H2020-ICT-2016-2

5G-MoNArch

Project No. 761445

5G Mobile Network Architecture

for diverse services, use cases, and applications in 5G and beyond

Deliverable D5.1

Testbed setup and 5G-MoNArch technologies demonstrated

Contractual Date of Delivery	2018-08-31
Actual Date of Delivery	
Work Package	WP5 – Testbeds
Editor(s)	Hendrik Roreger (HPA), Joanna Bednarz (ATOS)
Reviewers	Ekkehard Lang (NOK-DE), Ignacio Labrador (ATOS)
Dissemination Level	Public
Type	Report
Version	1.0
Total number of pages	70

Abstract: This document provides a draft on the two 5G-MoNArch testbeds that will be used to show the project's innovations, including (i) the general requirements, (ii) the description of the use cases, (iii) the specification and implementation, and (iv) the demonstration plans. These two testbeds are: the Smart Sea Port testbed, representing a vertical industry scenario, and the Touristic City testbed, representing a mobile operator's media and entertainment scenario. These two testbeds are built using the 5G-MoNArch common network functionality to instantiate two network deployments with customised slices. Additionally, they include the customised network functionalities needed to satisfy the specific requirements of each of the scenarios and use cases considered. The Smart Sea Port testbed implements the reliability and resilience functions, while the Touristic City testbed realises the network elasticity functions. This document provides a high-level overview of 5G-MoNArch's technical concepts involved in the testbeds, such as inter-slice control, network reliability and resilience, and resource elasticity. Alongside this, a first definition of the demonstrated use cases together with the general requirements and key performance indicators relevant for both scenarios is presented. Furthermore, technical descriptions are provided for both testbeds, including the envisaged hardware architecture, interface specifications, and the software architecture.

Keywords: 5G, demonstrators, testbeds, sea port, touristic city, network architecture, network slicing, network reliability, network resilience, network elasticity

Executive Summary

The key goal of the 5G-MoNArch project is to develop a flexible, adaptable, and programmable fully-fledged architecture for 5G mobile networks, building on the conceptual results achieved by the 5G-Public Private Partnership (5G-PPP) Phase 1 projects, and complementing them with the enabling and functional innovations developed within 5G-MoNArch. This architecture is being brought to practice in 5G-MoNArch with the development of two testbeds that represent relevant real-world use cases for 5G mobile networks. The goal of these two testbeds is to test, verify, validate and demonstrate the overall approach of network slicing, and the project's use case driven functional innovations. These real-world prototypes are directly run in the environment of two real verticals: (i) the Hamburg port, with the involvement of Hamburg Port Authority (HPA) provides access to the commercial sea port of the city of Hamburg, and (ii) Palazzo Madama of Turin, with the involvement of Turin's municipality through Telecom Italia.

Work Package (WP) WP5 focuses on the implementation and the setup of the testbeds for demonstrating the following innovations of 5G-MoNArch: (i) network slicing with inter-slice control and cross domain management, taking into account network flexibility and adaptability, and the performance demands like throughput, latency, compute and storage resources and processing management; (ii) network reliability and resilience, to ensure failsafe network operation and high network performance; and (iii) resource elasticity, to adapt to load changes and demands in the network without incurring unacceptable overheads in provisioning. These innovations are being implemented with pre-commercial hardware and research prototypes, which are customised to meet the testbeds' needs.

This deliverable provides an initial description of the use cases and the envisaged testbed architecture for testing and demonstrating the network slicing concepts of 5G-MoNArch along with the innovations described above. Specifically, the innovative features as well as the requirements and Key Performance Indicators (KPIs) attained by 5G-MoNArch are shown in the following two testbeds:

- The Smart Sea Port testbed, which implements and demonstrates the selected set of innovations and concepts from WP3 on the topics of network reliability and resilience. It focuses on industrial use cases and applications that aim at improving the operations of a large sea port. This testbed addresses the following three use cases: (i) the improvement of the logistics traffic management through connected traffic lights, (ii) the monitoring of the air quality within the sea port, and (iii) the support of engineering teams through augmented/virtual reality services. The testbed implements and shows a set of innovative concepts and functions in a scenario with three different network slices, which cover these three use cases. The goal is to implement failsafe network slices including high reliability access solutions that guarantee the required Quality of Service (QoS) in case of faults occurring in the network. The use cases will partly be embedded in the operational processes of HPA.
- The Touristic City testbed, which implements and demonstrates the selected set of innovations from WP4 on the topic of network elasticity. The use case addressed by this testbed focuses on enhancing the touristic experience using a virtual and augmented reality application. This includes an augmented/virtual reality service with the life stream of a 360° video, and a haptic feedback together with Voice over IP (VoIP) interactions between a tourist guide and a visitor to the touristic place. The testbed implements and shows a set of innovative concepts and functions for elasticity in a scenario with two different network slices. The goal is to improve the utilisation efficiency of computational resources by adapting the Network Function (NF) behaviour to the available resources without impacting performance significantly, and to optimise the allocation of computational resources to each slice based on its requirements. This will be tested with real end users which will experience this use case at Palazzo Madama.

This document explains how these testbeds show the flexibility of the 5G-MoNArch approach to deploy network slices satisfying a diversity of requirements. The specific KPIs and requirements that need to be satisfied by the testbeds are identified. These KPIs are associated to the 5G-MoNArch concepts demonstrated by the testbeds. Also, the envisaged technical setup of the two testbeds are described, including: (i) site and radio frequency planning, (ii) testbeds deployment, (iii) hardware components

required to demonstrate the chosen innovations, (iv) software modules, (v) required interfaces, (vi) the use cases, and (vii) their visualisation.

The next step during the project will focus on (i) improving the initial definition of the testbeds as described in this document, (ii) the continuation of implementing and deploying the testbeds, and (iii) performing real-life experiments that show the effectiveness of 5G-MoNArch in satisfying real use cases. The testbed setup, the architecture design and its implementation will be more concretised. This includes, among others, the following aspects: (i) the refinement of the selected elasticity scheme to be shown, (ii) the KPIs associated to reliability and resilience, (iii) the design of a refined software architecture and involved interfaces, and (iv) the refinement of the use cases definition and involved steps. These results, in addition to testbed implementation, validation and output of experiments, will be presented in the subsequent Deliverable D5.2, which is due at the end of the 5G-MoNArch project.

List of Authors

Partner	Name	E-mail
HWDU	Ali Ramadan Karthikeyan Ganesan	ali.ramadan@huawei.com karthikeyan.ganesan@huawei.com
NOK-DE	Peter Rost Lars Christoph Schmelz	peter.rost@nokia-bell-labs.com christoph.schmelz@nokia-bell-labs.com
HPA	Hendrik Roreger	hendrik.roreger@hpa.hamburg.de
SRUK	Mehrdad Shariat	m.shariat@samsung.com
CEA	Ghina Dandachi Antonio De Domenico	ghina.dandachi@cea.fr antonio.de-domenico@cea.fr
UC3M	Pablo Serrano Marco Gramaglia	pablo@it.uc3m.es mgramagl@it.uc3m.es
DT	Markus Breitbach	m.breitbach@telekom.de
TIM	Alessandro Trogolo	alessandro.trogolo@telecomitalia.it
ATOS	Joanna Bednarz Ruben Trapero	joanna.bednarz@atos.net ruben.trapero@atos.net
CERTH	Thanos Tsakiris	atsakir@iti.gr
MBCS	Dimitris Tsolkas	dtsolkas@mobics.gr

List of Acronyms and Abbreviations

3D	3-Dimensions
3GPP	3rd Generation Partnership Project
5G	5th Generation mobile wireless communication system
5G-MoNArch	5G Mobile Network Architecture
5G-PPP	5G-Public Private Partnership
5QI	5G Quality of Service Indicators
AMF	Access and Mobility Management Function
AN	Access Network
AR	Augmented Reality
AUSF	Authentication Server Function
BBU	Baseband Unit
BE	Best Effort
CAD	Computer-Aided Design
CAPEX	CAPital Expenditure
CI	Container Infrastructure
CMS	Content management system
CN	Core Network
CSCM	Cross-Slice Congestion Manager
CSMF	Communication Service Management Function
CPNFs	Control Plane (CP) Network Functions
DL	Downlink
DT	Deutsche Telekom
E2E	End-to-End
EM	Element Management
eMBB	Enhanced Mobile Broadband
FCAPS	Fault, Configuration, Account, Performance, and Security
FPS	Frames per second
GbE	Gigabit Ethernet
gNB	next Generation Node B
GS	Guaranteed Service
GUI	Graphical User Interface
HDMI	High-Definition Multimedia Interface
HLS	HTTP Live Streaming
HMD	Head-Mounted Display
HPA	Hamburg Port Authority
ICT	Information and Communication Technology
IDL	Interactive Data Language
ISC	Intra-Slice Controller
ITS	Intelligent Transport System
JIT	Just-In-Time
KPI	Key Performance Indicator
LCM	Lifecycle Management
LTE	Long Term Evolution
LZ77	Lossless data compression algorithm
MAC	Media Access Control
MANO	Management and Orchestration

MEC	Mobile Edge Computing
mMTC	Massive Machine Type Commination
NF	Network Function
NFV	Network Function Virtualisation
NFVO	Network Function Virtualisation Orchestrator
NSI	Network Slice Instance
NSSI	Network Slice Subset Instance
NSMF	Network Slice Management Function
NSSMF	Network Slice Subnet Management Function
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
OPEX	OPerational Expenditure
PAPR	Peak-to-Average Power Ratio
PCAP	Packet Capture
PCF	Policy Control Function
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
PLMN	Public Land Mobile Networks
PNF	Physical Network Function
PPN	Poly-Phase Network
PS	Protocol Stack
RAN	Radio Access Network
RLC	Radio Link Control
RORO	Roll-On-Roll-Off
RRC	Radio Resource Control
RSRP	Reference Signals Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RTMP	Real-Time Messaging Protocol
RTSP	Real Time Streaming Protocol
SCS	Sub-Carrier Spacing
SDAP	Service Data Adaptation Protocol
SDN	Software Defined Networks
SDO	Standards Development Organisation
SDR	Software Defined Radio
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SotA	State of the Art
SThD	Security Threat Detector
SThP	Security Threat Prevention
SThR	Security threat reaction
STZ	Security Trust Zone
ThIntEx	Threat Intelligence Exchange
TRxP	Transmission-Reception Point
TTI	Transmission Time Interval
TTT	Time-To-Trigger

UDM	Unified Data Management
UDP	User Datagram Protocol
UE	User Equipment
UP	User Plane
UPF	User Plane Function
UPNFs	User Plane (UP) Network Functions
URLLC	Ultra-Reliable Low Latency Communication
USB	Universal Serial Bus
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VMI	Virtual Machine Infrastructure
VNF	Virtual Network Function
VNFM	VNF Manager
VoIP	Voice over IP
VR	Virtual Reality
WAN	Wide Area Network
XSC	Cross-Slice Controller

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

The 5G-MoNArch (5G Mobile Network Architecture for diverse services, use cases, and applications in 5G and beyond) project aims to evolve the network slicing concepts developed in the first phase of 5th Generation infrastructure Public Private Partnership (5G-PPP) projects [5GPPP-I] to a fully-fledged architecture that provides the following enabling and functional innovative concepts: (i) network slicing, (ii) inter-slice control and cross-domain management, (iii) network reliability and resilience, and (iv) resource elasticity. One of the 5G-MoNArch project objectives is to prove, validate and demonstrate the technical feasibility of these concepts through the design and implementation of real-life testbeds. The network slicing technology plays a central role in these testbeds. The aim is to showcase the key 5G-MoNArch innovative features and technologies with a number of selected use cases and applications. This task is performed by Work Package (WP) 5, which focuses on the "testbeds".

The testbeds of the 5G-MoNArch project are run by two real verticals that are very good representatives of the potential customers of the 5G technology: Palazzo Madama museum of Turin and the Hamburg Port Authority. Within the framework of 5G-MoNArch, these two verticals run real-world testbeds, each involving multiple logical network instances in parallel satisfying a different set of requirements:

- Smart Sea Port testbed: To demonstrate how the network slicing concept can be employed in industrial environments, this testbed instantiates slices that provide reliable and resilient services with an extremely high level of reliability even under the occurrence of network failure situations. This testbed is deployed for the operations of the Port of Hamburg, involving the challenges occurring when integrating new communication techniques into a real operating environment.
- Touristic City testbed: To demonstrate the media and entertainment use cases, this testbed applies the elasticity concept, which allows the network to efficiently support high-quality services by using network slices that are able to adapt to workload changes by allocating and de-allocating resources in an automatic manner. This use case is tested with real end-users visiting the Palazzo Madama museum.

It is worth highlighting that the above testbeds are set up with building on pre-commercial equipment and research prototypes of the manufacturers involved in the project. This equipment has been modified to incorporate 5G-MoNArch features. This approach ensures that the two testbeds provide the opportunity to clearly impact the commercialisation of network slicing and the project's innovations.

1.2 Objectives

The goal of this document is to describe: (i) the project testbeds and associated use cases, (ii) their technical concepts, (iii) the requirements and KPIs that shall be achieved by these use cases, and (iv) the setup and implementation of the real-world testbeds. More specifically, the main objectives of this document are:

- Identifying and briefly describing those 5G-MoNArch innovations that shall be addressed by both testbeds.
- Describing the current definitions of the use cases that leverage the selected innovations, along with the related requirements and KPIs.
- Describing the planned setup and operation of each testbed infrastructure, along with all the involved stakeholders, and the site plans at different locations.
- Providing a high-level design overview of the functional components as well as the corresponding interfaces among them.
- Identifying the required hardware and software components together with the corresponding interface specifications.
- Determining the most suitable location for the various software components running in the testbed, thereby providing the basis for orchestration decisions in the demonstrators.

- Describing the planned test scenarios and procedures required to show and visualise the attained benefits for both network reliability / resilience and resource elasticity testbeds.

Furthermore, the plans and potential challenges for the implementation are identified. Note that the architecture design and implementation for both testbeds described in this document will be further evolved and validated in the second year of the project and will be presented in the next WP5 deliverable along with the validation and testbed results.

1.3 Structure of the document

This document is organised as follows.

Chapter 2 presents an initial description of the different use cases for both testbeds. It describes the testbeds themselves, the motivation for the selected use cases, the verticals' requirements that have to be satisfied, and the involved functions required to satisfy their needs. Also, the relevant innovations to be implemented in the testbeds are identified.

Chapter 3 provides the background on the concepts that will be shown in the testbeds. It describes the preliminary architecture components, including the role of the different entities and the 5G-MoNArch baseline architecture. The basic concepts and innovations on network reliability/resilience and the mechanisms for resource elasticity (defined in the Work Packages WP3 and WP4, respectively) are presented.

Chapter 4 addresses the requirements and KPIs resulting from the use cases. This includes some of the general requirements and KPIs for 5G networks, as well as specific ones for network reliability and resilience and for resource elasticity. The general requirements for 5G networks have been derived by initiatives outside 5G-MoNArch; among those, the ones were identified that are addressed by the project's testbeds. In contrast, the requirements on reliability and on elasticity are specific to 5G-MoNArch use cases and have been derived by the project. The latter are particularly relevant to the scope of the testbeds.

Chapter 5 describes the current planning for the hardware and software modules of the two testbeds. This includes, among others, the following issues: (i) site and radio frequency planning, (ii) hardware and software modules, (iii) interfaces, and (iv) overall integration. All these aspects are described along with some initial ideas for the test procedures and showcases.

Finally, the conclusions and future steps are provided in Chapter 6.

2 Definition of the use cases

In this chapter, the use cases that will be demonstrated in the Smart Sea Port and the Touristic City testbed are defined. First, a description of the context for each testbed is provided, and then the selected use cases are defined.

2.1 Hamburg Smart Sea Port testbed

2.1.1 Scenario description

Around 9,000 ship calls per year, almost 300 berths and a total of 43 kilometres of quay for seagoing vessels, more than 2,300 freight trains per week, four state-of-the-art container terminals, three cruise terminals and around 50 facilities specialised in handling roll-on-roll-off (RORO) and breakbulk and all kinds of bulk cargoes, along with about 7,300 logistics companies within the city limits – these are just a few of the factors making the Port of Hamburg (see map in Figure 2-1) one of the world's most flexible, high-performance universal ports. 138.2 million tons of cargo crossed the quay walls of Germany's largest sea port in 2016. That included around 8.9 million standard containers. Hamburg is accordingly the third largest container port in Europe and in the 17th place on the list of the world's largest container ports.

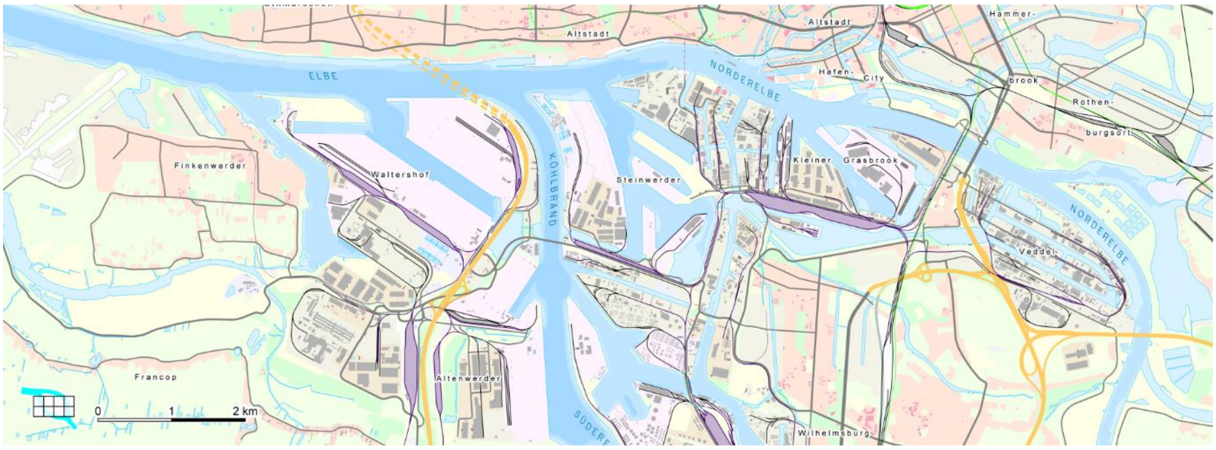

Figure 2-1: Port of Hamburg overview map

The port and the city of Hamburg are closely intertwined – and not just physically. The healthy development of the port ensures the growth and prosperity of the city and the entire metropolitan region. Port management and port development of the Hamburg port is Hamburg Port Authority's (HPA) responsibility. HPA realised early that the future of the port is tied to spatial development, but also to new, smarter approaches. It seemed clear that the various traffic and information flows must be merged to ensure efficient port operations. Efficiency may well be the most important differentiator for the Hamburg port when it comes to energy resources and environmental impact, infrastructure facilities, traffic control, and property management. The so-called HPA's *smartPORT* philosophy combines modern IT-supported transport and communications systems which help to accelerate traffic and trade flows in the port and coordinate them more efficiently. 5G-technology will further fuel this vision.

A sea port is responsible for the efficient traffic and trade management to maximise the throughput of goods. This is of particular relevance for domestic inner-city sea ports like Hamburg, where no possibilities for significant spatial extensions are available. Hence, the logistics within the sea port as well as with all connected infrastructure is of high importance and requires a well-designed information and communication technology (ICT) infrastructure to pave the way for intelligent transport systems (ITS). Such design particularly needs to consider the potentially international impact of failures in the ICT infrastructure, e.g., a stoppage at a central hub such as Hamburg may have an impact on goods transport in all central Europe.

The control systems used by the HPA up to date are already the most progressive in the world and the interplay between their sensor technology, analysis, forecasting and information systems delivers significant efficiency improvements. The 5G Smart Sea Port testbed represents the next step in creating a digital port – and in shaping Hamburg's future.

2.1.2 Use cases

5G-MoNArch intends to improve the operation of the Hamburg port by leveraging the 5G technology to support various use cases or applications that are useful for the port's operation. As the nature of these applications is quite different, the provided technology needs to provide support for service diversity. As the port operation would be severely harmed in case any of these applications did not work properly, the main requirements behind the various use case are:

- *Reliability*: It is essential to provide a guaranteed network connection that maintains a good quality.
- *Resilience*: It is also essential to guarantee availability all the time, even in the case of failures.

Specifically, the Smart Sea Port testbed incorporates the following three use cases, which will be deployed in the real operation of the Hamburg port:

- *Better traffic flow* (ITS and traffic light control): The testbed includes the control of all the traffic lights, which are connected to the central traffic control through the 5G mobile network with a dedicated slice.
- *Improved pollution control:* The testbed includes mobile sensors on barges for emissions measurement. These are environmental sensors installed on HPA ships providing real time data on the current air quality. The sensors are connected through the 5G mobile network with a dedicated slice.
- *Enhanced maintenance experience:* The testbed also includes Augmented Reality (AR) applications for the port's engineering team connected via mobile broadband. Engineers can access construction plans and information for buildings and installations within the port area, using AR applications on tablets or goggles, to simplify access to the data required on site. This equipment is connected to a central application server through the 5G mobile network, using again a dedicated slice.

In the following, each of the above use cases is described in detail.

2.1.2.1 Use Case 1 – Improve port traffic management

Today, traffic light management in the Hamburg port area is largely based on a fixed network infrastructure. Traffic light management is still performed by analogue legacy technology based on copper infrastructure which is up to 30 years old. This ageing infrastructure needs to be replaced in the next years. Traffic light's copper connections may be replaced by fibre. However, this poses the following issues: (i) the necessary fibre infrastructure is not available in all port areas; (ii) expansion of fibre infrastructure is a long-term project with high investments; (iii) digital connection requires replacing all cabling, and (iv) Capital Expenditures (CAPEX) and time efforts are very high (CAPEX for connecting a single traffic light with fibre are up to 100 K€). Due to construction work, mobile traffic lights are frequently used, but not connected to traffic management.

The 5G network slicing solution developed within 5G-MoNarch represents an excellent opportunity to connect all kind of traffic lights to the traffic management system in a safe, secure and high-performing manner, but with manageable CAPEX investments. This will allow to integrate all traffic lights into the sea port's overall traffic management environment.

In particular, 5G-MoNArch deploys a network slicing solution that provides high reliability and security and serves the following purposes: (i) monitoring the status of the traffic lights; and (ii) optimising traffic flow by connecting each traffic light in the Hamburg port area to the central traffic control.

To demonstrate this use case, it will be shown that the deployed system allows for managing the traffic lights with a separate network slice that can be operated independently from other parts of the port's network infrastructure carrying less critical traffic. It will be shown that even in extreme situations, e.g., in terms of radio channel degradations or peaks of network traffic within the sea port, the reliability of the network slice handling the traffic light related network traffic is not compromised.

2.1.2.2 Use Case 2 – Improve port operations using virtual- or augmented reality

Operations support at water gates or construction sites often need on-site expert assistance by engineers. Today, this assistance uses offline data or even paper documents. For example, engineers visit sites to discuss different construction options with stakeholders, often in areas without a physical network connection. The 5G-MoNArch network slicing solution enables those engineers in the field to use digital applications and solutions such as AR and Virtual Reality (VR) to this end. With more than 100 engineers employed, the Hamburg port engineering team is the largest in northern Germany. Increasing their efficiency by mobile network connected VR- or AR-headsets in a safe and secure manner provides considerable operational benefits and efficiency gains.

In particular, the Smart Sea Port testbed deploys a fast and flexible implementation of AR/VR and video streaming system to support the engineering teams remotely or directly on-site. The following operations are demonstrated by this use case, which rely on a network slice specifically supporting the Quality of Service and Quality of Experience (QoS/QoE) requirements of AR/VR and video streaming applications:

- AR- and VR-headsets connected through a mobile network connection and a corresponding network slice to HPA's databases.
- Engineers and experts are enabled to visualise Computer-Aided Design (CAD) construction plans along with real-time or simulated data.

The impact of the latency requirements for the above applications and use cases will be shown.

2.1.2.3 Use Case 3 – Reduce port pollution

The Hamburg port is the basis for the economic prosperity of many people, directly or indirectly supporting over 261,000 jobs. At the same time, it is a landmark of the city and makes Hamburg the proverbial gateway to the world. The HPA has prepared a climate protection concept that contains climate protection objectives, areas of action, activities to reduce emissions and suggestions on how to monitor and evaluate currently planned measures. The HPA's climate protection objectives are based on the regulations of the Free and Hanseatic City of Hamburg, which foresee a 40 percent reduction in carbon emissions by 2020 and an 80 percent reduction by 2050, based on 1990 as reference year. To ensure the continuous reduction of adverse impacts on the environment, a systematic environmental management system in line with the criteria of the international environmental management standard was introduced in 2011. The relevant impacts on the environment are analysed and quantified to identify further areas of action and develop appropriate measures.

To implement the above measures, more sensors need to be deployed in the Hamburg port. Today, there is no reliable infrastructure for mobile sensors. To overcome the current limitations, environmental sensors located on HPA ships are being connected through the 5G mobile network testbed, using a dedicated network slice, to be able to transfer real time environmental measurements. This involves the following challenges:

- Get fast, stable and secure transmission of sensor measurement data to the corresponding sea port Internet of Things (IoT)-Cloud through stable and secure connectivity.
- Moving sensors need to switch mobile network cells reliably.
- Visualisation of data, relationships, and identification of events and incidents.

Furthermore, there is the plan to connect many more sensors on moving vehicles (not only barges) in future, which adds a new challenge to the ones listed above.

2.2 Turin Touristic City testbed

2.2.1 Scenario description

The Touristic City testbed represents a use case of future advanced multimedia and entertainment services in an environment such as a touristic attraction. Indeed, touristic attractions like museum, historical palaces, etc. are particularly suitable for this type of services.

More in detail, this testbed deals with the provisioning of interactive VR content to the end-users visiting the touristic place. To provide an attractive and immersive user experience, this type of applications requires the deployment of dedicated slices able to deliver content at high speed and low delays. Furthermore, in order to provide the required services efficiently, it is essential to implement the necessary elasticity at the infrastructure side, in order to adjust the consumption of computing and network resources to the available resources without harming the resulting quality. In this way, it is possible to avoid overprovisioning resources and thus minimise the cost involved in providing this service.

