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 D6.1

Documentation of Requirements and KPIs and Definition of Suitable Evaluation Criteria

Contractual Date of Delivery	2017-09-30
Actual Date of Delivery	2017-09-30
Work Package	WP6 – Verification and validation
Editor(s)	Sina Khatibi (NOMOR)
Reviewers	Lars Christoph Schmelz, Christian Mannweiler (NOK-DE), Albert Banchs (UC3M)
Dissemination Level	Public
Type	Report
Version	1.0
Total number of pages	52

Abstract: This document describes the criteria and requirements necessary to achieve a flexible, adaptable and programmable fully-fledged architecture for 5G mobile networks. The architecture will incorporate the key enabling innovations (i.e., inter-slice control and cross-domain management; experiment-driven optimisation; and cloud-enable protocol stack) and the key functional innovations (i.e., resilience and security, and resource elasticity) of the 5G-MoNArch project. A brief high-level overview of the baseline architecture is presented, and the relevant general requirements and dedicated requirements for security, resilience, and resource elasticity aspects are defined. The initial collection of relevant Key Performance Indicators (KPIs) is provided. The considered services are briefly introduced, together with an elaboration on the service characteristics that will be in focus, especially with respect to the two testbed environments. In addition, a stakeholder model is defined, focusing on the stakeholder roles for the testbed environments. Finally, the initial draft for the envisaged verification and validation process is discussed.

Keywords: 5G Network Architecture, KPIs, Service Definitions, Validation, Verification

Executive Summary

The key goal in the 5G-MoNArch project is to achieve a flexible, adaptable, and programmable fully-fledged architecture for 5G mobile networks. This revised architecture shall be grounded upon and demonstrate the three key enabling innovations: *i)* inter-slice control and cross domain management *ii)* experiment-driven optimisation, and *iii)* cloud-enabled protocol stack. Moreover, resilience and security and resource elasticity are the two use-case specific functional innovations considered in this project. This document, as the first deliverable of Work package WP6 and the 5G-MoNArch project, provides the first design criteria and a set of requirements necessary to achieve the project's goals.

First, an overview of the service sets for 5G mobile networks defined by other 5G infrastructure Public Private Partnership (5G PPP) projects and Standards Developing Organisations (SDOs) such as the 3rd Generation Partnership Project (3GPP) and the European Telecommunications Standards Institute (ETSI) are presented. From these service sets, a subset which addresses the research focus and the testbeds of 5G-MoNArch is elaborated in more detail and therefore serves as the starting point for the further work in the project. Accordingly, updates to the aforementioned service sets that reflect the project's work progress will be reported in the next deliverables of WP6 during the course of the project.

Second, this document introduces the high-level 5G network architecture concept that serves as baseline for the further work in the project, in particular regarding work packages WP2 'Flexible and Adaptive Architecture', WP3 'Resilience and Security, and WP4 'Resource Elasticity'. With this architecture baseline, a definition for the corresponding stakeholder model is introduced, followed by an explanation of the role of each of the stakeholders for the testbed environments that will be implemented in WP5 'Testbeds'.

Third, this document describes the relevant requirements for designing the 5G mobile network architecture while taking care of novel technical enablers and innovations. The requirements are categorised into three main groups:

1. **General requirements**: represents a consolidated version of general requirements taken out of the output of 5G PPP Phase 1 (from projects like 5G NORMA and METIS-II and from project overarching 5G-PPP working groups (WGs)), of industry forums like NGMN, and of SDOs like 3GPP and ETSI;
2. **Resilience and security requirements**: represents specific criteria and requirements that are necessary for assessing whether the operation of a network implementing the considered architecture design meets the required standards in terms of resilience, i.e., the ability to keep the network and services up and running in case of infrastructure and radio link failures, and security, i.e., enterprise-level security even in shared infrastructure environments;
3. **Resource elasticity requirements**: represents specific criteria and requirements for assessing the ability of the network to dynamically adapt to load changes in addition to uncertainties regarding the availability of processing, storage and networking resources in a way that a targeted performance is guaranteed. The proposal for the resource elasticity is one of the project's novelties described in this document.

Based on these presented requirements, the related Key Performance Indicators (KPIs) are briefly introduced and described. In addition, techno-economic KPIs that are required to assess and validate the overall value proposition of the 5G-MoNArch approach and architecture, and application-specific KPIs related to the specific testbed use cases are briefly introduced and described.

Finally, the verification and validation process for the aforementioned requirements and design criteria is introduced at a high level. This process will be developed in more detail and implemented during the further work of WP6.

Note that the requirements, design criteria, KPIs and concepts described in this document will be improved and enhanced in the course of the project. The corresponding updates and improvements will be presented and described in the future deliverables of the project, in particular, D6.2 'Methodology for verification and validation of 5G-MoNArch architectural innovations' (due in Month 6 of the project) and D6.3 'Final report on architectural verification and validation' (due in Month 24 of the project).

List of Authors

Partner	Name	E-mail
NOK-DE	Borislava Gajic Diomidis Michalopoulos	Borislava.gajic@nokia-bell-labs.com Diomidis.michalopoulos@nokia-bell-labs.com
UC3M	Albert Banchs	banchs@it.uc3m.es
DT	Gerd Zimmermann Heinz Droste	ZimmermannG@telekom.de Heinz.Droste@telekom.de
TIM	Sergio Barberis Giorgio Calochira	sergio.barberis@telecomitalia.it giorgio.calochira@telecomitalia.it
ATOS	Rodrigo Diaz Beatriz Gallego-Nicasio Crespo	rodrigo.diaz@atos.net beatriz.gallego-nicasio@atos.net
CERTH	Anastasios Drosou	drosou@iti.gr
MBCS	Dimitris Tsolkas Odysseas Sekkas	dtsolkas@mobics.gr sekkas@mobics.gr
RW	Julie Bradford Simon Fletcher	julie.bradford@realwireless.biz simon.fletcher@realwireless.biz
NOMOR	Sina Khatibi	khatibi@nomor.de

Revision History

Revision	Date	Issued by	Description
0.1	01.07.2017	5G-MoNArch WP6	Initial draft
1.0	30.09.2017	5G-MoNArch WP6	Final version for delivery

List of Acronyms and Abbreviations

2G	2nd Generation mobile wireless communication system (GSM, GPRS, EDGE)
3G	3rd Generation mobile wireless communication system (UMTS, HSPA)
4G	4th Generation mobile wireless communication system (LTE, LTE-A)
5G	5th Generation mobile wireless communication system
3GPP	3rd Generation Partnership Project
5G-MoNArch	5G Mobile Network Architecture
5G PPP	5G infrastructure Public Private Partnership
AIV	Air Interface Variant
API	Application Programming Interface
AR	Augmented Reality
B2B	Business-to-Business
B2B2X	Business-to-Business-to-anyone
B2C	Business-to-Consumer
BF	Beamforming
CAGR	Compound Annual Growth Rate
CAPEX	CAPital EXpenditure
CN	Core Network
CP	Control Plane
CriC	Critical Communications
CU	Centralised Unit
D2D	Device-to-Device
DCSP	Data Centre Service Provider
DL	Downlink
DU	Distributed Unit
E2E	End-to-End
eMBB	Enhanced Mobile Broadband
FCAPS	Fault, Configuration, Account, Performance, and Security
FMC	Fixed Mobile Coverage
HW	Hardware
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
InP	Infrastructure Provider
ITS	Intelligent Transport System
KPI	Key Performance Indicator
LTE	Long Term Evolution
MANO	Management and Orchestration

MBB	Mobile Broadband
mMTC	Massive Machin Type Commination
mmW	Millimetre Wave
mIoT	Massive Internet of Things
MNO	Mobile Network Operator
MSP	Mobile Service Provider
MVNO	Mobile Virtual Network Operator
NE	Network Element
NEO	Network Operation
NF	Network Function
NFV	Network Function Virtualisation
NFVI	Network Function Virtualisation Infrastructure
NSI	Network Slice Instance
NSSI	Network Slice Subset Instance
N(S)-aaS	Network (Slice)-As-A-Service
NGMN	Next Generation Mobile Networks
IaaS	Infrastructure-As-A-Service
OPEX	OPerational EXpenditure
OTT	One-Trip Time
PAN	Personal Area Network
PaaS	Platform-As-A-Service
PNF	Physical Network Function
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RRC	Radio Resource Control
RRH	Radio Remote Head
RRM	Radio Resource Management
TT	Round-Trip Time
SA	System Architecture
SDMC	Software Define network Control
SDN	Software Defined Network
SDO	Standards Developing Organisation
SDU	Service Data Unit
SW	Software

TR	Technical Report
TRxP	Transmission-Reception Point
TSG	Technical Specification Group
UE	User Equipment
UL	Uplink
UP	User Plane
URLLC	Ultra-Reliable Low Latency Communications
V2I	Vehicle-to-Infrastructure
V2X	Vehicle-to-everything
VNF	Virtualised Network Function
VISP	Virtualisation Infrastructure Service Provider
VR	Virtual Reality
WP	Work Package

Table of Contents

1 Introduction

1.1 Objectives

A flexible, adaptable, and programmable mobile network architecture is an essential requirement for 5th Generation (5G) wireless networks, to be able to leverage and support the large diversity of services associated with the foreseen 5G use cases such as smart city [5GPPP17]. The 5G-MoNArch (5G Mobile Network Architecture for diverse services, use cases, and applications in 5G and beyond) project aims to evolve the network architecture concepts developed in the first phase of 5th Generation infrastructure Public Private Partnership (5G-PPP) projects [5GPPP-PI] to a fully-fledged architecture and complement it by three key enabling innovations, namely:

- inter-slice control and cross-domain management, to enable the coordination across slices and infrastructure domains,
- experiment-driven optimisation, to leverage experimental results to design high performance algorithms, and
- a cloud-enabled protocol stack, to gain flexibility in the orchestration of virtualised functions.

In addition, two use case specific functional innovations are considered:

- resilience and security,
- resource elasticity.

The key enabling and functional innovations will be evaluated in the two testbed scenarios intended within 5G-MoNArch, namely, the sea port and the touristic city testbed. With respect to the functional innovations, resilience and security are in focus of the sea port scenario, while resource elasticity plays the dominant role in the touristic city scenario to achieve highly efficient use of the resources.

Verification and validation will establish a framework for the overarching collaboration between Work Packages (WPs) in 5G-MoNArch to verify that the envisaged innovations are technically and economically feasible. Besides verification against Key Performance Indicators (KPI), which confirms whether the technical benchmarks are reached, the project's results also need to be validated against the expectations and needs of stakeholders to guarantee that the developed system satisfies their requirements.

The objective of this document is to provide the first design criteria and requirements to achieve the goals of project.

Although examining the range of potential 5G services is still an on-going task in research projects and standardisation, a basic set of service classes has already been identified. This set can be used as the starting point for the design choices to be made within the project. Furthermore, having the models for the stakeholders and their interaction within 5G wireless networks (i.e., individuals, entities or organisations that affect how the 5G-MoNArch system operates) is essential for defining the requirements and KPIs for the network architecture.

The stakeholder model and the first set of requirements and design criteria provided in this document build the basis for the flexible and adaptive architecture being developed in 5G-MoNArch WP2 'Flexible and Adaptive Architecture'. In addition to providing the general requirements, they are classified into, and enhanced by dedicated security and resilience requirements (taken from WP3 'Resilience and Security'), and resource elasticity requirements (taken from WP4 'Resource Elasticity'). Finally, a first summary of the KPIs needed for verifying and validating the aforementioned requirements and design criteria are described.

It worth noting that many of the requirements, design criteria, KPIs, and concepts described in this document will be improved and enhanced as part of the project work. The updates will be presented in future deliverables, in particular, D6.2 'Methodology for verification and validation of 5G-MoNArch architectural innovations' (due in December 2017) and D6.3 'Final report on architectural verification and validation' (due in June 2019).

1.2 Structure

The rest of this document is organised as follows:

- Chapter 2 addresses the services to be considered in 5G mobile networks. This section briefly presents the service sets defined within other 5G PPP projects and Standards Developing Organisations (SDOs); in particular, the 3rd Generation Partnership Project (3GPP) and the European Telecommunications Standards Institute (ETSI). In addition, the service characteristics that will be in focus of the 5G-MoNArch project, especially in the two testbed environments, have been elaborated.
- Chapter 3 presents a brief high-level overview of the initial 5G mobile network architecture considered in 5G-MoNArch. In addition, a stakeholder model is defined which is in line with the underlying architecture definition. The roles of each of the stakeholders for the testbed environments are explained.
- Chapter 4 addresses the relevant requirements for the 5G mobile network architecture which will be specified within 5G-MoNArch, taking care of novel technical enablers and innovations. After a short introduction in Section 4.1, Section 4.2 focuses on the general requirements to be fulfilled by 5G mobile networks (considering the output of SDOs or other 5G PPP projects), while Section 4.3 and Section 4.4 extend this requirement list by new criteria related to innovations on security and resilience aspects as well as on resource elasticity features that will be evolved in WP3 and WP4 of 5G-MoNArch, respectively. The requirements for resource elasticity has been proposed for the first time in this document. This proposal in addition to adaptation of the requirements for security and resilience are the novelties of this document.
- Chapter 5 lists the KPIs related to the requirements and the criteria explained in Chapter 4.
- Chapter 6 briefly describes the verification and validation process that will be performed during the project activity within WP6, which is also tightly connected to the testbeds to be established in WP5 of 5G-MoNArch.
- Finally, the document is summarised in Chapter 7.

2 Service Definition

2.1 5G networks claim to cover a diverse range of services

One of the key benefits claimed for 5G networks is the ability to support multiple services with diverse requirements from the same infrastructure set in an efficient and flexible way that:

- Saves costs compared with legacy networks;
- Allows the introduction and management of new services much faster than legacy networks;
- Unlocks new revenue opportunities from engaging with new industries and markets requiring wireless services beyond traditional consumer focused mobile broadband (MBB) services.

The range of potential 5G services have been and continue to be examined within SDOs, industry and research forums such as Next Generation Mobile Networks Alliance (NGMN), and European Commission projects. In particular, previous studies in this area have been reviewed in detail in [5GN-D21]. Figure 2-1 from this work summarises the range of service requirements arising from the 3 main service classes identified in this work. Here, in this project, three classes of services are evolved slightly to align better with latest 3GPP [3GPP-22186] [3GPP-22261] and ITU-R [ITUR-M2083] thinking of:

- Enhanced mobile broadband (eMBB) – where high throughputs are key;
- Massive machine type communications (mMTC) – where packet sizes and throughputs are usually small, but device densities are high;
- Ultra-reliable and/or low latency communications (URLLC) – where reliability and potentially latency are key as these services have a critical nature.

