HEXA-X-II

A holistic flagship towards the 6G network platform and system, to inspire digital transformation, for the world to act together in meeting needs in society and ecosystems with novel 6G services

Deliverable D2.2
Foundation of overall 6G system design and preliminary evaluation results

Co-funded by the European Union

6GSNS

Hexa-X-II project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101095759.

Date of delivery:	29/12/2023	Version:	0.3
Project reference:	101095759	Call:	HORIZON-JU-SNS-2022
Start date of project:	01/01/2023	Duration:	30 months

Document properties:

Document Number:	D2.2
Document Title:	Foundation of overall 6G system design and preliminary evaluation results
Editor(s):	Pawani Porambage (VTT)
Authors:	Pawani Porambage, Tao Chen, Henry Blue, Rafael Pires, Jere Malinen, Matti Laukkanen, Kimmo Ahola (VTT); Patrik Rugeland, Zhenhua Zou, Josue Castaneda, Flávio Brito, Mårten Ericson (EAB); Pol Alemany, Raul Muñoz, Ricard Vilalta (CTT), Jose Ordonez-Lucena, Riccardo Nicolicchia, Diego R. López, Antonio Pastor (TID), Anastasios Zafeiropoulos, Ioannis Tzanettis (ICC); Anton Krause, Sinuo Ma, Ana B. Martinez, Ahmad Nimr, Philipp Schulz (TUD); Sylvaine Kerboeuf, Abdelkader Outtagarts, Dinh Thai Bui, Huy Tran, Mathieu Boussard (NFR); Akshay Jain, Mohammed Elbamby, Mohamed Abdelaziz, Ozgur Akgul (NFI); Markus Staufer, Bernhard Kurz (NGE); Ignacio Labrador Pavón (ASA); Alperen Gundogan, Sameh Eldessoki, Panagiotis Botsinis (APP); Sławomir Kukliński (OPL); Pietro G. Giardina, Giada Landi (NXW); Pilar Andrés Maldonado, Troels Kolding, Abolfazl Amiri (NDK); Sonika Ujjwal (LMF); Christina Karousatou, Sokratis Barmpounakis, Vasiliki Lamprousi, Panagiotis Demestichas (WIN); Heikki Karvonen (SIS); Stefan Köpsell (BI); Elham Dehghan Biyar, Ferhat Karakoç, Leyli Karaçay (EBY); Milan Groshev (UC3); Bin Han (TUK); Rui Li, Bertrand Decocq (ORA).
Contractual Date of Delivery:	29/12/2023
Dissemination level:	PU
Status:	Final
Version:	1.0
File Name:	Hexa-X-II_D2.2

Revision History

Revision	**Date**	**Issued by**	**Description**
0.1	30.06.2023	Hexa-X-II WP2	Template for Deliverables/IRs
0.1	06.10.2023	Hexa-X-II WP2	Delivered to internal review.
0.2	07.11.2023	Hexa-X-II WP2	Delivered to external review.
0.3	08.12.2023	Hexa-X-II WP2	Delivered for the approval of general assembly.
1.0	29.12.2023	Hexa-X-II WP2	Submitted version.

Abstract

This document is the second deliverable of Hexa-X-II work package 2 – “Foundation for overall 6G system design and preliminary evaluation results”. The document provides the preliminary requirements of 6G end-to-end system, and the enablers related to radio interface and protocols, end-to-end management and automation, and security, privacy and system-level resilience. In addition, the document provides an analysis for a selected set of enablers for the integration in 6G end-to-end system and the management and orchestration view of the system blueprint with an end-to-end intent-based service management automation framework. It also presents the preliminary evaluation results of the proof-of-concepts.

Keywords

6G, blueprint, Digital Service Provider (DSP), End-to-End (E2E) system, intent-based service automation, Key Performance Indicator (KPI), Key Value Indicator (KVI), management and orchestration, privacy, Proof-of-Concept (PoC), radio interface, radio protocols, security, service management automation, system requirements, system-level resilience.

Disclaimer

Funded by the European Union. The views and opinions expressed are however those of the author(s) only and do not necessarily reflect the views of Hexa-X-II Consortium nor those of the European Union or Horizon Europe SNS JU. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This report is the second public deliverable of Work Package 2 (WP2) of Hexa-X-II, titled D2.2 “Foundation of overall 6G system design and preliminary evaluation results”. The first deliverable of WP2, D2.1 [HEX223-D21] provided the guidelines for the 6G end-to-end (E2E) system design including the design principles, the first draft of a blueprint for 6G E2E system proposed in Hexa-X-II project, the system design process, and the system proof-of-concept (PoC) evaluation plans. Relying on the guidelines provided in D2.1 [HEX223-D21], the focus of the current deliverable is to present the technical enabler development, further upgrades to the system blueprint, and an initial set of results from the evaluation framework.

First, the report D2.2 describes a preliminary set of requirements for the 6G E2E system, grouped as use case requirements and operational requirements. The use case requirements refer to the capabilities which the system should have to accomplish the needs of 6G use cases. The operational requirements complement the 6G functionalities not directly visible to end-users but required from the network operator’s perspective to efficiently fulfill the use case requirements.

6G aims to learn from the complexities and limitations of the 5G protocols, seeking improvements to support expanded capabilities, including new use cases, deployment scenarios, external technologies, and capability requirements. The report further describes a subset of enablers on the radio interface and protocol design including user plane, control plane and interaction with higher layers. The user plane is analysed to ensure reliable and spectrum efficient transmission. New mechanisms are proposed related to data recovery mechanisms as well as to ciphering and integrity protection. For the control plane, the current limitations are analysed in the areas of multi-layer down link (DL) radio resource control and of mobility procedure. New enablers in the area of mobility procedure are then proposed (e.g., user equipment (UE) initiating procedure, mobility robustness in 6G multi-connectivity, data-driven mobility). Interaction with higher layers can greatly improve the service differentiation and quality of service (QoS)/ quality of experience (QoE) management for the 6G latency critical use cases. In this direction, an enabler providing mechanism where the UE aids the radio access network (RAN) scheduling based on the applications/traffic characteristics is proposed for a great enhancement in the overall user experience. The report also provides an analysis on several protocol support of 6G enablers such as energy efficient radio design towards more sustainable operation, and new 6G sensing capability and its interplay with positioning.

Concerning the smart network management and automation aspects of the 6G E2E system, intent-based requests can be used to manage services of different administrative domains in an autonomous manner. Intents are specified by the users (e.g., tenants) with their desires without knowing how to accomplish them, whereas the management and control system is responsible for the technical implementation of those requests. The report proposes a preliminary intent-based digital service management architecture and provides nine enablers related to intent-based service management automation. These enablers are designed for intent translation and provisioning, data fusion mechanisms, closed loop coordination and conflict administration in intent management, human-machine intent interfacing, declarative intent reconciliation, intent reporting, and characterization of tenants.

As a pervasive functionality in the proposed system blueprint, security and privacy also play important roles in the 6G E2E system design. In the particular fields of security, privacy and resilience, enablers are intended to address specific threats, providing mechanisms to detect them and to mitigate their impact in system performance. Therefore, the proposed enablers are structured along the threat families they intend to address. The enablers are discussed considering the threats posed by the evolution of network architectural trends and the use of artificial intelligence, as well as the evolution of security-based technologies related to trust enabling techniques, cryptography, and physical layer security. Threats and specific enablers require to be validated in the early stages of design providing evidence on security, privacy and resilience properties. Thereby the security by design principle can be applied, avoiding the common situation of adding security features to existing design, what translates into intricate privacy and security solutions, difficult to be applied by users and service providers. To this purpose two kinds of validation mechanisms are proposed: the use of simulation for end-to-end resilience assessment and the evaluation of anomaly detection at the physical layer, and the

application of a Network Digital Twin (NDT) environment for evaluation of security and privacy threats and enablers.

The relevance and significance of the enablers developed in the technical WPs of Hexa-X-II in relation to E2E system design are of utmost importance. Adhering to the design process methodology identified in D2.1 [HEX223-D21], it is crucial to assess the integration of these enablers on E2E system design and ensure their alignment with the architecture design principles. As a preliminary analysis, this report describes a selected sets of enablers related to intent-management automation, smart network management, virtualization and cloud continuum transformation, and network function modularization. The analysis needs to be a checklist of what can be considered in technical enablers for the alignment with the E2E system performance and operation targets which can be used as feedback by WPs as a reference point for further development of enablers. Furthermore, the updates to the 6G E2E system blueprint are also presented. The view of the E2E system blueprint is discussed comprehensively with respect to the E2E service management and orchestration. As a continuation of this work, Hexa-X-II proposes an E2E intent-based service management automation framework, which results from evolving baseline telco architecture into a multi-stakeholder framework with a wider scope that include both communication and beyond communication services. This also indicates the role transformation of conventional communication service providers to digital service providers that can provide a wider range of digital services.

As part of the comprehensive E2E validation process, this report describes the design, implementation, and the preliminary evaluation results of the first system Proof of Concept (PoC) (named as system-PoC A), which is being developed with mobile collaborative robots (cobots) and extended reality related use cases. System-PoC A is described with respect to two application domains, such as warehouse inventory management and surveillance application. The enablers in system-PoC A are relevant to management and orchestration enablers that include intent-based network solutions, programmable and flexible network configuration, closed loop automation, and integration fabric. The targeted key value indicators and the key performance indicators of system-PoC A are measured to ensure the trustworthiness and the sustainability aspects as well as the performance of the management and orchestration mechanisms of the 6G E2E system. According to the first set of results obtained for power consumption, provisioning time and recovery time, the optimized management and orchestration workflows have better performance compared to the conventional ones. In addition to that, the virtual E2E system evaluation framework presented in D2.1 [HEX223-D21] which is based on digital twin-based approach for 6G connectivity enablers is further elaborated in this report.

Table of Contents

List of Tables

List of Figures

Acronyms and abbreviations

Term	Description
3GPP	3rd Generation Partnership Project
3P	3rd Party
4G	Fourth Generation
5G	Fifth Generation
6G	Sixth Generation
ACK	Acknowledgement
AI	Artificial Intelligence
AI/ML	Artificial Intelligence / Machine Learning
AM	Acknowledged Mode
AMR	Autonomous Mobile Robots
API	Application Programming Interface
ARQ	Automatic Repeat request
AS	Access Stratum
ASN.1	Abstract Syntax Notation One
BAN	Body-Area Network
BB	Baseband
BLER	Block Error Rate
BS	Base Station
BSR	Buffer Status Reports
CA	Carrier Aggregation
CAN	Car-Area Network
CG	Configured Grant
CHO	Conditional Handover
CIA	Confidentiality, Integrity, and Availability
CL	Closed Loop
CLC	Closed Loop Coordination
CLG	Closed Loop Governance
COP	Capability Operator
CP	Control Plane

CRC	Cyclic Redundancy Check
CSC	Communication Service Customer
CSP	Communication Service Provider
CU	Central Unit
D2D	Device to Device
DC	Dual Connectivity
DCI	Downlink Control Information
DDoS	Distributed Denial of Service
DL	Downlink
DLT	Distributed Ledger (Technologies)
DRL	Deep Reinforcement Learning
DSM	Digital Service Manager
DSP	Digital Service Provider
DT	Digital Twin
DTX	Discontinuous Transmission
E2E	End-to-End
ETSI	European Telecommunications Standards Institute
FL	Federated Learning
FTP	File Transfer Protocol
FHSS	Frequency Hopping Spread Spectrum
GBR	Guaranteed Bit Rate
GDPR	General Data Protection Regulation
gNB	gNodeB
GNSS	Global Navigation Satellite System
GUI	Graphic User Interface
HARQ	Hybrid ARQ
HSM	Hardware Security Module
HTTP	Hypertext Transfer Protocol
HW	Hardware
IaaS	Infrastructure as a Service
IBM	Intent-based Management

IBN	Intent-based Networking
ICD	Intent Conflict Detection
IEEE	Institute of Electrical and Electronics Engineers
IME	Intent Management Entity
IP	Internet Protocol
ISG	Industry Specification Group
ITIL	Information Technology Infrastructure Library
JCAS	Joint Communication And Sensing
KPI	Key Performance Indicator
KVI	Key Value Indicator
L1	Layer 1
L2	Layer 2
L3	Layer 3
L4S	Low Latency, Low Loss, Scalable throughput
LEO	Low Earth Orbit
LMF	Location Management Function
LoS	Line of Sight
LoT	Level of Trust
LPP	LTE Positioning Protocol
LTE	Long Term Evolution
LTM	Layer 1/2 Triggered Mobility
M&O	Management and Orchestration
M2M	Machine-to-Machine
MAC	Medium Access Control
MAC CE	Media Access Control - Control Element
MCG	Master Cell Group
MIMO	Multi-Input-Multi-Output
ML	Machine Learning
MnS	Management Service
MP-QUIC	Multipath Quick UDP Internet Connection
MP-TCP	Multipath Transmission Control Protocol

NACK	Negative Acknowledgement
NBI	North-Bound Interface
NDT	Network Digital Twin
NF	Network Function
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NG-RAN	Next Generation RAN
NLP	Natural Language Processing
NoN	Network of Networks
NR	New Radio
NRPPa	New Radio Positioning Protocol A
NS	Network Service
NSMF	Network Slice Management Function
NTN	Non-Terrestrial Network
NW	Network
NOP	Network Operator
OAM	Operation And Maintenance
OFDM	Orthogonal frequency-division multiplexing
OTT	Over-the-top
PaaS	Platform as a Service
PBM	Policy-Based Management
PCell	Primary cell
PDCCH	Physical downlink control channel
PDCP	Packet Data Convergence Protocol
PDU	Packet data unit
PHY	Physical
PKI	Public Key Infrastructure
PLC	Packet level coding
PLS	Physical Layer Security
PoC	Proof of Concept
PoT	Proof of Transit

PRS	Positioning reference signal
PSCell	Primary secondary cell
PT	Physical Twin
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
REST	Representational State Transfer
RL	Resource Layer
RLC	Radio Link Control
RLF	Radio Link Failure
RRC	Radio Resource Control
RSS	Received Signal Strength
RTK	Real-Time Kinematic
RTP	Real-time Transport Protocol
RX	Receiver
SaaS	Software as a Service
SBA	Service-Based Architecture
SBI	South-Bound Interface
SCell	Secondary Cell
SCG	Secondary Cell Group
SDAP	Service Data Adaptation Protocol
SDN	Software Defined Networking
SDR	Software Defined Radio
SDU	Service Data Unit
SEPP	Security Edge Protection Proxy
SLA	Service Level Agreement
SN	Sequence Number
SNS JU	Smart Network and Services Joint Undertaking
SotA	State of the Art
SpCell	Special Cell

SPIFFE	Secure Production Identity Framework for Everyone
SR	Scheduling Request
SRS	Sounding Reference Signal
SLA	Service Level Agreement
SSLA	Security Service Level Agreement
SU	Sensing Unit
TaaS	Trust as a Service
TB	Transport Block
TCP	Transmission Control Protocol
TEE	Trusted Execution Environment
TLA	Trust-Level Agreement
ToF	Time-of-Flight
TPM	Trusted Platform Module
TLS	Transport Layer Security
TSN	Time Sensitive Networking
TX	Transmitter
UAV	Unmanned Aerial Vehicle
UDP	User Datagram Protocol
UE	User Equipment
UM	Unacknowledged Mode
UP	User Plane
UPF	User Plane Function
VDU	Virtual Deployment Unit
V2X	Vehicle to everything
VDU	Virtual Deployment Unit
WP	Work Package
XAI	eXplainable AI
XR	Extended Reality
ZSM	Zero-touch network and Service Management

1 Introduction

Hexa-X-II is the 6G Flagship project under the European Union Horizon Europe research and innovation program Smart Network and Services Joint Undertaking (SNS JU) [HEXA2]. This document is the second public deliverable of Work Package 2 (WP2) – Foundation of overall 6G system design and preliminary evaluation results. In the first deliverable of WP2 (D2.1 Draft foundation for 6G system design [HEX223-D21]), three key terms were introduced as follows:

6G platform is presented as the external view of a set of technologies and interfaces delivering 6G services to applications, ecosystems, verticals, users etc., for enabling value.

6G end-to-end (E2E) system is defined as the technical realization of the 6G platform which includes the technology enablers and their interaction.

6G blueprint is considered as a reference architecture that meets the E2E system needs with respect to hardware, software, and applications.

Moreover, D2.1 [HEX223-D21] provided the first proposal of a 6G system blueprint and ten architecture design principles. The document unveiled the iterative E2E system design process in a two-fold manner. First, the design process considered the components and subsystems provided by the technology usage through key value indicators (KVIs) and the performance requirements through key performance indicators (KPIs). Then, an iterative design process with top-down versus bottom-up alignment was conducted. The document also presented a selected list of 6G innovations that would form the 6G E2E system along with the E2E system evaluation and validation framework.

This deliverable D2.2 will contain the early description of the components developed by WP2, the foundation of the 6G system blueprint, and the preliminary evaluation results at the system level in accordance with the objectives given below.

1.1 Objectives of the document

The objectives of this document can be classified into five main segments:

- **Objective 1**: To identify the 6G E2E system level requirements with respect to the use case requirements and the operational capabilities expected from the system.
- **Objective 2**: To present the early description of some selected technical enablers developed by Hexa-X-II and some of their preliminary component-level evaluation results. These enablers are related to radio interface and protocols, E2E service management and automation, as well as security, privacy, and system level resilience.
- **Objective 3**: To provide recommendations on enabler integrations in the E2E system for the consideration of enabler owners in the future iteration on the enabler design as framed in the top-down vs bottom-up alignment process.
- **Objective 4**: To unveil the novel updates to the 6G E2E system blueprint and to propose an E2E intent-based service management automation framework.
- **Objective 5**: To report the preliminary evaluation results from the first iteration of the system level proof-of-concept (PoC).

1.2 Structure of the document

The document is structured as follows: Chapter 2 provides a brief overview of the requirements identified in the Hexa-X-II 6G E2E system. Chapter 3 introduces the enablers related to the radio interface and protocols and describes the enablers that may have an impact on the radio interface and protocols. Chapter 4 discusses the enablers related to E2E service management and automation. Chapter 5 describes the enablers related to privacy, security, and resilience of the E2E system. Chapter 6 provides the updates to the E2E system blueprint, the details for the integration of a selected set of technical enablers in the 6G E2E system, and the E2E intent-based service management framework. Chapter 7 delivers the preliminary evaluation results for the first system level PoC. Finally, Chapter 8 concludes the report and highlights the next planned steps in WP2.

2 Requirements for the 6G E2E system

This section addresses the transformation of the use case and operational requirements. The use case requirements formalize the needs to be satisfied in a design-independent manner. They represent an application-oriented view of the system. into the Hexa-X-II 6G E2E system requirements. The use case requirements refer to capabilities of the system in terms of what it should do, whereas the operational requirements, which will not be directly visible to the end-users, provide functionality to efficiently fulfill use case requirements for operators.

2.1 Use case requirements

Table 2-1 lists the basic requirements for 6G and the corresponding representative use cases [HEX223-D12] for which the requirements can be relevant. The requirements are described below.

Table 2-1 System requirements and for which use cases the requirements are relevant.

Requirements\Use case	**Ubiquitous Network**	**Real-time digital twin**	**Seamless Immersive Reality**	**Cooperating mobile robots**	**Human centric services**	**Network assisted mobility**
Ubiquitous connectivity	X	X		X	X	X
Indoor coverage	X	X	X	X	X	
Extreme connectivity (high bitrate)			X			
Mobility support	X		X	X	X	X
Pervasive AI/ML		X	X	X	X	X
Efficient sleep states	X		X		X	X
Compute as a Service		X	X	X		X
Intent-based interfaces		X		X		
Reliability		X		X	X	X
Positioning/sensing		X	X	X	X	X
Ultra-low-cost	X					
Energy neutral	X					
Predictable low-latency E2E communication		X	X	X		X
Security/Privacy	X	X	X	X	X	X
Resilience	X	X		X		X
Service continuity	X		X			X

Ubiquitous connectivity:

As more and more mobile services become imperative to everyday life as well as indispensable for efficient industries, the expectation for connectivity wherever you are is continuing to rise. At least basic 6G services (e.g., video streaming, simpler low-tier XR services, etc.) should be available anywhere, to facilitate truly global services, and to provide connectivity to unconnected or under-connected areas and regions.

Indoor coverage:

A large share of the current data traffic is consumed indoor, either at home or in offices, whereas the majority of mobile connectivity is provided via outdoor base stations [Eri21]. Although the propagation characteristics and penetration losses from outdoor-to-indoor, which is further pronounced at higher carrier frequencies, limit the performance, the users will expect full 6G service indoor as well as outdoor.

Extreme connectivity (high bit rate):

Although most of the use cases require modest experienced bit rates, it is expected that a few niche applications, such as fully immersive merged reality, which will push the envelope towards extreme performance beyond what is possible in current generations [HEX223-D12].

Mobility support:

The distinctive feature of mobile networks is of course the mobility, which will naturally continue to be of prime relevance. This accounts for connectivity at high speeds, where Doppler effects begin to affect the radio channels, as well as the expectation for service continuity during mobility.

Pervasive AI/ML:

The development of AI/ML is progressing exceedingly fast, which is likely to impress expectations from the end-users on the availability of AI/ML service anywhere. This entails access to AI/ML services from any device, to optimize performance and enhance applications, as well as availability of the AI/ML services, wherever there is network coverage.

Efficient sleep states:

With sustainability and energy-efficiency as key driving forces for the development of 6G, it will be imperative to efficiently conserve energy when it is not needed and that devices and equipment can enter partial or fully power-saving sleep-modes to extend the operation, as well as minimizing the cost and environmental impact of the service.

Compute as a service:

As one of the key beyond-communication services, compute as a service from the end-user devices to the network is expected to enable more advanced services for longer durations on power- and computationally constrained devices, where the bulk of the processing is handled in the network.

Intent-based interfaces:

With networks becoming more and more heterogenous and complex, end-users will expect the network to discern the users' preferred connectivity option (e.g., necessary bandwidth, or utilization of JCAS or offloading) without any overt commands from the user.

Reliability:

With increasing reliance on mobile services for personal, commercial, and industrial applications, the expectation for service reliability will be much more pronounced and users will assume that services will remain available even during mobility or at unforeseen events.

Positioning/Sensing:

By leveraging on the radar properties of the radio propagation, 6G is poised to broaden the utility of mobile networks, to provide cost-efficient detailed positioning and sensing capabilities to end-users.

Ultra-low-cost:

While certain 6G use cases predict unprecedented performance, others lie at the other extreme end of the range, enabling near zero-cost devices to communicate via the network. These devices would be possible to distribute and embed nearly anywhere and provide simple connectivity to nearly anything.

Energy neutral:

A large detriment to current machine-type communication devices is their limited lifetime. Even if the battery-lifetime is ten years, the lifetime of the equipment and machineries they support are often several decades (e.g., in a factory) which will necessitate semi-regular battery replacements. By enabling energy neutral operation, e.g., through energy harvesting, the device lifetime could be extended beyond the lifetime of the systems they are embedded in. This could enable deployments in harsher environments, where access for battery replacements can be omitted.

Predictable low-latency E2E communication:

Although 5G systems are capable of delivering millisecond air interface latencies in specific deployments, in most cases it is not the best-case or even average latency which is relevant to provide the quality-of-experience (QoE) or to fulfill the delay and timing requirements for machine type communication, but instead the bounded worst-case or a predictable low-jitter latency would be more relevant to avoid processing chokeholds or sudden spikes in delay.

Security and privacy:

Similar to previous generations, security and privacy will continue to be paramount to avoid any malicious disruption (tampering, interception, impersonation, abuse, etc.) of networks, devices, and services. The network should support the definition and verification of security, privacy and resilience requirements, incorporating them into service level requests and agreements.

Resilience:

In order to ensure fulfillment of the service level agreement, it is important to ensure continuous operation and service continuity, even in case of unexpected disruptions. In particular, the resilience is the networks' ability to mitigate or quickly recover from disruptions, caused either by attacks impacting security or privacy, as stated above, or any other causes, e.g., in terms of natural disasters, power outages, or mechanical failures.

Service continuity:

Certain services and use cases can operate with relatively long service interruptions, e.g., file downloading or cashed video streaming, where several seconds of interruption can go unnoticed. However, for other use cases, even a few milliseconds interruption can have dire consequences, e.g., inducing nausea in an XR application or losing control of a remote-controlled vehicle. The service continuity is particularly important during mobility, when a device moves out of coverage of its current connection, the network need to seamlessly handover the connection to another node with sufficiently low service interruption.

2.2 Operational requirements

In addition to the use case requirements which are more focused on the end-user perspective, the 6G system is expected to provide functionality, which will not be directly visible to the end-users but required from the network operator's perspective to efficiently fulfill the use case requirements.