The testbed will be deployed in the touristic city of Turin, where a large number of visitors are present throughout the year. The specific location chosen for the testbed deployment is Palazzo Madama (Figure 2-2, http://www.palazzomadamatorino.it/en), one of the most representative monumental buildings of Piedmont, located downtown Turin. Today, Palazzo Madama houses the Museum of Ancient Art, whose collections include more than 70,000 works dating from the Middle Ages to the Baroque: paintings, sculptures, illuminated manuscripts, ceramics and porcelain, gold and silverware, furniture and fabrics that bear witness to the richness and complexity production of ten centuries of Italian and European art.

The objective of the Touristic City testbed is to demonstrate the benefits of the 5G functions developed by 5G-MoNArch to provide an immersive and interactive experience to the users. Such functions include: (i) network slicing, to provide the custom requirements needed for VR (e.g., very large throughputs), (ii) Mobile Edge Computing (MEC), to satisfy the latency requirements of haptic interactions between the users, and (iii) resource elasticity, in order to provide an efficient management of resources.

Visitors of the Palazzo Madama museum will be offered the possibility of using the testbed in order to complement their visit to the museum. In this way, it will be possible to gather their feedback and thus evaluate with real users how 5G technology can contribute to improve the user experience in a relevant use case for 5G such as the touristic one.

Figure 2-2: Palazzo Madama in downtown Turin, Italy

2.2.2 Use case – museum visit

The use case of the Touristic City testbed focuses on a multi-user real-time virtual guide of the Madama Reale chamber, which is one of the several representative rooms of Palazzo Madama. A 360° video camera is placed in this chamber to provide a real-time video stream of the room. The testbed is physically deployed and accessible to the end-users in the area in front of the bookshop. In this location, one user (the visitor) is connected via a 5G radio interface and interacts with a second user (which plays the role of a tour guide) connected through a normal Ethernet connection. Both users utilise Oculus Rift

VR headsets plus Oculus touch controllers (https://www.oculus.com/rift/#oui-csl-rift-games=mages-tale) to control virtual avatars that will appear in the real-time 360° environment, captured by the 360° video camera in the Madama Reale chamber.

The scenario involves direct voice communication between both users (VoIP, Voice over IP) as well as coordinated cooperative control of 3D scanned representations of cultural artefacts in the representative room. To provide a highly immersive experience, the users are able to move in the virtual environment through physical movements in the testbed deployment location (a 3 square metres area is provided to each user for an optimal space coverage). The application lets the user learn about this room in an interactive and fun way. Two possible options to this end are the following ones:

- A process where some artefacts in the room have been misplaced and the visitor (with the help of the guide) must coordinate their motions in order to pick up, move and manipulate the artefacts to their correct places in the room. The challenge will be to concurrently lift, move and rotate the objects together: if one of the users does not pick up the artefact at the same time as the other then this is considered "dropping" the object to the floor inducing damage to the priceless artefact.
- A process where the visitor is guided by the guide to perform a restoration of an artefact. The coordination here resides in two ways. First the guide will pass physical tools to the visitor for each step of the restoration process. This means that in the 3D space their hands must be synchronised so that when the tool is passed on from one user to the other the tool is not dropped. Second, when each step of the restoration occurs, there will be two identical copies of the artefact for both the visitor and the guide. The visitor must follow precisely the motions performed by the guide in the exact same location as the guide to successfully perform the action.

In both options, it will be important to guarantee a good coordination between the tourist and the guide. At the same time, haptic communication is involved in the communication between the user and the guide as well as with the system. This requires the provisioning of low latencies in the system, as well as the placement of the server close to the user, involving MEC as well as orchestration technologies.

An illustration of the high-level implementation of the Touristic City testbed is depicted in Figure 2-3. The system works as a client-server system based on two slices: the first slice carries the 360° video stream from the Madama Reale chamber, while the second slice handles all other client-server communication (VoIP, multi-user interaction, 3D model registration and control, etc.). For the first one, an Enhanced Mobile Broadband (eMBB) slice is appropriate for providing the service, while in the second case a URLLC slice is required that that provides the delay guarantees. Within the implementation, appropriate tools (such as, e.g., a traffic generator) are employed in order to simulate an overload in the network. This triggers the elasticity algorithm that guarantees the quality of the user experience.

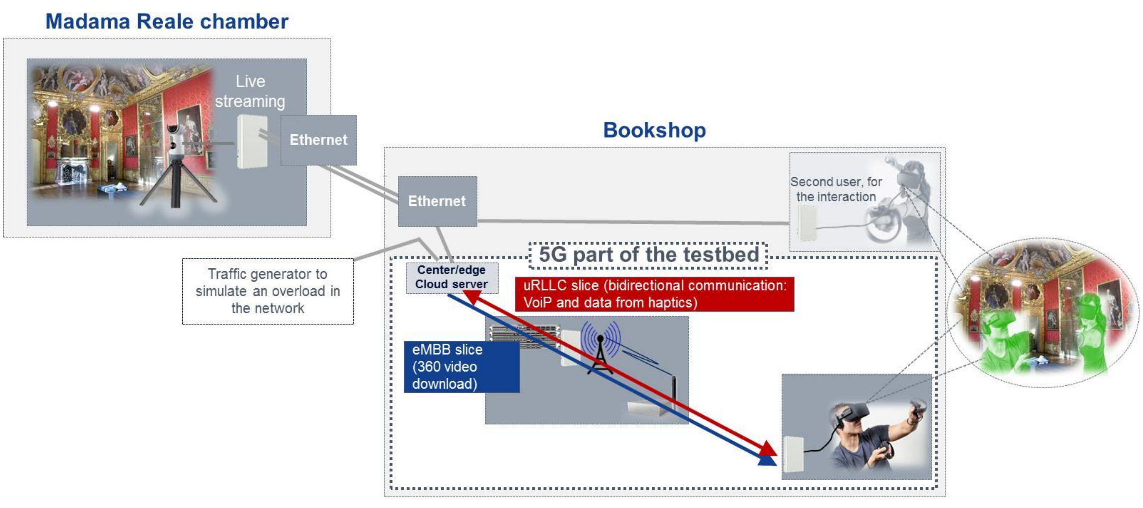

Figure 2-3: High level implementation of the Touristic City testbed

The use case described above is well suited to demonstrate the orchestration-driven elasticity concepts developed within 5G-MoNArch (see Section 3.3.2). Re-orchestration can be triggered as follows. Starting from a situation in which the Virtual Network Functions (VNFs) instantiated for the two slices are running on the edge cloud, a relocation of one of these VNFs may be triggered due to the lack of sufficient resources in the edge. For example, a network overload may be caused by additional tourists visiting the room from other locations, and this may impact the performance of the VR slices dedicated to the first user (in particular, the Ultra-Reliable Low Latency Communication (URLLC) slice that is expected to be the most impacted one). In this context, an orchestration-driven elasticity algorithm could trigger the relocation of the VNFs associated to the 360° video slice to the central cloud, releasing resources in the edge cloud. In this way, all the latency critical functions can be kept in the edge cloud to meet the requirements on latency; in particular, this includes the VNFs for the user interaction belonging to the URLLC slice.

More specifically, this use case involves the following functionalities:

- Two network slices are set up for VR, one for very low latency services and another one for services that are not latency constrained.
- Orchestration capabilities of the 5G-MoNArch architecture are employed to allocate VNFs at the most appropriate location for the service being provided.
- MEC is employed to provide haptic communications with very tight delay constraints – to achieve this, the corresponding VNF is placed at the edge cloud, very close to the user.
- In case of lack of resources, e.g., due to additional traffic in the network, orchestration-driven resource elasticity algorithms are triggered to re-orchestrate VNFs, satisfying the constraints for the MEC functions.

In summary, 5G-MoNArch provides the stringent throughput and delay constraints required for VR and haptic communications, and at the same time, by employing the elasticity concept, it meets these requirements while adapting to the limited resources of the underlying infrastructure. In this way, it can satisfy a very ambitious use case such as the touristic one without incurring very high costs, which fits the nature of this use case.

The reader is referred to Chapters 3 and 5 for more details about the concepts mentioned above; specifically, in Chapter the main 5G-MoNArch concepts involved in the testbeds are described, while Chapter 5 explains how these concepts are being implemented and deployed in the testbeds.

3 Selected 5G-MoNArch technical concepts

This chapter provides a high-level description of a subset of those 5G-MoNArch concepts and innovations that will be implemented in the project's testbeds. These innovations include inter-slice control solutions, experiment driven optimisation and a set of network reliability and resilience, and network elasticity solutions. Note that these innovations have been developed within WP2 "Flexible, adaptive architecture design", WP3 "Resilience and Security" and WP4 "Resource Elasticity" and are described in detail in the corresponding WP deliverables: D2.1 [5GM17-D21], D3.1[5GM18-D31] and D4.1[5GM18-D41]. In this document, only those concepts that are related to the testbeds are summarised, with the aim of providing a self-contained document.

3.1 Flexible adaptive architecture

Network slicing will be implemented and demonstrated in both 5G-MoNArch testbeds. Different slices have different service requirements, implement dedicated network functions, and require KPIs to be fulfilled depending on the use case. Depending on the respective services, different types of slices can be implemented, for example, eMBB, massive Machine Type Communication (mMTC), or URLLC) slices. In cases where services interact with each other, those slices have to be correlated. For example, in the Touristic City testbed where an AR/VR use case will be demonstrated, two slices have been defined: an eMBB slice for the 3D high resolution video streaming, and an URLLC slice for the haptic interaction of the visitors. These slices must be jointly managed and controlled in order to obtain good end-user experience. A detailed description of the slices to be implemented in the testbeds is provided in Section 5.1 for the Smart Sea Port, and in Section 5.2 for the Touristic City.

In this section the inter-slice control and experiment driven optimisation concepts are introduced as a set of the key enabling innovations in 5G-MoNArch that are relevant for all use cases and testbed implementations.

3.1.1 5G-MoNArch architecture concepts and instantiation

The novel 5G mobile network architecture shall provide an adaptive and flexible multi-service network, belonging to different stakeholders and with different service characteristics requirements. This architecture must enable the deployment and simultaneous management of different network slices, which is one of the key aspects of the 5G-MoNArch project.

The 5G-MoNArch mobile network architecture concept is based on the results of 5G-PPP Phase 1 projects, especially 5G NORMA [5GNOR17-D33], and is depicted in Figure 3-1 [5GM18-D22] below. The network architecture is divided into 4 layers: Service layer (responsible for business support systems, policy and decision functions as well as applications and services operated by a tenant or other external entities), Management & Orchestration layer (responsible for orchestration and management of multi-tenant and multi-service networks), Control layer (they translate decisions from control applications to VNFs and Physical Network Functions (PNF) and Network layer (comprises VNFs and PNFs in User Plane (UP) and Control Plane (CP). It can include 3GPP Rel. 15 CP functions such as: Access & Mobility Management Function (AMF), Session Management Function (SMF), Authentication Server Function (AUSF), Radio Resource Control (RRC) and UP functions: User Plane Function (UPF) and Packet Data Convergence Protocol (PDCP) [5GM18-D22]. Furthermore, the design distinguishes between intra-slice and cross-slice NFs.

The Management & Orchestration layer (M&O) is based on different network technologies such as the Network Function Virtualisation (NFV) Management and Orchestration (MANO) framework defined by ETSI [ETSI14] and 3GPP public mobile network management. This layer is enhanced to support End-to-End (E2E) network slicing and multi-tenancy concepts including network service requirements aspects. The M&O layer comprises the E2E M&O sublayer that hosts the 3GPP Network Slice Management Function (NSMF) and Communication Service Management Function (CSMF). These functions manage network slices and communications services, across multiple management and orchestration domains. Additionally, it contains Cross-slice and Cross-domain Orchestration & Management functions that implement orchestration and management of E2E network slices considering their corresponding KPIs. The Cross-slice Orchestration & Management function is responsible for the management of the functions that are shared between different network slices, while Cross-domain Orchestration & Management is in charge of management between different domains in a single slice. Furthermore, like in ETSI MANO, the M&O layer includes the Virtual Infrastructure Manager (VIM), the VNF Manager (VNFM) and the NFV Orchestrator (NFVO) as part of the Virtualisation MANO domain that is responsible for the Life Cycle Management (LCM) of Virtual Machines (VMs). In addition to the ETSI NFV MANO components, the 5G-MoNArch architecture is extended to support LCM of virtualisation containers. NFVO together with VNFM and VIM provide the functionality that allows the usage of either the Virtual Machine Infrastructure (VMI) or the Container Infrastructure (CI). The M&O layer communicates with the Control layer, which contains Intra-Slice Controller (ISC) and Cross-Slice Controller (XSC).

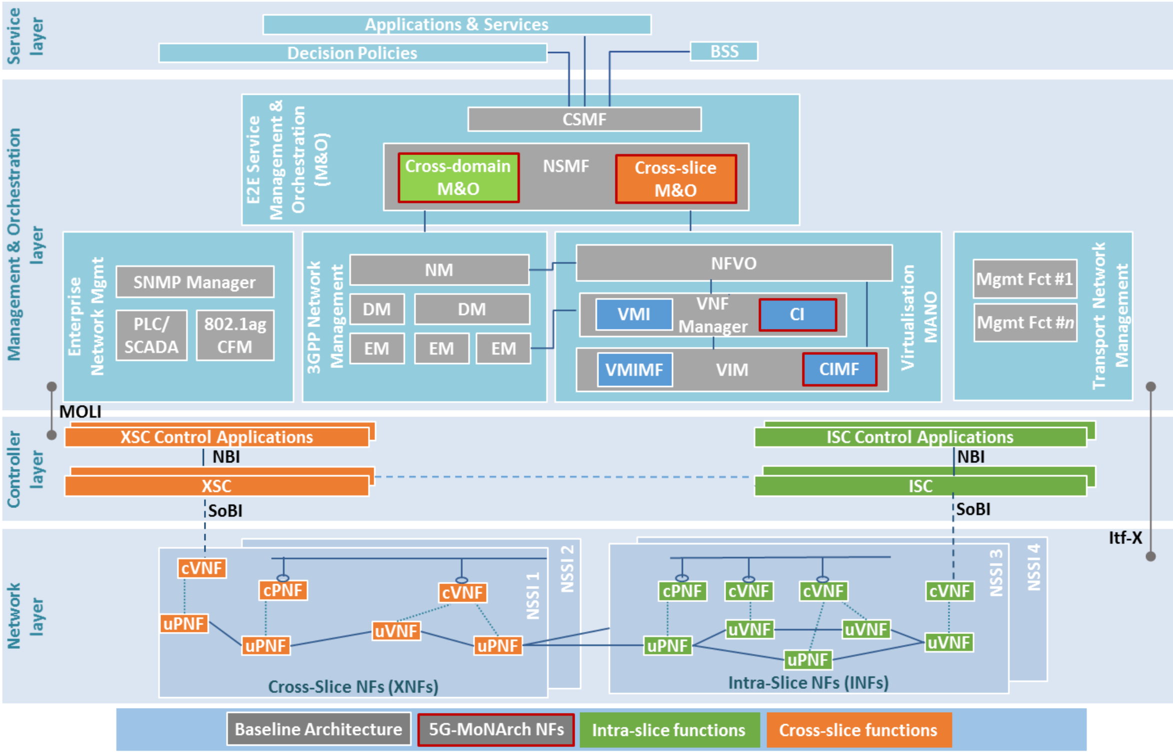

Figure 3-1: Initial high-level functional view of overall 5G-MoNArch architecture [5GM18-D22]

Additionally, the 5G-MoNArch architecture needs to support the three functional innovations, namely: resilience, security and resource elasticity demonstrated in the Smart Sea Port testbed and the Touristic City testbed respectively. Further details of the functional approaches are detailed below.

Smart Sea Port testbed architecture

For the use cases deployed in the Smart Sea Port testbed of 5G-MoNArch, a customised architecture instance of the general overall architecture is utilised. For this instance, a subset of the 5G-MoNArch enabling common network functionality as well as selected functions of the use-case specific functions developed in WP3 are utilised. The latter include cross-domain and cross-slice security and resilience management, 5G Fault Management (FM) functions as well as multi-connectivity-enabled Radio Access Network (RAN) for increased reliability (RAN reliability sub-plane.)

The Smart Sea Port testbed target architecture instance is depicted in Figure 3-2. It shows the network functions in each layer: Network layer, M&O layer and Service layer, respectively. Currently, the optional 5G-MoNArch control layer is not foreseen to be part of the Smart Sea Port testbed. The network functions are distributed across four locations in Hamburg and Nuremberg. In Hamburg, there are the Deutsche Telekom (DT) Data Centre, the HPA Data Centre (and private networks), and the TV Tower hosting the base station. In Nuremberg, one of the central office Data Centres (‘central cloud’) of DT is hosted. The four locations are depicted in yellow in Figure 3-2.

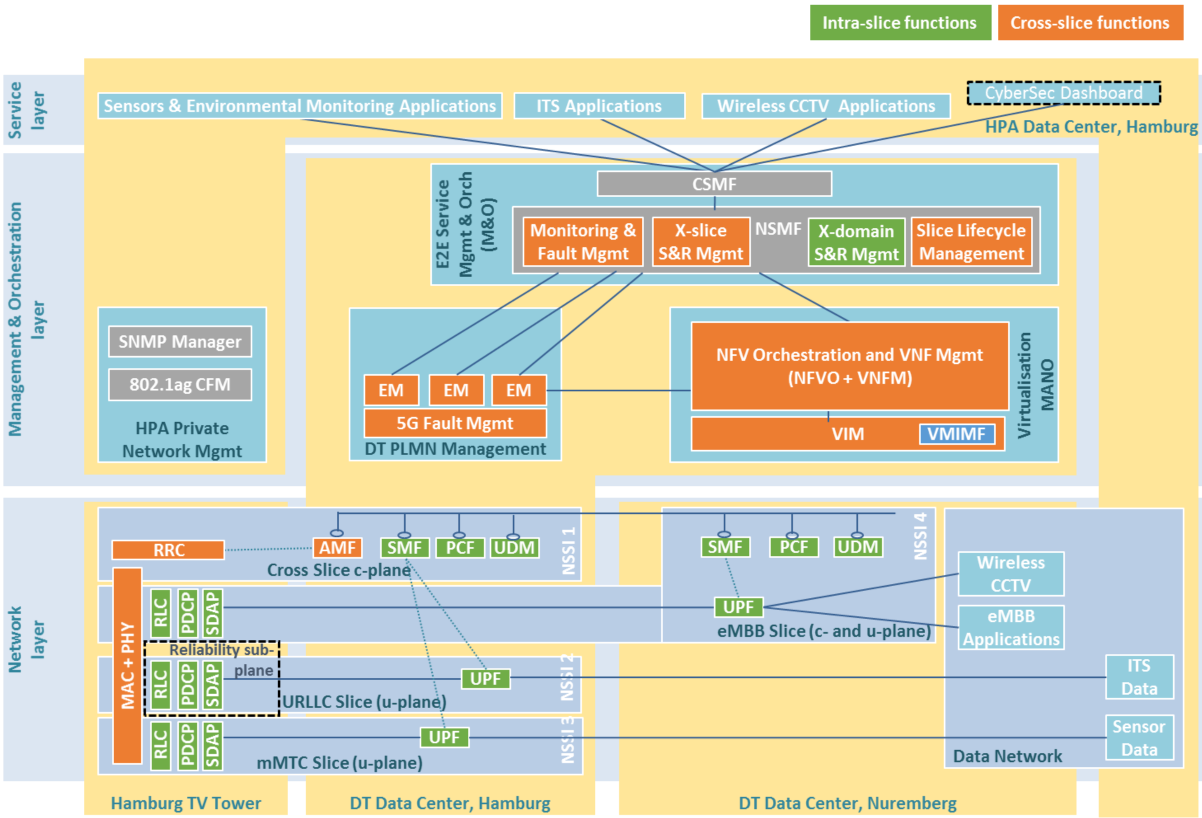

Figure 3-2: Targeted functional architecture for the Smart Sea Port testbed [5GM18-D22]

In the Network layer, the testbed implements three network slices, eMBB communication, URLLC, and mMTC delivering the AR, ITS, and Environmental Sensing use cases, respectively. They have the following deployment characteristics:

- **eMBB network slice**: The eMBB network slice is utilised to carry the AR traffic (e.g., augmented maintenance for HPA service staff) as well as providing eMBB services like Internet access or video streaming to cruise ship tourists.

  In the RAN, the slice uses the common Physical Layer (PHY) and Medium Access Control (MAC) layers of the testbed radio infrastructure. Service Data Adaptation Protocol (SDAP), PDCP, and Radio Link Control (RLC) layers are slice-specific due to customisations reflecting service requirements. Further, RRC is common for all deployed slices. In the CP, the AMF is shared with other network slices, while Policy Control Function (PCF), Unified Data Management (UDM), and SMF are dedicated to the eMBB slice. Core Network (CN) UPFs are dedicated and therefore service-specific. Besides AMF, all CN functions of the eMBB slice (from CP and UP) run in DT's central cloud data centre in Nuremberg, one of DT's central office sites. Further, the AR applications (and other eMBB-like applications) in the Data Network that process the incoming user data are also hosted in Nuremberg.

- **URLLC network slice**: The URLLC network slice is utilised for ITS applications in the sea port area, in particular for the traffic light control.

  Similar to the eMBB slice, the URLLC slice uses the common RRC and lower radio layers (MAC and PHY) and service-specific upper radio layers (RLC, PDCP, SDAP) in the RAN. One such service-specific customisation comprises the WP3 reliability sub-plane for multi-connectivity, thus increasing reliability in the RAN. In the CN, AMF is shared with all three deployed slices, while SMF, PCF, and UDM are shared among the slices deployed in the local edge cloud (DT Data Centre, Hamburg), i.e., URLLC slice and mMTC slice. An alternative deployment option would comprise separate PCFs for each of the two slices. Further details of PCF selection can be found in clause 6.3.7.1 of [3GPP18-23501]. The CN UP uses a dedicated and customised UPF instance. Due to latency requirements for traffic light control, all network functionality in CP and UP is deployed locally. Therefore, also the ITS Data applications in the Data Network are operated in the local HPA Data Centre in Hamburg.

- **mMTC network slice**: The mMTC slice is used to carry traffic from environment sensors deployed in the Hamburg sea port, particularly from the barges patrolling within the port area.

  The slice has the same setup as the URLLC slice in terms of deployment of network (CP and UP, RAN and CN) and application functions. Nevertheless, upper layer radio functions (RLC, PDCP, SDAP) and CN UPF are realised as dedicated instances with customised behaviour.

The M&O layer comprises DT's functionalities for managing public land mobile networks (PLMN), particularly Element Management (EM) functions and novel advanced 5G FM functions as developed in WP3. For the virtualisation MANO layer, the Smart Sea Port testbed utilises a VM-based virtualisation approach. The deployment uses a streamlined ETSI NFV MANO architecture, i.e., VIM and an NFV LCM component integrating NFVO and VNFM. Nevertheless, in general, container-based solutions could be incorporated. For E2E M&O, NSMF and CSMF incorporate according monitoring as well as FM and slice LCM functions. From WP3, cross-domain and cross-slice security and resilience management functions are incorporated into NSMF. A lightweight CSMF implementation provides mediation capabilities between NSMF and the Service layer. Beyond these M&O layer functions operated by DT in their Hamburg Data Centre, the deployment comprises the management functions for HPA's private networks, namely SNMP Managers and 802.1ag Connectivity FM running in HPA Data Centre in Hamburg. The latter functions manage the largely wireline network infrastructure of HPA which is also used to connect the UPs of the local network slices with the HPA Data Centre. More specifically, as depicted in Figure 3-2 , the *ITS* and *Environmental Monitoring/Sensor Data applications* run in the HPA Data Centre in Hamburg where the UP data coming from the URLLC and mMTC slice, respectively, are forwarded to. Form the mobile network perspective, these application functions belong to the Data Network outside the operator domain. Only in case of the eMBB slice, the application (*AR Data*) is hosted in the DT Data Centre in Nuremberg. Finally, each of the three applications also has a management and control component residing in the Service layer, executed in the local HPA Data Centre. They interact with the CSMF to provide the specific service requirements used to customise the slice instances and to receive latest performance and configuration details about the network slice hosting the respective service.

Touristic City testbed architecture

The Touristic City target architecture instance is depicted in Figure 3-3 that shows the NFs in each layer, i.e., Network layer, M&O layer, and Service layer. The NFs of this testbed are deployed in three main locations: the antenna site (where the radio PNFs are executed), the edge cloud, and a central cloud (that has a higher latency to the User Equipment (UE)). They are depicted in yellow in Figure 3-3. In practice, the edge and the central cloud are two dedicated cloud infrastructures, connected via fibre to the antenna site. The central cloud emulates a farther processing site, with an increased latency but a lower operational cost. Both sites are deployed in the premises of the demo but are owned by the operator.

In the Network layer, the testbed implements two network slices: an eMBB communication and an URLLC one. They are used to provide two different services: the high-res video streaming for the AR applications and the haptic server that connects the avatars for their interactions. They have the following deployment characteristics:

- **eMBB network slice**: The eMBB network slice delivers the high resolution 360 video to the mobile user. In the RAN, the slice uses the common PHY and MAC layers of the testbed radio infrastructure, while the higher layers are slice-specific due to customisations reflecting specific service requirements. The RRC instead, is common to both slices. The CP functionality is shared across slices, while the UPF is dedicated to each slice. In terms of deployment, the CN NFs are deployed in the central cloud, as well as the UPF. Also, the application server runs in the central cloud.
- **URLLC network slice**: The URLLC network slice is utilised for delivering the low latency haptic interactions among the avatars (one fixed and one mobile). The radio deployment is equivalent to the eMBB network slice. Also, the CN NF setup is similar in terms of sharing and

deployment. However, the UPF may be moved from one cloud to the other according to the specific load of the network, according to the inputs coming from the elasticity modules deployed in the NFVO.