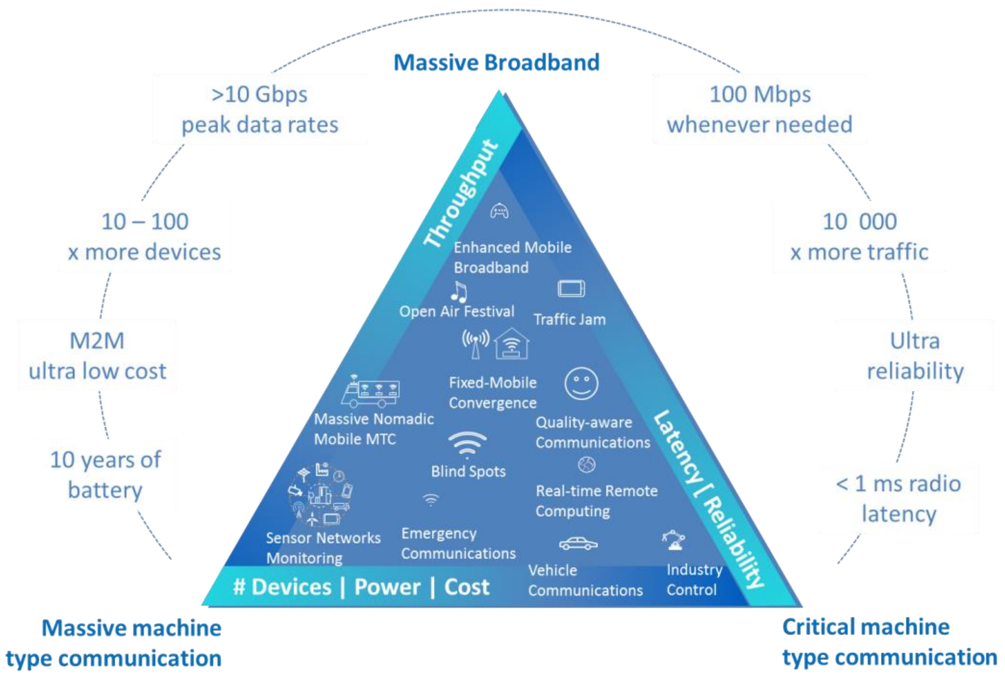

Figure 2-1: Main 5G service classes and possible use cases in the 5G ecosystem [5GN-D21]

2.1 Services of interest for assessment of 5G-MoNArch

The 5G-MoNArch project aims to evaluate the performance of the innovations that it generates in terms of:

- Improvements to relevant stakeholder experience of existing services or ability to deliver new services beyond today’s networks which may increase willingness to pay, unlock new revenue streams and/or deliver social benefits;

- Ability to deliver services more cost efficiently than in existing networks.

To quantify the impact on user experience of a given service and understand the required dimensioning of network infrastructure and functions running on it to deliver that service the services of interest in assessing 5G-MoNArch first need to be defined. This is done by:

- Focusing on the two testbed scenarios that are central to this project, and expanding on how 5G-MoNArch might enhance existing or deliver new services in these two example scenarios.
- Considering an example environment of a future smart city where there will be potential for the 5G-MoNArch platform to deliver a range of services to different user groups. This scenario does not directly reflect either of the two testbed scenarios but aims to capture an example scenario that reflects:
    - Roll out of the 5G-MoNArch concept and results in a larger network with a scope beyond the one considered in the testbed scenarios (which are relatively localised for practical reasons).
    - How 5G-MoNArch might provide an evolution of today's cellular networks from focusing on MBB for consumer devices to incorporating a wider range of services and value creation opportunities for both the mobile industry and wider society.
    - The mix of services that a 5G-MoNArch network may be required to deliver in addition to the more specialised services being focused on the testbed scenarios.

2.1.1 Touristic city and sea port testbeds

Within this project two 5G-MoNArch testbeds will be set up in the form of live deployments in

- A touristic city location to show how 5G-MoNArch can flexibly assign the resources needed to deliver demanding Augmented Reality (AR) and Virtual Reality (VR) services to localised demand hotspots around attractions to enhance a visitor's experience of a city (Figure 2-2).
- A large sea container port location to show how 5G-MoNArch can securely deliver a range of environmental sensor, logistics tracking, vehicle control, surveillance, congestion control and operational communications services to a setting with a huge volume of users (made up of both people and machines, Figure 2-3).

Figure 2-2: Illustration of immersive and integrated media anticipated for the touristic city testbed

These testbeds will demonstrate the key innovations of 5G-MoNArch in a real-world deployment, and allow for acquiring and measuring data at service level that can be compared with the defined KPIs.

Furthermore, practical implications of the developed concept and architecture on the instantiation and roll-out of 5G networks and services can be acquired. Some example features that are envisaged for the two testbed scenarios are given in Table 2-1 and Table 2-2. Further details on the requirements for the wireless services to support these features are expanded upon in the next sections.

Table 2-1: Touristic City Testbed

Demonstrated Feature	Description	Specific service
On-site Live Event Experience	Live events with 360-degree content video	eMBB supporting 360-degree video (high throughput but not necessarily low latency)
	Superimposing AR/VR overlay	AR/VR-based eMBB with low latency (i.e., <10 ms) and high throughput
Immersive and Integrated Media	Concurrent tour with people on-site and overlay of real people on-site. People will see a part of the touristic city area full of real and imaginary people.	AR/VR-based eMBB with low latency (i.e., <10 ms) and high throughput
Cooperative Media Production	Virtual visitors will record real-time 360 degrees AR/VR. People will cooperate in real time with imaginary and real people who are feeling the same VR experience.	AR/VR-based eMBB with low latency (i.e., <10 ms) and high throughput

Figure 2-3: Illustration of smart sea port testbed environment in Hamburg with some of the services intended to be considered (please note that there might be changes in the final definition of the testbed area according to upcoming specifications in WP5)

Table 2-2: Sea Port Testbed

Demonstrated Feature	Description	Specific service
Traffic Light Control (URLLC)	Traffic lights (static as well as mobile, e.g., in case of construction site along streets) connected through wireless links in a reliable and resilient way under consideration of data integrity	Intelligent traffic signal control (high reliability, low throughput MTC service)
Video Surveillance (eMBB)	Video control entrance to sea port area or parts of it, up-to-date status information related to those areas; also, data integrity and security as important aspects	eMBB service supporting 4k+ video (high throughput MTC service, but not necessarily low latency)
Sensor Measurements (mMTC)	Measurements about, e.g., environmental pollution on mobile barges connected through wireless terminals or at stationary locations	Low throughput, high density MTC for environmental data analysis, simply extendable also to logistics (when combined with other sensor requirements beyond those on barges)

2.1.2 Future smart city services

For the wider and more generic example scenario of a future smart city environment this project considers:

- eMBB services for consumer portable devices (such as smartphones, tablets and laptops).
  - This is largely driven by demand for 4k+ live streaming of video.
  - Support for challenging AR and VR applications may also be required in localised hot spots of usage.
- Vehicle-to-infrastructure (V2I)[1] services made up of:
  - Infotainment and advertising to passengers;
  - Information services on road and driving conditions as well as navigation (e.g., for parking purposes);
  - Assisted and automated driving services.
- City council and utility services made up of:
  - Environmental monitoring, intelligent transport system (ITS) infrastructure (e.g., road traffic congestion control), and waste management sensors;
  - Smart energy, covering, e.g., smart metering and smart grids.
- Logistics made up of:
  - Sensor data for tracking goods in transit.

2.2 User requirements for 5G-MoNArch services of interest

This section sets out the expectations and requirements of users against each of the services envisaged in the two testbed scenarios and the example future smart city scenario. These requirements for delivering acceptable service levels are described in terms of:

- Service requirements related directly to user experience;
- Capacity and coverage requirements.

[1] Note only V2I from Vehicle-to-everything (V2X) services are considered here as 5G-MoNArch focuses on infrastructure requirements rather than device requirements and device-to-device connectivity.

Under the above two categories the following metrics are defined:

- **Minimum required bit rate for this service** – this is the minimum guaranteed bit rate that must be delivered to ensure an acceptable user experience for this service. The user experienced data rate (described later in Chapter 5) for a network deployment will be compared against this to ensure an acceptable experience is being delivered.
- **Required end-to-end (E2E) latency** – this is the minimum guaranteed user perceived latency or reaction time of an application that must be delivered to ensure an acceptable user experience for this service. As this is the user perceived latency it covers the round-trip time (RTT) for data to successfully be transmitted, received and acknowledged at the application layer i.e., covering all layers of the radio protocol stack.
- **Data volume per device using a given service per day** – provides information on the typical message size and frequency or average amount of data consumed per device for a given service per day
- **Number of devices** – is the indicative expected device density for a given service. Combined with the above data volume per device this indicates the area traffic density expected for a service. This can then be matched with the area traffic capacity (see Chapter 5) of a given network.
- **Percentage of scenario with service coverage** – this is the percentage of the target demand locations in the scenario considered, where a device should be able to receive the service. This can be compared against the measured "Coverage" of a network as defined in Chapter 5.

The service requirements described here are an evolution of those already considered in the 5G NORMA project and detailed in [5GN-D33].

2.2.1 Touristic city testbed

In this section, the service requirements for the services to support the proposed features to be demonstrated in the touristic city testbed are presented in Table 2-3 and Table 2-4.

Table 2-3: Performance requirements for touristic city services

Service component	**Min. required bit rate**	**E2E latency**	**Reliability and security**
eMBB supporting 360-degree video (high throughput but not necessarily low latency)	50 Mbps	<100 ms	Best effort reliability and Consumer grade security
AR/VR-based eMBB with low latency (i.e., <10 ms) and high throughput	50 Mbps	<10 ms	Best effort reliability and Consumer grade security

Table 2-4: Capacity and coverage requirements for touristic city services

Service component	**Data volume per user per day**	**Number of visitors per km^2 per day (user density)**	**Percentage of tourist attraction area with coverage**
Enhanced MBB supporting 360-degree video (high throughput but not necessarily low latency)	1.125 GB (assumes typically 3 hours usage)	Up to 150k Note: The testbed will not demonstrate this high volume of users directly.	95% (outdoors at testbed area / point of interest)
AR/VR-based eMBB with low latency (i.e., <10 ms) and high throughput	1.125 GB (assumes typically 3hours usage)	Up to 150k Note: The testbed will not demonstrate this high volume of users directly.	95% (outdoors at testbed area / point of interest)

2.2.2 Sea Port Testbed

Here the service requirements for the services to support the proposed features to be demonstrated in the sea port testbed are presented in Table 2-5, Table 2-6, and Table 2-7.

Table 2-5: Performance requirements for sea port services

Service component	Minimum required bit rate	E2E latency	Reliability and security
Intelligent traffic signal control (high reliability, low throughput MTC service)	Minimum connectivity	>100 ms	High reliability and high security
eMBB service supporting 4k+ video (high throughput MTC service, but not necessarily low latency)	10 Mbps	<100 ms	Best effort reliability and high security
Low throughput, high density MTC for environmental data analysis or logistics	Minimum connectivity	>100 ms	Best effort reliability and Consumer grade security

Table 2-6: Capacity and coverage requirements for sea port services

Service component	Data volume per user or device per day	Number of devices or users per km^2	Percentage of port area with coverage
Intelligent traffic signal control (high reliability, low throughput MTC service)	1-byte messages with 1440 messages per day i.e., one per minute [IEEE-2011]	100s of road sensors in the port area.	99.9% (outdoors)
eMBB service supporting 4k+ video (high throughput MTC service, but not necessarily low latency)	1.8 GB (assumes 24-hour video surveillance)	10s of video surveillance points in the port area.	95% (outdoors)
Low throughput, high density MTC for environmental data analysis or logistics	200-byte messages, 100 messages per day i.e., updates every 15 minutes. [GSMA-2016]	10s of thousands of containers in the port area per day 100s of environmental sensors in the port area	95% (outdoors)

2.2.3 Example future smart city scenario

Services in the specific testbed areas would be provided in addition to a wide range of wireless services outside the locality of the testbed areas. Therefore, here the requirements for a busy future smart city scenario as would be expected as the baseline service requirement for future wireless networks are presented.

The services listed earlier for our example future smart city scenario map to the three-main service classes of eMBB, mMTC and URLLC as shown in Table 2-7.

Table 2-7: Categorisation of example services to service classes

Service		**eMBB**	**mMTC**	**URLLC**
eMBB for consumer portable devices	eMBB - 4k+ streaming	X		
	eMBB - AR/VR	X		
Vehicle-to-infrastructure	Infotainment	X		
City council and utility services	Information services		X	
	Assisted driving			X
	Environmental monitors, intelligent transport systems (ITS) and waste management		X	
	Smart energy		X	
Logistics	Tracking goods		X	

2.2.3.1 Future smart city - eMBB service requirements

The service requirements for all services envisaged in the future smart city scenario that are classed as eMBB type services are defined in Table 2-8 and Table 2-9.

Table 2-8: Performance requirements for future smart city eMBB services

Service component	**Minimum required bit rate for this service**	**E2E latency**	**Reliability and security**
eMBB – consumer portable devices (driven by video applications)	10 Mbps Downlink (DL)/Uplink (UL) (4k video quality experience)	<100 ms	Best effort reliability and Consumer grade security
V2I – infotainment (eMBB)	10 Mbps DL (4k video quality to at least one passenger)	<100 ms	Best effort reliability and Consumer grade security

Table 2-9: Capacity and coverage requirements for future smart city eMBB services

Service component	**Data volume per device per day**	**Number of devices**	**Percentage of city scenario with coverage**
eMBB – consumer portable devices (driven by video applications)	On average, each device consumes 0.25 GB per day in 2020 growing to nearly 3 GB by 2030, i.e., an approx. 30% CAGR[2] (Considering outdoor demand only which assumed to be 20% of overall eMBB traffic on average)	10s of thousands per km^2	95% (outdoors)
V2I – infotainment (eMBB)	On average 1 GB - 25 GB per day per car (2020 - 2030)	100s of vehicles per km	95% (vehicles, outdoors)

[2] Compound Annual Growth Rate (CAGR)

2.2.3.2 mMTC service requirements

The service requirements for all services in the future smart city scenario that are classed as mMTC type services are defined in Table 2-10 and Table 2-11.

Table 2-10: Performance requirements for future smart city mMTC services

Service component	Minimum required bit rate	E2E latency	Reliability and security
V2I – driver information service (mMTC)	0.5 Mbps DL/UL	< 100 ms	Best effort reliability and Consumer grade security
Environmental monitors, waste management and ITS (mMTC)	Minimum connectivity UL	> 50 ms[3]	Best effort reliability and Consumer grade security
Smart meters – sensor data, meter readings, individual device consumption (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security
Smart grid sensor data and actuator commands (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security
Logistics sensor data for tracking goods (mMTC)	Minimum connectivity UL	> 50 ms	Best effort reliability and Consumer grade security

Table 2-11: Capacity and coverage requirements for future smart city mMTC services

Service component	Data volume per device per day	Number of devices	Percentage of city scenario with coverage
V2I – driver information service (mMTC)	On average 50 MB consumed per day per car in 2020 growing to 1700 MB per day per car in 2030 due to growing service uptake (i.e., 40% CAGR).	100s of vehicles per km^2	95% (vehicles, outdoors)
Environmental monitors, waste management and congestion control (mMTC)	On average 230 bytes per day per roadside item (i.e., traffic lights, road signs, bins etc.) in 2020 growing to 1500 bytes per day per roadside item by 2030. This is a CAGR of 20%.	100s of devices per km^2	95% (outdoors)
Smart meters – sensor data, meter readings, individual device consumption (mMTC)	1600 bytes per smart meter per day i.e., 200-byte messages, 8 messages per day	10s of thousands per km^2	99% (indoors) Reflects smart meters likely in hard to reach locations
Logistics sensor data for tracking goods (mMTC)	4 MB per day per equipped vehicle based on 200-byte messages, 100 messages per day (i.e., updates every approx. 15 mins) per sensor and on average 200 items to track per goods vehicle	Up to 10k items to track per km^2	95% (vehicles, outdoors)

[3] This value is the latency tolerance and the delay could be more than 50ms and not impact them.