Flexible radio protocols:

For the 6G system to achieve its potential, one of the methods is to introduce a flexible protocol stack in the radio access network (RAN). This should consider approaches such as radio resource control (RRC) simplification, improving/optimizing the data recovery and reordering mechanisms, making security features more optimized and in general ensuring as much flexibility as possible.

Mobility procedures:

For the mobility aspects, the 6G system should support a solution wherein the mobility procedures may be unified. Additionally, if carrier aggregation (CA) is implemented for 6G RANs, then robustness in mobility would also be needed. Furthermore, it should be expected that the Hexa-X-II 6G system mobility procedures are agnostic to network architecture/deployment options. Moreover, prevalent methods should be further optimized for 6G including support for frequency range 2 (FR2, 24.25-71 GHz), centimetre-wave (7-15 GHz), sub-THz (100-300 GHz), as well as the inclusion of AI/ML.

Improved access convergence:

6G systems will have both terrestrial networks (TN) and non-terrestrial networks (NTN) as possible deployment solutions. Hence, the 6G system should consider convergence between TN-NTN systems with respect to how the spectrum resources should be shared. Furthermore, it should also consider the interworking between 3GPP and non-3GPP RATs.

Native AI/ML capabilities:
The 6G network should have AI/ML in-built in the functionalities of the network. The network should also support learned and/or customized waveforms in the form of predefined look-up tables or via other means of agreement between the transmitter and the receiver. Additionally, AI/ML will also be applied as an essential tool in management and orchestration, security, and privacy frameworks and in data analytics to improve the performance of the network as well as to serve the 3rd party applications/network applications.

Multi-connectivity:
The 6G system should support aggregation of contiguous and non-contiguous bands in all available frequency bands, from single and multiple transmission points for communication and beyond-communication services (e.g., sensing) to increase the instantaneous bandwidth and reliability. The solution should avoid complex UE configuration schemes.

Intent-based management:
Automation in 6G network operations should leverage a controller/manager architecture based on intent-based management and closed-loop automation. Intent-based management (IBM) can be applied at the different management layers. Within each management layer. intents are used to define the goal state which the Closed Loop (CL)-based controller then tracks. AI methods are employed within each layer to perform interpreting of the intent, modelling the network's state, and making decision to ensure the desired goal is achieved. Status information is provided as feedback to intents that are issued.
The 6G system should make service requests easy for non-experts (e.g., end-users, 3rd party, etc.) using an intuitive interface (e.g., natural language). It should allow the specification of the desired behaviour of the service in terms of targets specific to sustainability, security, and trustworthiness. It should provide means to automatically enforce the targets at the different layers and ensure that the system is always achieving the expected requirements. Moreover, the 6G system should be capable of detecting and resolving intent conflicts in real time that arise in a multi-tenant environment, applying the fairest conflict resolution actions with the main objective of respecting and minimizing the impact on the other tenants' services. Finally, it should support service intent in a multi-provider digital service environment. It should be capable of decomposing the intent into the capabilities of multiple digital service providers.

Seamless orchestration across the compute continuum:
The 6G system should support the deployment of 6G services over heterogeneous resources across the device-to-edge-to-cloud continuum. It should also support intent-driven orchestration for deployment of cloud-native applications with strict quality-of-service (QoS) requirements (e.g., latency-sensitive application components) across resources in the IoT-to-edge-to-cloud continuum. Furthermore, the 6G system should support interaction between network providers and over-the-top (OTT) players (edge/cloud providers) for optimal deployment of applications.

6G service delivery across multiple digital service providers:
Managing 6G E2E services requires the involvement of several service providers (e.g., network service providers, cloud service providers) to compose an E2E service. For example, an application may require 6G network resources such as RAN, transport, and CN services from one network operator, transport network resources from another network operator, and cloud computing resources from hyperscalers. Hence, the 6G system should support the composition and orchestration of multiple services coming from different digital service providers to provision and assure the 6G E2E service. It should also provide aggregator means for interfacing with different service providers and for exposing their services to third parties such as verticals.

New 6G capabilities exposure:
The 6G system should provision a generic and dynamic exposure functionality of beyond communication services comprising sensing, enhanced localization and tracking, and compute-as-a-service. Furthermore, it should be able to expose new and existing capabilities via APIs to developers of applications. Such APIs

should be simplified (abstracted) so they can be used by application developers without the need to be network experts. Those APIs can provide services that third-party application developers can consume with their familiar tools and processes while still providing them access to a wide set of network capabilities (e.g., application flows steering). Additionally, it should also enable the controllability by the network operator of the capabilities that are exposed to third parties (tenants).

3 Enablers related to radio interface and protocols

The radio interface and protocols concern the exchange of information over the air between, e.g., a device and a base station, or between two devices. The 5G radio interface and protocols are mostly an evolution of the corresponding radio interface and protocols in 4G. This resulted in a complex, and to some extent not useful interface and protocol design with too much configurability, out of which most deployment options weren't ever implemented. For 6G, those learned lessons inspire investigations in key directions on fixing limitations in 5G, e.g., one lesson learned is the lack of easy-to-use downlink resource control, which is further discussed in section 3.3.1. At the same time, the 6G radio interface and protocols need to support an expanded scope and capabilities compared to 5G, e.g., new use cases, new deployment scenarios, new external technologies, and new capability requirements. Examples include sensing (discussed in section 3.5.2) and computing offloading, sub-network, and distributed Multi-Input-Multi-Output (MIMO) (discussed in section 3.5.3).

In this chapter, the ambition for 6G radio interface and protocols if introduced, followed by the description of a subset of enablers on the radio interface and protocol design (including user plane, control plane and interaction with higher layers) and how to support other 6G enablers.

3.1 Ambition for 6G radio interface and protocols

In this section, we discuss the ambition for the 6G radio interface and protocols. This is done by mapping system design principles [HEX223-D21] to the radio interface and protocol design and specific considerations. Considering the lessons learned from 5G and the need to support new expanded scope and capabilities, the following aspect need to be taken into account when designing the radio interface and protocols:

- Consider essential functionality, features and options that go hand in hand with the fundamental 6G requirements, that are common and essential to all use cases (e.g., both high-end and low-end UEs covering a wide range of use-cases) from "day one",
- One flexible protocol stack for different scenarios that maximizes the use of single protocol components for different scenarios, which is easily extensible for further enhancements in the 6G timeframe, instead of over-optimized for the baseline; building on top of modular components that ease protocol scalability to cover more advanced radio requirements needed by the use cases,
- Simplicity in comparison to previous generations to implement and deploy,
- Optimize for actual scenarios and needs in the fields, not only for extreme performance or corner cases, e.g., to avoid unnecessary configurability,
- Ensure a fast and reliable protocol operation (e.g., supporting fail-safe mechanisms when needed),
- Keep separation of concerns in multi-layer protocol stack, without compromising performance by artificial boundaries,
- Better consideration of spectrum sharing of other technologies, privacy/security, resilience/availability, energy efficient operation of network and device, friendliness to cloud implementation, network scalability (e.g., adding/removing network nodes as needed).

3.2 Radio user plane

The radio user plane concerns the transmission of users traffic from a network node to a device, and vice versa. In what follows, the aspects are discussed related with data recovery due to transmission errors on the radio interface and ciphering and integrity protection of the transmitted data.

3.2.1 Data recovery mechanisms

3.2.1.1 Introduction

Data recovery and reordering capabilities are a key component in cellular radio design, to meet application requirements in terms of packet loss and to optimize the connection, in particular the air interface spectral efficiency. However, data recovery mechanisms are also evolving outside of radio networks, and the 6G data recovery and reordering mechanisms shall supplement transport and application layer mechanisms of the 6G-era also built for similar goals, to optimize the E2E performance. An overview of the relation of 6G and example external mechanisms is shown in Figure 3-1Figure 3-1. Modern Transmission Control Protocol (TCP) congestion control flavours are more resilient to packet losses and reordering errors compared to previous versions. Multipath techniques can be applied in both Internet and Ethernet, e.g., Multipath Quick UDP Internet Connection (MP-QUIC), Multipath Transmission Control Protocol (MP-TCP), and IEEE 802.1CB, to provide redundancy outside of the radio network. Finally, scalable congestion control is extended with mechanisms such as Low Latency, Low Loss, Scalable throughput (L4S) which allow applications to tune rates to ensure higher packet reliability within a delay window during congestion [BSB+23].

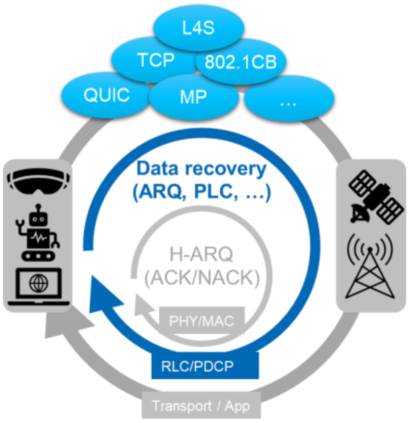


Figure 3-1: E2E mechanisms for data recovery and reordering (PLC is short for packet-level coding).

A secondary consideration is the specific use-cases for the 6G data recovery and reordering mechanisms. Across previous radio generations, Hybrid Automatic Repeat Request (HARQ) has seen a significant development culminating in 5G. However, operating a system at very high reliability levels has typically a penalty in terms of spectral efficiency and coverage limitations. For cases where performance of HARQ is insufficient versus the application requirements including chosen transport protocols, 6G needs to have a higher layer recovery mechanism available. More details on the performance of the recovery loops are detailed in section 3.2.1.3

In the following, first, several possible innovation directions for data recovery enhancements are listed. Next, the ARQ performance is evaluated with possible imperfections in the control channel. Finally, an example of a cross-layer approach for recovery loop enhancements is studied.

3.2.1.2 Considered innovations in higher layer data recovery

For cases where the higher layer data recovery mechanism is needed (i.e., Radio Link Control (RLC) layer and above), the disadvantages for RLC ARQ (e.g., high latency) should be considered when designing the procedures in 6G. Target 6G system latency are expected to be in the 10 milliseconds range, even for mobile broadband, to support an increasing number of interactive services. In this context, some design considerations for 6G higher layer (e.g., L2) data recovery to pursue in Hexa-X-II include:

- Reduce overall latency, reconsider currently established ARQ loop designs or consider new approaches, including packet level coding to bridge mechanisms such as ARQ and packet duplication known from, e.g., 5G.
- Reconsider packet reordering requirements (in-sequence delivery of packet to upper layers) and modes considering needs from application and transport level as well as overall design for low latency.
- Investigate novel flexible protocols and cross-layer interactions for an enhanced utilization of network resources to allow for overall improvements in the network performance.
- Further enhancements towards new use-cases and technologies, including non-terrestrial deployments and metaverse/XR applications.

These listed design considerations align with the principles in clause 3.1 such as simplification or optimization of the data recovery configuration to the actual scenario to serve. The study in this area will align with the enablers developed in [HEX223-D42] for the sustainable, trustworthy and inclusive holistic radio design.

3.2.1.3 Relation between ARQ data recovery performance and goodput

A potential 6G HARQ scheme should achieve performance better than earlier generations, the assumed performance is shown in Table for most use-cases. Note that for some systems, like non-terrestrial networks (NTN), the air interface latency may exceed application requirements (or memory capabilities of devices) and thus, the use of HARQ becomes even more restricted.

Table 3-1: Typical imperfections for downlink dynamic scheduling operation in 5G [SMP+14] [1].

HARQ reliability component	**Typical performance**	**Best achievable performance**
Downlink control channel error (dynamic), e.g., loss of dynamic downlink scheduling info	<1%, (<5% in problematic areas)	<0.1%
Discontinuous transmission interpreted as ACK by base station*	<1%	<0.1%
NACK interpreted as ACK by base station	<0.01%	<0.001%
NACK→ACK error by UE (assuming similar dynamic scheduling solution as in 5G, limited by CRC error)	Virtually zero	Virtually zero

*) Depends on number of bits for uplink control information (UCI) and if those scale up significantly in 6G

In recent generations, there has been a secondary retransmission mechanism on top of HARQ, e.g., at the RLC layer in 5G. Under bad radio conditions with high block error rate (BLER), the RLC AM uses retransmissions to recover lost packets. These ARQ retransmissions use other time slot opportunities, which adds latency to the initial transmission of new packets. In 5G, RLC acknowledged mode (AM) and unacknowledged mode (UM) have been defined for multiple different bearers to distinguish QoS flows in order to separate and prioritize different streams according to their needs. Using RLC UM provides tighter latency/jitter characteristics but is also less reliable compared to RLC AM. RLC AM provides 100% lossless links where a re-establishment procedure occurs in case of failures. However, this comes at the expense of added latency/jitter.

In addition, ARQ retransmissions may cause window stalling on the transmitter (TX) side and therefore leads to Head-Of-Line blocking and increased TX latency, as RLC AM must wait for all packets to be received in-sequence before moving the transmitter/receiver window. This increases the receiver (RX) latency, since it requires a reordering functionality (e.g., on the RLC layer in 4G and on the PDCP layer in 5G) for delivering all packets in-order towards the higher layers.

The following analysis aims at understanding the impact of adding a second data recovery layer on TCP user goodput, particularly in relation to the residual error rate after the HARQ process, as detailed in Table 3-1. In Figure , this relation is shown for both TCP Reno and Cubic in combination with using RLC UM and AM modes[2]. It is seen that it is important that the residual error is controlled. More specifically, if HARQ residual error rate is higher than approximately 1x10^{-5}, then having a second data recovery layer starts to have significant impact on experienced user goodput. The results of Figure 3-2 can be used together with the numbers in Table to guide the design of reliability for 6G user plane traffic. For services with more strict

[1] While these requirements are not yet defined for 6G, it is clear that future requirements point in the direction of increased reliability and lower latency support, see e.g., [ITUR23].

[2] Full scale system simulation using 5G physical layer and data recovery mechanisms as reference model. FTP data model 3 is used with average packet size and arrival rate adjusted to nearly 70 Mbps offered load. MSS is set at 1460 bytes. The scenario is 3GPP dense urban with a deployed bandwidth of 100 MHz and follows a typical 3GPP system evaluation model.

requirements, e.g., ultra-reliable-low-latency-communication (URLLC)-like services, the HARQ design should target much lower residual error rates, e.g., in the order of $1x10^{-7}$or less, to fulfil the requirements.

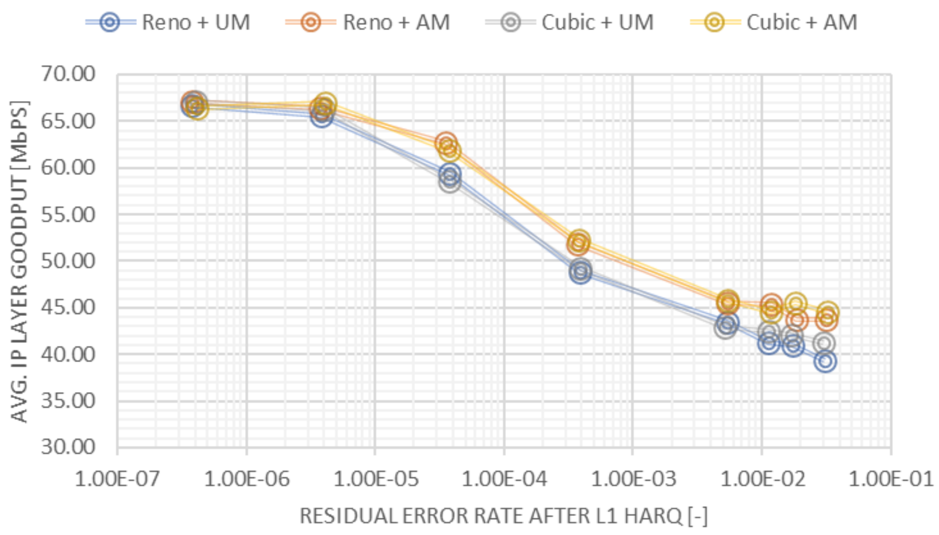


Figure 3-2: Example of relation between residual error of the L1 HARQ (at transport block level) and TCP goodput.

3.2.1.4 Cross-layer interaction enhancements example for data recovery

When all traffic is mapped to the same bearer (e.g., the non-GBR bearer), or when within a data stream there are different packet types with different importance (see Figure 3-3Figure 3-3), there are no means to overcome the aforementioned latency and jitter degradation of RLC AM while at the same time having higher reliability than RLC UM. This is because the separation of streams to AM and UM is based on e.g., source/destination IP addresses/ports as well as on header information and assumes those streams are homogenous within themselves, where each packet is of equal importance.

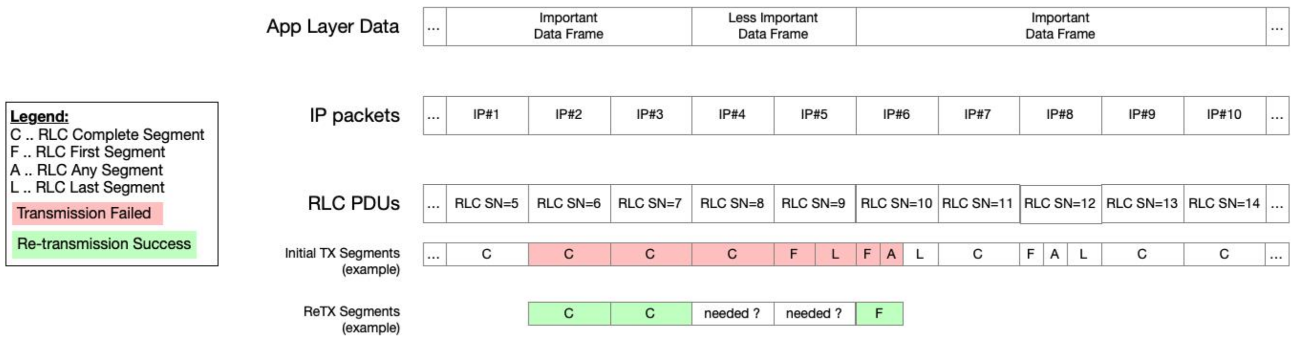

Legend:
C .. RLC Complete Segment
F .. RLC First Segment
A .. RLC Any Segment
L .. RLC Last Segment
Transmission Failed
Re-transmission Success
App Layer Data
...
Important Data Frame
Less Important Data Frame
Important Data Frame
...
IP packets
... IP#1 IP#2 IP#3 IP#4 IP#5 IP#6 IP#7 IP#8 IP#9 IP#10 ...
RLC PDUs
... RLC SN=5 RLC SN=6 RLC SN=7 RLC SN=8 RLC SN=9 RLC SN=10 RLC SN=11 RLC SN=12 RLC SN=13 RLC SN=14 ...
Initial TX Segments (example)
... C C C C F L F A L C F A L C C ...
ReTX Segments (example)
C C needed ? needed ? F

Figure 3-3: Example of data packets with different importance being transmitted and not being correctly received.

A considered direction for 6G related to optimization via cross-layer interaction, is to avoid RLC window blocking or unnecessary delays by allowing window moving operations based on certain additional delays. For example, the RLC window may move when a certain amount of retransmission attempts was not successful or when less important data is affected by BLER and RLC retransmissions. For the latter, the RLC layer may use higher layer (i.e., either in the radio protocol stack, network stack or Application layer) information to decide on data importance. Upon reception of NACKed Sequence Numbers (SN) from the RX RLC entity, the TX RLC entity may, depending on the above mentioned conditions (i.e. data importance), send information back to the RX RLC entity that the NACKed SNs shall be skipped. This way, the RLC RX/TX window can be moved forward and be unblocked, removing unnecessary RLC retransmissions that consume resources and result in additional latency and also removing data stalls due to RRC reestablishment procedure. Furthermore, RX reordering can be unblocked with this method by the RLC informing the reordering function about the gap(s) so that the reordering window may be updated. This would remove the unnecessary additional reordering latency as the waiting for packets can be skipped. For this method to succeed, the PDCP discard functionality could be extended to also become able to discard packets that have already been assigned an RLC SN on the TX side. Note that in 5G, only the packets that were not yet assigned an RLC SN on the TX side can be discarded. With this method, any RLC SN can be discarded even if it already resides in the RX reordering window. Figure 3-4 illustrates the operation of the described method.

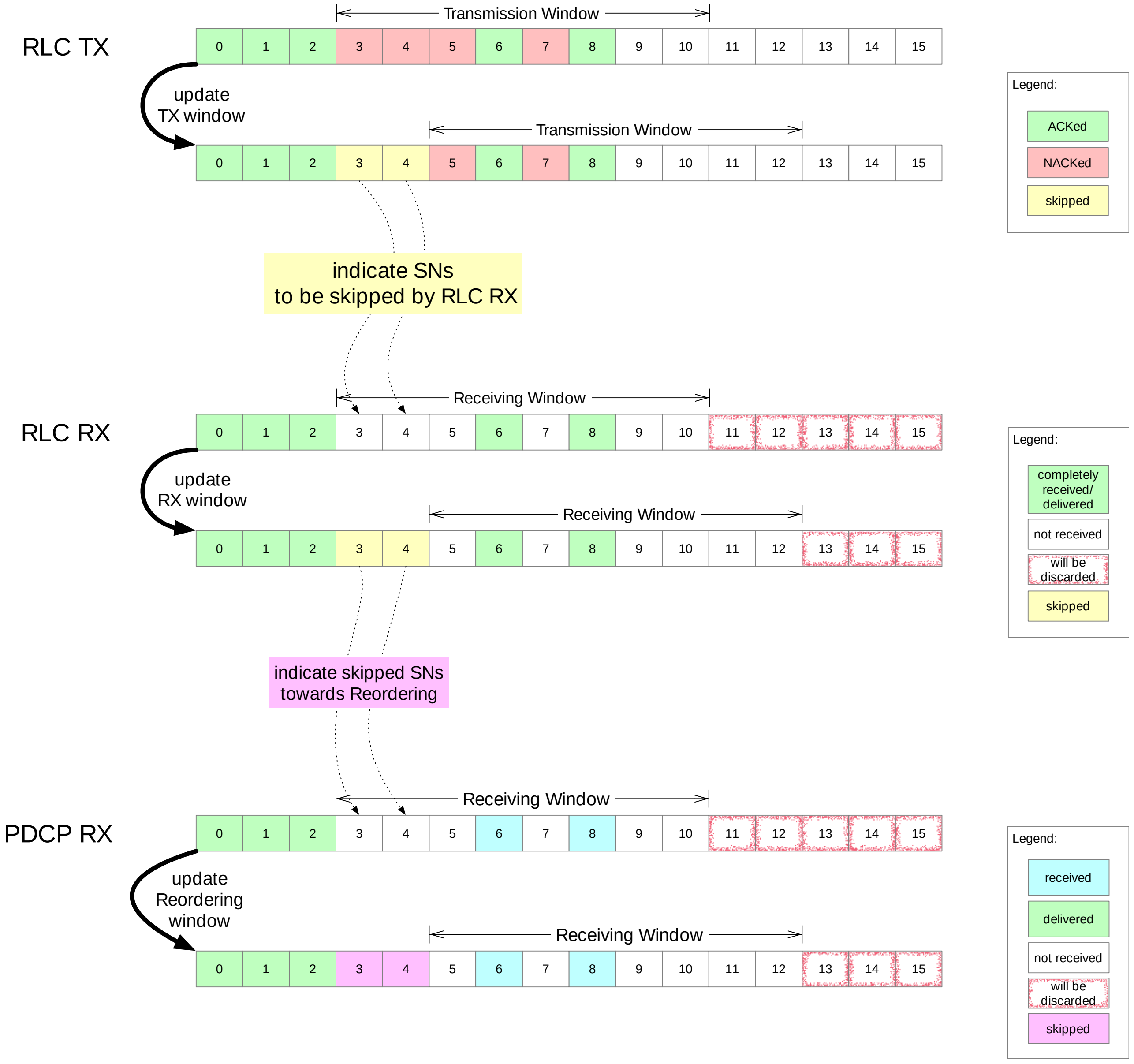


Figure 3-4: Indication of RLC PDUs to be skipped and moving of the receive window.

