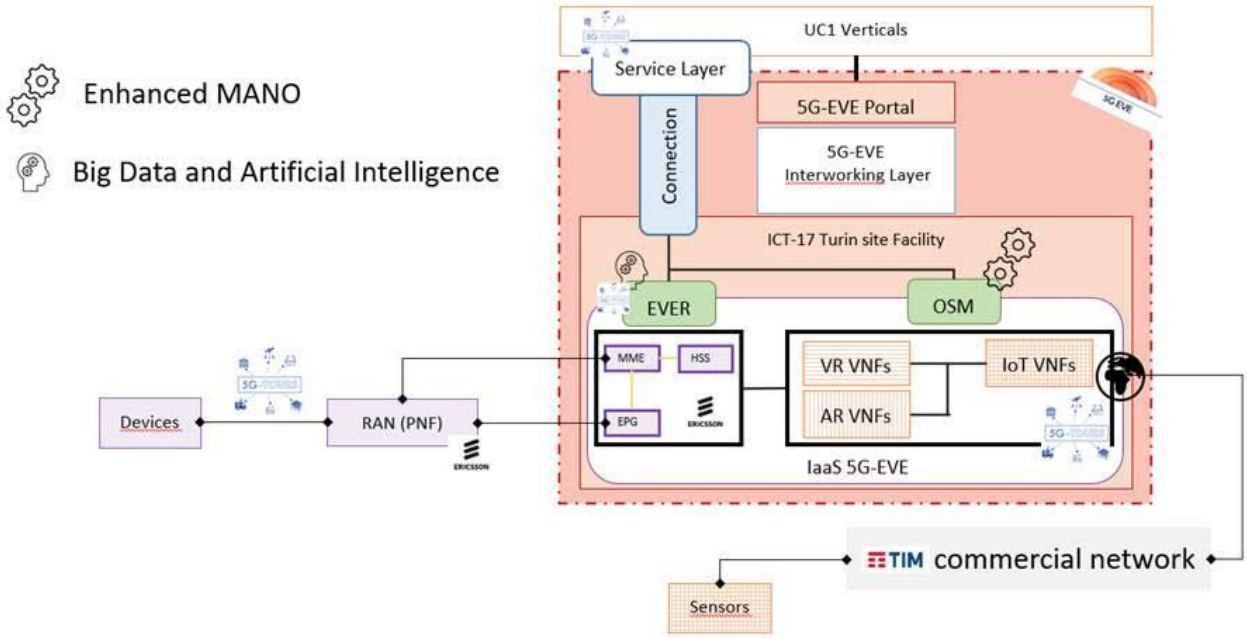


Figure 8. UC1 architecture instantiation.

2.2.3.2 Rennes site

The Rennes site hosts UC 7 and 8 which are part of the Safe City cluster in 5G-TOURS. The consortium worked on the integration of UC8 in 5G EVE covering two potential insertion points: a new location in CHU Rennes, and usage of 5G EVE blueprints to instantiate UC8 through 5G EVE portal. The integration of 5G-TOURS with 5G EVE is achieved as depicted in Figure 9. As for UC7, due to the existence of specific terminals that cannot be integrated with 5G EVE, the "Connection" option is chosen.

For the Safe City cluster ONAP is the primary "insertion technology" for new domains and services. ONAP (Open Network Automation Platform) is an open-source solution that gives the comprehension platform for real-time, policy-driven service-orchestration. Currently new site: BCOM Kubernetes (denoted as K8s) cluster in Rennes was declared successfully in ONAP in Chatillon. Moreover, new type of service instantiation was performed by ONAP in Chatillon. WEF 2.2 (5G CORE made by BCOM) was correctly deployed as CNF in K8s BCOM cluster in Rennes. A Translation Component (TC) has been managed via using new scripts for CNF onboarding with type macro as well as new scripts for CNF instantiation process. To distinguish the requests from VNF onboarding "on demand = à la carte" and CNF onboarding macro new parameter was added to the request from TC to ONAP: */onboarding/{service_name}?mode=macro*.

This new type of instantiation required new onboarding package of service for VNFs - much more complicated where all models (Heat, Helm, CBA are combined). Those tests were performed from the Translation Component (NS-Instantiation server) - it is a part of 5G EVE platform deployed on VM in Chatillon as application with microservices built on docker container. It was developed by ORA-PL in 5G EVE project and now is maintained integrated to 5G-TOURS uses cases.

The architecture of ONAP integration with 5G EVE Interworking layer by Translation component (NS-Instance server) is shown in Figure 9 below.

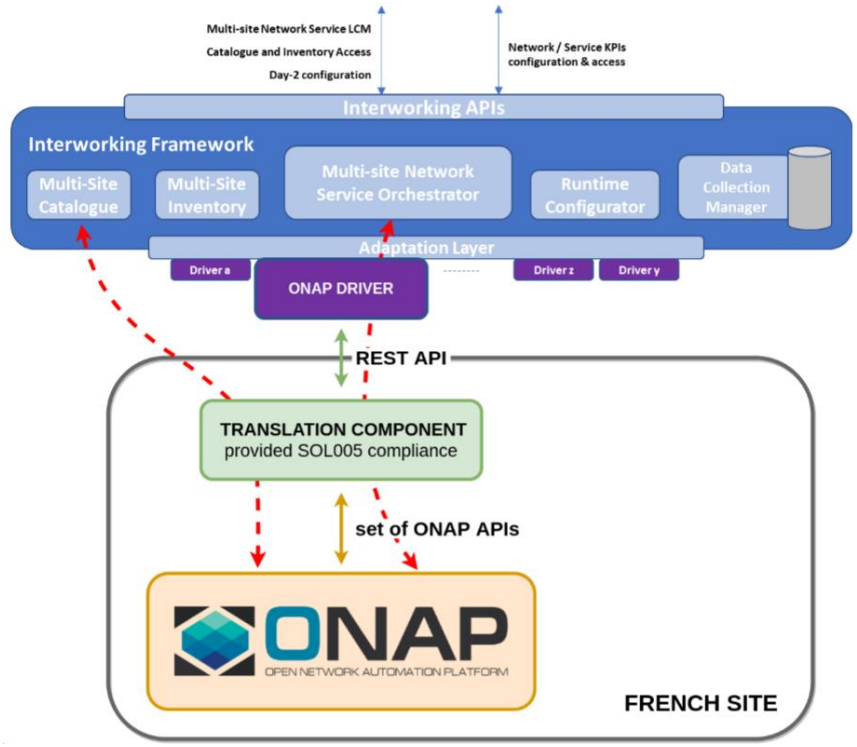


Figure 9. Integration of ONAP to 5G EVE interworking layer in the French Site.

Translation Component (TC) has been implemented in order to integrate French Site with Inter-working layer (IWL) 5G EVE component. Main responsibility of TC is to automate deployment process of network services (NS) and provide ONAP a communication compliance with ETSI NFV SOL005 standard. When some requests come from IWL through the ONAP driver to the French site – it must be proceeded by Translation Component before it gets ONAP. TC exposes REST API as external interface for communication with ONAP driver and IWL. TC triggers proper action using onapsdk [6] python package to interact with ONAP APIs to manage the life cycle of NS instances. TC provided the following basic functionalities required by 5G EVE IWL:

- create, instantiate, terminate and delete network service instances;
- retrieve a list of network services and life-cycle management operation occurrences;
- retrieve information about selected network service or life-cycle management operation occurrence;
- notification about life-cycle management operation occurrences of network services.

TC also selects the target cloud for NS instance therefore each cloud is treated as separate VIM with unique identifier *{vimShortId}*. That id has to be in line with the *{SiteId}* in the IWF Repository.

Recently the new site: k8s BCOM cluster in Rennes for 5G-TOURS project was declared by TC in ONAP. New Site has not been yet declared in IWF Repository and this has to be done to automatically instantiate UC7 from 5G EVE portal. The information about the selected localization for NS instance within the experiment is passed from portal to IWL. Next, the ID of the target sub-site is passed to TC by the ONAP Driver in the body of "create ns instance Id" request. Next, NS instance is deployed in a proper sub-site.

TC integrates ONAP with MS-Catalogue by synchronize services' descriptors managed by ONAP with IWL Catalogue. TC by dedicated API enables MS-Catalogue to access ONAP retrieve selected service specifications as a tosca csar archive. MS-Catalogue implements a dedicated tool to translate ONAP service model into the format required by IWL. It allows to receive and store specific information about services like topology, resource requirements, services management interfaces. Information retrieved in a process of ONAP-IWL catalogue synchronization can be utilized by other components, like portal - to present detailed information about use cases in French Site. According to instantiation of CNF from 5G EVE portal such integration tests should be proceeded with 5-EVE portal and TC and ONAP in 5G-TOURS project.

The next step to integration automatic orchestration with 5G EVE portal is to prepare correctly blueprints with correct service descriptors *{nsdId}*. Correct *nsdId* is taken from ONAP by TC with request: *GET /service_specification/{nsdId}*. This information is propagated to MS-Catalogue and IWL. In 5G EVE project such integration was done for VNFs successfully, for CNFs with macro-onboarding will be performed in 5G-TOURS project.

2.2.3.3 Athens site

The different use cases running in Athens site, are deployed in the AIA airport. It should also be noted that the use cases running in Athens will now also include the WP5 Safe City UC6 "Remote health monitoring and emergency situation notification" and UC9 "Optimal Ambulance routing". Trial of these UCs will be performed in Athens site as restrictions related to the current situation (pandemic) did not allow to perform them in Rennes.

The 5G coverage extension of the Athens site is already deployed, and its architecture is depicted in Figure 10. The 5G EVE is now enhanced by a new location in AIA connected to the 5G-CORE (part of 5G EVE platform) running in OTE Labs. The platform is based on OSM and it is realised using OpenStack (same infrastructure and software version with 5G EVE Athens site). As illustrated in the figure, 4 indoor and 2 outdoor pair antennas (3.5-3.6GHz) are connected to 2 BBUs (BBU1 and BBU2) inside different buildings (B2 and B11). The 2 BBUs are connected directly to a switch at AIA. Also, Small form-Factor Pluggable (SFP) probes are connected into the same switch for the need of real time measurements. A Streaming Server is connected also for the need of UC11 for transmitting emergency 4K video.

The OSN switch at AIA is connected through a 10Gbps line to another OSN switch at OTE Labs, where the 5G EVE infrastructure is developed. The network path for each UC is the following:

- UC6, UC9 (from WP5) (Figure 11) and UC10 work with outdoor antenna in B11 building, BBU2, OSN at AIA, OSN at OTE, 5G EVE infrastructure.
- UC11 works with outdoor antenna in B2 building, BBU1, OSN at AIA, OSN at OTE, 5G EVE infrastructure.
- UC12 works with 3 indoor antennas in Satellite terminal, BBU2, OSN at AIA, OSN at OTE, 5G EVE infrastructure.
- UC13 scenario a (AR) works with indoor antenna in B1 building, BBU1, OSN at AIA, OSN at OTE, 5G EVE infrastructure.
- UC13 scenario b (VR) works with outdoor antenna in B11 building, BBU2, OSN at AIA, OSN at OTE, 5G EVE infrastructure.

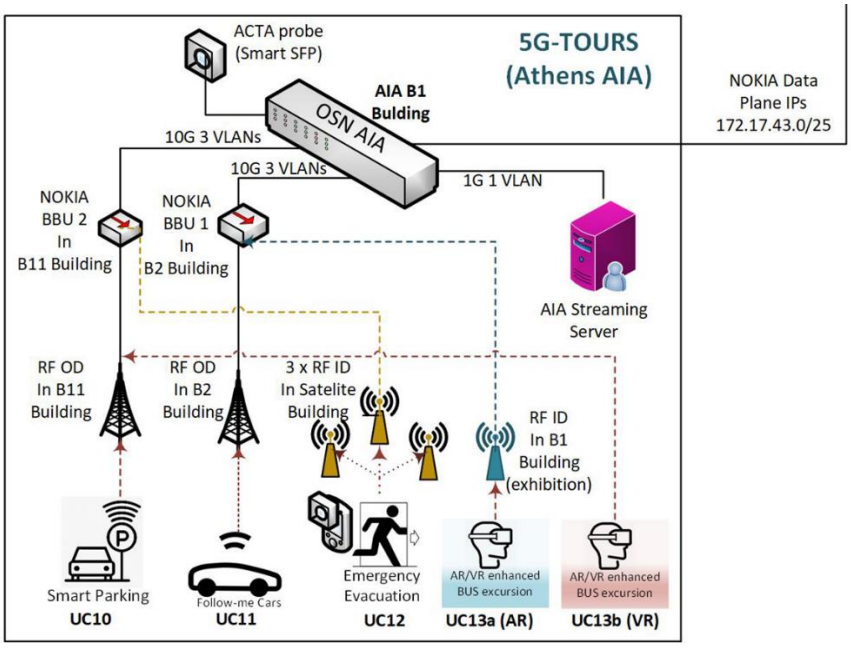


Figure 10. AIA extension location of Athens site for use cases 10, 11, 12 and 13.

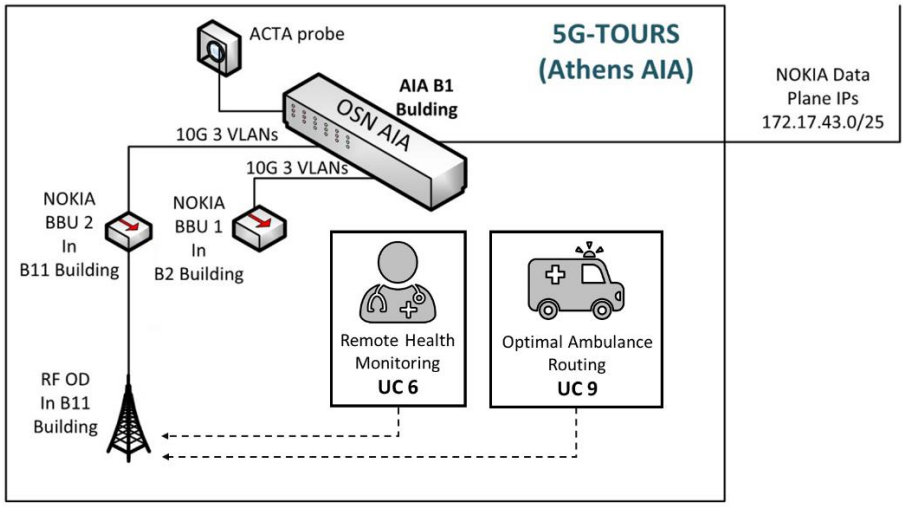


Figure 11. AIA extension location of Athens site for use cases 6 and 9.

In Figure 12, the interconnection for each one of the four UCs of the Athens site, are depicted in correlation with the NOKIA's 5G CORE assets of 5G EVE infrastructure. Similarly, in Figure 13 the interconnection of UC6 and UC9 with the Greek site is depicted. For the needs of Athens site extension an installation of 2 Virtual Machines (VMs) at OTE Labs, one for serving the back-end content needs of ATOS/Samsung AR/VR applications and one for serving the AR and Smart Parking/Evacuation application of WINGS. Also, a Streaming server has been installed at AIA for the needs of UC11. All the UCs use the same network path for the applications. App → 5G antenna → BBU → OTE Edge Network at AIA → Backhaul → 5G EVE EPC at OTE Labs. For the needs of UC10 the commercial 5G public network of OTE is used for sending parking spots status info to the WINGSPARK cloud at OTE Labs infrastructure.

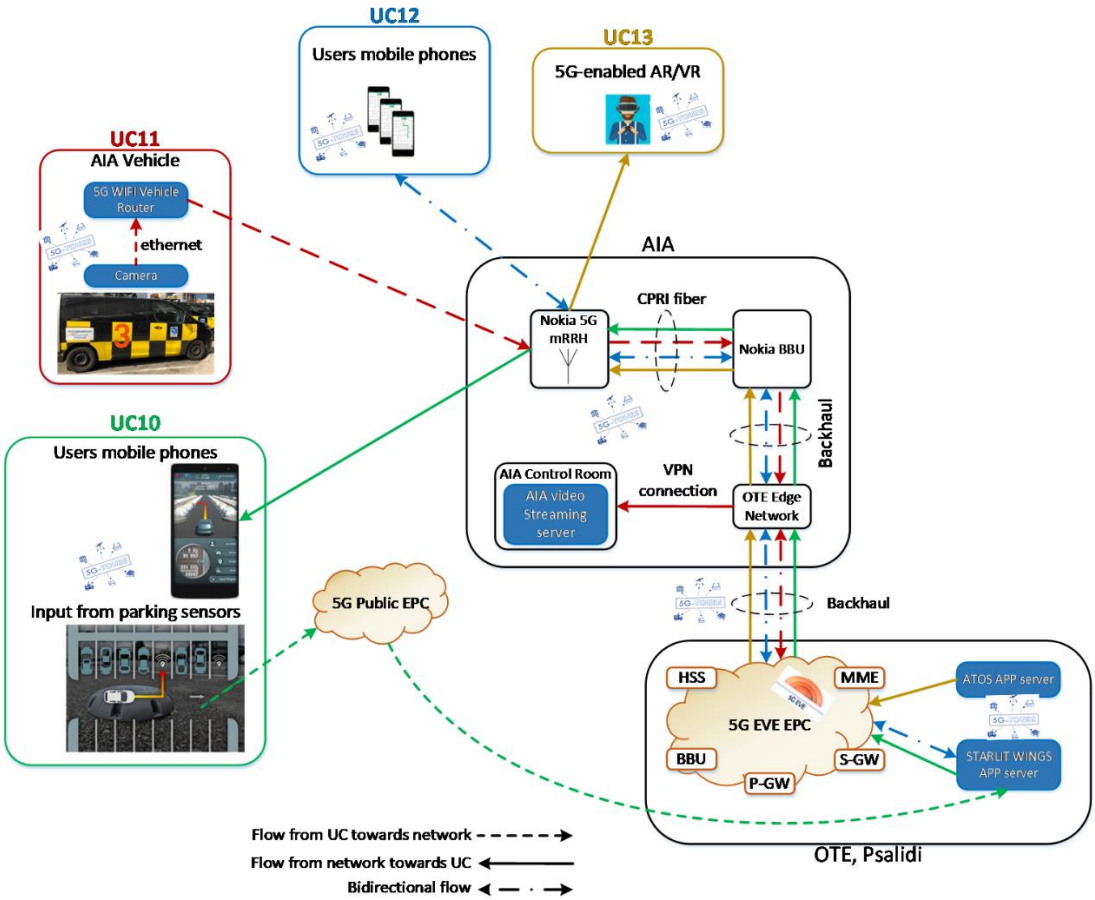


Figure 12. Use cases 10, 11, 12 and 13 integration in 5G EVE Greek Site.

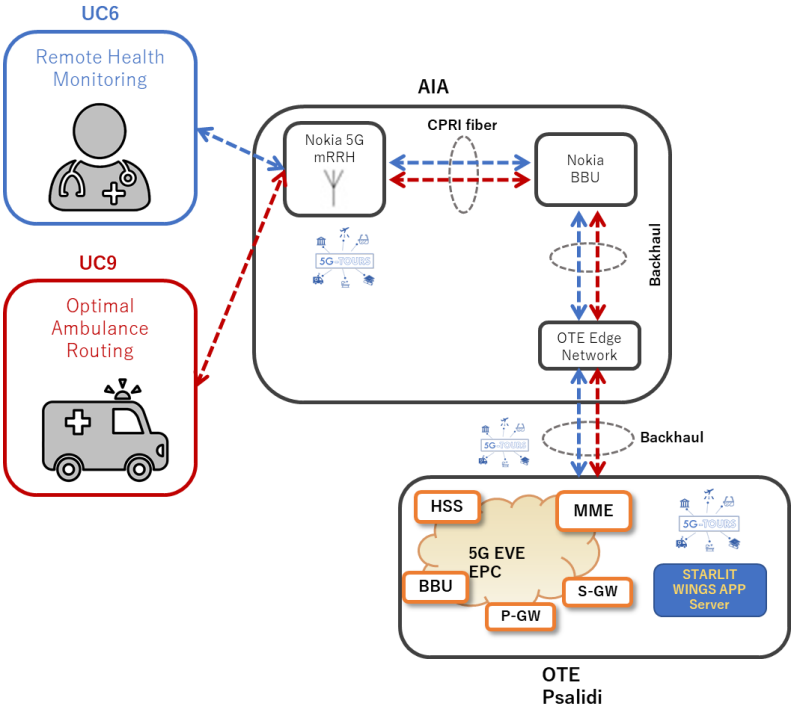


Figure 13. Use cases 6 and 9 integration in 5G EVE Greek Site.

2.2.4 5G security-by-design for verticals

Security aspects are of high importance for 5G deployments. 5G-TOURS user requirement Table 6 in Deliverable D2.3 [4] indicates that almost half of the use cases (6 out of 13) require high or ultra-high network security levels, while two specifically involve physical security operations center at the airport (UC12) or security officers at the museum (UC2c).

In this section, a so-called "security-by-design" approach to 5G security is explained, security improvements from 4G to 5G are briefly presented, and a threat analysis of Mobile Edge Computing (MEC) vulnerabilities is given, since several 5G-TOURS use cases intend to use MEC facilities to some extent.

2.2.4.1 "Security-by-design" approach

Wireless communication is inherently vulnerable and needs specific protection against interception and tampering. Consequently, ever since GSM, the second generation of mobile networks, encryption has been used on the radio interface to secure the user communication. In the following two generations of mobile networks, UMTS and LTE, respectively, the security architecture was significantly enhanced. Besides encryption of user traffic, these networks have also provided mutual authentication between mobile terminals and the network, as well as integrity protection and encryption for all control and management traffic. Overall, UMTS and LTE security features ensure not only high level of security and privacy for subscribers, but very importantly, also assure the resilience required to combat various forms of attacks against the integrity and availability of the services these networks provide. This raises the question: Are new security concepts required for the next mobile network generation? The answer is yes. On the one hand, the support of a variety of new use cases and, on the other, the adoption of new networking paradigms has made it necessary to reconsider some current elements in the approach to security. Figure 14 visualizes the main drivers for 5G security.

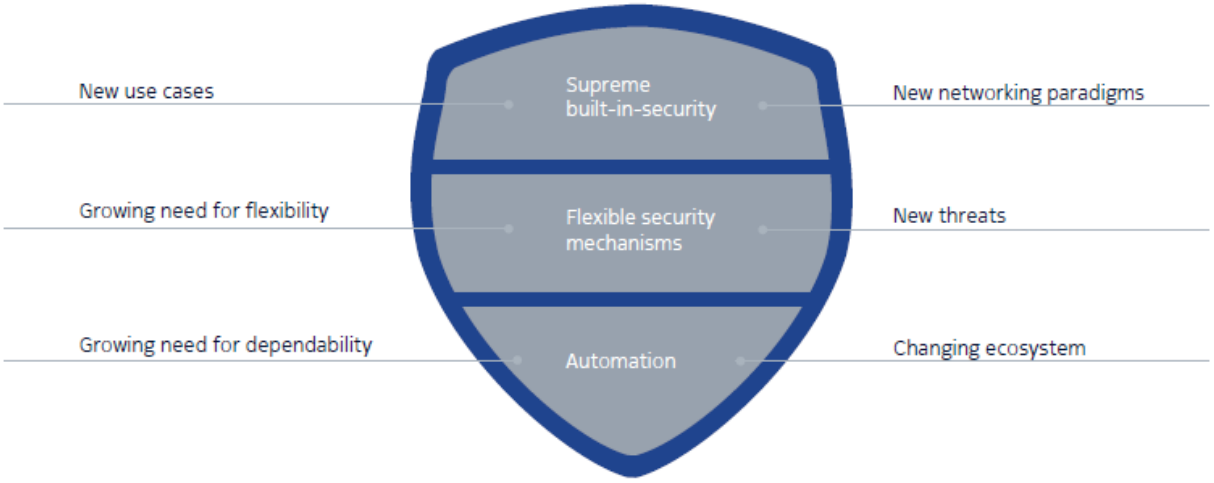


Figure 14. 5G security drivers.

While 4G was designed primarily to support the mobile broadband use case (i.e., broadband access to the Internet), 5G targets a variety of additional use cases with a variety of specific requirements. These cases include support of an enormous density of mobile devices or the need for ultra-low latency in the user communication. Use cases, such as vehicular traffic control or industry control, place the highest demands on the dependability of the network. Indeed, human safety and even human lives depend on the availability and integrity of the network service.

To support each use case in an optimal way, security concepts will also need to be more flexible. For example, security mechanisms used for ultra-low latency, mission-critical applications may not be suitable in massive Internet of Things (IoT) deployments where mobile devices are inexpensive sensors that have a very limited energy budget and transmit data only occasionally.

To efficiently support the new levels of performance and flexibility required for 5G networks, it is understood that new networking paradigms must be adopted, such as NFV and SDN. At the same time, though, these new techniques also bring new threats. For example, when applying NFV, the integrity of virtual network functions (VNFs) and the confidentiality of their data may depend to a larger degree on the isolation properties of a hypervisor. More generally, they will also depend on the whole cloud software stack. Vulnerabilities in such software components have surfaced in the past quite often. In fact, it remains a major challenge to provide a fully dependable, secure NFV environment. SDN, for its part, bears the threat that control applications may wreak havoc on a large scale by erroneously or maliciously interacting with a central network controller.

Another driver for 5G security is the changing ecosystem. LTE networks are dominated by large monolithic deployments—each controlled by a single network operator that owns the network infrastructure while also providing all network services. By contrast, 5G networks may see a number of specialized stakeholders providing end-user 5G network services, as illustrated in Figure 15.

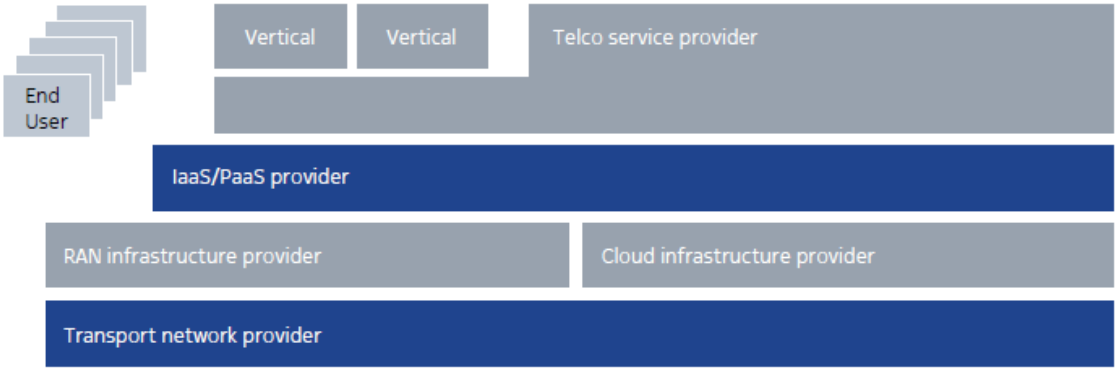


Figure 15. Example of multiple stakeholders involved in providing end-user 5G.

In particular, there may be dedicated infrastructure providers decoupled from telco service providers that host several service providers as tenants on a shared infrastructure. Cloud as the infrastructural choice on its own brings a new set of important 5G security considerations and dilemmas to be solved, such as whether to build/utilize a private cloud (IaaS/PaaS) infrastructure or to make use of external could service providers, how to ensure secure communication in cloud, how to leverage cloud high availability and resilience etc.

In another case, telco service providers may offer not only end-user communication services, but also provide complete virtual networks or "network slices" specialized for specific applications, such as IoT applications. These may be operated by verticals. For example, a manufacturing company could run a virtual mobile network specialized for industry control applications for its own plants. The relevant security issue here is the building and maintenance of new trust relationships among all stakeholders. The aim would be to ensure a trusted and trouble-free interaction resulting in secure end-user services.

Obviously 5G networks must support a very high level of security and privacy for their users (not restricted to humans) and their traffic. At the same time, networks must be highly resistant to all kinds of cyber-attacks. To address this two-fold challenge, security cannot be regarded as an add-on only; instead, security must be considered as part of the overall architecture and **built into the architecture right from the start ("security-by-design")**. Based on a secure architecture, secure network function implementations are also essential in order to ensure a high security network. Security assurance methods are therefore essential so that operators can ensure the required security level for different network functions.

5G security must be flexible. Instead of a one-size-fits-all approach, the security setup must optimally support each application. This entails the use of individual virtual networks or network slices for individual applications, as well as the adjustment of the security configuration per network slice. Security features subject to this flexibility may comprise the mechanisms for identifying and authenticating mobile devices and/or their subscriptions, or for determining the way that user traffic is protected. For example, some applications may rely on security mechanisms offered by the network. These applications may require not only encryption, as in LTE, but also user plane integrity protection. However, other applications may use end-to-end security on the application layer. They may opt out of network-terminated, user-plane security because it does not provide additional security.

2.2.4.2 Security improvements from 4G to 5G

Implementing the security architecture for mobile network functions as standardized by 3GPP is an essential pillar for building highly secure and reliable 5G networks. 3GPP TS 33.501 is the key document providing a detailed description of 'security architecture and procedures for 5G system' [7]. The specification defines a model of a security architecture, consisting of six security domains, as depicted in Figure 16:

- **Network access security (I)** – security features that enable a user terminal to authenticate and access the network by providing protection on the radio interfaces.
- **Network domain security (II)** - security features that enable network nodes to exchange signalling and user data securely.
- **User domain security (III)** - security features that enable the secure user access to mobile devices.
- **Application domain security (IV)** - security features that enable user and provider domain applications to exchange messages securely. 33.501 specifications do not cover application domain security.
- **Service Based Architecture (SBA) domain security (V)** - a new set of security features that enable network functions of the SBA to communicate securely within serving and other network domains.
- **Visibility and configurability of security (VI)** - security features that enable the user to be informed regarding which security features are in operation or not.