2.2.3.3 URLLC service requirements

Finally, here the service requirements for all services in the future smart city scenario that are classed as URLLC type services are defined in Table 2-12 and Table 2-13.

Table 2-12: Performance requirements for URLLC services

Service component	Minimum required bit rate	E2E latency	Reliability and security
V2I – assisted driving	0.5 Mbps DL/UL	< 100 ms	High reliability

Table 2-13: Capacity and coverage requirements for URLLC services

Service component	Data volume per device per day	Number of devices	Percentage of city scenario with coverage
V2I – assisted driving	On average 50 MB consumed per day per car in study area in 2021 growing to 1500 MB per day per car by 2030, i.e., 45% CAGR	100s of vehicles per km^2	99.9% (vehicles, outdoors) High coverage target reflects high reliability requirement

3 Stakeholder Models

3.1 A brief high-level overview about the 5G-MoNArch architecture

The 5G-MoNArch architecture is based on a flexible and programmable, highly virtualised network infrastructure where the so-called network slices can be readily formed to meet the specific requirements of a given service or target market [NGMN15]. Figure 3-1 illustrates the idea of forming network slices by implementing chains of Network Functions (NFs) related to one or more services across a shared common infrastructure set. The underlying infrastructure layer may span across antenna sites (which is the radio site containing as a minimum the antennas and potentially some radio protocol stack processing), edge cloud sites (small, locally located data centre centres with processing capacity close to the antenna site) and central cloud locations (centrally located data centres hosting a significantly large collection of processing, storage, networking) connected via Software Defined Network (SDN)-based transport networks (covering both backbone and x-haul, i.e., front-/mid-/backhaul) [5GN-D31] [5GN-D32] [5GPPP16]. It is further assumed that network slice provisioning may span across different Infrastructure Provider (InP) domains to allow wider service coverage. This may also include infrastructure of vertical players, such as an enterprise campus or factory building.

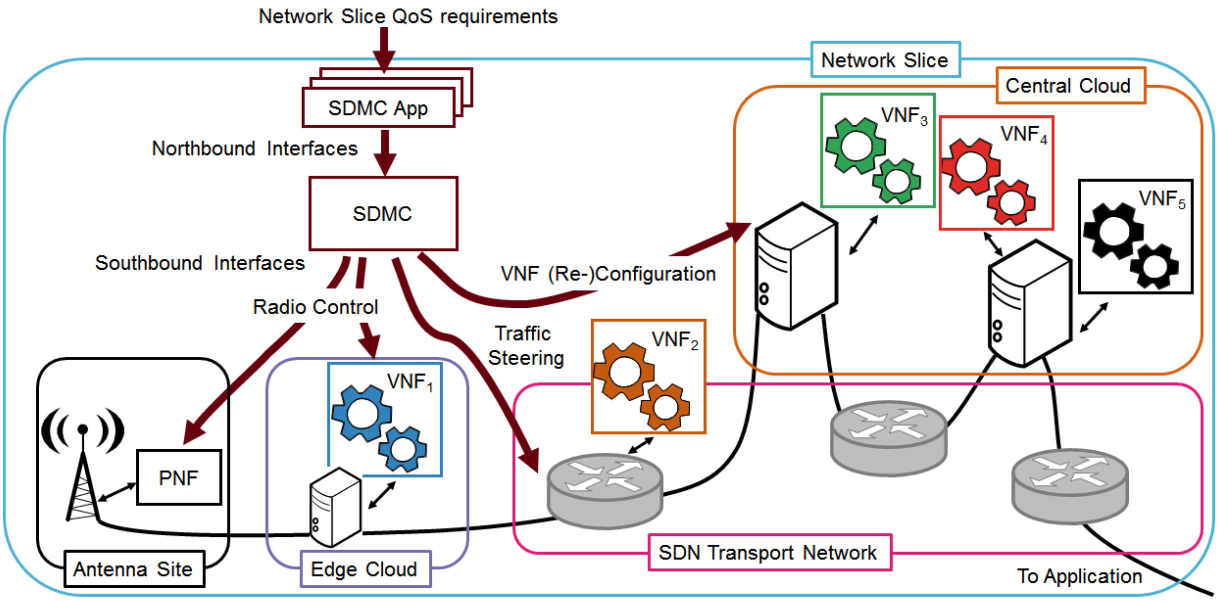

Figure 3-1: High level architecture initially considered in 5G-MoNArch (enhanced from [5GN-D32])

According to the approach of Network Function Virtualisation (NFV) [ETSI-NFV] there can be a separation between the hardware (HW) infrastructure layer and the software (SW) running on it, which allows Virtualised NFs (VNFs) to be flexibly placed across the infrastructure nodes according to the needs of the services covered. Due to performance and energy efficiency reasons, there may still be some Physical NFs (PNFs) that are coupled with underlying HW (especially related to lower layer processing in the RAN protocol stack), so not all NFs in a system will be fully virtualised.

The SDN principle of separating Control Plane (CP) and User Plane (UP) is generally applied in the architecture concept by logically centralised Software Defined Mobile network Control (SDMC). Dependent on the Network Slice Instances (NSIs) to be implemented dedicated control applications will be orchestrated via the northbound interface of the SDMC to the network CP supervising the underlying UP handling of data flows. The Network Management and Orchestration (MANO) functionality (not explicitly shown in Figure 3-1) considered in 5G-MoNArch is seen as an extension of the ETSI NFV MANO concept [ETSI-MAN001]. More details on the architectural view can be found in [5GN-D32] and [5GPPP16] and will be described in Deliverable D2.1 of 5G-MoNArch [5GM-D21].

Overall, the virtualisation of the network means that a separation can exist between those stakeholders providing the sites, physical equipment and inter-site connectivity and those implementing the required functionality to provide E2E network connectivity. This provides the opportunity for today's mobile ecosystem to grow and accommodate new stakeholders as introduced in the next subsection.

3.2 Definition of the stakeholder model

In today's cellular networks the Mobile Network Operator (MNO) typically owns the spectrum, antenna sites and core network sites including the corresponding equipment. The networks also implement the required functionality at each site to deliver the required service level to either their subscribers and/or a Mobile Virtual Network Operator (MVNO). Occasionally, the MNO may also own the inter-site transport network (integrated operator) or be leasing the corresponding lines from another operator. Within 5G virtualised networks as proposed by 5G-MoNArch there is the opportunity to move away from this highly integrated stakeholder model, towards a model with multiple layers. With these multiple stakeholder layers, opportunities for new market entrants are introduced, which allows to introduce new business models in the mobile network market, and to provide customised equipment or service implementation wherever and whenever needed. This ability to customise will ideally lead to the integration of new verticals into the mobile ecosystem, enable opportunities for new revenue streams for mobile service providers, and knock-on benefits to society more generally.

One view of the tiered stakeholder model, enabled through a flexible 5G network architecture, is shown in Figure 3-2. This model is largely taken from the 5G-NORMA project [5GN-D32], but aligned to the terminology used currently at 3GPP [3GPP-28801]. The definition of the stakeholder roles within this model are presented in the following, continued by an elaboration on how the two 5G-MoNArch testbed scenarios can be mapped to this tiered model.

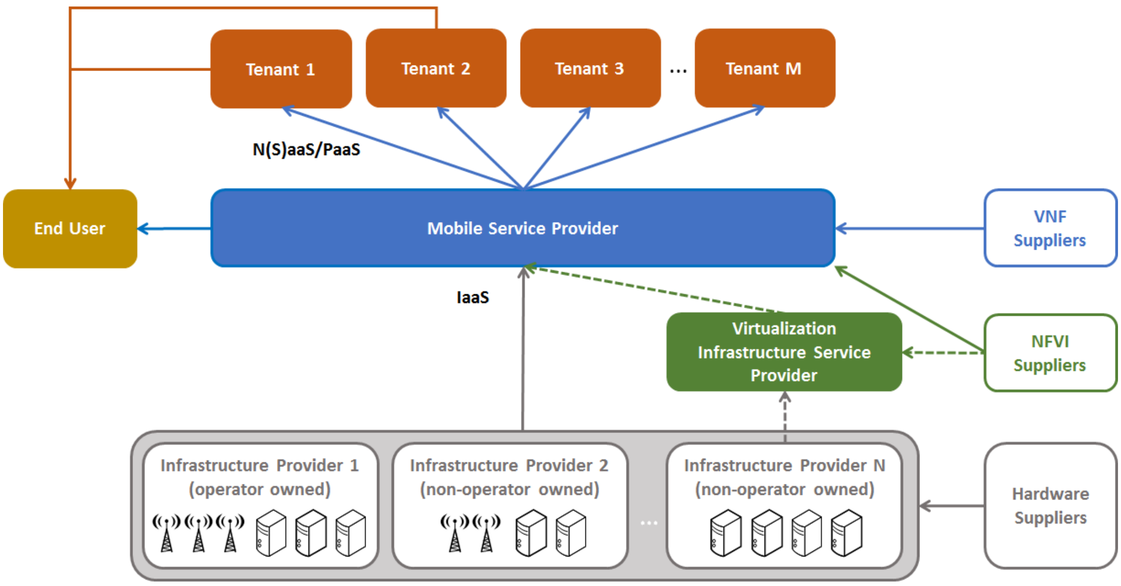

Figure 3-2: Example of the tiered stakeholder model that 5G virtualised networks enable (modified from [5GN-D32])

Stakeholders are individuals, entities or organisations that affect how a system implementing the 5G-MoNArch approach operates. Where appropriate, and as guided by business model analysis, some stakeholders will be actors in the cost or revenue structure.

A 5G-MoNArch **Mobile Service Provider (MSP)** provides mobile internet connectivity and telecommunication services to either end users directly, i.e., through a business-to-customer (B2C) relationship, or via an intermediate "tenant", i.e., a business-to-business (B2B) or business-to-business-to-anyone (B2B2X) relationship; see next stakeholder description. The dedicated logical mobile network resources offered by an MSP are based on Network Slice Instances (NSIs) realising the relevant NF chains to support the instantiated telecommunication services, e.g., eMBB or mMTC. In case of intermediate tenants, the MSP's offerings are Network (Slice)-as-a-Service (N(S)aaS) or Platform-as-a-Service (PaaS). An MSP is responsible for design, build and operation of its service offerings.

A **5G-MoNArch tenant**, usually a business entity, buys and leverages a 5G-MoNArch network slice and services provided by the MSP. A tenant can, for example, be an MVNO, an enterprise (e.g., from a vertical industry) or other organisations that require telecommunications services for their internal business operations or for offers to their customers.

A **5G-MoNArch Infrastructure Provider (InP)** is the entity/company that owns and manages parts of or all of the infrastructure of the network under consideration and offers it to the MSP, i.e., Infrastructure-as-a-Service (IaaS). With respect to the architectural model in 5G-MoNArch, the InP role may be further sub-divided into antenna site infrastructure provider, transport network provider, and data centre service provider (DCSP). The former owns the physical infrastructure such as the antenna sites, the HW equipment for the antennas and Remote Radio Heads (RRHs), monolithic base stations, etc. (i.e., infrastructure related to PNFs). The latter is represented by the collapsed roles of an entity/company that owns and manages local and/or central data centres. Within 5G-MoNArch, there are two types of data centre operators, infrastructure providers acting on small/medium size data centres (in terms of resources to be deployed and geographical presence) and big players (like Amazon) having big data centres deployed world-wide.

In 5G-MoNArch terminology an MNO is an entity that operates and owns the mobile network, i.e., it vertically integrates the roles of MSP and InP into a single stakeholder entity.

In practice, there may be also a so-called **Virtualisation Infrastructure Service Provider (VISP)** which designs, builds and operates its virtualisation infrastructure(s) on top of InP services provided by one or more DCSPs. The VISP offers its infrastructure service to the MSP.

Further roles in the stakeholder model to be mentioned are the **HW supplier** offering HW to the InPs (server, antenna, cable, …), the **NFV Infrastructure (NFVI) supplier** providing the corresponding NFV infrastructure to its customers, i.e., to the VISP and/or directly to the MSP, respectively, and finally the **VNF supplier** offering virtualised SW components to the MSP.

3.3 Examples for stakeholders

Given the definitions of the stakeholder model above, this section provides some examples for mapping this model to dedicated roles (such as individual persons or companies) in a real mobile wireless network environment. This mapping is performed for the economic sample area of a smart city scenario, see Table 3-1, as well as for the two testbed scenarios, see Table 3-2 and Table 3-3, respectively. While the roles for the testbed scenarios are rather concrete in terms of dedicated 5G-MoNArch partners, the smart city scenario roles are more generic and cover a wider scope and perspective.

The stakeholder model and roles gain a particular importance for the economic and technical validation of the project's conceptual work, to be conducted as part of the work in WP6, but also for the analysis and evaluation of the results provided by the testbed implementations. While in WP6 a cost-benefit analysis can be conducted based on simulation results of the research work in WP2, WP3 and WP4, the testbeds will deliver results from practical implementations, allowing to incorporate measures that cannot be achieved through simulations (e.g., equipment and services roll-out cost and complexity).

It is to be noted again that, in this deliverable, only a first high-level idea of the stakeholder model is presented, which will be worked out in further detail in the course of the project.

Table 3-1: Stakeholder roles in future smart city scenario

Stakeholder Role	**Fulfilled by**
InP	City council assets, city property owners, existing MNOs, fixed network operators, DCSPs
MSP	Existing MNOs
Consumer	Pedestrians and passengers in vehicles using consumer handheld devices, drivers (assisted driver services), logistics companies, city councils (smart city applications), energy companies (smart metering and smart grids)
Tenant	Vehicle manufacturers, public transportation companies, city councils, energy companies

Table 3-2: Stakeholder roles in the sea port testbed scenario

Stakeholder Role	**Fulfilled by**
InP	MNOs (e.g., Deutsche Telekom/DT (also w.r.t. fixed network), with Nokia as possible HW supplier), Hamburg Port Authority/HPA (own network infrastructure), venue owner, city council
MSP	DT, HPA
Consumer	HPA, logistics management company, train operator
Tenant	Port authority (here HPA)

Table 3-3: Stakeholder roles in touristic city testbed scenario

Stakeholder Role	**Fulfilled by**
InP	MNOs (e.g., Telecom Italia/TIM (also w.r.t. fixed network), with Huawei as possible HW supplier)
MSP	TIM
Consumer	Tourist
Tenant	Venue owner, city council

4 Relevant Requirements

4.1 Introduction

Achieving a fully-fledged 5G mobile network architecture in addition to enhancing the network architecture with a set of key-enabling innovations are the two key goals of the 5G-MoNArch project. In the first step toward achieving these goals, different requirements and design criteria for the 5G network have to be defined and described. Focusing on the 5G-MoNArch innovations, this section addresses this first step through three categories:

1. Defining a comprehensive set of general requirements for 5G mobile networks based on the output of 5G PPP Phase 1 projects, industry forums, and SDOs.
2. Defining a set of security and resilience specific design criteria and requirements which are particularly related to WP3 of the project.
3. Defining a set of resource elasticity design criteria and requirements which are particularly related to WP4 of the project.