3.2.2 Ciphering and integrity protection

The ever-increasing peak data rates and real-time demands, lead to a greater demand on the required hardware (HW) to perform ciphering and integrity protection on a per packet basis on the user equipment (UE) modem side. For example, in New Radio (NR), ciphering and integrity protection are done in the Packet Data Convergence Protocol (PDCP) layer, where every byte above Service Data Adaption Protocol (SDAP) (aka. SDAP Service Data Unit (SDU)) is optionally ciphered and integrity protected in the RAN (i.e., UE-gNB interface) [38.323]. Note that ciphering and integrity protection are both optional and are configured independently by the network via RRC signalling [38.331]. This results in ciphered and optionally integrity protected Internet Protocol (IP) headers and payload, while the L2 headers (i.e., SDAP, PDCP, Radio Link

Control and Medium Access Control (MAC)) are uncyphered and partially not integrity protected (i.e., only PDCP and SDAP headers are integrity protected). This means that the IP payload, which is in most cases E2E encrypted on the application layer, is ciphered, while some L2 headers are kept unciphered and partially not even integrity protected. As a result, eavesdroppers can see this kind of metadata and may discover traffic patterns and used applications and services, insert forged MAC CEs, RLC Headers to break the connection (e.g., Secondary Cell (Scell) deactivation, Sequence Number (SN) window stalling, etc.) and do active payload alteration (e.g., aLTEr attack [aLTEr]). Additionally, assuming a 1500 Byte SDAP SDU, around 99% of a MAC Packet Data Unit (PDU) is ciphered. With increasing throughput this would mean that more data needs to be ciphered, requiring baseband (BB) modem HW ciphering block design to scale accordingly. This scaling of the HW block would need to accommodate for peak throughputs, even though such peak throughput is a very rare case and as mentioned previously, application data in most cases is E2E encrypted on the application layer. Figure 3-5 gives an overview of the NR transport block (TB) highlighting which parts are ciphered and integrity protected (i.e., Access Stratum (AS) Security) as well as the unciphered parts.

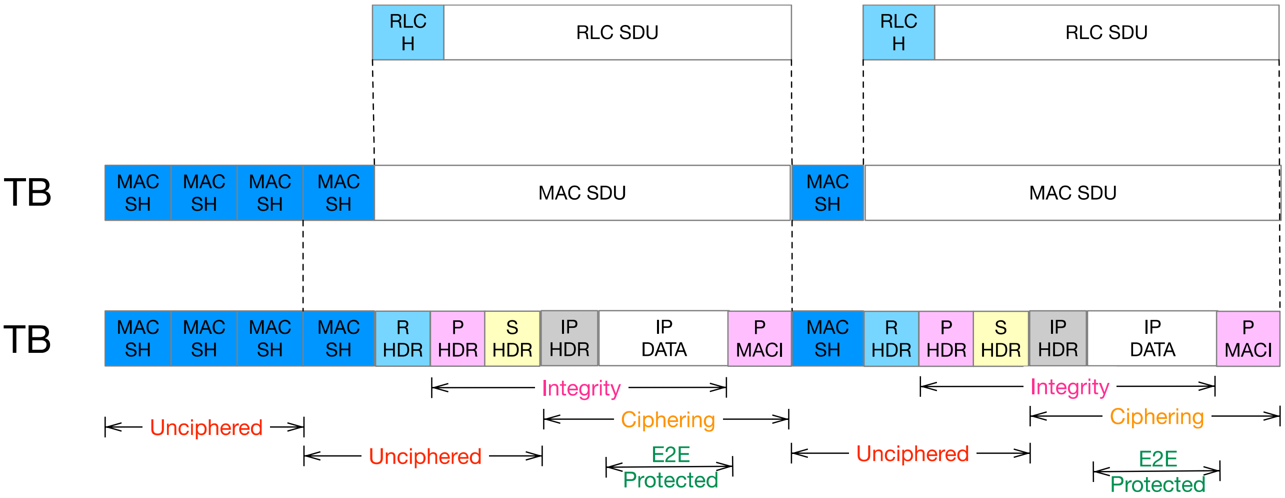


Figure 3-5: NR access stratum security.

To overcome the aforementioned issues, a new ciphering and integrity protection mechanism is proposed, where such functionality would be moved to the MAC layer (i.e., from PDCP in case of NR). By this, all L2 headers (i.e., SDAP, PDCP, RLC and MAC) are integrity protected and additionally the MAC SU part containing SDAP, PDCP and RLC headers would be ciphered. Moreover, ciphering and integrity protection offsets are introduced, which defines the portion of the SDAP SDU that is ciphered and integrity protected of the higher layers (i.e., SDAP SDU) is protected. Figure 3-6 highlights the newly proposed TB headers, highlighting the parts of the TB that would be ciphered and integrity protected. In the figure, the I-Offset and C-Offset are arbitrary examples (e.g., C-Offset needs not necessarily enclose I-Offset) showing how they could span into the SDAP SDU to cover for example, IP Headers (e.g., outer IP layer only), IP and TCP headers, IP, UDP and RTP headers, Ethernet headers or even the protection of the complete payload (i.e., same as NR legacy) in most extreme cases. Note that with the moving of the ciphering and integrity protection functionality to the MAC layer, MAC headers would now require the introduction of an SN.

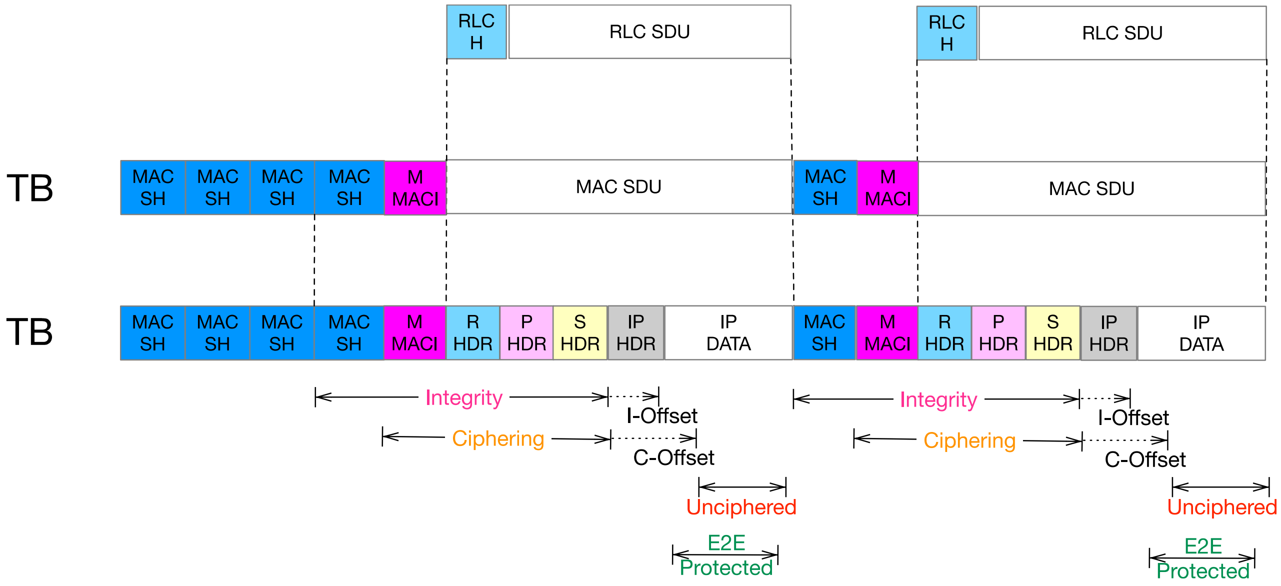


Figure 3-6: 6G Access Stratum Security Proposal.

In conclusion, in the newly proposed security concept for 6G, layer 2 headers (i.e., SDAP, PDCP and RLC) are ciphered and integrity protected compared to legacy. This means that there is no possibility for L2 header manipulation, where MAC sub-headers (i.e., including MAC CEs**)** may now be integrity protected compared to legacy. Additionally, IP Header and potentially even IP Data may be ciphered and integrity protected. Finally, the amount of bytes to be encrypted and/or integrity protected is reduced extremely (i.e., less than 1% for 8000 Byte and 4% for 1500 Byte IP Packets), resulting in a significant reduction of the required HW capabilities for ciphering (i.e., area and power) as ciphering throughput is drastically reduced.

3.3 Radio control plane

The radio resource control from the network in the downlink (DL) as well as the requests and responses from the device to the network in uplink (UL) are crucial components in a cellular communication system in which the radio resources are controlled and coordinated by base station. Additionally, support of mobility and handover of devices has been an essential component in the control plane protocol. In section 3.3.1,the current limitations in the area of multi-layer DL radio resource control are discussed. In section 3.3.2, the current limitation and introduction of new enablers in the area of mobility procedure are discussed.

3.3.1 Multi-layer DL radio resource control

For a base station to control how a device accesses radio resources for transmission and reception, the base station transmits commands to the device. In addition, the device can transmit requests, status information, or confirmations to the network, however, this section only discusses the DL aspects. There are three such mechanisms with various purposes, whose differences depend on which radio protocol layer the command is transmitted. However, these mechanisms are inherited from previous generations (e.g., 4G), which turn out to be already outdated in 5G due to an increasing number of use cases, scenarios and requirements. In particular, these control signals need to be backward compatible, i.e., an older device need to be able to discern which part of the signalling concerns it, while still enable extensions to be signalled to newer devices. Thus, the continuous extensions and interdependencies between different signalling exacerbate the complexity.

Table 3-2: Summary of the difference between PDCCH (Physical Downlink Control Channel), downlink MAC CE (Medium Access Control – Control Element) and RRC (Radio Resource Control), all for Enhanced mobile broadband traffic (expanded from TS 36.300 [36.300]).

Signalling	PDCCH	DL MAC CE	RRC message
Mechanism	Transmitted on Layer 1 as DCI; no direct feedback and retransmission	Transmitted on layer 2 of MAC as transport block; there is HARQ feedback and potential HARQ retransmission; residual errors occur due to false detection of HARQ NACK/DTX to HARQ ACK	Transmitted on layer 2 of PDCP as PDCP PDU; always uses RLC AM mode, and there is RLC AM status PDU feedback and potential RLC retransmission; residual errors occur due to CRC check error of the Transport Block that carries RLC AM status PDU
Reliability	1 - 10^{-2}	1 - 10^{-3}	1 - 10^{-6}
Delay	Very short (a couple of OFDM symbols)	Short (a couple of slots)	Longer (dozens of milliseconds)
Signalling extensibility	None or very limited	Limited	High
Security	No integrity protection; No ciphering	No integrity protection; No ciphering	Integrity protection and Ciphering

PDCCH

The principle of the PDCCH processing is that the device tries to blindly decode candidate PDCCHs transmitted from the network. The devices are provided with information on where to look for PDCCH and a condition to check whether the decoded information is correct. To limit device complexities, the device can only perform a limited number of such blind decoding. By design, PDCCH misdetection probability is 1% and false alarm probability is around 10^{-6}. Multiple UEs can monitor the same resource and a UE needs to blind decode multiple candidates. This implies that:

- PDCCH has currently in 5G a very limited extensibility due to the blind decoding principle. For example, in NR, the device can only monitor up-to four DCI sizes, and it is not obvious to increase the DCI size to a larger size because it would impact the PDCCH performance.
- DCI was designed for one-shot commands, e.g., UL/DL data transmission scheduling. (Semi)-persistent commands change the "operation mode" of the UE. If, after such a command is sent, the network and the UE misunderstand each other on the state, then it may need to recover from such errors. This state mismatch happens because there is no explicit feedback for PDCCH transmission, HARQ feedback error (e.g., DL scheduling type of PDCCH), or false alarm at an unintended UE.

RRC configuration:

RRC was designed as a generic control plane protocol by which the UE conveys e.g., its capabilities and measurements and by which the network configures the UE [38.331]. RRC configuration is transmitted using RLC AM mode and thus reliable and providing in-sequence delivery (i.e., can avoid race conditions). The message is also integrity protected and ciphered. RRC uses ASN.1 encoding, and extensibility is natively supported.

On the other side of the coin,

- RRC message may take longer time to deliver due to waiting for RLC status PDU, acknowledgement message, segmentation on RLC layer, or a large message size due to hierarchical message structure.

- RRC message processing delay is comparatively longer than MAC CE and DCI. The two latter ones are assumed to have zero processing delays. The RRC processing delay defined in section 12 of TS 38.331 [38.331] is 10-16 milliseconds for potentially large reconfigurations.

DL MAC CE

MAC CE is transmitted as a transport block, which means that, if lost, it can be recovered by HARQ retransmission, but it suffers from HARQ NACK/DTX to HARQ ACK residual errors and there are potential security risks (e.g., no ciphering nor integrity protection). MAC CE is initially intended for less frequent signalling than DCI with no need of reliable delivery. For example, the mechanisms based on the MAC CEs defined in the early releases of LTE can recover from the MAC CE loss, like timing advance command, or DRX command.

Features like MIMO require a fast and a frequent control from the network to the devices on, for example, indication of transmission/reception parameters and reference signals for channel state estimation. The number of configurable parameters is higher than what a DCI command can accommodate, and some are (semi)-persistent commands which may not be suitable in DCI. RRC configuration does not fit either due to the delay associated with the RRC message, e.g., large message size, or long time to deliver/process.

Consequently, many such DL controls are added as MAC CE commands. To ensure reliable transmission, implementation-specific things are needed to make it as reliable as RRC configuration, such as a better coding for DL transmission if the base station knows that it contains a DL MAC CE, adjusting the detection threshold to reduce the HARQ NACK/DTX to ACK error, or waiting for confirmation to avoid state mismatch.

In 6G, with an expectation of an increasing number of use cases, scenarios and requirements, the patch-on solution of using MAC CE to perform DL radio resource control will not scale. Thus, there is a need to develop a simplified and easy-to-use high-performance Radio Resource Control (RRC) framework that can handle the ever-increasing flexibility of L1/L2.

3.3.2 Mobility procedure in RAN

3.3.2.1 Analysis of the State of the Art (SotA) and its limitation

Mobility procedures have been under constant developments with enhancements in each 3GPP release of the 5G. Those are tailored for specific scenarios with fragmented supports. This section analyses the pros and cons of each solution (e.g., legacy baseline layer 3 (L3) handover, conditional handover (CHO), layer 1 / layer 2 triggered mobility (LTM)) and provides a guideline on how a unified procedure can be.

In the case of legacy L3 handover, the UE is configured by the network to measure reference signals from neighbour cells and calculate the L3 measurements (e.g., Reference Signal Received Power, RSRP, Reference Signal Received Quality (RSRQ), or Signal to Intensity and Noise Ratio, SINR) based on configured reference signals. A measurement reporting from the UE to the network is triggered based on the conditions like quality of the received DL reference signals, which are further configured by the network. After receiving the measurement reports, the source gNB sends a Handover request to the target gNB. If the handover request is granted by the target gNB, the source gNB sends an RRCReconfiguration message to the UE, which contains information on how to access to the target cell. Thereafter, the UE detaches from source (i.e., stops data Tx/Rx) and syncs to the target gNB.

There are two noticeable drawbacks of legacy L3 handover, each is tackled by a separate enhancement:

1. A fast-moving UE can move out of the source cell coverage so that the UE is not able to send the measurement report nor able to receive RRC reconfiguration.
   - CHO is introduced for robustness. The procedure is triggered based on a network configured mobility event and the UE sends RRCReconfigurationComplete to the selected target cell. The mobility event has an associated condition to trigger the procedure and a set of conditional configurations the UE applies when the condition is met.
2. There is an interruption for data communication when the UE is under the process of syncing to the target.
   - LTM is introduced to reduce latency and interruption. In this procedure, the UE is pre-configured with a set of candidate configurations. After L1 measurement reporting from UE,

which can be more frequent and lightweight than L3 measurement reporting, the handover is NW triggered in which the gNB sends MAC CE to switch the UE to the selected target cell. In this procedure, the UE can pre-sync to the target gNB.

For both CHO and LTM, the network pre-configures the UE with a target RRC configuration in advance of execution. The differences lie in that for CHO, the execution is UE triggered based on a network-controlled condition; for LTM, the execution is networked triggered by layer 1/ layer 2 signalling and DL/UL pre-synchronization for contention-free random access. For both approaches, one of the drawbacks (among many) is that there is a cap on the maximum number of candidates, as the resources may need to be reserved in the target gNB, e.g., contention free random access resources.

CHO and LTM are introduced in two different releases in an incremental fashion. As can be seen, it is indeed not clear why pre-sync cannot be used in the CHO-like procedure. A unified signalling structure (at least for CHO and LTM) is thus preferable to allow for more flexible usage of CHO and LTM features. The details of the solution will be further elaborated in the next deliverable.

3.3.2.2 Enhanced SpCell change procedure with UE initiation

In 6G, it is expected that the number of handovers will increase with the potential increase in the number of cells given the use of high frequency bands with typically smaller cell coverage and with the increase of the connection density. Even using CHO/LTM mobility procedures, it would be very challenging to be able to manage the mobility of connected UEs efficiently and reliably within this highly dense environment with many cells. The UE may need to either perform handover to a non-prepared cell, which brings the drawbacks of the normal handover, or to be configured with many potential target cells, which leads to increasing handover preparations signalling for both UE and network, increases energy consumption and inefficient utilisation of network resources.

To overcome these issues, a new connected mode mobility procedure is required that is more scalable than conditional mobility and more reliable than baseline mobility, where the UE can initiate a handover to a non-prepared cell directly without waiting for the configuration from the serving cell (see Figure 3-7). The new connected mode mobility could be applied for handovers in both MCG and SCG (PSCell change), therefore the term “Special Cell (SpCell)” is used in the following. In this procedure, it is assumed that the UE does not have any AS-Context dedicated configurations for the target SpCell and that the target SpCell has no information about the UE. The network may still control UE mobility by configuring neighbouring SpCell lists or execution conditions with generic broadcasted configurations applicable on all UEs having this feature enabled. The figure below illustrates a high-level signalling diagram of the proposed mobility procedure. At step 1, the UE requests establishing AS-Context with the target SpCell.

Upon the request, the target cell retrieves the UE context from the source SpCell, prepares the UE AS-Context and request from the source SpCell taking over the UE at step 4. The source SpCell encrypts the UE AS-Context of the target SpCell via the source SpCell-UE security context and provides it back to the target SpCell. At step 6, the target SpCell sends to the UE the established target SpCell-UE AS-Context.

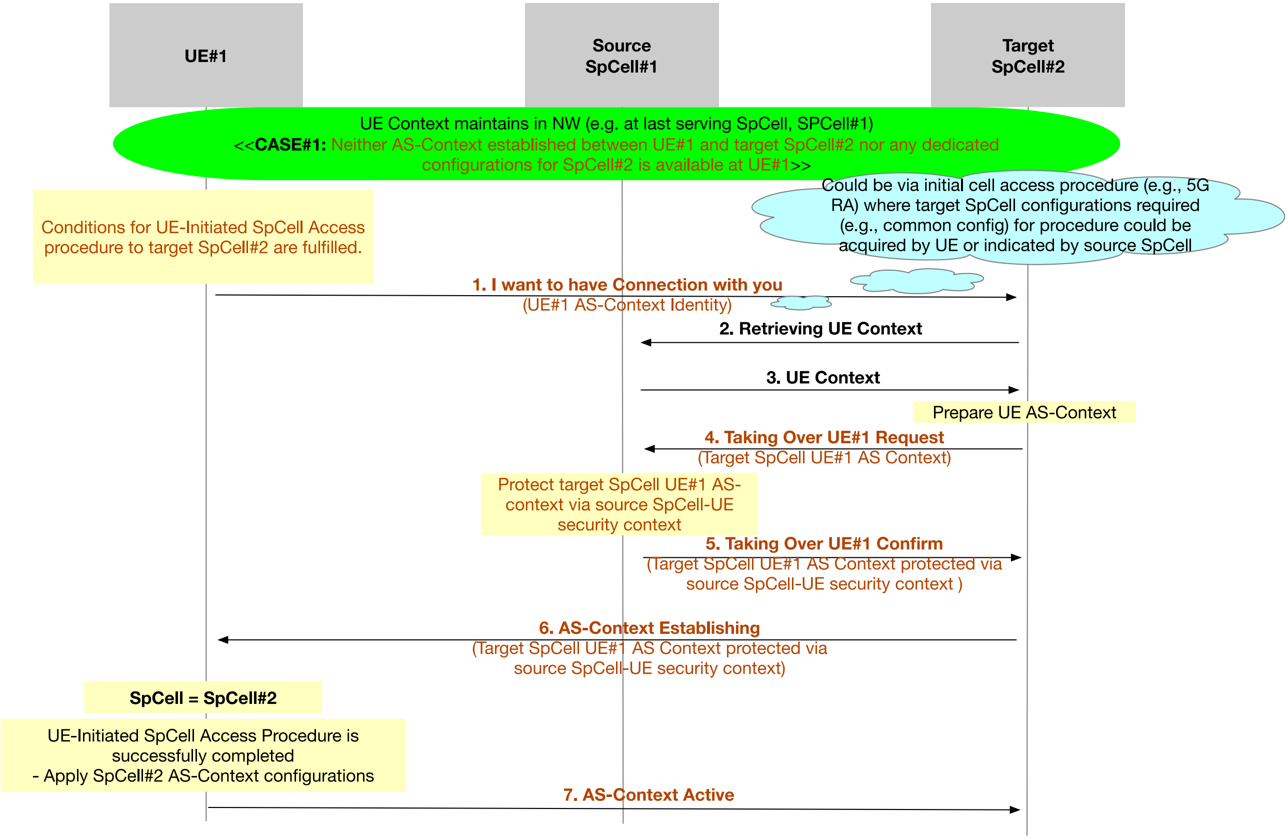

UE#1
Source SpCell#1
Target SpCell#2
UE Context maintains in NW (e.g. at last serving SpCell, SPCell#1)
<<CASE#1: Neither AS-Context established between UE#1 and target SpCell#2 nor any dedicated configurations for SpCell#2 is available at UE#1>>
Could be via initial cell access procedure (e.g., 5G RA) where target SpCell configurations required (e.g., common config) for procedure could be acquired by UE or indicated by source SpCell
Conditions for UE-Initiated SpCell Access procedure to target SpCell#2 are fulfilled.
1. I want to have Connection with you
(UE#1 AS-Context Identity)
2. Retrieving UE Context
3. UE Context
Prepare UE AS-Context
4. Taking Over UE#1 Request
(Target SpCell UE#1 AS Context)
Protect target SpCell UE#1 AS-context via source SpCell-UE security context
5. Taking Over UE#1 Confirm
(Target SpCell UE#1 AS Context protected via source SpCell-UE security context)
6. AS-Context Establishing
(Target SpCell UE#1 AS Context protected via source SpCell-UE security context)
SpCell = SpCell#2
UE-Initiated SpCell Access Procedure is successfully completed
- Apply SpCell#2 AS-Context configurations
7. AS-Context Active

Figure 3-7: UE initiated SpCell change.

The new proposed UE-initiated SpCell change procedure is less complex and more efficient in high connection density environments for UE as the cell preparation is not required and reservation of dedicated resources is reduced in comparison to CHO. Moreover, complexity shall not be increased as the AS-Context transfer, security update and the path-switch are anyway performed in both the baseline and conditional handover procedures. Additionally, the complexity of the proposed solution may be less than CHO/LTM as the network does not need to prepare any cells. Although the UE-initiated SpCell change may increase the latency during handover due to signalling between the source SpCell and the target SpCell, it can co-exist with conditional mobility. CHO can still be applied in certain use cases where mobility latency is very critical.

3.3.2.3 Mobility robustness in 6G multi-connectivity

Drawbacks of the dual connectivity (DC) solution have been extensively discussed in Hexa-X-II deliverable 3.2 [HEX223-D32], including the need to coordinate between the master node and the secondary node for flow control and UL transmission power split, the coupling of UL and DL in any one of the legs, and the specification complexity to support various connection options. The goal in 6G is to combine the best features from carrier aggregation (CA) and DC with only one multi-connectivity solution. This section analyses how mobility robustness from DC should be considered in such a single multi-connectivity solution.

For DC, upon either master cell group failure or secondary cell group failure, the other leg can take over the data transmission and additionally the recovery of the failed leg can be done via the functioning leg. The connection interruption time in most cases is zero in the case of a low probability that both links fails at the same time.

Assuming the new multi-connectivity solution is based on CA as a baseline. In what follows, the available options are explored. The first candidate solution direction is to allow Primary Cell (Pcell) recovery by Secondary Cell (Scell) which might still be working (see Figure 3-8). This is similar to the approach for DC.

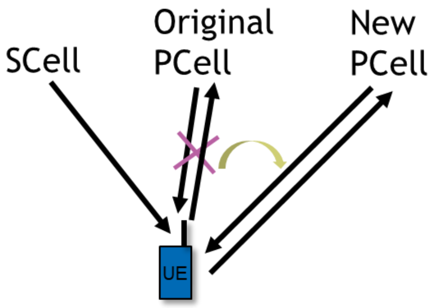


Figure 3-8: Pcell recovery.

Technically, if Pcell fails, the UE can still recover from the failure by RRC re-establishment procedure but it takes long time, due to a conservative Radio Link Failure (RLF) triggering condition, cell re-selection and RRC re-establishment procedure delay. Another solution direction is to reduce the delay of these procedure, and solutions are already discussed in 5G.

Figure 3-9 shows an example of these solutions.