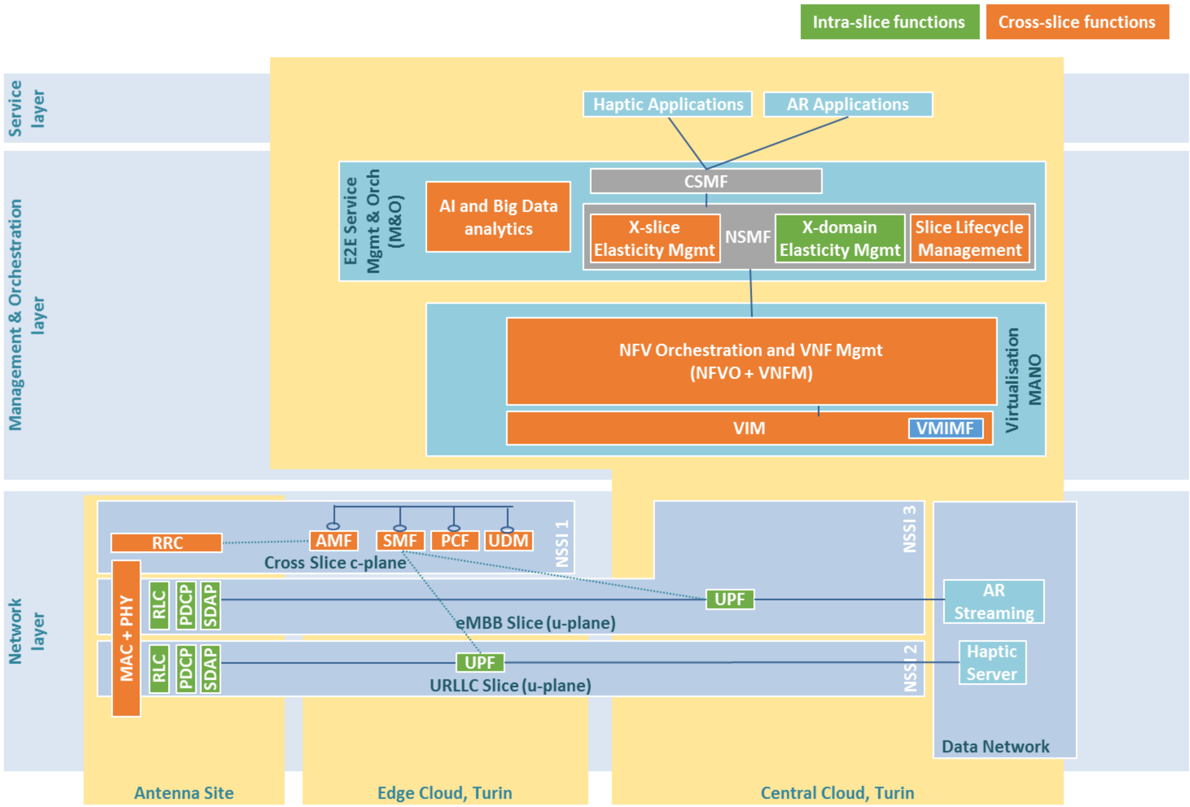

Figure 3-3: Targeted functional architecture for the Touristic City testbed [5GM18-D22]

The management of the network comprises an implementation of the 3GPP elements CSMF, NSMF and Network Slice Subnet Management Function (NSSMF) that, in turn, include specific elasticity modules. The testbed includes both PNFs and VNFs that are managed by a VM-based virtualisation approach and the related MANO modules. The MANO stack is a simplified one, that relies on a VIM and an ad-hoc implementation of the VNFM and NFVO. Nevertheless, container-based virtualisation may be included, in particular for the radio functions (this will be assessed after the final integration tests). The loop is closed by monitoring modules that report the current load to the management modules that use this information (e.g., network load, CPU load) to trigger both cross-slice and intra slice elasticity algorithms.

3.1.2 Network slicing and inter-slice control

Network slicing and inter-slice control [5GNOR17-D33], [5GM17-D21] is one of the key enablers that allows the 5G mobile network architecture to achieve greater flexibility, and to efficiently adapt to and support services with different characteristics and requirements. By mapping different services to different slices, it is easier to satisfy service-specific demands within each slice, for example, related to latency, resilience or elasticity. This allows a higher level of optimisation and flexibility as the NF can be dynamically reallocated across slices and can belong to more than one slice.

3GPP [3GPP18-28801] defines a Network Slice Instance (NSI) as a set of NFs (and the hardware / computing / storage resources for these NFs) which are arranged and configured, forming a complete logical network to meet certain network characteristics. An NSI is complete in the sense that it includes all functionalities and resources necessary to support a certain set of communication services thus serving a certain business purpose. The NSI contains NFs belonging to both (R)AN and CN. If the NFs are interconnected, the 3GPP management system [3GPP18-28801] contains the information relevant to connections between these NFs such as topology of connections and individual link requirements.

An NSI may be composed of network slice subnets of PNFs and/or VNFs. A Network Slice Subnet Instance (NSSI) constituent may include at least one NF. An NSSI may be shared by two or more NSIs, this is called a shared constituent of NSI. Figure 3-4 shows two NSIs with partly shared NSSIs; NSSI A is connected to NSSI B and NSSI C whereas NSSI B and NSSI C are independent. NSI X is composed of NSSI A and NSSI C whereas NSI Y is composed of NSSI A and NSSI B.

5G-MoNArch follows the 3GPP approach related to network slice management [3GPP18-28801], which foresees new management functions for Network Slice and Network Slice Subnet management. As E2E network slices can be composed of a subset of dedicated and shared resources such as: storage, computation or radio access, 5G-MoNArch introduces the concepts of Inter-Slice and Cross-Domain Management Functions that compose an E2E service management and orchestration function. These components can be mapped to CSMF or NSMF functions within 3GPP framework. To manage the NFs and their resources properly 5G-MoNArch foresees corresponding control elements: ISC and CSC. Cross-slice control is introduced to improve the systems efficiency on using shared infrastructure resources and fulfil the slices' KPIs.

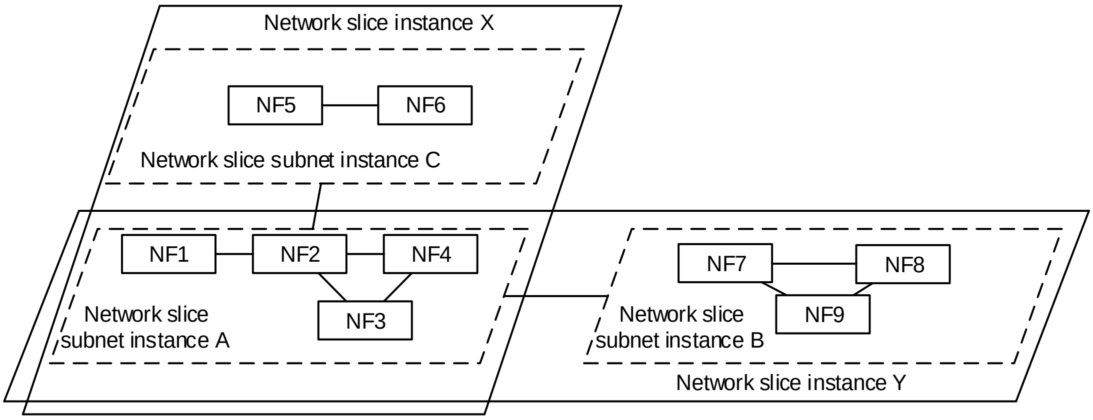

Figure 3-4: NSI X and Y composed by NSSI A, B and C [3GPP18-28801]

3.1.3 Experiment-driven optimisation

Experiment-driven optimisation is one of the key enablers that are going to be implemented in 5G-MoNArch testbeds. The idea of this innovation is to build and adjust 5G-MoNArch algorithms based on the experiments performed on real environment. The goal is to improve the performance and achieve a high level of optimisation of the cloud infrastructure.

To design novel algorithms, it is at first necessary to understand the real-time behaviour of the functions and the infrastructure. This knowledge serves to adjust the models of the function behaviour and thus enhance them significantly. This approach is especially applied to orchestration algorithms, which are demonstrated in both testbeds. Indeed, to create a flexible and adaptive architecture, the orchestration algorithms need to be enhanced. For example, in the Touristic City testbed, the orchestration algorithms move NFs across the infrastructure to achieve flexible adaptation to network conditions (such as changing network load). Depending on these conditions and the availability of the resources at the edge cloud, the re-orchestration of determined NFs from central cloud to edge cloud might be performed (see Section 3.3.2).

The optimisation of the E2E cross-slice orchestration requires exhaustive measurements throughout the network [5GM17-D21], for example measurements of latency and throughput as well as of the proper NFs and their computational aspects such as processing power, memory and storage.

When optimising the algorithms, it is necessary that all the experimental procedures take into account the QoS restrictions of all the VNFs. The VNFs can either belong to one particular slice or be shared among several slices. The E2E cross slice optimisation is especially challenging as it has to ensure the requirements of each VNF optimally across slices. To ensure the QoS, continuous measurements and monitoring will be made based on quantitative KPI verification [5GM17-D21].

The measurements are performed at network slice level, physical infrastructure level and/or at VNF level. The following aspects are considered when refining the model for resource consumption behaviour:

- Network level aspects:
  - Latency
  - Packet error rate
  - Download time
  - End-to-end delay
- Physical infrastructure level aspects:
  - Available resources at a node
  - Spectrum
  - NFV level
- VNF level aspects:
  - Shortage on disk storage/CPU/RAM

The observation results of the experiments will be then evaluated and used to modify the models in order to achieve higher level of optimisation.

Table 3-1 summarises the enabling innovations derived from WP2 [5GM17-D21] that are going to be implemented in the testbeds:

Table 3-1: Key innovations derived from WP2

	Innovation areas	**Challenges**	**Proposed Solutions**
Innovations derived from WP2	Inter-slice control and management	E2E cross-slice optimisation	Development of network control functions in order to achieve a flexible and programmable inter-slice control and management framework
	Experiment driven optimisation	Lack of experiment-based E2E resource management of VNFs	Measurement campaigns, VNFs and Network Slices performance monitoring

3.1.3.1 Initial test results for experiment-driven optimisation

In order to apply the experiment-driven optimisation approach in the Touristic City testbed, individual real-time tests have been performed to the different modules that compose the testbed. The complete E2E test of the integrated components (PHY, protocol stack (PS), and application) will be performed in a later stage of the project.

Test results of PHY layer

The PHY layer of Touristic City testbed implements two bandwidth parts as shown in the Figure 3-5 with different PHY layer numerology configuration (e.g. subcarrier spacing, cyclic prefix, Transmission Time Interval (TTI) length, etc.). The PHY layer numerology for eMBB is 60 KHz subcarrier spacing with a 1ms slot duration while that of URLLC is 90 KHz subcarrier spacing and a 500 µs slot duration. The cyclic prefix for both is 2 µs.

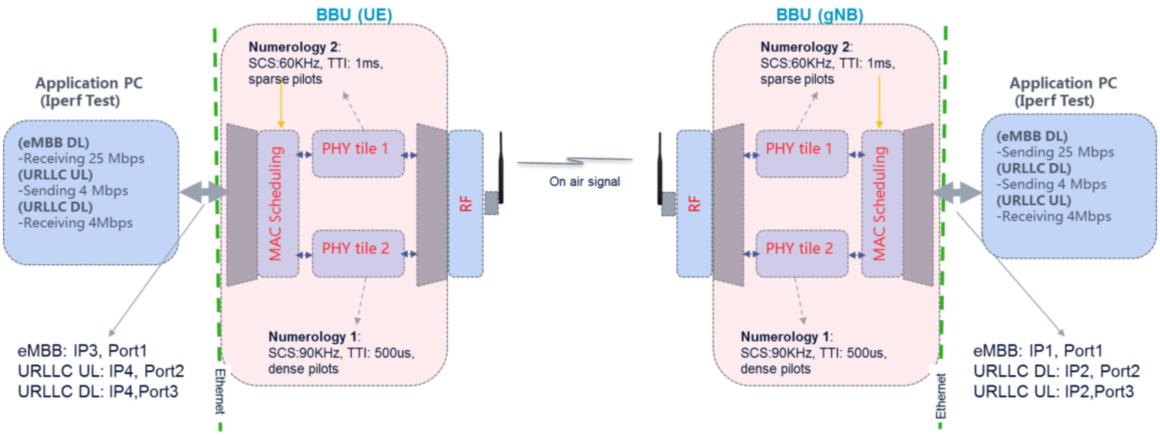

Figure 3-5: PHY with two bandwidth partitions with different numerologies

Figure 3-6 shows the real-time test results of the two slices in terms of latency, throughput and signal quality. The test has been performed in an indoor environment with transmit power of ~3 dBm. The measured throughput shows around 25 Mbps unidirectional traffic for eMBB while 4 Mbps bidirectional traffic for URLLC. The round-trip latency of each slice was measured as 2.8 ms for URLLC and 5 ms for eMBB. The obtained values meet the strict latency and throughput requirements in 5G systems, such as 5Mbps in case of URLLC slice and 30 Mbps for eMBB slice.

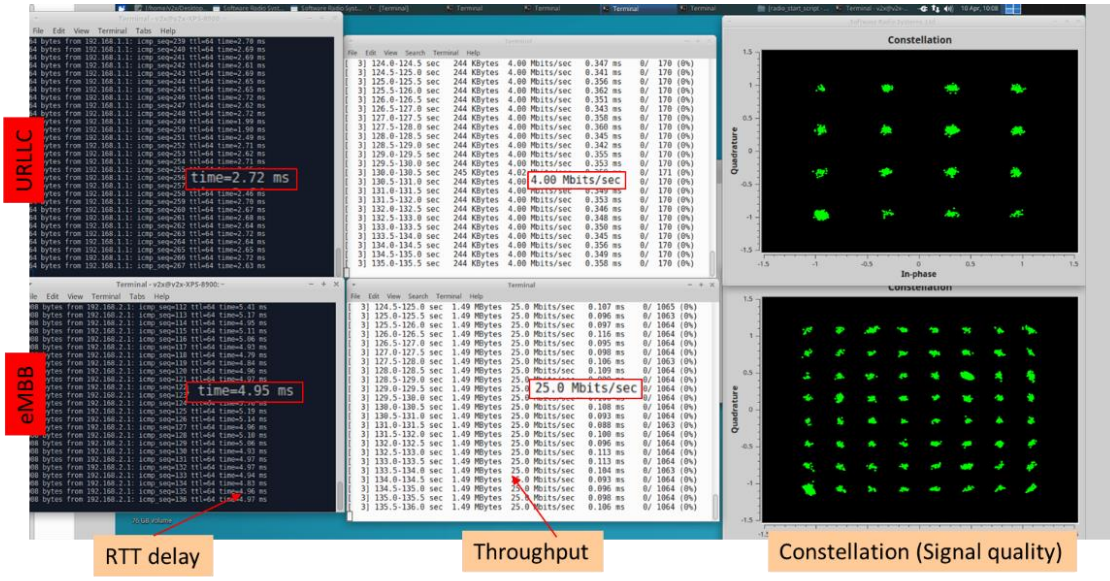

Figure 3-6: Throughput and latency measurement using IPERF

Test results of the higher layers of the protocol stack

The tests of the higher layers of the protocol stack were focused on the performance of the UP NFs implemented in the PDCP and RLC protocols for UpLink (UL) and DownLink (DL) flows, at both the Next Generation Node B (gNB) and UE sides. The illustration given in Figure 3-7 depicts the software components used for the test. The gNB and UE parts of the higher layer's software were implemented in two different VMs, respectively, while both VMs are hosted in a single physical machine. The characteristics of each of the VMs and the physical machine were as follows:

- Processor type CPU(s) 4 x Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz
- Cache memory size 8 MB SmartCache
- Memory assigned 4 GByte
- hypervisor PROXMOX Virtual Environment 5.1-41

- OS Ubuntu 16.04.4 kernel:4.4.0-31 generic

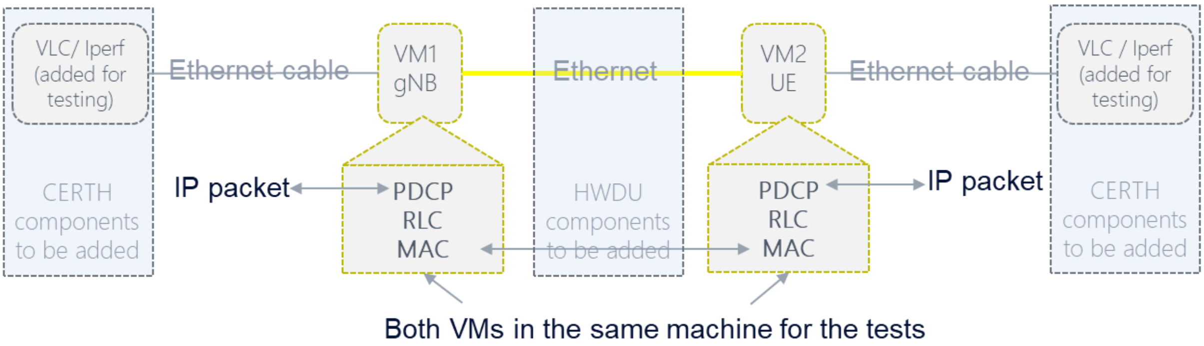

Figure 3-7: Illustration of the set up used for the higher layer's tests

To quantify the performance of the protocol stack implementation, the focus is on the gNB and more precisely on the DL traffic that it serves. For this purpose, the CPU load has been monitored for various traffic loads (from 5 to 30 Mbps, see Figure 3-8) together with the time that is needed to process the packets that enter the stack (Figure 3-9).

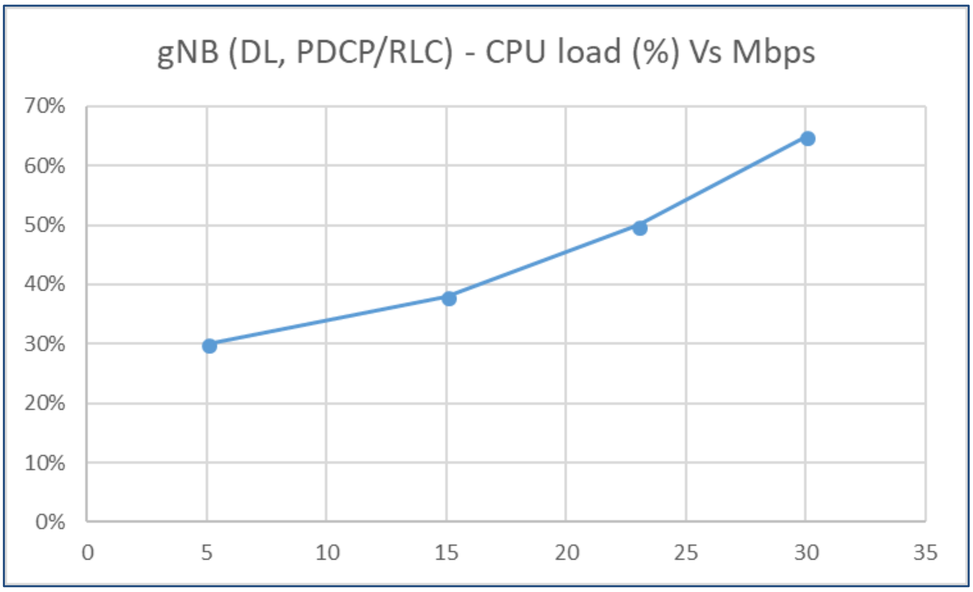

Figure 3-8: CPU load versus the traffic load (in Mbps) that the higher layers serve (eMBB slice, gNB part, DL flow)

As depicted in Figure 3-8, the current implementation of the higher layers of the protocol stack can support up to 30 Mbps by consuming up to 65% of one processor core. Additionally, for the time needed for packet processing, it can observed in Figure 3-9 that the higher layer at the gNB part is expected to burden the E2E performance with up to 1 ms latency (600 microseconds on average). The results validate that the eMBB and the URLLC slices [5GM18-D62], foreseen for the Touristic City demo, can be supported efficiently by the current implementation of the higher layers. Moreover, they provide a preliminary assessment of the impact of the function split between MAC and RLC to the overall performance. The key outcome of this assessment is that the higher layer functions are not highly demanding in terms of CPU and RAM and the MAC - RLC split may be reconsidered in scenarios where the latency is a critical KPI and the extra delay due to the MAC - RLC Ethernet-based interfacing can be avoided. The next step is to evaluate the higher layers in an E2E implementation where all the components foreseen for the Touristic City testbed are available (including potentially a platform where CPU demanding functions will be virtualised and be subject to VIM services in order to showcase resource elasticity).

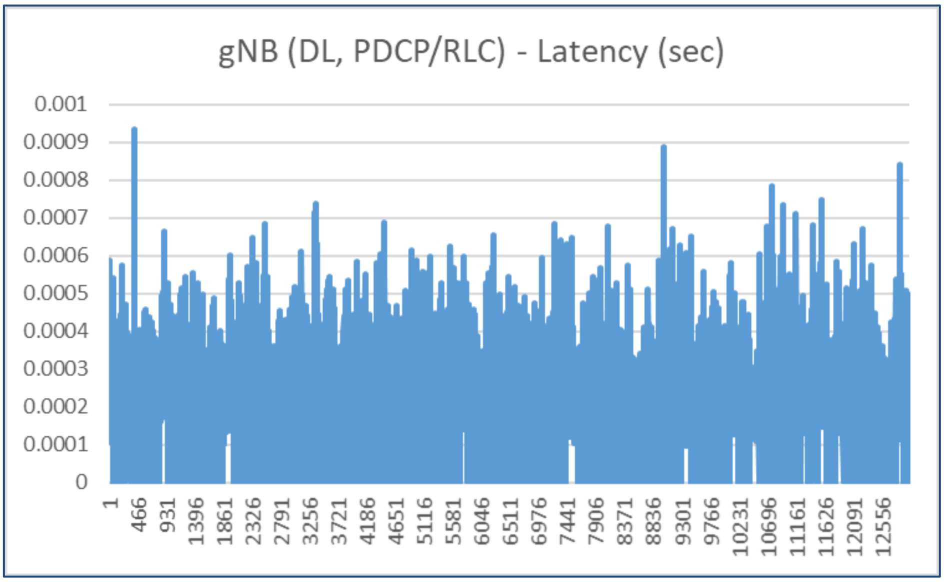

Figure 3-9: Time needed for the higher layers processing per received packet (eMBB slice, gNB part, DL flow)

3.2 Network reliability, resilience and security

The Smart Sea Port specific reliability and resilience functions to be demonstrated are chosen among different solutions and innovations described in WP3. These selected concepts consider the use case requirements described in Section 2.1.

3.2.1 Network reliability concepts and mechanisms

Ensuring high RAN reliability is a challenging task, especially for certain use cases such as high-mobility scenarios. Reliability in RAN refers to the success probability of transmitting a certain packet to its destination within a given delay requirement. To improve RAN reliability, two approaches are studied within 5G-MoNArch WP3 (see further details in D3.1 [5GM18-D31], Chapter 2): multi-connectivity and network coding, where only multi-connectivity is planned to be implemented into the Smart Sea Port testbed.

Multi-connectivity uses data duplication, where the same packets are transmitted multiple times to minimise the probability of erroneous reception. A mobile terminal is connected to two macro-cell base stations and data is duplicated along both paths. Hence, this approach represents a specific case of macro-diversity using a distributed antenna system. Besides increasing the link reliability, the approach also reduces the service interruption during handover caused by the time to trigger (TTT) as well as the necessary re-connection to the target cell. Therefore, a handover in 3GPP Long Term Evolution (LTE) implies an interruption of few 10 ms, which is supposed to be reduced to less than 1 ms in the envisioned implementation. This will allow guaranteeing that no packets are lost even during handover.

The multi-connectivity approach is illustrated in Figure 3-10, i.e., it shows schematically two cells implemented in the Smart Sea Port testbed. When a mobile terminal starts moving in cell 1 towards the cell border, it will at some point in time add the second cell as a secondary base station (point 1). From this moment on, the mobile terminal is able to receive duplicated data via both base stations. As the mobile terminal moves closer to the second cell, the secondary base station will be at some point stronger than the master base station, target and source base station will change the roles (second cell will become new master cell and vice versa) and finally the mobile terminal will only be connected to the new master cell (point 2).

Network coding is a technique with high potential in improving the performance and throughput of networks. In traditional networks, signals from different nodes are treated separately, and the intermediate nodes within the network are only allowed to perform routing operations, i.e., the intermediate nodes forward their received signals to their destinations without performing any kind of processing. It could be shown that for certain networks the performance can be improved if the intermediate nodes are allowed to perform operations, where they combine their received packets and forward these combinations to their destinations, where they are decoded. Depending on the application as well as the requirements involved, this improvement in the performance can be converted into gains in terms of transmission rate, reliability, as well as transmission power. In the framework of 5G-MoNArch WP3, network coding is mainly studied to improve the reliability, where network coding approaches for UL and DL are investigated separately, as both cases impose different structures.

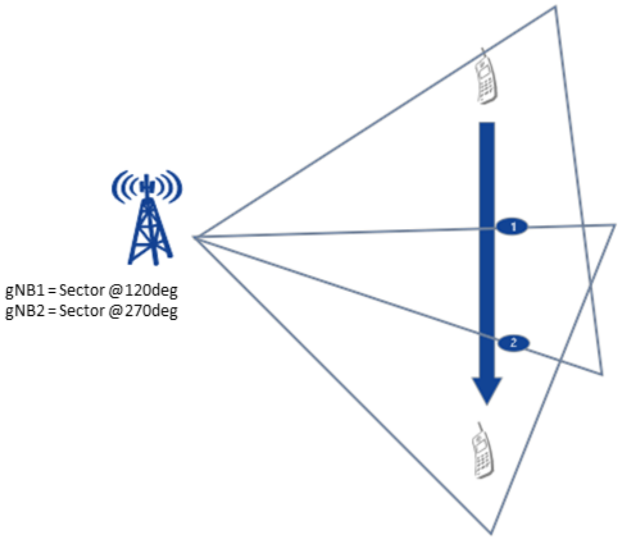

Figure 3-10: Illustration of multi-connectivity approach implemented in the Smart Sea Port testbed

3.2.2 Telco cloud resilience concepts and mechanisms

Resilience is the ability of the network to continue operating correctly during and after unexpected disturbance, such as the loss of mains power. More specifically, telco cloud resilience describes the ability to provide and maintain an acceptable quality level of services in case of faults, in particular for network functions running on virtual infrastructure, i.e., the telco cloud as denoted within the 5G-MoNArch project.