Note that the requirements and design criteria described in this chapter represent a first collection and summary performed during the start-up phase of 5G-MoNArch and will be improved and enhanced in the course of the project, in particular with Deliverable D6.2 'Methodology for verification and validation of 5G-MoNArch architectural innovations' (due in Month 6 of the project), and with Deliverable D6.3 'Final report on architectural verification and validation' (due in Month 24 of the project).

4.2 General Requirements

In this section, a set of comprehensive requirements for the 5G system being considered within the 5G-MoNArch project are given. This set represents a consolidated version of general requirements taken out of the output of 5G PPP Phase 1 (from projects like 5G NORMA and METIS-II and from project overarching working groups (WGs)), of industry forums like NGMN and of SDOs like 3GPP and ETSI. The style of requirement descriptions is the one used by 3GPP in their specifications. Most of the requirements are taken one-to-one from those specifications and reports whereas others are from other forums but aligned to 3GPP ones. Partly, requirements addressing the same issue are merged together, but without changing the meaning behind them. There were also some editorial changes, e.g., abbreviations and terminology, due to reasons of a harmonised presentation in this deliverable.

The requirements are collected according to the objectives that they address into the following five blocks (represented as Sections 4.2.2 - 4.2.6):

- Generic requirements on the overall 5G system;
- Requirements on network slicing;
- RAN-related requirements;
- Requirements w.r.t. capability exposure;
- Security requirements.

Please note that only fully-fledged 5G-related aspects are covered. The issues which may be relevant only for initial deployment of 5G based on non-standalone (NSA) approach of 3GPP (use of 4G EPC and LTE-A eNBs as the main anchor for New Radio (NR) gNBs) (e.g. [3GPP-23799] and [3GPP-38801]) are not considered here.

Initially, some background information is given in Section 4.2.1 to highlight the corresponding sources for the requirements listed.

4.2.1 Background information

NGMN has set out initial requirements for 5G within its 5G White Paper from March 2015 [NGMN15] and also taken into account the output of EU FP7 projects on the 5G topic like METIS and 5GNOW. In

June 2016, updated KPIs and requirements were addressed by NGMN in [NGMN16-1] as input to the 5G standardisation process at 3GPP.

5G PPP Phase 1 projects like 5G NORMA and METIS-II defined performance and functional requirements as well as KPIs in their initial phase in the second half of 2015 (see [5GN-D21] and [MII-D24]) considering also the NGMN output at that point in time. It is worthwhile to note that METIS-II focused on the 5G radio access network (RAN) design, whereas 5G NORMA addressed the overall system architecture including core network (CN) and management & orchestration (MANO).

3GPP started their work on 5G within technical specification group (TSG) "System Architecture" (SA) with the feasibility study SMARTER performed by WG SA1. Several technical reports (TRs) were produced describing 5G use cases (UCs) and potential service and operational requirements. An overall collection is given in TR 22.891 [3GPP-22891], whereas a more detailed analysis of the four UC categories i) massive Internet of Things (MIoT; aka mMTC), ii) critical communications (CriC; aka URLLC), iii) enhanced mobile broadband (eMBB), and iv) network operation (NEO) has been performed in TRs 22.861-864 [3GPP-22861]-[3GPP-22864]. During their normative work, SA1 provided a new collection of 5G service requirements in the technical specification (TS) 22.261 [3GPP-22261] by grouping them according to high-level requirements and basic capabilities as well as performance, security, and charging requirements. TS 22.261 provides also some insights into the relation of communication service availability and reliability (see Annex C of [3GPP-22261]). In TS 22.185 [3GPP-22185] basic scenarios and requirements for vehicle-to-anything (V2X) services are described, which are already addressed by LTE-A technology. TS 22.186 [3GPP-22186] goes beyond that status by introducing novel scenarios (e.g., platooning, advanced and remote driving) and their requirements w.r.t. enhancements of radio technology by 5G.

In study on architecture for the next generation (aka 5G) system, the WG SA2 described architectural requirements and assumptions as well as an initial high-level architecture (see TR 23.799 [3GPP-23799]). The 5G system architecture as currently defined in TS 23.501 [3GPP-23501] supports data connectivity and services enabling deployments to use techniques such as NFV and SDN with corresponding requirements.

High-level security requirements are described by WG SA3 in its study on security aspects for the 5G system in TR 33.899 [3GPP-33899]. This TR also includes an overview of the initial 5G security architecture in an informative annex. During their normative work on 5G SA3 added more detailed requirements related to the architecture in TS 33.501 [3GPP-33501].

SA5, the responsible WG for telecom management aspects, has listed in TS 28.500 [3GPP-28500] a set of business level related requirements to be fulfilled for applying NFV as technology enabler in 3GPP-based systems. Potential requirements on management of VNFs being part of NR air interface are stated in TR 32.864 [3GPP-32864]. TR 28.801 [3GPP-28801] covers results of an SA5 study on management and orchestration for network slicing on the network slice instance layer and for non-virtualised network elements also on the resource layer. The TR additionally lists dedicated use cases and their requirements. TR 28.802 [3GPP-28802] contains results of an SA5 study on potential management related requirements for 5G in terms of network architecture and high-level features. Please note that the work on TRs 32.864, 28.801 and 28.802 is still in progress.

The 3GPP TSG RAN has collected its requirements w.r.t. 5G in the TR 38.913 [3GPP-38913]. Except for requirements, this report includes also the KPIs defined as well as evaluation scenarios.

4.2.2 Generic requirements

The list of requirements described in the following contains generic requirements on the overall 5G system, including all parts of the network (radio, core, transport, etc.):

- 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.

- The 5G system shall support the concept of dedicated network slices, understood as the allocation of dedicated network resources to serve a defined business purpose, customer, or use case.
- The dependencies between 5G RAN and CN should be minimised to allow separate evolutionary steps, i.e., the architecture should be defined with a converged access-agnostic CN with a common RAN - CN interface which integrates different wireless access types (3GPP, non-3GPP), but which may also cover fixed access types (targeting Fixed-Mobile Convergence (FMC)[4]).
- The 5G system design shall support infrastructure sharing and multi-tenancy.
- The 5G system shall support the separation of User Plane (UP) functions (UPNFs) from Control Plane (CP) functions (CPNFs), allowing independent scalability, evolution and flexible deployments, e.g., centralised location or distributed (remote) location.
- 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 network slicing by adaptive placement of those NFs at different locations within the network infrastructure.
- 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.
- Wherever applicable, procedures (i.e., the set of interactions between NFs) shall be defined as services so that their re-use is possible.
- 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.
- The 5G system shall be able to manage both VNFs and PNFs. Fault, configuration, account, performance, and security (FCAPS) management, as well as lifecycle management (LCM) of VNFs, shall be based on NFV-MANO principles [ETSI-MAN001] [ETSI-WP5G].
- The 5G system shall support "stateless" NFs, where the "computational" resource is decoupled from the "storage" resource.
- The 5G system shall support a unified authentication and ID management framework across different access networks. Special provisions shall be offered for IoT devices (e.g., low-cost sensors) with limited capabilities.
- 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 user equipment (UE) or group of UEs.
- The 5G system shall be able to self-configure and self-heal in case of failures by the implementation of self-organising/optimisation network (SON) procedures.
- 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.
- The 5G system shall support different levels of resilience for the services provided.
- 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).
- 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.

4.2.3 Requirements on network slicing

Requirements regarding network slicing in 5G systems, including creation, removing, or modification of network slices are listed in the following:

---

[4] Note: FMC not in focus of 5G-MoNArch

- The 5G system should allow the realisation of Network Slice Instances (NSIs) across several infrastructure domains which may be owned by different parties (multi-domain operation).
- The 5G system shall allow the operator to create, modify and delete an NSI (or network slice subset instance (NSSI), and to define and update the set of services and capabilities supported in an NSI.
- The 5G system shall allow the operator to configure the information which associates a UE or a service to an NSI.
- The 5G system shall allow the operator to assign a UE to an NSI, to move a UE from one NSI to another, and to remove a UE from an NSI based on subscription, UE capabilities, the access technology used by the UE, operator's policies and services provided by the NSI.
- The 5G system shall enable a UE to be simultaneously assigned to and access services from more than one NSI of one operator.
- Traffic and services in one NSI shall be logically isolated from traffic and services in other NSIs in the same network.
- Creation, modification and deletion of an NSI shall have no or minimal impact on traffic and services in other NSIs on the same network.
- The 5G system shall support scaling of an NSI, i.e., adaptation of its capacity.
- 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.
- 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.
- The 5G system shall support means by which the operator can differentiate policy control, functionality and performance provided in different NSIs.
- In a shared 5G network configuration, each operator shall be able to apply all the requirements from this clause to their allocated network resources.
- The 5G system shall provide the implementation of suitable network slice management functions (NSMFs) allowing efficient fault, configuration, performance, lifecycle, and policy management of NSIs and NSSIs, respectively, also for automated (i.e., SON-based) reconfiguration, optimisation, and healing.

4.2.4 RAN-related requirements

In this section, requirements related to 5G RAN and UEs are listed. These requirements are as follows:

- The 5G RAN shall be highly scalable with respect to parameters like throughput, number of devices or number of connections to support a wide range of 5G service types with diverging requirements (like those spanned by the triangle of eMBB, mMTC, and URLLC).
- The 5G RAN shall be designed to operate in a wide spectrum range with diverse characteristics such as bandwidths and propagation conditions. For higher frequency bands such as millimetre wave (mmW), beamforming (BF) will become essential.
- 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 5G RAN should enable a tight interworking between different 5G air interface variants (AIVs) incl. LTE-A Pro evolution.
- 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 AIV (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 5G system shall support UEs with single and multiple radio capabilities.
- When a UE is able to use two or more AIVs simultaneously, the 5G system shall be able to select between AIVs in use, taking into account, e.g., service, traffic characteristics, radio characteristics, and UE's moving speed.

- The 5G RAN shall natively support network-controlled device-to-device (D2D) communication (side link, i.e., point-to-point, multicast and broadcast), both in as well as out of coverage of TRxPs (important for, e.g., V2X scenarios).
- The 5G RAN shall enable operators to support wireless self-backhauling using 5G AIVs with flexible resource partitioning between access and backhaul functions. The self-backhauling functionality shall be multi-hop capable and support topologically redundant connectivity.
- The 5G RAN should support a radio resource management (RRM) covering both AIV-specific and AIV-overarching aspects for data traffic steering/aggregation and interference management.
- The 5G RAN design must be energy efficient. This means that permanently active NFs or signals transmissions (e.g., reference symbols) have to be avoided.
- The 5G RAN shall minimise the signalling that is required prior to user data transmission.
- The 5G RAN shall optimise the CP and UP resource usage for data transmission taken care of UE capabilities (mobility type, location, …), communication pattern (e.g., Tx-only, frequent or infrequent), payload characteristics (e.g., small or large size data), application characteristics (e.g., provisioning operation, normal data transfer) etc.
- The 5G system shall be able to support seamless mobility/handover between the 5G AIVs and/or TRxPs.
- The 5G RAN shall support different means for reducing UE power consumption while UE is in periods with data traffic as well as in periods without data traffic.
- The 5G RAN shall support the operation of downlink only broadcast/multicast over specific geographic areas (e.g., a cell sector, a cell or a group of cells) to both stationary and mobile UEs.
- Different options and flexibility for splitting the 5G RAN architecture shall be allowed. This shall cover both the horizontal split (split between or inside radio protocol stack layers) and the vertical split (i.e., CP/UP separation).
- The 5G RAN shall be designed such that it can maximally leverage from the centralised processing of radio layers, but also operate well in the case of distributed TRxPs with imperfect x-haul (back-/mid-/front-haul) infrastructure (split between a centralised unit (CU) and distributed units (DUs) near the antenna sites).
- The RAN-CN interfaces and RAN-internal interfaces shall be open for multi-vendor interoperability.

4.2.5 Capability Exposure Requirements

In this section requirements for 5G systems related to network capability exposure and UP handling in the case of local service hosting are given.

- The 5G system shall support capability exposure to third parties through a set of open Application Programming Interfaces (APIs), allowing different provider business models to be implementable (e.g., XaaS).
- Based on operator policy, the 5G system shall provide suitable APIs to allow a trusted third party to create, modify, delete and monitor NSIs used for the third party as well as to manage the set of services and capabilities within and access of UEs to those NSIs (incl. capacity and QoS features).
- 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.
- 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.
- Based on operator policy, the 5G system shall provide suitable APIs to allow a trusted third party to manage and monitor this trusted third party’s owned application(s) in the operator's service hosting environment (e.g., for offloading purposes close the UEs' location).
- 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.

- 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.

4.2.6 Security requirements

This section lists some more general high-level requirements with respect to security issues a 5G system has to fulfil. As extension of that list some more dedicated security topics will be described in following Section 4.3 that will be covered in more detail by innovative solutions to be derived within 5G-MoNArch.

- The 5G system shall be designed in a way that it secures the network, its users and their traffic effectively against cyber-attacks, and may provide flexible security mechanisms that can be tailored to the needs of the different use cases that are supported.
- The 5G system shall provide mechanisms to verify the integrity of radio messages. These mechanisms shall allow the detection of unauthorised radio messages, detection of "false base stations" and verification of an authorised network. The mechanisms defined should cater for high-speed communications envisioned in 5G and for battery efficient low volume data as well.
- The security mechanisms defined in the 5G system shall be able to be configured to comply with local lawful interception laws and regulations.
- The security mechanisms defined in the 5G system shall be able to be configured to confidentially protect voice, data and signalling, as well as subscriber's privacy.
- The security mechanisms defined in the 5G system shall be able to be configured to provide authorisation services for users, devices and networks both at a bearer level and at a services level.
- The security mechanisms defined in the 5G system shall be able to be configured to provide authorisation, integrity protection and confidentiality between network elements and between networks.
- The security mechanisms defined in the 5G system shall be able to be configured to provide authorisation, integrity protection and confidentiality for new 5G services.
- As the 5G system networks may be active up to and beyond 2030 and as the ability to attack security mechanisms increases over time, the security mechanisms specified for the 5G system shall be extensible to new algorithms and procedures that will be defined during the lifetime of the specifications, where appropriate.
- The CP shall be protected against denial of service attacks from UEs. Mechanisms should be specified which limit the effect which signalling attacks may cause to the network. Signalling caused by UEs should not be able to degrade the network performance for other end users and the network itself.
- UEs shall be protected against denial of service attacks from the network. Mechanisms should be specified which limit the effect which signalling attacks may cause to UEs. Signalling caused by the network should not be able to degrade the network performance for end users.
- UEs and the 5G network should be protected against denial of service attack from external networks, e.g., the internet, and from other UEs. The impact to network and end user signalling or data processing due to external attacks should be minimised. Signalling and data processing caused by external network traffic should not degrade the network performance for end users and the network itself, as well as the UE performance, e.g., the power consumption.