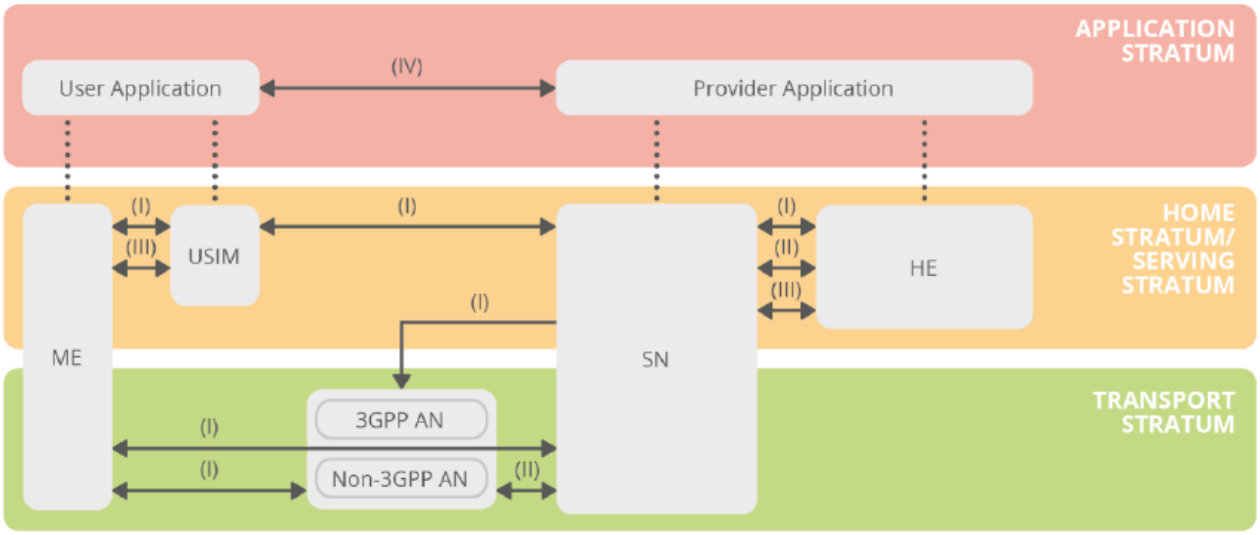


Figure 16. Security architecture model as defined in TS 33.501
(acronyms used are: ME=Mobile Equipment, SE=Serving Network, HE=Home environment)

5G builds on the well-proven security concepts of 4G, such as maintaining separated security associations for the access stratum and the non-access stratum, using a choice of well scrutinized crypto algorithms, using temporary identifiers to protect user location privacy and so on. However, 5G also introduces some significant enhancements and improvements.

At a glance, the new 5G security features are [13]:

- New access-agnostic authentication framework with improved home network control for roaming
- Enhanced subscription privacy (protection against "IMSI-catching")
- User plane integrity protection
- EAP-based "secondary authentication"
- Security for service-based interfaces
- Enhancements for interconnection security

Table 3 provides an extended overview over the new 5G mechanisms and a comparison with the respective 4G mechanisms [13]. At the same time, however, some of these security controls are defined as optional or there is a degree of flexibility left to suppliers on how to implement and for operators on how to interpret and utilise the controls.

Table 3. 3GPP Security Features – 5G versus 4G.

4G (LTE) Security	5G Security
UE is authenticated by two different authentication methods depending on the access network type (EPS AKA on LTE access and EAP AKA'on Wi-Fi access).	**UE is authenticated using either 5G AKA or EAP AKA'**, irrespective of access type (access agnostic). The home network decides on the method.
A serving network can fraudulently request authentication info for a UE from the UE's home network, even if the UE is not at all roaming in the serving network. This info can then be abused.	**Improved home network control**: When a serving network authenticates a roaming UE, the home network gets a proof that the UE is indeed present in the serving network.
The UE subscription identifier (IMSI) is in some cases transmitted as plain text without encryption over the air; fake base stations can force UEs to reveal the IMSI ("IMSI-catching").	The Permanent Subscription Identifier (SUPI) is not sent in clear over the air in any network procedures; instead, the **Subscription Concealed Identifier (SUCI)** is used, which is an encrypted form of the SUPI.
No integrity protection of user plane traffic; this allows certain attacks although the data are encrypted.	**Integrity protection of user plane traffic** is mandatory to support by the UE and the network (optional to use).
The network provides very restricted support for authentication of UEs to connected packet data networks ("PCO-based authentication").	**EAP-based "secondary authentication"** provides a flexible and strong concept for authentication between UEs and connected data networks.

Packet core network control signaling uses Diameter, and IPsec protection is mostly not applied for it.	The **5G Service Based Architecture (SBA) uses HTTP/2 with IETF Transport Layer Security (TLS)**, and the OAuth 2.0 framework to authorize access to restful APIs.
No flexible security for interconnection of different PLMNs via an IP Exchange (IPX) network. Use of E2E IPsec tunnels is specified, but this is in conflict with the operational requirements of an IPX, so it is mostly not used.	The **new function SEPP (Security Edge Protection Proxy)** protects the edge of the network and provides flexible interconnection security. It allows to protect selectively sensitive information while making other information visible to entities in the interconnection network, as required.

2.2.4.3 MEC threat analysis

MEC technology and its security aspects are important in 5G-TOURS, as the user requirement Table 6 in Deliverable D2.3 [4] indicates that almost half of the use cases (6 out of 13) are supposed to employ either Edge computing or/and Edge storage. This section proposes a summary of a survey of the latest MEC threats and their possible countermeasures.

Mobile Edge Computing (a.k.a. Multi-access-Edge Computing in the ETSI MEC Industry Specification Group) [8] complements the 5G architecture and allows applications to be executed close to the Radio Access Network (RAN) and in proximity of the User Equipment (UE). This is critical for applications requiring ultra-low latency. While latency to a core network (executed in a regional data center) is about 20ms to 40ms, latency for an edge node is less than 10ms (and even down to 200µs for a far edge). Moreover, as UEs such as IoT have very limited computation and processing capacities, the execution of many tasks needs to be performed in an infrastructure having much powerful resources. For this purpose, it is required to locally host both the data and the compute-intensive processing. MEC is also applicable for numerous and various markets (e.g., eHealth, smart-cities, industry 4.0, transport, connected & autonomous cars) and it allows tremendous business opportunities [9], [10].

Security is of course a key challenge regarding MEC network. 5G network security has been widely investigated [11]-[14] and is of course applicable to MEC. Security of public MEC (e.g., MEC belonging to a telco) is expected to be more secure as it benefits from the operator security infrastructure. However, private MEC (e.g., for a hospital, an airport) will be deployed within the enterprise infrastructure which are usually not full secure and can therefore be more vulnerable to new attacks based on the MEC.

The following Figure 17 depicts a simplified view of the architecture as defined in the ETSI-MEC framework [15]. MEC architecture has two main levels: the Mobile Edge Host Level and the Mobile Edge System Level. The Mobile Edge Host Level mainly contains the local Apps executed upon a virtualized infrastructure; the Mobile Edge System Level contains the orchestration and life-cycle functions handling the apps over the different MEC hosts, as well as some other OSS and customer facing functions.

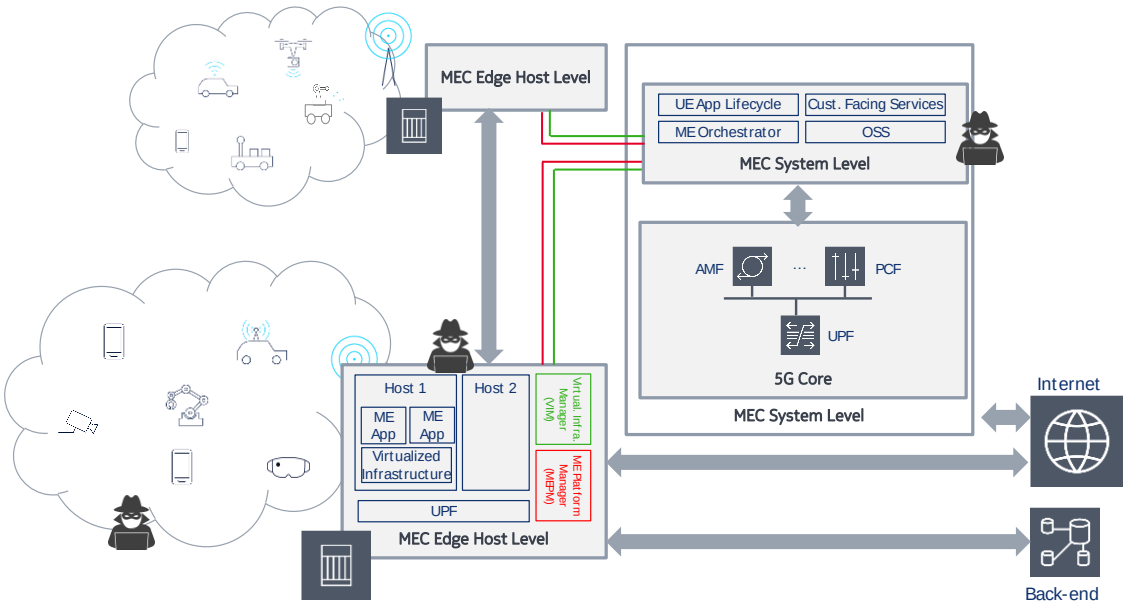


Figure 17. Network scope for vulnerability analysis.

Figure 17 also presents the areas of the main vulnerabilities [16]:

1. Vulnerabilities of the local radio MEC network.
2. Vulnerabilities in the MEC host Edge cloud (including between MEC Hosts).
3. Vulnerabilities in the MEC System Level.

1. Vulnerabilities of the local radio MEC network

It is well known that numerous UEs are highly vulnerable (e.g., IoT such as sensors, robots, AR/VR glasses) and are subject to various attacks (e.g., physical tampering, malicious code injection, hardware trojans). A compromised UE can threaten the MEC system in various ways. An attacker can:

- Convey malicious content from a tampered UE to the MEC host [17]-[22].
- Deplete the resources of the UE, disrupting the normal execution the Applications/Services running in the MEC [19].
- Manipulate the data volume sent to the MEC applications ("offloading tampering") to allocate more resources and degrade the performances of the application or even the MEC itself.
- Use the potential huge numbers of IoT to perform critical large DDoS attacks on the applications.
- Eavesdrop the offloading channels from UEs may not be secure since computation tasks can be overheard by malicious eavesdroppers. Traffic encryption between the UE (that may have limited resources) and the MEC can increase the propagation delay as well as the execution delay, thus reducing the application performance in a level not acceptable for applications.

2. Vulnerabilities in the MEC host Edge cloud (including between MEC Hosts)

Beyond the MEC applications providing by the enterprise holding the MEC, various applications from third parties or telco operators can also run on the MEC platform. This is particularly the case for verticals as they have often their independent ecosystem and developer community.

MEC network and platform API can then be a source of attacks notably unauthorized access and privilege escalation, sensitive data leakage, malicious use of the MEC NFV functions [21]. For some vertical market (e.g., entertainment, smart cities) applications can be downloaded by consumer end-user from a marketplace. While these applications will run in the UE, it can be necessary (e.g., for time-sensitive and/or high bandwidth services) that the server part of the application will run locally in the MEC. This can be an attack vector to the MEC, notably if the user downloads the application from an official marketplace.

Large scale MEC deployment (e.g., smart cities) can be severely affected by the tampering of just one MEC. The possible heterogeneity and diversity of MEC environment also increase vulnerabilities. The different MEC applications may not have the same image version, or all updated with all the required patches. An attacker can compromise an unpatched MEC to later compromise the other MEC having the same vulnerability.

For some use-cases, the device density in MEC network could be up to 1 million devices for every square kilometre. Managing device credentials and detect their possible compromising (for credential revocation) will request scaling capabilities that have not previously been uncounted.

New attack vectors can also compromise privacy (e.g., location) and could constitute a major threat for privacy sensitive application and regarding regulation policy [22], [23].

For some use-cases (e.g., car traffic control and autonomous driving), MECs can be located along a path (e.g., highway) and an application can migrate in real-time from one MEC to another. If compromised, the application can tamper the other crossed MECs. Besides, crossing borders can induce data regulation policy violation [24].

3. Vulnerabilities in the MEC System Level

Attacks can also be performed on the link between the MEC and the core (backhaul link) [25]. MEC hosts support a much smaller traffic compared to the centralized core and are then more exposed to DoS attacks.

Moreover, the backhaul link transports critical data from the core or other servers accessed through internet. This is for instance the case of sensitive data exchanged between a MEC and a central office (e.g., company headquarter) which are exposed to attacks as eavesdropping. Communication between MECs (e.g., for automotive) also extend the possible attacks.

This link is also critical for the execution of applications as it allows migrating VM, offloading tasks that required high processing resource, transporting of critical network control information (MEC network control), transporting statistics and service logs information (from the MEC network and applications).

3 5G-TOURS NETWORK INNOVATIONS

3.1 INTRODUCTION

This section details the network and system related innovations devised within 5G-TOURS WP3. They usefully complement applications-related innovations showcased in WPs 4,5,6 and 8. Among these network innovations, AI-based orchestration and Enhanced-MANO solutions are presented as they are cornerstone for modern virtualized deployments of 5G network functions and allows for more efficient core and radio resource usage / optimization. Then, specific capabilities for 5G broadcast support were developed for point-to-multipoint media transmissions and deployed in two UCs. Finally, solutions for 5G-TOURS Service layer facing the verticals and responding to their specific requirements are described, and, for each of them, open-source SDKs (APIs and/or code) are presented.

3.2 ENHANCED MANO

3.2.1 AI-Agent functionality

3.2.1.1 **Overview**

OSM is delivering an open-source Management and Orchestration (MANO) stack aligned with ETSI NFV Information Models [27]. As a community-led community, OSM offers a production-quality MANO stack that meets operators' requirements for commercial NFV deployments.

Regarding this, the overall project objective at the beginning of the project was to contribute to OSM with ideas and concepts discussed as part of the 5G-TOURS Project and, if possible, even with code. In practice, the envisaged potential contribution was the development of the AI-Agents service component itself (code included).

In the context of network services management and orchestration AI-Agents are software entities able to trigger orchestration actions based on gathering relevant network metrics (from the network itself and/or the network services running on it) and processing them by means of AI/ML algorithms. Those agents can be executed individually (e.g., associated to a specific service or O&M procedure) or distributed through the network.

Aligned to this, during the 5G-TOURS project lifetime a significant activity has been developed for implementing the "AI Agents for OSM" functionality. The development of the concept has been shared and discussed within the ETSI OSM Community (mainly within the Service Assurance team) during their regular meetings, providing several presentations and interchanging communications to share the design ideas and the main concepts related to this functionality.

A central part of the discussion was the alignment of the functionality with the OSM functional architecture. During the discussions it was initially evaluated to modify some of the OSM architectural modules (mainly MON and POL); however, the final approach has been to deploy the AI Agents as embedded elements in the VNFs using their Execution Environment, and the usage of the so-called SOL005 interface [28] through the OSM North-Bound Interface (NBI) to perform the AI-based Service Assurance operations (VNFs scaling and alerting). Based on this approach a practical implementation was developed, targeting the objective of contributing to OSM with ideas and concepts discussed in 5G-TOURS, even with code.

The result of the implementation has been made available to the Open-Source community through the "readthedocs.io" public repository (URL: https://ai-agents-for-osm.readthedocs.io/en/latest), where the technical documentation and the source code of the solution can be accessed under the Apache 2 open-source license.

Also, besides the AI-Agents contribution itself, the 5G-TOURS team also contributed to the OSM Community on its regular activities, attending and participating in their regular meetings and contributing to certain bug fixing tasks on the OSM platform, which also facilitated the implementation of the AI Agents functionality itself.

3.2.1.2 Implementation

The implemented "AI-Agents for OSM" functionality is intended for automating the deployment of Artificial Intelligence Agents on the VNFs orchestrated from the OSM platform[1]. The aim is to enable the Virtual Network Functions (VNFs) deployed in OSM with artificial intelligence capabilities. The following figure illustrates the general concept:

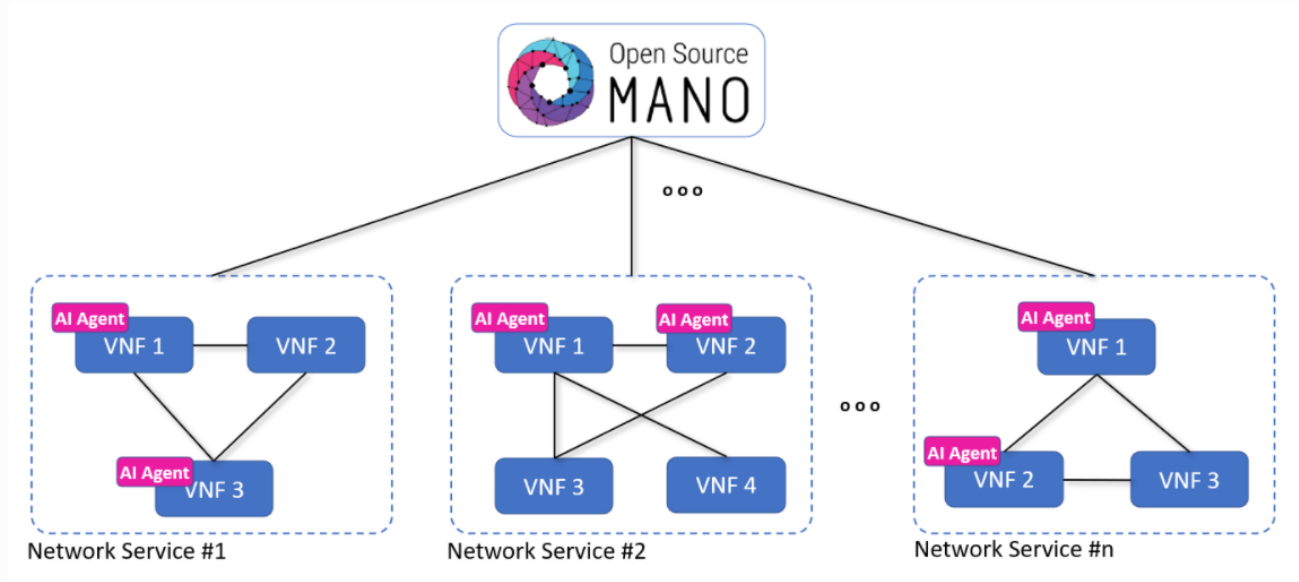


Figure 18. AI-Agents deployed on different VNFs in different Network Services managed from OSM.

As Figure 18 shows, individual AI-Agents can be attached to specific VNFs associated to different Network Services to perform AI-related tasks. Of course, attaching AI-Agents to VNFs is optional, so there can be VNFs with or without AI-Agent attached.

VNFs with AI-Agents attached will have additional AI/ML capabilities associated to the Service Assurance (SA) set of actions already available in OSM, namely:

- Alerting actions (VNF metrics can trigger threshold-violation alarms), and
- Auto-scaling actions (automatically increase or reduce the number of VNF instances).

Without relying on the new "AI-Agents for OSM" functionality these actions are regularly executed based on the definition of simple rules or thresholds in OSM, associated to certain by-default metrics (e.g., CPU, RAM or network usage), but using AI-Agents for OSM this behavior can be enriched, enabling the possibility of triggering those alerting or scaling actions based also on more complex metrics leveraging on AI/ML algorithms.

In fact, the AI-Agents for OSM functionality can be used to enhance the VNFs deployed on OSM, allowing (for instance) to define proactive alerting or scaling behaviors based on forecasting certain service usage patterns; also, to trigger scaling actions based on image recognition or data clustering algorithms, among other common AI/ML-based applications.

In summary, AI-Agents for OSM allows to go beyond the purely reactive OSM behaviour and the limited set of metrics defined on it, allowing the integration of AI/ML capabilities so expanding the by-default OSM functionality.

[1] The content from this paragraph until the end of this subsection has been taken from the above-mentioned AI-Agents documentation repository: https://ai-agents-for-osm.readthedocs.io/en/latest

The deployment of the AI-Agents for OSM functionality is based on the Execution Environments (EE) feature introduced in OSM Release 8 [29]. At high-level, the architecture of the AI-Agents for OSM solution is divided into two main functional blocks:

- The AI Agents themselves, which can be attached to VNFs to provide them with AI/ML capabilities.
- The so-called AI Models Server, where the AI Models (AI algorithms) used by the AI Agents are hosted (see Figure 19 below).

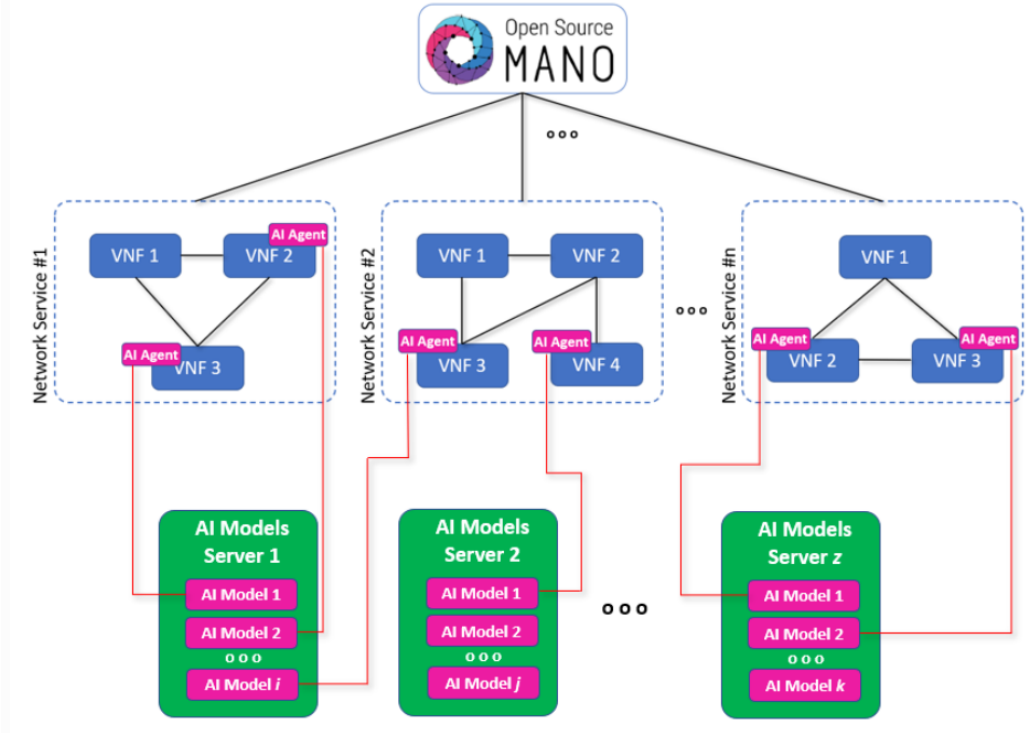


Figure 19. AI-Agents connected to AI Model Servers.

As shown, different AI Models Server instances can host different AI models. AI-Agents communicate with these models to perform their AI-related functions. Since some AI Models may have highly demanding computational requirements, It was considered better to have them hosted on an external server (the AI Models Servers) in order to avoid overloading the primary function of the VNFs to which they are associated.

According to this, the AI-Agent component can be seen as a proxy between the VNF and the AI Models. I.e., while "the intelligence" is actually in the AI Models Server, the AI-Agent works as "the interface" between the VNF and the AI Models hosted on the AI Models Server.

The AI Models Server has been envisioned as an external component to OSM. So, similarly to the VIM, it has to be installed besides OSM to provide the required AI/ML functionalities. However, the AI-Agents for OSM solution does not require the usage of any specific software to implement the AI Models Server functional block. The only condition is that the AI models hosted on the AI Models Server are accessible through a REST API.

This allows an agnostic implementation: the AI Models Server can be based on common Open Source solutions in the field of Artificial Intelligence (TensorFlow Serving[30], PyTorch [31] or others), or also on proprietary solutions that the AI Agents user may already own.

However, although the AI Agents intelligence is actually hosted in the AI Models Server, the role of the AI-Agents is not minor: they collect and normalize the data for training the AI models (during the training stages) and pass them to the AI Models Server in the proper format (during the production stages). They also interface with OSM in two ways: First, OSM prepares the Execution Environment with the required runtime variables to deploy the AI Agents. And second, the AI Agents request the necessary Service Assurance actions (VNFs scaling or alerting) when necessary.

As mentioned, this "AI-Agents for OSM" solution assumes that there will be an "AI Model" (or a set of them) deployed on the AI Models Server. So, we could imagine the AI Models Server as a regular database hosting a

set of data models to support the AI Agent actions. The difference here is that these data models are special models based on artificial intelligence algorithms.

Unlike data models in conventional databases, these AI Models usually need to be trained. In short, training consists on feeding the models with example data from which they are able to learn and generalize in an algorithmic way. As we know there are different machine learning paradigms in the field of Artificial Intelligence, mainly supervised, unsupervised and reinforcement learning, but also others derived from these.

The "AI Agents for OSM" solution is initially designed to work with supervised and unsupervised models, or any other model in which training and production stages are clearly separated. Other learning models (such as reinforcement) that rely on a continuous update of the deployed model are not yet considered (although they could be addressed in a next development stage). Anyway, it is considered this is not a major issue for implementing practical applications, since supervised and unsupervised models (and many other derived from them) already offer a quite wide range of practical use cases.

In any case, what the AI-Agent expects to find in the AI Models Server during the production phase is a valid trained AI Model. However, in order to have this model already trained, it is necessary to train it (of course) using data taken from the production environment (or simulated data in the proper format). The following Figure shows this way of working in which the training and production stages are clearly separated.

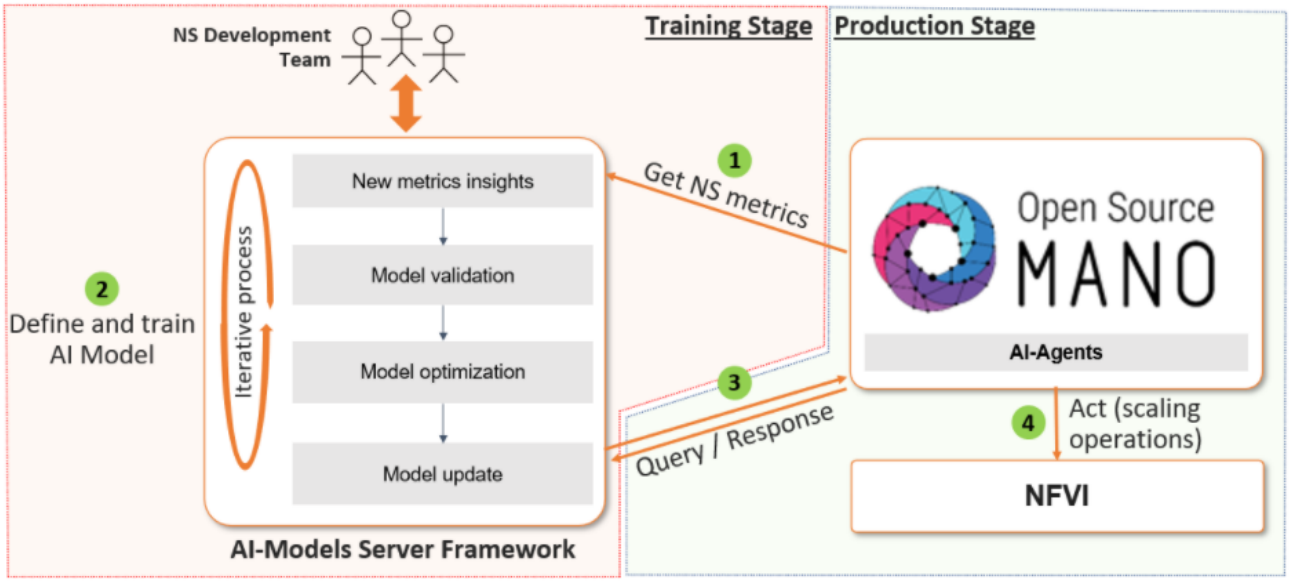


Figure 20. Integrating AI Models.

As shown during the training stage the AI-Agent is used to get the necessary Network Service data to train the models hosted in the AI Models Server. Those data could be collected from the staging or even the production environments (it is important to have access to real training data in order to generate accurate AI models). These data would be stored by the AI Agent in a volume that can be accessed by the AI Models development team during the training stage. Once the model is trained and ready for production it can be queried from the AI-Agents. Based on the responses, the AI Agents would apply (or not) the necessary Service Assurance actions (scaling or alerting).

3.2.1.3 AI-Agents Deployment

AI Agents are modelled as part of the VNF package through a Helm Chart, so they can be instantiated by OSM through its North-Bound interface (NBI - see Figure 21). This approach relies on the dockerized VNF "Execution Environments" (EE) introduced in OSM Release 8, which are modeled through the mentioned Helm Charts implementing a lightweight API server able to configure (or reconfigure) the rest of the elements included in the environment in the form of OSM Day-1 and Day-2 primitives (labels from '1' to '4' in Figure 21). The EE is intended to provide a deployment environment with default configurations such as network connectivity with the OSM core elements as well as with the attached VDU. The EE is also in charge of providing the runtime configuration that AI Agents require to correctly locate the instantiated VNF information within OSM.

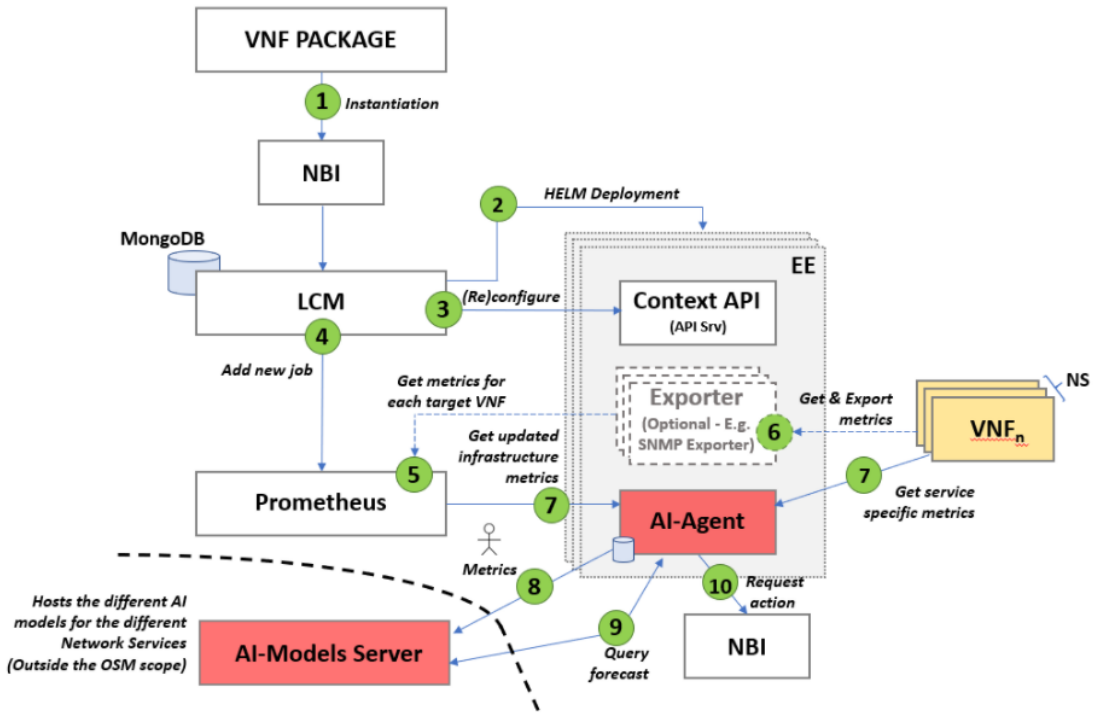


Figure 21. AI-Agents solution deployment.

Once deployed, the AI Agents are able to get metrics from different sources (label '7'), e.g.:

- VNF metrics, that could be accessed in different ways: directly from the VNF (through a specific API) or through a dedicated metrics exporter (e.g., an SNMP exporter - labels '5' and '6');
- Infrastructure metrics from the OSM Prometheus timeseries database (still to be implemented).

Label '8' represents the access to the training data collected by the AI Agent from the production or the staging environments. These data would be stored in a shared volume to be collected by the AI Models development team to be used during the training stage.