To improve the resilience in telco clouds, four approaches are studied within 5G-MoNArch WP3 (see further details in D3.1 [5GM18-D31], Chapter 3):

- *5G Island*: Dimensioning and configuring edge cloud resources to allow autonomous operation of basic network services at the edge without requiring continuous connectivity to the central cloud. The envisaged approach enables autonomous failsafe operations through distributing VNF implementations to the edge clouds, instead of relying on the centralised servers in the CN (central cloud).
- *VNF redundancy* can improve resilience through offering different levels of availability depending on the redundancy model and the redundancy strategy (active, passive). The redundancy model can combine active and standby replicas of hosted VNFs. In active strategy, there are no standby replicas and all the replicas work in parallel. When one node fails, tasks executing at the failed node can be resumed in any remaining node. In passive redundancy, there is one working replica whereas remaining replicas are standby.
- *FM* techniques to identify, trouble shoot and isolate occurring network faults; this includes monitoring tools for the detection of changes, potential problems and anomalies in network behaviour, the root-cause analysis enabling the localisation of the actual problem, and problem isolation such that the propagation of the fault effects and impact to the rest of the network can be minimised.
- Methods for supplementary *hardening* of the resilience of critical NFs.

The expected outcome of the corresponding tasks in WP3 are two-fold: It involves an investigation of the trade-off between availability and the required level of resilience, along with a specification of edge cloud dimensioning and study and deployment of fault management techniques.

With respect to the implementation of resilience mechanisms into the Smart Sea Port testbed, the analysis of the possibilities is still ongoing, as a number of preconditions for each of the approaches are to be fulfilled. Under main consideration at the current point in time is the implementation of dedicated FM techniques.

3.2.3 Security mechanisms

WP3 has designed a complete model for the protection of 5G infrastructures. The approach of this model is based on the instantiation of several components (intra- and inter-slice) for the monitoring of the services running within a network slice. Intra-slice components include elements for the detection, reaction and prevention of incidents, while inter-slice components include elements that coordinate the individual elements instantiated in every network slice and prevent the propagation of incidents across other network slices (Figure 3-11 represents this schema; details are given in D3.1 [5GM18-D31]).

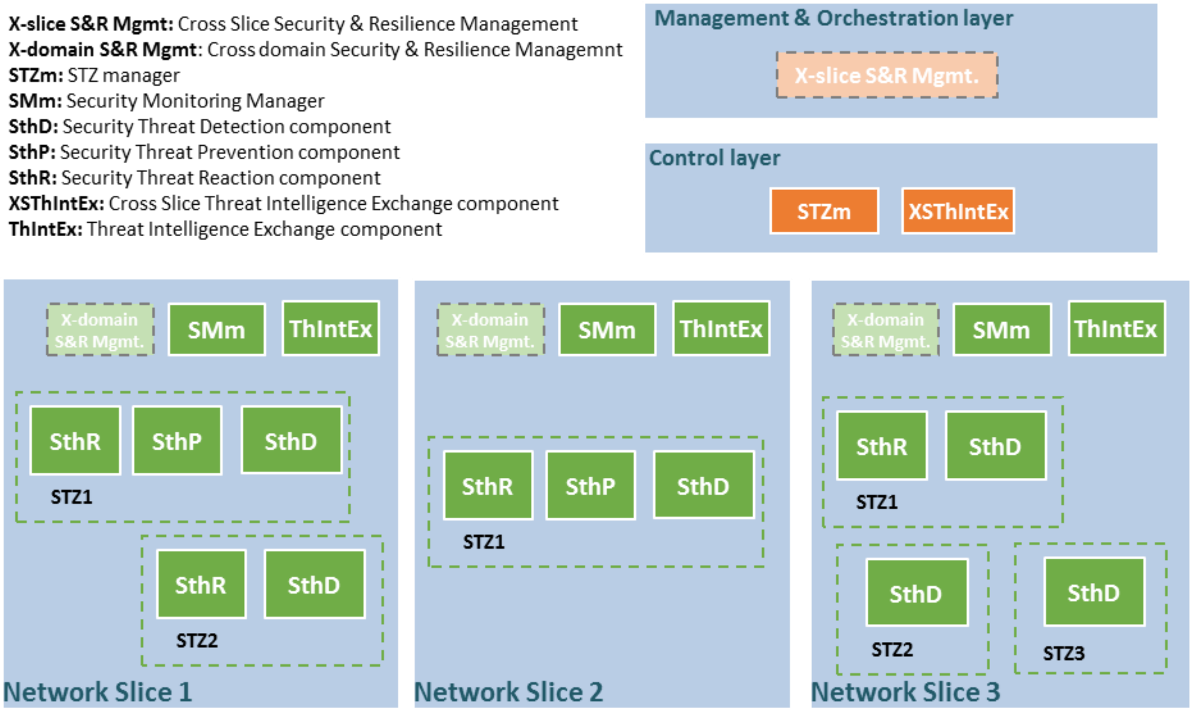

Figure 3-11: Slice-aware STZ deployment strategy (taken from D3.1 [5GM18-D31])

The main characteristic of this security model is based on adapting the security protection components to be deployed to the security requirements of the network slice. To this end, the deployment of the security components will be adapted to the testing infrastructure based upon an incremental process that will increase complexity as long as the development and deployments within the pilot progresses. Three phases will guide this process:

- Phase 1: Monitoring and alerting. In this phase a network slice will be instantiated. The security monitoring components will be deployed and configured to detect incidents affecting the resources available at the network slice. A customised attacker will be deployed to simulate several types of attacks: denial of service, brute force attacks, etc. Several sensors will be deployed within the network slice in order to detect anomalies. New rules and plugins will be implemented at the Security threat Detector (SthD) in order to be adapted to events detected within a network slice. Additionally, in this stage the incidents detected will be exported to prevent the propagation of incidents across slices (as part of the Threat Intelligence Exchange – (ThIntEx)- component, as described in D3.1[5GM18-D31]).

- Phase 2: Evaluation of incidents over a network slice to design reactions. Based on the incidents detected in a network slice the Security threat Reaction (SthR) will evaluate potential countermeasures to mitigate incidents within a network slice. The characteristics of the network slice will determine the possible countermeasures to enforce. The level of criticality of the services available at the network slice will also be used to provide with risk values that will estimate the most suitable reaction.
- Phase 3: Deployment of security trust zones adapted to specific requirements of a complete network slice. In this stage the Security Trust Zone (STZ) concept will be introduced (see D3.1[5GM18-D31] for details). Several security trust zones templates will be available. These templates will be used to instantiate concrete STZs. The STZs will be tailored to the characteristics of the network slices available (for instance, a separate STZ for every use case demonstrated within the Smart Sea Port testbed). In this phase the prevention mechanisms (through the Security threat Prevention –SthP) will be used based on the security information exchanged between slices through the threat intelligence capabilities defined in WP3.

Figure 3-12 represents a graphical representation of the aforementioned validation phases. The security alert shown in the figure is piece of information (i.e., a message) that notifies about an incident within an infrastructure and includes details about the characteristics of the incident (origin, destination, type of incident, severity).

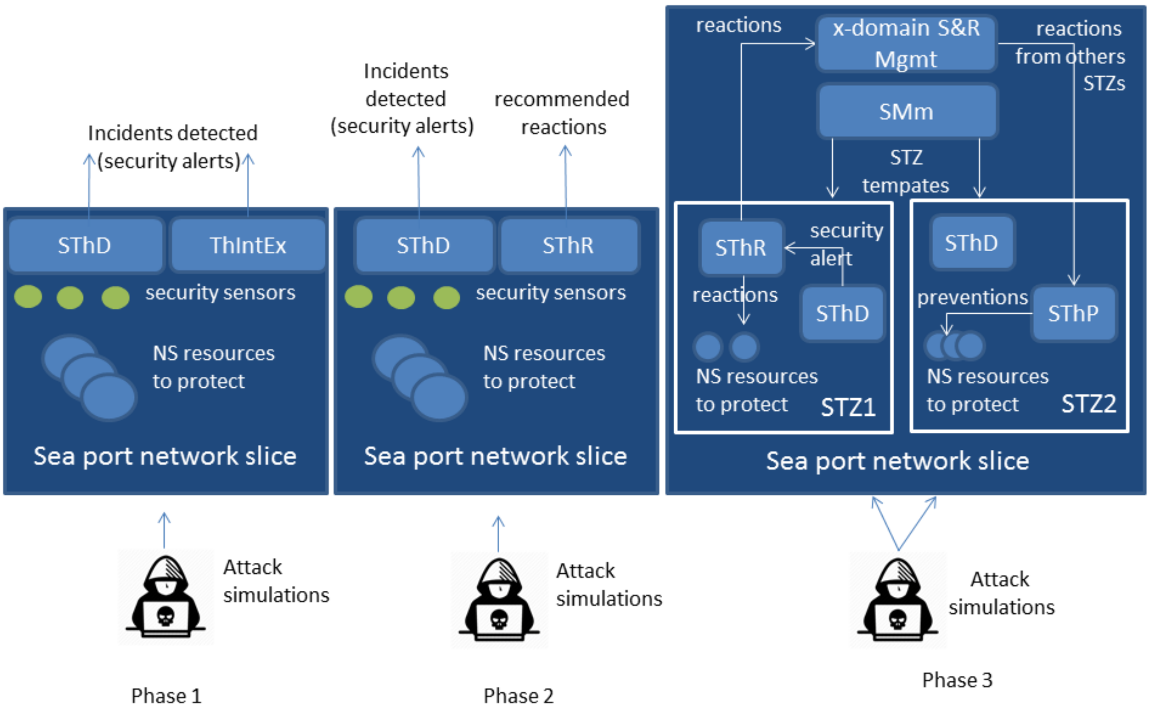

Figure 3-12: Security mechanisms strategy for validation over the Smart Sea Port testbed [5GM18-D31]

3.2.4 Reliability, resilience and security key innovations

The specific implementation, test procedure, and measured KPIs for reliability and resilience are further discussed in Section 4.2. In Table 3-2, the described key innovations, challenges, and the proposed solutions are summarised with a focus on those solutions that are likely to be implemented into the Smart Sea Port testbed.

Table 3-2: Key innovations, challenges and proposed solutions towards network reliability and resilience

	Innovation areas	Challenges	Proposed Solutions
Network reliability and resilience	RAN reliability	Ensure high network reliability to ensure packet transmission even in case of mobility	Multi-connectivity
	Resilience in telco clouds	Provide high level of service quality even in case of faults in the network	Fault management
Security mechanisms	Network slice adapted security	Dynamically adapt the security mechanisms to the security requirements of different networks slices	Security Trust Zones templates
	Inter-/Intra-slice incident reaction and prevention mechanisms	Provide techniques to prevent the propagation of incidents by exchanging incident information across network slices	Threat Intelligence Exchange, Security Threat Reaction/Propagation

3.3 Resource elasticity

The Touristic City specific selected elasticity functions to be demonstrated are chosen among different solutions and innovations described in D4.1 [5GM18-D41]. The selected concepts take into account the use case requirements described in Section 2.2.2.

3.3.1 Resource elasticity dimensions

In this section, the key ideas on how to provision resource elasticity are described, with particular emphasis on the technical challenges in the virtualised architecture of 5G systems that resource elasticity is meant to address (see Table 3-3) as well as design hints on the type of solutions or mechanisms that will address those challenges within the 5G-MoNArch project.

Table 3-3: Key innovations, challenges and proposed solutions towards elastic 5G architecture

	Innovation areas	Challenges	Proposed Solutions
Resource-elastic virtual functions	Computational Elasticity	Graceful scaling of computational resources based on traffic load	Elastic NF design and scaling mechanisms
	Orchestration-driven Elasticity	NF interdependencies	Cloud aware protocol stack
	Slice-aware Elasticity	E2E cross-slice optimisation	Monitoring and provisioning of resources exploiting multiplexing gain

As shown in the table, the resource elasticity mechanisms are classified along three dimensions: (i) computational elasticity, (ii) orchestration-driven elasticity, and (iii) slice-aware elasticity. In the following, each of these dimensions and the mechanisms are further explained:

- *Computational Elasticity*: The goal of exploiting computational elasticity is to improve the utilisation efficiency of computational resources by adapting the NF behaviour to the available resources without impacting performance significantly. Furthermore, this dimension of elasticity addresses the notion of computational outage, which implies that NFs may not have sufficient resources to perform their tasks within a given time. To overcome computational outages, one potential solution is to design NFs that can gracefully adjust the amount of computational resources consumed while keeping the highest possible level of performance.

RAN functions in particular have been typically designed to be robust only against shortages on communication resources; hence, the target should be directed at making RAN functions also robust to computational shortages by adapting their operation to the available computational resources. An example is a function that chooses to execute a less resource-demanding decoding algorithm or number of iteration in case of resource outages, admitting a certain performance loss. Additional virtual users and services must be added to the live users in the testbed in order to generate stress on the system and to increase the computational load on the used hardware. Suitable measures will be triggered to balance the KPI metric to guarantee the promised and required QoE.

- *Orchestration-driven Elasticity*: This innovation focuses on the ability to re-allocate NFs within the heterogeneous cloud resources located both at the central and edge clouds, considering service requirements, the current network state, and implementing preventive measures to avoid bottlenecks. The algorithms that implement orchestration-driven elasticity need to cope with the local shortage of computational resources by moving some of the NFs to other cloud servers which are momentarily lightly loaded. This is particularly relevant for the edge cloud, where computational resources are typically more limited and more cost efficient than in the central cloud. Similarly, NFs with tight latency requirements should be moved towards the edge by offloading other elastic NFs without such tight timescale constraints to the central cloud servers. Section 2.2.2 describes how the orchestration-driven elasticity concept can be used in Touristic City testbed. To efficiently implement such functionalities, special attention needs to be paid to (i) the trade-off between central and edge clouds and the impact of choosing one location for a given function, and (ii) the coexistence of MEC and RAN functions in the edge cloud. This may imply scaling the edge cloud based on the available resources, clustering and joining resources from different locations, shifting the operating point of the network depending on the requirements, and/or adding or removing edge nodes [EOB +16]. Also important are the resources available on the network interfaces e.g. backhaul links or network segments connecting edge and central cloud servers especially in the live setup of the testbed where the traffic shares public network resources.
- *Slice-aware Elasticity*: Finally, this innovation addresses the ability to serve multiple slices over the same physical resources while optimising the allocation of computational resources to each slice based on its requirements and demands, a challenge earlier referred to as E2E cross-slice optimisation. Offering slice-aware elastic resource management facilitates the reduction of CAPEX and OPEX by exploiting statistical multiplexing gains. Indeed, due to load fluctuations that characterise each slice, the same set of physical resources can be used to simultaneously serve multiple slices. Adaptive mechanisms that exploit multiplexing across different slices must be designed, aiming at satisfying the slice resource demands while reducing the amount of resources required. Hence, the solutions must necessarily dynamically share computational and communications resources among slices whenever needed.

3.3.2 Mechanisms to provide elasticity

In this section, some of the mechanisms that are considered for bringing elasticity into the testbed are described. The focus is thereby on orchestration-driven and slice-aware elasticity algorithms. Computational elasticity algorithms are evaluated in the framework of WP4 and WP6, with a special focus on their large-scale deployment.

A possible solution for both orchestration-driven and slice-aware elasticity is implemented at the NSMF by dynamically controlling resource availability, slice priorities, and queue state. The NSMF may decide, based on the service level requirements of a class, to scale down the allocated resources to one or multiples slices in order to accept a larger number of requests, which have high priority. The proposed Cross-Slice Congestion Manager CSCM [5GM18-D41] has to be able to forecast the impact of a decision on the overall system performance [PJD+15]. This intelligence is ensured by using Reinforcement Learning (RL) techniques that allow to make the optimal decisions maximising resources utilisation [GBL+12].

For preliminary results, two slice classes are defined: Best Effort (BE) and Guaranteed Service (GS). In order to prioritise the deployment of GS requests, a higher reward is assigned for accepting their requests. It is important to note also that negative rewards will be considered when dropping a GS request so that the policy is pushed toward deploying more GS requests rather than BE slices. In this first study Q-learning techniques are used to compute the optimal strategy to implement at the CSCM [GBL+12]. In the results shown in Figure 3-13, the proposed solution is compared with a Greedy policy in term of accepted and dropped slice requests. The results show that the proposed solution is able to improve the resource utilisation enabling to increase the percentage of accepted slice request without negatively affecting the performance at the GS slices.

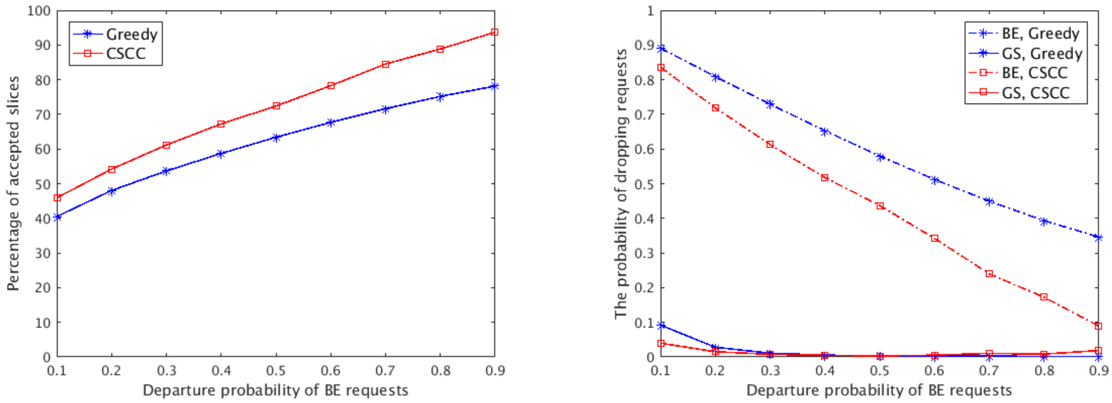

Figure 3-13: Percentage of accepted slices for Greedy and the proposed QL-based CSCM as a function of BE departure probability (left) and Dropping probability for Greedy and QL-based CSCM as a function of BE departure probability (right).

At a more general level, given the current trend of softwarising NFs to run in the cloud, the used definition of elasticity is based on concepts from cloud computing. According to [ODCA14], a cloud application has reached the maximum maturity level (see Table 3-4), namely, is *elastic*, if it "has been designed to fully exploit the unique qualities of a cloud environment," being able "to migrate seamlessly from one cloud environment to another, either partially or completely, without any interruption in service." Also, they can scale up or down depending on the load, this scaling being horizontal if more instances of a service are provided, and vertical if more resources (CPU, memory) are added to the same instance.

Table 3-4: Maturity levels of cloud applications

Maturity Level	Description
Level 3: Elastic	Application can dynamically migrate across infrastructure providers w/o interruption of service
	Application can elastically scale out/in appropriately based on stimuli
Level 2: Abstracted	Services are stateless
	Application is unaware and unaffected by failure of dependent services
	Application is infrastructure agnostic and can run anywhere
Level 1: Loosely Coupled	Application consumers one or more cloud services: compute, storage, network
	Application services are discoverable by name
	Application compute and storage are separated
Level 0: Virtualised	Application runs on virtualised infrastructure
	Application can be instantiated from an image or script

The investigated case deals with instances of a complete network (i.e., slices) instead of just a single application. Note that a service is composed by a set of inter-connected (i.e., "chained") NFs, which run over a set of different shared resources. More specifically, the 3GPP specifies in [3GPP18-28801] four-stages concerning the network slice LCM, which are: (1) preparation, (2) instantiation, (3) run-time, and (4) decommissioning, see Figure 3-14. Elasticity therefore has to also refer to the ability to dynamically alter the operation of a slice during the run-time phase, where this altering of the slice might happen along two axes:

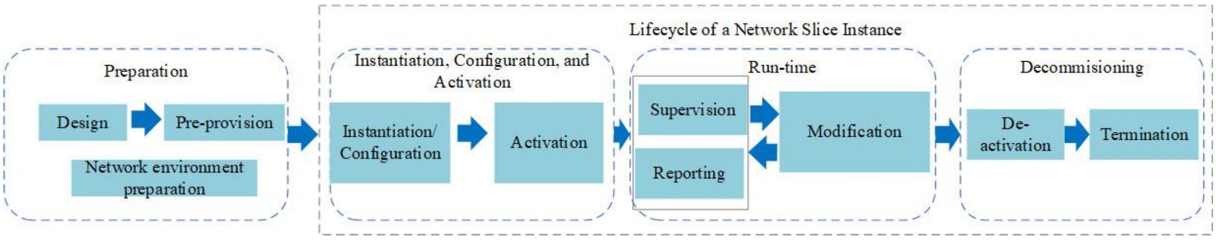

Figure 3-14: Lifecycle of a network slice instance [3GPP18-28801]

- The "function chaining" axis: the way in which networking functions are connected can be dynamically re-configured without interruption of the service. Note that this includes both the substitution and the re-ordering of functions.
- The "resource assignment" axis: the reallocation of resources to a given slice can be dynamically changed, also without interruption of the service. Note that this goes beyond the horizontal and vertical scaling, as a network slice can be re-allocated over different physical elements.

The adoption of adaptability requires supporting a high dynamicity, higher than the one already introduced with network slicing. Network slicing, with the above four-stages, supports a rearrangement of the network in a more automatic and more thorough way than with the *classic* Fault, Configuration, Accounting, Performance, Security (FCAPS) model. However, to fully support adaptability, the Management and Orchestration of the network shall be able to perform updates and amendments to the NFs providing a given communication services, especially during the run-time stage and without introducing any disruption of the service. This results in the following set of features:

- **Function Scaling**: the horizontal or vertical scaling of cloud environments shall be supported, but with the ability to instantiate new resources in different server farms.
- **Service Function Chain amendments**: it shall be possible to add, substitute or remove a NF from a network slice.
- **Network Function sharing**: it shall be possible to share, or un-share, a given NF among network slices.
- **Network Function relocation**: it shall be possible to move (not necessarily "scale") a NF between different location of the physical infrastructure.

One key adjective of the definition of adaptability is "seamlessly," i.e., without any interruption of the service. In this way, the definition is associated with a service, which is defined by a set of explicit or implicit KPIs. These should be taken into account when introducing a modification into the function chain, with the consequence that some services would be "more adaptable" than others, depending on the way the NFs are programmed (if they are properly designed for a cloud environment), provisioned (i.e., how they make use of the available resources), and physically deployed (in particular, if there are stringent timing constrains). The 5G-MoNArch Touristic City testbed elastic functionality will embrace the aforementioned methods to provide elasticity.

3.4 Selected 5G-MoNArch innovations summary

The use cases described in Section 2.1.2 and in Section 2.2.2 for the Hamburg Smart Sea Port testbed and the Touristic City of Turin, respectively, are used to demonstrate a selected set of key innovations of 5G-MoNArch. These innovations have been described in the previous sections. In the following it is specified in detail which innovations are implemented in each of the testbeds.

The objective of Smart Sea Port testbed is to demonstrate the following 5G-MoNArch innovations:

- Network Slicing (i.e., URLLC, eMBB, and mMTC) to support different types of services.
- Network Resilience using 5G islands and FM methods.
- Network Reliability by means of Multi-connectivity and Network coding techniques.

Table 3-5 provides the mapping of the use cases to the corresponding 5G-MoNArch innovations implemented in Smart Sea Port testbed.

Table 3-5: The selected 5G-MoNArch innovations for the Smart Sea Port testbed

Use case	Innovation	Deployment
Improve port logistics traffic management	• Network slicing: Orchestrator and edge/central loud network functions mapping • Network reliability: RAN multi-connectivity • Network resilience: fault management	• Traffic lights connected through reliable wireless links. • Intelligent traffic signal control (highly reliability, low throughput URLLC service)
Improve port operations using AR/VR and video streaming	• Network slicing: Orchestrator and edge/central cloud network functions mapping • Network resilience: fault management	• eMBB service supporting 4k+ video (high throughput service, but not necessarily low latency) • Data analytics service to provide end-users with visualisation and awareness of the port status at different levels
Reduce port pollution using 5G technology	• Network slicing: Orchestrator and edge/central cloud network functions mapping • Network reliability: RAN multi-connectivity • Network resilience: fault management	• Machine type communication slice (low throughput from multiple mobile terminals, but high reliability for mobility required) for measurements on, e.g., air pollution through wireless connected sensors on mobile barges or at stationary locations

On the other hand, based on the specific use cases described in Section 2.2, the Touristic City testbed aims to demonstrate the following features:

- Flexible creation of a specific eMBB and URLLC slices including the configuration of 5G New Radio (NR) (e.g., robust numerology for tactile closed loop control) and 5G Core (5GC).
- Resource elasticity upon changes of the overall network conditions (e.g., latency).
- Solutions related to the dynamic placement of functions (orchestration-driven elasticity).
- Assessment of the system's KPIs in terms of latency and throughput and QoE.
- Suitable interface for real-time monitoring of the network status, computational load and latency.
- Control interface to reconfigure/re-parametrise NFs of 5G NR and 5GC.
- Flexibility to adapt / integrate control and decision algorithms devised within WP4.

Table 3-6 lists the selected 5G-MoNArch innovations of the Touristic City testbed.

Table 3-6: The selected 5G-MoNArch innovations for the Touristic City testbed

Use case	Innovation	Deployment
Enhanced touristic experience	• Network slicing: VNF Orchestration and edge/central cloud network functions mapping • Resource elasticity: Orchestration-driven elasticity • Experiment-driven optimisation	• Client-Server deployment providing a VR visit of a museum room with the interaction between the tourist (wireless 5G connection) and the guide (wired, Ethernet connection) located in two different places.