4.3 Resilience and Security Requirements

Besides the general requirements listed in Section 4.2, the 5G-MoNArch project puts particular attention to innovation aspects related to security and resilience as part of WP3, as well as part of the sea port testbed. In this respect, specific criteria have been set for assessing whether the operation of the considered architecture design meets the required standards in terms of resilience and security.

For a consolidated study of the above criteria, they are listed in three major sets, depending on whether they reflect the ability of the network

1. **Prevent design space:** to prevent by design that a problem or malfunction takes place or at least to minimise its probability;
2. **Problem space:** to detect such a problem or malfunction in an efficient manner, ideally before it has any drastic impact on the system or service performance;
3. **Solution space:** to effectively react to such problems or malfunctions after they have been detected.

In this context, the three sets of criteria mentioned above correspond to different system requirements. Specifically, we distinguish between malfunction **protection requirements**, **detection requirements**, and **reaction requirements**. These three major sets are employed for quantifying and evaluating the design of the network and are elaborated separately below.

A. **Protection requirements,** refer to requirements used to define how efficiently the network can protect itself from encountering any type of malfunctions. This category of requirements falls into the "prevention design space", implying the set of design criteria that prevent the occurrence of network faults. The major requirements in this regard that will be considered by 5G-MoNArch are as follows:
   - 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. 5G-MoNArch will take care of that requirement by deriving and evaluating specially tailored multi-connectivity based solutions. Those may rely on efficient traffic flow steering between different available links or on approaches involving data duplication across the radio links with or without applying network coding schemes [MVD16] [GDK08].
   - The design of the telco cloud to be integrated into the 5G system shall be in a way that it prevents the network performance to degrade or at least keeps the degradation on a minimum level. Similar to the approach for the RAN design, 5G-MoNArch will counteract network faults in the telco cloud with the use of prior added redundancy for the sake of resilience [WLA10].

It is important to note that such redundancy both for the radio link and for the telco cloud does not come for free, implying that the cost of such network infrastructure design should be minimised also. Therefore, any measure used in this respect has to reflect not only the achievable performance w.r.t. reliability and availability required to keep certain service level agreements (SLAs) but also the cost efficiency of such a redundancy-based design. Those techno-economic evaluations are also part of the verification and validation process within WP6 of 5G-MoNArch.

B. **Detection requirements** are associated to requirements related to the so-called "problem space". That is, the requirements for identifying a problem to the network of any kind. These include:
   - The 5G system shall support the detection of network faults or malfunctions, ideally before those have any drastic impact on system or service performance. 5G-MoNArch will work on correlation algorithms that are directed towards detecting a problem or a fault in network components and identifying its root cause. The assessment of such algorithms will be carried out on the basis of their effectiveness to detect and diagnose a malfunction [ZOY03].
   - The 5G system shall support the detection of security threats. To identify security threats at an early stage 5G-MoNArch will evaluate specially designed algorithms for anomaly detection. Similarly, as with network malfunctions, classifying such anomalies as security threats is imperative for a healthy operation of the network. The requirement of detecting network faults includes their ability to spot anomalies in cases of massive threat attacks. This implies that such criterion is not a fixed one, but is flexible enough to take into account that the required level of security is maintained even for unexpected situations where an unusual number of security threats is present [SWG+16].

C. The **reaction requirements** correspond to the performance measure used to assess the ability of the system to recover after a problematic functionality has occurred. This is also referred to as the "solution space". More precisely, the reaction requirements include:
   - The 5G system shall be designed in a way that in case of a network malfunction the service restoration time (cf. Section 5.2) is minimised. A network malfunction may entail an abrupt

interruption of a network or customer service. Within the concept of network resilience, it is crucial that such service interruption is limited to the minimum possible extent. The design of the 5G system by 5G-MoNArch will guarantee that the effect of network malfunction is mitigated in a way that is required by the service. The network's service restoration ability is measured via the service restoration time which corresponds to the time between when a malfunction has started (independently of whether this has been diagnosed or not, c.f. network fault detection requirement), until the service has been completely recovered [LSW+16].
- The 5G system shall provide measures to react to security attacks and to mitigate their impact. When the network is under security attack, it is imperative that certain measures are taken to ensure that such attacks affects its performance as little as possible. 5G-MoNArch will design the system to be able to isolate security intrusions into given areas by utilising the concept of security trust zones. This concept states that parts of the network are isolated from each other to make sure that security intrusions remain geographically, as well as functionally, limited [HWM+17].

In summary, the **protection requirements** refer to a targeted network design towards minimising network faults, the **detection requirements** reflect the ability of the system to detect and diagnose a malfunction or anomaly, while the **reaction requirements** focus on the ability of the network to withstand such undesirable situations. All the above groups of requirements affect an important part of the system evaluation, as far as resilience and security are concerned. To evaluate the performance and effectiveness of solutions to be derived within 5G-MoNArch to fulfil these requirements special related KPIs are defined in Chapter 5.

4.4 Resource Elasticity Requirement

According to the Merriam-Webster dictionary, one definition of elasticity is "the quality of being adaptable," while another one relates to resilience, i.e., the "ability to recover from or adjust easily to misfortune or change" [Merr17]. Building on these definitions, elasticity considering both axes, namely, adaptability and recoverability is defined.

The elasticity of a network slice can be defined as its ability to gracefully adapt to load changes in an automatic manner such that at each point in time the available resources match the demand as closely and efficiently as possible. In this way, there are a number of key concepts that have to be addressed when dealing with elasticity:

- The network slice under study;
- One or more resources, supporting the network slice;
- One or more outputs, resulting from the execution of the slice.

The concept of resource elasticity can be analysed from three different perspectives, each of them being a fundamental piece needed to bring elasticity to the network operation:

- Elasticity at the VNF Level;
- Elasticity at the Network Slice Level;
- Elasticity at the Infrastructure Level.

The first item is probably the more relevant to the project, as elastic network slices that run on an elastic infrastructure are built on elastic VNFs. The latter two are related to the orchestration levels. The description of the evaluation criteria for each of them is provided in the following.

4.4.1 Elasticity at the VNF Level

The concept of VNF elasticity is also related to both scalability and efficiency. Scalability of a system (i.e., its ability to meet a larger load demand by adding a proportional amount of resources) is a prerequisite for elasticity but it does not consider temporal aspects of how fast and how often scaling actions can be performed. Efficiency and elasticity are related as better elasticity results in higher efficiency, but the implication is not always true in the other direction as efficiency can be influenced by factors not related to elasticity [HKR13].

A general framework to discuss the key concepts and their relations, prior to defining the criteria for the evaluation of elasticity, is provided in Figure 4-1.

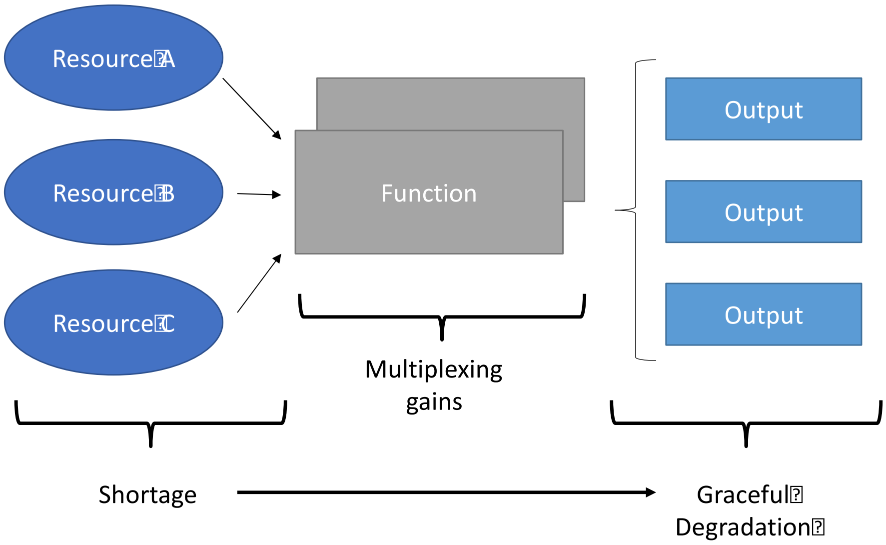

Figure 4-1:Context for the execution of an elastic function.

The concept of elasticity for a NF was not directly applicable to a legacy PNF. There, and especially for the distributed ones, the functionality is provided by a physical box that is the result of a thorough joint HW/SW design. Therefore, they have traditionally been designed without any major constraint on the available execution resources: they were expected to be always available *by design*.

In networks with centralised VNFs, the joint HW/SW design is not possible anymore: VNFs are pieces of SW that run on virtual containers on heterogeneous (with standard interfaces) cloud platforms. Therefore, in this new but already widely adopted scenario, expecting that all the needed resources are always available by design is not the case anymore.

Unfortunately, current VNFs (especially RAN ones) have been designed under the assumption that required computational resources are always available and they may not be prepared to manage this situation. Indeed, when such resource (e.g., CPU) outages occur, current virtualised RAN implementations [NMM+14] just drop the frame being processed, and as a result they see their performance severely degraded.

Under ideal circumstances, i.e., no shortage or variation of the resources, a function has to operate reliably as in the case of a "classical" NF. Given that functions will interact with each other, we follow the framework of [ETSI-REL003] for models and end-to-end calculations and define **availability** as the relative amount of time that a given function is providing the expected output (i.e., its uptime):

$$Availability = Uptime / (Uptime + Downtime)$$

It is to be noted that in the classical approach this reliability referred to the availability of a communication resource (e.g., "five nines reliability" for an availability of 99.999%), and therefore was challenged by congestion or degradation of links, but in the 5G-MoNArch vision there are more resources that impact performance.

In fact, in general the **resources** supporting the execution of the considered function will be heterogeneous, and therefore care should be taken when performing experiments, in order to compare apples to apples. Specifically, VNF execution is supported by IT resources (CPU, Disk, RAM), as well as networking and communications resources (e.g., transport network, spectrum), so comparing them will need a careful design of comparable metrics. Moreover, as there is already a comprehensive state-of-the-art available on how to cope with shortages of networking resources, the focus here will particularly be on IT resources.

An elastic VNF should thus be able to cope with variations in the availability of these resources, without causing an abrupt degradation in the outputs provided by the function. Depending on the relation between this shortage and the reduced performance that is obtained, a **graceful degradation** of a function can be defined if, e.g., the percental decrement of resources is bound by a similar or smaller percental reduction in performance. Furthermore, a function will require a **minimum set of resources** (minimum footprint) to be able to provide any output. Similarly, when the load in the system is low, elastic VNFs should behave in an elastic manner by avoiding overprovisioning of unnecessary resources (e.g., by consolidating virtual machines (VMs) on a common HW platform).

The ability of a function to adapt to variations in the available resources could result in a more efficient operation, which can be interpreted in terms of two related concepts: on the one hand, to serve the same set of VNFs an elastic system requires less (IT but also energy) resources, i.e., it can be better dimensioned and operated, thus leading to a **cost efficiency gain**, related to the "resource footprint" of, e.g., an elastic function or an orchestration mechanism, and the inelastic counterparts; on the other hand, with the same set of resources, it may support a larger number of VNFs, improving the **resource utilisation efficiency**, by exploiting multiplexing gains or deploying VNFs where resource are available.

Building on the above, there are different axes to evaluate the elasticity of a NF:

- In terms of its output: reliability;
- In terms of its minimum set of resources needed: its minimum footprint;
- In terms of the impact of resource shortage on its output: the degradation;
- In terms of system-wide perspective: the cost efficiency or resource utilisation efficiency gains.

4.4.2 Elasticity at Network Slice Level

The elastic design of a VNF, as described in the previous section has an impact on the elasticity of a network slice, defined as the chain of VNFs that provide a telecommunication service. Indeed, chaining a sequence of VNFs with different elastic KPIs (as described above) will result into an overall elasticity associated to a tenant running a service using a network slice. This ultimately affects the QoE/QoS perceived by users, that may experience different performance degradations according to the elasticity level provided by the tenant.

Following the analytical framework provided in [ETSI-REL003] (namely, "modelling of composed systems"), service-related metrics can be derived based on the performance of the individual involved modules. In this way, if a service is composed of two independent components that run in a *serialised* manner, the output depends on the contemporary availability of the two modules (i.e., both must be active for the service to work), as illustrated in Figure 4-2:

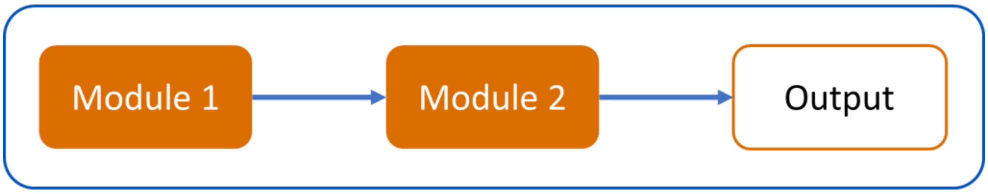

Figure 4-2: Service composition: serialised modules

Then its **availability** is given by the product of the corresponding single availabilities:

$$Availability = Availability_1 \times Availability_2$$

In contrast, if a service is composed of *redundant* modules, running in parallel to improve resilience (at the cost of potentially an increased resource consumption), as illustrated in Figure 4-3:

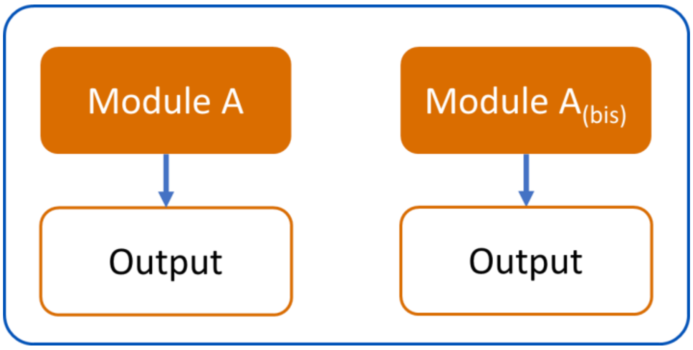

Figure 4-3: Service composition: parallelised modules

Then the availability of the service would be given by:

$$Availability = 1 - (1\text{-}Availability_A) \times (1 - Availability_{(bis)})$$

Depending on the specifics of the service under consideration, this including the behaviour of each of the modules composing the service, more complex analysis can be performed to derive the corresponding KPIs of interest, such as the resource consumption, the function of performance degradation, or the minimum set of resources required.

4.4.3 Elasticity at the Infrastructure Level

The last dimension of elasticity is the one that involves the infrastructure on which elastic NFs run. The choice of how many network slices are hosted in the same infrastructure depends on the infrastructure provider, that runs, for example, admission control algorithms to guarantee that the SLA with the various tenants are always fulfilled.

Elasticity at the infrastructure level is a metric that involves both business and technical KPIs. By leveraging multiplexing gains, more network slices can be hosted on the same infrastructure (providing hence higher revenues), but this comes at the cost of having to resort to more elastic VNFs.