Once the necessary AI Models are trained and deployed on the AI Models Server, the AI-Agents can execute Service Assurance actions by querying the AI Models Server using a REST API (label '9') and by requesting the necessary Service Assurance actions to OSM through its NBI (label '10').

3.2.2 Service level assurance in 5G RAN MANO enhancement

As described in [3] ML-enhanced analytics on network and performance may be used to identify and monitor, with the help of proper data visualization tools, patterns, trends, outliers and other remarkable features and to assist closed loop logic in adopting the right decision policy in order to achieve service assurance. To this end, some principles of ML-based analytics for mobile network behavior characterization will be described in this paragraph.

The first step is modeling the system behavior of a mobile access network wrt the following KPI categories:

- traffic characteristics,
- user distribution,
- network resource usage,
- experienced QoS / QoE.

These KPIs can be obtained by the observation of several elementary performance counters derived from the network elements. In the context of a 5G mobile, however, the performance analysis has to be performed breaking down the end-to-end performance into different domains, e.g. service level (application servers), Core Network and RAN functions, virtualization and cloud infrastructure.

Moreover, performance characterization of the above mentioned KPI categories should take into account different dimension, e.g. spatial distribution (e.g. down to cell level or portion of cells) and time periodicity (daily, weekly, monthly, yearly).

The following figure is an example representing the number of connected users in a cell: a weekly periodicity can be highlighted, with different load in working days or weekends, clearly showing two peaks in working days.

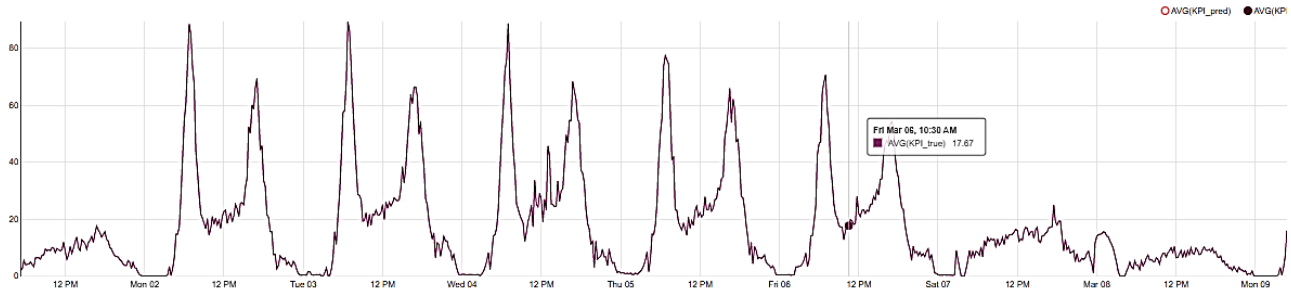


Figure 22. Example plotting the number of connected users in a given cell during a certain period of time.

This kind of trend, however, is not typical of all the cells, as it strongly depends for instance on the cell type, on the observed KPI and on the long-term trend (e.g. months and seasons). In this context, the adoption of Machine Learning techniques applied to network and service KPI may help in characterizing the system behavior.

The following graph of Figure 23 compare the true values of the previously mentioned KPI (in violet), with the values (in red) predicted using an algorithm based on Linear Regression approach.

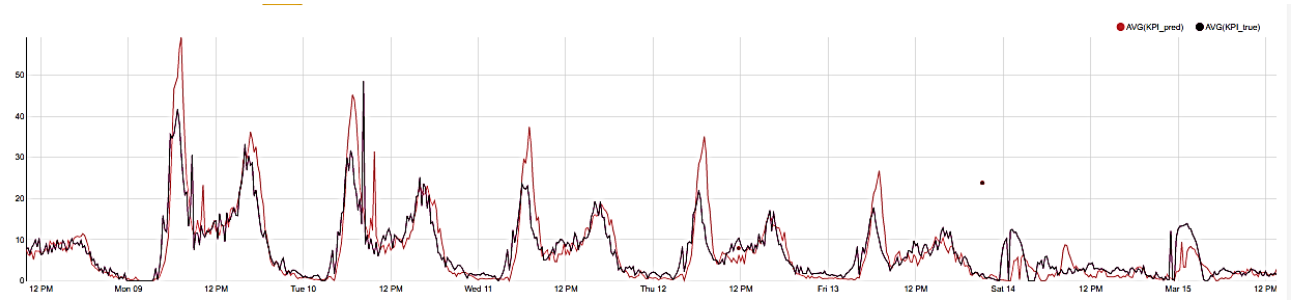


Figure 23. Example of predicted (red) and real (violet) values regarding the number of connected users in a cell.

The possibility to, for example, predict the trend, in particular the traffic peaks, using some margins to take into account an estimation error (over-estimation, in the given example), may be used to adopt algorithms of near real time system optimization.

As described above, behavior at cell level may have different characteristics and trends and it is important to be able to train ML models accordingly. The table in Figure 24 depicts a Heatmap that reports an example of the monthly averages for the Normalized Mean Square Error (NMSE) of the previously described prediction algorithm for 19 different cells[2].

The highlighted row and columns show that even with individual training, the prediction error may vary over different cells /cell types. Moreover, the heatmap highlights cell behaviors over time: in the example given, two cells shows a completely different average performance in the month of August (when typical season fluctuations occur in the considered area due to summer holidays), resulting in higher values of Normalized Mean Square Error (NMSE).

[2] The actual values here are to be considered just as examples to clarify how the modelling and the usage of proper tools like the heatmaps can help to understand and characterize an access network behaviour down to cell level. These values are not related to the project UCs performed in Turin.

Therefore, when using not aggregated per cell data, more subject to unpredictable fluctuations than aggregated average values of areas containing several cells, is of paramount importance to determine correctly the possible errors and the shortcomings of the prediction algorithm concerning certain cells or specific time of year and to apply some countermeasures in the modeling such as specific margins.

As stated above, the ML algorithm is a Linear Regression. Linear Regression is a Machine Learning algorithm based on supervised learning to predict a dependent variable value (in this case the number of connected users) based on a given independent variable (time: period of the day, week…). As the model is an input to an example of closed loop mechanism for service level assurance in a network MANO, it can be applied locally in 5G-TOURS Network architecture.

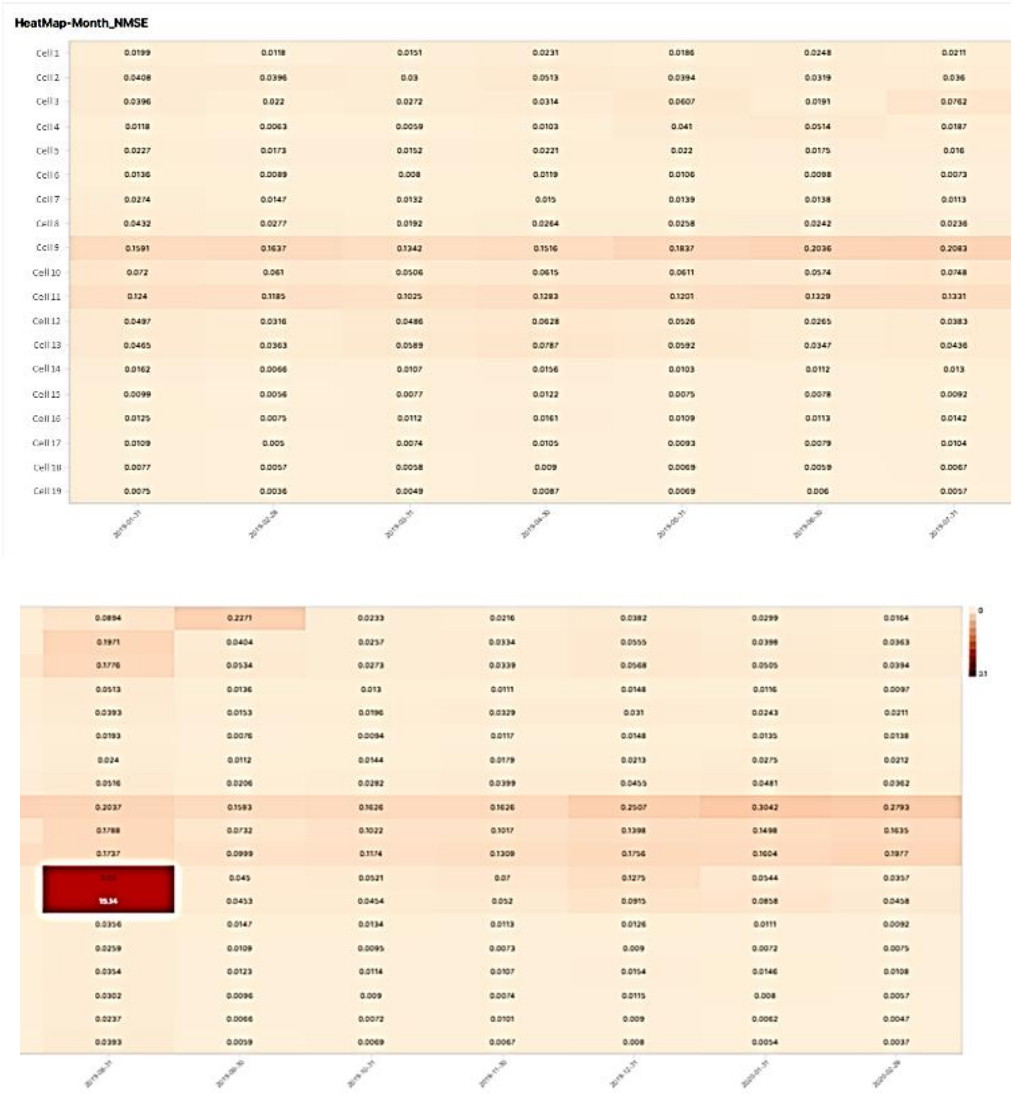


Figure 24. Example of a HeatMap depicting the NMSE monthly average of 19 cells for 14 months.

3.3 AI ORCHESTRATION

3.3.1 Outlook on the integration of AI into the major standard architectures

The integration of AI into the major network architecture has reached now almost all the levels of the network domains, as discussed next. For instance, the standardization work on intelligent elements in the networks shall include root cause analysis or impact analysis. Individual data analytics capabilities can be combined to process incoming events and other information related to faults and performance of the monitored objects, aggregate,

and analyze the information and derive novel information required for enhanced operations, e.g., failure prediction to prevent faulty network states. Policy recommendations exposed through this capability are eventually consumed by other elements in each domain (as discussed in Section 2) or across domains.

Examples of these behavior are countless. Also starting from 3GPP Release 15, the mobile network started to provide analytics services, mostly through the NWDAF in the 5GC, which, e.g., provides information about the load of a NFs which may be used by other NFs to adjust sensible settings for e.g. load balancing purposes. Similar data analytics elements are envisioned by O-RAN through the Radio Network Information Base (RNIB). Those data analytics services are also consumed by management functionality for network slice related decision. Recently, 3GPP Release 18 preparation work has specified ML as one of the work items to be studied in Release 18 both for RAN and CORE.

The management system also produces analytics information through the Management Data Analytics Service (MDAS) (Table 4). This service is consumed by other management entities or forwarded to other domains such as the orchestration or the network function domain. Still, in current standardization efforts, this interaction is limited to very few metrics. Together with the 5G Core analytics the Operation and Management is one of the producer and consumer of the 3GPPP defined network analytics.

Table 4. Management Data Analytics Service.

Analytics	Description	Consumer	Producer
Slice Load Level	Subscriber agnostic load level of a network slice instance	PCF and NSSF	NRF
Observed Service Experience	QoE established in the Service Level Agreement	Any NF, OAM	AF (NEF), AMF, SMF, UPF, OAM
NF Load Information	NF resource status, load, virtual resource usage, including UPF	Any NF. OAM	NRF, OAM, UPF
Network Performance	Network load in an area of interest	Any NF, OAM	NRF, AMF, OAM
UE Mobility	UE Mobility statics or predictions (e.g., location, trajectory)	Any NF, AF, OAM	AMF, AF, OAM
UE Communication	UE Communication pattern (time, period, volume).	Any NF, AF, OAM	SMF, AF, UPF, AMF
Abnormal UE Behavior	Identifies a group of UEs being misused or hijacked	Any NF, AF	AMF, SMF, AF, UPF
User Data Congestion	Congestion in a geographical area or for a given UE	Any NF, AF	AMF, OAM
QoS Sustainability	Potential QoS change in a geographical area	Any NF, AF (V2X)	OAM

Finally, orchestration and control solutions usually provide "AI as a Service"' features (for instance, the inclusion of ACUMOS[3] into ONAP) or analytics services such as root cause analysis or VNF placement suggestions.

[3] www.acumos.org

Summarizing, while there is a large amount of data generated and consumed by the orchestration domain, most of it is restricted to the domain. Very recently, Linux Foundation has also launched the Akraino[4] project, which aims at simplifying the interaction among service providers and edge orchestrators, through the usage of blue-prints.

However, one of the most important aspects of the AI that is studied in the project is the interaction between the business intelligence and the network intelligence through the usage of the service layer, which has been showcased with the ETSI ENI PoC, which is discussed in Section 3.3.3. That is, in a context in which NetApps will be every time tighter, especially with the usage of the AF in the 5G Core.

3.3.2 AI approach in resource forecasting in 5GC

A sufficiently accurate resource forecasting in 5G control plane, can give the operator the chance to reduce its cost and ensure that the virtualized resources are not over or under-loaded. Moreover, resource prediction techniques provide better QoS – that is less UE requests are rejected, so the service is more available to multiple users. This strategy also gives the ability of 5G to support multiple vertical use cases simultaneously on the same infrastructure which meets 5G-TOURS project's scopes. Therefore, it makes sense in trials (Rennes, Athens, etc.) where prediction how to scale in or scale out as output of AI module is deemed to be particularly important. 5G-TOURS UCs that could benefit of having this AI driven tool for adaptive CNF/VNF scaling are the following ones: WP6 – UC10, WP5 - UC6, UC9.

3.3.2.1 PoC Assumptions

Our main assumption is to provide the solution for scaling containerized network function by using AI driven resource prediction. It is related to the analysis of performance problems in the control plane of 5G networks. The main bottleneck is AMF/MME server receiving all connection and session related information from the User Equipment (UE). This issue is more visible in case of IoT traffic characterized by peaks and large volumes of UE traffic. Among the 5G-TOURS use cases we also have IoT traffic like in use cases 6, 9 and 10, where optimal resource prediction could improve the infrastructure utilization.

Our second innovation is about rescaling pods rather than virtual machines. According to the recent telecommunication trends, usually User Plane and Data Plane, as more latency sensitive, are implemented as VM, and CP is migrated as cloud native using containers and microservices. Also, in 5G-TOURS project in the French site WEF2.2 the control plane is deployed as CNF. In our PoC we deploy on k8s cluster our own bases on microservices deployed on several pods and one of it simulates the AMF server.

Our solution takes advantages of Kubernetes API to scale out and scale in one of the cluster pods. We have to reject the solution to engaged the orchestrator ONAP in rescaling one of CNF elements, as it does not support currently any solution that does it, especially that this rescaling is triggered by an extended AI model, which is unknown to ONAP.

Initially, we assumed that our solution would be integrated with the 5G-TOURS architecture in French site. However, it turned out that we would not be able to use the WEF2.2 implemented in French site for rescaling process. WEF2.2 as control plane supports two use cases from French Site: use case 7 and use case 8. Both uses cases are dedicated only for one or a few UEs, so it would not be possible to overload the AMF server. Moreover, the process of rescaling the WEF2.2 currently would require to reset the entire cluster, which would complicate the whole process.

Currently, the data from 5G-TOURS is not available, as the use cases are still under deployment. Moreover, 5G-TOURS UCs will be used only by a few UEs, do not have KPIs connected with the number of UE requests and do not fit the occurrence when AMF server is overload. Therefore, our AI model is trained on mobile

[4] https://www.lfedge.org/projects/akraino/

network traffic dataset collected during the Big Data Challenge organized by Telecom Italia[5]. The data set is Open Source and rich in information.

3.3.2.2 Data analysis and ML model choice

The AI model was trained on mobile traffic datasets. Open source datasets from the Big Data Challenge organized by Telecom Italia, serving as measure of the level of interaction between the users and the mobile phone network, were used to predict the evolution of the load in a core network simulated in the PoC.

The datasets contain information on outgoing and incoming calls, outgoing and incoming sms, and the number of internet call requests (number of CDRs registered in the observed sector) calculated for time intervals of 10 minutes, as well as several features irrelevant to our study. The data are available as separate txt files for each day of 2 months of observations, from November 1st, 2013 until January 1st, 2014.

The initial analysis consisted of cleaning the data, extracting features, and examining the behavior of the resulting time series, as well as formatting the dataset so that it can be used to train the prediction model.

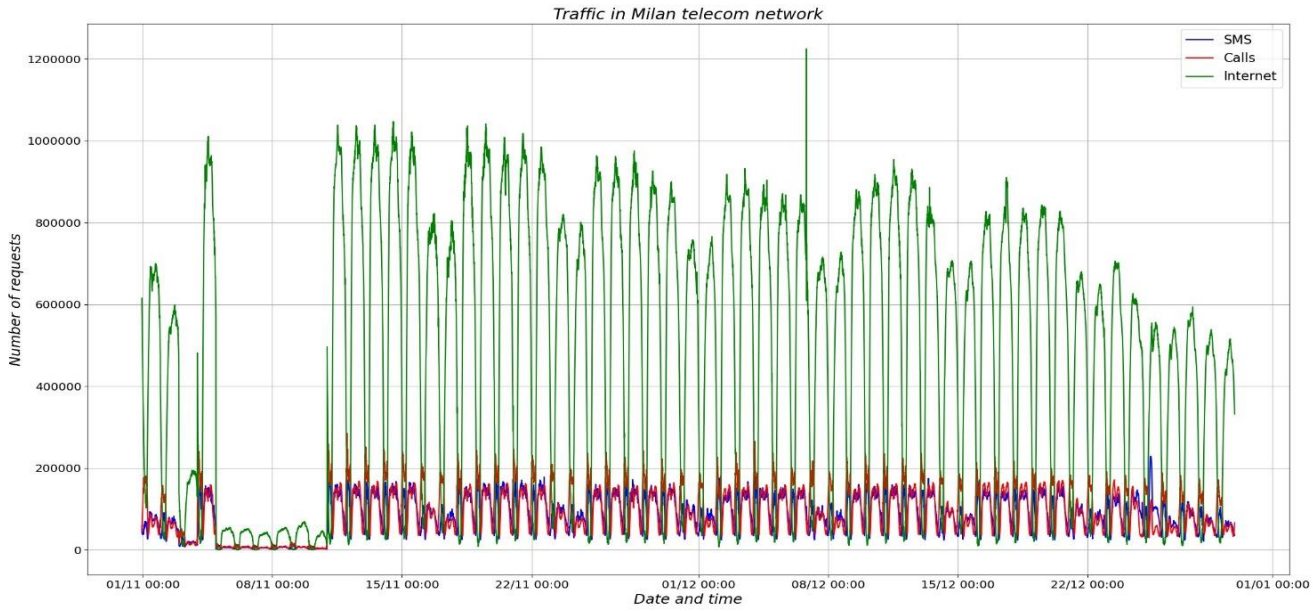


Figure 25. Traffic components in the Telecom Italia dataset.

From the preliminary observation of the data, it appears that all interactions between users and the mobile network can be treated collectively, as they equally load the virtual resources to be multiplied in the PoC.

The first visualization of the data shown an odd behavior of data from the first 10 days and the last 10 days of observations, so for the prediction purposes these periods of observations were removed and the and only data from the period between November 10th and December 22nd were considered.

For further work, hourly data was taken as an average of 6 10-minute samples. This approach did not result in a significant deterioration in data quality, instead it greatly simplified further analysis and the use of prediction to rescale virtual resources.

[5] http://www.telecomitalia.com/tit/en/bigdatachallenge/contest.html

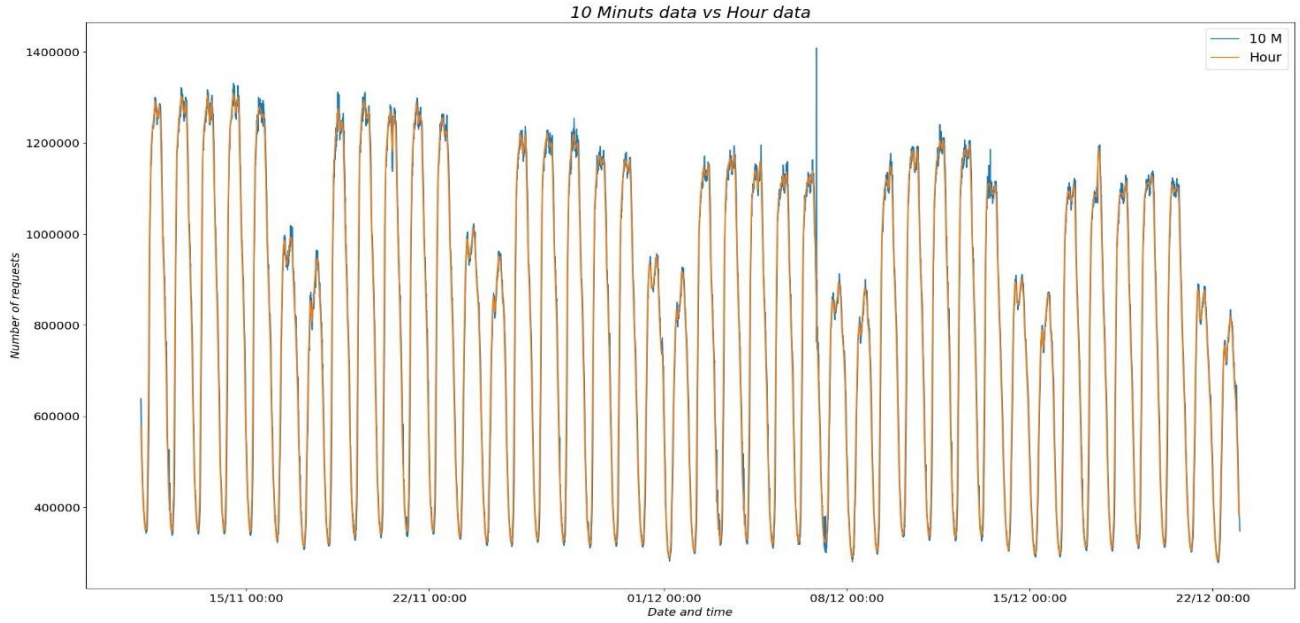


Figure 26. Comparison of 10-minutes samples to 1H samples.

The automated Augmented Dickey-Fuller (ADF) [32] test performed on the dataset showed that the series may not be stationary. However, the decomposition of the time series using STL (Seasonal-Trend decomposition using Locally estimated scatterplot smoothing) [6] revealed the multi-seasonality (daily pattern and weekly pattern) rather than a trend in the data.

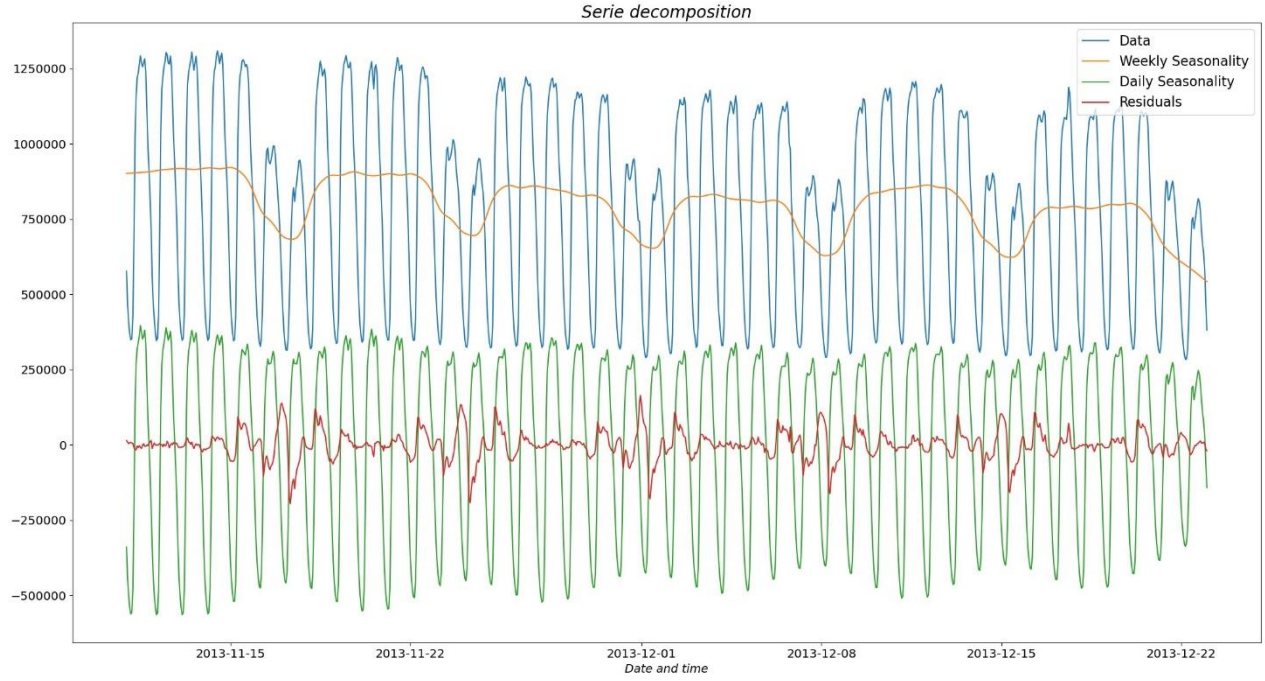


Figure 27. Time series decomposition using STL

Therefore, the choice of ML prediction model was restricted to autoregressive models adapted to seasonal data, derived from the ARMA model.

[6] https://www.statsmodels.org/stable/examples/notebooks/generated/stl_decomposition.html

The Autocorrelation function (ACF) and Partial autocorrelation function (PACF) were applied to determine the seasonality and the autoregressive terms of ARMA model.

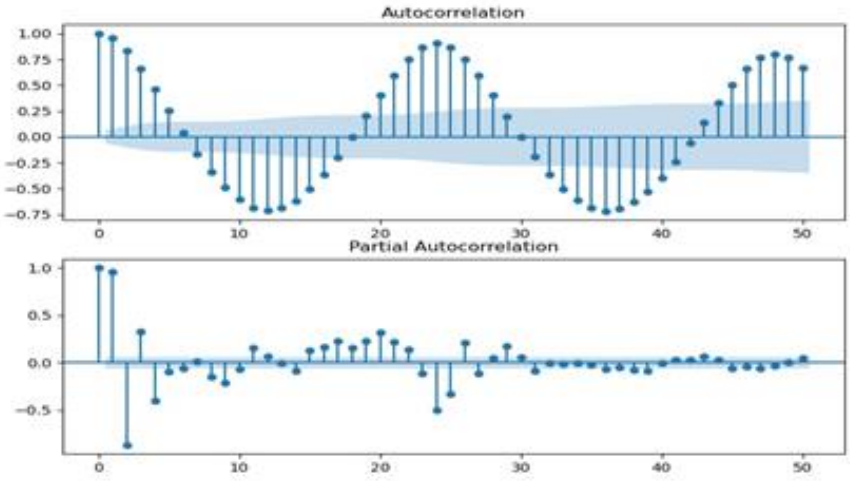


Figure 28. ACF/PACF analysis.

The suitability of the SARIMAX[7] model was investigated, but it proved to be not very accurate in the situation of double seasonality of the data as it overestimates the effect of a decreasing trend, which results in the prediction of underestimated values.

The TBATS[8] (Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trigonometric Trend and Seasonal components) performed much better and it was this one that was eventually adopted as the ML model in the PoC.

For the comparison purposes, both models were trained with 90% of data and tested with the remaining data as well as the prediction for the next 7 days was calculated.

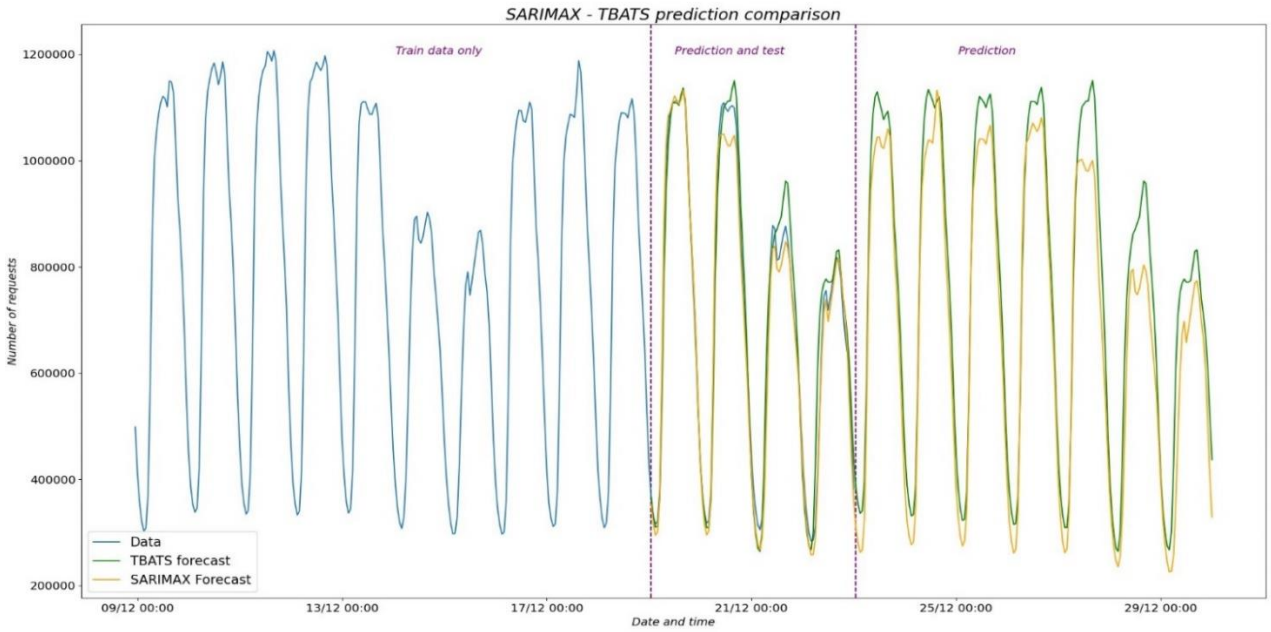


Figure 29. SARIMAX and TBATS forecasts comparison.

[7] https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

[8] https://pkg.robjhyndman.com/forecast/reference/tbats.html

A visual comparison of the predicted time series runs is more favorable for the TBATS model. The measured RMSE (Root Mean Square Error) took the values:

SARIMAX_RMSE = 89033 (approx. 8%) and TBATS_RMSE = 50166 (approx. 4,5 %), which confirm the better prediction with TBATS model (lower error value).

Given the large data values, it is more appropriate to compare the RMSLE (Root Mean Squared Log Error) values, which only takes into account the relative error between the predicted and actual values, with the magnitude of the error being negligible.