4 Requirements and KPIs

Building on the use cases defined in Chapter 2, this chapter provides an overview of the qualitative functional and operational requirements and, if applicable, quantitative performance KPIs that are relevant for these use cases, and in how far these requirements and KPIs are addressed in the corresponding testbeds.

To this end, the requirements and KPIs identified in D6.1 [5GM17-D61] are considered. The objective of this chapter is to set the focus on the subset of requirements and KPIs that are related to the 5G-MoNArch technologies and features to be deployed in the testbeds, and therefore serves as a baseline for the future assessment in how far the testbed implementations have been successful.

Due to the nature and limited scope of a testbed, it is not possible to prove all the defined 5G requirements and KPIs as defined in D6.1. Indeed, the focus of the testbeds is on showcasing and test the key technologies and features developed within the scope of 5G-MoNArch, which correspond to a subset of the overall 5G requirements and KPIs and primarily requires a qualitative assessment. Some of the 5G KPIs may be indirectly addressed based on the measurement results recorded during the operation of the testbed.

The KPIs addressed by the testbed fall within the general context of the verification and validation activities of 5G-MoNArch. The objective of this activities is to demonstrate and prove – in real-world 5G deployments – the feasibility and operation of network slicing and the capabilities such as resilience, security, low latency and resource elasticity that allow supporting service types such as eMBB, MTC and URLLC. Furthermore, it is shown that the key innovations of 5G-MoNArch meet the requirements and provide benefits to stakeholders from both a functional as well as an economical perspective.

The results of WP5 in terms of KPI's satisfaction will be provided to WP6, which (as part of 5G-MoNArch project has) defined the methodology and framework that is used to evaluate, verify and validate the architecture and features based on the two testbeds deployed by the project. This allows to qualitatively and, where applicable, quantitatively evaluate the operation and performance of the proposed innovations, it validates and verifies the overall architecture, and it provide a techno-economic analysis of the project's concepts and results. Part of this methodology consists in the definition of the evaluation cases, selection of enablers chosen for verification, and the definition of the requirements and KPIs that shall be fulfilled. These requirements and KPIs are divided to different categories: functional, operational, and performance.

In the following sections, the requirements and KPIs are provided along the following categories: Section 4.1 introduces the general requirements on 5G, Section 4.2 requirements relevant for network reliability and resilience, and Section 4.3 the requirements relevant for resource elasticity.

4.1 Requirements on general 5G operation

This section provides the qualitative functional and operational requirements corresponding to general aspects of a 5G network. These include features such as control/data plane separation, support for network slicing, support for multi-tenancy, or technologies such as NFV or software defined networks (SDN). Out of the complete list of requirements and KPIs provided in D6.1, this section focuses only on those requirements that are relevant for the two testbeds (resilience and security, and resource elasticity), and are qualitatively demonstrated in at least one of them. An overview about the requirements addressed by each testbed is provided in Table 4-1.

The following general qualitative functional and operational requirements are relevant for the two testbeds:

1. *The 5G system shall leverage novel technology enablers (e.g. NFV and SDN) to reduce the total cost of ownership and to improve operational efficiency, energy efficiency, and simplicity and flexibility for offering new services.*
2. *The 5G system shall support the concept of dedicated network slices, understood as the allocation of dedicated network resources to serve defined business purposes, customers, or use cases.*
3. *The 5G system design shall support infrastructure sharing and multi-tenancy.*

4. *The 5G system shall support the separation of UP network functions (UPNFs) from CP Network Functions (CPNFs), allowing independent scalability, evolution and flexible deployments, e.g. centralised location or distributed (remote) location.*
5. *The 5G system shall allow a modularised function design (based on the decomposition of RAN and CN NFs), e.g., to enable flexible and efficient adaptive placement of NFs at different locations within the network infrastructure. This is a pre-condition for network slicing.*
6. *The 5G system shall allow for deployment flexibility e.g. to host relevant RAN and CN NFs and application functions close together at the edges of the network, when needed, e.g. to enable context aware service delivery, low latency services, etc.*
7. *NFs shall be enabled to interact with other NFs directly if required. The architecture should not preclude the use of an intermediate function to help to route CP messages.*
8. *The 5G system shall be able to manage both VNFs and PNFs.*
9. *FCAPS management, as well as lifecycle management (LCM) of VNFs, shall use the ETSI NFV MANO architecture [ETSI14] [ETSI17] as baseline.*
10. *The 5G system shall allow operators to optimise network behaviour (e.g. mobility management support) based on the mobility patterns (e.g. stationary, nomadic, spatially restricted mobility, full mobility) of a single UE or a group of UEs.*
11. *The 5G system shall efficiently support network resource utilisation and optimisation based on system information (context awareness), providing mechanisms to collect such information (e.g. network conditions, information on served UEs, user subscription profiles, application characteristics) within an operator configured time scale.*
12. *The 5G system shall support different levels of resilience for the services provided.*
13. *The 5G system shall allow flexible mechanisms to establish and enforce priority policies among the different services and users (subject to regional or national regulatory and operator policies).*
14. *The 5G system shall be able to provide the required E2E QoS (e.g. reliability, latency, and bandwidth) for a service and support prioritisation of resources when necessary for that service.*

In addition to the above general requirements (1-14), the following qualitative functional and operational requirements (15-22) on network slicing are addressed by the testbeds:

15. *The 5G system shall allow the operator to create, modify and delete an NSI or NSSI, and to define and update the set of services and capabilities supported in an NSI.*
16. *The 5G system shall allow the operator to configure the information which associates a UE or a service to an NSI.*
17. *The 5G system shall enable a UE to be simultaneously assigned to and access services from more than one NSI of one operator.*
18. *Traffic and services in one NSI shall be logically isolated from traffic and services in other NSIs in the same network.*
19. *Creation, modification and deletion of an NSI shall have no or minimal impact on traffic and services in other NSIs on the same network.*
20. *The 5G system shall enable the network operator to define a minimum available capacity as well as a maximum capacity for an NSI. Scaling of other NSIs on the same network shall have no impact on the availability of the minimum capacity for that NSI.*
21. *The 5G system shall enable the network operator to define a priority order between different NSIs in case multiple NSIs compete for resources on the same network.*
22. *The 5G system shall support means by which the operator can differentiate policy control, functionality and performance provided in different NSIs.*

The following qualitative functional and operational requirements (23-26) related to network capability exposure and UP handling in the case of local service hosting are also relevant for the defined use cases:

23. *The 5G system shall support concurrent access to local and centralised services. To support low latency services and access to local data networks, UPNFs may be deployed close to the RAN.*
24. *The 5G system shall enable a service hosting environment provided by an operator, support configuration of that environment and be able to interact with applications in that environment for efficient network resource utilisation and possible offloading of data traffic.*
25. *Based on operator policy, application needs, or both, the 5G system shall support an efficient UP path between UEs attached to the same network, modifying the path as needed when the UE moves during an active communication ("UE-specific UP path").*
26. *Based on operator policy, the 5G system shall be able to support routing of data traffic between a UE attached to the network and an application in a service hosting environment for specific services, modifying the path as needed when the UE moves during an active communication ("service mobility").*

Table 4-1: General requirements addressed in testbeds

No.	Addressed by Smart Sea Port testbed	Addressed by Touristic City testbed
1	Yes (NFV and SDN)	Yes (NFV and SDN)
2	Yes, including full slice LCM (slice configuration, instantiation, operation, deletion)	Yes, there will be two network slices dedicated to URLLC and eMBB services
3	Yes	Yes, infrastructure sharing
4	Yes	Yes, especially in the CN
5	Yes (flexible placement of CN NFs at different locations: Edge / Central Cloud)	Yes, to enhance their flexible placement
6	Yes (flexible placement of CN NFs at different locations: Edge / Central Cloud)	Yes, supported with elastic orchestration algorithms
7	Out of scope	Yes, especially in the core through the SBA
8	Yes (CN fully virtualised, RAN only PNFs)	Yes. Lower radio layers are PNFs, the rest are VNFs
9	No – baseline is the 5G-MoNArch concept for slice lifecycle management, but a dedicated VIM used for this testbed	Yes, but implementing 5G-MoNArch enhancements e.g. for the orchestrator
10	No	Large scale evaluation of mobility aspect is left to WP6
11	No	Yes. Intelligent orchestration mechanisms will be based on the availability of probed data
12	No, as only a single radio site is used, and no redundant transport network is available	No, the testbed is focused on elasticity
13	Yes, by using different 5G Quality of Service Indicators (5QI) e.g. for the URLLC and mMTC slices, and by enabling the enforcement of different priorities	Yes, slices will be separated directly at the spectrum level
14	Yes, by defining different QoS properties for the different slices in the radio, and by enabling the enforcement of different priorities	Yes, through elastic algorithm at cross-slice level
15	Yes	Yes, the NSI onboarding will be an automated process
16	Yes	No, due to the controlled environment, the number of kind of UEs are limited

17	Yes	Yes, UEs will connect to both slices
18	Yes	Yes, in both the RAN and CN
19	Yes	Yes, slices will be independently orchestrated
20	Yes, but this is only implemented per slice type but not per NSI	Yes, although not for radio resources
21	Yes	Yes, although not for radio resources
22	No	Yes, core functions and orchestration will have a software defined approach
23	Yes	Yes, for the URLLC slice
24	No	Yes, orchestration will be based on open source software
25	No	Yes, UEs belonging to the URLLC slice will have optimised UP NFs for low latency
26	No	Yes, the haptic feedback service will follow this paradigm

4.2 Requirements on network reliability and resilience

In addition to the general requirements described in the previous section, there are some additional requirements that are specific to each of the testbeds. This section provides the specific requirements of the Smart Sea Port testbed, related to reliability, resilience and security.

The functional and operational requirements related to network reliability and resilience (corresponding to the concepts and features developed in WP3) are demonstrated by the Hamburg Smart Sea Port testbed. High network reliability and resilience are especially important in critical communication type services, which is the case for industrial URLLC and mMTC slices. An example for such an URLLC slice is the traffic light control functionality in the Smart Sea Port testbed, and an example for an mMTC slice are the sensor-based environmental measurements on the mobile barges. In the latter case, the mobility plays an important role and requires dedicated features to achieve the requirements.

Network reliability and resilience are assured at multiple levels: (i) RAN, (ii) transport network and (iii) CN. 5G-MoNArch pays special attention to RAN reliability that can be increased by multi-connectivity and network coding. It can be measured by the rate of package loss in the network. Telco cloud resilience is an important factor to provide and maintain an acceptable quality level for services in case of faults. This is supported through the implementation of FM mechanisms.

In the following, the subset of qualitative requirements from D6.1 relevant for the operation of a reliable and resilient network according to the Smart Sea Port testbed is provided. For each of the requirements, a description is first given, followed by an explanation how the requirements are addressed within the testbed.

1. *It shall be possible for efficiency purposes to run one or more NSI with varying service characteristics on the same frequency band by sharing of time-frequency resources.*

   The Smart Sea Port testbed implements network slicing and enables the instantiation, deployment, operation and deletion of customised network slices with different service characteristics, in particular, an URLLC slice type for the traffic light control, an eMBB slice for the AR/VR-based planning and construction application, and an mMTC slice type for the environmental measurements.

2. *The 5G RAN shall natively and efficiently support multi-connectivity, i.e., the case when a UE is connected to more than one transmission-reception point (TRxP) (inter-site, i.e., not co-located) and/or more than one air interface variant (which may be co-located or not). Multi-connectivity shall be supported for both throughput increase via aggregation of parallel data flows as well as for link reliability improvement via data duplication and/or network coding features.*

The Smart Sea Port testbed implements a dual-connectivity approach with data duplication in order to increase link reliability and to prevent service interruption during mobility events, i.e., handovers. This is particularly shown through the mMTC slice type for the environmental measurements, where the UEs are installed on moving barges.

3. *The design of the 5G RAN (inclusive of radio protocol stack) as well as the deployment options resulting from this, shall allow the minimisation of the radio link outage probability with the aim to achieve high reliability and availability values targeted especially by vertical industries for URLLC use cases.*

   As mentioned in point 2 above, the Smart Sea Port testbed applies a dual-connectivity approach, which represents a form of macro-diversity using distributed antenna systems in order to prevent link outages.

4. *The telco cloud to be integrated into the 5G system shall be designed in a resilient way such that it prevents the network performance to degrade or at least keeps the degradation on a minimum level.*

   The Smart Sea Port testbed allows for customising the NF placement within the CN, thereby addressing resilience QoS requirements. Furthermore, fault detection mechanisms are implemented to timely identify degradations.

Beyond the conceptual network reliability and resilience requirements described above, the Smart Sea Port testbed allows to validate some testbed-specific qualitative functional and operational requirements as well. From the network reliability and resilience-related requirements provided in D6.1 [5GM17-D61] (Section 5.2), those being of particular relevance for the Smart Sea Port testbed are listed in the following:

5. *The Smart Sea Port testbed shall provide end-to-end reliability in terms of packet error rate, in particular for mobile UEs.*

   The testbed implements a dual-connectivity approach with data duplication in order to increase link reliability and to prevent service interruption during mobility events, i.e., handovers. With this approach, it is possible to minimise the maximum latency in a deterministic way and prevent from handover-induced latency peaks and hence packet loss. Such mechanism is not available in, e.g., LTE networks. This is particularly shown for the mMTC slice type for the environmental measurements, where the UEs are installed on moving barges.

6. *The Smart Sea Port testbed shall provide resilience in the telco cloud in terms of packet error rate.*

   The testbed implements measurement processes and procedures that allow to identify the impact of network failures on the operation of the mobile network. However, it is not planned to implement mechanisms that modify the configuration of the network such that the failure cause can automatically be resolved.

7. *The Smart Sea Port testbed shall provide service continuity for all implemented services*

   The testbed implements processes and procedures that allow to measure the end-to-end latency as well as the packet error rate, during failure events as well as during mobility events. From the acquired measurements it is possible to infer on the service continuity of the URLLC, eMBB and mMTC services. However, it is not planned to implemented mechanisms that modify the configuration of the network such that potential shortcomings in the service continuity can be resolved automatically.

4.3 Requirements on resource elasticity

This section provides the specific functional and operational requirements related to resource elasticity, which are relevant for the Touristic City testbed. Resource elasticity is defined as the ability to gracefully adapt to load changes in an automatic manner such that at each point in time the available network resources match the services’ demand as closely and efficiently as possible. The corresponding requirements are specifically tailored to the cloud infrastructure. In particular, there are a number of key requirements that have to be addressed when dealing with elasticity:

1. *Some of the VNFs that compose an elastic network slice shall consider elasticity since the very beginning.*

   Some of the VNFs included in the Turin city testbed have specific interfaces to the Orchestration platform to enable their elastic orchestration.

2. *The radio access network functions shall support 5G Numerology and efficient bandwidth partitioning.*

   The designed radio access functions are optimised for the specific services included in the testbed.

3. *Orchestration algorithms shall be aware of the characteristics of the infrastructure a network slice is running on.*

   In order to take advantage of potential multiplexing gains, orchestration allocates resources to VNFs when it is most suitable (e.g., low latency NFs shall n be located close to the edge) and eventually re-orchestrate them when changes take place.

4. *The orchestration algorithms shall rely on precise information about the computational profile of VNFs.*

   One of the inputs provided to the orchestration framework is the fine-grained information about how a VNF implementation reacts to different loads in terms of CPU or memory consumption. This allows for a more accurate orchestration of resources, especially in locations where resources are more expensive (i.e., in the edge).

5. *Fine grained monitoring of network slice status shall be available.*

   The VNFs included in the network slices will support an efficient use of resources by means of infrastructure monitoring (i.e., spectrum, CPU, memory, bandwidth) and VNF internals (i.e., used physical resource blocks, number of active users, …)

6. *Algorithms shall take advantage from past utilisation statistics.*

   The orchestration algorithms developed for the Touristic City testbed will rely on Machine Learning (ML) techniques that, building on the past values, will devise optimal resource assignment patterns.

4.4 Testbed requirements

In this section, further information is provided on requirements and test procedures that go beyond those provided from Section 4.1, 4.2 and 4.3. These requirements are related to the specific technical implementation of the two testbeds (which have completely different characteristics), the commercial interests of the involved partners (related to the respective exploitation plans), the operation of the testbeds, and – where applicable – the applications that shall be implemented.

4.4.1 Smart Sea Port testbed

Besides the functional requirements provided in Sections 4.1 and 4.2, there are a number of additional functional and non-functional requirements for the Smart Sea Port testbed. This testbed aims at providing a networking environment applicable in particular to industrial wide-area services and applications. This means that the testbed implements a macro outdoor network coverage such that network services can be provided to multiple different applications within a rather large area of the Hamburg port. Furthermore, the testbed network will support macro mobility of the UEs.

With respect to the applications implemented in the testbed, these are real-world industrial applications related to the day-to-day operation of the Hamburg port. All these applications can only be realised through a 5G mobile radio network, as their service characteristics and the related QoS requirements cannot be fulfilled by State of the Art(SotA) networks (3G, LTE).

Based on the above, the following further functional and operational requirements are applicable to the Smart Sea Port testbed:

1. *The Smart Sea Port testbed shall proof and showcase the technical feasibility of network slicing on commercially available equipment with certain enhancements.*

The testbed has been planned and implemented using commercial networking equipment (Nokia AirScale base station, Nokia AirFrame servers) as basis, and has been integrated with the transport network and data centre infrastructure of DT.

2. *The Smart Sea Port testbed shall implement, proof and showcase solation in terms of isolation between logical networks as well as resource isolation*

   To prove isolation between logical networks (and data privacy), the mechanisms to separate and isolate data have to be described and implemented for the whole end-to-end communication paths. For testing resource isolation mechanisms, it may be necessary to apply special testing procedures to cause a reaction of the system proving that the expected functionality is really working as expected. For instance, to test resource isolation, it may be necessary to generate and inject additional test traffic to operate the system at high traffic load.

3. *The Smart Sea Port testbed shall proof and showcase the ability of the 5G mobile network architecture, network slicing, and the reliability and resilience features developed in 5G-MoNArch to support actual real-world industrial use cases using the specific example of a Smart Sea Port (cf. Section 2.1.2).*

   The selected use cases (URLLC: traffic light control: mMTC: sensor-based environmental measurements on mobile barges; eMBB: AR/VR-enabled construction and maintenance support for port operations) have been directly defined by HPA as the vertical representative of this testbed. During the operation of the testbed, and the corresponding application and service tests, HPA can directly evaluate if the intended functionality can be achieved, and if the performance of the applications and services corresponds to their requirements.

4. *The Smart Sea Port testbed shall implement, proof and showcase processes and means for network slice lifecycle management, i.e., the definition, instantiation, management and deletion of network slices.*

   A corresponding GUI-based testbed operations application is currently developed, that allows for remotely performing all tasks related to slice LCM.

4.4.2 Touristic City testbed

The Touristic City testbed also has some specific functional requirement related to the specific services provided by that testbed. The main goal of the testbed is to provide enhanced services such as VR and haptic communication, which impose severe and diverse requirements on the network that would be impossible to achieve on a non-sliced, traditional network. In particular, VR services require very high bandwidth while haptic communications require ultra-low latency.

In particular, the Touristic City testbed includes network slicing-enabled network functions for the radio access and an orchestration platform that fulfils the elastic requirements described in Section 4.3. In this context, the following additional requirements can be identified:

1. *Different network slices should run on the same infrastructure.*

   The devised cross-slice orchestration algorithm is capable of re-assigning resources to network slices when and where it is needed, possibly using unutilised resources from idle network slices, in order to maximise multiplexing gains (and network monetisation). The Touristic City testbed contains two network slices: an eMBB slice type for demonstrating AR/VR application and an URLLC slice type for tactile feedback.

2. *The Touristic City scenario shall showcase the advantages of the elastic management of the network.*

   The Touristic City testbed serves as direct outcome of the elastic network solutions envisioned and designed by WP4. More specifically, the orchestration architecture includes the elastic elements designed such as the ML-based ones. As a matter of fact, the Touristic City testbed serves as Proof of Concept for some of the use cases defined by the ETSI Experiential Networked Intelligence (ENI) working group, that is precisely working on these topics.

3. *The URLLC slice shall enable haptic feedback communications.*

Providing haptic feedback is probably one of the most challenging services, as it requires very specific orchestration patterns and resource assignment such as dedicated bandwidth in the radio access network and the instantiation of network functions closer to the users. The Touristic City testbed provides slice-aware RAN functions that are properly configured to support an URLLC slice (with specific 5G numerology) by the orchestration framework.

4. *Seamless orchestration of network functions*

   In the implementation of network elasticity, it is necessary to be able to re-orchestrate network functions on the fly to adapt to changes in the network and the user traffic. To this end, orchestration mechanisms are required that are capable of re-locating network functions with very small latencies and without affecting the ongoing services on the network. This will require novel orchestration mechanisms for the live migration of network functions.

5. *Data analytics and AI engine for orchestration*

   The decision on network orchestration will be taken based on applying artificial intelligence / machine learning techniques to collected data, in line with the ENI architecture of ETSI. To this end, machine learning algorithms and suitable interfaces are necessary that are capable of collecting the required data and learning the most appropriate orchestration decisions.

5 Description of testbeds implementation and setup

To realise the use cases described in Chapter 2, it is necessary to deploy the network infrastructure comprising the different hardware components, develop and integrate the required software modules and perform the corresponding tests and show cases. In this chapter, this is described for each of the two 5G-MoNArch testbeds.

5.1 Smart Sea Port testbed

5.1.1 Context of the testbed

The Smart Sea Port testbed in Hamburg is intended to trial network slicing and multi-connectivity in the context of industrial applications. It proves that essential components of the architecture concept for network slicing described in [5GM17-D21] and in Section 3.1.1 can be implemented with 5G-MoNArch technology. This demonstrates the potential that network slicing technology provides to industrial applications. Furthermore, the testbed is used to explore multi-connectivity mechanisms for increased E2E reliability in a realistic environment.

In addition to the requirements resulting from the 5G-MoNArch research objectives and already stated in Section 4.2, some working assumptions and constraints have to be taken into account in the design of the testbed:

- Commercially available equipment shall be used as far as possible; it has to be modified to implement features specifically needed for network slicing. This applies in particular to the radio base station and the terminals.
- For testing network slicing, the radio standard of the air interface is of secondary importance. 3GPP Rel. 14 / LTE-Advanced Pro are sufficient, and 3GPP Rel. 15 / 5G NR is not required. As a matter of fact, at the time of the testbed installation, 5G NR was not yet available in commercial products, and hence an LTE-Advanced Pro compliant air interface is used in the testbed.
- The (geographical) distance between the Edge Cloud data centre and the Central Cloud data centre shall be sufficiently large to demonstrate the benefits of an Edge Cloud deployment over a fully centralised network deployment.
- Since this testbed is a network for research purposes, it has to be expected that it cannot yield similar stability, reliability and availability as a regular production network. Note that this lower reliability is merely caused by the fact that this is a research testbed and thus des not contradict to the fact that 5G-MoNArch technology aims at improving the reliability of mobile radio networks. This shall be taken into account in the selection of applications / use cases.
- For security and privacy reasons, the testbed will be set up in an isolated way, without connection to any other publicly accessible communication network. The usage will be restricted to the testbed activities and conducting the corresponding test cases.

5.1.2 Infrastructure setup

5.1.2.1 Site and radio frequency planning

Figure 5-1 shows the main physical components of the testbed. The characteristics of all these components are described next.

Terminals: In order to implement the required network slicing functionality together with the reliability and resilience features, the testbed uses commercial router boards with LTE modems, and Software Defined Radio (SDR) terminals with small-form-factor PCs. Only these terminals allow to exploit the full set of features required for the three slices of the Smart Sea Port testbed. Depending on the installation site (traffic light control switch box, or barges), different commercial antenna types will be used to connect the terminals to the base station.

Base stations: The testbed comprises two base stations that cover the Hamburg port almost completely. The eNBs are Nokia AirScale base stations with the following technical data:

- Radio standard: LTE-Advanced Pro / 3GPP Rel. 14 compliant.
- Radio parameters:
    - FDD
    - Single Carrier, bandwidth 10 MHz
    - DL: 713 – 723 MHz
    - UL: 768 – 778 MHz

The base stations and antennas are mounted at the Heinrich-Hertz Tower in Hamburg. Radio components are mounted aside the antennas and connected via optical fibre to the base band unit. The antenna is installed at a height of 182 metres above ground and there are two radio cells, one per eNB.

Edge Cloud: The Edge Cloud consists of two Nokia Airframe servers that are mounted in the data centre of DT in Hamburg.

Central Cloud: The Central Cloud comprises servers of DT in a data centre in Nuremberg, Germany. The geographical distance is approx. 460 km and thus sufficiently long for a realistic assessment of the benefits of Edge Cloud deployments.

5.1.2.2 Transport network and protocol architecture

Figure 5-1 shows the transport network connections between the physical components described before. All transport network connections consist of optical fibres with a capacity of 1 Gbit/s.