For the elasticity at infrastructure level, KPIs will be used that provide insight to the effects of overbooking, i.e., the average performance loss in comparison with the monetary gains that a larger number of network slices may provide, or specific numbers on how many network slices can be hosted by a given infrastructure deployment.

5 Key Performance Indicators

This chapter provides an overview about KPIs which are relevant for the verification and validation process within 5G-MoNArch. Section 5.1 includes a collection of KPI definitions already provided by SDOs like 3GPP and ETSI, regulatory bodies (e.g., ITU-R), and other forums being important for specification of the 5G system (e.g., 5G PPP and NGMN). 5G-MoNArch will take into account those definitions as far as possible; and, if required, additional KPIs will be defined for evaluation of innovations to be derived within the project (see Sections 5.2 and 5.3). In addition, Section 5.4 covers application-specific KPIs to be used especially for the assessment of achievable service performance in the two testbeds. Section 5.5 lists KPIs to be applied in the techno-economic evaluations of 5G-MoNArch.

5.1 General KPIs

The general KPIs given in this section are based on a collection taken from SDOs, regulatory bodies and forums like 5G PPP and NGMN.

Please note that the listing is made in alphabetical order, i.e., the sequence is not according to the relevance of the different KPIs.

Area traffic capacity

(based on 3GPP/ITU-R; aka **traffic volume density** by 5G PPP)

The total traffic throughput served per geographic area (in bps/m^2). This metric can be evaluated by two different traffic models: Full buffer model and non-full buffer model.

- By full buffer model: The computation of the total traffic throughput served per geographic area is based on full buffer traffic.
- By non-full buffer model: Again, the total traffic throughput served per geographic area is computed, but in addition to the area traffic capacity also the **user experienced data rate** need to be evaluated at the same time using the same traffic model.

The area traffic capacity is a measure of how much traffic a network can carry per unit area. It depends on site density, bandwidth, and spectrum efficiency. In the case of full buffer traffic and a single layer single band system, it may be expressed as:

$$\textit{Area traffic capacity (bps/m}^2\textit{)} = \textit{site density (site/m}^2\textit{)} \times \textit{bandwidth (Hz)} \times \textit{spectrum efficiency (bps/Hz/site)}$$

Site here refers to a single transmission and reception point (TRxP). It is proposed to perform full buffer evaluation, using the spectrum efficiency results together with assumptions on available bandwidth and site density in order to derive a quantitative area traffic capacity KPI for information.

Area traffic capacity is typically evaluated through system level simulations. Note that D2D traffic should be evaluated independently from the cellular one. Besides, the link between source and destination may cover multiple hops especially when non-ideal backhaul is taken into consideration.

Availability

(based on 3GPP/5G PPP/NGMN/ETSI)

Percentage value (%) of the amount of time a system is in condition to deliver services divided by the amount of time it is expected to deliver services in a specific area.

The availability may be specific for a communication service. In that case, it refers to the percentage value of the amount of time the end-to-end communication service is delivered according to an agreed QoS, divided by the amount of time the system is expected to deliver the end-to-end service according to the specification in a specific area.

Note 1: The end point in "end-to-end" is assumed to be the communication service interface.

Note 2: The communication service is considered unavailable if it does not meet the pertinent QoS requirements. If availability is one of these requirements, the following rule applies: the system is considered unavailable in case an expected message is not received within a specified time, which, at minimum, is the sum of end-to-end latency, jitter, and survival time.

The RAN availability is characterised by its availability rate [x], defined as follows: a TRxP is available for the targeted communication [x]% of the time. Unavailable communication for shorter period than [y] ms shall not be counted. The 5G specification's ability to provide URLLC services shall not be compromised by the functions defined to improve the network or UE energy efficiency, or by system reconfigurations and software upgrades. The availability may be targeted for the whole coverage area of a TRxP or for [z]% of the locations within the area.

Considering NFV-based implementations of a system the service availability may be also defined as the total service available time divided by the sum of total service available time and total restoration time required to set up the service again [ETSI-REL003]. In that case, typically only the communication path in the network infrastructure is considered, i.e., without involvement of UE and radio link.

In the classical resilience theory, the availability of a system/component is defined as the relative uptime, or more precisely as the mean time between failures (MTBF) divided by the sum of MTBF and mean time to repair (MTTR) (e.g., [ETSI-REL003] w.r.t. availability considerations for systems consisting of NFV-based components).

Bandwidth

(based on 3GPP)

Bandwidth means the maximal aggregated total system bandwidth. It may be supported by single or multiple RF carriers. It is a quantitative KPI.

Cell-edge user throughput

(based on 3GPP)

The cell edge user throughput is defined as the fifth percentile point of the CDF of user's average packet call throughput

Connection density

(based on 3GPP/ITU-R)

The total number of connected and/or accessible devices per unit area (per km2). Connectivity or accessibility refers to devices fulfilling a target QoS, where the target QoS is to ensure a system packet drop rate less than [x]% under given packet arrival rate [l] and packet size [S]. The packet drop rate is equal to the number of packets in outage divided by the number of generated packets, where a packet is in outage if this packet failed to be successfully received by destination receiver beyond a packet dropping timer.

Analytical, link level, and system level evaluations can be performed to derive the connection density in a certain deployment scenario (note: 3GPP proposes to use an Urban environment for mMTC related evaluations).

Control plane latency

(based on 3GPP/5G PPP)

Time to move from a battery efficient state (e.g., IDLE) to the start of continuous data transfer (e.g., ACTIVE).

Analytical evaluation is typically used as the evaluation methodology.

Coverage

(based on 3GPP)

Maximum coupling loss (MaxCL) in UL and DL between UE and TRxP (antenna connector(s)) for a data rate of [x] bps, where the data rate is observed at the egress/ingress point of the radio protocol stack in each direction.

Link budget and/or link level analysis are typically used as evaluation methodology.

CL is defined as the total long-term channel loss over the link between UE antenna ports and the TRxP antenna ports, and includes in practice antenna gains, path loss, shadowing, body loss, etc. The MaxCL is the limit value of CL at which the service can be delivered, and therefore defines the coverage of the service. The MaxCL is independent of the carrier frequency. It is defined in the UL and DL as:

UL MaxCL = UL Max Tx power - TRxP Sensitivity

DL MaxCL = DL Max Tx power - UE Sensitivity

Note: 3GPP proposed a target for coverage of 164 dB for an mMTC service assuming 160 bps. For a basic MBB service characterised by a DL/UL data rates of 2(1) Mbps/60(30) kbps for stationary users, 3GPP proposed a target MaxCL of 140(143) dB. For mobile users 3GPP assumes a DL data rate of 384 kbps as acceptable. At a coupling loss of 143 dB relevant DL/UL control channels should also perform adequately.

Coverage area probability

(based on 5G PPP)

The coverage area probability is defined as the percentage of the area under consideration, in which a service is provided by the mobile radio network to the end user in a quality (i.e., data rate, latency, packet loss rate) that is sufficient for the intended application (QoS/QoE level). The RAN may consist of a single radio cell or of a multi-cell deployment. For services of different types and QoS/QoE levels the coverage area probability will be also typically different. For radio network planning purposes or coverage measurements the area under consideration will be usually divided into 2-dimensional pixels or segments (e.g., along roads) with the same size.

End-to-end latency

(based on 3GPP/5G PPP)

The time that takes to transfer a given piece of information from a source to a destination, measured at the communication interface, from the moment it is transmitted by the source to the moment it is successfully received at the destination. It is also referred to as one trip time (OTT) latency.

Another latency measure is the round-trip time (RTT) latency which refers to the time from when a data packet is sent from the transmitting end until acknowledgements are received from the receiving entity.

Energy efficiency

(based on 3GPP/ITU-R)

Energy efficiency means to sustain a certain data rate while minimising the energy consumption. It has two aspects:

- On the network side, energy efficiency refers to the quantity of information bits transmitted to/received from UEs, per unit of energy consumption of the RAN (in bit/Joule).
- On the UE side, energy efficiency refers to quantity of information bits per unit of energy consumption of the communication module (in bit/Joule).

It is a qualitative KPI. The evaluation methodology should be based on inspection, but more detailed quantitative assessment can be performed.

Latency for infrequent small packets

(based on 3GPP)

For infrequent application layer small packet/message transfer, the time it takes to successfully deliver an application layer packet/message from the radio protocol layer 2/3 SDU ingress point at the UE to the radio protocol layer 2/3 SDU egress point in the RAN, when the UE starts from its most "battery efficient" state.

Analytical evaluation is the baseline evaluation methodology and system level evaluation can be considered if needed.

Mean time between failures (MTBF)

(by ETSI)

The MTBF is the statistic mean uptime of a system/component before it fails.

Note: The MTBF is a property of the system/component and hardly can be influenced by the operator (except by ensuring that it is kept within its specified operational limits).

Mean time to repair (MTTR)

(by ETSI)

The MTTR is the statistic mean downtime before the system/component is in operations again.

Note 1: In contrast to the MTBF the MTTR is determined solely by procedural aspects, namely how long will it take to organise a spare part and to get the repair work done.

Note 2 (extension by 5G-MoNArch): In contrast to former telco network generations, where systems are primarily characterised by exchange of HW boxes in the repair case, for more SW-oriented systems using NFV approaches the time for SW fault localisation, remediation, and recovery without the need to exchange the underlying HW will be considered which may drastically reduce the resulting MTTR and therefore the reliability (see also the definition of the **service restoration time** in Section 5.2).

Mobility

(based on 3GPP/ITU-R)

Maximum speed at which a defined QoS and seamless transfer between TRxPs which may belong to different deployment layers and/or radio access technologies (multi-layer/-RAT) can be achieved (in km/h).

The evaluation methodology should be link level evaluation with deployment scenario specific operating point.

Mobility interruption time

(based on 3GPP/5G PPP)

The shortest time duration supported by the system during which a UE cannot exchange UP packets with any TRxP during transitions. This KPI is for both intra- and inter-frequency mobility as well as for intra- and inter-AIV mobility.

Analytical evaluation is typically used as the evaluation methodology.

Ideally, the mobility interruption time should be 0 ms, which may be achievable by multi-connectivity and CP/UP decoupling.

Peak data rate

(based on 3GPP/ITU-R/5G PPP)

Highest theoretical single user data rate (in bps), i.e., assuming ideal, error-free transmission conditions, when all available radio resources for the corresponding link direction are utilised (i.e., excluding radio resources that are used for physical layer synchronisation, reference signals or pilots, guard bands and guard times).

Analytical evaluation is typically used as the evaluation methodology.

Peak spectral efficiency

(based on 3GPP)

The peak data rate normalised by the bandwidth applied. Higher frequency bands could have higher bandwidth but lower spectral efficiency and lower frequency bands could have lower bandwidth but higher spectral efficiency. Thus, peak data rate cannot be directly derived from peak spectral efficiency and bandwidth multiplication.

Analytical evaluation is typically used as the evaluation methodology.

Reliability

(based on 3GPP/ITU-R/5G PPP/NGMN)

Percentage (%) of the amount of sent network layer packets successfully delivered to a given system node (incl. the UE) within the time constraint required by the targeted service, divided by the total number of sent network layer packets.

Note 1: The reliability is evaluated only when the network is available.

Note 2: Dependent on the targeted service the RTT latency instead of the E2E (OTT) latency may be applied.

The RAN reliability can be evaluated by the success probability of transmitting X bytes within a certain delay of [t] ms, which is the time it takes to deliver a data packet from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point of the radio interface. The target communication range and reliability requirement is dependent of the selected deployment and operation scenario, i.e., by taking into account a certain channel quality (e.g., at the coverage edge).

Link level evaluation with deployment scenario specific operating point and system level simulations are to be performed (e.g., Indoor Hotspot and Urban Macro for eMBB; Highway and Urban grid for connected cars/URLLC).

In the classical resilience theory, the reliability of a system/component over time is directly related to its MTBF. In the simplified case that the MTBF will not change over the system's/component's lifetime, it can be calculated as follows:

$$R(t) = exp(-t/MTBF).$$

It should be noted, however, that the MTBF of most systems/components (respectively their failure rate = 1/MTBF) will change significantly over time (e.g., [ETSI-REL003] w.r.t. reliability considerations for systems consisting of NFV-based components).

Resilience

(based on ITU-R)

Resilience is the ability of the network to continue operating correctly during and after a natural or man-made disturbance, such as the loss of mains power.

Service continuity

(based on 3GPP)

The uninterrupted user experience of a service that is using an active communication when a UE undergoes an access change without, as far as possible, the user noticing the change.

Note 1: In particular service continuity encompasses the possibility that after a change the user experience is maintained by a different telecommunication service (e.g., tele- or bearer service) than before the change.

Note 2: Examples of access changes include the following. For EPS: CS/PS domain change. For EPS and 5G: radio access change, switching between a direct network connection and an indirect network connection. Indirect in that sense means using a relay node (TRxP or another device) in the connection to the end customer device (e.g., via D2D/slide link mode).

Spectral efficiency per cell/transmission and reception point (TRxP)

(based on 3GPP/ITU-R)

TRxP spectral efficiency is defined as the aggregate throughput of all users (the number of correctly received bits, i.e., the number of bits contained in the service data units (SDUs) delivered to Layer 3, over a certain period of time) within a radio coverage area (site) divided by the channel bandwidth divided by the number of TRxPs. A 3-sector site consists of 3 TRxPs.

In case of multiple discontinuous "carriers" (one carrier refers to a continuous block of spectrum), this KPI should be calculated per carrier. In this case, the aggregate throughput, channel bandwidth, and the number of TRxPs on the specific carrier are employed. It is a quantitative KPI.

Spectrum and bandwidth flexibility

(based on ITU-R)

Spectrum and bandwidth flexibility refers to the flexibility of the 5G system design to handle different scenarios, and in particular to the capability to operate at different frequency ranges, including higher frequencies and wider channel bandwidths than today.

UE battery life

(based on 3GPP)

Life time of the UE battery to be evaluated without recharge. Analytical evaluation is typically used as the evaluation methodology.

Note: For mMTC, 3GPP proposed that UE battery life in extreme coverage shall be based on the activity of mobile originated data transfer consisting of 200 bytes UL per day followed by 20 bytes DL from MaxCL of 164 dB, assuming a stored energy capacity of 5 Wh.

User experienced data rate

(based on 3GPP/ITU-R; aka **experienced user throughput** by 5G PPP)

The achievable data rate that is available ubiquitously across the coverage area to a mobile user/device (in bps). It is usually related to the minimum data rate required to achieve a sufficient quality experience (dependent on the selected service type).

The user experienced data rate can be evaluated for non-full buffer traffic and for full buffer traffic, but non-full buffer system level simulations are preferred for the evaluation of this KPI taking care of respective deployment scenarios and using bursty traffic models.

For non-full buffer traffic, the user experienced data rate is the 5%-percentile (5%) of the user throughput. User throughput (during active time) is defined as the size of a data burst divided by the time between the arrival of the first packet of a burst and the reception of the last packet of the burst.