The obtained RMSLE values were respectively:

SARIMAX_RSMLE = 0.15619 and TBATS_RSMLE = 0.08755, so again a lower error value for TBATS.

The comparison of metrics validated therefore the intuitive assessment of a better fit of the TBATS model for the given data.

The current PoC assumes a resource scaling levels difference much higher than the prediction accuracy obtained, i.e. the potential difference between predicted and observed number of requests is negligibly small compare to the number of requests triggering the scale-in or scale-out, we succeed to identify a good model and train it to reflect traffic variation over time.

3.3.2.3 Example of model implementation in Kubernetes cluster resource usage optimization

In order to utilize the prediction made by the AI model to scale a Kubernetes cluster we've needed a tool. One that would allow us to both rescale the number of replicas of a pod, but also run said AI model in the first place.

As for the rescaling part, there are only two ways of changing the number of replicas of a pod officially supported by Kubernetes: an automatic and a manual one. The automatic one is based on available resource thresholding (like current percent of CPU usage), which has nothing to do with AI or prediction-based scaling, so we could not use that method. The manual method requires the usage of configuration files. In Kubernetes applications can be deployed by using a configuration file in JSON or YAML format. The number of replicas can be changed in said file and re-applied to the cluster, changing the number of pods.

Since the AI model is a piece of Python code and we've needed a way to dynamically change the contents of a configuration file and apply the updated version to the cluster on the fly, we've made the decision that the best way to accomplish this is with a REST type API. That kind of software could be an easy way to deliver data to the AI model and receive a configuration file back, through a GET type endpoint. With that we were able to design an automatic loop that feeds current network traffic data to the AI model, said model returns a prediction to the API and said API returns an updated configuration file as shown on the infographic below.

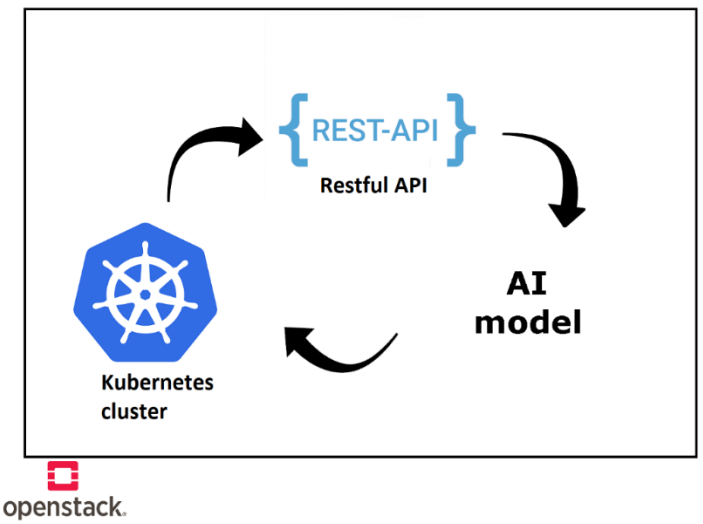


Figure 30. Diagram showing the interaction between the API, Kubernetes cluster and the OS.

The whole process is shown in the graph below. Host OS (on which the Kubernetes cluster is running) sends cluster log data to the API, which directs it to the AI model and awaits a prediction. After said prediction is delivered (which is the peak number of requests), API calculates how many pods are required and then updates the configuration file and exposes it on a specified HTTP endpoint. Simultaneously the prediction alongside the input data is saved to the database for analysis at a later date. At this point new configuration can be applied to the cluster. Should a simulation request be sent to the API without any logfile attached, the whole operation will fail.

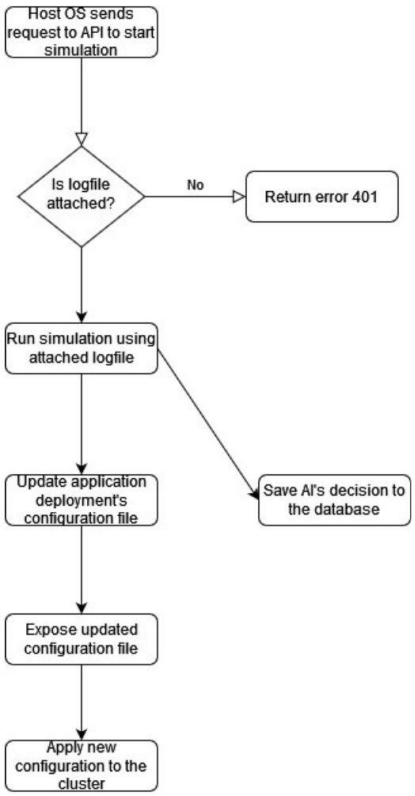


Figure 31. Diagram showing the application loop.

A loop like that is executed once every hour automatically using Linux build-in crontab as a task scheduler. While this interval could be changed it is not advised for it to be too short. While parameters like the amount of time it takes to make a prediction or any latency issues could be worked around with more powerful CPU or better bandwidth, the process of cluster rescaling is unfortunately very unpredictable when it comes to how long it takes to finish. Occasionally pods may linger in a "terminating" state for even a few minutes. Constantly adding or removing pods could lead to unpredictable behavior in the cluster itself. We believe hourly predictions to be the minimum time interval that is stable in a long-term usage while still being very effective at resource optimizing, since within a single hour changes in traffic peaks are not very massive.

Adding more resources to the application (either by physically adding more RAM/CPU or in a cloud environment by adding more pods) increased the amount of requests per second the application can manage. Out testing suggests that the increase in applications ability to manage said requests is rather linear and is a basis for the rest of our proof of concept.

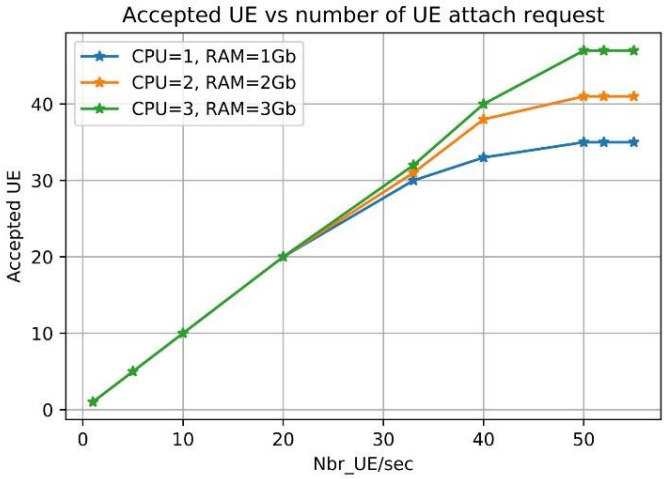


Figure 32. Diagram showing how many more request can be accepted with more resources.

Finally a very important part of the system is the fact our testing suggests that no system can just be scaled by adding more resources (CPU/RAM/bandwidth) indefinitely. As shown on the graph above just doubling the amount of resources does not double the amount of requests per second accepted. Our internal testing suggests that each next pod performs about 5% worse than the previous one.

3.3.2.4 Conclusions from PoC

Predictions with AI algorithms of virtual resource usage in 5G networks is a crucial matter for current and future research works. In some cases, the idea is ahead of the technological possibilities. That was in our Use Cases for French site where ONAP is not ready yet to cooperate with external AI systems and rescale only special, individual pod in orchestrated CNF service. Therefore we were forced to choose Kubernetes API. After studies, it appears that API does not support a large variety of pod scaling methods in k8s cluster. Aside from simple thresholding, the only way to rescale the number of pods is to reload the whole application configuration (or charts for that matter). It is a very suboptimal way. While reloading configuration file with just a changed number of replicas should not interfere with the application, it still creates an enormous security gap. Because the configuration is constantly reapplied it is very vulnerable to attacks with malicious Intent. After all, any change other than the number of replicas in said configuration file would cause the application to reload losing all of its temp data. However, since it is the only way to scale pods in native Kubernetes we had to use it. Predictions of time series process, included requests coming to RAN control network, is also complicated problem. The detailed analysis of the data made it possible to understand the behavior of the observed time series, its seasonality and dynamics of changes. This, in turn, allowed a selection of the optimal prediction method in terms of accuracy and speed of calculation. In our PoC we used the TBATS method, which offers prediction with satisfactory accuracy while not requiring much computational time. The accuracy of the prediction was measured by a RMSLE metric that takes into account the fact that the sample values are very big.

We believe that our PoC is the entrance point to the undiscovered paths for technology improvement and development in supporting 5G virtual infrastructure.

3.3.3 AI for Zero-Touch Network Slicing

When developing the intelligence for the automated allocation of resources in zero-touch network slicing, forecasting holds a fundamental role. Indeed, the orchestrator needs to know in advance the capacity that will be required by each slice to take informed decisions and maximize resource utilization. Unlike traditional prediction, network capacity forecasting is driven by monetary costs: errors lead to resource misconfigurations that entail different economic penalties for the operator [33].

Current state-of-the-art solutions for capacity forecasting in network slicing take into account the costs due to

1) the allocation of unnecessary resources that go unused, and

2) the insufficient provisioning of resources that cannot accommodate the demand and lead to violations of the Service-Level Agreements (SLA) with the slice tenant. Hence, they aim at minimizing overprovisioning while avoiding SLA violations.

However, limiting the problem to this simple trade-off implicitly assumes that resource instantiation and reconfiguration occurs at no cost. While this may hold for some types of resource (e.g. CPU time within the same bare metal machine), it is not generally valid for slice resource management scenarios. Instantiation and reconfiguration costs are capital in NFV technologies that enable the cloudification of the access and core networks by entrusting many network functions to Virtual Machines (VMs) running in datacenters. Examples include baseband processing in Cloud Radio Access Networks (C-RAN), interconnection functionalities towards the external packet networks through the User Plane Function (UPF), or central office operations.

In all the above cases, resource instantiation does not take place for free: VM boot times in prominent public cloud services like Amazon AWS or Microsoft Azure consistently exceed 40 seconds, topping at 400 seconds in worst-case scenarios [34]; even in recent tests, booting a lightweight VM containing an Alpine Linux takes around 30 seconds in a local deployment [35].

Reconfiguring already allocated resources has also a non-negligible cost: modern software architectures such as Kubernetes need several seconds to execute new pods, e.g. on VMs that are already running. In addition, re-orchestration often implies recomputing paths on the transport networks and implementing them via, e.g Software Defined Networking (SDN) architectures: the latency is in the order of hundreds of milliseconds in a small five-switch topology and with precomputed routing, and it has to be scaled to thousands of switches with on-the-fly path re-calculation.

All unavoidable delays above entail monetary fees for the operator, in terms of both violations of the SLA with the tenants e.g. due to infringement of guarantees on end-to-end latency), and user dissatisfaction (with ensuing high churn rates). By neglecting these sources of cost, present capacity forecast solutions risk to introduce uncontrolled data flow latency once deployed in operational networks, ultimately causing economic losses to the operator.

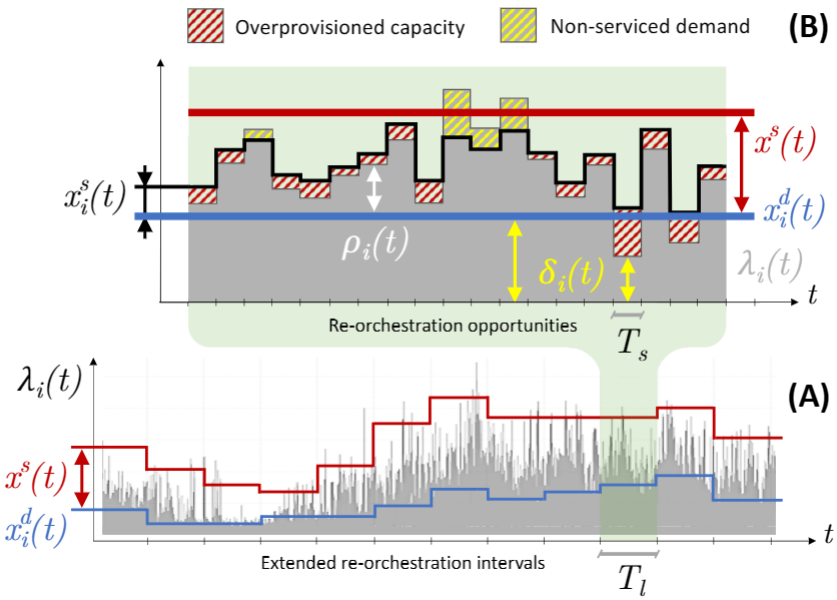


Figure 33. The Orchestration Model.

In [36] we consider a different orchestration model that overcomes these issues, as depicted in Figure 33:

(A) represents long-timescale orchestration. The background time series represents the traffic demand generated by slice i (grey). The curves portray the time evolution of the dedicated capacity allocated that

slice (blue), and of the shared capacity (red) over extended intervals of duration. Note that the shared capacity is added to the dedicated resources to determine the total available capacity, and, unlike the dedicated one, is not reserved for slice i but available to all slices.

(B) is short-timescale orchestration during one extended interval. At every time interval a portion (black solid curve) of the (fixed) shared capacity is allocated to slice i, based on the residual demand not satisfied by the (fixed) dedicated resources. The figure also highlights the volume of overprovisioned capacity and non-serviced demand (pattern regions), and the slice traffic below dedicated capacity.

Therefore, in [36] we propose a deep learning architecture, which we detailed in D3.3 [3] and also implemented for the ETSI ENI PoC. Here, we show further evaluation of the system with an extensive dataset of mobile data traffic collected at 470 eNodeBs of a real-world network serving a large metropolitan region in Europe. The measurement data concerns a set of five popular and heterogeneous mobile services, namely YouTube, Facebook, Instagram, Snapchat, and iTunes, whose traffic flows were classified by the operator using proprietary Deep Packet Inspection (DPI) techniques.

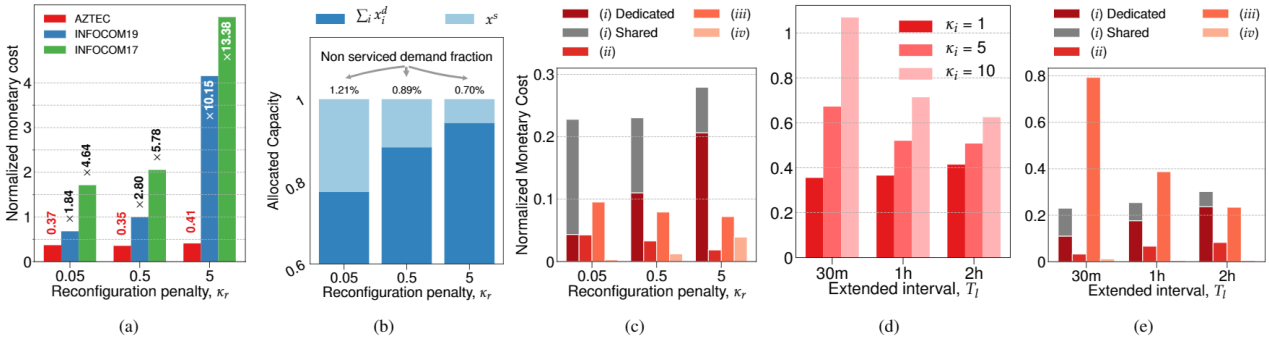


Figure 34. Evaluation results.

The evaluation results are depicted in the figure above. In (a) we show the normalized monetary cost of our solutions and two benchmarks versus the reconfiguration cost scaling factor, which models the ratio of the cost between shared and dedicated resources. Numbers denote the exact cost for our model, and the added cost factor for the benchmarks. In (b) we show the total dedicated capacity and shared capacity allocated by our solution. Numbers denote the fraction of re-orchestration opportunities with insufficient allocated resources. Then, in (c) we breakdown the normalized monetary cost by penalty type. Here we also split the overprovisioning cost into the contributions of the dedicated capacity and of the shared capacity. Subfigure (d) depicts the normalized monetary cost vs the duration of the extended re-orchestration interval, for different scaling factors of the resource instantiation cost. Finally. We show in (e) the breakdown of the normalized monetary cost by penalty type, for a scaling factor equal to 10.

All these results show the capability of our proposed system to adapt to zero touch network slicing scenarios.

3.4 BROADCAST SUPPORT

Multicast/broadcast communications are considered a key capability in the context of 5G systems. The use of point to multipoint transmissions will translate into higher efficiencies in the use of the network resources. 5G-TOURS Task 3.2, Broadcast Support, focuses on the creation of new technologies aiming the delivery of high-quality content distribution through Use Cases 4.b and 4.c [37] which use LTE-Based 5G Broadcast and 5G Native Broadcast respectively. The first Use Case will use 5G Rel.16 compliant equipment. However, Use Case 4.c will use beyond state-of-the-art simulated equipment aligned with the latest updates proposed by 3GPP.

3.4.1 LTE-based 5G Broadcast

UC4.b focuses on the transmission of high-quality video distribution using broadcast delivery and it is further detailed in deliverable 4.2 [37]. The delivery of the content will use RAI's broadcasting network, with a HPHT (High Power High Tower) topology, to all users at once offering great performance independently of the number of connected users.

This use case is split in two phases:

- Phase 1: live or pre-recorded video distribution, using Rel-14 FeMBMS, received in RAI production center premises or Palazzo Madama and transmitted via Wi-Fi multicast to all users using a broadcast tower.

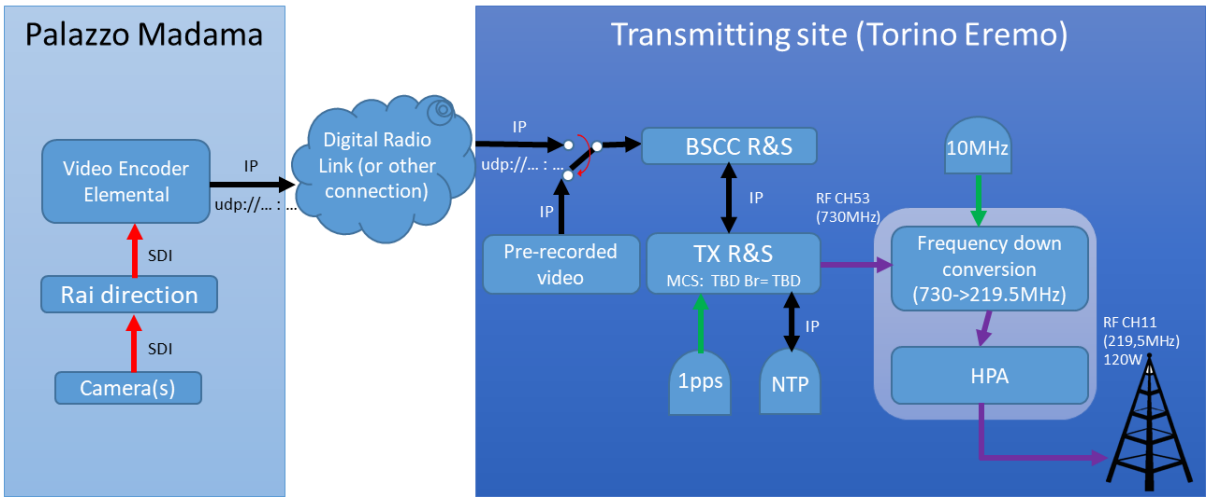


Figure 35. Phase 1 trial's configuration.

- Phase 2 has been updated to live or pre-recorded video distribution provided by UC5 demo (joint demo) using 5G Broadcast Rel-16 updated equipment in two different scenarios. The trial tests first larger coverages from Palazzo Madama. The second one is an in-car scenario to test higher speed mobility. The figure of both subcases of Phase 2 can be found below:

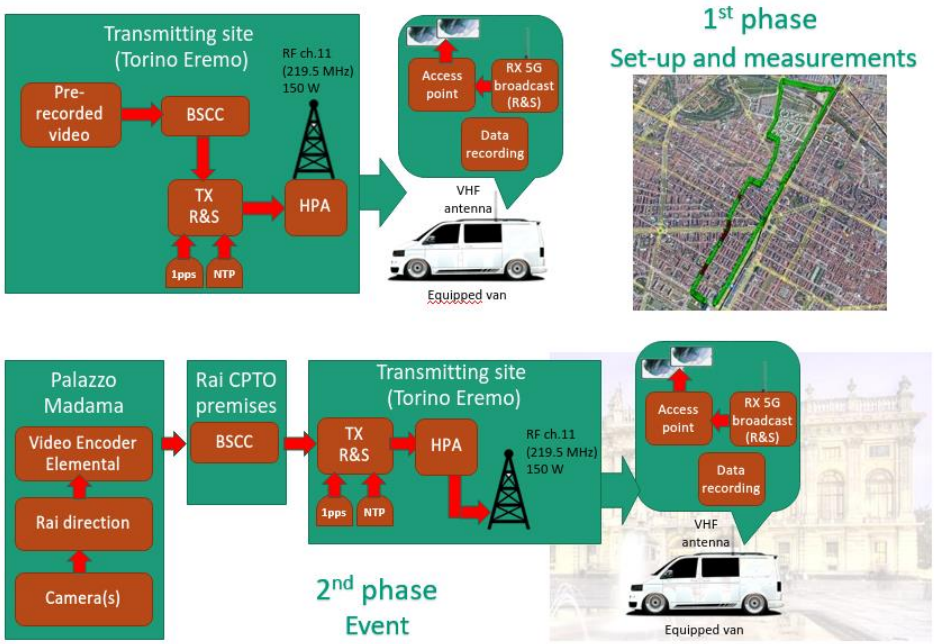


Figure 36. UC4.b phase 2 trials architecture.

Further information about this Use Case is provided in the previous deliverable [3].

3.4.2 5GC Multicast

Contrasting with Use Case 4.b, Use Case 4.c is focusing on the implementation of a 5G Native Multicast system, using 5G technology in both Core and RAN part of the system. Even though the implementation started without a standardized mark from 3GPP, the system has been modified and aligned with the last 3GPP updates.

3GPP standardization

The proposed architecture has been updated in the latest version of TR 23.247 "Architectural enhancements for 5G multicast-broadcast services" [39]. The figure with the architectural graph can be seen bellow:

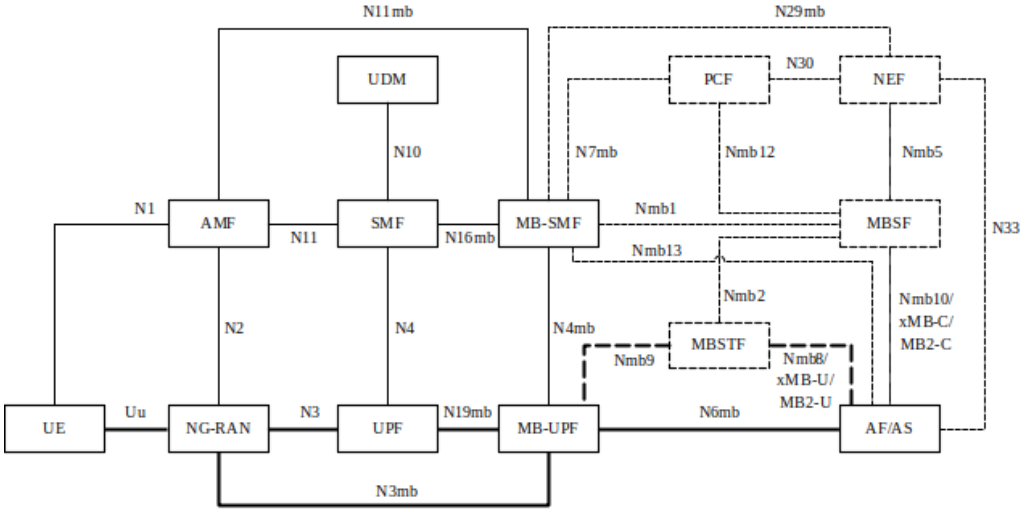


Figure 37. Reference point representation of the architecture for 5G multicast/broadcast services.

In this sense, the implementation focuses on the software simulating the new entities defined in the 5GC transport layer (MB-SMF and MB-UPF) and the updated Network Exposure Function enabling the secure exposure of the new broadcast network services towards 3rd party applications. The new modules will be connected by the standard interfaces: N29mb (NEF - MB-SMF), N11mb (MB-SMF – AMF), MB-N3 (MB-UPF – gNB) and N4 (MB-SMF – MB-UPF).

Since last deliverable, the main innovations from 3GPP come from the aforementioned update of TR 23.247 [39] and from the new TR 29.532 [40] released in August 2021.

On one hand, within TS 23.247 an overview of the different Network Functions Services and the message flows of the different MBS procedures are included.

On the other hand, TS 29.532 specifies the stage 3 protocol and data model for the Nmbsmf Service Based Interface. Stage 2 specifications about architecture and procedures are included into 3GPP TS 23.501 [41] and 3GPP TS 23.502 [42]. The Multicast 5G Core implementation includes the following services:

- Nmbsmf_TMGI Service: Consisting in the allocation or deletion of TMGI values that will be assigned to different services.
- Nmbsmf_MBSSession Service: Consisting in the creation, update and release MBS sessions.
- Nmbsmf_Reception Service: Enables to request the start or termination of MBS data reception for a multicast session (this service is not compatible to broadcast).

The document also includes a detailed explanation about Nmbsmf_Information Service but this service is not considered necessary for this use case and will not be implemented. OpenAPI specifications for these services are also published in the TS. However, these specifications are not completed and will be further detailed in future releases of the document.

Developing process

As it was explained in the previous deliverable, this project is aligned with the idea of broadcast as a service, hence it aims the architecture proposed in TR 26.802 [49]. The mentioned architecture is shown in Figure 38:

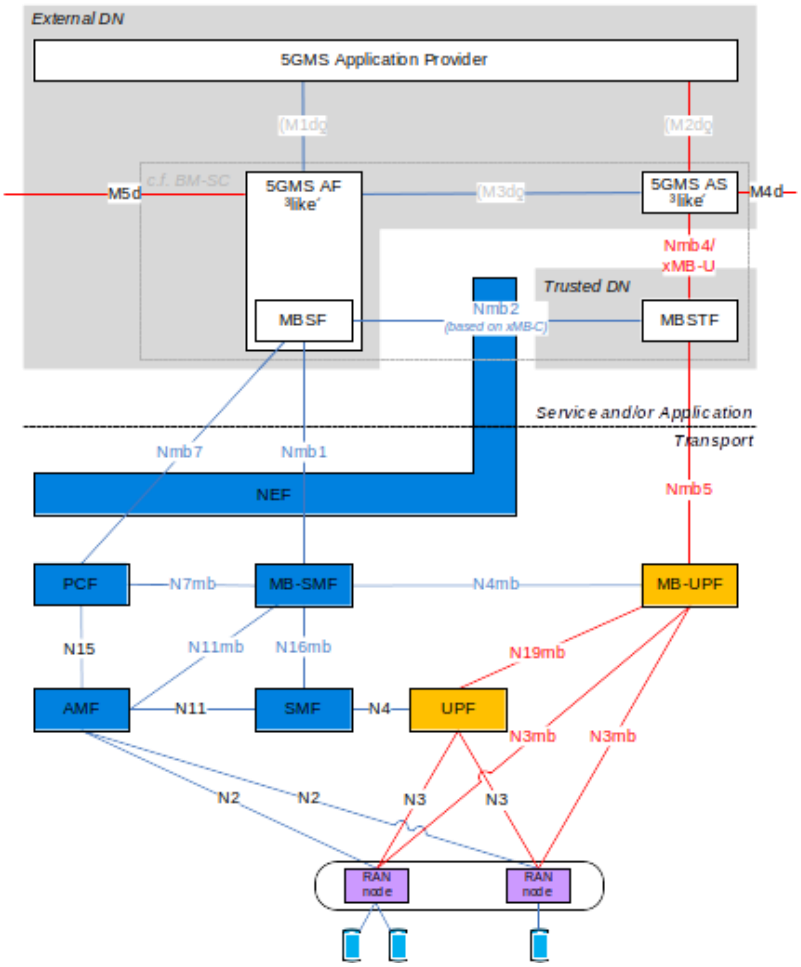


Figure 38. TR 26.802 proposed architecture.

In the use case, UPV is implementing an equivalent model adapted to the mark of 5G-TOURS. In addition, the trial required to be updated from the model proposed in deliverable D3.3 involving 5G EVE premises in them. In this sense, Enensys' Service Layer will be placed into 5G EVE premises. To conclude with the configuration, UPV and 5G EVE lab networks have been updated in order to enable the connection. The final proposed model is shown in following figure.

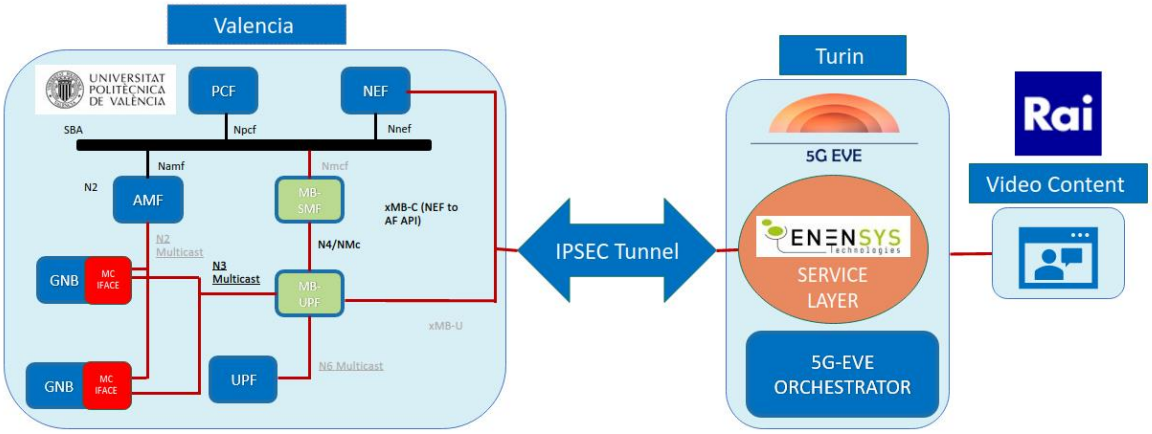


Figure 39. 5G-TOURS adapted Use Case 4.c architecture.

The 5G Core Multicast software development is nearing completion and is expected to be finished in the next months. The system has been implemented into UPV premises and included inside the software Open5GCore (O5C). Open5GCore is a professional implementation simulating a 5GC whose last version is Rel.16 compliant. MB-SMF, MB-UPF and NEF modules will be implemented adding some enhancements to the original SMF, UPF and NEF modules provided by O5C.

In the scope of the data plane, MB-N3 will use 5GC shared MBS traffic delivery developed using the Internet Group Management Protocol (IGMP) enabling one-to-many communication between the NG-RAN and the MB-UPF. In addition, Packet Forwarding Control Protocol (PFCP) has been extended in order to provide the correct functioning and support of the multicast-broadcast features in the N4mb interface. The modified Open5GCore will be tested using simulated multicast gNBs due to the lack of Rel.17 commercial equipment. As an alternative, an IGMP endpoint is being developed. This new component will provide unicast gNBs with the capability to receive the 5GC shared MBS traffic using IGMP by converting multicast to unicast and the correct functioning of the system can be checked. The aforementioned services will be included in MB-SMF and in NEF in order to expose the service to the external non-trusted applications functions.