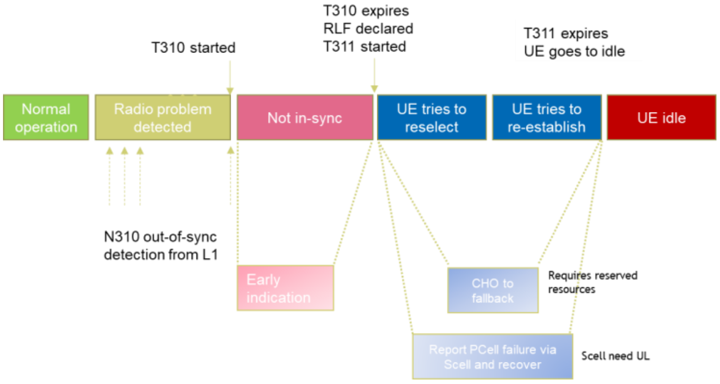


Figure 3-9: PCell recovery delay component and how to reduce the delay.

With all these in mind, the proposal for 6G mobility procedure in the case of a single multi-connectivity solution should not allow a single point of failure and long recovery time, as was the case of Pcell failure in 5G. The detailed description of the enablers will be in the next deliverable.

3.3.2.4 Data-driven Mobility

The movement of the UE may result in the degradation of the serving cell signal strength and lead the UE to perform handover to the strongest cell, in terms of reference signal measurements observed by the UE. The handover decisions for both baseline and conditional handover procedures rely on the pre-configured signal measurement events, e.g., the A3 event (neighbour signal strength becomes offset better than SpCell) [38.331] where the accuracy of the measurement affects the handover success ratio. However, the handover decisions which are based on signal measurement only may increase the energy consumption of the UE as the UE needs to perform frequent measurements on reference signals to increase the measurement accuracy. The L1/L3 filtering mechanisms cause the UE to capture abrupt changes in the signal strength late, e.g., when a user enters a tunnel. Thus, it may result in radio link failure and may increase the service interruption time. In addition, the handover decision is agnostic to the user and mobility context and may lead to unnecessary handovers.

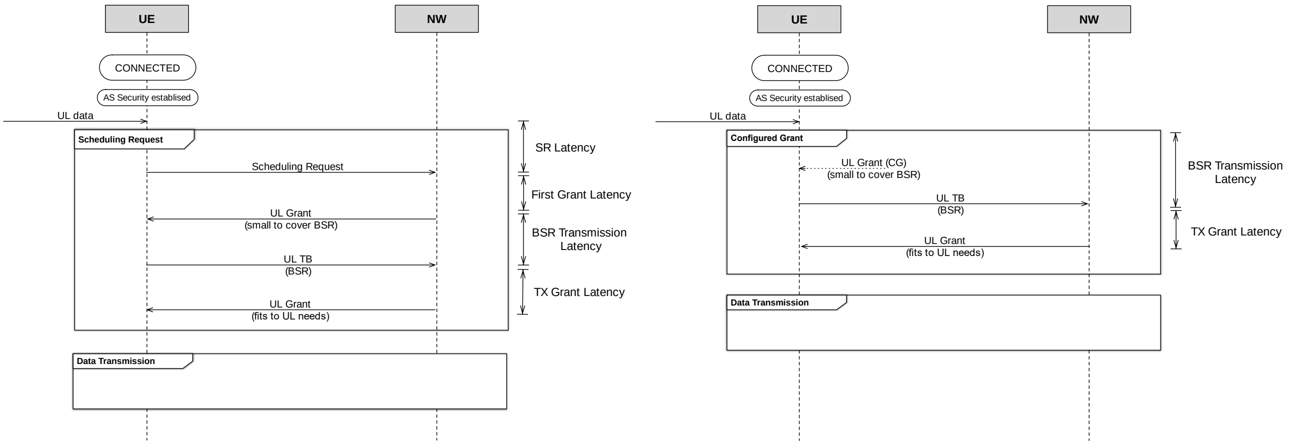
In 6G, the enhanced capabilities of UEs in terms of computation and sensing facilitate the execution of data-driven decisions using advanced AI/ML techniques at the UEs. For mobility, the UE can improve the connected mode mobility decisions by having information about its radio environment and application usage, which is not available in network. To overcome the aforementioned issues and leverage capabilities of UE devices, a new mobility decision mechanism is required, which considers not only the signal measurements but also UE contextual information. In addition to traditional mobility signal measurement events (A1, A2, A3, A6, B1, etc.), (see Section 5.5.4 of 38.331 [38.331]) new measurements events can be introduced, which rely on contextual information at the UE. The type of events can be divided into two groups.

- Radio/Environment Events:
    - Depend on UE capability of understanding/sensing e.g., indoor condition, tunnel detection, etc.
- User Traffic Events:
    - Depend on expected UE traffic over time e.g., if high throughput is expected to be exchanged between the UE and the network, UEs can be prepared/handed over proactively.

Note that, these events should be quantized to ensure user privacy. The confidence level calculation and thresholds can be agreed between UE and the network to manage the information sharing e.g., model output, etc. and/or execution of the events.

3.4 Application-NW interaction for service differentiation and QoS/QoE management

For latency critical use cases, like immersive telepresence, UL latency (i.e., from UL packet arrival to actual transmission) shall be minimal. Additionally, the “setup” latency needs to be minimized as it may become the predominant latency addition factor in short data sessions (e.g., HTTP bursts). In 5G, there are two mechanisms for UL grant scheduling to the UEs, which are Scheduling Request (SR) and Configured Grant (CG) (which was introduced as an UL latency reduction feature). Both of those mechanisms rely on the UE Buffer Status Reports (BSR) to do the resource planning and allocation. Figure 3-10 provides an overview figure of the Message Sequence Chart (MSC) and latency adders in both SR (left) and CG (right) UL grant scheduling mechanisms from NR. As seen from the figure, the CG mechanism offers a reduced overall latency in comparison to SR, but of course at the expense of potential resource waste as CG would require the reservation of dedicated UL resources. For different services and applications or for a combination of them, the UEs may know specific characteristics of upcoming traffic, like number of upcoming streams, protocols, packet sizes, cadence, jitter constraints, delay budgets among many other things. Those traffic characteristics are application layer driven and getting more dynamic, and the access network is not aware of such characteristics as it only sees the reported BSRs. Note that the access network should not be aware of all traffic characteristics to ensure user privacy and scheduler simplicity. In addition, traffic characteristics may change dynamically over time due to the huge number of different applications on the UE side. However, from the UE’s perspective such traffic may still be predictable for certain applications and more precise scheduling can be beneficial for the overall user experience. Hence, in 6G a mechanism where the UE aids the RAN scheduling based on the applications/traffic characteristics, may offer a great enhancement in the overall user experience.

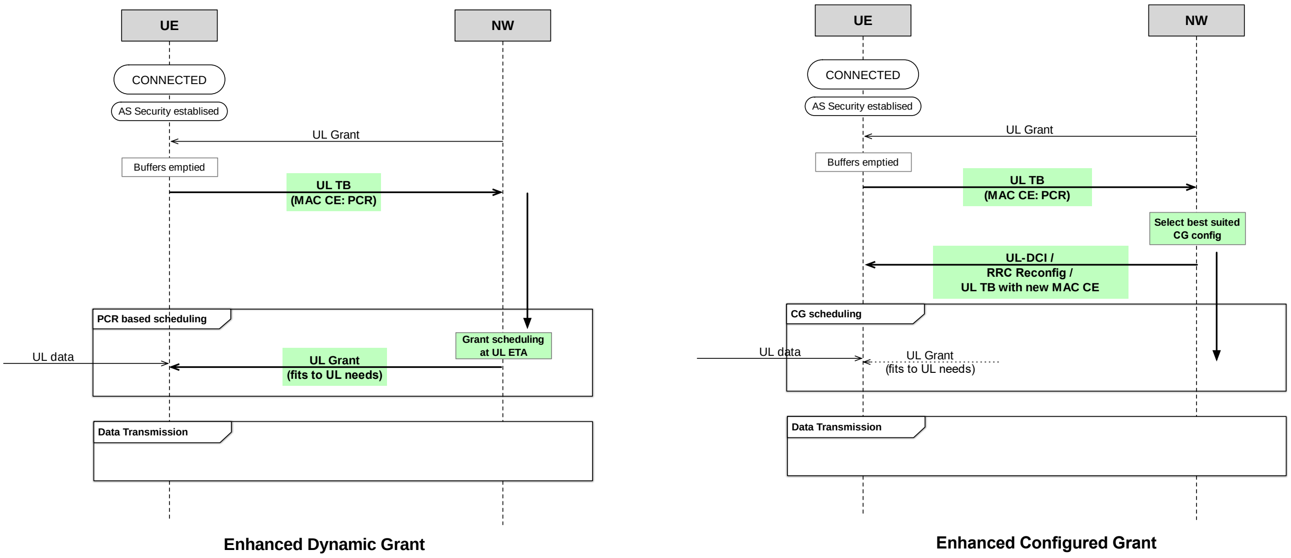


Scheduling Request

Configured Grant

Figure 3-10: NR UL Grant scheduling mechanisms.

A proposal in this direction is to employ the so-called zero-latency scheduling. In this approach, the UE indicates what traffic characteristics to expect via a new reporting mechanism called Periodic Cadence Report (PCR) sent to the NW, that is expected to enhance the dynamic grant as well as the configured grant scheduling (i.e., see Figure 3-11). For dynamic grants, this mechanism may deem the legacy SR mechanism as obsolete, since the NW would be able to offer a better opportunistic UL scheduling given the additional information provided in the PCR. Even before the arrival of the first packet in the modem’s UL buffer, the application may inform the UE's modem about such traffic parameters (e.g., during the connection setup phase). For the CG enhancement, the NW may setup, reconfigure or release the CGs based on PCR information or activate a different pre-configured, better-suited CG based on the PCR. The activation and de-activation of CGs could take place via RRC for Type1 CGs, DCI for Type2 CGs or even with a newly introduced DL MAC CE, which would indicate the pre-configured CG to be activated.



Enhanced Dynamic Grant

Enhanced Configured Grant

Figure 3-11: PCR enhanced UL Grant scheduling.

Such PCR may be sent by the UE either periodically or based on specific events, such as when the buffers are emptied to inform when the next UL grant is expected or upon change of traffic patterns or upcoming scheduling changes (e.g., to allow apps to adapt to jitter).

In comparison to BSR, this new PCR mechanism may overcome the BSR's issue of providing only quantized snapshots of the current UE buffer state. Additionally, it avoids the waste of resources with zero padding, where the NW schedules the UE until its buffers are empty**.** PCR information may also help the NW to know the traffic of different UEs and how this traffic would change, in order to achieve better resource utilization and lower latencies, as well as to enable a greater role for the UEs in shaping the UL assignments more towards their application needs. In comparison to SR and CG, this scheme would offer a considerable saving in terms of latency. This is achieved by skipping the whole SR and BSR logic by being proactive and either aligning the UL grant reception with the data expected time of arrival from higher layers and/or with the cadence of the data packets or informing applications with the expected UL grants to sync the traffic accordingly. Additionally, this is still a NW controlled dynamic grant that avoids waste of network resources.

3.5 Protocol support of 6G radio enablers

This section gives an analysis on protocol support of 6G radio enablers, while the previous sections are related with enablers on protocols. More specifically, detailed analysis is provided on energy efficient radio design towards more sustainable operation and a new 6G capability of sensing and its interplay with positioning. The section ends with an initial analysis of more enablers, and the details will be provided in the next deliverables.

3.5.1 Energy efficient radio design

From the radio perspective, a significant number of enablers play an important role in fulfilling the targeted 6G KVIs. In particular, when seeking sustainability, an energy and cost-efficient radio design is a crucial enabler that has to be carefully considered towards the mentioned key value.

In contrast to previous generations of wireless communications, which predominantly focused on maximizing the spectral efficiency and peak data rates, a new concept in the radio design is necessary to satisfy the diverse requirements foreseen for the near future. It should be mentioned that a wide variety of applications demand low to moderate data rates. Furthermore, as higher frequency ranges are considered for communication and the spectrum availability increases, the required data rate could be achieved in many cases without the necessity of resorting to high spectral efficiency schemes. This fact, together with the high amount of energy consumption associated to the radio access, leads to a fundamental change in the radio design, from maximizing the spectral efficiency and the peak data rate to maximizing the energy efficiency considering the spectrum availability and providing the required data rate.

This concept can be realized through several operation modes, each one comprising a predetermined hardware configuration together with adaptable software settings to guarantee a certain range of energy efficiency. The number of configuration possibilities will be limited, thus ensuring flexibility while avoiding an overcomplicated structure that might limit the deployment, maintenance and cost-efficiency. Besides, the radio design should not be perceived as an isolated enabler but rather as a component that can be seamlessly integrated into the E2E system, following the system design principles. This integration necessitates dynamic interactions with several layers of the system, ensuring alignment with the desired performance of the system. Appropriate communication and exchange of information with respect to the hardware and radio resources, application requirements and wireless channel conditions will guarantee the radio compliance with the architecture, complementing it without any restrictions.

Since this radio design involves the incorporation of new operation modes, which include the use of different hardware, software settings and signal processing, it is expected that the radio interface and protocols have to be adjusted to include the information needed for proper configuration and operation, for example, from a network node to a device. A more detailed analysis of the impact of this enabler on the radio interface and protocols will be conducted in the near future.

3.5.2 Sensing and positioning

3.5.2.1 Positioning in 6G

Sensing functionality is envisioned to be an integrated part of the communication network by reusing the communication spectrum and the communication infrastructure (e.g., deployed network nodes). Previous generation cellular networks support estimation of the location of a wirelessly connected device. Sensing goes beyond that, e.g., estimation of the presence of passive objects, environments, to know where the clutter is and feature detection. This section discusses the similarities and differences between the two from the radio protocol point of view.

5G supports various RAT dependent positioning methods. In essence, when positioning of a connected device is performed, the signalling exchange on the radio interface is reflected by transmitting reference signals either from the network or from the device, reporting measurement results of the received reference signals and the configuration of these. The measurements can be used to understand multi-path, power, angle, and time estimations. Two types of specific reference signals for positioning were defined in 5G, Positioning Reference Signal (PRS) in DL and Sounding Reference Signal (SRS) in UL.

The main concerns in RAT dependent positioning methods are the latency and accuracy. In 6G era, positioning techniques will play critical roles in many industry use cases, for time sensitive applications. Some of these use cases are identified in Hexa-X-II project [HEX223-D12]. Under current 5G positioning architecture and radio protocol design, the average time from the positioning request to retrieve location information falls in the order of 100 – 250 ms, which is far beyond the requirements of many mission-critical industry use cases. 6G needs to redesign the positioning architecture, signalling and measurement process at the PHY layer.

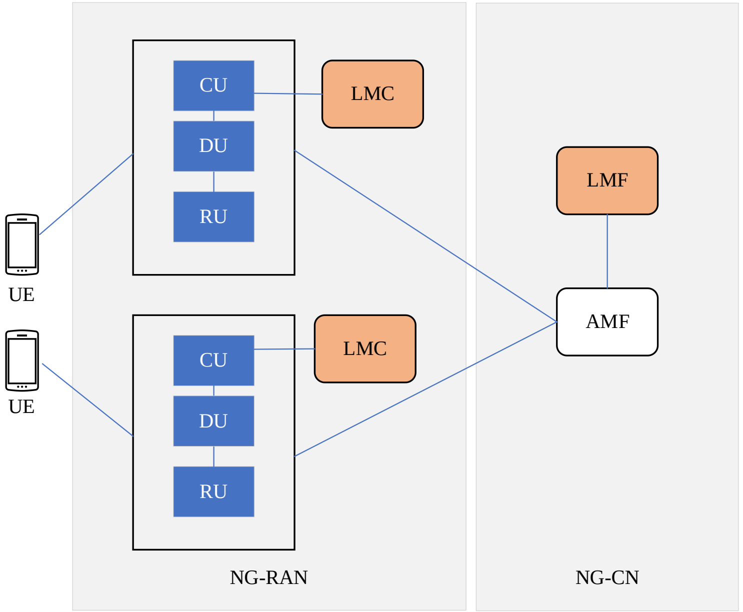
The positioning architecture in 5G is illustrated in Figure 3-12. The key function for positioning is the location management function (LMF) in the core network. The LMF collects measurement reporting from user equipment (UE), calculates the location, and offers location information to the requested client. More details of the architecture description can be found in [38.305]. For positioning with the aid of new sensing functionality, the information related to positioning will be expected to be processed under the similar positioning architecture.

A typical location request procedure is shown in Figure 3-13. The LTE positioning protocol (LPP) originated from 4G is still used in 5G to exchange positioning information between LMF and UE. The New Radio Positioning Protocol A (NRPPa) is used to carry the positioning information between gNB and LMF. It is expected that 6G will follow the similar procedure, but with significant improvements to reduce the latency.

With the positioning architecture and signalling procedure in mind, it can be easily understood the prominent positioning latency problem in 5G networks. The whole E2E latency includes the measurement at the device to collect positioning-related information, including PHY-layer related processing of control signalling, the signalling between the device, gNB and LMF, and location calculation. The new system and radio protocol design for latency enhancement will consider the following aspects in 6G positioning:

- Optimizing PHY layer procedure and the measurement of reference signals, e.g., PRS, SRS, for location estimation of devices by leveraging the higher available bandwidth in 6G spectrum to speed up the time-of-arrival acquisition and angle-of-arrival measurement with improved precision.
- Redesigning the signalling process between a device and different network components to avoid unnecessary signalling and combine multiple messages into one for less message exchange.
- In 5G, the location information is transmitted between device and LMF and is transparent to the RAN node. In 6G, the radio and LPP protocol need to be redesigned to allow the direct access of the location information by gNB for positioning latency reduction.
- Moving LMF closer to RAN nodes. 6G will have a flexible architecture design to allow core network functions deployed closer to the edge. The LMF will be located to the edge node for latency improvement. The new location management component (LMC) and mechanism can be directly introduced in RAN nodes for future latency reduction, which is shown in Figure 3-13.

Note that positioning architecture provides the reference design to implement sensing functionality in 6G. The same principles can be applied to sensing architecture design for drastic latency reduction. The integration of sensing and positioning architecture under the same framework is expected in 6G.



AMF: Access and Mobility Management Function, CU: Central Unit, DU: Distributed Unit, LMC: Location Management Component, LMF: Location Management Function, RU: Radio Unit, UE: User Equipment

Figure 3-12: Positioning architecture beyond NG-RAN [38.305] with LMC.

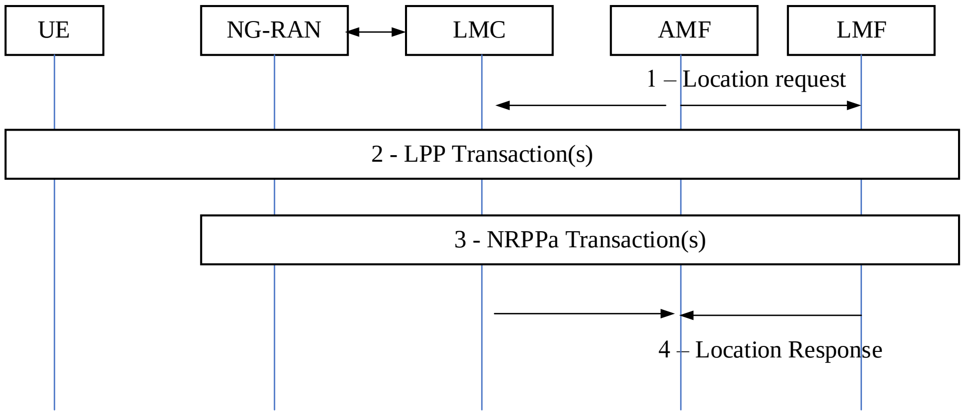


Figure 3-13: Positioning flow diagram with new local management component (LMC).

5G uses signal strength, time-of-arrival or angle measurement to estimate positioning. In Rel-17, the target accuracy reaches up to 20 cm. The cm-level is expected to be required in certain 6G use cases.

The accuracy of the signal strength measurement is limited due to the uncertainty in path loss of the radio channel. The accuracy of the angle-based approach is limited by the antenna design. The time-based positioning has accuracy specified by the bandwidth of the PRS. In addition, the accuracy of these approaches is prone to reflections in radio propagation environments.

With higher bandwidth, the pulse of reference signals can be shortened, and thus reduce time estimation error. As 6G may have a large BW due to higher frequency ranges, new time-based approaches will be introduced to offer better accuracy due to the evolution towards higher frequency bands. In particular, carrier aggregation will be exploited in 6G for positioning, in which positioning reference signals from different carriers can be combined into a signal with high overall bandwidth. The coherent combination of multiple PRS will require knowledge of the phase between other carriers, which is typically unavailable in different RF chains. It needs the new radio design to make available such information.

In Global Navigation Satellite System (GNSS), real-time kinematic (RTK) achieves cm-level accuracy. A similar approach, called carrier-phase based positioning, can be adopted in 6G to achieve highly accurate positioning. The radio signal propagates as a wave, where its phases can be translated to the distance, as a complete phase cycle equals a wavelength of the carrier signal. The detection of the phase difference at the receiver can indicate a slight difference in distance. A highly accurate distance estimation can be achieved by combining with the time-based approach, which can be used to determine the total number of phases the propagation has taken.

A carrier-phase positioning approach is illustrated in Figure 3-14. Reference gNBs know their positions and relative clock offset, which enables phase and time coherency between them, and can assist the target UE in performing differential time and carrier-phase measurements on pairs of carrier-phase PRS transmitted by multiple gNBs. The measurements are sent to the LMF to calculate the difference between measurements from the target UE and reference gNBs. Then, the LMF estimates the position of the target UE.

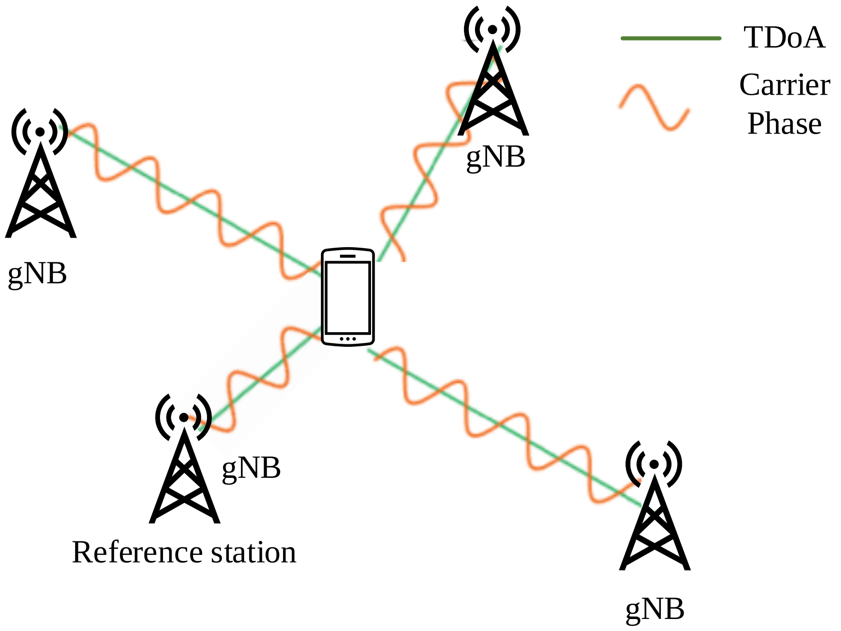


Figure 3-14: Carrier-phase positioning approach [FAN22].

3.5.2.2 Similarity and difference between sensing and positioning

6G should strive for a unified approach on the radio protocol to support both positioning and sensing. Both Positioning and sensing can be achieved by a node (either a device or a network node) transmitting a reference signals (or in some cases rely on transmitted user data), measuring the received reference signals (which can be reflected, have LoS component or even back to the transmitter) by the same node or another node, and (if needed) reporting measurement results. However, there are several distinct differences between sensing and positioning, for example:

- design of sensing reference signals to get an even richer information set of the objects beyond location, e.g., movement, orientation, or material properties.
- post-processing of the received sensing measurement from one or more devices and base stations will be different from positioning.
- third party objects not connected to the network, the privacy perspective is different from the location service of a user device connected to the cellular network, since the third-party objects does not have the opportunity to opt-out to the sensing.

In what follows, a non-exhaustive list of the simplest to more complicated sensing scenarios are discussed.

1. **Mono-static sensing with BS only**

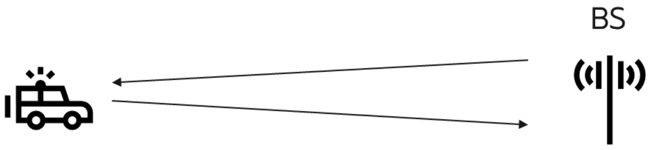


Figure 3-15: BS only mono-static sensing

In the scenario depicted in Figure 3-15, all radio resources are under network control and only network nodes are involved. The base station allocates its own DL transmission resources and is prepared to receive reflected signals in the UL.