For full buffer traffic, user experienced data rate is calculated as:

User experienced data rate = 5% user spectrum efficiency × bandwidth

Here it should be noted that the 5% user spectrum efficiency depends on the number of active users sharing the channel (e.g., 10 in the evaluations in ITU-R Report M.2135), and that the 5% user spectrum efficiency for a fixed transmit power may vary with bandwidth. To keep a high 5% user spectrum efficiency and a few users sharing the channel, a dense network is beneficial, i.e., 5% user spectrum efficiency may vary also with site density (site here refers to single TRxP).

5% user spectrum efficiency means the 5% point of the cumulative distribution function (CDF) of the normalised user throughput. The (normalised) user throughput is defined as the average user throughput (the number of correctly received bits by users, i.e., the number of bits contained in the SDU delivered to Layer 3, over a certain period of time), divided by the channel bandwidth and is measured in bps/Hz. The channel bandwidth for this purpose is defined as the effective bandwidth times the frequency reuse factor, where the effective bandwidth is the operating bandwidth normalised appropriately considering the uplink/downlink ratio. In case of multiple discontinuous "carriers" (one carrier refers to a continuous block of spectrum), this KPI should be calculated per carrier. In this case, the user throughput and channel bandwidth on the specific carrier are employed.

The user experienced data rate is calculated separately for DL (transmission from TRxP(s) to UE), UL (transmission from UE to TRxP(s) and (potentially) for D2D (transmission directly between involved UEs).

User plane latency

(based on 3GPP/5G PPP)

The time it takes to successfully deliver an application layer packet/message from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point via the radio interface in both UL and DL directions, where neither UE nor TRxP reception is restricted by DRX.

Analytical evaluation is typically used as the evaluation methodology. The evaluation needs to consider all typical delays associated with the transfer of the data packets in an efficient way (e.g., applicable procedural delay when resources are not pre-allocated, averaged HARQ retransmission delay, impacts of network architecture).

Note: that the reliability KPI also provides a latency value with an associated reliability requirement.

5.2 Resilience and Security KPIs

In this section, the specific KPIs are listed which are relevant to the network design with respect to resilience and security (see the corresponding requirements in Section 4.3).

In addition to the KPIs listed in Section 5.1, namely **reliability** (based on 3GPP/ITU-R/5G PPP/NGMN), **resilience** (based on ITU-R), and **mean time to repair** (by ETSI) the following indicators are foreseen to play an important role in the network design from the resilience and security point of view.

End-to-end reliability

This KPI equals the probability that all network components, including the virtualised and non-virtualised part of the network, are capable to support a required function (taken from the set of computation; networking; storage) for a given time interval (c.f. "Reliability" and "Reliability of telco cloud" KPIs discussed in Sections 5.1 and 5.2, respectively).

Note: This KPI refers to the end-to-end performance, since if any of the network parts (i.e., RAN, transport, telco cloud) are not capable of meeting the requirements, the QoS as seen from the UE perspective is dropped. As such, this is a combined KPI, where all network elements from the access network to the core network are involved.

Reliability of the telco cloud

Probability that a telco cloud component can perform a required function (taken from the set of computation; networking; storage) under stated conditions for a given time interval.

For example, in case of networking, this can be the percentage (%) of time that a defined set of I/O ports (of a data centre or a NFVI-PoP[5]) supports I/O packet processing at a given data rate for a given time interval. That is, the set of ports would be the "telco cloud component", I/O packet processing would be the "required function", and the given data rate would be the "stated condition(s)".

This KPI reflects the ability of a telco cloud to withstand any network faults or malfunctions which might have negative impact on system or service performance.

Note: This KPI refers to the infrastructure level, and is not directly related to the RAN reliability defined in Section 5.1.

Service restoration time

Time span required between a point in time when a service related malfunction has started (independently of whether this has been diagnosed or not, c.f. network fault detection requirement), until the service has been completely recovered.

With this KPIs the 5G system will be assessed in terms of its ability to restore an affected service within a given, usually strongly limited time.

Security threat identification

Percentage (%) of security threats (where any type of security intrusion attempt is regarded as security threat) that are identified by threat identification algorithms.

With this KPI the effectiveness of security threat algorithms for anomaly detection derived within 5G-MoNArch will be evaluated.

[5] Point of Presence

This KPI is evaluated on a simulation basis, where artificial threats are created in order to assess the effectiveness of the proposed threat identification algorithms.

Security failure isolation

Percentage (%) of propagated security failures, i.e., of security failures that pass the security zone (i.e., the zone where certain security measures to be implemented).

This KPI is also evaluated in a simulation environment, where the ability of the 5G system to isolate artificially security failures is assessed.

Besides the KPIs introduced above, end-to-end KPIs will be considered, which are related to the overall system performance as described below.

5.3 Resource Elasticity KPIs

For resource elasticity, a scenario is assumed where an NF is running over a set of resources to produce a set of outputs. Furthermore, it is assumed that for each novel function under consideration there is an inelastic counterpart, i.e., the "classical" version that provides the required functionality. Furthermore, we note that the period of time (when applicable) is implicit in the definitions below, and its actual length may have notable implications when performing a characterisation of a system (e.g., a system that is not available during 1 second over a period of 10 seconds has the same relative unavailability as a system that is not available during 1 hour over a period of time of 10 hours).

The following KPIs in alphabetical order are given with respect to assessing the fulfilment of requirements listed in Section 4.4:

Availability

Availability is defined as the relative amount of time that the function under study produces the output that it would have produced under ideal conditions. This is aligned with the availability concept described in Section 5.1, but with a specific focus on the resource provisioning.

Cost efficiency gain

This metric measures the average cost of deploying and operating the network infrastructure to support the foreseen services. An elastic system should be able to be optimally dimensioned such that less resources are required to support the same services; in addition, in lightly loaded scenarios the elastic system should avoid the usage of unnecessary resources and reduce the energy consumption (thus limiting the operation expenditure).

Elasticity orchestration overhead

The amount of resources required for realising orchestration functions, i.e., functions that enable NF elasticity (such as the re-placement of a VNF) and are not part of the traditional architecture. An example could be the vector that includes the amount of CPU, RAM, and the amount of networking resources (e.g., the minimum bandwidth needed for the outlet links in a VNF), consumed by the orchestration function.

Minimum footprint

Given a set of resources to execute a function, the minimum footprint is the set of combinations of these resources that are needed to produce any output. Depending on the heterogeneity of these resources, it might be the case that there is a "region" of minimum footprints, which includes all the possible combinations of resources that results in a successful execution of the function.

Multiplexing gains

Number and kind of functions that can run in parallel over the same set of resources with a certain performance level.

Performance degradation function

This KPI characterises the relation between a reduction in the available resources (from 100% until the minimum required) and the reduction in performance of a function. In this case, an elastic NF should achieve graceful performance degradation, avoiding abrupt breakdown under peaks.

Rescuability

When a resource shortage occurs, scaling out or up the VMs that are executing VNFs is the most likely solution to be adopted. Still, re-orchestration processes usually operate at larger time scales (i.e., seconds), which may not be sufficient for certain services. Even with a graceful resource degradation, the overall QoE metrics may not be fulfilled. This will classify VNFs (and the slices using them) according to the capacity of providing graceful performance for a certain interval before new resources come in. This metric should hence measure how “fragile” a VNF is w.r.t. the orchestration: for how long an elastic function can maintain the KPIs before incurring into a SLA violation and the kind and amount of resources needed to be rescued. If a VNF can maintain acceptable levels for a very short time and needs a large amount of resources to restore the previous SLA, then it has low rescuability. Conversely, if a VNF can maintain an acceptable level for a long time and need few resources to re-gain a normal operation, then it has a high rescuability.

Resource consumption

Given a resource (CPU, RAM, others), its consumption is defined as the percentage of time it is occupied because of the execution of a function.

Resource savings

The amount and type of resources consumed by an elastic function to perform a successful operation as compared vs. its inelastic counterpart (e.g., percentage of saved resources while providing 99% of the performance of the inelastic counterpart).

Response time

Time required for resources to be provisioned when demand changes. The shorter the response time is, the better the elasticity.

Resource utilisation efficiency

It is a way to measure how resources are efficiently utilised to provide the desired output. An elastic system should be able to lead to a larger resource utilisation efficiency, since it can deploy a higher number of VNFs over the same physical infrastructure.

Let us consider a set of resources D related to different domains, such as IT and radio access resources, then the system resource utilisation efficiency U_e can be computed as:

$$U_e = \frac{\sum_{d\in D} u_{e,d}}{|D|},$$

where $u_{e,d}$ is the resource utilisation efficiency of the domain $d \in D$. $u_{e,d}$ is calculated as

$$u_{e,d} = \frac{\int_{t=0}^{T} C_d(t)}{T \cdot C_{max,d}},$$

where T is the period in which the resource usage $C_d(t)$ is measured and $C_{max,d}$ describes the maximum resource availability in the domain $d \in D$.

Service Creation Time

This parameter is defined as the time from the arrival of a request to setup a network slice at the network operator’s management system until the slice is fully operational.

Time for reallocation of a device to another slice

This KPI is defined as the duration from the request to connect a terminal device to a certain network slice until this device can start communication. For example, in case of an emergency, a dedicated list of phones or sensors may be pushed into a privileged slice to guarantee crucial communication. The time until these devices can start to communicate in the privileged slice shall be measured by this KPI.

5.4 Application-specific KPIs

This section lists parameters that are relevant to describe the performance of various applications in the two testbeds, i.e., a sea port and a touristic city. Possibly not these parameters can be assessed in the context of the testbeds; details, including precise specifications for the measurement of these parameters, are for further study and will be documented in the subsequent Deliverable D6.2. These KPIs in alphabetic order are as followed:

Frame Rate Judder

The flicker of the entire environment is a common one in VR using headsets such as Oculus Rift. It has been termed judder, and it occurs when the frame rate drops below 75 fps. This judder can induce motion sickness and general discomfort in VR applications. To measure the judder, the following metric is defined:

$$\frac{n/75}{\sum_1^n t_n \in (t > 1/75)},$$

where *t* is the time required for each frame to render and *n* the total number of rendered frames. The formula above represents the percentage of time during a VR application where the framerate was less than 75 frames per second. Minimising this time reduces the probability of motion sickness.

Maximum number of simultaneously active IoT devices

It is expected that in the future cargo containers will be equipped with smart sensors monitoring and reporting environmental conditions (e.g., temperature, humidity, bumps, etc.) online during their journey. Container ships today carry up to 20,000 containers. When such ship coming from overseas enters the coverage area of the very first mobile radio cell, possibly all 20,000 containers will attempt to access the radio cell almost simultaneously. This KPI shall measure the maximum number of sensors within the given deployment area that can be supported by the network slice.

Task Success Rate

Also known as task[6] completion rate, task success rate is the percentage of correctly completed tasks by users. This is probably the most commonly used performance metric that reflects how effectively users are able to complete certain tasks. As long as the task has a clearly defined goal or end point, such as completing a registration form, buying a certain product, etc. the success rate can be measured. So before collecting data, it is important to define what constitutes success. The formula used for Task Success Rate is as follows:

$$\frac{CT}{A},$$

where CT is the number of correctly completed tasks and A is the total number of attempts. It is also important to track the first-time users' success rate and then track the progress: how the rate changes through time, when users gain more experience with the service. This will give you an understanding of system's learnability, which is another indicator of user experience success. The higher the task success rate, the better.

[6] The Task is any activity supported by the end-user application, e.g. the user tries to access information on a cultural exhibit while using natural interaction techniques (gesture recognition, virtual selection of elements in the point of view of the user).

Time on Task

Time on task is sometimes referred to as task completion time or task time. This metric is basically the amount of time it takes the user to complete the task, expressed in minutes and seconds. Time on task data can be analysed and presented in different ways, but the most common way is to present the average time spent on each task.

This can be a useful metric for diagnosing problems. But the time-on-task metric gives more insight in a dynamic view, when comparing the same metric for different iterations.

Generally, the smaller the time-on-task metric, the better the user experience.

Use of Search vs. Navigation:

This is a valuable metric for evaluating the efficiency of information architecture and navigation. Usually when users try to find something through navigation and get lost, search is their final option. Using this KPI, the user perception of network failure can be measured and then correlate it to the underlying problem.

$$\frac{CT_n}{T}, \frac{CT_s}{T},$$

where CT_n is the number of tasks completed through navigation and CT_s is the number of tasks completed search and T the total number of completed tasks. So, this metric can be tracked the following way: for example, a usability task can be set up to find information on a cultural exhibit and track how many users used search and how many used navigations.

5.5 Techno-economic KPIs

Performance against the technical KPIs listed already with respect to the service expectations, listed earlier in Chapter 2, will set the quality of the user’s experience. This user experience will drive a user’s willingness to pay for a service leading to potential incremental revenues for service providers. Additionally, service performance will impact how a service is used which will potentially deliver less tangible social benefits. Conversely to these benefits this improved service performance or even introduction of new services may come at higher network deployment costs. The techno-economic evaluation on this project will use KPIs such as those suggested here to measure overall value delivered by 5G-MoNArch.

Incremental revenue per GB

The revenues expected to be received for services on the 5G-MoNArch platform per unit of data delivered. Some services will be highly commoditised and have a low revenue per GB whereas other newer B2B services may perform better against this metric. Revenues will be considered as an increment to those already expected from today’s 4G networks and their evolution to LTE-A Pro.

Incremental cost per GB

The costs expected to be incurred in delivering services on the 5G-MoNArch platform per unit of data delivered. This will increase with the service requirements. For example, high reliability of a service will come at increased cost. Ideally, the incremental revenues will outweigh the incremental costs. Also, ideally incremental costs will decrease compared with today’s networks for some services due to the 5G-MoNArch innovations.

Social benefits

A financial assessment of the social benefits achieved through new 5G services from the 5G-MoNArch platform. These may include for example savings from more efficient energy usage via smart meters or reduced costs from road traffic accidents due to the introduction of vehicular services. These social benefits need to be considered in association with incremental revenues and network costs to understand the full value proposition for 5G-MoNArch. In the cases of some services the incremental revenues may be low but social benefits high implying some form of government intervention might be required.

6 Verification and Validation Process

Chapter 5 comprises KPI definitions based on existing ones defined by different SDOs and projects as well as extensions to novel KPIs defined by WP3 and WP4 to take care of 5G-MoNArch functional innovations on security, resilience, and resource elasticity. To prepare the definition of one or more assessment metrics an early fixing of the final targets of verification and validation activities within the project and in that context looking ahead to the methodology definition (intended for Deliverable D6.2) is needed.

Verification and validation of 5G-MoNArch architecture will be based on evaluation cases to be defined in Task T6.2 of WP6 and described in D6.2. Each evaluation case compares the underlying baseline system with an evaluation case specific system[7] integrating one or more 5G-MoNArch innovations into the baseline system (Figure 6-1).