These functions will be implemented in the context of the WP4. 5G EVE will be enhanced with multicast/broadcast support with the additions included by 5G-TOURS.

3.5 SERVICE LAYER

In this section, we discuss the instantiation of novel technologies linked to the service layer. 5G-TOURS provides solution for the Enhanced MANO (Section 3.5.1), which is embodying the needed extensions to incorporate the usage of AI. We present also the design of the "AI-Agents for OSM" (section 3.5.2) and PoC for Autonomous Network Slice Management (section 3.5.3). We have also developed a state-of-the-art solution for the broadcast support (Section 3.5.4). Finally, in section 3.5.5, we present 4 open-source initiatives developed for the Service layer to support some of these technologies.

3.5.1 AI-enhanced MANO

3.5.1.1 Overall Architecture

The 5G-TOURS Greek site architecture descripted also in D3.3 [3] has been updated and the set of components are illustrated in Figure 40. The Inter-Working Layer (IWL) is connected to the site's orchestration infrastructure via the Multi-Site Network Orchestrator (MSNO). The Greek site's orchestration infrastructure introduces an AI-enhanced Management and Orchestration (MANO) component, utilizing AI algorithms that work together with the Open Source MANO (OSM) to provide intelligence to the system. In addition, a Virtual Infrastructure Manager (VIM) (i.e., Openstack) and an open-source stream-processing platform (i.e., Kafka Cluster) are also available for deploying VNFs and capturing data respectively. The Runtime Configurator at the IWL is responsible for any last-minute configuration of the VNFs that are deployed with the help of the AI-enhanced MANO and OSM at VIM. Also, a Central Kafka broker and a Monitoring module are included in the IWL to monitor the deployed VNFs and gather the required data for further analysis. These data can be used by the Performance Diagnosis tool, which has been developed based on data analytics processes to guarantee the full exploitation of the service metrics and KPIs and enable the acquisition of in-depth insights per vertical category and use case. At the final architecture, the AI-enhanced MANO component acquires real-time information from VNFs, applications, infrastructure and also receives in real-time feedback from the Performance Diagnosis tool regarding the quality offered and the performance achieved by the infrastructure. Finally, the main 5G EVE Portal is involved through the onboarding procedure, by stating various information in specific forms allows the design, creation, scheduling and deployment of vertical network services (NS).

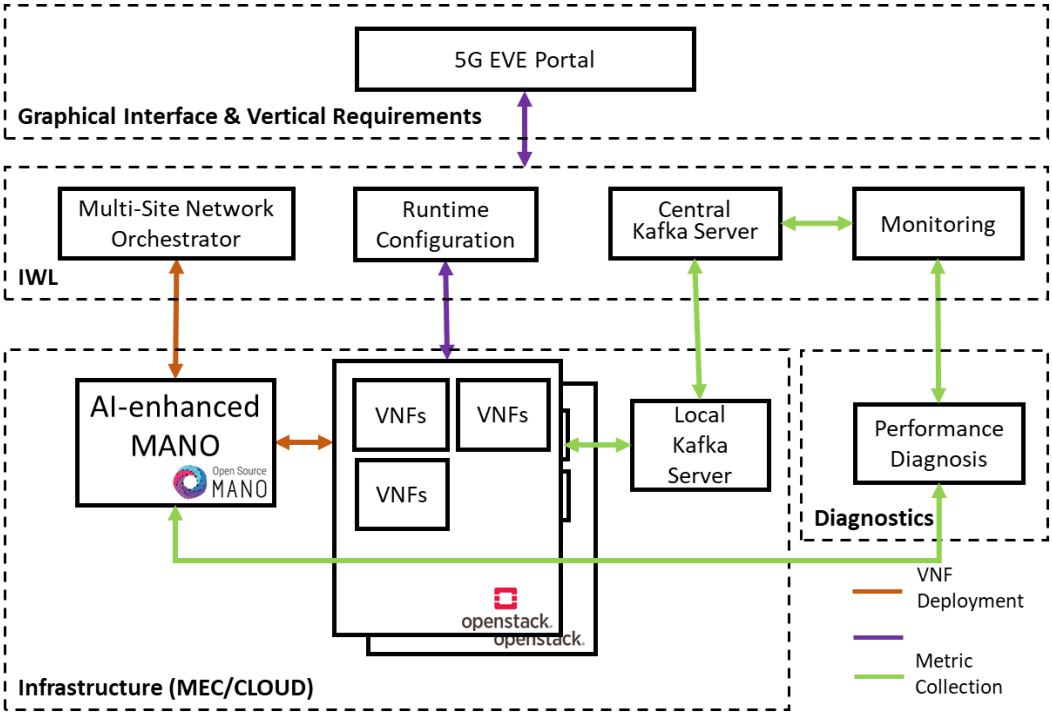


Figure 40. 5G-TOURS Greek site architecture.

For an experiment to be executed on the 5G-TOURS platform the necessary blueprints, Vertical Service Blueprint (VSB), Network Service Descriptor (NSD), test case, experiment and execution descriptors need to be onboarded on the system. The AI-enhanced MANO will orchestrate and Manage the NS that will be deployed as VNFs at VIM. When all steps are completed, the experiment will run for the allocated time slot. These steps are performed from the main graphical user interface of the project (5G EVE Portal). In addition to this GUI a more specific web interface to showcase the AI-enhanced MANO system, has been designed and developed as Figure 41 illustrates. This interface retrieves information from the AI-enhanced MANO component and presents in bar, pie and graph charts the different metrics from the VNFs and Infrastructure also presenting the migrations between different MECs or the optimal distribution of different VNFs in an automatic way when the AI deems necessary to do so.

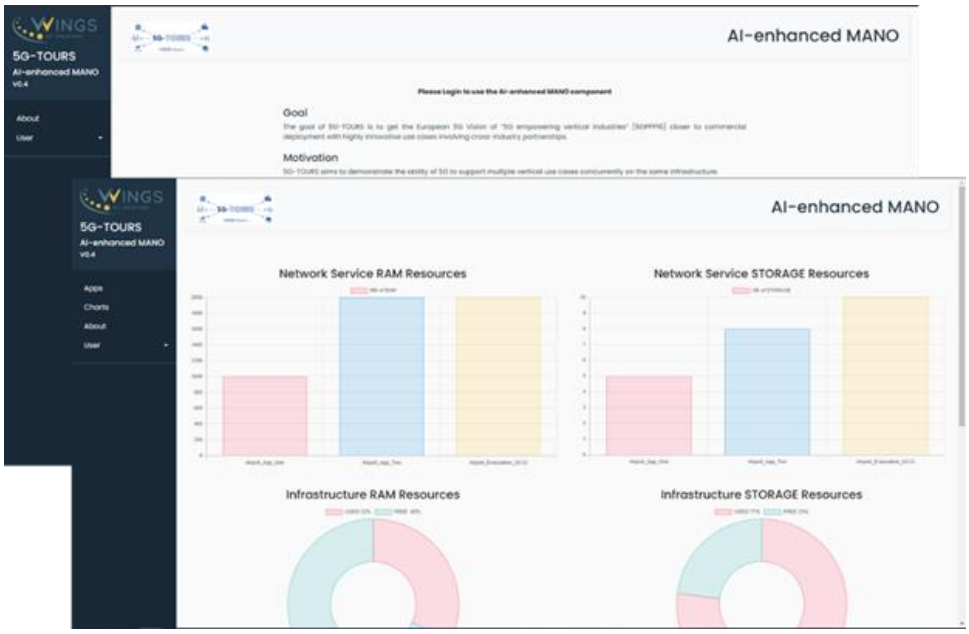


Figure 41. AI-enhanced MANO Graphical User Interface.

3.5.1.2 Performance Diagnosis

At D3.3 Chapter 5.5.1 [3] the AI –enhanced MANO architecture, interfaces and their connections have been presented. There the connection between AI-enhanced MANO and Diagnostic/Metrics has been descripted extensively. At the final architecture and deployment of the 5G-TOURS system the diagnostic component is deployed as a specialized Performance Diagnosis tool which is responsible for collecting information during an experiment, parsing that information, analysing it, and producing insights on the performance of the various elements that comprise the experiment. It is also responsible for detecting anomalies in the behaviour of the various elements as well as identifying the root cause of possible problems or performance degradations during the experiment. To accommodate the real-time operation of the AI-enhanced MANO mechanism, the diagnostic component has been updated in terms of synchronization with the overall workflow. Instead of reading and analysing the collected information (metrics, KPIs) at the end of an experiment, its operation (analysis, performance diagnosis) is now in real-time as well, to provide the proper insights as an input for the AI-enhanced MANO mechanism, when performance degradation is detected in a deployed service.

For the latter, interfaces are developed for the various internal components of the Performance Diagnosis tool to communicate with each other in real-time. This allows the processes of collection, pre-processing and analysis of the generated data to take place on the spot, without any functional overlaps. In addition, the Performance Diagnosis structure allows for different diagnostic algorithms to be deployed. These algorithms can be tested and deployed completely interchangeably, thus exploiting the modular layout of the tool. Finally, this modular layout allows chaining different algorithms together for more in-depth analysis and conclusion. In this case, the results of the anomaly detection module (produced by Machine Learning algorithms) are sent to the fault localization module to determine the root of the detected anomaly.

To utilize the modular layout described above, the Performance Diagnosis tool used in 5G-TOURS will embody an extra Machine Learning algorithm (compared to the respective tool used in 5G EVE). Comparison tests will take place to evaluate the performance of the ML diagnostic algorithms implemented, or any other algorithms that could be deployed in the future. To this point, the health status of each element of a service is determined by an anomaly detection component based on a Self-Organizing Maps (SOM) algorithm, as developed for the 5G EVE platform, but parameterized (in terms of the component's I/O and learning process) to operate in real-time. Alongside the latter, an implementation of the Density-based spatial clustering of applications with noise (DBSCAN) algorithm is deployed for 5G-TOURS, to operate as an interchangeable component of the PD tool used for anomaly detection as well. Both algorithms can utilize pre-trained models that will be enriched as the respective deployed service runs, and will be used to determine the health status of each node, based on the node's previous states (historical data).

A low complexity Root Cause Analysis algorithm has been implemented to localize a problematic node that causes performance degradation to other nodes and the deployed service in general. The basic principle of the algorithm is to check the reachability between the nodes (in our case, instances that belong to Virtual Network Functions), using each node's health status that is determined by the health status calculated in the anomaly detection module, and the network topology.

Besides the main output of the PD, used for monitoring and diagnosing a deployed service, a secondary output is fed to the AI-enhanced MANO component's interface. The produced diagnosis is used by the AI-enhanced MANO to make decisions on a service's migration (or even termination) during the service operation, using the health status (or the root cause status, if such arises) of each service node, as diagnosed by the PD tool. An overview of the PD tool's workflow can be shown in Figure 42.

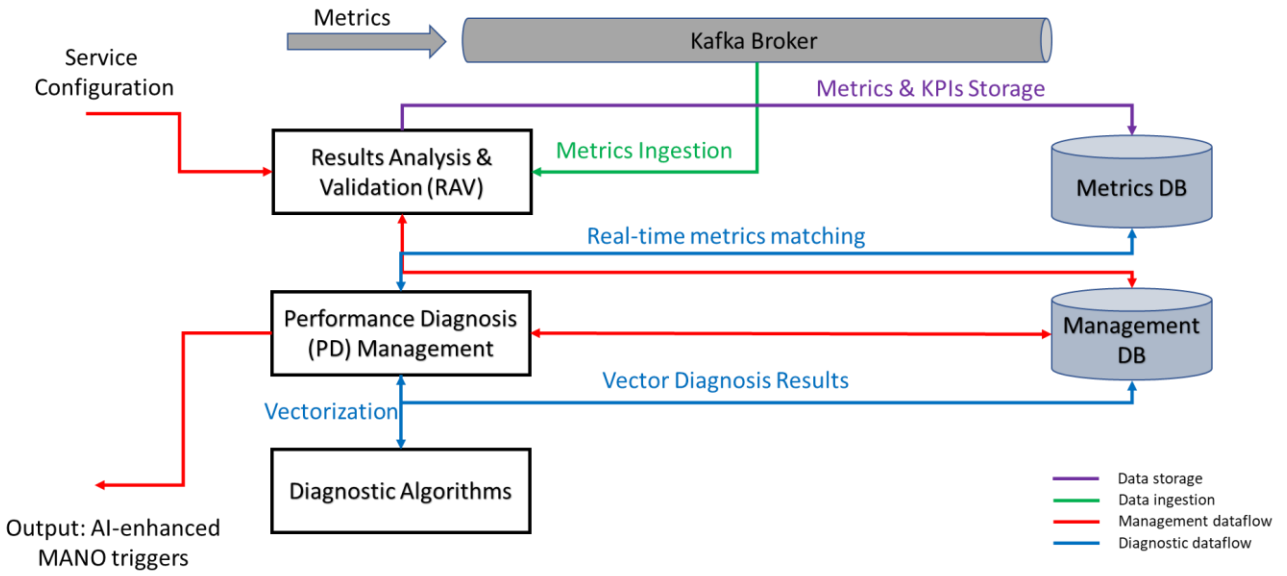


Figure 42. Workflow overview of the Performance Diagnosis tool.

3.5.1.3 Cooperation of the AI-enhanced MANO & Performance Diagnosis tools

As described in D3.3 [3], the AI-enhanced MANO embedded tool will analyze the various data and historical information to find out if changes to the deployment of VNFs need to be done. If a metric does not meet the vertical requirement for a specific VNF, OSM will make the necessary changes to fix it, either by relocation of the VNF, scaling or other appropriate actions. Some of these metrics are vertical requirements, real-time infrastructure and application metrics. User applications, each of which is composed of a set of Virtual Network Functions (VNFs) can be deployed at different Mobile Edge Computing (MEC) and Cloud infrastructures. VNFs typically do not require constant resources but need resources based on the traffic and computational load. In some cases, the allocation or re-allocation of a service's resources is based on the particular service priority, or the priority of the services that are already deployed in the respective infrastructure, as described by the vertical. The AI-enhanced MANO's functionality is to determine the re-allocation of a service's resources (based on the current load at the infrastructure and the performance of the service), the migration or termination of another service with lower priority than the service that needs to be deployed at a given moment (based on the services' SLAs), and the overall oversight of the deployed services.

Apart from the decisions made by the AI-enhanced MANO tool, regarding the resource allocation of a service prior to its deployment, an important feature of the tool is the resource re-allocation decisions made as a deployed service operates. A major factor in those decisions is the knowledge on a service performance, as produced by monitoring and constantly diagnosing any anomalies. Thus, the output of the Performance Diagnosis tool plays an important part in the AI-enhanced MANO operation. When the PD tool detects a certain anomaly that directly correlates with a specific resource (i.e. a system or network resource), the AI-enhanced MANO is triggered with a form of alert, to take action on overcoming the diagnosed anomaly.

To demonstrate the functionality of the AI-enhanced MANO tool, as well as its cooperation with the Performance Diagnosis tool, an example scenario follows:

As a specific number of services run on an infrastructure, in the form of VNFs implementing 3 5G EVE & 5G-TOURS Use Cases (UC 1-3), a request is made by a vertical, for a new Use Case to be deployed (UC 4). A lack of resources is detected, thus an action on managing the infrastructure's resources need to be taken. AI-enhanced MANO decides the termination of a low-priority service, in order for the new, high-priority service to be instantiated, as shown in the next figure:

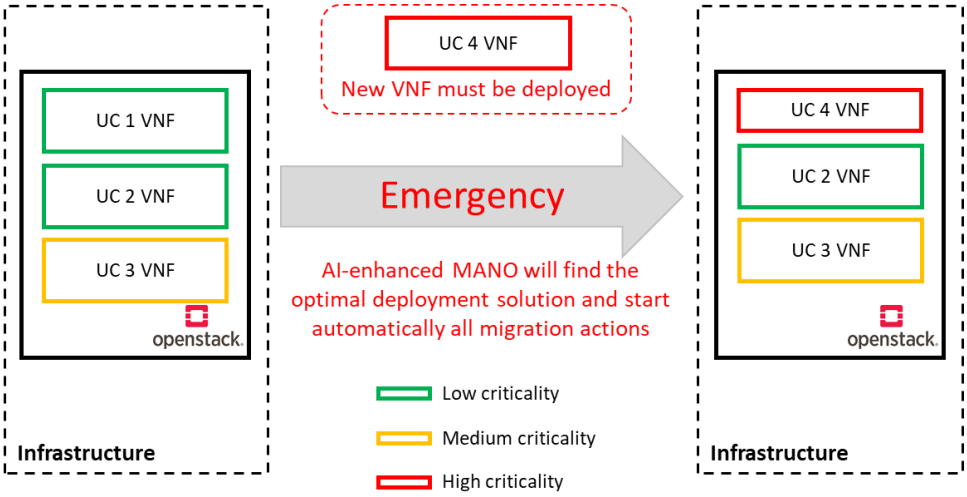


Figure 43. AI-enhanced MANO migration operation of critical use case.

As UC 3 VNF is instantiated and operating normally, a system anomaly on this VNF is detected by the PD tool. The real-time diagnosis produced, points out that a specific resource of the service is highly utilised. To address the issue, AI-enhanced MANO decides to increase the specific resource. As shown in Figure 44, UC 3 VNF is re-instantiated with the new, increased resource, once the viability of this re-allocation on the infrastructure is confirmed by the tool.

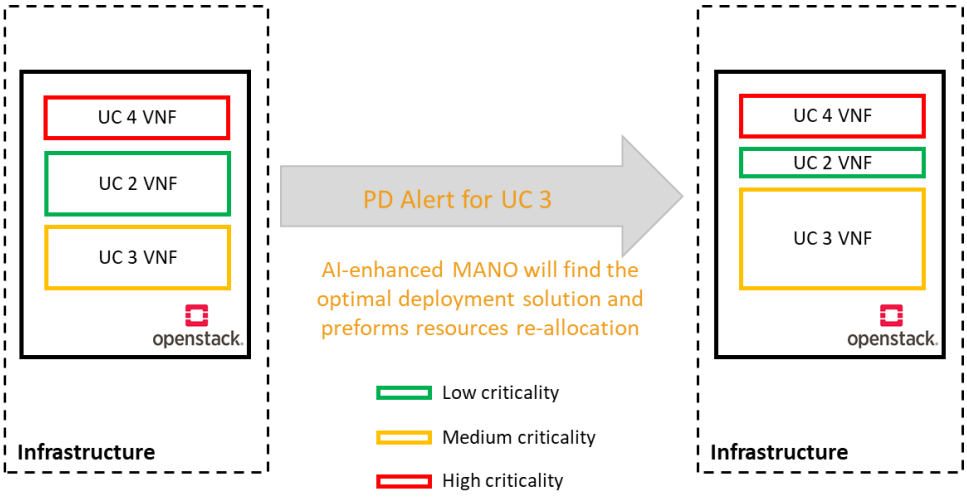


Figure 44 AI-enhanced MANO performance diagnosis operation.

3.5.2 AI-Agents in OSM

As already introduced in the previous Deliverable D3.3 [3], the design of the "AI-Agents for OSM" functionality with two main functional blocks (the AI-Agents themselves, and the AI-Models Servers - see Section 3.2.1) makes possible to deploy this solution with a clear separation of competence areas: AI-Agents themselves would be deployed attached to the VNFs orchestrated from the OSM platform (within the 5G EVE scope), while the AI Models Server would be deployed as an ancillary component outside the MANO layer (e.g., in the Verticals scope).

With a deployment performed in this way the communication between these two domains (the AI Agents deployed from the OSM orchestrator and the AI Models Server in the Verticals scope) would be performed through the 5G-TOURS Service Layer (see Figure 45).

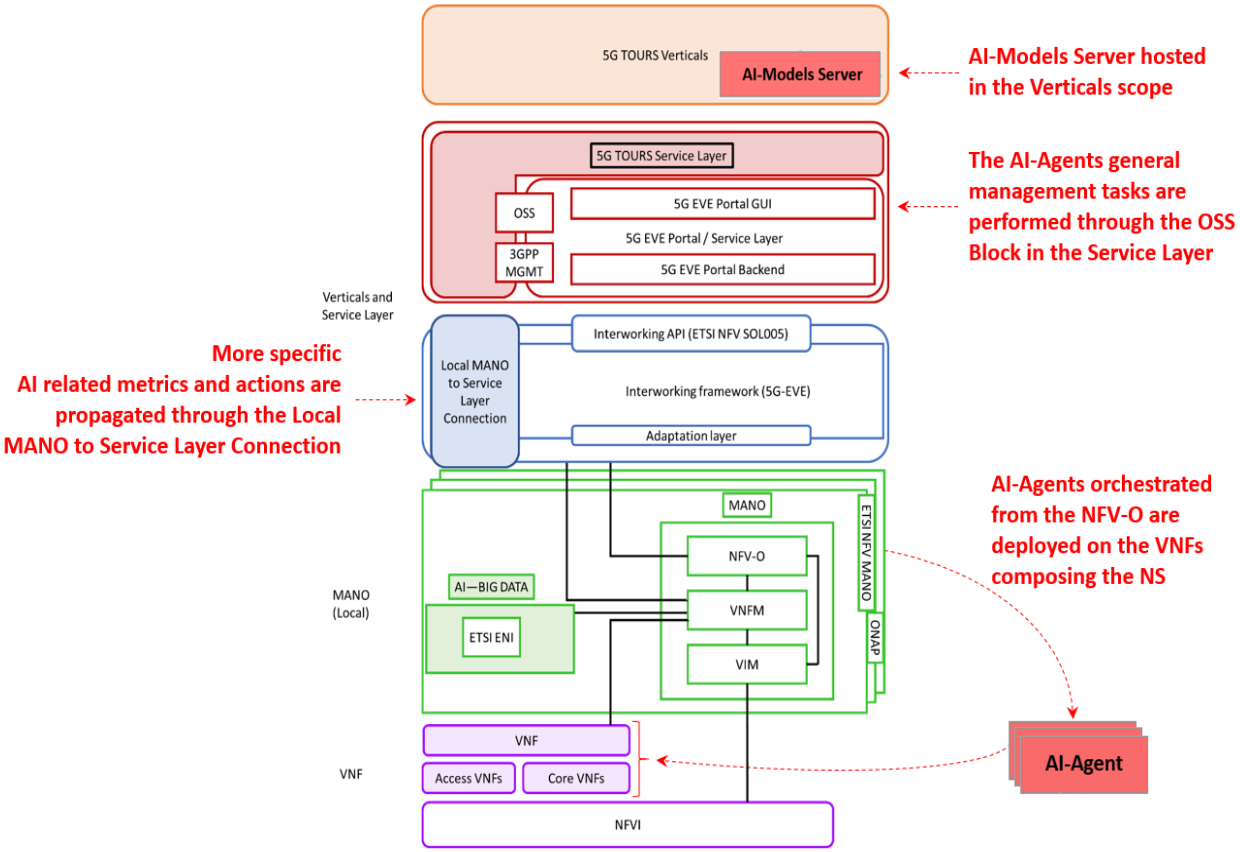


Figure 45. Service Layer integrating AI Agents and AI-Models Server.

In this context, two different workflows would be considered:

a) The regular life-cycle management (LCM) tasks associated to the AI Agents (i.e., the regular CRUD operations).
b) The AI/ML related operations (AI models design and training).

For the first (regular CRUD operations), since AI-Agents are designed as software components that are executed on the VNFs Execution Environment, the general LCM tasks would be performed from the NVF-O in the MANO layer (OSM in or case) and may be handled by from the OSS block in the 5G TOURS Service Layer (through the 5G EVE Portal).

Regarding the second (the more specific actions related to the design and training of the artificial intelligence models), they would be delegated to the vertical. Of course, this second workflow would be executed separately whenever it was necessary to integrate a new AI Model or (re)train an already deployed one. The connection between the AI Models (in the vertical's scope) and the associated production AI Agents would be enabled only when the model was ready to go into production.

For implementing these two workflows two communication channels would be used: the ETSI NFV SOL005 Interworking API for the first one, and the Local MANO to Service Layer Connection for the second (see blue layer in the middle of the Figure 45). This second communication channel would be used by the AI Agents to query the AI Models Server, and also, for gathering metrics that could be necessary to train AI Models. Since the "AI Agents for OSM" functionality does not impose any specific technology for implementing the AI Models Server, the design of the interface for implementing this second workflow would be at the discretion of the implementation model decided in each specific case (the implementation performed in this 5G TOURS project has been based in on TensorFlow Serving for the AI Models Server [43] and a specific REST API).

3.5.3 Service layer for the vertical closed-loop integration

3.5.3.1 Example of closed loop service level assurance

Predictive analytics as the ones described in Section 3.2.2 can be used as input to a policy decision and execution framework at the management and orchestration functions within the Network Domain.

In particular knowledge about future states of the network may help in estimating the corresponding quality of service / user experience and to decide the point in time and the nature of the policy governing the action to be taken in order for the network fulfill the expected performance requirements.

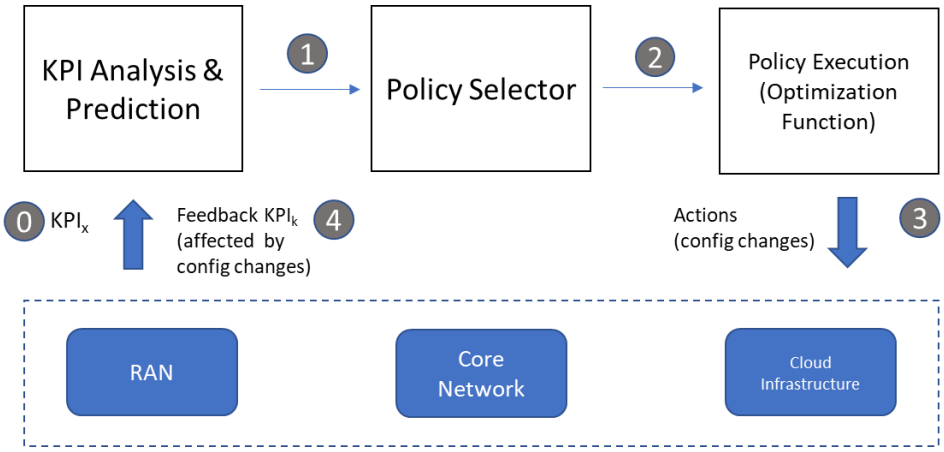


Figure 46. Closed loop architecture.

The following steps describe a simple use case:

- Step 0: the ML model is trained for cells Cell 1, Cell 2, Cell 3, Cell 4, etc. on the past period T.
- Step 1: at time t_0, predicted KPI_1, ..., KPI_n (for instance regarding traffic, service and user characterization) are calculated for each cell. Predicted values, together with estimated errors, are sent to a Policy Selector.
- Step 2: Based on the evaluation of the KPI, the Policy Selector evaluates the proper policy to adopt, e.g. no action, temporary load balancing between 2 cells (with estimated time to revert to the previous configuration), stepwise load balancing, etc.
- Step 3: actions related to the selected policy are indicated to the optimization function which decide and execute the changes on the network settings.
- Step4: feedback KPI_k affected by the actions previously decided are fed back for a KPI analysis that helps the Policy Selector to determine the effectiveness of the configuration changes. For instance, such KPIs may be related to the network resource usage

3.5.3.2 ETSI ENI PoC for Autonomous Network Slice Management

The PoC presented for ETSI ENI has been integrated in the network architecture, as depicted in Figure 47 below.

The functionality is spread among three different layers of the architecture: the vertical business intelligence, which is out of the scope of this work, is hosted in the vertical domain. Then, the PoC technology is using the connection, for the exposure of the parameters (the knob, as discussed in D3.3) as this functionality has not been integrated into the 5G EVE portal. Then the specific resource orchestration resides in the Local MANO, leveraging on the existing ETSI interfaces, as discussed in the PoC Report [52] and in D3.3 [3].

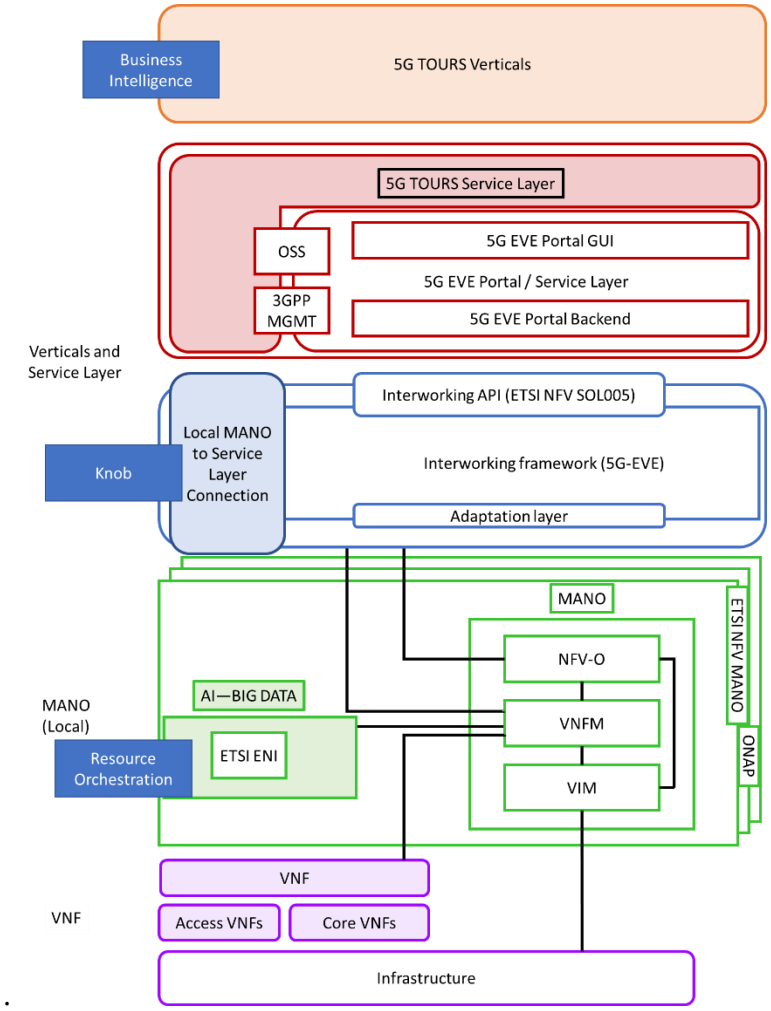


Figure 47. The integration of the PoC in the overall architecture.

3.5.4 Multicast/broadcast functionality

The work on the service layer for multicast/broadcast support follows two axes: the development of the new network functions being specified at 3GPP, and the optimization of the transport delivery stack.