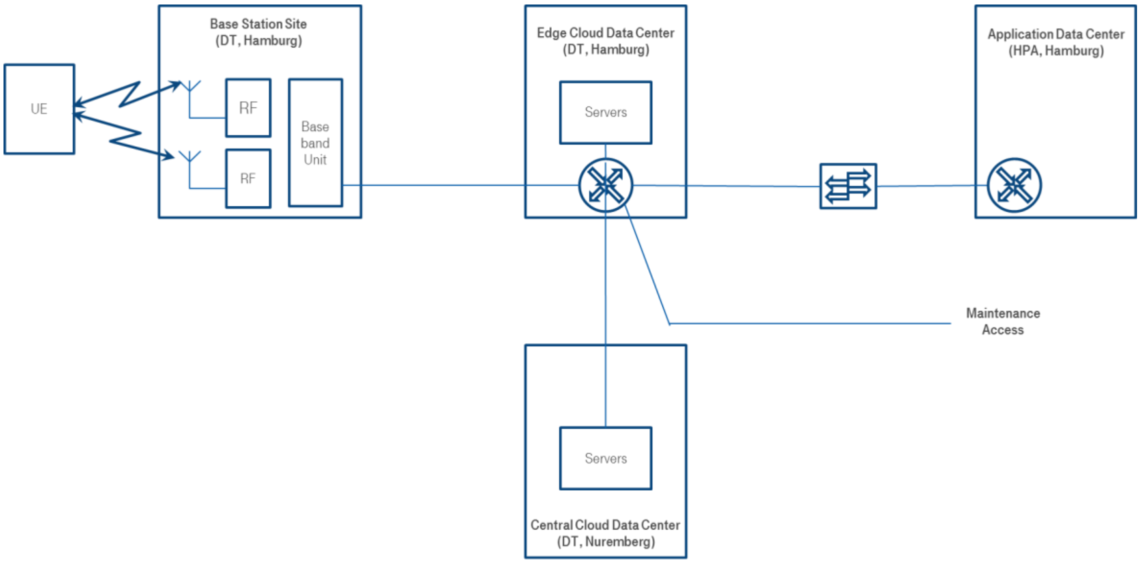

Figure 5-1: Physical components of the Smart Sea Port testbed

The main task of the transport network is to exchange user data and control data between the network sites mentioned in the previous section respectively between the NFs deployed in these sites. In addition to the typical requirements, network slicing requires that the data from different slices are handled strictly isolated on all protocol layers (i.e., transport network and above).

Figure 5-2 shows a simplified protocol structure of the user plane communication path from the UEs to the application servers of HPA.

The protocol stack builds on the legacy stack as standardised by [3GPP18-23401], with additional layers on top of that. The transport network protocols are shown in red colour for the MNO part; protocols for data transport in the tenant's data centre are depicted in blue. UDP / TCP layers on top of the IP layer have been omitted for simplicity. The radio protocol stack for communication between terminal and eNB is shown in green colour. These protocols are based on LTE Rel. 14, with some modifications for the support of network slicing.

Yellow colour has been used for the protocols under control of the Evolved Packet Core (EPC; 4G CN), namely the GTP tunnelling protocol and the IP on top of it. In this IP layer, a single IP tunnel per mobile terminal (not per sensor) is provided. Since this IP connection is terminated in both the mobile terminal as well as at the SGi interface behind the EPC, the address space of this IP connection can be assigned independently from the transport network (red).

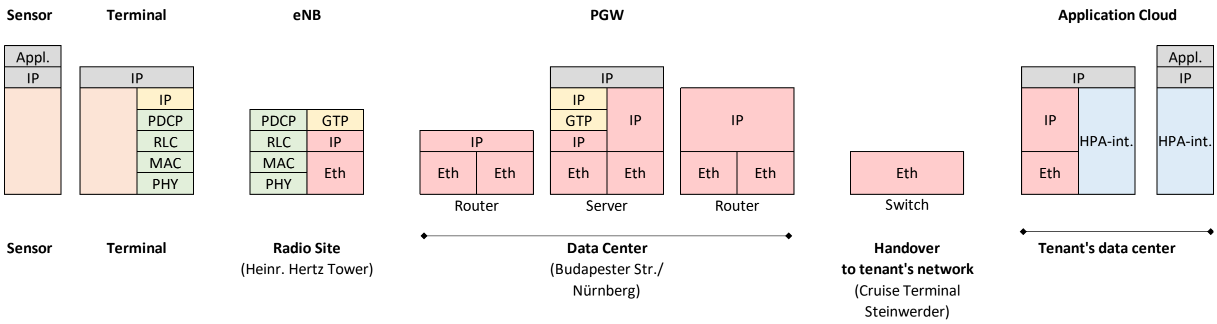

Figure 5-2: Simplified user plane protocol structure for communication between UEs and application servers.

Since multiple sensors can be connected to the same mobile terminal and since each of them shall have its own IP address (allowing the application server to access each sensor individually), the grey IP layer has been added on top of the legacy LTE protocols. The mobile terminal acts as router that routes IP packets coming from the application server on the right side to one of multiple sensors connected to the same mobile terminal. Since the "grey" IP packets are treated as payload of the yellow / red / blue network, the address range for the grey IP layer is independent from the yellow and the red IP layers.

So, from an application viewpoint, the network slice looks like an IP network with both the sensor on the left side and the corresponding server on the right side acting as end devices and the mobile terminal and the Packet GateWay (PGW) in the EPC acting as IP routers.

A key requirement arising from the network slicing paradigm is the traffic isolation between the slices:

- The mobile terminal is authenticated by the eNB, and traffic is encrypted on the air interface. In this way mixing of traffic originating from different mobile terminals on the air interface is prevented, although all mobile terminals are sharing the same radio resources.
- The eNB then forwards the traffic from the mobile terminal to the PGW assigned to the respective network slice via a point-to-point IP connection in the transport network. Since the IP addresses for these connections are assigned by the network operator and cannot be modified by the mobile terminal, IP packets from different slices are routed to different PGWs and hence traffic from these slices is kept separate. As an additional security measure, the network operator can apply encryption at the IP layer for this hop.
- The IP addresses of the SGi interface of the PGW and the slice gateway into the tenant's data centre are assigned by network operator and tenant, respectively, and cannot be modified by the mobile terminal. In this way, the connection from PGW to the gateway into the tenant's data centre is also a static point-to-point IP connection and traffic from one slice cannot be injected into another slice.

5.1.3 Hardware and software modules

5.1.3.1 Physical deployment

The actual physical deployment of the Smart Sea Port testbed located in Hamburg is described in Figure 5-3. The radio access hardware is installed at the Hamburg TV tower (Heinrich Hertz Tower) at an elevation of 182 m above ground, using two sectorised antennas, operating at 700 MHz and using a Nokia AirScale base station. The radio equipment is connected to the DT data centre in Hamburg where 2 Nokia Airframe servers are deployed as well as with the DT data centre in Nuremberg where additional VMs are deployed. Furthermore, HPA provides access to a traffic light switch box as well as two mobile barges, which are cruising within the port area. The devices installed at the traffic light switch box and on the barges will be connected wireless to the corresponding application running locally at the operations centre of HPA. Due to the particular requirements of the use cases, and the 5G-MoNArch features to be implemented, customised devices will be used, see Section 5.1.2). The installation and commissioning of the devices has already started for the URLLC and the mMTC use cases, the mounting on site took place in July and August 2018.

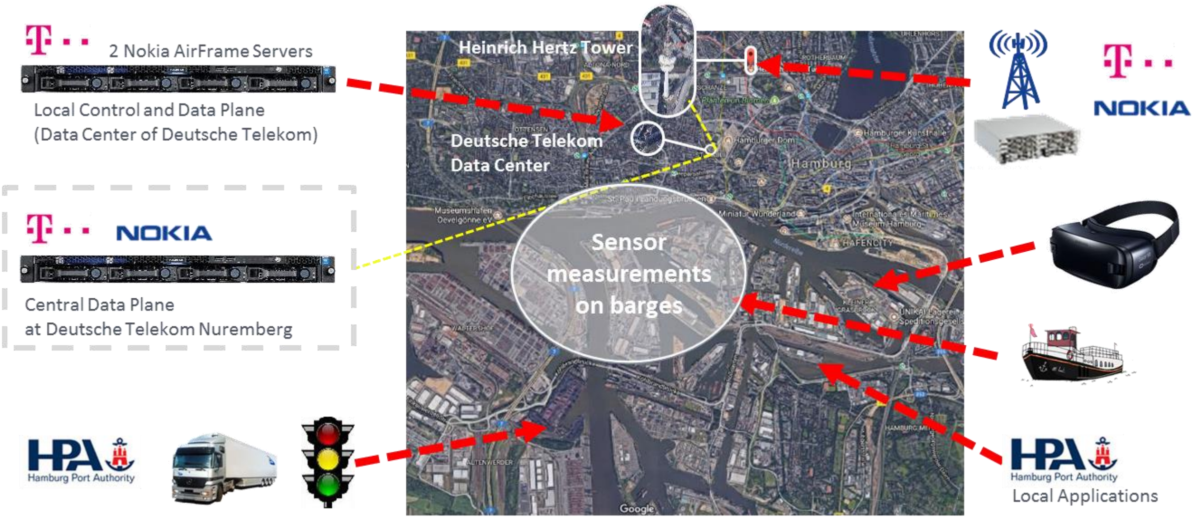

Figure 5-3: Physical deployment of Smart Sea Port testbed

5.1.3.2 Logical deployment

Based on the physical deployment described in the previous section, the logical setup shown in Figure 5-4 is implemented. In general, the testbed shows network slicing in an actual industrial deployment (see Section 2.1 for more details) and the increased robustness of mobility due to multi-connectivity. In particular, the testbed deploys multi-slice capable devices which connect to more than one network slice (logical network) providing different and isolated services. Those services may be performed either locally (local UP) or at a more central location (remote UP). The network slices will be set up E2E, i.e., both RAN and CN will apply slice-specific configurations and functionality depending on the slice-type (eMBB, URLLC, and mMTC services).

Furthermore, CP and UP are separated and only a single CP instance is deployed. The management and control of UP and CP are performed by a dedicated LCM, which allows for defining, commissioning, and managing the installed network slices. The testbed is operated in an isolated network in order to avoid any kind of cross-talk with services using actual customer or provider information.

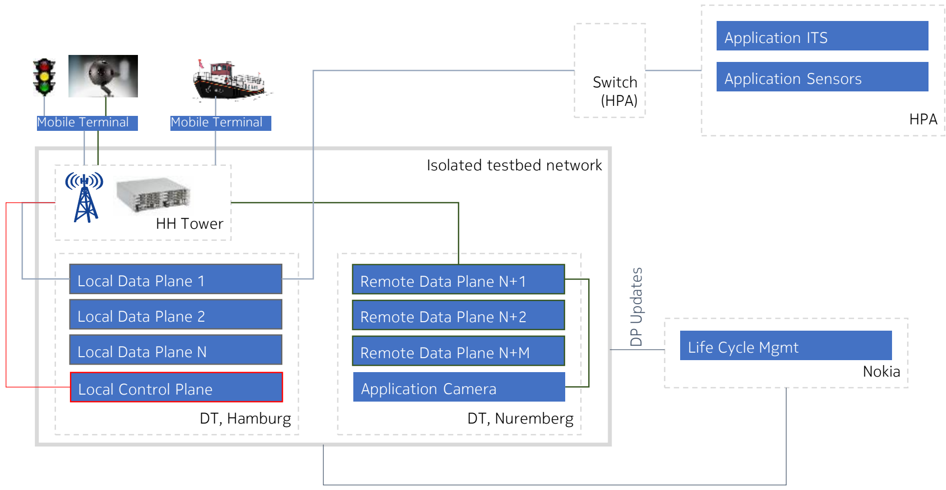

Figure 5-4: Logical deployment of Smart Sea Port testbed

5.1.3.3 Software deployment

Based on the above describe physical and logical deployment, the following software components are deployed and modified.

1. **Mobile terminal:** The terminal uses Android OS and is modified to support network slicing via dedicated OS calls (no Kernel or OS modifications are planned). It further provides an interface for defining and setting up the networking slices, which are accessed. The mobile terminal can be accessed remotely using a host-laptop connected to the mobile terminal.
2. **Radio access network:** The RAN is modified to provide individual and customised radio network configurations/operation based on the individual network slices currently active.
3. **Core network:** The CN is split in CP and UP components, which are independently deployed across different physical locations. Each UP instance is associated with a single network slice, while all network slices and user terminals are under a common CP instance. The instantiation of network slices is dynamic.
4. **Lifecycle management:** At M&O layer, a LCM interface (see Figure 5-5) is provided to define, commission, and monitor the individual network slices including all of the above-mentioned components. The interface allows for setting up network slices, assigning them to individual network slice categories, associating devices with these network slices, and to monitor the operation of the network slices. The underlying LCM tool is specifically developed by Nokia for the purpose of the Smart Sea Port testbed, providing the basic required functionality. Figure 5-5 shows a screenshot of the planned Graphical User Interface (GUI), which is used to control the network slice LCM. This GUI will also provide information about relevant KPIs as detailed in Chapter 4.

Figure 5-5: Screenshot of GUI used to control the network slice lifecycle management

5.1.4 Setup Use Cases

The deployment of the Smart Sea Port testbed infrastructure to implement the three use cases takes place in HPA's data centre, HPA's Cruise Centre Steinwerder (CC3), at the traffic light location in the port area and on the barges.

5.1.4.1 Use Case 1 – Traffic management

The challenge on traffic light hardware integration is the highly crucial nature of traffic management systems. Therefore, the 5G UE must be integrated with certified traffic light controllers in a safe and secure manner. The physical hardware is integrated with the following equipment:

- The 5G-Modem will be integrated in the control cabinet, being connected via Ethernet to the control unit (see Figure 5-6).
- DT's infrastructure is routing the network traffic to HPA's network.
- The operational data of the traffic light is integrated in HPA's Port Road Management Centre (see Figure 5-7).

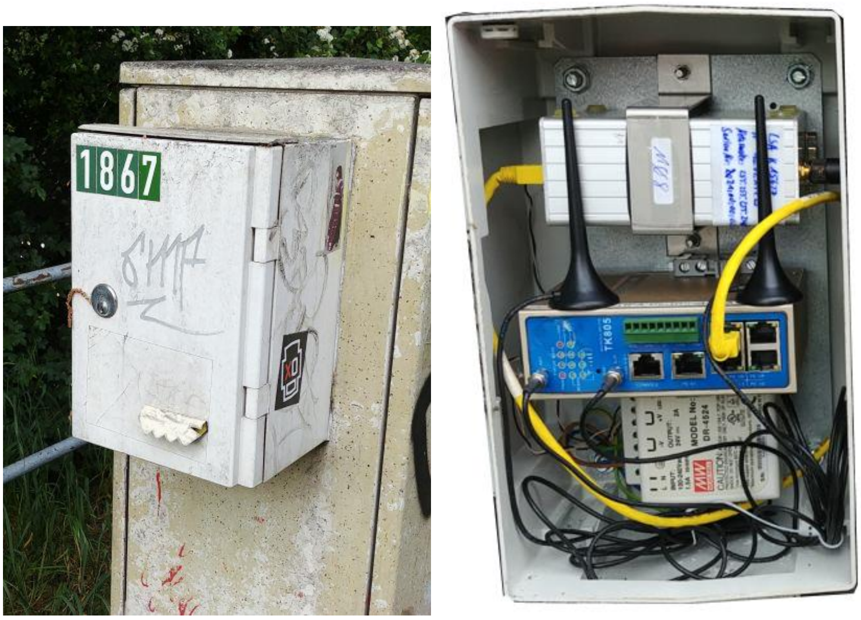

Figure 5-6: Hardware and containment for traffic light integration

Figure 5-7: HPA's Port Road Management Centre (PRMC) visualising traffic light status

5.1.4.2 Use Case 2 – 5G with HoloLens for better planning and construction

The setup of the second use case is the least challenging one, but it therefore represents a "rapid deployment" scenario of 5G network slices. The AR/VR-Hardware is capable of Wi-Fi-connections. Therefore, the system is set up by using Wi-Fi-tethering of the experimental UEs. The AR-Headset is integrated with the following steps:

- Connection of the Microsoft HoloLens via 5G;
- 5G-Smartphone-Prototype enables connectivity via Wi-Fi-tethering;
- HPA-Engineers will use the equipment to be supported in activities in the field (Figure 5-8).

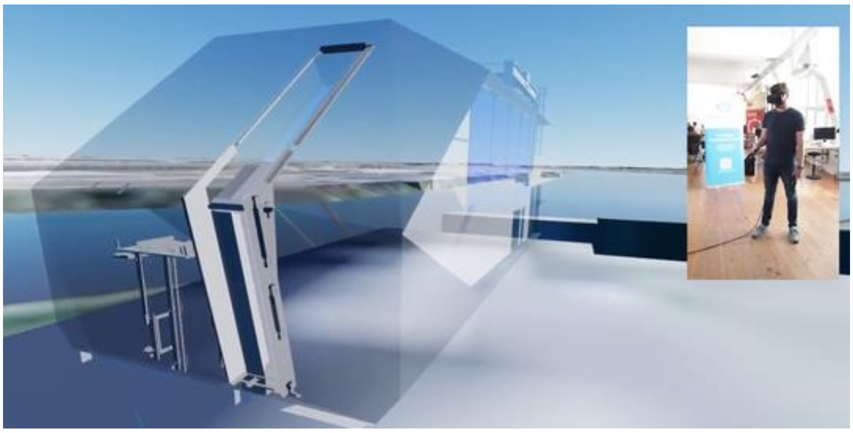

Figure 5-8: HPA's engineer evaluating constructions in virtual reality

5.1.4.3 Use Case 3 – Environmental sensors on barges

Use Case 3 is the only use case where custom UEs are installed on a moving platform. Since the barges are equipped with a 230V power supply, it is not necessary to use battery-powered UEs. The UE itself is integrated with the environmental sensors in a weatherproof (IP67 grade) box. This box has two external connectors: one connects the environmental sensor with its controller; the other connecting the antenna plug for the UE. The antenna itself is mounted on top of the ships pole to allow optimal reception. The setup scenario for the third use case:

- Equipping three ships of the Flotte Hamburg fleet with 5G access points
- Prototype environmental sensors are installed and connected via 5G (see possible installation points in Figure 5-9)
- The raw data is processed in HPA's data centre and visualised for employees of HPA environmental strategy team (see Figure 5-10)

Figure 5-9: Antenna and sensor-box installation points on barge

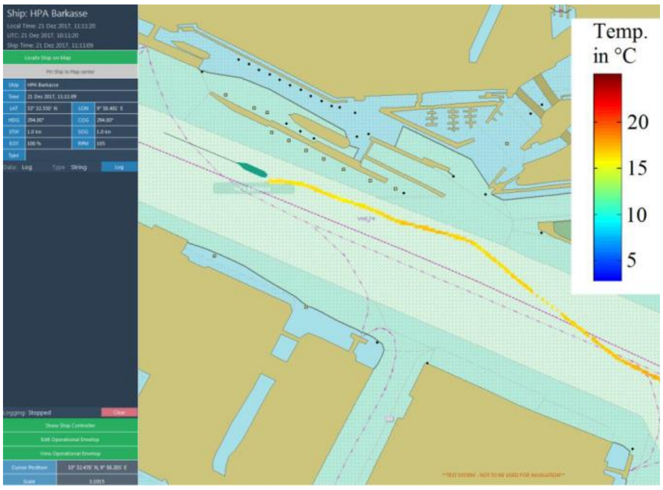

Figure 5-10: User interface mock-up depicting environmental data

5.1.5 Initial measurement campaigns

For the Smart Sea Port testbed, the mobile network infrastructure setup has progressed significantly, and first feasibility measurements have been performed. These measurements were performed to ensure that the envisaged areas of the Hamburg port are sufficiently covered by the testbed's mobile radio network. These measurements included Reference Signals Received Power (RSRP), Reference Signal Received Quality (RSRQ) and Received Signal Strength Indicator (RSSI) measurements, as well as latency measurements using ping. It has been verified that a reliable service can be provided at the relevant areas of the sea port.

The measurements have been performed at a traffic light located about 6.5 km from the Heinrich-Hertz-Tower where the radio equipment is installed. In addition, measurements on a barge moving within the port were performed. Results of the measurement campaign within the Hamburg port area are shown in Figure 5-12. The measurement was started downtown Hamburg (Eastern tip of shown path), and then it followed along the cruise ship terminal, container terminal, and the watergate, to the southern tip of the port area at a distance of about 8.5 km from the Heinrich-Hertz-Tower. On the return path, the measurements were extended to the west until the signal dropped (blue part). The areas of the port which are relevant for the testbed are located in the centre of the figure and are well covered through the testbed's radio site.

In the next step, the UEs and applications planned for the Smart Sea Port testbed will be implemented, installed in the testbed, and then tested. This will require additional measurement campaigns to evaluate the functioning and performance.

Figure 5-11: Measurement at traffic light within port area

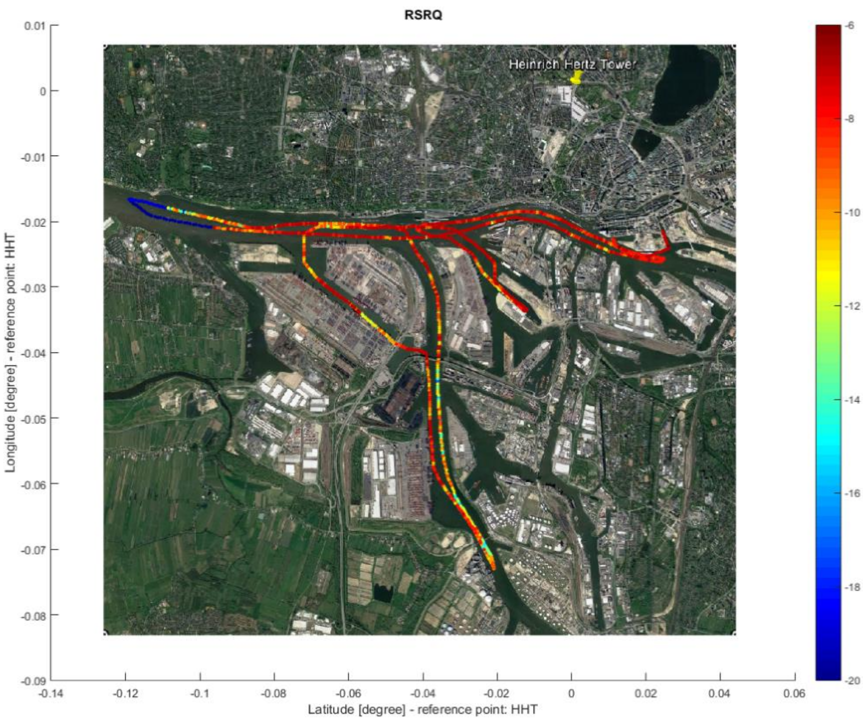

Figure 5-12: RSRQ measurement results within Hamburg port area

5.2 Touristic City testbed

5.2.1 Site and radio frequency planning

The testbed is going to provide an indoor coverage for which a suitable frequency carrier needs to be decided. Since Palazzo Madama is located in Turin downtown, any interference with the commercial network operating in that area must be avoided. For example, some mobile network operators are currently operating on 900 MHz and 1900 MHz to provide GSM coverage as well as on 800 MHz, 1800 MHz and 2600 MHz for LTE. In addition to that, some other commercial service may operate on different frequencies. Candidate frequency bands are 3.4-3.6 GHz or 3.6-3.8 GHz (Band 42 and 43), depending on the presence of other services operating on those ranges. On this specific issue, an interaction with the Italian regulatory authorities regarding an application for the appropriate temporary usage license is currently ongoing. The outcome will also depend on the 5G frequency auction that will take place in September 2018.

According to the actual testbed hardware that will be provided for the radio part, the base station can operate using four antennas that need to be fixed at a high position (around three metres) within the testbed site, in order to provide the appropriate radio coverage of the room. Based on the initial setup measurements, one or two antennas could be sufficient to guarantee a good coverage with a sufficient signal quality.

5.2.2 Hardware and software modules

In the following the various hardware and software modules are described that are included in the Touristic City testbed.

5.2.2.1 PHY/MAC modules

The physical layer of the 5G system to be used for this demo is an SDR-based system with highly reconfigurable system parameters. The hardware components include:

- Lightweight 5G UE device with two main components: Baseband Unit (BBU) for baseband processing of PHY and MAC layers implemented with embedded C/C++ on a x86 platform, and RF unit (RFU).
- Base station (gNB) with three main components: BBU for PHY and MAC connected to higher protocol layers via 10 Gbit/s Ethernet, RF unit connected to BBU with 10 Gbit/s Ethernet, and RF power amplifier for coverage extension.
- The carrier frequency for both gNB and UE can be tuned between 400 MHz to 4 GHz with different bandwidth options (5 / 10 / 20 MHz).
- Delock antennas 88571 (10 cm) for the UE.
- MMDS 8.5 dbi antennas (40 cm).

In order to demonstrate the resource elasticity for the Touristic City testbed, a reconfigurable 5G system even in the lower layers is used. The following are some key physical layer features of the testbed:

- The system supports multiple carrier frequencies, e.g. 800 MHz, 2.6 GHz, 3.5 GHz.
- Different bandwidth options: 5 / 10 / 20 MHz
- Cellular schedule access avoiding interference and collision enabling high reliability.
- The frame structure has a flexible TTI length: 125 µs, 250 µs, 500 µs, 1 ms enabling scalable latency and throughput.
- Adaptive numerology based on service (subcarrier spacing, TTI length, cyclic prefix length etc.).
- Flexible Waveform based on Orthogonal Frequency Division Multiplexing (OFDM): Cyclic Prefix OFDM (CP-OFDM), Windowed OFDM (W-OFDM), Pulse shaped OFDM (P-OFDM) enabling high reliability & low latency.

- Supports Time Division and Frequency Division Duplexing schemes (TDD/FDD).
- Different subcarrier spacing: 15 kHz, 30 kHz, 60 kHz.
- Channel coding: TB-CC for CP, Turbo-code for UP.
- Dynamic reconfiguration of reference signal: scattered or preamble, based on channel conditions.
- Reliability enhancement by retransmission (blind retransmission).
- Supports Peak-to-Average Power Ratio (PAPR) optimisation based on Discrete Fourier Transform (DFT) spreading.