2. **Bi-static sensing with BS/UE**

In the case of bi-static sensing shown in Figure 3-16 and Figure 3-17, one requirement that influences the performance of sensing heavily is the level of synchronization between the transmitter and the receiver. In the presence of a LoS propagation path between the transmitter and the receiver, a common approach of synchronization relies on the estimation of the time-of-flight (ToF) of the LoS path. This information can be used as a form of synchronization for measuring the ToF of propagation paths between the transmitter, targets, and the receiver. An alternative form of synchronization relies on the use of an available clock, such as the one provided by GPS. Furthermore, for the case of Figure 3-16, conventional approaches of synchronization of Bs can be used.

There is also a requirement that the transmitter needs to know the location of the receiver to correlate with the measurement results to fulfil the intended sensing service request.

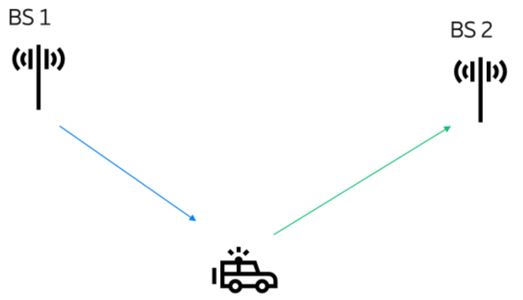


Figure 3-16: BS only bi-static sensing

Figure 3-16 is an extension of the BS-based mono-static sensing. What comes additionally is the need to coordinate between the two base stations, e.g., BS1 indicates to BS2 to mute uplink transmissions in BS2 to prepare for reception of reflected signals, or BS2 indicates to BS1 which time and frequency resources are free, i.e., can be used for sensing. This can be done via direct communication between two base stations or via a central entity. Regardless of which approach, it is similar to BS coordination in the case of positioning (e.g., UL SRS reception at the neighbouring cell).

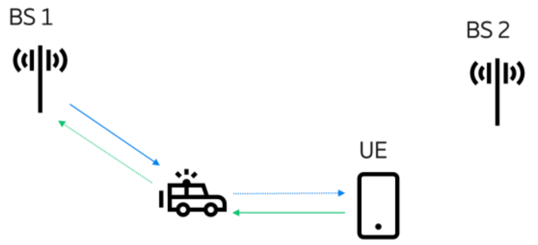


Figure 3-17: Bi-static sensing with BS and UE

In the scenario shown in the Figure 3-17, UEs are present in areas of interest and can most likely reach coverage that may otherwise be occluded from a base station only and thus it seems that there is a need to have multi-static sensing in which both base station and UE are involved.

As can be seen from the figures above, this is very similar to the positioning framework in which the reference signals are transmitted, measured at a connected device or network, and the measurement results are reported.

For positioning framework, there are multiple base stations involvements and so the time sync is upon the network side. In the above case, in order to make sense of the measurement (which include also time difference), an accurate time sync between BS1 and UE may be needed. One can use a separate over-the-air time sync on top, e.g., accurate time sync with 1 nanosecond in 5G. One noticeable difference is that when the base station is the final consumer of the sensing results, the base station might need to know where the UE is located (assuming that the UE is moving or stationary but with initially unknown position).

3. Mono-static sensing with UE only

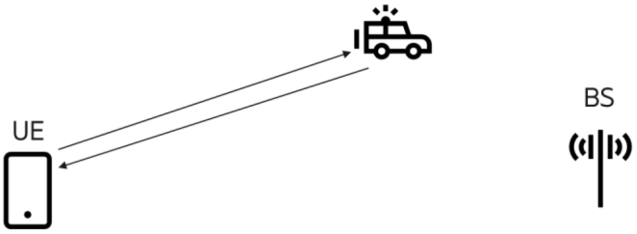


Figure 3-18: UE-only sensing

In the scenario shown in Figure 3-18, the base station configures the resources (e.g., time, frequency, and beam relations) for the UE to operate, i.e., to transmit and to measure the reflected signals. The argument to support this is similar to the sidelink in the sense that the sensing area is rather in the vicinity of end-user devices than close to a base station. The configuration should be done in such a way not to interfere with other ongoing transmissions. The sidelink framework with network-based coordination can be a good starting point to investigate how to support this operation, e.g., NW assign side link resources. A more complicated scenario is explicitly studied in WP4, see section 9.2.6 from Hexa-X-II D4.2 [HEX223-D42]. The resource allocation methods and protocols, and also synchronization aspects to enable inter-UE sensing are an open question.

Summary

When a base station is involved in the sensing operation either in sensing signal transmission or sensing signal reception, there are similarities between sensing and positioning from the radio protocol interface/protocol point of view and thus a unified approach is preferred. Nevertheless, some specifics about sensing need to be considered:

- Differences lie in detail on what to configure, what measurement results to send, which entities to collect the measurements etc.
- In bi-static sensing with UE, it is of importance to be able to synchronize the base station and UE and also to know the location of the UE at the time of sensing.

On the other hand, when only UE are involved, the sidelink-based approach of allocating resources might be relevant and can be a basis for further analysis. Lastly, the analysis of a multi-static sensing setup is left for future work.

3.5.3 Other enablers

As details of enablers are not settled down yet, this section provides an initial analysis of enablers that may have impacts on the radio protocol. This is not an exhaustive list and more enablers, and more analysis will be provided in future deliverables.

Compute offloading: In order to support the new requirements on distributed computing latency and quality of compute, different functional nodes (i.e., computing, offloading and compute offload controlling nodes) need to be introduced as well as some potential modifications to the traditional CP/UP procedures might be needed. The scope of these changes will depend on the network architecture adopted and the split functionality between RAN and CN. Moreover, the deployment of the different functional nodes in the NW, might entail the need for some further extensions to the radio protocol stack. In Hexa-X-II D3.2 [HEX223-D32], some of the procedures already introduced in the compute offload architecture already hint at the need of further analysis of the current protocol stack to support new procedures like node discovery and computation offload.

AI: Enabling distributed ML among cellular nodes requires data and model exchange between the cellular network and the UEs. This implies new requirements related to data collection, training, and inference such as privacy preservation and coordination of data and learning among cellular nodes. On-device/UE machine learning training and inference could be partially offloaded in a distributed fashion to the network or other UEs, while preserving user privacy requirements.

This requires the introduction of novel architectural elements that enable for privacy aware data collection and learning [HEX223-D32]. The main architectural component is the UE aggregation unit, which performs privacy-preserving data aggregation by leveraging secure aggregation techniques to collect UE data. UE aggregated data is shared with the data-driven network control unit, another architectural component, which can be used by the network to improve its control and decisions.

The protocol stack may be affected by coordination and signalling of these novel architectural components. Moreover, the coordination mechanism for keeping the models and data up to date and synchronized among the cellular nodes may additionally have an effect on the protocols.

Subnetworks: A subnetwork is a collection of nodes operating with the control of a management node (MgtN) under coordination with the overlay network. The management node is a new UE role in a subnetwork, and it is the primary node of the subnetwork, which can communicate with the BS and other UEs. Various subnetwork architectures will be considered, based on the connections between the BS, the management node and the UEs. Based on these architectures, the investigation will focus on how the management node may assist the UEs in its subnetwork to reduce the complexity of various control plane and user plane procedures. The operation of the subnetwork may have an effect on the radio protocol stack, regarding the air interface and new procedures between the management node-BS and management node-UE (e.g., based on legacy UE to base station interface, legacy UE to UE interface, other access networks) and based on what spectrum the subnetwork operates on (i.e., licensed or unlicensed) and on the role of the management node and on whether the subnetwork is transparent or non-transparent to the NW.

Distributed-MIMO (D-MIMO): The architecture modelling of the D-MIMO has not concluded, e.g., from the physical layer point of view, the difference between D-MIMO and UL/DL coordinated multi-point (CoMP) is not clear. Further, in 5G, there is a concept of multi transmission-reception point (multi-TRP) and there is a need to check whether the multi-TRP framework can be used as a baseline and extend those to support D-MIMO. Lastly, there is a need to check the mobility procedure depending on the assumption if D-MIMO is across different cells or rather in one cell.

Reflective intelligent surface (RIS): From radio protocol and interface point of view, this is very similar to a network-controlled repeater as discussed in 3GPP Rel-18. The open question is whether there will be UE-controlled repeaters and there is the follow-up question on whether the network is aware of such a RIS.

Energy neutral device: This has been categorized in deliverable D5.2 [HEX223-D52] and with technological enablers to support low-overhead and energy-aware operation. The challenge on the radio protocol design is how to incorporate the aspects of energy neutral devices in the baseline. If not included in the baseline, how to ensure easy extensibility in the 6G time frame for those devices.

4 Enablers related to E2E service management and automation

One of the key technical pillars within the Hexa-X-II project is the study and evaluation of intent-based management. To this end, and from an E2E point of view, an investigation has been done looking for the elements necessary for the design of an E2E solution that is able to manage the services of different administrative domains as autonomously as possible using intent-based requests. By using intents, the responsibility of the users to know how to implement the desired services is given to the management and control system and the users only focus on what they want. This chapter is organized in two main sections: the first section focuses on the intent-based Digital Service Manager (DSM) (i.e., the management solution belonging to a Digital Service Provider (DSP)) functional architecture design, and the second section describes the multiple proposed enablers that are based on the multiple functional blocks previously presented. Finally, in this chapter some concepts are used (e.g., DSP, tenants, etc.), which are properly introduced in subsection 6.2.2.2.4.

4.1 Preliminary intent-based digital service manager functional architecture

The use of intents should simplify the services and networks resources management, by allowing to give commands to the control and management systems with simple orders such as: “Deploy service AA between points X and Y with low latency”. Other formats to express intents are possible as described later. Regarding the previous example, once the intent is received, a translation from human to machine should be applied in order to generate the data object to be used by the control and management system and which allows to be compared with other deployed or simultaneously requests with the objective to evaluate if the incoming intents are feasible (with the available resources) and if the may generate any kind of conflict among them.

Based on these introduced concepts (i.e., request, translation, and conflicts), this section describes the State of the Art (SotA) focused on intent-based network architectures, the interfaces necessary to manage intents between the entities requesting them and the entities managing them, and finally the requirements related to the architectures and solutions to manage the intents. Following the presented SotA, a framework to control and manage intent-based requests is presented with the high-level architecture defining the functionalities and capabilities required to manage properly the life cycle of an intent.

4.1.1 Intent-based management architectures and interfaces (SotA)

The main objective of IBM is to remove the need to know the details about the infrastructure resources and which specifications are required to manage (i.e., create, configure and terminate) services within a network. By using intents, the user only uses an abstracted request what is the desired service, and it is the system itself which must translate from the “human” request towards the “system” request in an autonomous way. In terms of initial research and standardization work, there are different works focused on IBMs.

Intents have their origin in the work related to Policy Based Management (PBM), a topic well studied by multiple SDOs such as the ITU, TMForum and the IETF. Among their main aspects to study, they defined important elements such as the PBM architecture [HSW05] or the classification of policies in imperative policies (i.e., event-condition-action) and declarative policies (i.e., intents) [MES+01]. Based on these references, the work done within the Hexa-X-II project regarding intents focuses on how to use intents by an E2E management and orchestration system for the deployment of services and the management of the resources involved in multiple domains in order to deliver the expected E2E service.

One of the first documents in terms of standardization is presented in the RFC 9315 [CCG+22], where the basic concepts and definitions about IBM are presented. An even more recent work was done by the ETSI Zero-touch Network and Service management (ZSM) group, which presented their first group report [ZSM011] about how intent driven autonomous network should be and work. In there, they use their experience on the ZSM architecture to define how the IBM should be across an E2E multi-domain scenario. Among the multiple aspects within the IBM field, there are two important concepts to consider: the translation from human requests to “intent-based” request as presented in [JPR21] and the intent conflict resolution. IBM systems rely on intent definitions to automatically configure and manage the network, but conflicts [ZLF22] can arise when these definitions are not consistent, or mutually compatible. Conflicting intents can lead to inconsistent network behavior, reduced performance, or other issues negatively impacting the network.

To manage these two tasks, the ZSM group presented the evolution of their ZSM architecture [ZSM011] that defines how intents should be managed taking into account the characteristics of each management domain involved together with the overall (i.e., E2E) point of view. Now the ZSM architecture considers new elements: the Intent Management Entity (IME) and the (Domain/E2E) Intent Handling.

Regarding the IME, and as also defined by the TMForum in IG1253C [TMF_IG1253] and the IRTF [CCG+21], it may play two different roles: the intent owner and the intent handler. The first is the intent source/requester and must take care of its intent lifecycle, being the only one that can manipulate the intent itself. Instead, while the intent handler has the responsibility to extract requirements, goals and constraints and to operate to make the next state of the system compliant with the intent specifications defined/requested by the intent owner. It is worth noting that one IME may play simultaneously both roles, but only one of the two roles per a given intent. Finally, the intent handling is the interface used by the different IMEs to interact among them as owners or handlers.

Based on the previous intent roles, the intent interface is the one between the two intent management functions covering the roles of owner and handler respectively. On one side, the handler exposes the intent interface, also called Intent Handling Management service, that provides the owner with all the features required for the lifecycle management of the intents. Furthermore, the intent interface reports on the status of the intent, creating a control loop between the handler and the owner that is closed when the intent succeeded. On the other side, as presented in [ZSM011], the owner is the Management Service (MnS) Consumer of the interfaces exposed by the handler (called MnS Producer). In complex systems, multiple sources of intents can coexist, as well as multiple handlers and depending on where an intent is originated, each interface will vary. For example, intent objects coming from the Communication Service Customer will be served by a different interface that intent objects coming from the Communication Service Provider (CSP) [CCG+21].

One owner may interact with multiple handlers and vice versa and both owners and handlers can manage more than one intent at the same time. Nevertheless, there is an important constraint: the intent used between a couple owner-handler is always a unique object so each intent has only one owner and cannot be sent to multiple handlers. The intent management functions can cover the role of either owner or handler with respect to the same intent interface. However, a handler can fulfil an intent request by creating other intents to be sent to its subordinate domains, so also becoming an intent owner.

When dealing with intent-based interfaces, the set of APIs exposed by an intent-based system are indicated, enabling the IBM and the separation of concerns between the system that defines the intent and the system that manages the intent. Moreover, TMForum [TMF_IG1253] also defines a set of mandatory and optional procedures that an intent interface should support, this work regarding the definition of the related APIs is a work in progress and is expected to result in the new API TMF 921 providing all the interface details. Meanwhile, [TMF_IG1253] assumes that intent interface will be asynchronous and REST-based, and in addition, TR-523 [ONF_TR523] introduced the discussion of two important properties:

- A request is non-prescriptive when it does not specify any explicit information related to the system (or sub-system) fulfilling it, such as resource definition and type, virtualization techniques, allocation, strategies, etc. The intent interface accepts such kind of requests that specifies what are the services requested without detailing how to establish them, leaving this task in charge of the service provider.
- The provider independence allows the consumers of the intent interface to apply the same request to any intent interface provider, without any variation. This property descends directly from non-prescription.

Considering the aforementioned statements regarding intent-based interfaces and characteristics and challenges that are inherent to IBM systems, it is of paramount importance to have a clear view of the requirements that these kinds of interfaces will breed. In [Szi21] the author extensively analyses the motivations and requirements associated to intent-based human-to-network interfaces. Below, a list summarizing its main requirements is given:

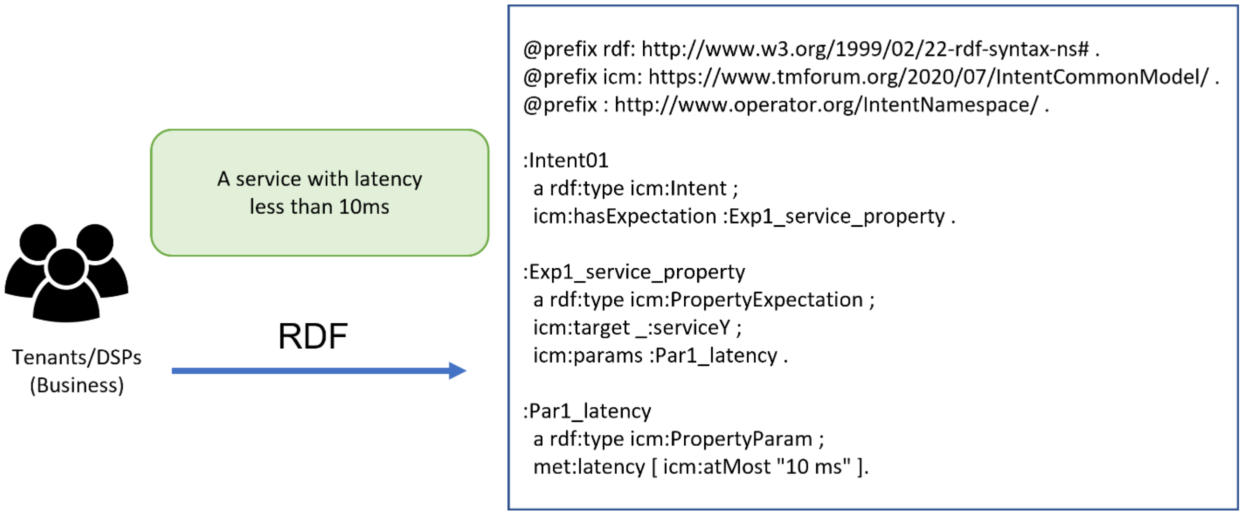
1. **Multi-stakeholder roles support:** Support the standard decomposition of systems and services used by all multi-stakeholder roles (i.e., verticals, CSPs, etc.), as they specify their goals using a language or API that is similar to their own domain abstractions.
2. **High-level interfaces:** Verticals seek to boost income by improving the value of their own businesses and enhancing the productivity of their own technologies. High-level interfaces, between industrial users and telco technologies, are required in order to facilitate the integration of telco technologies without an extensive telco technical expertise.
3. **Bi-directional human-machine interfaces:** Human to machine and machine to human interfaces are required to enable the proper definition of business, service, or resource level objectives as well as to get information on system status, intent fulfilment, and assurance.
4. **Proper feedback:** Intent-based interfaces should be able to provide meaningful feedback, using the same communication method (i.e., natural language, etc.) as the original party that expressed the intent, instead of reporting extremely technical error/alarm logs.
5. **Simplicity, safety & trustworthiness:** The complexity of the system should be abstracted from the intent-based end users giving sufficient security mechanisms so that the management operations can be fully delegated to the system which provides the necessary level of trustworthiness.
6. **Manual override**: The execution of management operations is expected to be delegated to the automated mechanisms behind intent-based interfaces, however, mechanisms of direct control, monitoring, and manual override of the operations should be available through the usage of intent-based interfaces to face unexpected critical events e.g., enabling direct interaction with certain resources and, in the event of unanticipated events, overriding auto-derived contextualized targets.
7. **Automation & fault management:** Depending on the context/state of the intent users, devices, resources, and prospective actions, automatically accomplish the goals stated by the given intent. Harmonize action to achieve numerous objectives (connected to different intentions) in an optimized manner, resolve conflicts and indicate their existence (see requirement 4 proper feedback).

Together with the interface definitions and their properties, it is also important to identify the sources of the managed intent-based objects. The research literature for IBM has considered the description of intent objects in the following areas [ZSM011]:

- The Graphical User Interface (GUI) approach allows users to express their intent by providing a simple template. It is the easiest to adapt to IBM systems since users select options through drop down/drag menus to describe their requirements. Since the GUI is tied to the expertise of the users, these interfaces need to be tailor-made to accommodate the knowledge of the users. For example, a data-center administrator will likely have more control than an average user with simple requirements. Thus, this is a semi-flexible approach for expressing the intent.
- Natural Language Processing (NLP) allows for the conversion of human natural language to machine language [ZSM011]. Moreover, it extracts different context depending on the processing need to be made. For example, it is possible to extract semantic relations between two or more entities in a given text. In IBM systems, NLP provides a chatbot-like conversational interface. However, even though natural language is very flexible, the NLP tools expect to receive a specific grammar format to extract information and create a syntactically correct intent.
- An intent-based language should be readable and abstract the technical details; but it should also be flexible enough to be extended and adjusted according to the business scenario [28.312]. There can be custom languages, such as NEMO, or approaches based on existing data modelling languages. The languages strive to remove ambiguity of intent expression.

The grammar/keyword-based approach follows a verb-object-subject format. The verb can be the operation/action the user wants the network to do, the object specifies the network service, and the subject can identify a service/object (a modifier can parameterize or quantify the subject too). An example is the Resource

Description Framework, which allows to describe sub-graphs (containing expectation, target and conditions [W3C14]) to create an intent. For example, a simple intent with one latency expectation, coming from a tenant, can be defined using the Resource Description Framework as shown in Figure 4-1 below:



```
@prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# .
@prefix icm: https://www.tmforum.org/2020/07/IntentCommonModel/ .
@prefix : http://www.operator.org/IntentNamespace/ .

:Intent01
  a rdf:type icm:Intent ;
  icm:hasExpectation :Exp1_service_property .

:Exp1_service_property
  a rdf:type icm:PropertyExpectation ;
  icm:target _:serviceY ;
  icm:params :Par1_latency .

:Par1_latency
  a rdf:type icm:PropertyParam ;
  met:latency [ icm:atMost "10 ms" ].
```

Figure 4-1: Resource Description Framework example to describe an intent.

Based on all the previously presented introduction on intent architectures and interfaces, one of the main topics within the Hexa-X-II project is the design of an intent-based framework focused on the E2E service management to reach a proper level of autonomy and make the interaction between the system providing services and its clients more agile and faster. To this end, the following sections present the initial set of outcomes from the work done on studying the requirements defined by multiple organizations such as the 3GPP or the ETSI, which resulted on the functional architecture with the essential capabilities to have all the aspects around the intent management under control. Moreover, using the functional architecture, an introduction of a set of possible enablers to manage different intent aspects is presented.

4.1.2 High-level description of the Hexa-X-II intent-based framework

The main outcomes related to the intent-based aspects at the current Hexa-X-II project status are focused on two main points: first, the placement of the intent-based aspects within the project to identify how from an E2E domain point of view, the use of intents relate with other technologies within the same domain and influence to the domains below, and secondly, the main functionalities and the enablers.

While the first topic outcomes are presented in Section 6.1.2.1 with the rest of the E2E system vision of the project, the discussion related to the second point (i.e., internal intent-based framework functionalities and related enablers), produced the following functional architecture illustrated in Figure 4-2 for an intent-based DSM;

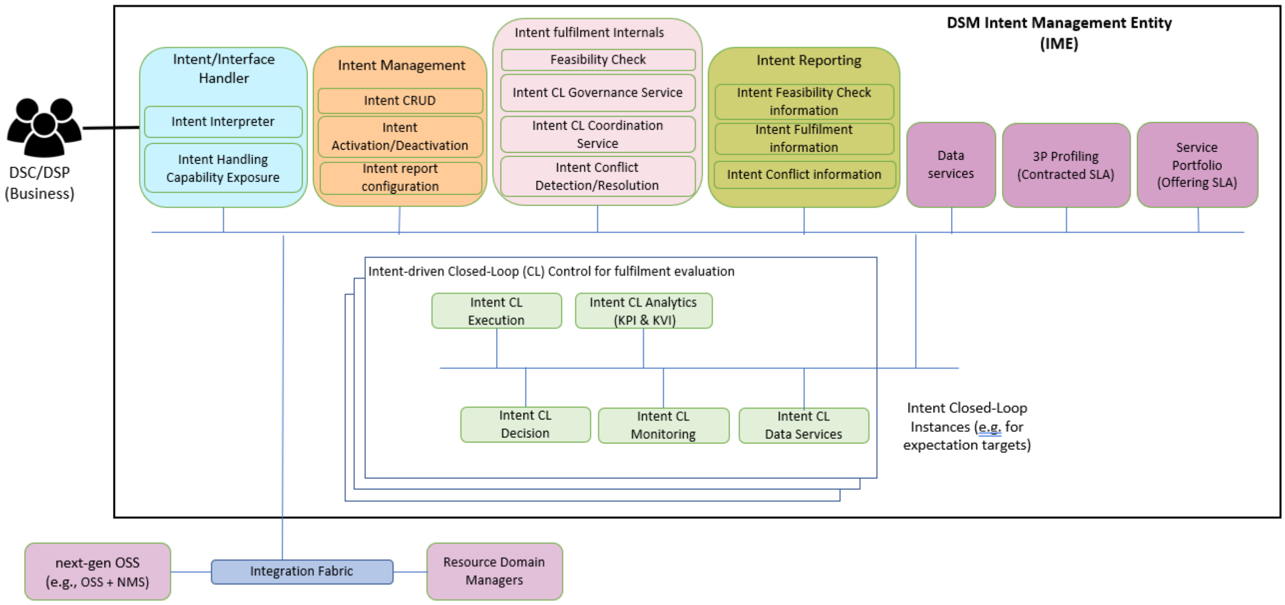

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-2: Hexa-X-II Intent-based DSM IME functional architecture.