3.5.4.1 **5G architecture for the multicast/broadcast service layer**

The devised architecture will be compliant to 3GPP architecture as described in Section 3.4.2. As the 3GPP architecture is not yet finalised, the architecture will certainly evolve during the 5G-TOURS project. According to 3GPP work in TS 26.502 [55], 2 new entities are defined in the 5GCore Service Layer:

- **Multicast Broadcast Service Function (MBSF)**: The functionality of the MBSF is defined in clause 5.3.2.11 of TS 23.247 [39]. The MBSF performs the following functions to support MBS:
    - Service level functionality to support MBS, and interworking with LTE MBMS
    - Interacting with AF and MB-SMF for MBS session operations, determination of transport parameters, and session transport.
    - Selection of MB-SMF to serve an MBS Session.
    - Controlling MBSTF if the MBSTF is used.
    - Determination of sender IP multicast address for the MBS session if IP multicast address is sourced by MBSTF.

It receives provisioning and control commands either directly at reference point Nmb10 or at reference point Nmb5 (via the NEF). The MBSF invokes MBS Session operations on the MB SMF at reference point Nmb1. The MBSF configures the MBSTF at reference point Nmb2.

- **Multicast Broadcast Service Transport Function (MBSTF):** The functionality of the MBSTF is defined in clause 5.3.2.12 of TS 23.247 [39]. The MBSTF performs the following functions to support MBS if deployed:
    - Media anchor for MBS data traffic if needed.
    - Sourcing of IP Multicast if needed.
    - Generic packet transport functionalities available to any IP multicast enabled application such as framing, multiple flows, packet FEC (encoding).
    - Multicast/broadcast delivery of input files as objects or object flows.

    It receives User Plane data traffic at reference point Nmb8 and sends MBS data packets to the MB UPF via reference point Nmb9.

The multicast broadcast service functions provide the transport functionalities offered in 4G by the BM-SC. The inclusion of these service functions has been discussed a lot within the 3GPP SA2 working group, as only a transparent transport of IP flows was initially considered. The 5G-TOURS service layer for multicast/broadcast service layer implementation will be based on the current virtualised BM-SC, nevertheless a continuous look at 3GPP activities is needed to verify the progress on the new interfaces specified by the 3GPP SA2 working group.

3.5.4.2 Low latency MABR delivery

A 3GPP multicast/broadcast Service is usually based on DASH/HLS. DASH and HLS are also widely used for Over The Top (OTT) services in order to offer high quality streaming of media content over the Internet, delivered from conventional HTTP web servers. DASH and HLS works by breaking the content into a sequence of small segments containing typically few seconds of data.

ABR technique suffers from a significant latency (up to 30 seconds for HLS) compared to broadcast delivery, mainly due to the need of buffering segments at the point of reception, which can jeopardize the viewing experience, especially for live event, such as sport. The support of low latency has also been specified by DVB for M-ABR DVB using this solution.

3.5.4.3 Service Layer error correction with RaptorQ

Raptor codes are the first known class of fountain codes with linear time encoding and decoding. Raptor codes perform software-implemented Forward Error Correction (FEC), thus, there is no need to integrate any additional dedicated hardware in the system to perform AL-FEC. The increment of reliability in data transmission derived from Raptor Codes determined its standardization for MBMS AL-FEC.

However, it exists a new generation of fountain codes named RaptorQ [44]. RaptorQ, developed by Qualcomm, is the most recent and improved version of Raptor codes. This code effectively reduces the redundant FEC information outperforming conventional Raptor codes and offering near-optimal properties with minimum processing overhead and excellent flexibility in packet lost recovery.

In reception, RaptorQ decoder only needs slightly more information than the original to recover from high Bit Error Rate and even packet loss. In practice, RaptorQ only needs 2 extra symbols comparing to the original quantity to reduce Bir Error Rate to less than one in a million. In addition, its excellent flexibility enables the same decoding performance no matter if the received symbols are source or repairing symbols.

RaptorQ will be included inside the service layer to perform the application layer forward error correction process. High quality video will be encoded with RaptorQ technology before being injected into the MB-UPF. The encoded data will be decoded inside the SDR-based UE. Multicast sessions enhanced with AL-FEC RaptorQ technology will optimize both efficiency and reliability allowing high quality video transmission even in the most challenging environments.

3.5.5 Service Layer SDK

The project contributes to 4 main open-source initiatives linked to the service layer SDK (Software Development Kit). The 5G-TOURS Service Layer SDK is provided as an interface composed by a set of APIs exposed to the verticals. The consortium is currently producing 4 different APIs for the Service Layer, according to the specific targeted functionality and subject to partners' policies. These are described in the following subsections.

3.5.5.1 AI-Agents in OSM

The "AI Agents for OSM" functionality has been released to the Open-Source community through a specific GitLab public repository[9]. Also, a complete user's guide has been made available through the public online documentation service 'readthedocs.io' [10]. From these repos, the technical documentation and the full source code of the solution can be accessed under the Apache 2 open-source license.

The collection of the software resources and the information available from these repositories constitutes an SDK by itself, which can facilitate the development of Network Services based on the "AI Agents for OSM" functionality described in Section 3.2.1.

3.5.5.2 ETSI ENI PoC for Autonomous Network Slice Management and Vertical closed-loop integration

One of the key innovations of the project is the introduction of AI, which is needed to support large-scale deployments such as the ones envisaged in 5G-TOURS. To this end, the project has been working on AI-based management of the network and pushing them to these relevant standardization fora, in order to promote the wide commercial adoption of the project solutions.

As part of the effort in open sourcing all the relevant project technologies, we released on the project website the interface used by the northbound service layer to provide the Autonomous Network Slice Management functionality, which is based on the AI for Zero-Touch network slicing capability described in Section 3.3.3.

Also, for the interested practitioners, all the deep learning models we used for the PoC are also released as Open Source, and are available on Github[11].

3.5.5.3 Open Northbound API for the multicast service

3GPP MCC has set up a Git repository on a web-hosted platform called 3GPP Forge. The repository element is based on a private instance of GitLab. This provides a fancy front-end for exploring Git source code repositories.

- As a member of the public, I can browse the 3GPP Forge repositories without needing to log in. I can clone any branch (e.g., the Rel-16 branch) of a repository (e.g., the 5G API repository) and start developing code against the interface definitions I have downloaded.
- As a 3GPP participant, I can log in the 3GPP Forge platform using my ETSI Online credentials and request read/write access to a repository. I can then create new branches of the repository (e.g., to develop a Change Request), commit changes to my branch and then push the commits on my branch back to 3GPP Forge. I can also request that my branch is merged into a superior branch as as result of a Change Request approval.
- (The traditional form-based Change Request process operates in parallel with this to make updates to the paper specification. Where the paper specification carries a copy of some Forge-hosted asset, such as an OxpenAPI interface definition file, the paper specification is considered definitive. In this case,

[9] https://scm.atosresearch.eu/ari/pub-osm/pub-osm-ai-agent

[10] https://ai-agents-for-osm.readthedocs.io/en/latest

[11] https://github.com/wnlUc3m/AZTEC

the Change Request form sent to TSG Plenary meeting for approval references a commit ID or tag in the 3GPP Forge repository and the MCC Technical Officer is responsible for ensuring that the two agree.)

This represents the default branch of the repository (also known as the "master" branch) and includes approved changes contributing to the current 3GPP release (Release 17 at the time of writing).

Older releases are represented in the repository by a different branch. For example, the current snapshot of Release 16 (including all TSG-approved essential Change Requests to Release 16 specifications after the API freeze date) can be found at 3GPP site[12].

The work has been initiated by the work Item MBUSA [45] and should be finalised in 2022.

3.5.5.4 AI-enhanced MANO

AI-enhanced MANO can automate the deployment of a new critical service (e.g. Airport Evacuation use case) in a secure and robust way to achieve the use case specific requirements and optimize resource usage and services downtime due to network or general infrastructure problems. For the deployment of a new critical service the steps of AI-enhanced MANO are:

- Critical service requests resources from MEC;
- Various metrics, KPIs and SLAs are gathered;
- AI-enhanced MANO algorithm decides and performs all deployments and migration actions.

AI-Mano REST APIs using the OpenAPI V3 standard is located at Wings public repository[13].

The specification includes:

- Available endpoints (e.g. /kpis) and operations on each endpoint (e.g. GET /kpis, POST /kpis, DELETE /kpis, etc.);
- Operation parameters Input (e.g. kpiID, Identifier of KPI);
- Responses and output for each operation (with samples);
- Schemas, description and example values;
- Authentication methods.

[12] https://forge.3gpp.org/rep/all/5G_APIs/tree/REL-16

[13] https://ai-mano.5gtours.wings-ict-solutions.dev/

4 NETWORK INFRASTRUCTURE & DEPLOYMENT

4.1 TOURISTIC CITY DEPLOYMENT UPDATE

The overall physical architecture of the Turin site differs between Phase 1 and Phase 2. Phase 1 was described in D3.3 [3] while at this stage of the project Phase 2 in ongoing and some goals are already achieved and described in the next paragraph.

4.1.1 Deployment of Physical Infrastructure Phase 2

The phase 2 deployment consist of two network architecture instantiations:

- In-the-field network solution, where the 5G indoor RAN coverages is connected to the TIM commercial core network. In terms of museum infrastructure integration, this can be considered as an evolution of Phase 1.
- The TIM laboratory Network, using the end-to-end 5G NSA network solution provided by 5G EVE [46].

Palazzo Madama

The phase 2 in-the-field final technical solution and positioning of the indoor equipment is now in place in Palazzo Madama. It is an innovative and "zero impact" indoor 5G radio solution within a museum, the first in Italy. To install radio equipment inside a historic structure protected by the Superintendency of Archaeology, Fine Arts and Landscape, required a considerable innovative effort in terms of design to cope with the intrinsic complexities of the environment in which the solution would operate. For the insertion of the technological components within the museum context, especially in the courtly rooms of Palazzo Madama, a UNESCO heritage site, a not impacting extremely flexible radio solution without supports or wiring on the wall and ceiling has been designed mainly using the existing museum furnishing for the radio installation. Furthermore, this kind of solution allowed the usage of the same technical solution adopted for Phase 1.

The new Radio positioning were approved with some modification compared with the previous one; Figure 48 shows the new positioning at ground and first floor:

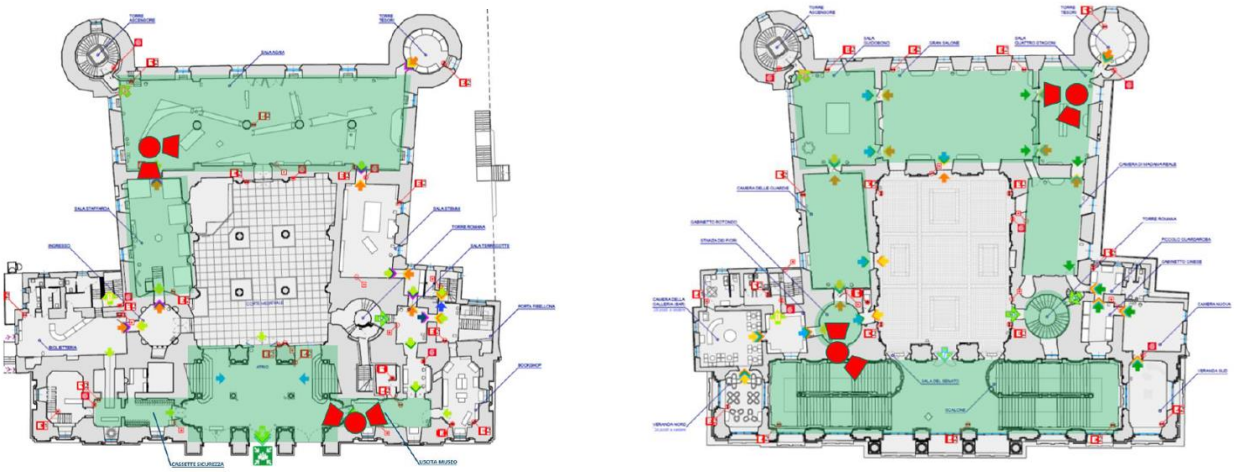


Figure 48. Radio positioning at Palazzo Madama ground and first floor.

From a technical perspective the same performances of Phase1 were granted, while from an aesthetical point of view the solution was fully compatible with aulic context of this location.

Figure 49 shows two integrated Radio station examples as they are located at the moment in the museum: the current radio station position allowed the trial of the Use Case 5 and Use Case 4b.

Figure 49. Installation in Palazzo Madama Phase 2.

After Use Case 5 completion, a further modification has been deployed in December 2021: to respond to the requirement coming from Use Case 3 at second floor, the installation in Sala Acaja has been moved to the second floor in Sala Ceramiche. This solution has been already approved by the Cultural heritage officer and foresee the positioning of the radio and antennas in the external walkway of the building (see Figure 50).

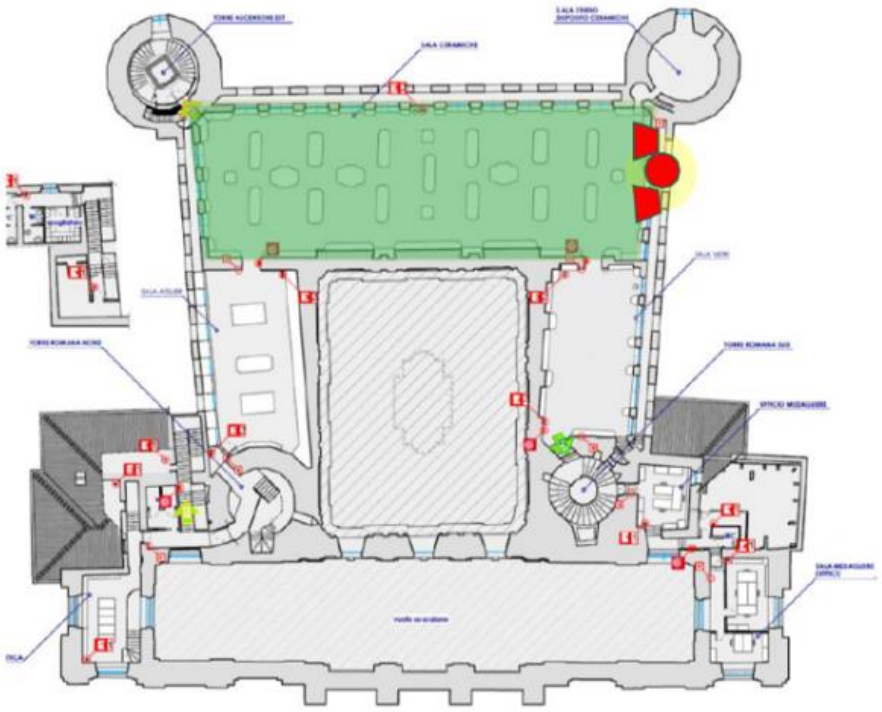


Figure 50. Radio positioning at Palazzo Madama second floor.

GAM

GAM solution is being examined by the Superintendency for Cultural Heritage in order to define how to insert the radio installation in the modern art museum. The identified solution is based on the Ericsson DOT technologies. To be more precise the models that will be used in the museum will be composed by the IRU8846, Indoor Radio Unit, and eight 5G DOT 4479 (N78) located along the museum corridors. The IRU will be connected to the same baseband BB6630 used for Palazzo Madama inserted in the TIM Network. Figure 51 shows the DOT positioning in the museum:

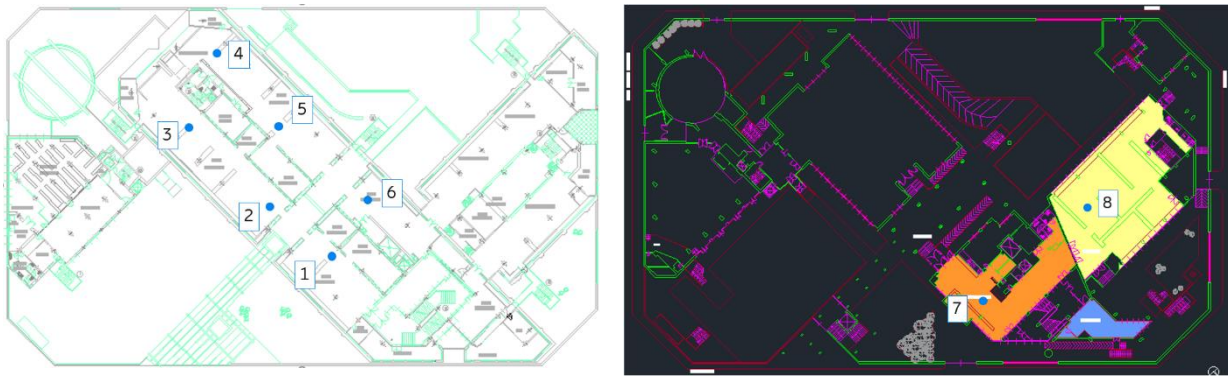


Figure 51. Radio positioning at GAM.

From an aesthetical solution perspective, the context of the modern museum allows to adopt a non-masked solution. Figure 52 represent a simulated view of the installation:

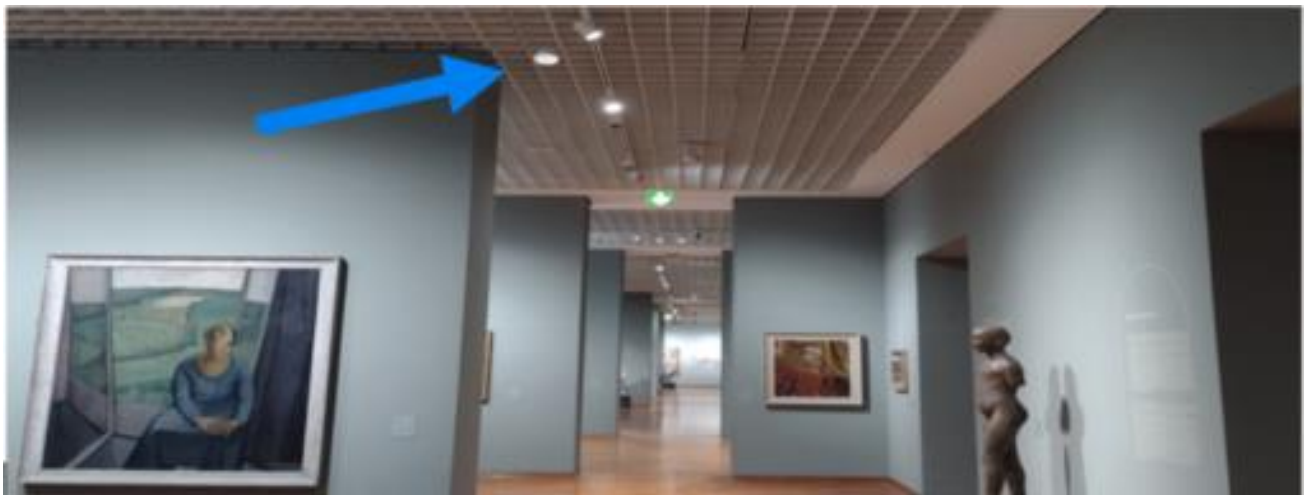


Figure 52. DOT Radio installation simulation at GAM.

4.1.2 Network Equipment

TIM Lab

The phase 2 laboratory infrastructure consists of a 5G RAN (with NSA nodes), fronthaul and backhaul and open-source tools for the management and orchestration of the NFV infrastructure.

RAN infrastructure is providing an indoor coverage in shielded rooms. It is composed by Baseband 5216 for 4G LTE service connected to Radio 2217 (B7 2600 GHz), while Baseband 6630 for 5G service, is connected to the AIR6488 5G NR B42 (3.5 – 3.6 GHz).

Transport infrastructure is based on Router 6672 for X2 connection between baseband interface toward Core Network.

The CORE solution adopted is using the 5G EVE setup based on the virtual EPC installation, and it is currently providing the same set of features because 5G-TOURS use cases do not require specific functionalities.

Virtual EPC in a box solution means that all the core components are virtualized (vMME, vEPG, vHSS-FE, vCUDB), and contained in a Virtual Machine as VNF. Cloud environment hosts the entire system and manages the cross functions towards RAN components.

As described in 5G EVE deliverable D2.3 [26], this in-lab network is a full mirror of the field network architecture. By a dedicated Private APN, it provides to the 5G TOURS experiments a controlled and ideal environment for execution.

Table 5 lists the installed network capabilities.

Table 5. TIM Lab Core Capabilities.

Capabilities	Features
5G Services	Enhanced MBB (eMBB)
	URLLC (URLLC) Rel-15
	Massive IoT (mMTC)
5G Architecture Options	Rel15-GNR + EPC in NSA mode
5G Access Features	Flexible Numerology
	Massive MIMO
	Multi-User MIMO
	Latency Reduction Rel-15
Core Network	vEPC supporting 5G
	Interworking with LTE
Slicing	Network Slicing (std 5G Services: eMBB, URLLC, mMTC)
	Service Slicing (cloud orchestration level)
	Multi-site Slicing
Virtualization	NFVi support
	SDN control
	Vertical Virtualized Application deployment support
Orchestration	VNF, CNF, PNF

Two different orchestrators are installed in the Italian facility:

- the MANO Orchestrator (OSM) is located in the "Politecnico di Torino";
- the RAN Orchestrator (EVER) is located at TIM-labs premises.

The Radio and Core EVER orchestrator is available in the Trial Core environment and is provided by Ericsson. The Figure 53 shows the building blocks namely Resource abstraction and advertisement, Orchestration of virtual resources, Instantiation of VNFs or VNs to deploy network services, Management of physical mobile transport network, computing and storage infrastructure, Selection and configuration of transport and radio.

The interested reader can find in [26] the detailed definition of each block detailed meaning.

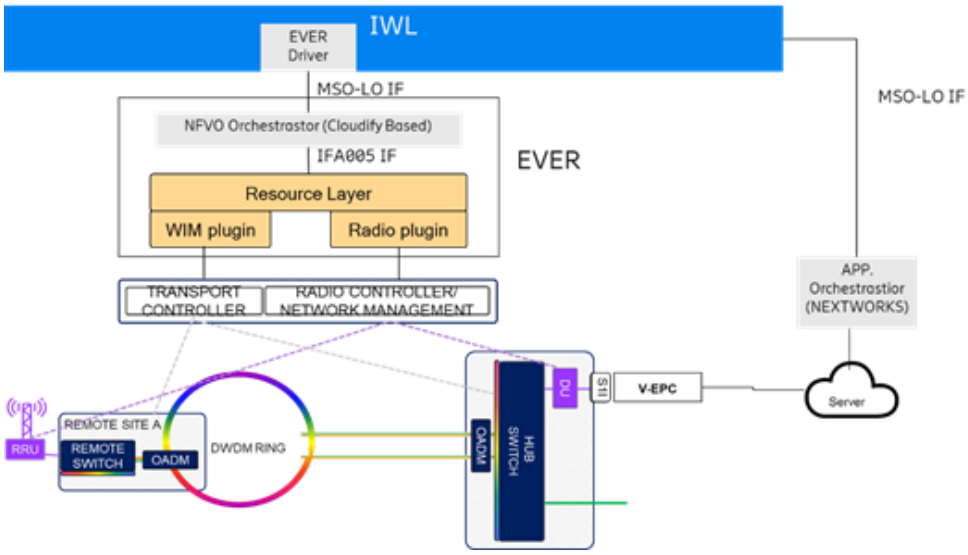


Figure 53. EVER building blocks [34].

To get a complete view of the Network across the Project phases, Figure 54 shows the overall Network Infrastructure.

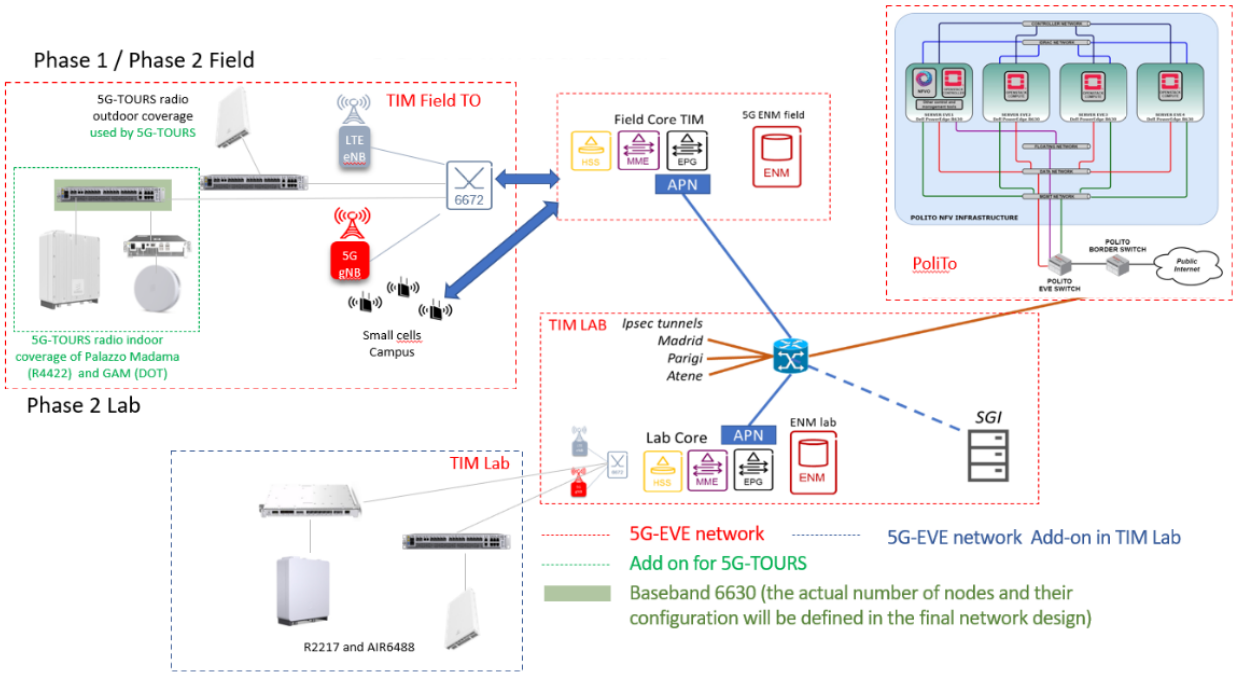


Figure 54. Overall Turin Network infrastructure.

4.2 SAFE CITY DEPLOYMENT UPDATE

The Safe City use cases will be trialled in 2 locations:

1. **Rennes**, using the mobile network infrastructure of Orange and Nokia at BCOM's and CHU premises. This is applicable to use cases 7 and 8.
2. **Athens**, using the mobile network infrastructure deployed in the WP6 at OTE premises (see Section 4.3). This location will host use cases 6 and 9 which use a standard eMBB slice provided by the WP6 node.

4.2.1 Deployment of Physical Infrastructure

Two mobile network deployments are ongoing for 5G NR:

1. Outdoors: at the BCOM premises, for the connected ambulance, as shown in Figure 55. The Nokia 5G NR antenna is now installed on the roof of the BCOM building, using primarily the 26 GHz frequency band. An important integration work aiming to run the proprietary manufacturer-based antenna (Nokia) took place using an opensource-based core network from BCOM for 4G and 5G. The UC7, the targeted usage for this setup, confirmed successful first testing.
2. Indoors: at the Wireless Operating Room at CHU Rennes to provide high-speed, low-latency, millimetre-wave wireless access for medical imaging equipment, using 26 GHz for data transmission and 2.6 GHz as the anchor frequency band, see Figure 56. For this location also, the Nokia antenna has been selected. The current planning indicates that the necessary work to install the antenna take place as soon as the sanitary situation allows for the engineers (not essential staff) to enter the hospital for the needs of the deployment.

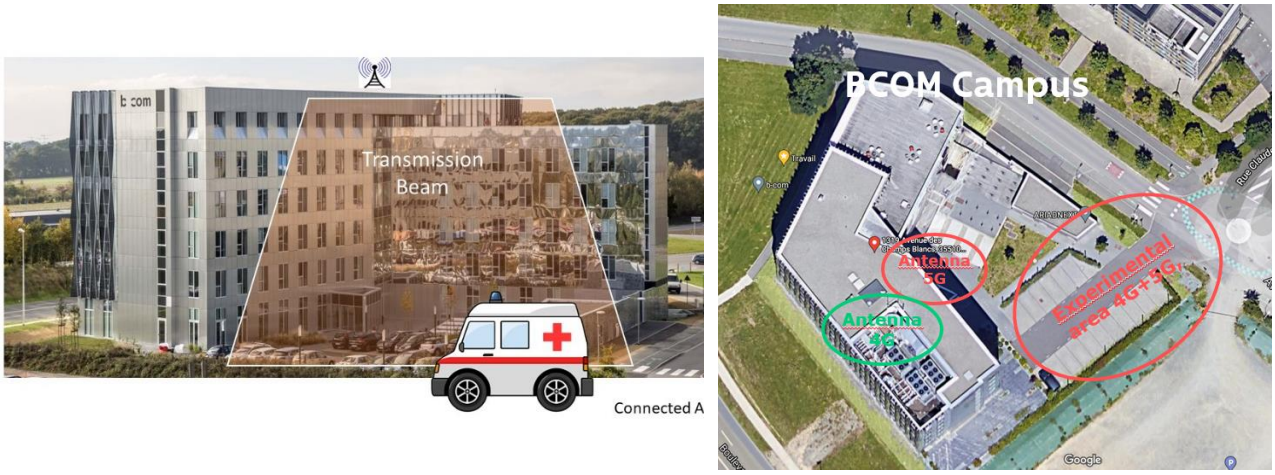


Figure 55. 5G-TOURS 5G NR NSA wireless coverage at BCOM.

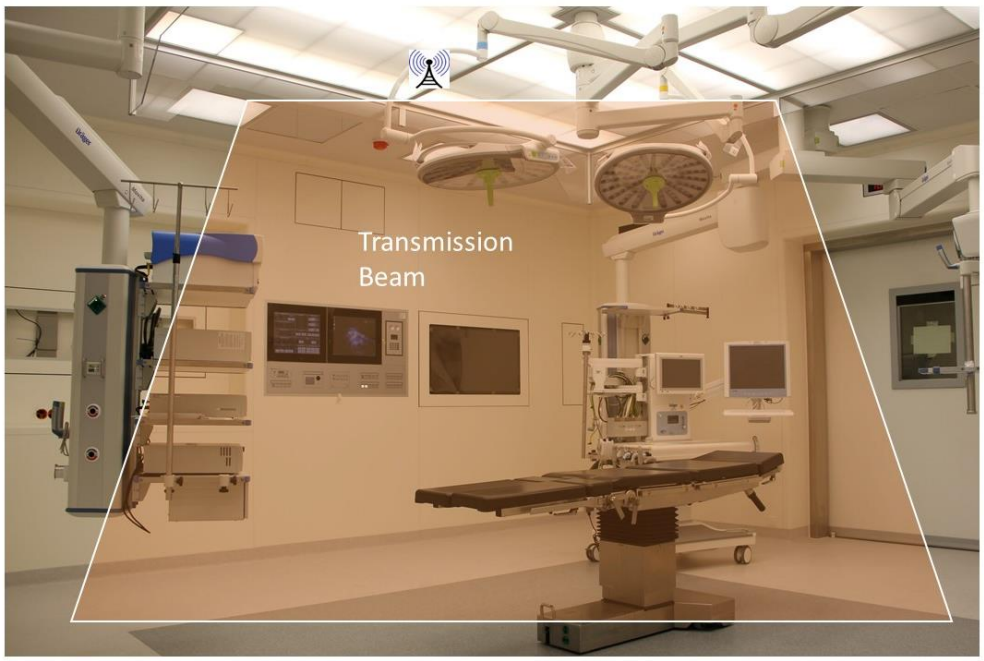


Figure 56. 5G TOURS 5G NR wireless coverage in the Wireless Operating Room at CHU.