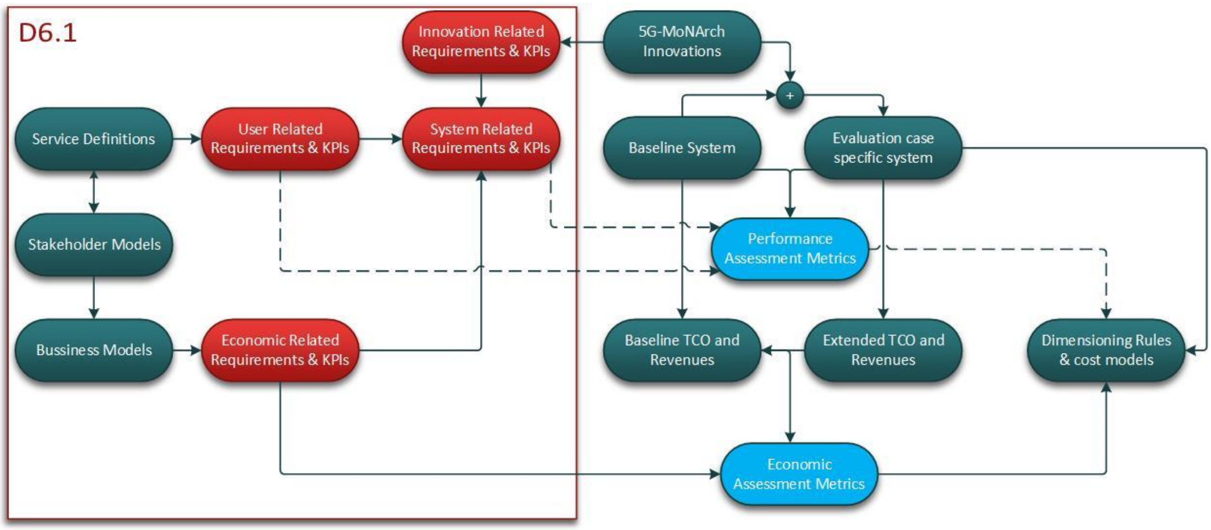

Figure 6-1: 5G-MoNArch assessment model.

The chosen KPIs will be continuously verified during the project work. The type of analysis will vary according to the KPI and it will potentially be of three types:

1. **Analytical evaluation:** The verification process is performed through an analytical evaluation based on theoretical assumptions and values of the considered system.
2. **Simulation:** The verification process is performed through a SW simulation of the considered system that is modelled according to the goals of the verification.
3. **Testbed measurements:** The verification process is performed through experimental measurements during trials in the testbeds. The collected data is processed statistically according to the goals of the verification. Data can be objective (collected from systems) or subjective (collected from users).

Based on service definitions, (end) user related performance metrics can also be derived from the KPIs described in Chapter 5. These user-related metrics are quantified through Quality of Experience (QoE) estimations and measurements [ITU-T G1011]. QoE represents the overall quality of a provided service, as it is perceived by the (end) users, and as such, it is a very appealing alternative for evaluating the quality of a provided service. Similarly, to QoS, QoE may be incorporated in network mechanisms and specifically in network decision processes. Thus, the “QoE-driven” or “QoE-aware” design is an integral feature of 5G systems (and beyond). QoE-awareness may also drive a more resource-efficient or elastic network operation, by helping recognise when the provision of extra resources would not improve the QoE that is eventually perceived by the (end) user. In this context, the existence of proper QoE

[7] Both systems should have clearly arranged complexity (testbed or system level simulation).

estimation methods will allow stakeholders of 5G systems (and beyond) to benefit from the shift from the QoS to the QoE era. The QoE estimation methods enable the translation of application, system, and network related requirements to QoE metrics. A primary classification of the available approaches in the literature is based on whether these metrics are derived directly by client-side questionnaires or automatically through mathematic formulas that apply upon measurements application, system, and network related KPIs. In the first case, specific assessment processes are used, referred to as subjective tests, while in the second case the mathematical formulas and the algorithms that are used, are referred to as objective models.

It has to be noted that there are also some system related requirements that are independent from service definitions, e.g., related to basic functionalities a 5G system has to support. This list of system related requirements and KPIs is complemented by requirements and KPIs which need special emphasis due to technical innovations derived within 5G-MoNArch. There are also some additional system related requirements initiated by economic examinations (e.g., with respect to demand on interfaces (APIs) between stakeholders and their infrastructure, respectively, to realise certain business models). These requirements and KPIs will finally lead together with the systems under inspection to the derivation of performance assessment metrics. During that process testbed measurements and/or respective system level simulation results may be taken to verify the technical KPIs.

Economic requirements and KPI which are used as inputs for economic evaluations are derived based on stakeholder and business models. These will include ensuring that the network can meet service level agreements or provide a consistent QoE that is in line with the expectations of tenants or end users across a range of verticals and applications. The ability of the network to meet or indeed exceed these service expectations will feed into economic assessment metrics such as expected incremental revenues or social benefits beyond today's cellular networks by affecting:

- Willingness of end users or tenants to pay for a given service.
- The opportunity to offer new services not feasible either technically or economically via today's networks.
- The opportunity to offer new services sooner than via today's networks due to improved cost efficiencies or avoiding high initial investment costs by re-using a shared infrastructure set rather than developing new standalone networks.

As well as metrics related to revenue and social benefits as mentioned above, the economic assessment metrics will also examine network cost savings and in particular impact on CAPEX, OPEX and total cost of ownership over an investment period. Economic assessment metrics may be defined by iteratively performed economic evaluations measuring whether the achievement of certain performance figures is in line with guarantying economic viable solutions. In this way, the baseline as well as the evaluation case specific system are investigated both from technical and economic point of view demonstrating the performance improvements of 5G-MoNArch technical innovations and their potential savings with respect to resources and costs. In a final step, dimensioning rules for the 5G network as well as corresponding cost models are developed on the basis of performance and economic assessment metrics.

7 Summary

This document focused on the first design criteria and requirements to achieve a fully-fledged architecture for 5G mobile networks by means of the key innovations (i.e., inter-slice control and cross-domain management, experiment-driven optimisation, and cloud-enabled protocol stack) in addition to the use case specific functional innovations (i.e., resilience, security, and resource elasticity). This deliverable covered the service definitions, stakeholder definitions, requirements and KPIs in addition to verification and validation model. Thereby, this deliverable sets a baseline for the further work in the project.

The definition of the services for 5G systems is covered by a brief overview of the potential service set for 5G mobile networks defined by 5G PPP projects as well as SDOs such as 3GPP and ETSI. Since considering and finally implementing the complete range of potential 5G services is not within the scope of 5G-MoNArch, a subset of this complete set has been selected to be used as input for design choices within the development of the overall network architecture in WP2 and the dedicated functions in WP3 and WP4. The characteristics (i.e., minimum required data rate, end-to-end latency, and resilience/security requirements) of services related to the testbeds of the project (touristic city and sea port) in addition to the services of interest for project innovations have been comprehensively described.

Moreover, the requirements of 5G mobile networks have been listed and categorised into general requirements, requirements related to resilience and security and requirements related to resource elasticity. In addition, techno-economic KPIs and application-specific KPIs have been explained.

The procedures for verification and validation that will be applied within the project have been briefly introduced. Three potential verification approaches that are planned to be used within that process have been briefly introduced, namely, analytical evaluations, simulations and testbed measurements.

The design criteria and KPIs described in this deliverable will be continuously updated and enhanced during the course of the project. The corresponding progress and updates will be described in the following deliverables of WP6, namely, D6.2 ‘Methodology for verification and validation of 5G-MoNArch architectural innovations’ and D6.3 ‘Final report on architectural verification and validation’.

References

[3GPP-22185] 3GPP TS 22.185, "Service requirements for V2X services; Stage 1 (Release 14)", V14.3.0, Mar. 2017.

[3GPP-22186] 3GPP TS 22.186, "Enhancement of 3GPP Support for V2X Scenarios; Stage 1 (Release 15)", V15.1.0, June 2017.

[3GPP-22261] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1 (Release 16)", V16.0.0, June 2017.

[3GPP-22861] 3GPP TR 22.861, "Feasibility Study on New Services and Markets Technology Enablers for Massive Internet of Things; Stage 1 (Release 14)", V14.1.0, Sep. 2016.

[3GPP-22862] 3GPP TR 22.862, "Feasibility Study on New Services and Markets Technology Enablers for Critical Communications; Stage 1 (Release 14)", V14.1.0, Sep, 2016

[3GPP-22863] 3GPP TR 22.863, "Feasibility Study on New Services and Markets Technology Enablers - Enhanced Mobile Broadband; Stage 1 (Release 14)", V14.1.0, Sep. 2016.

[3GPP-22864] 3GPP TR 22.864, "Feasibility Study on New Services and Markets Technology Enablers - Network Operation; Stage 1 (Release 15)", V15.0.0, Sep. 2016.

[3GPP-22891] 3GPP TR 22.891, "Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)", V14.2.0, Sep. 2016.

[3GPP-23501] 3GPP TS 23.501, "System Architecture for the 5G System; Stage 2 (Release 15)", V1.0.0, June 2017.

[3GPP-23799] 3GPP TR 23.899, "Study on Architecture for Next Generation System (Release 14)", V14.0.0, Dec. 2016.

[3GPP-28500] 3GPP TS 28.500, "Management concept, architecture and requirements for mobile networks that include virtualised network functions (Release 14)", V14.1.0, Mar. 2017.

[3GPP-28801] 3GP TR 28.801, "Study on management and orchestration of network slicing for next generation network (Release 15)", V1.2.0, May 2017.

[3GPP-28802] 3GPP TR 28.802, "Study on management aspects of next generation network architecture and features (Release 14)", V0.7.0, May 2017.

[3GPP-32864] 3GPP TR 32.864, "Study on management aspects of virtualised network functions that are part of the New Radio (NR) (Release 15)", V0.4.0, May 2017.

[3GPP-33899] 3GPP TR 33.799, "Study on the security aspects of the next generation system (Release 14)", V1.2.0, June 2017.

[3GPP-33501] 3GPP TS 33.501, "Security Architecture and Procedures for 5G System (Release 15)", V0.2.0, May 2017.

[3GPP-38801] 3GPP TR 38.801, "Study on new radio access technology: Radio access architecture and interfaces (Release 14)", V14.0.0, Mar. 2017.

[3GPP-38913] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14)", V14.2.0, Mar. 2017.

[5GM-D21] ICT-761445 5G-MoNArch, Deliverable D2.1, "Baseline architecture based on 5G-PPP Phase 1 results and gap analysis", October 2017 (planned).

[5G-NORMA] 5G Novel Radio Multiservice adaptive network Architecture (5G NORMA) Project, https://5gnorma.5g-ppp.eu/

[5GN-D21] ICT-671584 5G NORMA, Deliverable D2.1, “Use cases, scenarios and requirements”, October 2015

[5GN-D32] ICT-671584 5G NORMA, Deliverable D3.2, “5G NORMA network architecture – Intermediate report”, January 2017

[5GPPP16] 5G PPP Architecture Working Group: White Paper “View on 5G Architecture”, Version 1.0, July 2016

[5GPPP17] 5G PPP, Report, “5G-PPP use cases and performance evaluation”, Version 2.0, July 2017

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

[ETSI-NFV] European Telecommunications Standards Institute Industry Specification Group Network Function Virtualisation (ETSI ISG NFV), http://www.etsi.org/technologies-clusters/technologies/nfv

[ETSI-WP5G] ETSI ISG NFV, White Paper “Network Functions Virtualisation (NFV) – Network Operator Perspectives on NFV priorities for 5G”, Feb. 2017.

[ETSI-REL001] ETSI GS NFV-REL 001, “Network Functions Virtualisation (NFV); Resiliency Requirements”, V1.1.1, Jan. 2015.

[ETSI-REL003] ETSI GS NFV-REL 003, “Network Functions Virtualisation (NFV); Reliability; Report on Models and Features for End-to-End Reliability”, V1.1.1, Apr. 2016.

[ETSI-MAN001] ETSI GS NFV-MAN 001, “Network Functions Virtualisation (NFV); Management and Orchestration”, V1.1.1, Dec. 2014.

[GDK08] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability Gain of Network Coding in Lossy Wireless Networks," INFOCOM’08 – 27th IEEE Conference on Computer Communications, Phoenix, AZ, USA, Apr. 2008.

[GSMA-2016] GSMA - Mobile IoT Industry Alignment Group. (n.d.). 3GPP Low Power Technologies - GSMA White Paper. http://www.gsma.com/connectedliving/wp-content/uploads/2016/10/3GPP-Low-Power-Wide-Area-Technologies-GSMA-White-Paper.pdf

[HRK13] N.R. Herbst, S. Kounev, R. Ruessner, “Elasticity in cloud computing: what it is, and what it is not.”, in Proc. of ICAC’13 – 10th International Conference on Autonomic Computing, pp. 23–27, San Jose, CA, USA, June 2013.

[HWM+17] B. Han, S. Wong, C. Mannweiler, M. Dohler, and H. D. Schotten, "Security Trust Zone in 5G networks," In Proc. of ICT’17 – 24th International Conference on Telecommunications (ICT), Limassol, Cyprus, May 2017.

[IEEE-2011] IEEE, "IEEE 802.16p Machine-to-Machine (M2M) Evaluation Methodology Document (EMD), " Technical Report IEEE 802.16p-III0014, May 2011.

[ITU-T G1011] ITU-T, Recommendation G.1011, “Reference guide to quality of experience assessment methodologies”, June 2015.

[ITUR-M2083] ITU-R, Recommendation M.2083, “IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond”, Sep. 2015

[ITUR-E800] ITU-R, Recommendation E.800, “Definitions of terms related to quality of service”, Sep. 2008

[LSW+16] J. Li, X. Song, Y. Wang, X. Zhang and W. Tang, "Service Restoration for Distribution Network Considering the Uncertainty of Restoration Time," Proc. 3rd International Conference on Systems and Informatics (ICSAI), pp. 188-192, Shanghai, Chana, Nov. 2016

[Merr17]	Merriam Webster Dictionary, available at: http://merriam-webster.com
[MII-D24]	ICT-671680 METIS-II, Deliverable D2.4, “Final Overall 5G RAN Design”, June 2017
[MVD16]	D. S. Michalopoulos, I. Viering and L. Du, "User-plane multi-connectivity aspects in 5G," ICT’16 – 23rd International Conference on Telecommunications, Thessalonikē, Greece, May 2016.
[NGMN15]	NGMN, White Paper, “5G White Paper”, v1.0, Feb. 2015
[NGMN16-1]	NGMN, Deliverable, “Recommendations for NGMN KPIs and Requirements for 5G”, June 2016
[NMM+14]	N. Nikaein, M. Marina, S. Manickam, A. Dawson, R. Knopp and C. Bonnet, “OpenAirInterface: A flexible platform for 5G research," ACM SIGCOMM Computer Communication Review, pp 33-38, Oct. 2014
[SWG+16]	R Shu, P. Wang, S.A Gorski III, B. Andow, A. Nadkarni, L. Deshotels, J. Gionta, W. Enck, and X. Gu, “A Study of Security Isolation Techniques.” ACM Computing Surveys (CSUR), Vol 49, Issue 3, Dec. 2016
[WLA10]	L. Wu, H. Luo and Z. Ai, “Proactive fault management in large scale computing systems”, Journal of Huazhong University of Science and Technology, vol. 38, no. S1, pp. 20-24, June 2010
[ZOY03]	S. Zhou, Z. Ou, and Y. Yuan, “The technology of on-line diagnosis, fault isolation and dynamic rebuilding for network”, Application Research of Computers, vol. 20, no. 1, pp. 92-93, 2003