This architecture identifies 8 functional blocks listed in the following items; the first 5 items are specially identified for the intent management and, the last 3, to give support on the intent-based management:

1) Intent/Interface Handler: The gateway for the user to interact with the whole intent management solution and trigger those actions available for the user. It offers two main capabilities: the “Intent Interpreter” capability to process the agreement reached between user and the intent-based solution into the right set of data objects to deploy the required services across the system, and the capability called “Intent Handling Capability Exposure” which allows to the system to show to the user what the last one may request to the first one.
2) Intent Management: This functional block is the core element of the architecture. It takes care of the intents data objects with the multiple actions to control the intent’s life cycle. To do so, it uses the CRUD (Create/Read/Updated/Delete) operations (i.e., “Intent CRUD”). Moreover, it triggers the activation and deactivation of an intent to make it work or keep it in standby until the user requires it (i.e., “Intent Activation/Deactivation”). Finally, the “Intent Report Configuration” capability offers the actions related to the expected reporting of intents and, it has a strong relationship with the Intent Reporting functional block.
3) Intent Fulfilment Internals: This functional block is composed of those capabilities that a user should never be able to access but that are key to those capabilities visible by the user. The “Feasibility Check” capability focuses on ensuring that once there is an intent request, this one may be properly applied. The “Intent Closed Loop (CL) Governance Service” takes care to trigger the corresponding Intent-driven CL Control for fulfilment evaluation Instance associated to each requested intent, to ensure the correct deployment and to fulfill the expected requirements. Moreover, the “Intent CL Coordination Service” offers the capability to control in a coordinated way the multiple CLs alive within the intent-based solution. Finally, the “Intent Conflict Detection/Resolution” aims to identify if a new intent or an action over an existing intent may affect negatively to other existing intents and in case there is an intent conflict between two or more intents, to resolve it.
4) Intent Reporting: This functional block offers three capabilities that focus on the multiple types of intent-related information. These are the “Intent Feasibility Check Information” that presents whether it is possible to deploy an intent or not with the current resources, the “Intent Fulfilment Information” such as the intent fulfilment status and the associated target achieved values, and finally, the “Intent Conflict Information” with the associated conflicts and their resolution.
5) Intent-driven CL Control for fulfilment evaluation: This functional block offers the capabilities to manage the life cycle of the Intent CL instances and ensure they are fulfilled at any time. Once an Intent CL instance is activated (using the Intent Management functional block), it has the following capabilities associated: a) the “Intent CL Governance Service” offers the orchestration capability to control any action involving its intent instance, b) the “Intent CL Execution” executes any task

commanded by the governance capability, c) the “Intent CL Analytics (KPI & KVI)” is the capability to process monitored data and generate insights from it, d) the “Intent CL Decision” make use of the insights generated to select the most suitable decision about the intent CL instance status, e) the “Intent CL Monitoring” is the capability to acquire data and metrics related to defined KPIs to be properly analyzed by the “Intent CL Analytics (KPI & KVI)” capability, and finally, f) the “Intent CL Data Services” is the capability to store the information-related to each specific intent CL instance.

6) Data Services: This functional block is in charge to store the intent data objects and other possible information such as SLAs and policies.
7) 3rd Party (3P) Profiling: This functional block allows providing a full characterization of every Hexa-X-II tenant (i.e., a 3rd party) through a 3P profile. This profile captures tenant-specific information on security (supported credentials and access control solutions), trustworthiness (relevant in federation scenarios), contracted services and SLAs, and end-users.
8) Service Portfolio: This functional block is focused on offering the available 6G services information to the tenants, so based on the available services and their information, tenants may request with more knowledge better intent-requests. Among the service information available, there are aspects such as their status (i.e., defined, designed, built, tested, released, etc.), their owner, variabilities over the same service (i.e., SLA offerings, etc.), costs or dependencies with other services, etc.

Finally, all these functional blocks interact with the rest of the E2E management system. Within the Hexa-X-II project, and as presented in section 6.2.2.2, the architecture illustrated will interact with other management and orchestration enablers defined such as the Integration Fabric and others which are part of the OSS and the Resource Domain Managers located in the domains below, which are described in the Hexa-X-II deliverable D6.2 [HEX223-D62].

4.2 Enablers proposed for intent-based management automation

Based on the previous DSM architecture (section 4.1) and its functional blocks, a set of new intent-based enablers was identified. Each of these enablers group a set of the previous functions to reach specific tasks on the management of the intents. For each enabler, a SotA focused on the enabling technologies is presented, a beyond SotA with the high-level description is described and the mapping of the selected functions from the previously introduced architecture is illustrated.

4.2.1 Intent translation and provisioning

This enabler focuses on the development of a micro-service-based solution with the main functionalities to understand, manage and orchestrate an intent request (i.e., receive, translate, deploy) and to generate the specific service requests to deploy services (i.e., network slice, network service, etc.).

4.2.1.1 Enabling technologies (SotA)

The use of intents within the Hexa-X-II project is based on documents defined by ETSI [ZSM011] or 3GPP [28.312] in which they define how the intents need to be managed during their life cycle or possible scenarios where their usage brings more advantages. Based on these references, the “Intent Translation & Provisioning” enabler depends mainly on the following technologies and architectures:

- The Natural Language Processing (NLP) is a branch of machine learning (ML) that studies and processes incoming human requests and translates them to generate machine-specific requests. Using a dictionary model, NLP can identify the words composing an incoming human sentence (i.e., intent) and select those bringing the true and important words with the right meaning that define necessary information to generate the machine-based service request.
- The Zero touch network & Service Management (ZSM) architectures are able to create and manage the actions to achieve the proper close-loop process for each intent. The ZSM architecture defines how the two layers (E2E and specific domains) interact among them to achieve an autonomous resources management. To this end, and in parallel to the other interfaces within the ZSM architecture, ETSI defined an intent interface in [ZSM011] for the exchange of information between the E2E and the specific domains allowing to define roles of intent owner and intent handler. Moreover, in order to

achieve a complete management of intents, the use of the closed-loop concept will allow to have the full and continuous control of a managed object across all the steps of its life cycle.

- The Service Level Agreements (SLA) are the reached outcome after a discussion between a service client and a provider. The SLAs may be created ad-hoc in a real-time interaction between the two interested parties or it may be pre-defined based on what providers may want to offer and so, clients have a closed set of options to choose. The original SLA focus on the service performance (i.e., to maintain a certain capacity threshold), but this concept has evolved on other directions such as security performance (i.e., how the system reacts towards a threat) with Security SLA (SSLA) and trustworthiness performance (i.e., how good is a provider fulfilling expectations) with agreements related to the definition of trustworthiness. The use of these agreements is a key element within the management of intents as at the end, the outcome between an intent owner and an intent handler will be an agreement defining multiple aspects related to service, security and trustworthiness performances.

4.2.1.2 High-level description (Beyond SotA)

This enabler aims to contribute as one of the first intent-based solution to manage service resources located across multiple domains and to select the most suitable combination of them to deploy the E2E service and to achieve the expectations and targets defined by the service requester, while keeping under consideration the specific characteristics of each involved domain. The main objective of this enabler is to reach an agreement with the user and create the final intent data object that it can be translated (considering the service, security and trust levels and the constraints of the resources domains) into a set of machine-based requests depending on the different resource domain managers available behind the integration fabric.

To do so, the use of NLP and SLA (and SSLA and Trust-Level Agreement -TLA-) will be necessary to achieve a conclusion accepted by both intent-based parties (i.e., customer and provider). Once the intent is defined and translated, the enabler will make use of the ZSM entities to manage the different intent-based instances that should result on the E2E services provisioning through the shared Integration Fabric (i.e., an enabler later described in subsection 6.1.2.2.3) across all the resource domains using their managers.

As later presented in subsection 6.2.2.2.4, the intent-based E2E service management automation framework is considering two approaches: a Digital Service Provider (DSP) aggregation and a DSP federation when the collaboration of DSPs is needed. In the second option some service management actions could use the business operations between DSPs. These operations will use an east/west bound interface with federation management and a subset of the northbound interface exposed to the verticals. These are high level operations that need to be combined with detailed operations in the domain management.

4.2.1.3 Analysis of mapping with functional blocks of the architecture

The "Intent Translation and Provisioning" enabler aims to offer the following functions and capabilities defined in the intent-based DSM functional architecture illustrated in subsection 4.1.2. Based on the architecture illustrated in Figure 4-2, the selected functions and capabilities are illustrated (with a red square) in Figure 4-3:

- Intent Interface/Handler:
  - Intent Interpreter to translate the user (human) requests into machine-language data objects towards the final agreement between the user and the enabler.
  - Intent Handling Capability Exposure: Together with the Intent Interpreter, it must assist the user on showing the services offered by the enabler towards the user, so the user may select the best action at any moment.
- Intent Management:
  - Intent CRUD contains the set of the main and public actions to manage intents such as Create, Read, Update and Delete actions.
  - Intent Activation/Deactivation to control the status of the intents objects that are feasible, deployed, and ready to be used by the user.
- Intent Fulfilment Internals:

  - Feasibility Check to assist on reaching an agreement between user and system, this capability should check that there are the resources and elements required to achieve what the user is asking.
  - Intent Conflict Detection/Resolution to validate if once the user and system reach an agreement, the defined intent does not generate any conflict with other existing deployed intents. In case of a conflict, a resolution method should be applied to solve the situation.

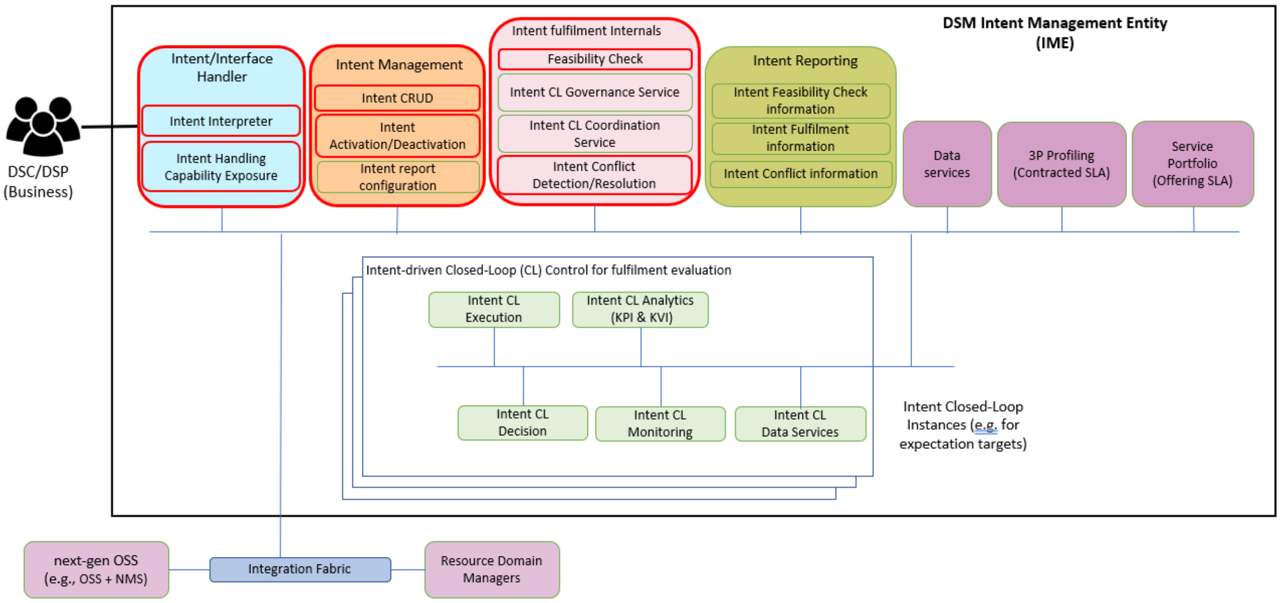

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-3: Intent translation and provisioning mapping to functional blocks of the DSM IME.

4.2.2 Data fusion mechanisms based on telemetry data

One of the biggest obstacles in effectively managing distributed services and applications while guaranteeing the satisfaction of the corresponding user intents is the need to efficiently monitor the different aspects of their deployment, which in turn can be expressed in terms of the achieved measurable values that can validate the intent fulfilment. Thus, the transition from traditional monitoring tools to modern cloud-native and network observability tools becomes crucial for providing a fine-grained view of application/infrastructure/network performance. Traditional tools are not tailored to distributed microservice environments and interactions among containers, as they are typically geared towards monolithic applications. Additionally, while tracing techniques have long been employed by developers to track an application’s behaviour-related metrics, they are not well-suited for microservices-based applications, and do not account for the horizontal scaling abilities of these applications. The concept of cloud-native observability has emerged as a means of providing insight into the health and status of applications within cloud-native elements like microservices, containers, and orchestration tools [HEX223-D21]. Cloud-native observability of application/network functions components has also to be combined with data coming from network telemetry mechanisms, enabling the examination of both application and network-oriented metrics and the identification of the main causes of faults or delays. This enabler aims to provide a methodology as well as an implementation of such a tool that will support a fully observable view of all application, cloud and network resources involved, as well as their interactions. The work and methodology for its functionality is closely related and complementary to the functionality of enabler “programmable network monitoring and telemetry” [HEX223-D62].

4.2.2.1 Enabling technologies (SotA)

Multiple considerations are taken into account such as various observability signals and the need for data fusion of the collected information, providing a unified state of function performance, facilitating the analysis of otherwise sparse, heterogeneous data sources. The signals may refer to compute resource usage metrics (e.g., CPU, memory), QoS metrics, application-specific metrics, software traces and logs. It should be noted that modern approaches to observability need to take into account the necessity of integrating with existing or emerging monitoring tools [TAZ+23]. The existing monitoring mechanisms in open-source orchestration engines are not designed to support advanced monitoring that caters to the unique characteristics of distributed applications and the observation of metrics related to interactions among application components. Additionally, multiple third-party tools exist to support distributed tracing and logging mechanisms [TAZ+23], albeit with limited integration with the aforementioned monitoring tools. Over the collected data from modern observability tools, various analysis pipelines can be executed, including the pipelines that are based on machine learning techniques.

An efficient and flexible observability data collection methodology is crucial for correlating signals of different nature. OpenTelemetry [OTL23] provides a structured framework for the collection of different types of signals that can be exploited for this purpose. Not only it provides a series of code instrumentation libraries, but also defines a standard which can be adopted for building mechanisms on top of it. Additionally, the OTLP (OpenTelemetry Protocol) Collector is provided for receiving, processing and exporting signals, while different deployment scenarios are available to fit the corresponding architecture.

Distributed applications deployed across the continuum include a lot of different observed resources and thus, can generate a huge list of different signals. Identifying issues in such complex sequences of datasets is next to impossible when faced by developers or infrastructure/network operators. A methodology for automating such processes is needed and this can include multiple approaches.

One such approach is the identification of anomalies in the operation of the deployed services. Techniques for exploiting all kinds of signals can be found [SB22]. For example, [NMA+16], [JCY+17], and [JYC+17] utilize unsupervised learning techniques for identifying anomalies outside of the distributed service graph’s normal operation zone. Similar methods can be found for the analysis of metrics which are studied as time series, so the time series analysis field provides a variety of techniques to tackle the problem. Supervised and unsupervised learning are very popular [MMP+18], [WTE+20], while SLO threshold checks are also used in various works such as [GLC19]. Similar work is found for distributed tracing analysis based on unsupervised [NCK19] and supervised learning [NCK19], [GZH+19] techniques, while trace comparison methods with pre-production measured traces are also considered [MJS+21]. Another approach for automating the reliable operation and management of distributed services in the continuum, is root cause analysis. Work in this direction can be found based on monitoring, logging and distributed tracing [SA22]. Monitoring-based and tracing-based cause analysis explore statistical analysis methods [MWZ+13], [WZC+20] as well as topology graph-based ones [WTE+20], [KSS13], while causality graphs are built and analysed for logging and monitoring-based methods [WTE+20], [LCZ18]. Especially for distributed tracing, visualization methods are also used for guiding manual resolution [GPW+20].

While a lot of work can be found on the individual signal analysis, there is not a lot of work when considering different kinds of signals and in the cases where two different signal types are used, the goal is quite specific. Additionally, recent work focused on specific domains, not considering application and network observability in a joint manner. The embedding of heterogeneous signals from different resources to a common space, taking into account the interactions between them, is crucial for the identification of hidden patterns in the data and for understanding the end-to-end flow characteristics that may influence application performance and functionality in the continuum.

4.2.2.2 High-level description (Beyond SotA)

The enabler aims to provide a fused viewpoint of the 6G services operation when deployed and managed in the 6G infrastructure. For this purpose, a series of components are considered that will allow the proper collection and analysis of such heterogeneous data as well as a set of interfaces to allow the broadcasting of the results. These are shown in Figure 4-4.

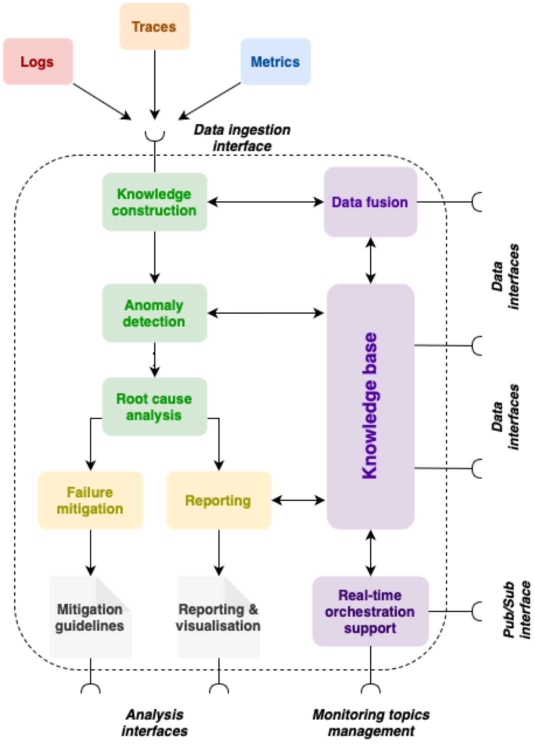


Figure 4-4: Observability data fusion architecture.

The *Data ingestion interface* will be built to provide an entry point of the data to the system. This will align with the OpenTelemetry standards and will make use of the OTLP protocol to work in synergy with the OTLP Collector that may be used interchangeably with any other third-party tools that may also be considered. Specifically, the OTLP collector is provided both as a centralized gateway, collecting signals from different locations, and as an agent collecting signals from the local deployments and forwarding them at a later time, in an aggregated manner to a centralized point. Thus, the exporters from application, infrastructure and network resources push observability data either to third-party tools or to OTLP Collectors. In both cases, data will be collected by the tool. The different data collection architectures considering integration or not with the OTLP Collector and the utilization (or not) of third-party tools are briefly shown in Figure 4-5.

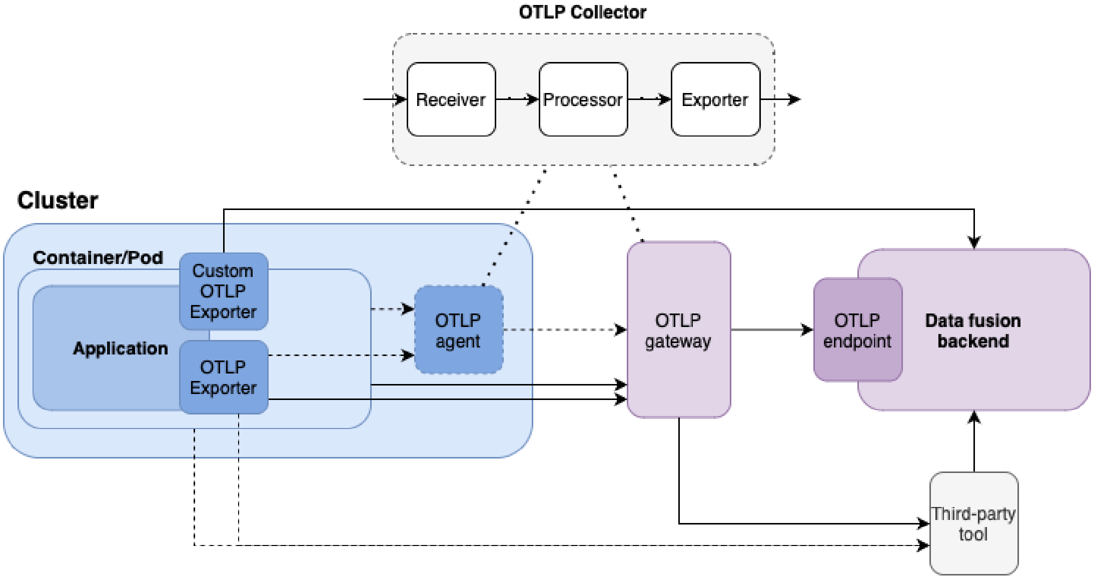


Figure 4-5: Data collection using Open Telemetry.

The complex relations between application, cloud and network infrastructure can create a big overhead when analyzing possible failures and thus, it is important to properly capture the different elements constituting the operation environment as well as the interlinking among the denoted entities. For this purpose, an analytical data model will be considered to collect the necessary information and continuously build a *Knowledge Base* from the collected signals (Figure 4-6). Application, as well as computing and network infrastructure details will be covered by the model, while information from all types of signals (e.g., application and QoS metrics, traces, logs) should be sufficiently expressed.

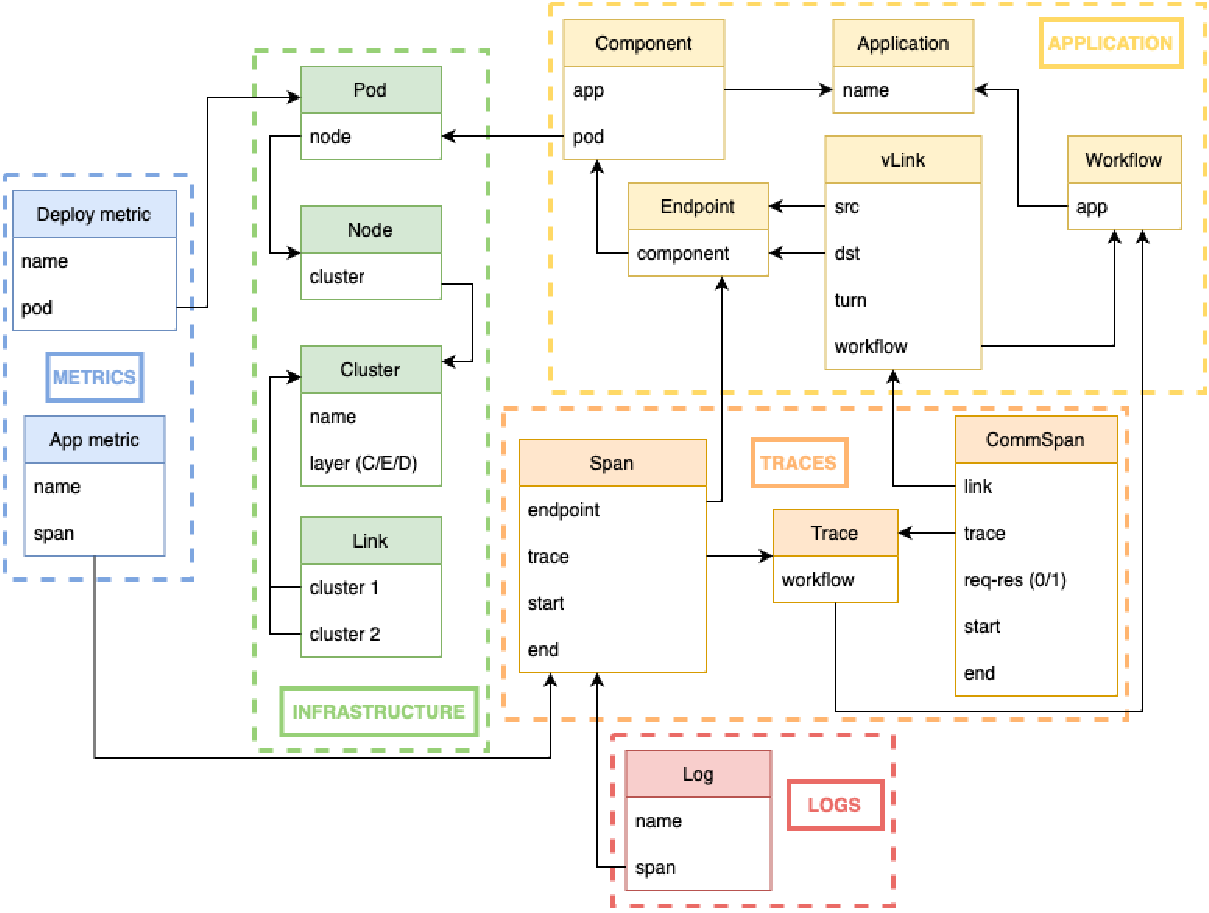


Figure 4-6: Knowledge base data model.