At the BCOM premises, the 5G base station with a local virtual UPF, part of the so-called "Wireless Edge Factory" (WEF) is currently being integrated and tested. Similarly, there will be a WEF UPF at the hospital that connects to the WEF core network hosted in the BCOM datacentre through a dedicated VPN backbone. This is depicted in Figure 57. This will enable the setting of end-to-end network performance KPIs and the prioritization of data traffic between the ambulance and the hospital to guarantee the required quality of service. Furthermore, the WEF Core Network deployed in BCOM datacenter will manage the WEF UPF at the hospital to connect the 5G terminals of the Wireless Operating Room.

In addition, for the non-critical overall network orchestration and automatic deployment of the WEF core network, Orange provides an ONAP orchestrator in their Châtillon datacenter as part of their 5G EVE infrastructure. ONAP enables the user or the experimenter to deploy and configure the WEF Core Network on demand. It could also be used to deploy the user plane part of the WEF.

The Orange datacenter has already been connected to the BCOM datacenter in the scope of the 5G EVE project. This is also shown in Figure 57.

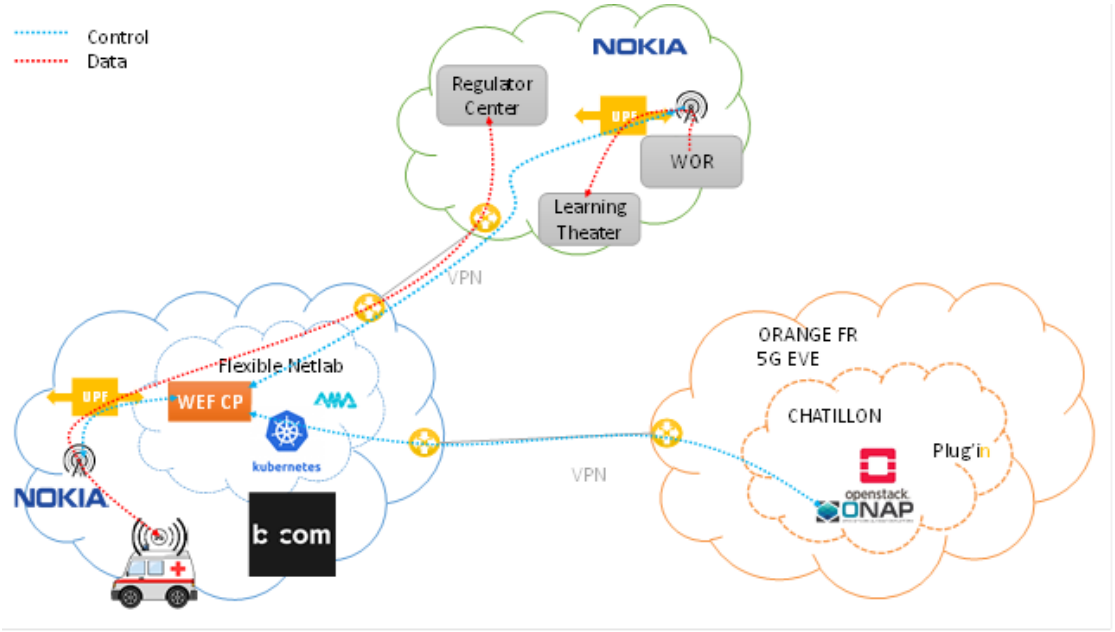


Figure 57. Overall network architecture and physical deployment of network equipment and functions.

4.2.2 Network Equipment

Control plane network equipment

The control plane is a virtual 4G/5G Core Network compatible with the 5G NSA standard (3GPP Rel-15). The Control Plane is part of the WEF solution developed by BCOM [47]. It is deployed as a set of Docker containers managed by a Kubernetes cluster. This cluster is hosted on the Flexible Netlab platform in the BCOM datacenter [48]. The Control Plane is deployed and orchestrated by an instance of the ONAP orchestrator hosted by Orange.

User plane network equipment

The user plane equipment provides connectivity between the RAN equipment and the data network (Internet). The main component is the User Plane Function (UPF) component of the WEF provided by BCOM. Two instances of the UPF will be deployed as part of 5G-TOURS.

The first instance will be a VNF i.e., a purely virtual UPF deployed in BCOM datacentre as a virtual machine hosted on an OpenStack cluster provided by Flexible Netlab. This virtual machine hosts an OpenVSwitch (OVS) virtual switch that acts as a tunnel endpoint for the GTP tunnels coming from the RAN equipment deployed at BCOM for use case 7 (UC7). It is thus used to connect this RAN equipment to the Rennes CHU through the VPN. The WEF Control Plane manages the virtual switch under control of the OpenDaylight SDN controller that is deployed in the control plane. The second instance is a PNF i.e., an appliance built from a COTS network switch and a COTS 1U server. The server is a KVM hypervisor that hosts an OVS-based virtual machine similar to the one deployed in Flexible Netlab. It will be installed in the technical room of the Rennes CHU and will interconnect the RAN equipment deployed there with the various components required by use case 8 (UC8). The same WEF Control Plane will manage this switch through the VPN established between BCOM and the CHU.

RAN equipment

For 5G-TOURS, the Nokia Small Cell technology is the RAN equipment chosen. Two small cells will be deployed: one at the Rennes CHU to provide coverage for the Wireless Operating Room operated by Nokia and one at BCOM premises to cover the outside area for UC7, operated by BCOM. Both will use the 26GHz/2.6GHz bands in 5G NSA mode. Both deployments will be combining the Nokia RAN with BCOM Core Network.

Initial testing and integration was conducted for the first integration tests of UC8 in the BCOM showroom using Amarisoft Classic Callbox RAN equipment. This equipment uses the 3.5GHz band and is also compatible with 5G Non Stand Alone (NSA) mode. At this point, we do not plan to move to a 5G Stand Alone (SA) network

within the duration of the project. For this initial testing, all medical equipment that requires 5G wireless connectivity has been connected to this RAN equipment through compatible CPEs.

Further tests are currently carried out with the NOKIA 5G RAN using the required frequencies (n257 band for 5G and B38/41 band for 4G which are allowed by ARCEP, the French regulator), and the BCOM WEF Core network so that the two uses cases can rely on a complete E2E 5G NSA network. Such integration will allow the validation of the 5G NSA network which must be deployed in both sites, at the ThérA-Images room of the CHU Rennes and at BCOM parking. Using the same type of BBU (Base band unit) on each site connected to one Core network deployed in BCOM datacenter, an indoor antenna will be deployed in the CHU operating room and operated by Nokia while an outdoor antenna will be deployed on BCOM building and operated by BCOM to cover its parking and the ambulance.

For the final tests, all medical equipment such as the ultrasound probes and smart glasses will be connected via a specific Android smartphone - Asus 5G Phone on the mmWave spectrum [56] built on top of a Qualcomm® Snapdragon™ 888 5G Mobile Platform and using a B41+n257 frequency (2.6GHz + 26 GHz) on the Nokia mmWave RAN attached the BCOM WEF core network [56].

RAN solution is composed by RRH (Remote Radio Head) and BBU are depicted in Figure 58.

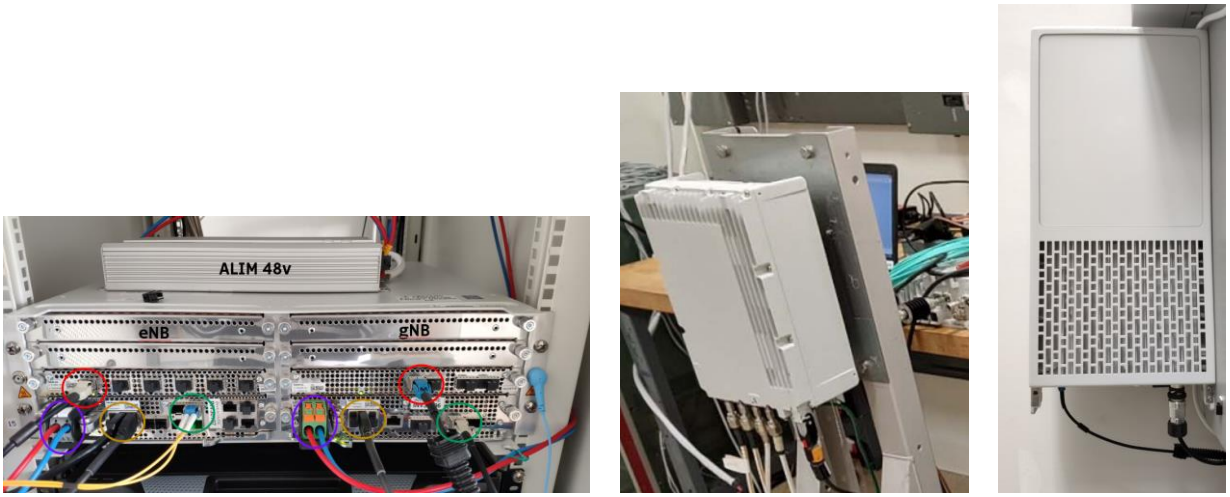


Figure 58. BBU, RRH 4G and RRH 5G under integration phase in BCOM labs.

Integration with the 5G EVE project

The integration of 5G-TOURS with 5G EVE is achieved as depicted in Figure 59. The Service Layer interacts with the 5G EVE Portal through a programmable REST API to request the deployment and instantiation of the whole vertical service by the 5G EVE platform.

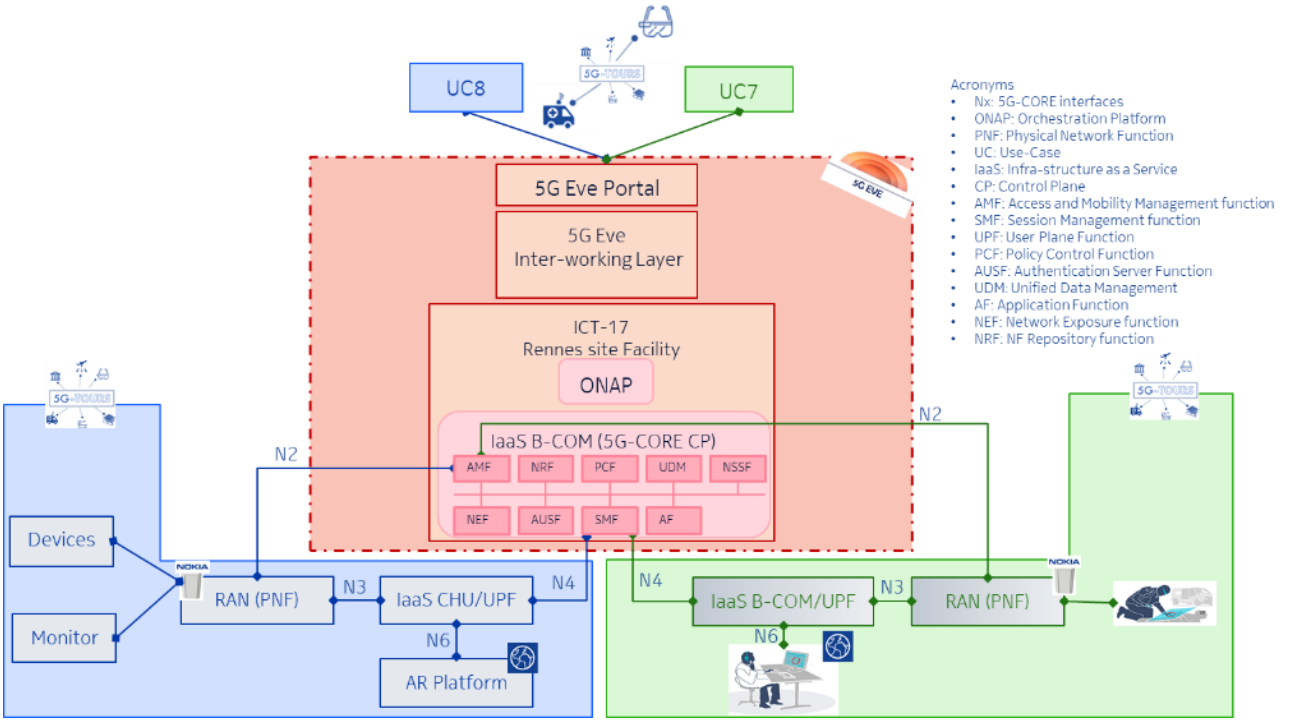


Figure 59. 5G-TOURS integration with 5G EVE in Rennes.

The 5G EVE Portal API enables a programmable interaction between 5G-TOURS and 5G EVE at the portal level. Such API documentation is available in 5G EVE D4.2, which includes the general description and the functionalities of the first version of the portal, and in 5G EVE D4.3, which includes the functionality extensions made to the first version [49], [50]. The 5G EVE Portal API supports experiment lifecycle management operations (e.g., instantiation, termination, polling status, etc.), whilst all the experiment design operations are available only through the 5G EVE Portal GUI. This means that a pre-liminary offline step will be needed through the 5G EVE Portal GUI to create blueprints and descriptors for the experiments associated to the vertical service in 5G EVE platform.

In addition, it is assumed that a pre-provisioning of connectivity between 5G-TOURS and 5G EVE sites is already in place through a secure VPN.

The integration relies on the interworking capabilities of the 5G EVE platform for handling multi-site services and experiments. Following this concept, the coordination of the provisioning of the end-to-end service is entirely delegated to the 5G EVE platform.

The first step is to define the vertical service and its subcomponents and onboard the related blueprints on the 5G EVE platform, using the 5G EVE Portal GUI.

As depicted in Figure 59, the 5G CORE control plane will be part of 5G EVE infrastructure. The 5G CORE user plane named UPF will be instantiated in the EDGE node deployed in CHU Rennes and in the BCOM datacentre. Table 6 lists the prerequisites for the UPF execution environment named IaaS BCOM/UPF.

Table 6. Prerequisites for IaaS BCOM/UPF.

Operating System	Ubuntu 16.04
CPU	1 vCPU, RAM: 512 MB, Network Interfaces: 4
Management	1 (For administration purpose) • SDN-MNGT: 1 (For the SMF-SDNC to manage the UPF) • GTP interface: 1 (To assure the connection in between the RAN and the UPF) • WAN interface: 1 (To provide access to the Data Network)

The UPF execution environment is deployed as VMs using a KVM hypervisor and OpenStack as IaaS manager, see Figure 60. OpenStack provides an API to manage the provisioning and deployment of the VMs as well as its network configuration. It is compatible with the ONAP orchestrator used in the 5G EVE project.

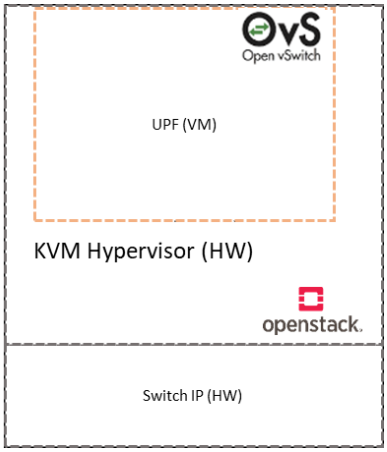


Figure 60. UPF deployment over OpenStack execution environment.

In terms of monitoring, the 5G EVE platform will be responsible of providing the collection and visualization functionalities for the monitoring of data of the entire vertical service, provided that the VNFs developed by 5G-TOURS support the required extensions to publish monitoring data into the 5G EVE monitoring platform. The 5G EVE platform supports the visualization of monitoring data through the 5G EVE portal GUI and provides internal functionalities for performance validation and evaluation based on KPIs.

4.3 MOBILITY-EFFICIENT CITY DEPLOYMENT UPDATE

The Athens node of Mobility efficient city consists of the 5G EVE Greek Site infrastructure and the extended site at the Athens International Airport. The overall infrastructure is already completed during this 2nd phase of the project and is depicted in the section below.

4.3.1 Deployment of Physical Infrastructure

The mobility-efficient city presents a set of use cases that improve the tourism and tourism-related experiences from various perspectives. During the 2nd phase of the 5G-TOURS, all the infrastructure components of the Mobility-Efficient City Athens node are already installed and completed. These components are separated into two geographical areas (described in D3.3 [3]):

- 5G EVE Greek site infrastructure installed at the OTE's facilities at Psalidi area;
- Extended AIA site infrastructure at the AIAs facilities.

5G EVE Greek Site final infrastructure at OTE's premises

The installed components of the final Greek site are the following:

- OTE's Lab network/data center infrastructure;
- a NOKIA 5G NSA platform is installed and fully functional;
- a NOKIA 5G SA platform is installed and fully functional;
- Two pairs of 5G RANs;
- an OSM interconnected with NOKIA's platform;
- Interconnection of Greek site with IWL in Turin;
- Interconnection of Greek site with portal in Spain;
- Interconnection of the Greek site data plane with the corresponding data plane of Rennes for the needs of running Multisite UCs.
- a Kafka Broker for collecting KPI measurements, interconnected with central Kafka;

- an AI-Enhanced MANO / Diagnostic Tools for post-process analytics on the collected metrics and KPIs for dynamically allocate resources to the running apps;
- Clouds serving the needs of UC's applications;
- VMs for Content Servers for serving the back-end content needs of applications;
- Data shippers from NOKIA core for shipping the network metrics to the Kafka;
- Probe Server farm and probes for real time KPI measurements;
- TWAMP/VIAVI software for Latency, Throughput, Jitter, packet loss measurements of different segments of the E2E Network;
- Blueprints for UC10 and UC12;

Extended final infrastructure at Athens International Airport

For the needs of 5G-TOURS WP6 UCs implementation an extended site was installed at the AIA premises, which is fully interconnected with the 5G EVE infrastructure at OTE's premises. The four UCs of WP6 are implemented in this extended site utilizing the 5G Nokia's NSA core of 5G EVE. This extended AIA implementation comprises of the following components:

- 2 outdoor NOKIA's antennas designed and installed, that utilise the spectrum band of 3450-3500MHz, 50 MHz bandwidth are up and running;
- 4 indoor NOKIA's antennas designed and installed, that utilise the spectrum band of 3450-3500MHz, 40 MHz bandwidth are up and running;
- All the antennas are connected via optical fibre with the NOKIA's BBUs at the airport;
- OTE L2/L3 OSN switch is interconnected to the OTE IP Core, using a 10 Gbps capacity line, that interconnects 5G EVE Greek site infrastructure located at OTE Labs in Psalidi-Attika with AIA;
- E2E interconnection with 5G EVE infrastructure has been configured and tested successfully;
- Smart devices are used and specific innovative applications developed for the implementation of the 4 UCs (Smartphones, tablets, AR/VR headsets, IP-cameras, IoT sensors, etc.);
- For UC10, 55 parking sensors have been installed at AIA;
- For UC11, a streaming server has been installed at the AIA place and is interconnected on the OSN. This server steams the 4K video of the Van UHD cameras;
- For UC12 and UC13 4 indoor antennas have been installed into 2 different AIA buildings, that have been tested E2E;
- Also, KPI measuring probes have been installed between antennas and the BBUs, as well as between the BBU and OTE's Core-switch (at AIA);
- TWAMP/VIAVI software for Latency, Throughput, Jitter, packet loss measurements of different segments of the E2E Network;
- 4 UCs run at the Athens node that belong to the WP6: Smart Parking, Ground-based vehicles, Airport Evacuation, and AR/VR student bus excursion;
- Moreover, two more UCs belonging to WP5 will be hosted and run at the Athens node: Remote Health Monitoring and Optimal Ambulance Routing;

All the above infrastructure is depicted in the following figure:

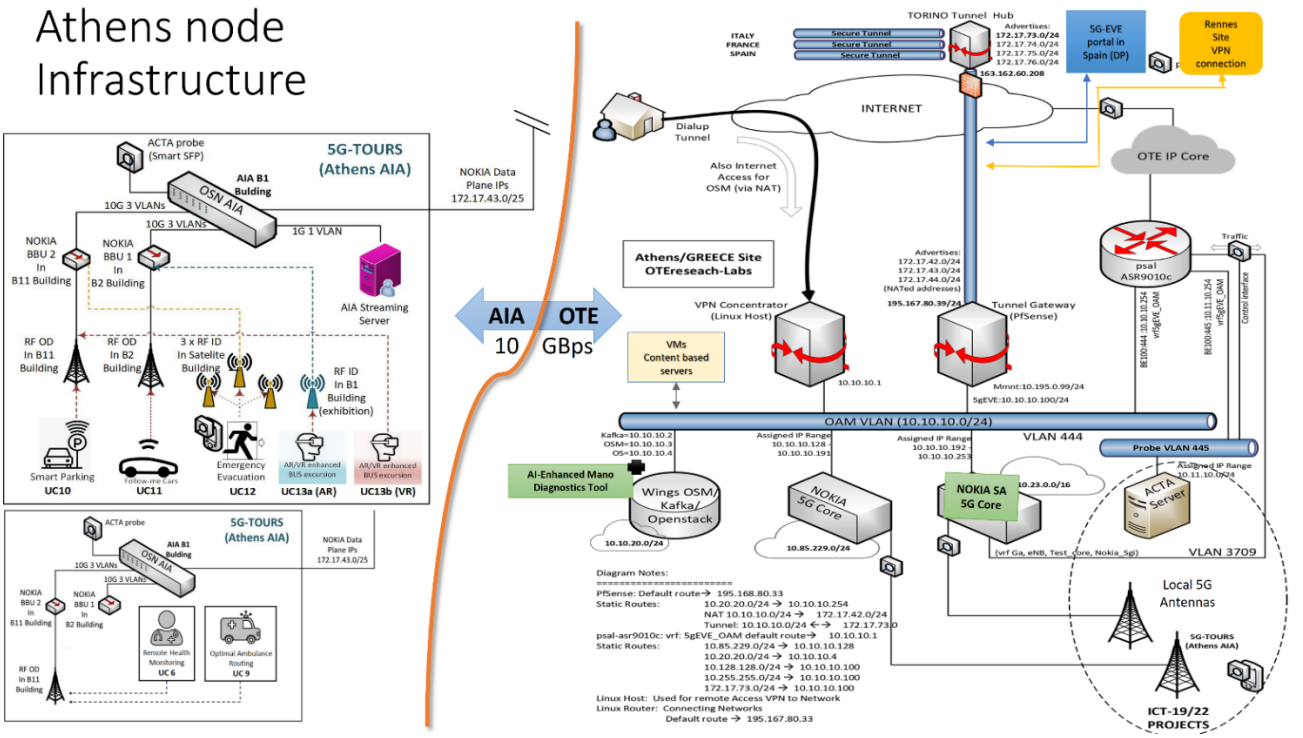


Figure 61. Final Athens node infrastructure for the needs of 5G-TOURS.

The following figures depict the outdoor radio equipment installed at AIA to be used as part of the Use Cases demonstration, as well as probes and analytics server.

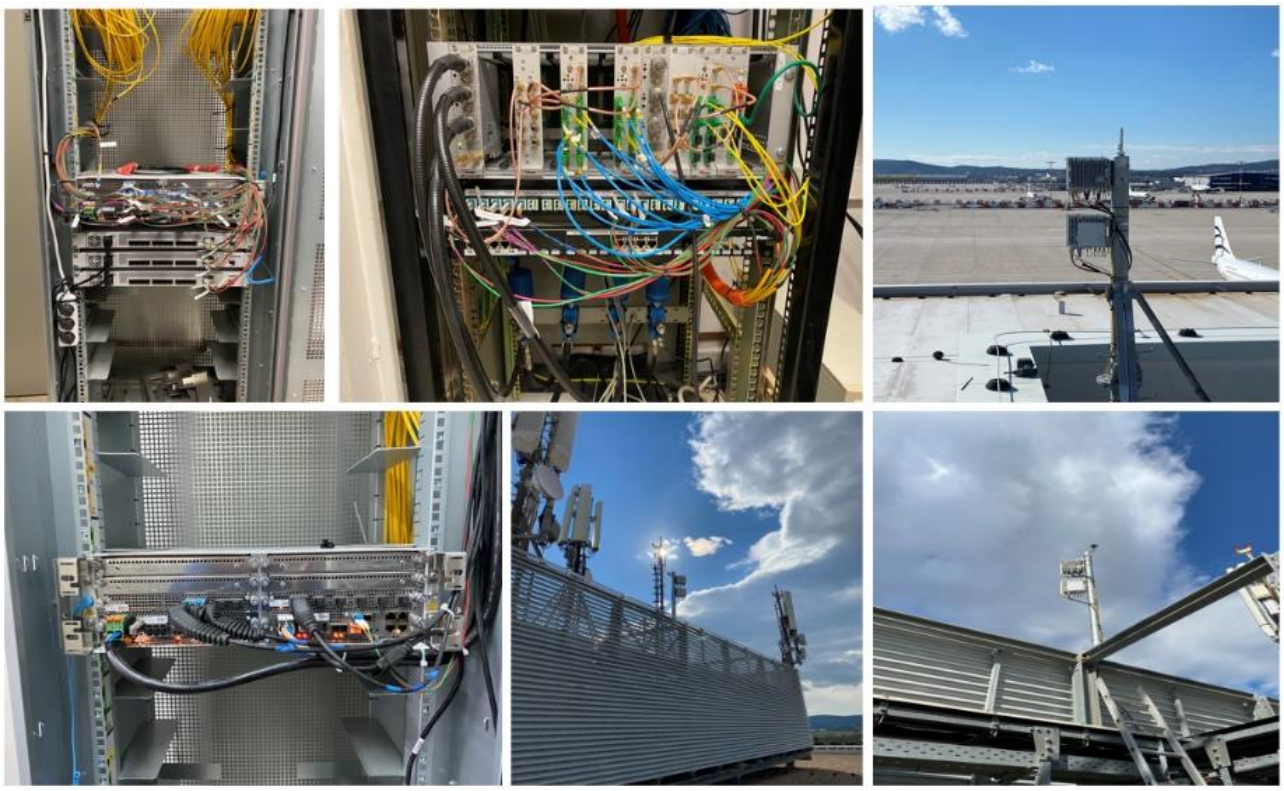


Figure 62. Outdoor Radio equipment installation at AIA.

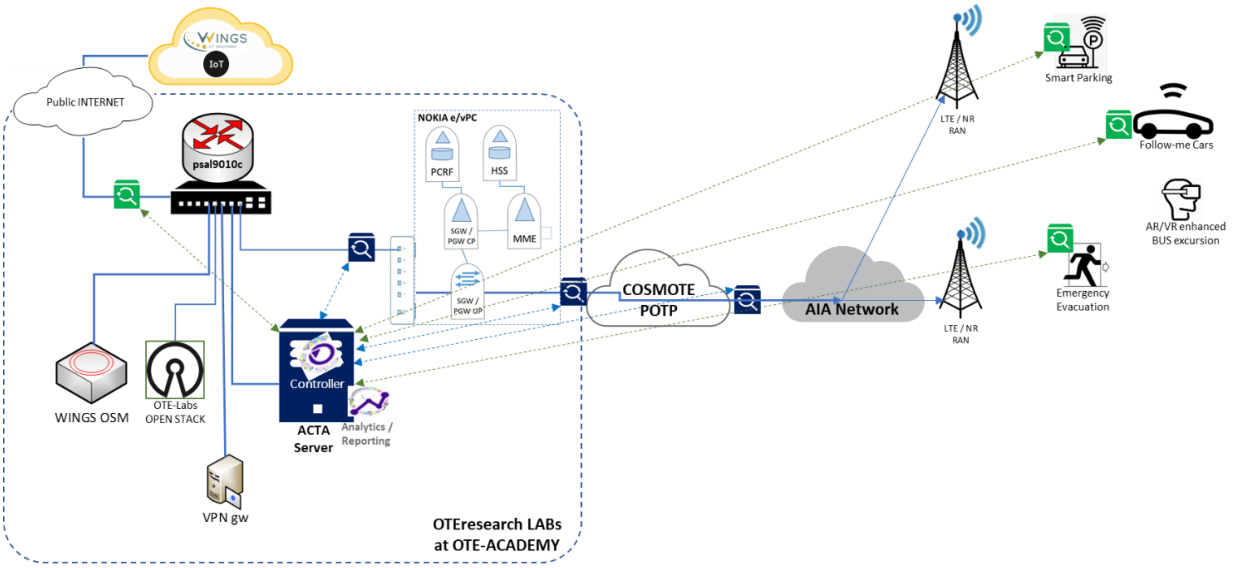


Figure 63. ACTA Probes and Server topology in Athens node.

Figure 64. Pictures of actual probes and ACTA's server (KMVaP) at OTE Labs.

The Athens node network infrastructure deployment is based on a two phases' implementation approach in order to meet the coverages, the performances and Quality of Service requirements of the different use cases, their development and trials roadmaps, as well as the integration with the 5G EVE platform/infrastructure. The main aspects of the two implementation phases are summarized hereafter.

As part of phase 1 deployment, a vEPC Core Network is deployed consisting of MME, SPGW, HSS, PCRF offering 4G with Option 3X Radio and Core Network capabilities. Regarding phase 2 deployment, a 5GC Core Network is deployed consisting of AMF, SMF/UPF, UDM, NRF/NSSF offering 5G Standalone Radio and Core Network capabilities.

The same Radio hardware is used as for both phase 1 and phase 2, given that the latter can be configured to work either as Option 3X or as 5G Standalone RAN solution. The network implementation of phase 2 will validate the need of employing the 5G technology and demonstrate the benefits of 5G-TOURS innovations in terms of enhanced network functionalities and capabilities.

Core 5G-TOURS network solution at AIA and Psalidi (OTE's Labs)

In line with the phased approach described above, Nokia GR and OTE started to work on the network solution to provide the Radio Coverage (LTE and/or NR) at Athens International Airport (AIA).

The 5G-TOURS Core Network reuses the 5G EVE Core Network deployment for both Phase 1 and Phase 2 of the project, as described above.

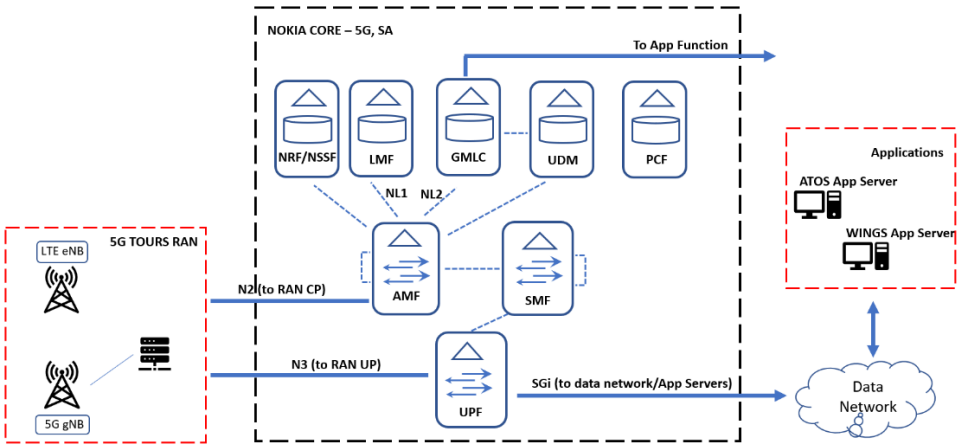


Figure 65. High level view of 5G-TOURS RAN and CORE Network infrastructure at Greek Node.