Figure 5-13 illustrates the block diagram of the physical layer of the system. The basic functionality of Physical Downlink Shared CHannel (PDSCH) are similar to that of LTE. However, extra 5G based components are added to achieve high reliability and configurability to the system. Unlike LTE, both DL and UL channels use OFDM based waveform. To achieve different level of reliability and spectral efficiency, a Poly-Phase Network (PPN) module is used to shape the OFDM signal to different waveform types such as P-OFDM, W-OFDM, and CP-OFDM. The system configuration module is used to tune the numerology of the system, e.g. changing the sub carrier spacing, cyclic prefix length, TTI length or pilot distribution

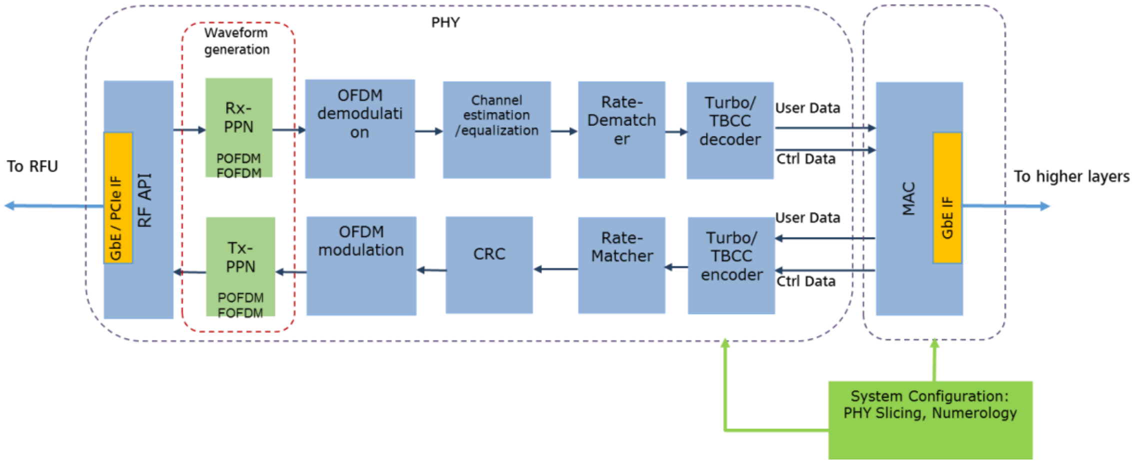

Figure 5-13: Block diagram of the baseband unit (PHY layer)

Based on the requirements of the different slices used for the demonstration, a single physical channel carrying all slices or, a separate physical data channel for each slice can be used.

The physical data channels are configured with different numerologies (TTI length, subcarrier spacing, bandwidth, cyclic prefix, etc.), and can be mapped to one component carrier, or to multiple component carriers to extend the bandwidth. Each slice will be assigned a unique IP address as depicted in Figure 5-14. The control channel can be assigned a different IP address or mapped to one of the slice IPs.

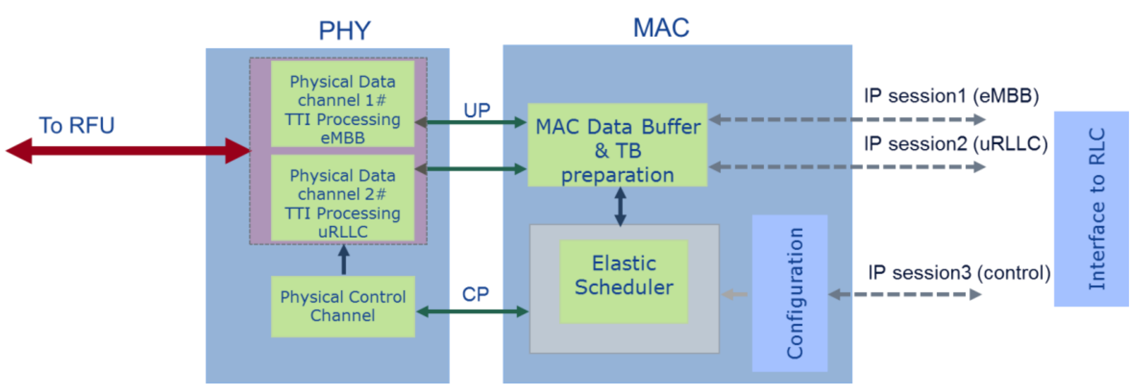

Figure 5-14: High-level overview of the PHY/MAC module

5.2.2.2 Higher layer modules

For the protocol stack the following hardware components will be used: 1 PC for the Edge Cloud and 1 PC for the protocol stack of the 5G UE. At the RAN, referring to the wireless connection between the gNB and the UEs, a set of protocols are defined by 3GPP [3GPP18-38300] for both UP and CP in order to perform procedures above the MAC and PHY layers (Figure 5-15a, and Figure 5-15b) provided by the L1 module. These protocols are integrated in the high layer modules required for the Touristic City testbed. Focusing on the gNB and UE side of the wireless link, the higher layers' module A and module B are defined, respectively.

High layers' module A (the gNB side): This module performs the standardised functionalities of the RLC, PDCP, and RRC protocols needed for the demonstrated scenarios at the gNB side. It is connected via 10 Gigabit Ethernet (GbE) to the L1 modules All the functionality of this module is realised in a single virtualised entity (e.g., in one VM) enabling a spit in the protocol stack. The backhaul connection, i.e., the link towards the MAC layer at the gNB side can be with reconfigurable in terms of reliability (controlled latency and bandwidth).

High layers' module B (the UE side): This module performs the standardised functionalities of the RLC, PDCP, and RRC protocols needed for the demonstrated scenarios at the UE side. It is connected to the BBU component of the UE module and feeds the application layer of the VR client module.

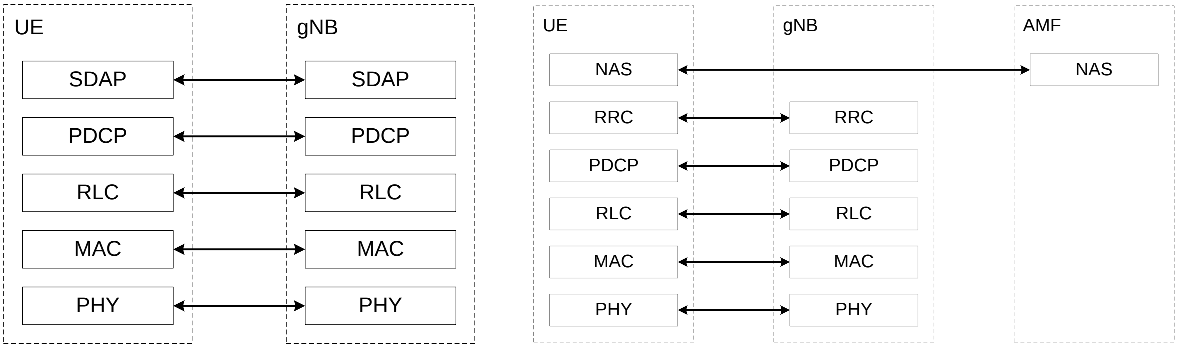

Figure 5-15: a) user plane protocol stack (left), b) control plane protocol stack (right)

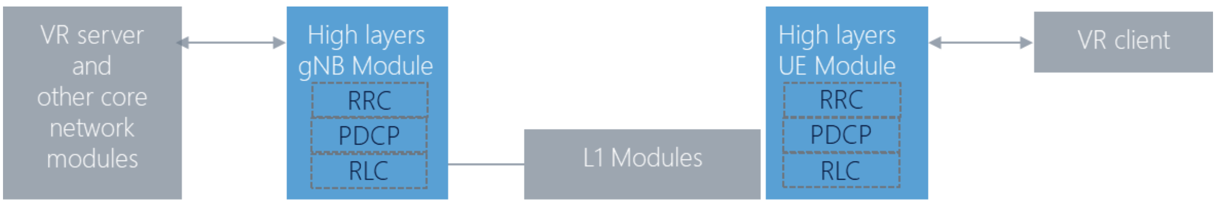

Figure 5-16: High layers' modules

5.2.2.3 AR/VR application

Also, for the VR/AR application, the following hardware will be used:

- 1 360° VR streaming camera (Insta360 Pro).
- 2x Oculus Rift + Oculus Touch.
- 2x Graphics Workstation PCs.
- 1 Video stream + multi-user communication/synchronisation Server PC.
- 2x 1GB Ethernet switches.

The AR/VR application will be implemented as a multi-user client/server system where the clients will be two Oculus Rift Head Mounted Displays (HMDs) with Oculus Touch haptic controllers, each connected to a graphics workstation capable of real-time VR, and located at each user's location. The remotely located PC will be connected to the network through the 5G UE, while the second PC will be connected to the network via a fixed Wide Area Network (WAN) IP connection. The multi-user server

PC, responsible for streaming the 360° VR video stream as well as for the communication and synchronisation of interactions between end users, will be located at the testbed site and connected to the network via a fixed WAN IP connection. During the tests it will be determined if the communication and synchronisation operations handled by the server will need to be moved to the eMBB edge cloud to provide better latency and less delay. In detail, the following software is planned to be installed:

- 2 End-user 3D VR applications implemented in Unity3D with support for VoIP (probably SIP).
- 1 Video stream server.
- 1 Content Management System (CMS) for 3D Models, avatars, artefact multimedia and information.
- 1 Communication and Synchronisation server application.
- 1 VoIP server.

Video streaming server

The video streaming server is used as a live broadcasting solution incorporating the museum's interior inside the VR application via an RTSP video stream over UDP. Figure 5-12 shows the physical connection between the video streaming server and the rest of network. The 360° video stream from the Insta360 Pro camera is connected via Wi-Fi to the streaming server that in turn sends the RTSP video stream to the VR application, controlled by the synchronisation server PC.

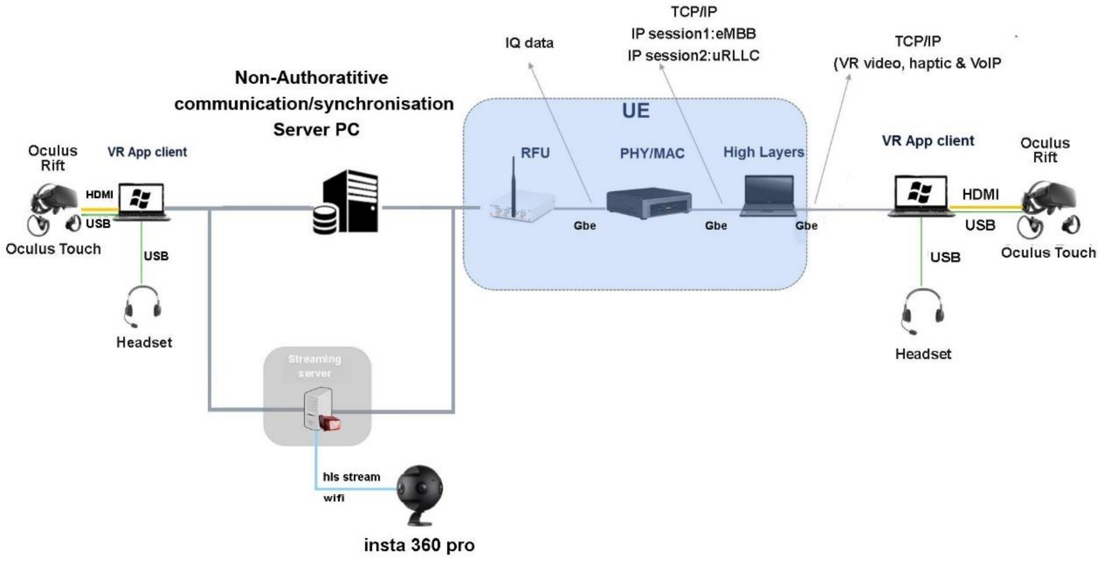

Figure 5-17: Physical connection of video streaming server

The actual function of the video streaming server is to control and define the desirable streaming protocol and method which ends in the network. The live streaming mode of the camera has specific protocols and non-editable scripts limiting the configuration options to the default parameters. A streaming server with transcoding capabilities can restream different video formats into a desirable one and over a specific protocol. Using the VLC[1] Player as the end-user enables to test the server in two ways: (i) different video feeds (resolution, bitrate, live feed, recordings), and (ii) different stream protocols (HTTP Live Streaming (HLS), RTMP, RTSP).

First tests started with Full HD recordings as the minimum accepted resolution, in order to confirm stable connections and the feasibility of stereo video. During these tests the quality of the feed was checked using the VLC (lost packages, frames etc.) to allow for optimising the settings of the stream. Regarding the requirement of UDP packages for the video distribution, RTSP turned out to be the best

[1] VLC is a free and open source cross-platform multimedia player and framework that plays most multimedia files and various streaming protocols.

option for the live stream mode. RTSP delivers continuous streams of requested data without actually storing the data on the hard disk, making it highly advantageous in all respects since a live stream event can be provided to many people simultaneously, regardless of their location. So, RTSP combined with UDP's advantages, such as broadcast/multicast connections and no restrictions in connection-based communication model, provides the best results for the live stream video use case.

Insta360 Pro

The images inside the museum are captured by an Insta360 Pro camera which is connected with the video streaming server as it can be seen in Figure 5-18.

Figure 5-18: Physical connection of Insta360 Pro

The Insta360 Pro can capture 360° videos, mono and stereo, meeting the demands of VR applications. Regarding the live broadcast function of the camera, maximum resolutions of 4K (3840x2160) with 30 frames per second, and 4K 3D (3840x3840 pixels) with 24 frames per second can be achieved for 360° panorama and 360° stereo mode, respectively. Other supported resolutions are for 4K (3840x1920) and 1440p 3D (2880x2880 pixels). In any case, the tested bitrate fluctuated between 15 and 40 Mbps.

A VLC Player was again the media player for the end-user during the tests. Insta360 player is also a good solution for 360° videos. Some problems were identified at higher bitrates (25 Mbps+) due to limitations of the network. Using the built-in HLS Server, no problems occurred except the aforementioned network limitation. On the other side, RTMP and RTSP default streaming options had many problems during broadcast and were non-playable in some cases. So, HLS turned out to be the most appropriate solution.

VR Application

The stereo video from Insta360 Pro is top-bottom model based. So, additional settings were applied in the VR application. Two cameras and two spheres were created representing each eye and their corresponding angle of vision with slightly different positions, matching the feeds (top feed – left sphere, bottom feed – right sphere). Each sphere is actually a media player displaying the incoming footage from the museum and the viewer is at the centre of the sphere. The final result is a video with depth, using Oculus Rift. Figure 5-19 shows the procedure.

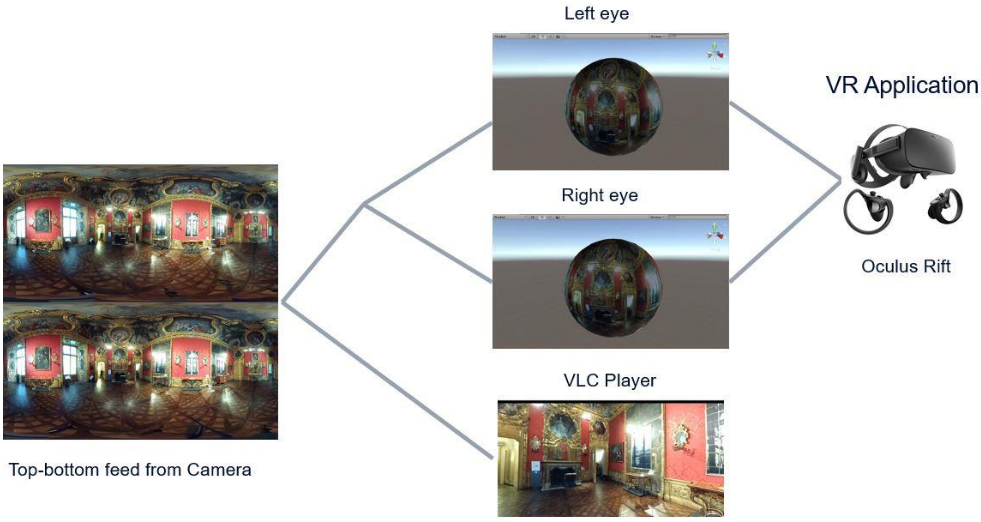

Figure 5-19: Video Stream to VR application

The VR application will utilise 3D models of the cultural artefacts exhibited in the Palazzo Madama Reale room presented in the live stream, where these models will be superimposed on the background video stream such that the two users can manipulate the artefacts in coordinated actions. The application scenario is drafted in such that the capabilities of slicing can be demonstrated in a way where different requirement exists for each slice.

In order to generate the 3D models of the antique furniture exhibited in the Palazzo Madama Reale room, a photo session was conducted in June 2018 where each artefact was thoroughly photographed from multiple angles and heights, to create a 3D scan of each through photogrammetry, see Figure 5-20. In total, 904 photographs were taken of the artefacts:

- Artefact 1: 165 Photos
- Artefact 2: 195 Photos
- Artefact 3: 144 Photos
- Artefact 4: 131 Photos
- Artefact 5: 134 Photos
- Artefact 6: 135 Photos

Figure 5-20: The testbed artefacts from Palazzo Madama

These photos were processed through a photogrammetry reconstruction application which registers each photograph in space to create a raw point cloud for each object, as shown in the example in Figure 5-21. After the point cloud is generated, a texturing stage is applied to derive the appearance of the scanned object from the multitude of photos as shown in Figure 5-22.

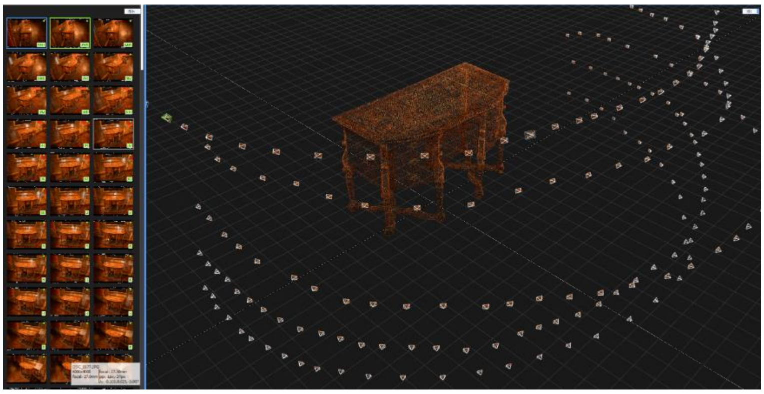

Figure 5-21: Point cloud generation

Figure 5-22: Texture generation for the reconstructed 3D objects

The final result is a high-resolution 3D model of each artefact consisting of 1.5 to 2.5 million vertices. This kind of detail however is not feasible in real-time applications, especially in VR applications that require very high framerates. Thus, the next stage is an optimisation of the models to reduce the number of polygons to ~100000.

In certain cases, due to the complexity of the objects and their reflecting surfaces (especially the table tops) in combination with the overhead lighting in the room, the reconstructed 3D model contained holes that the photogrammetry application was not able to fill based on the algorithms it utilises. Furthermore, some badly reconstructed vertices did not correspond to the actual geometry of the physical object, as in the example in Figure 5-23. In order to fix these problems a final manual editing of each object was performed in a 3D editing application and the 3D models were exported in the FBX file format used by the unity platform. The final 3D models are shown in Figure 5-24.

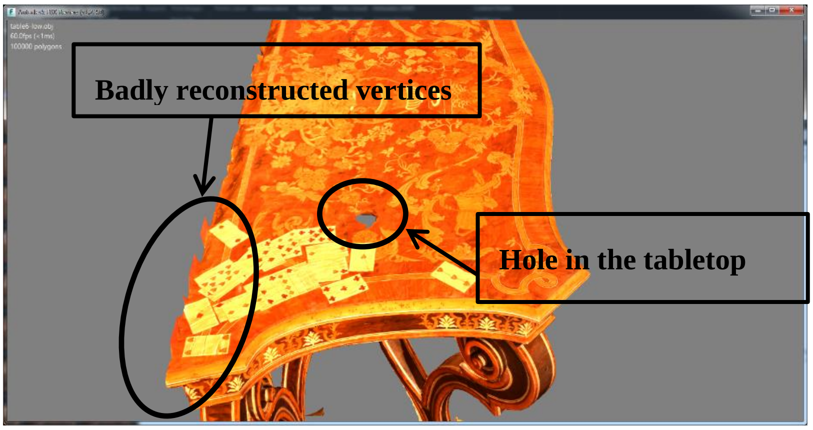

Figure 5-23: Texture generation for the reconstructed 3D objects

Figure 5-24: The final 3D models to be used in the VR application

The next step regarding the 3D reconstructed artefacts will be to embed them into the VR application environment and superimpose them with exact dimensions and locations over the artefacts displayed in the real-time backroad 360° video stream. Starting from there, in order to include the artefacts into the multi-user interaction of the VR application scenario, each object will be assigned physical properties such as weight, collision areas, interaction points etc., allowing for haptic manipulation.

AVATAR

The design steps concerning the implementation of the avatar consists of two main stages:

- The Avatar's movement (placement, animation, haptic interactions with objects);
- Multiuser Interaction.

Movement

One of the main objectives was to design an avatar that acts in a human-like manner. To achieve this, each main part of the (humanoid) biped avatar was matched to a corresponding human part. The next step was to assign to each tracking device used by Oculus Rift the appropriate puppet piece. Inside the VR application, the Oculus Touch controllers must be located over the hands and the fingers of the avatar and the headset in front the head, representing the eyes. Some basic hand and finger animation is included in the Oculus Package and was imported to the avatar's animation controller. Moreover, positional displacement above a defined threshold is translated to a walking animation. Connecting all above functions creates a humanoid avatar capable of walking, moving head and hands, and capable of interacting with objects. Corrections regarding the absolute simulation of the human body were considered, in order to avoid causing disorder to the user. For example, constraints in the rotation of the hand or spin were imposed, offering the user a more realistic experience. Additional improvements were introduced regarding interactions, which are necessary for the projects goals, such as grabbing an object. Figure 5-25 shows some of the main features reported above.

Figure 5-25: Making a humanoid AVATAR

After the puppeteering procedure, the avatar was imported into the VR application which already included the museum's interior as described above. Further optimisations especially regarding the avatar's size had to be made in order to fit to the application's environment and achieve a realistic representation of a humanoid.

Multiuser

The multiuser related requirements of the testbed include features such as almost real-time communication among the users, both visual and haptic, the quality standards set for a commercial VR application, and more. More specifically, the network requirements that ensure an immersive VR experience are:

- Server update: 40 network ticks/s.
- A low round trip time (below 200 ms).
- No server/client reconciliation (minimal caching of state data).
- A stable minimum display frame rate of 80 fps.
- Low packet loss (since TCP is used)
- Reliable ordered packet transfer
- Minimising the server's CPU load

Messaging Layer

To achieve a multi-platform solution, a language-neutral, platform-neutral interface description languate (IDL) was required. Google's protocol buffers (protobuf) (see https://developers.google.com/protocol-buffers/) turned out to be the best solution, since it offers a lightweight serialisation with small overhead for small messages, such as the user's update messages. A set of messages is created and compiled to the client's and the server's corresponding languages with the protobuf compiler and then the code generated files are stored to each project. This allows to overcome platform issues.

Interest Management

Interest management attributes to the reduction of the information amount that needs to be exchanged between users and the resources it consumes, on the other hand the information filtering for each individual has a computational cost as a trade-off. The information relevant to the user for a VR application is:

- The transform of the tracked points:
    - Head Mounted Display (HMD)
    - Tracked Controllers
- The state of the user (e.g.: Is the user wearing the HMD?)
- The hand gestures of the user.
- The state of the interaction with any objects.

More specifically for the above states, the transform refers to the position orientation and velocity of HMD and controllers of the user. The state of the user refers to the situation of the user at the time being connected. So, when a user does not use the HMD, or when there are connection issues as high delay or frequent disconnections, the Avatar stands still. Moving the controllers and pressing specific buttons means that there will be movement in the those hands the user sees through the HMD. For example, the Avatar can clench its fists pressing some button. All these movements are considered as hand gestures. Finally, the state of interaction refers to interactions between the Avatar and the mounted objects, e.g., to grab and pick up an object or pass it to another Avatar.

As the space of the application is limited, no occlusion was required, so at each point in time the user is updated for the whole of his environment.

Interest management was extensively used in order to reduce the average message size, the techniques used were:

- Ignoring data that didn't change since the last update.
- Compressing spatial data representation, as the positional displacement or change in rotation since the last update.
- Compressing quaternions[2]:
    - Normalised quaternions[3] can be reconstructed by using their length and isolating one of the real numbers $x = \sqrt{1 - y^2 - w^2 - z^2}$. Always the biggest number is chosen for reconstruction because of its larger byte size representation.
    - Quaternions q and –q give the same rotation since they represent a change in rotation with the same orientation as a target. This modified orientation can be achieved by rotating the appropriate axis by the required angle in one of the 2 directions. Thus,

[2] Quaternions are used to represent rotations. A quaternion rotation is made up of 4 numbers, whose values all have a minimum of -1 and a maximum of 1.

[3] A normalised quaternion is a quaternion that moves along a unit sphere and its first-time derivative is tangent to that sphere. The magnitude of a quaternion is calculated as $||q|| = \sqrt{x^2 + y^2 + w^2 + z^2}$ and for a normalised quaternion it is known that $||q|| = 1$

always the positive rotation is chosen for transmission, so there is no need to transmit the sign of the orientation.
  - Limiting floating point precision by using bounded unsigned values for the *x,y,z* components, so floats are not sent in full. Afterwards, an analysis of the values that are required to be transmitted a specific precision range is performed, in order to limit the byte size representation of a single rotation. This is of big importance since rotational data contributes with a considerable share to the sent data. Hence, there is a trade-off between network traffic and precision.
- Limiting the precision required for spatial data representation. The *y* component, representing the height, can be represented with a more restricted bounded range since the players will be not be able to jump, or throw items very high.
- Compressing the message in the wire with a LZHL[4] family algorithm and decompressing it on receipt.

Server-Client Synchronisation

The requirements to achieve the desired synchronisation among users are:

1. Floating point determinism: Due to the non-deterministic nature of floating points[5], no floats are used; instead, a pre-agreed precision range of 7 digits is used and all values are converted to integers until they reach the next user.
2. The server decides which is the current state
3. Each message is referenced on an acknowledged state basis: The clients try to reach the server's state, in case a client has missed the last update due to the message never reaching, the client doesn't halt the information flow towards the server.
4. Client and Server buffering on receipt:In order to counter package loss both client and server accept messages that refer to the last 10 and the following 10 states. On a package reconstruction, the client is updated accordingly with diminishing returns based on the state lapse between the current state and the new-found state or is scheduled for update.
5. Interpolating between agreed states: Interpolation between known states allows for fluid movements even with a 10% package loss.
6. Client-side extrapolation/prediction: The client does not stop the movements abruptly in case it is not updated, instead the linear and angular velocity is dampened until it reaches a lower bound.
7. Jitter Buffering: Temporary storage buffer that stores the incoming data packets for a small period of time in order to smoothen the network delay inconsistency.