Taking into consideration interoperability with external tools, the fused information as well as the raw data will be available through a series of *Data interfaces*, while real-time data will be broadcasted using *pub/sub* mechanisms to also provide *Real-time orchestration support*.

Finally, an analysis pipeline of the heterogeneous signals will be built to take advantage of the collected interlinked data and provide intuitive guidelines regarding the identification of present or future failures as well as their mitigation. For this purpose, the specifically designed knowledge base is going to be exploited in order to build efficient data structures that will facilitate the complex analysis of the interlinked data. *Anomaly detection* will handle the identification of failures in the operation, root cause analysis will find the relevant resources that caused the failure, while *Mitigation* will provide a resolution if possible. A series of corresponding interfaces will allow external entities to access these results in descriptive, reporting or visualization forms.

4.2.2.3 Analysis of mapping with functional blocks of the architecture

This enabler considers covering the Intent CL Monitoring / Decision / Execution / Analytics blocks of the architecture as the monitored parts of the continuum will provide different signals, which will be analysed in correlation to the corresponding SLOs, while mitigation plans will also be extracted from the analysis after fault identification for facilitating decision making and resolution execution. Additionally, the enabler will connect with the Intent Fulfilment Information block to contribute to the calculation of the *achieved values* of the intent (Figure 4-7).

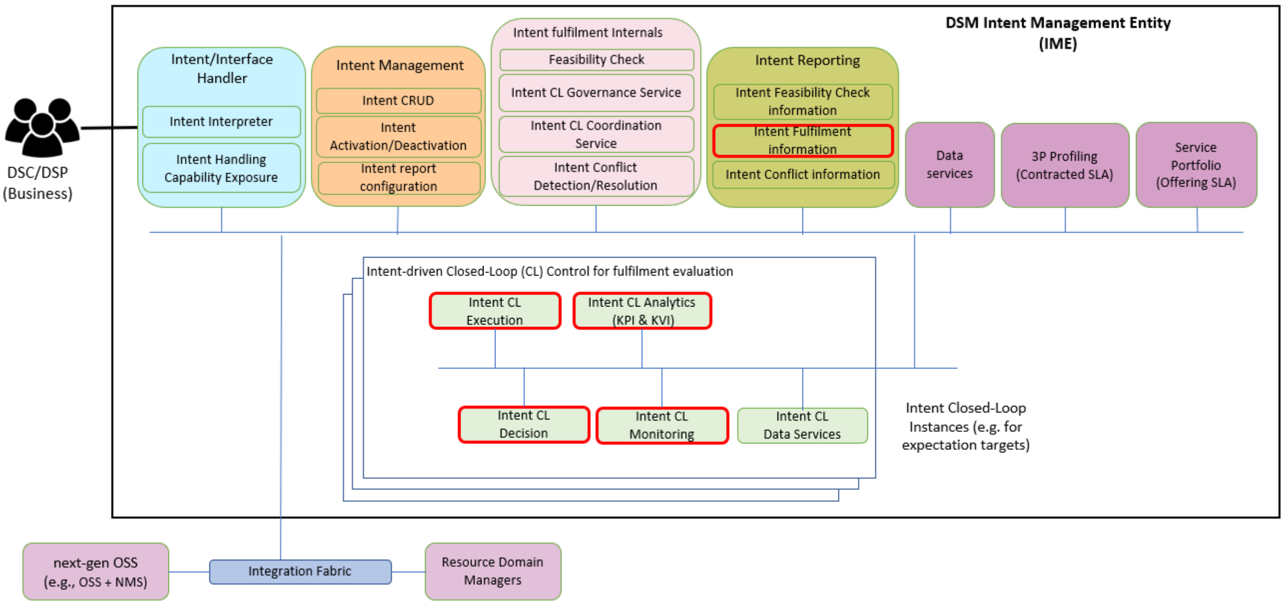

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-7 - Data fusion mechanism mapping to functional blocks of the DSM IME.

4.2.3 Closed loop coordination for intent management

4.2.3.1 Enabling technologies (SotA)

6G systems are characterized by a high-level of automation that requires the usage of control closed-loops at all levels, including the intent management. ETSI ZSM in [ZSM009-3], [ZSM011] consider the management of the intents as intrinsically based on CL automation, since the IME (see section 4.1.2) is able to figure out if an intent has been fulfilled and in the negative case, perform corrective actions, if possible. It is important to note that an IME manages the lifecycle of several intents and this alone already implies the presence of multiple-closed-loops to be managed. Another important aspect is related to its fulfilment i.e., the corrective action to be executed to move the underlying system towards a desired state: it can happen by exploiting conventional management interfaces or, by triggering a new subordinate intent (hence, a new CL) with certain requirements.

In [ZSM009-1], ETSI discusses the application of CLs in ZSM architecture and poses the basis for their automatic management, introducing the concepts of CL Governance and Coordination. The Governance provides a set of functionalities to directly interact with the CL and its components for e.g., loop start, stop, removal, get status information, etc. The CL Coordination (CLC) is the set of capabilities that allow the coexistence of multiple closed loops and their interaction to maximise their effect while avoiding possible conflicts. The use cases described are related to closed loops executing within the ZSM framework in a generic manner, i.e., the coordination methodology can be applied to any loop careless of its own nature e.g., service, network, intents, etc.

In hierarchical scenarios, the subordinate loops can enforce conflicting optimizations or local optimization that might not lead to an E2E optimum. In this case, the coordination with loops "1", that are the E2E one, can exploit two different techniques, alone or in combination (hybrid operational mode): Delegation and Escalation. With Delegation, the E2E CL properly configure the local CL to make it autonomous to fulfill the goal(s) of both CLs. On the other way around, with the Escalation, when a local CL is not able to achieve its own the goal(s), It can escalate to the E2E CL.

In peer scenarios the loops run at the same system level (e.g., insist on same network resources) and may benefit from sharing loop information reports. Loops can exchange goals, models, health status, values of CL attributes, and goal fulfilment information. Peer CLs may perform local operations with the risk of conflicting operations done by another peer. To resolve conflicts, peer CLs may cooperate to align their intents and/or policies, which may also utilize the impact assessment service. If multiple CLs are acting on the same managed entity, concurrency coordination is needed, e.g., race condition avoidance. This may be done for example by comparing the closedLoopPriority attribute. In peer scenarios, a CL may also request the resolution of an issue from another peer CL.

Recent academic research includes various approaches and methodologies of CL coordination, ranging from methodologies for coordinated and optimal instantiation and operation of multiple closed loops within the above described ETSI ZSM framework [GBH+21] [XGH+20] to independent methods such as using multi-agent reinforcement learning agents to implicitly incentivize loops to cooperate without human interaction by prioritizing selected KPIs [PMD+22].

4.2.3.2 High-level description (Beyond SotA)

This enabler aims at investigating and defining an IME integrated solution for the automatic coordination of multiple CL generated for the lifecycle management of the intents. Furthermore, the main idea is also to explore the possibility of interacting with other CL Coordination functions outside the IME scope. As discussed in the SotA, Intent CL can be peers (e.g., two or more intent executing in parallel) or part of a hierarchy (an intent whose fulfilment implies the execution of another intent). This opens for different forms of coordination mechanisms such as Cooperation (for peers), Escalation and Delegation (for hierarchical closed loops).

Conflict detection and resolution is one of the crucial features characterizing the CLC. Nevertheless, in the architecture proposed in Figure 4-8 both of them are delegated to the Intent Conflict Administration (ICA) enabler, discussed in Section 4.2.4, while CLC maintains the logic required to enforce policies and configurations for the resolution/mitigation of the conflicts, following the instruction provided by the ICA.

Conflicts can be detected at i) Intent creation and/or ii) Intent runtime (when the closed-loop is in the Execution state).

In i) the ICA can decide to reject the intent or invoke the conflict resolution, providing the CLC with specific instructions: the CLC executes in turns specific actions (e.g., request a new scheduling) to mitigate the conflict, if possible. In ii) CL concurrency issues may occur. In this case, the CLC needs to interact with the CL Governance(s) managing the potentially conflicting CLs, and tuning their execution (e.g., giving priority to a given CL by starting/stopping other CLs). Unfortunately, manipulating the execution of a CL may have a negative impact on CL actions. The validation of such impact is in charge of ICA, while the application of specific policies to minimize the degradation is a task of CLC.

Another topic that this enabler will investigate, is the possibility to go beyond the Escalation and Delegation strategies for coordinating hierarchical loops. In the Escalation, a subordinate CL escalates to a superior one where it is not able to achieve its goal. Conversely, a superior CL delegates its own goal(s) to a subordinate one. This enabler aims to define a mechanism to coordinate different instances of CLC belonging to different domains.

4.2.3.3 Analysis of mapping with functional blocks of the architecture

Intent CLC is a service belonging to the Intent Fulfilment Internals group, according to the IME architecture defined in Section 4.1.2. Several interactions with other components of such are foreseen:

- Intent CL Governance
- Intent Activation/Deactivation
- Intent Conflict Information
- Intent Conflict Detection

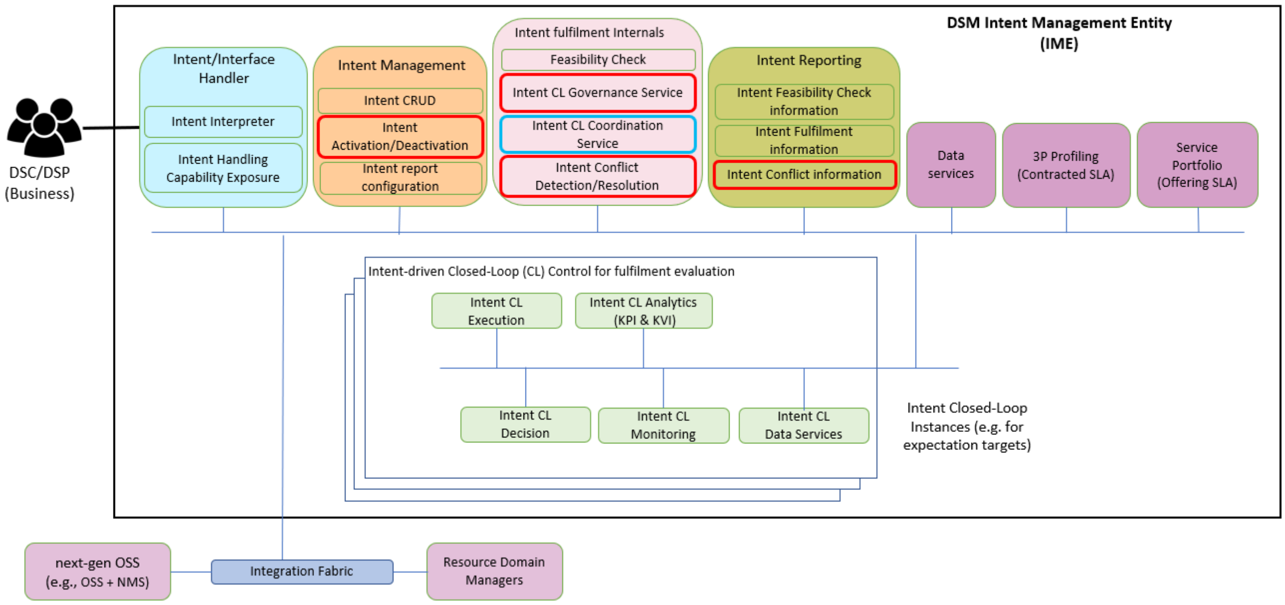

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-8: Enabler “closed loop coordination for intent management” interactions within DSM IME

4.2.4 Intent conflict administration

4.2.4.1 Enabling technologies (SotA)

An enabler combined with algorithms and entities to avoid conflicts is one of the essential components in autonomous networks management. Therefore, accurate, scalable and fast conflict administration is critical to meet the expressed requirements from intents. According to [ZSM011], based on the level of abstraction, we can categorize intent conflicts into 3 groups:

1. Syntax-level conflict: This type of conflict compromises conflicts among various KPIs or components within an intent expression.

2. Action-level conflict: This type of conflicts arises if the actions of different intents cannot be performed concurrently.

3. Impact-level conflict: In this case, the actions can be executed simultaneously, however they can cause different negative impacts.

Considering different conflict levels, the first way is to ensure that negative impacts of actions cannot happen by designing a system in a way that KPIs do not interfere. A second way is to detect conflicting effects at deployment time. The third way to handle the conflicts is at the run-time. There exists work that considers this approach for conflict detection [BMC21, BMC21a], however the focus is on managing conflicts in a single intent. The proposed architecture in [BJZ+22] can detect and resolve conflicts arising among multiple intents at runtime in a scalable manner. In [ZLF22], the main goal is to create algorithms to identify conflicts and resolve policy issues to avoid conflicts for active intents. In case of arrival of a new intent, the proposed framework is capable of checking, identification, and resolution of the conflicts.

4.2.4.2 High-level description (Beyond SotA)

In autonomous and future network systems that are designed to support multiple tenants and use cases numerous intents will co-exist. According to the MAPE-K (Monitor, Analyse, Plan, Execute, and Knowledge) framework for building autonomic and self-adaptive systems, these systems can employ closed loops composed of different components to fulfil multiple intents. To achieve this vision, a closed loop can be established for fulling expectations associated with an intent, wherein each expectation represents a specific target KPI to be addressed. However, due to limited and shared nature of the network resources, the closed loops may influence one another. For example, one closed loop that is responsible for fulfilling a particular target KPI can have positive impact on other active KPIs coming from multiple intents whereas, there may be situations where one KPI negatively impacts other present KPIs. In this respect, when it comes to real system deployments, managing conflicts among different closed loops can be challenging. Such scenarios with closed loops competing for a set of limited resources in a shared network can lead to unstable states. Hence, conflict management and administration between closed loops is an essential component of network automation.

The problem we want to solve by the proposed enabler can be tackled in different ways. In general, there are three ways to solve this problem.

- A first way is to ensure that negative side effects cannot happen. This can be done by designing the system such that KPIs do not interfere. This approach is possible, but easily becomes restrictive. It also puts the burden on the designer to be aware of all relations between KPIs.

- A second way to avoid negative side effects is to detect them at deployment time. This means that the system would analyse the feasibility of each new intent. When a new intent is received, the system would check if these new expectations may interfere with already existing expectations. The system would detect a potential interference and reject the new intent in advance. The detection of possible interference would be based on a model of the environment under control, a model of all possible actions that may be taken for the new expectations, and a model of the results of these actions. These could then be correlated to possible actions for existing expectations. If there is an overlap, there is a risk that interference may happen. In that case the new intent is rejected. This approach is possible, but with the disadvantage that it rejects too many intents. Even though there is a potential interference, it cannot be known at deployment time if this interference really will happen at run-time.

- A third way to avoid negative side effects is to detect conflicts at run-time and to solve the situation when it happens.

It should be noted that the three alternatives mentioned above are not mutually exclusive. A system may for example partly implement the first and second way to avoid the most obvious conflicts and use the third way for optimisation.

4.2.4.3 Analysis of mapping with functional blocks of the architecture

The Intent Conflict Administration considers two components such as conflict detection and resolution. This means that the enabler maps to the “Intent fulfilment internals”, the “Intent-driven Closed Loop Control”, and slightly to the Intent Reporting. For the first block, “Intent fulfilment internals” the Intent Conflict detection and Intent CL Coordination service are used to give assurance that in case of a conflict, the Intent-based digital service manager is able resolve the conflict without creating additional conflicts with existing intents. Thus, for each intent instance there will be a closed loop. By having closed-loop(s) per intent, the Intent Conflict Administration enabler focuses on the Action-level conflict and Impact-level conflict which are captured by the “Intent-driven Closed Loop Control”. Once a conflict has been detected, independently if it comes from an action or is explicit, and/or is being addressed (as some time might pass before the conflict is solved), the tenant has to know about this conflict; thus, the enabler also maps to the “Intent Conflict Information” block from the Intent Reporting of the functional architecture. For Syntax-level conflicts, we consider other functional blocks that are out of the scope of the Intent Conflict Administration. Figure 4-9 maps the components of the functional blocks of architecture to the Intent Conflict Administration enabler.

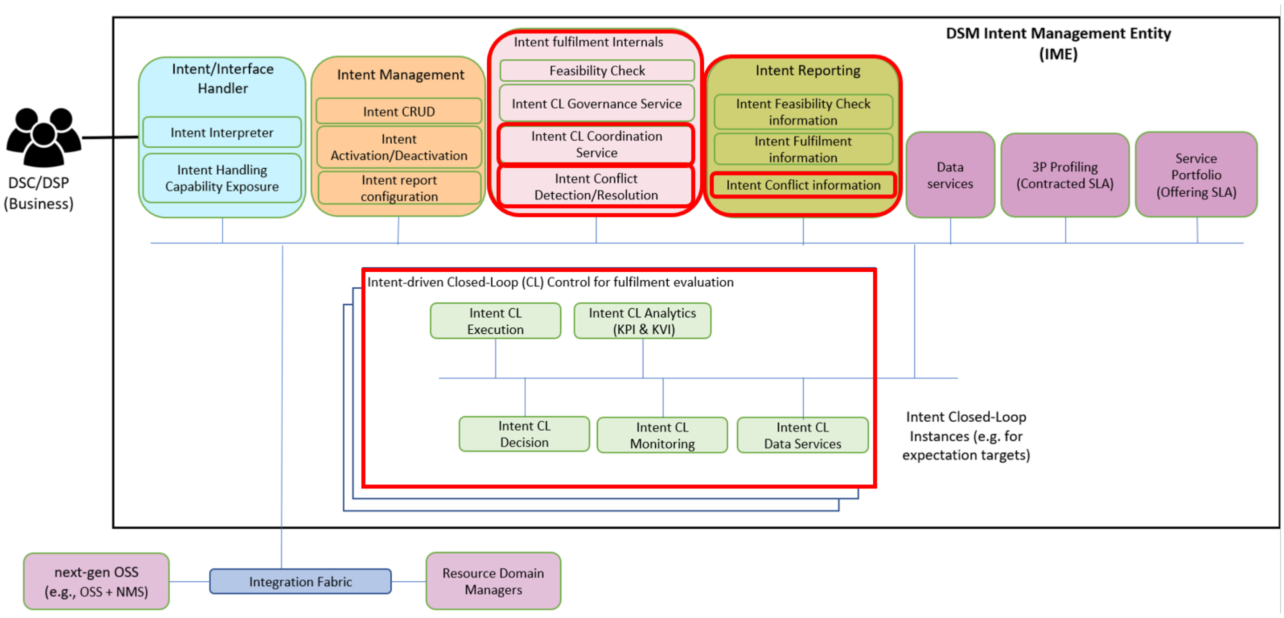

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-9: Intent Conflict Administration mapped to DSM IME.

4.2.5 Human-machine intent interface design

6G services are being consumed by increasingly different types of applications, each with their specialized requirements and domain-specific context. An application is defined as an external entity to a network that consumes network services and communicates needs, requirements, configurations, and status through appropriate interfaces. This ranges from vertical industrial applications, e.g., extended reality, vehicle-to-everything (V2X), as well as over-the-top (OTT) mobile applications.

To communicate with the network, applications requirements need to be expressed in a format that reflects the "what" rather than the "how".

4.2.5.1 Enabling technologies (SotA)

As mentioned earlier in section 4.1.1, requirements and intents can be expressed, currently, using a GUI, NLP, an IBN language, or a grammar/keyword-based approach. These approaches limit the users and applications to express intents not supported as an option, in case of a GUI. They require a specific grammar format, in case of NLP, and others. Moreover, and most importantly, they expect a network domain language to be used when expressing the intents which assumes an in-depth telco knowledge of the users. High-level description (Beyond SotA)

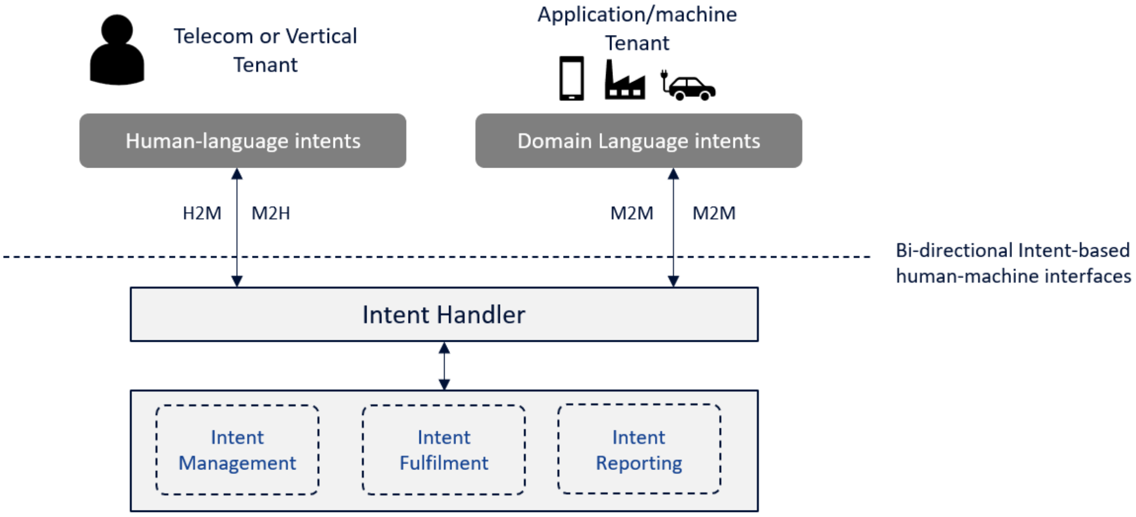


Figure 4-10: Human-Machine intent interface

There exists a trade-off between expressing the application requirements in network or application domain language. The former allows for easier understanding and processing at the network, whereas the latter allows for more accurate expression of the application's real needs.

Further, application's requirements are expressed to the network from a tenant that is either a human operator or a machine (e.g., a running application code). Feedback from the network is then sent back to the human or machine operator. This feedback includes configuration and fulfilment status, as well as potential actions based on these status data.

These points motivate the need for human-machine interfaces (HMI) that enable efficient expression of intents from applications and exposing network data to applications. An overarching goal of an intent-based HMI is to bridge the gap between humans, machines, and networks, facilitating seamless bi-directional communication and interaction between users and applications on the one side and networks on the other. It aims to provide a way for both humans and applications to articulate and contextualize their network requirements in their respective domain languages, while maintaining meaningful and actionable feedback from the network, as illustrated in Figure 4-10.

In order to realize the intent-based HMI concept, key features and requirements are needed, as follows:

- Bi-directional Interfaces:
  A key feature of an intent-based HMI is establishing bi-directionality. It enables both expression of needs and requirements to networks, and system insights, feedback and actionable items back to applications. Interfaces should enable dynamic interactions where networks can adapt to human and application inputs and provide timely and meaningful feedback. The bi-directional interfaces are categorized as follows:
  - H2M (Human to Machine): Targets to enable the definition of business, service, resource level, or management objectives by a human operator.
  - M2H (Machine to Human): Provides insights on system state, fulfilment, and assurance to human operators. Signal and resolve conflicts.
  - M2M (Machine to Machine): Targets allowing applications to express their needs in their domain language, not necessarily in a native network language, as well as providing insights and actionable feedback to applications.
- Requirements for Verticals and Enterprise:
  - Providing intent-based interfaces in application's language is of utmost importance to verticals and Enterprise applications, particularly, with emphasis on industrial networks and private networks.
  - Aims to increase revenue by adding value to businesses and enhancing production efficiency.

  - Emphasizes the need for clear interfaces between industry applications and telco technology to support telco integration without in-depth telco knowledge.
- Experts in Different Roles at CSP
  - Intent-based HMI design should involve planning, OSS, BSS.
  - Experts should be able to define specific intents supporting their goals of a certain domain abstraction.

4.2.5.2 Analysis of mapping with functional blocks of the architecture

As shown in Figure 4-11, this enabler is mapped to the following functional blocks:

Interface handler:

- Intent-based HMIs affect how intents are ingested and handled at the DSP. Interface types (H2M, M2H, and M2M) affect what kind of content and syntax is exchanged between the operator and the network.

Intent interpretation:

- Interpreting intents should be expanded to allow for specialized intents that are tailored towards particular types of applications.
- Methods that efficiently translate application intents, map it to the right network objectives, and provide actionable feedback to the applications are needed.

Intent CL execution, analytics, monitoring, and decision:

- CL handling blocks need to take actions that are aligned with application's objectives, expressed through intent-based HMIs. In this regard, appropriate translation of application-level quality metrics to network QoS is needed. Further, monitoring analytics functions need to keep track of the validity of the application's quality metrics with respect to the offered network KPIs. In case of deviation from the application-level goals, mechanisms to report actionable feedback to applications are needed. Analytics should be able to differentiate between deviation due to inaccurate application requirement translation and that due to insufficient network capabilities and execute the right management decisions accordingly.