The radio access network is deployed at AIA site (see below for details on radio deployment locations for all Use Cases and deliverable D6.2 [51] for RAN details) and then connected to either vEPC or 5GC, located at Psalidi area (OTE premises) by dedicated 10Gbit line. The 5G EVE platform is then offered access to both OTE intra-network and public internet so that it may connect with Application Server(s) needed for the Athens Use Cases.

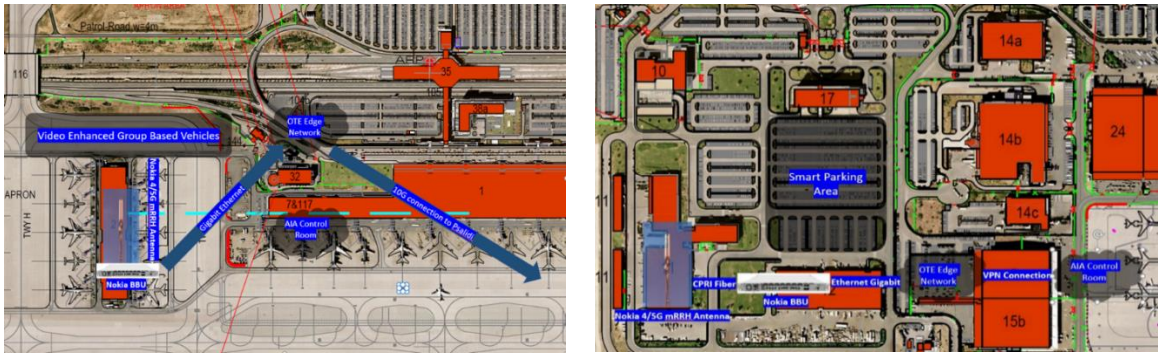


Figure 66. RAN coverage of AIA for Smart Parking UC10 and Video enhanced airfield vehicles UC11.

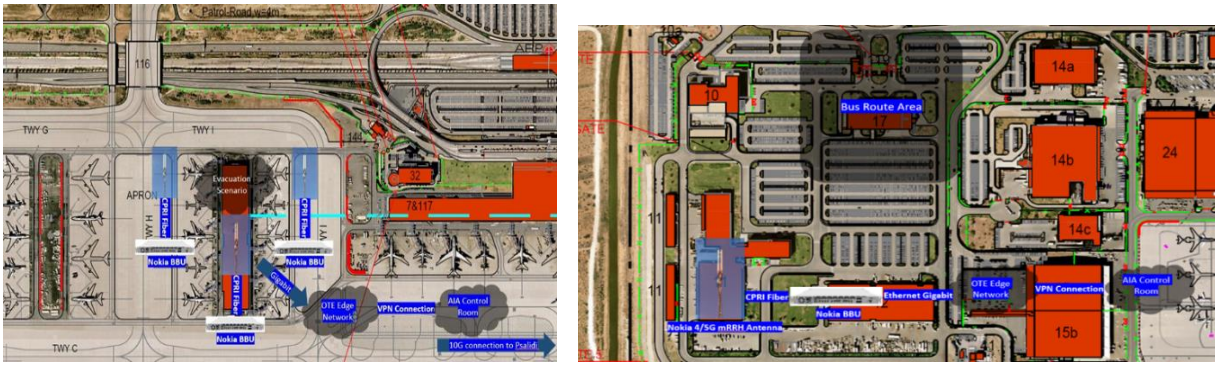


Figure 67. RAN coverage of AIA for Evacuation UC12 and AR/VR Bus Excursion UC13.

4.4 5G EVE Blueprints summary

This section presents some examples of the so-called 5G EVE blueprints used for 5G-TOURS use case integration with 5G EVE platform.

In order to integrate and onboard the network functions related to the different use cases developed by the project we have to design the 5G EVE blueprints [50] that will be used to onboard them on the infrastructure. The 5G EVE blueprints are composed by a set of files that have to be designed to perform the full onboarding operation, as discussed below:

- 5G EVE VSB file: this file provides the high-level description of the Vertical Service, indicating all the VNFs that compose the service and their arrangement.
- 5G EVE NSB file: this file contains the low-level specification of the VNFs matching interfaces and subnet to the high-level VSB.
- 5G EVE TCB file: this file contains the Test Case definition for each experiment. In here, the context conditions such as the background traffic or the commands to enable the data exposure towards the Service Layer are defined.

The following subsections give some typical examples of EVE blueprints and descriptions. The repository for complete descriptions is available on 5G-TOURS website, dedicated subsection of the page http://5gtours.eu/open-source-releases/. Also, UC1 Blueprint v8 and UC10 Blueprint v15 are available on 5G EVE portal.

4.4.1 Turin blueprints

For WP4, blueprints, in this phase, we just compiled the VSB and NSB file for one exemplary use case. Other use cases and their TCBs will be filled in conjunction with the QoE / QoS evaluation methodology and the test cases defined in WP7. We selected Use Case 1 (including all of its sub use cases) for this analysis.

Before entering the details of the blueprint, we depict in below the high-level infrastructure instantiation of the Use Case 1:

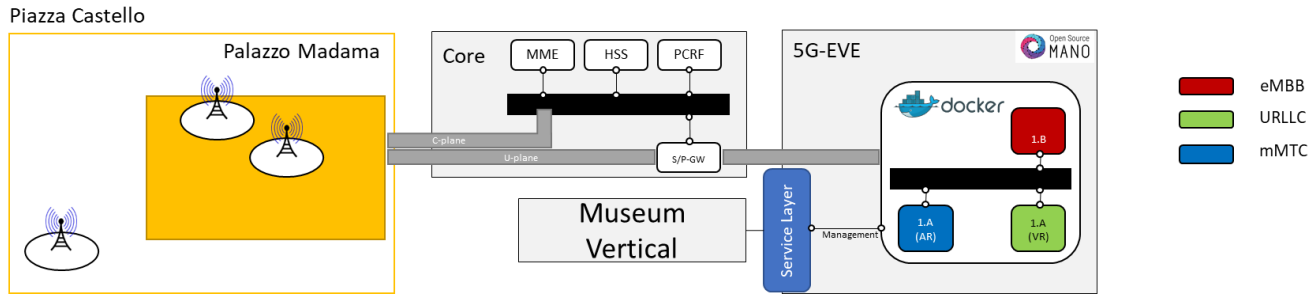


Figure 68. UC1 high-level infrastructure instantiation.

On the left hand side, the Radio Deployment is providing wireless coverage inside and outside the Palazzo Madama building, and it is connected via fiber with the rest of the 5G EVE infrastructure, especially the Core Network available in the TIM Lab. The three slices used for UC1 (eMBB and URLLC for AR and VR respectively, and mMTC for the sensors deployed in the city) are then onboarded using the 5G EVE portal and the Service Layer on the 5G EVE NFVI infrastructure.

In 5G-TOURS we adopted a cloud native architecture, still owing to the implementation choices made by the 5G EVE project, that provides in Turin an NFV Infrastructure based on VM and Orchestrate with the Open Source MANO software. Hence, we host our Container based solutions for vertical applications into a fully-fledged VM containing the VNF. More specifically, three containers are hosted in the VNF:

- The HTTP Server for the AR server, as discussed in IR4.2, will serve the rendered object to the terminal;
- The AR server for the UC1.a;
- The data gathering server for the Internet of things scenarios of UC1.b;

All of them use the docker-compose engine to orchestrate the cloud native part, into a single VNF which is then orchestrated, through the 5G EVE portal. In the following, we briefly discuss the main part of the blueprints we created for UC1.

```
nsdIdentifier: "5G-TOURS-UC1-NSD"
designer: "NSD generator"
version: "1.0"
nsdName: " NSD for 5G-TOURS UC1 VR and AR"
nsdInvariantId: "5G-TOURS-UC1-NSD"
```

This part includes the main signatures of the file, which is used by the 5G EVE portal to identify the Network Slice (or the Vertical Service).

```
parameters:
- parameterId: number_of_user_equipments
  parameterName: Number of user equipments
  parameterType: number
  parameterDescription: Number of User Equipments requesting con-
tent from the AR and VR
  applicabilityField: User Equipments
```

The VSB files offer the possibility to specify parameters associated with the vertical services. In this case, we identify the number of UEs in the network.

```
- componentId: vrar_vm_1
  serversNumber: 1
  endPointsIds:
  - cp_vrar_vm_1_mgmt # ssh, web interafaces
  - cp_vrar_vm_1_data # for the real app
  - sap_vrar_vm_1_mgmt # to the portal / outside
  - sap_vrar_vm_1_data # to the portal / outside
```

The VSB file lists the component in the vertical service. As previously discussed, at VM level, we just have one instance (which contains the docker images providing the services). The component has different endpoints (interfaces) that provide connectivity either to the terminals or the management services and portals.

```
- endPointId: cp_vrar_vm_1_data
  external: true
  management: false
  ranConnection: true
```

Each endpoint has different attributes which specify whether they are used for management (e.g., a web interface or ssh connections), external (i.e., with connection to the Internet), or connected to the RAN (hence to the UEs).

The different low level details of the interfaces are then specified in the NSD blueprint, as follows:

```
sapd:
  - cpdId: "eth1" # from the VNFD
    layerProtocol: "IPV4"
    cpRole: "ROOT"
    addressData:
      - addressType: "IP_ADDRESS"
        iPAddressAssignment: false
        floatingIpActivated: true
        management: true
        iPAddressType: "IPv4"
        numberOfIpAddress: 1
    sapAddressAssignment: false
```

```
    nsVirtualLinkDescId: "vl_vrar_vm_1_mgmt"
```

Here, the low-level details such as the interface name or the number of IP addresses are specified in this section.

4.4.2 Rennes blueprints

The work on the French node 5G EVE blueprints in still ongoing and will be finalized in line with the WP5 timeline. The final blueprints will be uploaded on 5G-TOURS website in 2022, on a dedicated subsection of the page http://5gtours.eu/open-source-releases/.

4.4.3 Athens blueprints

One of the use cases that will be deployed on the Greek site is UC10: Smart airport parking management. A high-level overview of the UC10 can be seen in Figure 69.

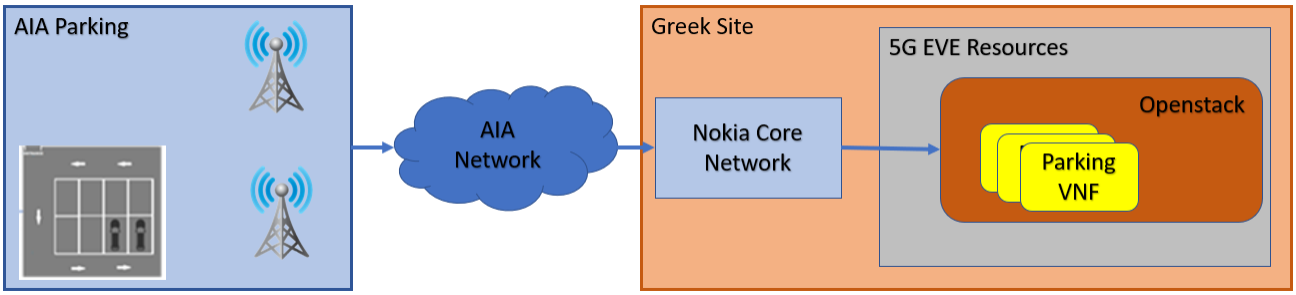


Figure 69. UC10 high level overview.

Smart sensors have been deployed on the parking of the Athens International Airport to monitor the parking spots. These sensors relay information over the 5G network, provided by Nokia, to the UC10 VNFs deployed on the Greek site's 5G EVE resources. These resources are managed and orchestrated using Openstack and OSM deployed by WINGS. The service implemented for UC10 is designed based on the architecture and guidelines of 5G EVE and comprises of two VNFs handling the business logic of the service. These VNFs are connected to both the Control network of 5G EVE, for overall management and monitoring, and the Data network, utilizing the 5G links, for service traffic.

As mentioned above, the the blueprints supported by 5G EVE are the VSB, the NSB and the TCB. The VSB provides the high-level view of the described service, specifically the service identifiers and a list of the components of the service, in this case the two VNFs handling the Routing and Spots Management business logic, as show in Figure 70.

```
{
   "blueprintId":"3df309e0-87ce-42ea-90d1-004eae5a0832",
   "version":"1.0",
   "name":"WINGS TOURS UC10 NS v10",
   "description":"WINGS TOURS UC10 NS with VNFs v10",
   "parameters":[
      {
         "parameterId":"number_of_spaces",
         "parameterName":"Number of parking spaces",
         "parameterType":"number",
         "parameterDescription":"Number of parking spaces",
         "applicabilityField":"num spaces"
      }
   ],
   "atomicComponents":[
      {
         "componentId":"routing",
         "serversNumber":1,
         "endPointsIds":[
            "vnf-routing-mgmt",
            "vnf-routing-data"
         ]
      },
      {
         "componentId":"spots",
         "serversNumber":1,
         "endPointsIds":[
            "vnf-spots-mgmt",
            "vnf-spots-data"
         ]
      }
   ],
```

Figure 70. UC10 VSB overview.

Also, a list of available parameters for this service can be included here, which for this UC is the number of parking spots monitored. Another important part of the VSB is the list of application metrics that can be included in the monitoring workflow for the analysis and diagnosis components to take them into account. An example of that list can be seen in Figure 71.


```
"applicationMetrics":[
    {
        "metricId":"routing_route_calc_time",
        "name":"route calculation time",
        "metricCollectionType":"GAUGE",
        "unit":"ms",
        "interval":3,
        "metricGraphType":"LINE",
        "topic":"/app/routing_route_calc_time"
    },
    {
        "metricId":"routing_sens_stat_req",
        "name":"sensor stats request",
        "metricCollectionType":"GAUGE",
        "unit":"ms",
        "interval":3,
        "metricGraphType":"LINE",
        "topic":"/app/routing_sens_stat_req"
    },
    {
        "metricId":"spots_free_spots_calc_time",
        "name":"free spots calculation time",
        "metricCollectionType":"GAUGE",
        "unit":"ms",
        "interval":3,
        "metricGraphType":"LINE",
        "topic":"/app/spots_free_spots_calc_time"
    },
    {
        "metricId":"FA_Count_F-N",
        "name":"ACTA metric_1",
        "metricCollectionType":"GAUGE",
        "unit":"ms",
        "interval":1,
        "metricGraphType":"LINE",
        "topic":"/app/FA_Count_F-N"
```

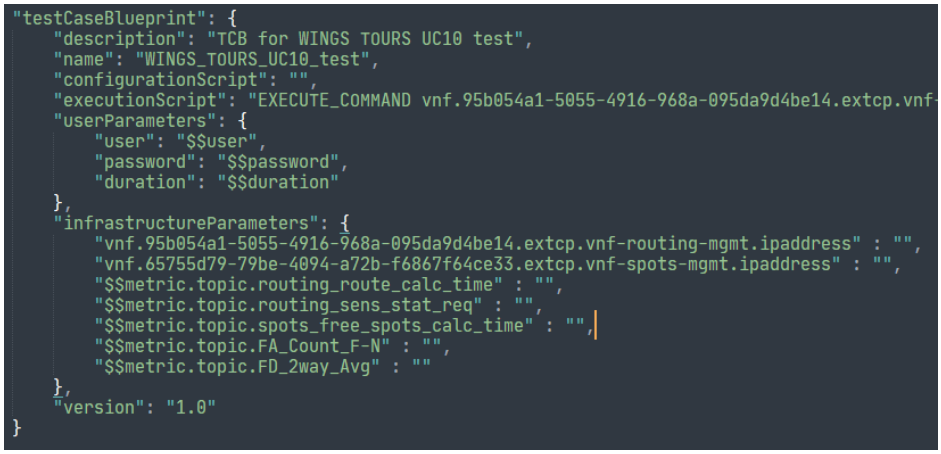

Figure 71. UC10 VSB list of application metrics.

While the VSB is more of an abstract description of the service, the NSB, part of which is shown in Figure 72, is a more direct representation of the deployment of the service and is purpose is to do exactly that, deploy the service, leveraging the orchestration tools provided by the 5G EVE platform. The NSB structure is based on the SOL-005 standard and is a typical network descriptor, containing all information regarding the components, endpoints, and connectivity of the service along with some additional fields that handle the mapping of the descriptor to the other blueprints for managements purposes of the platform.


```
"sapd": [
],
"virtualLinkDesc": [
  {
    "virtualLinkDescId": "mgmtnet",
    "virtualLinkDescProvider": "NSD generator",
    "virtualLinkDescVersion": "1.0",
    "connectivityType": {
      "layerProtocol": "IPV4",
      "flowPattern": ""
    },
    "virtualLinkDf": [
      {
        "flavourId": "mgmtnet_df",
        "serviceAvaibilityLevel": "LEVEL_1"
      }
    ]
  },
  {
    "virtualLinkDescId": "datanet",
    "virtualLinkDescProvider": "NSD generator",
    "virtualLinkDescVersion": "1.0",
    "connectivityType": {
      "layerProtocol": "IPV4",
      "flowPattern": ""
    },
    "virtualLinkDf": [
      {
        "flavourId": "datanet_df",
        "serviceAvaibilityLevel": "LEVEL_1"
      }
    ]
  }
],
"nsDf": [
  {
    "nsDfId": "vsb_wings_tours_uc10_df",
    "flavourKey": "vsb_wings_tours_uc10_df_fk",
    "vnfProfile": [
      {
        "vnfProfileId": "routing_vnfp",
        "vnfdId": "95b054a1-5055-4916-968a-095da9d4be14",
        "flavourId": "routing_vnf_df",
        "instantiationLevel": "routing_vnf_il",
        "minNumberOfInstances": 1,
        "maxNumberOfInstances": 1,
        "nsVirtualLinkConnectivity": [
```


Figure 72. UC10 NSB overview.

Finally, the TCB, shown in Figure 73, includes all the information necessary for the execution of the requested operations, or an experiment as it is defined in 5G EVE. Such information consists of a set of parameters that might be necessary to configure the VNFs comprising the service, parameters from the infrastructure such as the dynamically generated topic to push the generated metrics on and the experiment actions. The experiment actions are a set of actions to be executed for the configuration and execution of a scenario whose operation the vertical wished to test and validate.



```
"testCaseBlueprint": {
    "description": "TCB for WINGS TOURS UC10 test",
    "name": "WINGS_TOURS_UC10_test",
    "configurationScript": "",
    "executionScript": "EXECUTE_COMMAND vnf.95b054a1-5055-4916-968a-095da9d4be14.extcp.vnf-
    "userParameters": {
        "user": "$$user",
        "password": "$$password",
        "duration": "$$duration"
    },
    "infrastructureParameters": {
        "vnf.95b054a1-5055-4916-968a-095da9d4be14.extcp.vnf-routing-mgmt.ipaddress" : "",
        "vnf.65755d79-79be-4094-a72b-f6867f64ce33.extcp.vnf-spots-mgmt.ipaddress" : "",
        "$$metric.topic.routing_route_calc_time" : "",
        "$$metric.topic.routing_sens_stat_req" : "",
        "$$metric.topic.spots_free_spots_calc_time" : "",
        "$$metric.topic.FA_Count_F-N" : "",
        "$$metric.topic.FD_2way_Avg" : ""
    },
    "version": "1.0"
}
```


Figure 73. UC10 TCB overview.

Depending on the number of components, parameters, metrics, and scenario actions these blueprints can get quite big in size.

5 Conclusions

This document describes the final achievements on the network architecture work and physical deployments in 5G-TOURS. The architecture has been designed to accommodate all the novel technologies developed during the project that support and enhance the full set of 13 use cases. Architectural 5G-TOURS design is considered as foundational basis to host and nurture these innovations. Baseline 5G-TOURS architecture, while encompassing 5G EVE functional layers, fulfils the requirements posed by the use cases from the vertical actors. Architectural "instantiations" are also made, describing the functional assets that are common to all use cases per trial site, and how different WP3 innovations are linked over them.

The physical implementations on the three trial sites - Turin, Rennes and Athens - are explained. A description of full capabilities is given, including use case on-boarding and 5G EVE integration. This includes the technology and insertion strategies of genuine 5G-TOURS components – in addition to what 5G EVE offered – also discussed in the document. In particular, continuous check of consistency on alignment between network infrastructure and use case descriptions was done, as well as work on the architecture to onboard the core network and 5G RAN, physical and logical connectivity, specific terminals, application features etc.

The novel mechanisms brought by the project are presented. In particular, an in-depth discussion and status report is given of Enhanced MANO, AI based orchestration, 5G broadcast/multicast and a new service interface for verticals to deploy services, manage slices and monitor network KPIs. Several implementations of 5G-TOURS technical innovations as part of the expanded infrastructure are described in detail, namely:

- Service Layer to provide an interface to vertical customers: 5 different implementations, providing open-source SDK descriptions;
- Enhanced MANO: 2 implementations (AI-agents implementation and 5G RAN assurance solution);
- AI-based data analytics and orchestration: 2 implementations (AI/ML-based 5GC forecasting and AI for Zero-touch network slicing);
- 5G broadcast support: 2 tracks (LTE-based broadcast and 5GC multicast).

Exploitation and market impact of these network innovations will be tackled in conjunction with WP8 "Business validation and exploitation" during the remaining months of the project, particularly by applying Market & Technology Readiness Level (MTRL) scales, working with technology experts to determine the readiness of the network capabilities both from technology and business points of view, which will be described in the final 5G-TOURS WP8 deliverable D8.4.

Acknowledgment

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No. 856950.

References

[1] M.Gramaglia et al., "5G-TOURS D3.1: Baseline Architecture and deployment objectives", 2019. [online]. Available: https://5gtours.eu/documents/deliverables/D3.1.pdf

[2] C.Barjau et al., "5G-TOURS D3.2: Technologies, architecture and deployment initial progress", 2020. [online]. Available: http://5gtours.eu/documents/deliverables/D3.2.pdf

[3] C.Thiénot et al., "5G-TOURS D3.3: Technologies, architecture and deployment advanced progress", 2021. [online]. Available: http://5gtours.eu/documents/deliverables/D3.3.pdf

[4] S.Castro at al., "5G-TOURS D2.3: Technical requirements of the use cases, economic and deployment implications", 2021. [online]. Available: http://5gtours.eu/documents/deliverables/D2.3.pdf

[5] 5G PPP Architecture Working Group - View on 5G Architecture, Version 4.0, 2021. [online] Available: https://zenodo.org/record/5155657

[6] Python ONAP SDK, sdk to use ONAP programmatically with python code" https://gitlab.com/Orange-OpenSource/lfn/onap/python-onapsdk

[7] 3GPP TS 33.501, "Security architecture and procedures for 5G System", available at http://www.3gpp.org/ftp//Specs/archive/33_series/33.501/33501-g30.zip

[8] Standard - ETSI ISG MEC and 3GPP specifications - Harmonizing standards for edge computing.

[9] "Advanced Research Report: Multi-Access Edge Computing." Del'Orro, 2020.

[10] P. Donegan, "Security Requirements for Deploying MEC at Scale," Heavy Read., 2017.

[11] M. Liyanage, I. Ahmad, A. B. Abro, A. Gurtov, and M. Ylianttila, A Comprehensive Guide to 5G Security, Com édition. Hoboken, NJ: Wiley-Blackwell, 2018.

[12] INSPIRE-5Gplus Consortium "D2.1: 5G Security: Current Status and Future Trends," p. 101, 2019.

[13] Peter Schneider and Josef Urban, "Security in 5G Networks," Breitbandversorgung Dtschl., Mar. 2020.

[14] J. Hodges, "5G Security Strategy Considerations," Heavy Read., 2019.

[15] Standard - ETSI GS MEC 003 - Framework and Reference Architecture.

[16] P. Ranaweera, A. D. Jurcut, and M. Liyanage, "Survey on Multi-Access Edge Computing Security and Privacy," IEEE Commun. Survey. Tutor., pp. 1–1, 2021, doi: 10.1109/COMST.2021.3062546.

[17] Q.-V. Pham et al., "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," ArXiv190608452 Cs Math, Jan. 2020, Accessed: Dec. 09, 2020. [Online]. Available: http://arxiv.org/abs/1906.08452.

[18] R. Tourani, A. Bos, S. Misra, and F. Esposito, "Towards security-as-a-service in multi-access edge," in Proceedings of the 4th ACM/IEEE Symposium on Edge Computing, Arlington Virginia, Nov. 2019, pp. 358–363, doi: 10.1145/3318216.3363335.

[19] N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum, and N. Ghani, "Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations," IEEE Commun. Surv. Tutor., vol. 21, no. 3, pp. 2702–2733, 2019, doi: 10.1109/COMST.2019.2910750.

[20] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez, "A Survey of IoT-Enabled Cyberattacks: Assessing Attack Paths to Critical Infrastructures and Services," IEEE Commun. Survey. Tutor., vol. 20, no. 4, pp. 3453–3495, 2018, doi: 10.1109/COMST.2018.2855563.

[21] Guardtime company, "Enabling Multi Party Trust in the Era of 5G and Multi-Access Edge Computing." 2020.

[22] R. Borgaonkar, L. Hirschi, S. Park, and A. Shaik, "New Privacy Threat on 3G, 4G, and Upcoming 5G AKA Protocols," Proc. Priv. Enhancing Technol., vol. 2019, no. 3, pp. 108–127, Jul. 2019, doi: 10.2478/popets-2019-0039.

[23] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu, "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues," IEEE Access, vol. 6, pp. 18209–18237, 2018, doi: 10.1109/ACCESS.2018.2820162.

[24] A. Masood, D. S. Lakew, and S. Cho, "Security and Privacy Challenges in Connected Vehicular Cloud Computing," IEEE Commun. Survey. Tutor., vol. 22, no. 4, pp. 2725–2764, 2020, doi: 10.1109/COMST.2020.3012961.

[25] G. Choudhary, J. Kim, and V. Sharma, "Security of 5G-Mobile Backhaul Networks: A Survey," J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl., vol. 9, no. 4, pp. 41–70, Dec. 2018, doi: 10.22667/JOWUA.2018.12.31.041.

[26] 5G EVE Deliverable 2.3: https://zenodo.org/record/5070253#.YcCNHFnSI2w

[27] ETSI NFV, https://osm.etsi.org
[28] "Network Functions Virtualisation (NFV) Release 2; Protocols and Data Models; RESTful protocols specification for the Os-Ma-nfvo Reference Point", ETSI GS NFV-SOL 005 V2.4.1 (2018-02)
[29] Lavado G., "OSM Service Assurance," 2019. [Online]. Available: https://osm-download.etsi.org/ftp/osm-6.0-six/8th-hackfest/presentations/8th%20OSM%20Hackfest%20-%20Session%208%20-%20OSM%20Service%20Assurance.pptx.pdf [Accessed 02 12 2021].
[30] Olston C. et al., "TensorFlow-Serving: Flexible, High-Performance ML Serving", 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. [Online]. Available: http://learningsys.org/nips17/assets/papers/paper_1.pdf
[31] https://pytorch.org
[32] Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. Journal of Busi-ness & Economic Statistics, 13(3), 277–280
[33] D. Bega, M. Gramaglia, M. Fiore, A. Banchs and X. Costa-Perez, "DeepCog: Cognitive Network Management in Sliced 5G Networks with Deep Learning," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, 2019, pp. 280-288, doi: 10.1109/INFOCOM.2019.8737488
[34] M. Mao and M. Humphrey, "A Performance Study on the VM Startup Time in the Cloud," 2012 IEEE Fifth Interna-tional Conference on Cloud Computing, 2012, pp. 423-430, doi: 10.1109/CLOUD.2012.103.
[35] 5G-Coral, Deliverable D3.2
[36] D. Bega, M. Gramaglia, M. Fiore, A. Banchs and X. Costa-Perez, "AZTEC: Anticipatory Capacity Allocation for Zero-Touch Network Slicing," IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, 2020, pp. 794-803, doi: 10.1109/INFOCOM41043.2020.9155299.
[37] L.Vignaroli et al., "5G-TOURS D4.2: First Touristic City use case results", 2020. [online]. Available: http://5gtours.eu/documents/deliverables/D4.2.pdf
[38] 3GPP TR 23.757: "Study on architectural enhancements for 5G multicast-broadcast services," 2020. [Online].
[39] 3GPP TR 23.247: "Architectural enhancements for 5G multicast-broadcast services " 2021. [Online].
[40] 3GPP TR 29.532: "5G Multicast-Broadcast Session Management Services; Multicast architecture," 2021. [Online].
[41] 3GPP TS 23.501: "System Architecture for the 5G System; Stage 2" [Online].
[42] 3GPP TS 23.502: "Procedures for the 5G System; Stage 2" [Online].
[43] Olston C. et al., "TensorFlow-Serving: Flexible, High-Performance ML Serving", 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. [Online]. Available: http://learningsys.org/nips17/assets/papers/paper_1.pdf
[44] IETF, "RaptorQ Forward Error Correction Scheme for Object Delivery" 2011. [Online]
[45] 3GPP SP-210376 - New WID on 5G Multicast-Broadcast User Service Architecture and related 5GMS Extensions - https://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGs_92E_Electronic_2021_06/Docs/SP-210376.zip
[46] 5G EVE deliverable D3.1 (5G EVE.eu)
[47] Wireless Edge Factory, https://b-com.com/fr/bcom-wireless-edge-factory
[48] Flexible Netlab platform, https://b-com.com/fr/bcom-flexible-netlab
[49] First version of the experimental portal and service handbook, 5G EVE: 5G EVE - D4.2 (5G EVE.eu)
[50] Models for vertical descriptor adaptation, 5G European, 5G EVE: 5G EVE - D4.3 Models for vertical descriptor adaptation (5G EVE.eu)
[51] E.Giannopoulou et al. "5G-TOURS D6.2: First mobility efficient city use cases implementation results", http://5gtours.eu/documents/deliverables/D6.2.pdf
[52] Autonomous Network Slice Management for 5G Vertical Services PoC description, https://eniwiki.etsi.org/images/2/27/ENI%2821%29000010_POC_9_Final_Report.pdf
[53] 5G EVE project website: https://www.5g-eve.eu/
[54] 3GPP TR 26.802: "5G Multimedia Streaming (5GMS); Multicast architecture", 2020. [Online].
[55] 3GPP TS 26.502: "5G Multicast-Broadcast User Service Architecture", 2021 [Online].
[56] ASUS 5G Smartphone for Snapdragon Insiders specifications: https://estore.asus.com/de/90ai0073-m00030-smartphone-for-snapdragon-insiders.html, https://www.qualcomm.com/snapdragoninsiders/smartphone.