Login architecture and reconnects

The requirements to achieve synchronised users are described in Section 4.4.2. The server recognises clients using a public/private key scheme, which is exchanged on first log in. The public key is known by all the users and is used to address each in broadcast messages, while the private key is used for the point to point communication.

Moreover, some of the design decisions (buffer on receipt in particular) allows for easy client-side reconciliation in case of a disconnect, since big packages can be reconstructed easily, and is regarded as a simple state update from the client.

---

[4] LZ77 algorithms achieve compression by replacing repeated occurrences of data with references to a single copy of that data existing earlier in the uncompressed data stream. A match is encoded by a pair of numbers called a length-distance pair, which is equivalent to the statement "each of the next length characters is equal to the characters exactly distance characters behind it in the uncompressed stream".

[5] Floating-point determinism: Floating-points can be made non-deterministic due to manufacturers using optimisations or different design decisions on their compilers or instruction set architectures that strive for speed of executions in trade-off for consistency (like most JIT compilers or .Net CLR compiler).

Other Design Decisions

For testing purposes, the client application is devised in 2 sub-applications, a Unity client and a C# console application. The console application's purpose is to flood the network and stress the server by creating virtual clients and emulating multiple user's behaviour by providing randomly generated update messages, mingled with disconnects and reconnect attempts. The Unity client is designed as described above, by bringing together the avatar, the environment and the client-side network application.

The next step will be the implementation of an interactive networked environment where users are able to act together and accomplish haptic tasks in almost real-time. The tasks will be designed with simple physics in mind, in order to not bottleneck the throughput with simulation data and will require precise movement and synchronised cooperation to complete. As a final result, two humanoid Avatars will be available, connected in the same room communicating and interacting with each other as shown in Figure 5-26 and Figure 5-27.

Figure 5-26: View of Client A

Figure 5-27: View of Client B

5.2.2.4 Interfaces and integration

Terminal side

Figure 5-28 shows the different physical and logical interfaces between the testbed components at the UE side. The Oculus Rift glass will be connected to a Windows PC that runs the client AR/VR application via HDMI connection. Both, the Oculus Touch controller and the headset for VoIP are connected via USB to the same machine. A software module is implemented in the Windows PC to convert between TCP/IP and USB/HDMI. The application PC is connected via GbE to a laptop running the higher layer's protocol stack of the UE. This laptop is connected to the BBU via GbE. The Windows PC used for AR/VR application contains a module to combine/split the haptic feedback and VoIP services and conversion to/from Ethernet packets. The protocol stack receives/forwards the eMBB and URLLC services to MAC in two IP sessions. While on the BBU laptop, a Packet Capture (PCAP) module for IP inspection and forwarding is implemented. The two IP sessions are mapped to different PHY tiles with different numerologies before sent to the RF unit.

Network side

Figure 5-29 shows the physical connection between the different components for both the network and the UE sides. The VR video stream (eMBB) from the Insta360 Pro camera is connected via HDMI to the VR application server. In order to perform multi-user communication/synchronisation, a server PC is used to stream a synchronise VR video to the guide UE that is connected via GbE to a switch and to the 5G UE as described in Section 2.2. A traffic generator PC is connected to the network via GbE to generate IP packets used to overload the eMBB slice. Several measurement interfaces are required to provide inputs to (a) the graphical visualisation and (b) the algorithms running in the management/orchestration server. Several additional interfaces are required to control the behaviour of the radio, edge and cloud-based functions and to select different modes regarding elasticity. For simplification, these interfaces are build based on records stored in a data base or implemented as UDP messages.

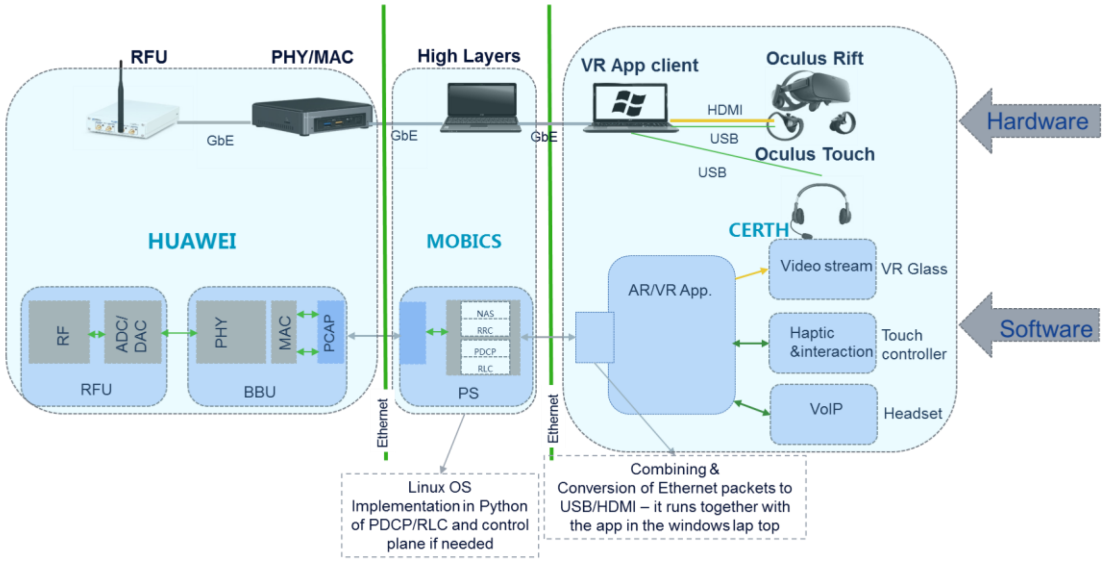

Figure 5-28: Interfaces between different hardware/software modules at UE side

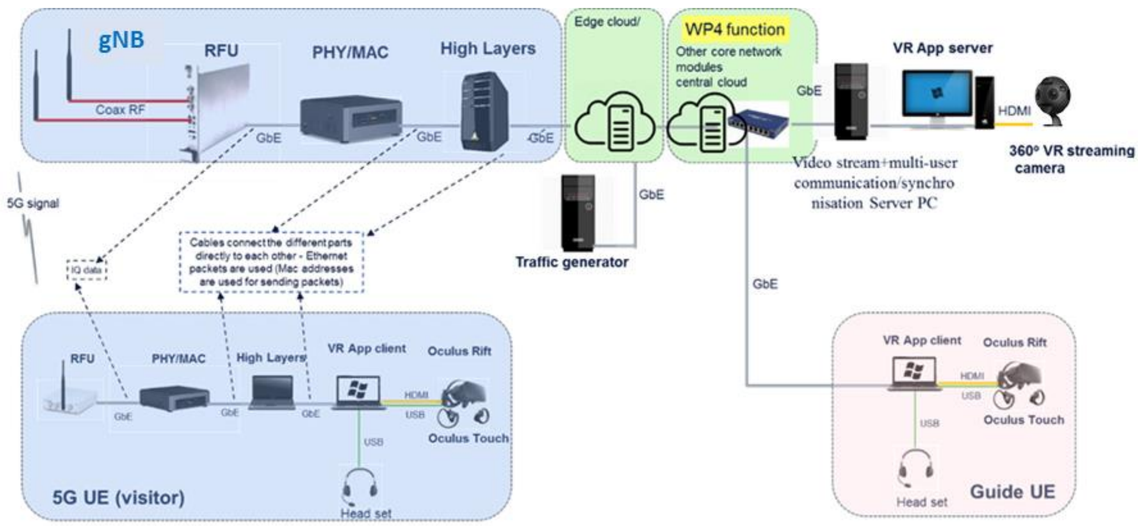

Figure 5-29: Physical connection at network side

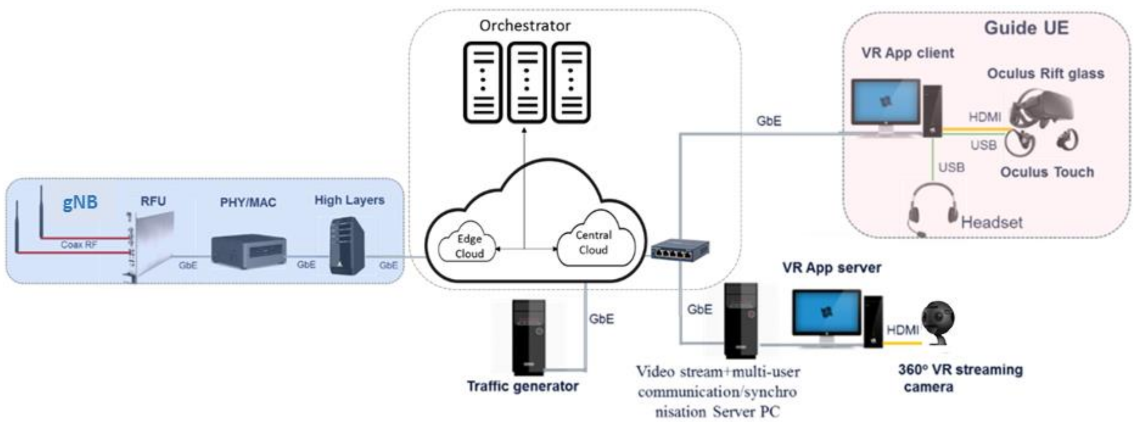

Figure 5-30: Physical connection between orchestrator and cloud modules

Orchestrator and cloud modules

Figure 5-30 clarifies the physical connection between the orchestrator and the cloud modules at the network side. The orchestrator has interfaces with both edge and central cloud infrastructures. These interfaces allow the orchestrator to monitor resources availability and to move NFs from edge to central cloud or inversely resulting in an increased elasticity. The software modules of the network side are depicted in Figure 5-31.

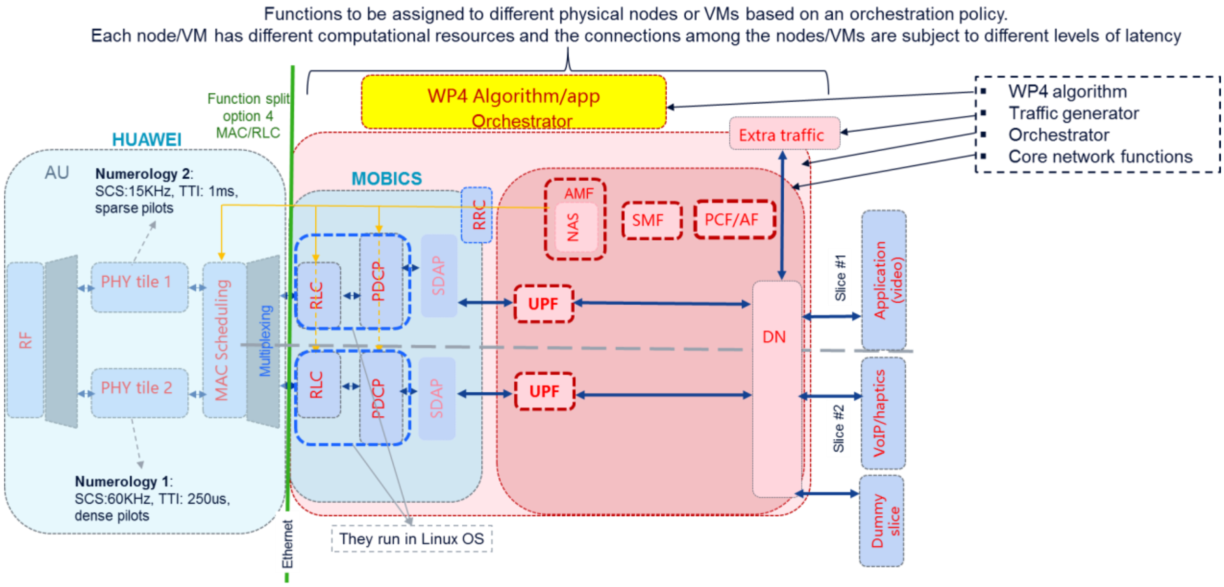

Figure 5-31: Software modules and interfaces at the network side

The Access Unit (AU) of the testbed contains PHY and MAC. The PHY layer is split into two tiles, each with a specific numerology to support two different types of services. The first tile is configured with low subcarrier spacing, long TTI, and sparse pilot distribution for eMBB service. The second one, dedicated to the URLLC service, is configured with high subcarrier spacing, short TTI, and dense pilot distributions. Parts of protocol stack functions are virtualised and can be moved between edge and central cloud. NFs are assigned to different physical nodes or VMs based on the orchestration policy. Each node/VM has different computational resources and the connections among the nodes/VMs are subject to different levels of latency. E.g., UPF and other time-critical processing of RLC and PDCP can be moved between the edge and central cloud, while other CP functions such as AMF can remain in the central cloud.

The orchestration framework will include elements of cognitive network management in the context of the WP4 activities within the ETSI ENI working group. Specifically, ML-based algorithms will be implemented for the cloud resources assignment, especially for the CN functions.

Performance monitoring GUI

For the network performance monitoring and the network elasticity showcasing, a GUI is being developed based on the following principles:

- Illustrates the infrastructure and the functional view of the network in real time.
- Depicts in real time CPU and RAM load of all the testbed components/modules.
- Depicts application layer performance in terms of throughput and delay.
- Allows the interaction with end users. They will be able to a) trigger a slice setup b) add preconfigured traffic flows to the system.
- Depicts with animations the effect of elasticity to the infrastructure and the NFVs of the slices.

An initial version of the GUI screens is depicted in Figure 5-32.

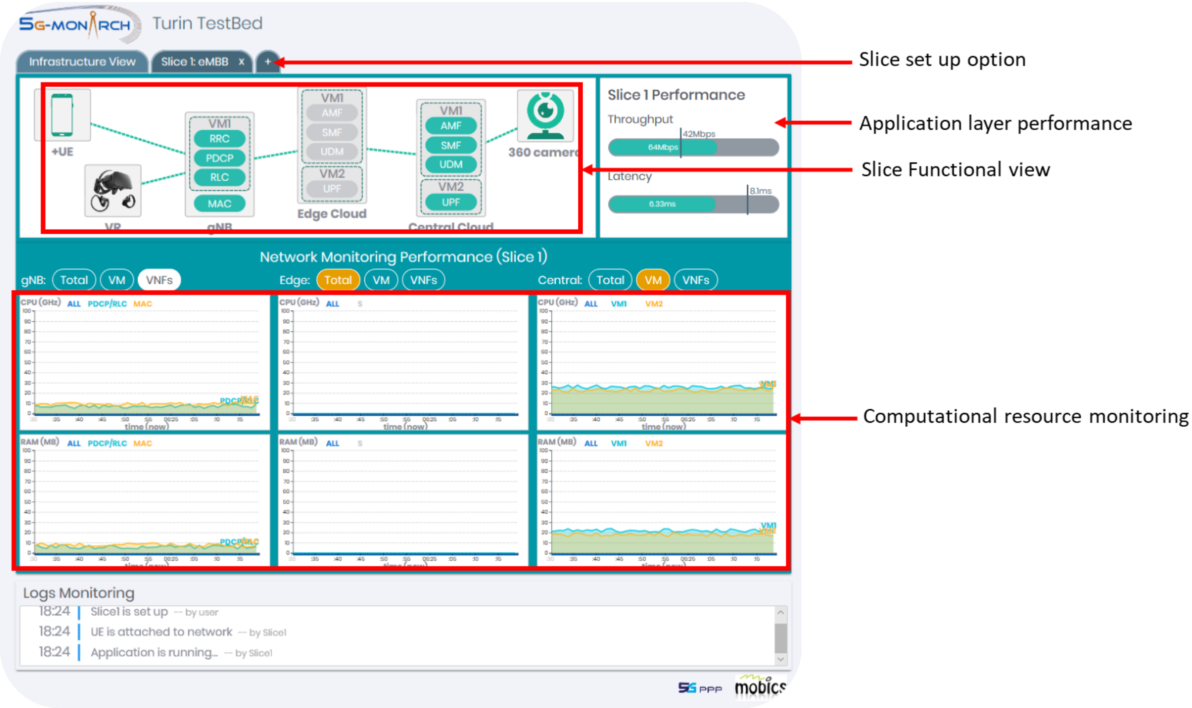

Figure 5-32: Basic screen of the GUI under development for the Touristic City testbed

5.2.3 Test procedure and showcase of network elasticity

The testing of the 5G network slicing capabilities of the project regarding the Touristic City testbed revolves around the following scenario:

- The first end-user acting as the visitor and the second user acting as the touristic guide, both located in different rooms in the museum wearing Oculus Rift, connect to the 3D content CMS and download locally the 3D models situated at the remote cultural site.
- When the models are downloaded, the live video stream from the cultural site starts providing a 4K video of the site and registering the 3D models at locations around the 360° video stream.
- The end-user acting as the visitor downloads multimedia information regarding the 3D artefacts on-demand from the CMS.
- The two users' avatars locations, gestures and interactions with the 3D objects are transmitted in real-time to the communication and synchronisation server and are visible to both users.
- The two users are asked to collaboratively perform manipulations and haptic interactions with the 3D artefacts while at the same time communicating with each other via VoIP.

Through the operations described above bandwidth capacity can be monitored, furthermore transfer times, response delays between clients and server, and latency in the communication between clients and server. The aim is to determine the best 5G network slicing strategy and transfer of operations between edge and remote server that will provide the minimum latency for time-critical operations that affect the user VR experience (avatar placement, animation and haptic interactions as well as voice communication with minimal delay) while at the same time keep the 360° video stream high bandwidth requirement intact without introducing detrimental delays or quality degradations in the streaming of the real-time video feed between both users. Regarding this, the user will able to monitor the connection's quality and speed inside the application and if speed falls below a threshold value a message will appear giving the option to switch to edge cloud.

6 Conclusions

The objective of WP5 is to develop and implement the two testbeds and their components to showcase the key innovations of 5G-MoNArch. In particular, two real-world testbeds are being deployed: (i) the Smart Sea Port testbed, and (ii) the Touristic City testbed. The current state of these activities is reflected by this document, reporting the progress of both testbeds in terms of use case definition, software and hardware design and implementation, as well as the setup and deployment of the testbeds.

It is particularly worth highlighting the following aspects regarding the testbed activities described in this document:

- The selected testbeds involve very relevant use cases for 5G, such as industrial-type of applications (Smart Sea Port testbed) and media/entertainment (Touristic City testbed).
- For the operation of the testbeds, 5G-MoNArch involves real verticals, such as the Hamburg Port Authority and the Municipality of Turin; these are potential customers of the 5G technology which can give valuable feedback on the value that the technology provides to them.
- The setup of the two testbeds is being performed by the operators involved in the project (Deutsche Telekom and Telecom Italia), which provides them with hands-on experience on the issues involved in deploying 5G networks.
- The network equipment used in the testbeds corresponds to pre-commercial equipment and research prototypes of two manufacturers involved in the project (Nokia and Huawei). The use of pre-commercial equipment provides a strong basis for incorporating the features developed by 5G-MoNArch into commercial products.
- The Smart Sea Port testbed is being deployed in an operational network, which provides valuable feedback on the performance of the 5G-MoNArch technology on a real network with stringent requirements.
- The Touristic City testbed will offer real users visiting the Palazzo Madama the possibility of using the testbed deployed to improve the museum experience. Their feedback will provide valuable insights into the improvements that 5G technology can bring to the end-users.

This document has defined a set of use cases to demonstrate the innovations of the 5G-MoNArch project. Each of these testbeds includes the following use cases:

- The Smart Sea Port testbed is deployed in the Hamburg port, one of largest sea ports in Europe. It demonstrates three use cases: (i) the traffic management within the sea port area, (ii) sea port operations through virtual and augmented reality applications, and (iii) environmental measurements to contribute to the reduction of pollution. These use cases show the potential of the network slicing, and network reliability and resilience innovations.
- The Touristic City testbed is deployed in Palazzo Madama Museum in Turin. It focuses on one use case: the improved touristic experience. Within this use case, it demonstrates the following features (i) orchestration-driven elasticity, and (ii) slice-aware elasticity. These features show the potential of network slicing and resource elasticity.

One of the main contribution of this document has been to identify the hardware and software components of the two testbeds. These components have been described in detail, including the integration and the expected interfaces between these components. On top of the baseline components, key aspects of 5G-MoNArch architecture include (i) network slicing management, (ii) orchestration mechanisms and (iii) elasticity algorithms, which are essential support the envisaged use cases.

So far, the various components (namely PHY/MAC layer, higher layer protocols, and applications) have been developed independently. Once the integration of all these components is finalised, the testbeds will be completed deployed and, measurements on performance metrics such as latency, throughput, packet loss rate and resource utilisation efficiency will be performed. These measurements will be fed into the WP6 verification framework and will be used to perform experiment-driven optimisation.

The Smart Sea Port testbed will be used to verify the following targets (i) technical feasibility of network slicing on commercially available equipment including definition of commissioning and management of the slices; (ii) reliability of mobile and wireless communication for industrial services to be provided; and (iii) isolation schemes for a) slice operation, b) resources, and c) reliability. On the other hand, the

Touristic City testbed demonstration results will be used to validate: (i) the optimal network slicing strategy and transfer of operations between edge and central clouds; and (ii) strategies to accomplish minimum latency along with streaming high-volume data, e.g. a high quality 360º video.

Beyond the targets mentioned above, the two testbeds will be used to validate some of the key requirements and KPIs satisfied by the 5G-MoNArch architecture. The requirements and KPIs selected for the testbeds correspond to a subset of those that have been listed in D6.1 [5GM17-D61]. These KPIs cover a number of key innovations of the project on (i) network slicing, (ii) network reliability and resilience, and (iii) resource elasticity. Among others, the KPIs addressed are related to E2E latency, resilience, peak data rate, availability, response time, spectrum and bandwidth flexibility.

The results included in this report provide a solid basis for the development and deployment of the testbeds, having clearly identified the scope of the showcases and the required components. The next steps that will be pursued towards the finalisation of the testbeds are the following:

- The use cases associated to the two testbeds will be refined to include more details on the different steps that will be followed to show the desired functionality.
- The internals of the hardware and software modules comprised within the testbed architecture will be defined in further detail, and so will the interfaces between modules.
- The functions and algorithms executed by each module in each of the individual steps will be identified for the use cases, as well as the communications between the modules.
- The implementation of the various software and hardware components that carry the required functionality will continue.
- Once the implementation of the different modules is complete, they will be integrated and deployed in the testbeds.
- Plans for testing and validating the testbeds and measuring the corresponding KPIs will be further elaborated in conjunction with WP6.

The initial testbed implementation along with preliminary experiments will be reported in IR5.2, and the final results will be provided in D5.2.

7 References

[3GPP16-23799] 3GPP, “TR 23.799, Study on Architecture for Next Generation System,” Release 14, Dec. 2016.

[3GPP18-23401] 3GPP, TS 23.401, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access,” Release 8, June 2018

[3GPP18-23501] 3GPP, “TR 23.501, System Architecture for the 5G System,” Release 15, June 2018.

[3GPP18-28801] 3GPP, TR 28.801, “Study on management and orchestration of network slicing for next generation network,” Release 14, Jan. 2018.

[3GPP18-38300] 3GPP TS 38.300, “NR and NG-RAN Overall Description,” Release 15, June 2018

[5GM17-D21] 5G-MoNArch project, deliverable D2.1, “5G-MoNArch baseline architecture based on 5G-PPP Phase 1 results and gap analysis,” Nov. 2017

[5GM17-D61] 5G-MoNArch project, deliverable D6.1, “Documentation of Requirements and KPIs and Definition of Suitable Evaluation Criteria,” Oct. 2017

[5GM18-D22] 5G-MoNArch project, deliverable D2.2, “Initial overall architecture and concepts for enabling innovations,” July 2018

[5GM18-D31] 5G-MoNArch project, deliverable D3.1, “Initial resilience and security analysis,” June 2018

[5GM18-D41] 5G-MoNArch project, deliverable D4.1, “Architecture and mechanisms for resource elasticity provisioning,” June 2018

[5GM18-D62] 5G-MoNArch project, deliverable D6.2, “Methodology for verification and validation of 5G-MoNArch architectural innovations,” July 2018

[5GNOR17-D33] 5G NORMA project, deliverable D3.3, “5G-Norma 5G NORMA network architecture – Final report,” Oct. 2017

[5GPPP15] 5G-PPP, 5G Vision Brochure, online available at: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[5GPPP-I] 5G-PPP Phase I projects overview, https://5g-ppp.eu/5g-ppp-phase-1-projects.

[EOB +16] E. Strinati, J. Oueis, S. Barbarossa, “Distributed mobile Cloud Computing: A multi-user Clustering Solution,” IEEE Int. Conf. on Communications (ICC 2016), Kuala Lumpur, Malaysia, May 2016.

[ETSI14] ETSI-MAN001, “Network Functions Virtualisation (NFV); Management and Orchestration”, V1.1.1, Dec. 2014.

[ETSI17] ETSI ISG NFV, White Paper “Network Functions Virtualisation (NFV) – Network Operator Perspectives on NFV priorities for 5G,” Feb. 2017

[GBL+12] I. Grondman, L. Busoniu, G. A. D. Lopes et R. Babuska, “A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), pp. 1291-1307, 2012.

[ODCA14] Open Data Center Alliance, “Best Practices: Architecting Cloud-Aware Applications,” Rev. 1.0, 2014

[PJD+15] S. Parsaeefard, V. Jumba, M. Derakhshani et T. Le-Ngoc, “Joint resource provisioning and admission control in wireless virtualised networks,” IEEE WCNC, pp. 2020-2025, 2015.

[WAL12] B. Walder, “Cloud Architecture Patterns,” O’Reilly Media, Sept. 2012.