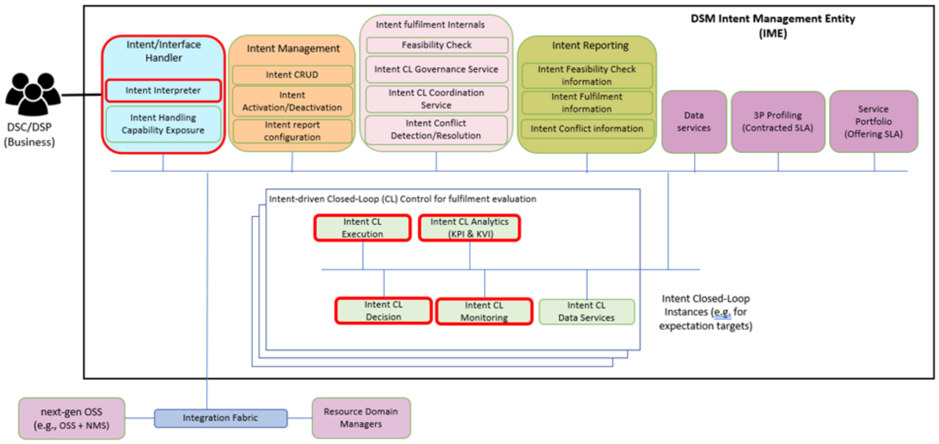

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-11 Human-Machine Intent Interface design mapping to functional blocks of the DSM IME

4.2.6 Intent-driven placement

4.2.6.1 Enabling technologies (SotA)

With the advent of both Network Virtualization (relying on Network Function Virtualization and Software-Defined Networking) and Edge Computing ([MEC21]), 5G has leveraged the joint orchestration of network and compute resources, both for network management and operation (Virtual Network Function Orchestration) and for the realization of services with advanced requirements. However as pointed in [NGA22], “the communication and computing aspects are still decoupled, and it is not expected that holistic integration of ICT and IT domains will be supported in 5G releases. Such integration will require reconsidering the boundaries of the network to include the edge and the cloud domains and to allow for joint optimizations across multiple domains belonging to different stakeholders”. It is expected that 6G will change this, “[enabling] a large-scale distributed cloud in heterogenous and ubiquitous computing environment, and the incorporation of device compute”, in essence changing the focus of telecommunication networks “from communication-centric to becoming communication-computing-data centric” [Eri23].

In an intent-based approach, this means that considered overarching intents may not only include *what* to execute, but also explicitly or implicitly (e.g., through performance requirements) indicate *where* to execute services compute elements (e.g., containers) along a wider Compute Continuum including execution domains ranging from extreme edge devices to edge cloud to central cloud. Challenges that arise when dealing with containerized orchestration of different types of applications in such heterogeneous infrastructures, each with their own characteristics and policies or even belonging different authorities, have been identified in the literature. In particular, the inclusion of the end-devices as available compute resources as extreme-edge domain poses extra challenges regarding its very volatile nature. To address those challenges [*KLM+22*] proposes to use a multi-agent approach to compute continuum orchestration that is clearly relevant, in particular to represent execution domains.

4.2.6.2 High-level description (Beyond SotA)

The goal of this enabler is to study and propose intent description extensions and associated analysis and decision mechanisms to steer high level (intra-DSP/cross-DSP) compute placement and associated necessary network configuration across the whole compute continuum.

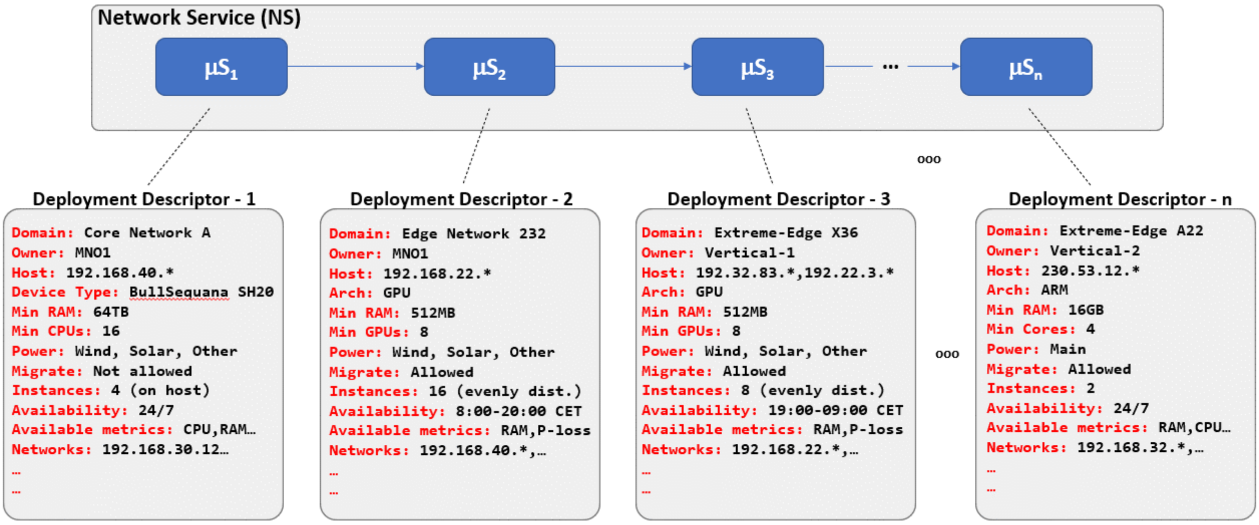
- Extensions to intent description may be explicit references to target execution domains (e.g., requesting execution on this extreme edge device or on the edge cloud serving the latter)
- In some cases, the target execution domain may have to be derived from broader intent description requirements, e.g., requesting execution on resources trusted by or belonging to a given actor, or more indirectly requesting an SLA that implies the selection of certain execution domains and specific networking configuration.
- Whether explicit or derived from the intent description, the Intent-driven placement enabler will be in charge of deriving the execution domain (agent) to contact and request orchestration from, while maintaining a closed-loop in charge of reacting to changes (e.g., through the Data fusion mechanisms based on telemetry data enabler or other sources of context information) to adjust placement according to the intent expression.

As an initial high-level approach, Figure 4-12 showcases how, from the perspective of the operations teams, the deployment of a network service composed of different microservices (from μS_1 to μS_n) could be requested using a declarative intent-based approach in the deployment descriptors of the microservices. As it can be seen in this illustrative example, for each microservice, the deployment descriptors would define the target features of the infrastructure components on which these microservices are required to be deployed, including a list of parameters that define the features of the required nodes, such as the number and type of CPUs, available RAM, a range of IP addresses, the networks to which the device should be connected, the stakeholder to which the node belongs (e.g., an MNO, a hyperscaler, an industry, etc.), the network domain to which it is associated (cloud, edge, or extreme-edge), geographical information, etc[3].

The definition of this set of generic parameters for the different microservices would be provided in a declarative intent-based way, focusing on the desired final result regarding the deployment from the perspective of the operations teams (or other stakeholders), and not in an imperative manner, stating "how" the orchestration system would deploy the service (the step-by-step procedure to achieve the required end state). The data model for defining the descriptors would allow to deploy the micro-services on different kinds of

[3] This list and those parameters in the figure are just to illustrate this initial conceptual approach. A more complete and consistent definition of the specific list of parameters is planned to be provided in Deliverable D6.3.

target nodes, with the required computing architectures, network domains, kind of power supplies, etc. Of course, a mechanism in charge of processing these high-level declarations in the descriptors should exist, as well as to select the appropriate nodes in the available infrastructure pool to perform the deployment of each microservice during the network service provisioning. Additionally, once the network service is provisioned, associated closed-loop assurance mechanisms should be continuously executed to keep the deployment consistent with the intent declarations in the deployment descriptors. This concept is planned to be further developed regarding the definition of the so-called Deployment Node (DN) component, already introduced in deliverable D3.2 [HEX223-D32], which would be in charge of processing and executing those intent-driven deployment requests. In alignment with the intent-based framework (see section 4.1.2) such component would be deployed as part of the "next generation OSS" block.



Deployment Descriptor - 1

```
Domain: Core Network A
Owner: MNO1
Host: 192.168.40.*
Device Type: BullSequana SH20
Min RAM: 64TB
Min CPUs: 16
Power: Wind, Solar, Other
Migrate: Not allowed
Instances: 4 (on host)
Availability: 24/7
Available metrics: CPU,RAM…
Networks: 192.168.30.12…
…
…
```

Deployment Descriptor - 2

```
Domain: Edge Network 232
Owner: MNO1
Host: 192.168.22.*
Arch: GPU
Min RAM: 512MB
Min GPUs: 8
Power: Wind, Solar, Other
Migrate: Allowed
Instances: 16 (evenly dist.)
Availability: 8:00-20:00 CET
Available metrics: RAM,P-loss
Networks: 192.168.40.*,…
…
…
```

Deployment Descriptor - 3

```
Domain: Extreme-Edge X36
Owner: Vertical-1
Host: 192.32.83.*,192.22.3.*
Arch: GPU
Min RAM: 512MB
Min GPUs: 8
Power: Wind, Solar, Other
Migrate: Allowed
Instances: 8 (evenly dist.)
Availability: 19:00-09:00 CET
Available metrics: RAM,P-loss
Networks: 192.168.22.*,…
…
…
```

Deployment Descriptor - n

```
Domain: Extreme-Edge A22
Owner: Vertical-2
Host: 230.53.12.*
Arch: ARM
Min RAM: 16GB
Min Cores: 4
Power: Main
Migrate: Allowed
Instances: 2
Availability: 24/7
Available metrics: RAM,CPU…
Networks: 192.168.32.*,…
…
…
```

Figure 4-12: Example of intent-driven deployment of a network service on a multi-domain infrastructure.

4.2.6.3 Analysis of mapping with functional blocks of the architecture

As depicted in Figure 4-13, work on this enabler will focus on impacts to the Intent Interpreter (and Intent Handling Capability Exposure) to provide (and advertise) the support mechanisms translating placement-related intent parameters into actual compute placement-oriented control loops, and on studying these placement-oriented control loops (and impacts to execution, analytics, monitoring and decision functions of the latter to support placement adaptation to contextual changes). Links to Intent Fulfilment and Intent Reporting functions will also be considered.

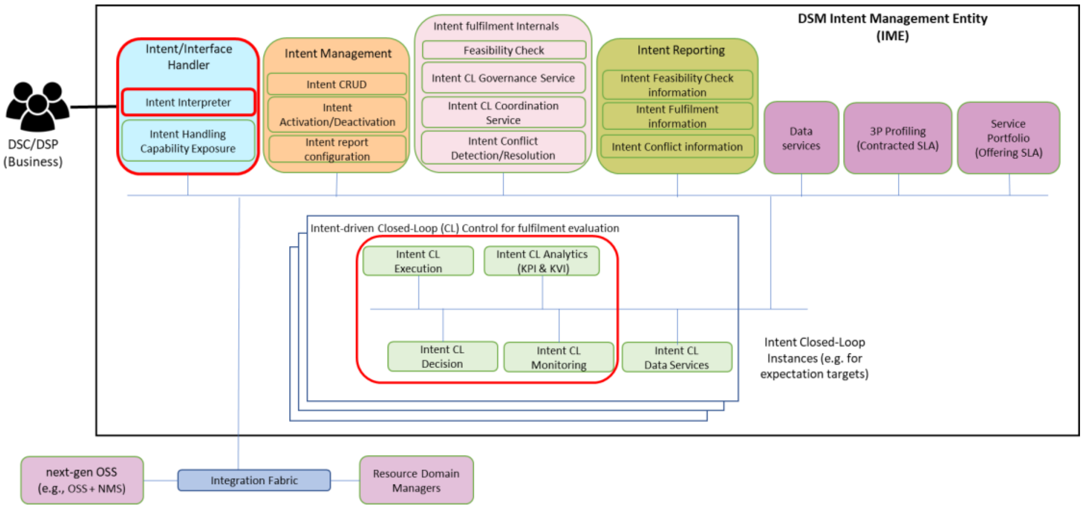

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-13: Intent-driven placement design mapping to functional blocks of the DSM IME.

4.2.7 Declarative intent reconciliation

This enabler employs a declarative approach to i) describe, modify and version-control intent specifications; and ii) define pipelines to streamline the intent life cycle management across multiple management domains operated by different stakeholders.

4.2.7.1 Enabling technologies (SotA)

Declarative Intent Reconciliation (DIR) adopts Infrastructure as Code (IaC) and GitOps principles. IaC is an approach to declare, manage, and provision resources (i.e., IT infrastructure, cloud platform, etc.) through a structured text-based code format. With IaC, resource specifications are described in configuration files, which facilitate the modification and distribution while avoiding undocumented changes. In addition, GitOps is a methodology that extends the principles of IaC with continuous delivery pipelines. GitOps involves using the Git version control system as a single source of truth to manage resource declaration. Changes to the resource specifications can be made through Git actions such as committing new parameters to update the existing ones on the Git repository. Once a change is made, collaborators in the Git repository can review and approve before being realized by pipelines. This approach enables rapid, reliable, and auditable resource modifications, ensuring consistency across stakeholders, reducing the risk of configuration errors and security vulnerabilities.

As mentioned in section 4.1.1, there has been efforts from SDOs (i.e., ETSI ZSM, TMF, etc.) to define intent models and interfaces between the intent owner and handler. By considering the intent as the primary resource to be managed, IaC and GitOps principles are particularly well-suited for the intent-based framework. This adoption ensures continuous reconciliation of management domains to keep up with the changes of the underlying infrastructure owned by various stakeholders and the dynamic requirement of tenants. Nevertheless, there exists challenges such as intent information access control, reconciliation progress monitoring, etc. To address these concerns, DIR has been introduced.

4.2.7.2 High-level description (Beyond SotA)

6G systems aim to address a wider range of use cases by employing more granular intent specifications. Each intent instance is defined by a set of parameters. As the number of intent instances grows, a robust and scalable solution becomes essential to effectively store intent information and track intent configuration changes. Moreover, an intent could be managed by intra-DSP or cross-DSP components of multiple stakeholders to deliver end-to-end services. Thus, a declarative pipeline for a particular intent delivery task (intent translation, intent provisioning, intent conflict coordination, etc.) can help define the involvement of corresponding management blocks, specify the expected input/output of each block, and monitor the workflow progress.

The Declarative Intent Reconciliation (DIR) enabler is introduced to address those challenges by employing two key components (see Figure 4-14): **Git** and the **Reconciliation Engine**. These components work together to ensure that intent specifications are consistent, up-to-date, and effectively translated into actionable configurations across multiple management domains.

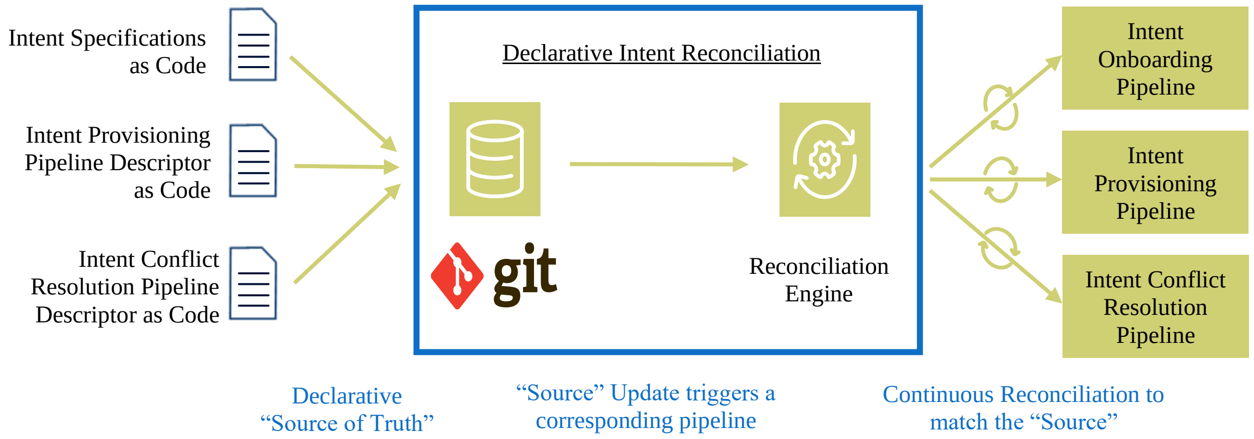

Intent Specifications as Code
Intent Provisioning Pipeline Descriptor as Code
Intent Conflict Resolution Pipeline Descriptor as Code
Declarative Intent Reconciliation
git
Reconciliation Engine
Intent Onboarding Pipeline
Intent Provisioning Pipeline
Intent Conflict Resolution Pipeline
Declarative "Source of Truth"
"Source" Update triggers a corresponding pipeline
Continuous Reconciliation to match the "Source"

Figure 4-14: Declarative Intent Reconciliation.

- **Git** provides a simplified user interface, a database to store source codes (i.e., intent and pipeline instances) and a versioning mechanism to keep track of the changes. In addition, each party must be authenticated by the Role-based Access Control (RBAC) to become an authorized contributor, if it wants to access the repository. Depending on its role, each contributor may have different permissions (i.e., view, modify, etc.).

- **Reconciliation Engine** (RE) is in charge of executing pipelines to reconcile the current intent deployment (actual state) to any updates declared in the source code (desired state). Depending on the type of events triggered by Git, RE queries corresponding intra-DSP/cross-DSP functional blocks to fulfil the new state. Furthermore, it monitors the life cycle of the intent and reacts to any changes of both intent instances and delivery pipelines**.**

4.2.7.3 Analysis of mapping with functional blocks of the architecture

The interactions between DIR and other functional blocks are illustrated as follows in Figure 4-15:

- Interface Handler: providing a simplified interface towards the tenant.
- Intent Management: storing consistent intent information.
- Intent CL Coordination Service: storing consistent CL instances information.
- Intent CL Governance Service: reconciling CL instance.

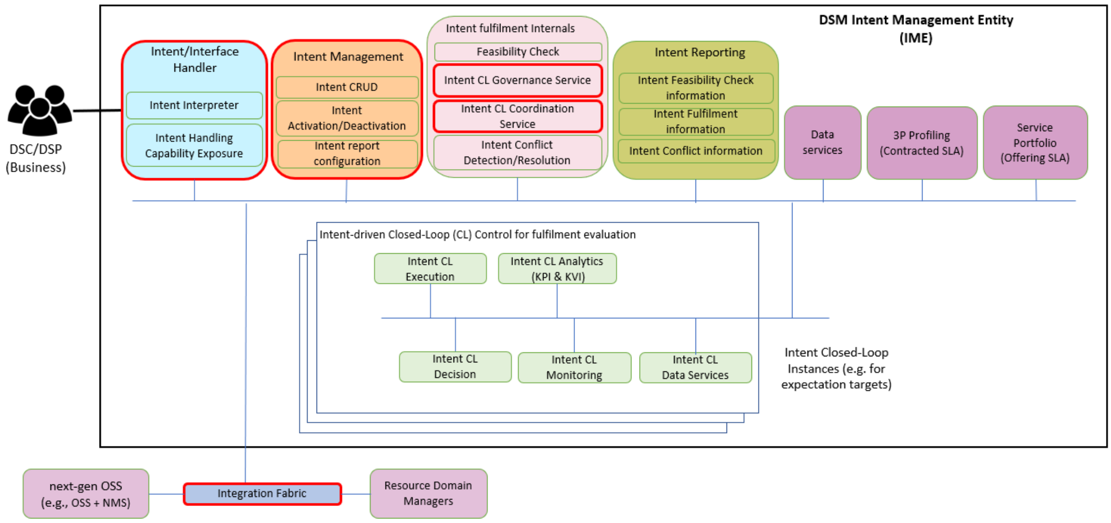

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-15: DIR design mapping to functional blocks of the DSM IME.

4.2.8 Intent reporting

This enabler provides the intent owner with the ability to consume intent reports from the intent handler, using either query or subscribe-notify consumption patterns. Which information an intent report should contain, and how it can be captured into an information model, are issues that will be discussed in the scope of this enabler.

4.2.8.1 Enabling technologies (SotA)

Intent reporting is a feature that provides means for the tenant (intent owner) to verify and audit that the intent, processed by the DSP (intent handler), gets fulfilled across the entire lifecycle. The DSP reports on progress according to the reporting conditions the owner has specified. ETSI ZSM [ZSM011] sets the basis with regards to intent reporting. TM Forum IG1253 series [TMFIG1253] and 3GPP SA5 [28.312] take ZSM directions, and translate them into actual solutions eligible for business and service/network layers, respectively.

Below there is a summary of the main points that define the governance of intent reporting:

- The DSP shall have the capability to report on the following information elements:
    - Intent fulfilment information, which represents the properties for an aspect of the intent (e.g., either an expectation, a target, or the whole intent), including fulfilment status and achieved values for targets.

  - Intent conflict information, which represents conflict type (e.g., intent conflict, expectation conflict, target conflict) and possible solution recommendation to address the conflicts.
  - Intent fulfilment feasibility check information, which indicates that the intent is feasible or infeasible. In the latter case, the reason why the intent is unfeasible (e.g., the intent conflict, the satisfaction of intent fulfilment lowering than threshold) can be reported.
- The DSP shall have the capability to enable the tenant to request intent report information.
- The DSP shall have the capability to enable the tenant to specify the content of the report it wants to get.
- The DSP shall have the capability to enable the tenant to configure the frequency of the intent reporting.

4.2.8.2 High-level description (Beyond SotA)

The aim of this enabler is two-fold. On one hand, the aim is to define an information model for the intent report. This model needs to allocate information elements (fulfilment information, conflict information, and fulfilment feasibility check information) into well-defined constructions connected through class relationships, with attributes configured with readable/writable/notifiable permissions that allow the DSP (intent handler) to offer Hexa-X-II tenants (intent owners) the listed capabilities. On the other hand, the aim is to extend the baseline functional requirements listed in the previous section to enrich intent reporting features, in order to cope with the specific needs of Hexa-X-II tenants. Examples of these needs are listed below.

- Different Hexa-X-II tenants (see Section 6.2.2.2.4 for the definitions of tenants) may have different requirements for intent reporting. For example, some tenants only want that the intent report only includes fulfilment information, while others want also details on intent conflicts. Likewise, different tenants might want to get intent reports with different frequencies, because their assurance systems want to calculate/monitor the performance values in different periods.
- While the intent report is provided at the end of each observation period, the tenant may also wish to know whether the fulfilment info was consistent for the entire observation period. For example, the intent expectation may be reported FULFILLED at the end of the observation period. However, it may be possible that within observation period the intent expectation was NOTFULFILLED (see Figure 4-16). This information can be important for the tenant to understand whether the observation period they specified needs an update (e.g., shortened) or not, and how likely it is for the DSP to keep expectations stable for such an observation period.
- A Hexa-X-II tenant can require different intent reports to be generated in different situations. Based on the content selection criteria, the tenant can obtain reports of different contents according to different conditions. For example, it is possible to ask for a report about all elements of the intent when the system is getting degraded. If the system complies again, a shorter report might be sufficient.
- Reports also can be generated and sent based on events, rather than periodically. Events describe significant situations in the operation of intent and indicate that the intent has reached a particular state. For example, these events can include intent being accepted, intent being rejected, or intent being degraded, etc.

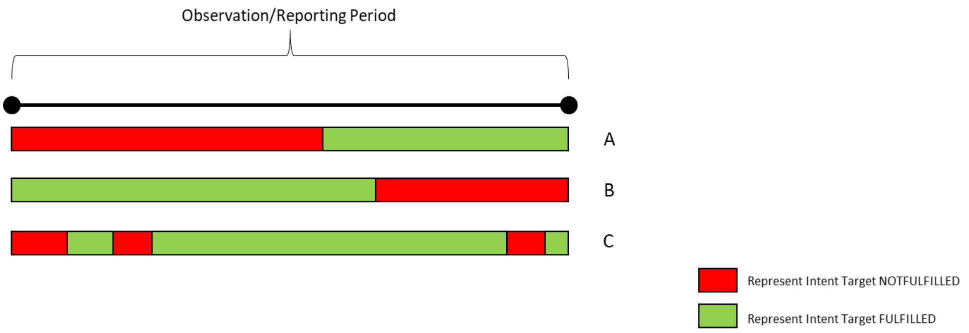

Observation/Reporting Period
A
B
C
🟥 Represent Intent Target NOTFULFILLED
🟩 Represent Intent Target FULFILLED

Figure 4-16: Intent not fulfilled within the observation/reporting period.

4.2.8.3 *Analysis of mapping with functional blocks of the architecture*

The impact of this enabler in the DSP architecture is illustrated in the figure below (Figure 4-17).

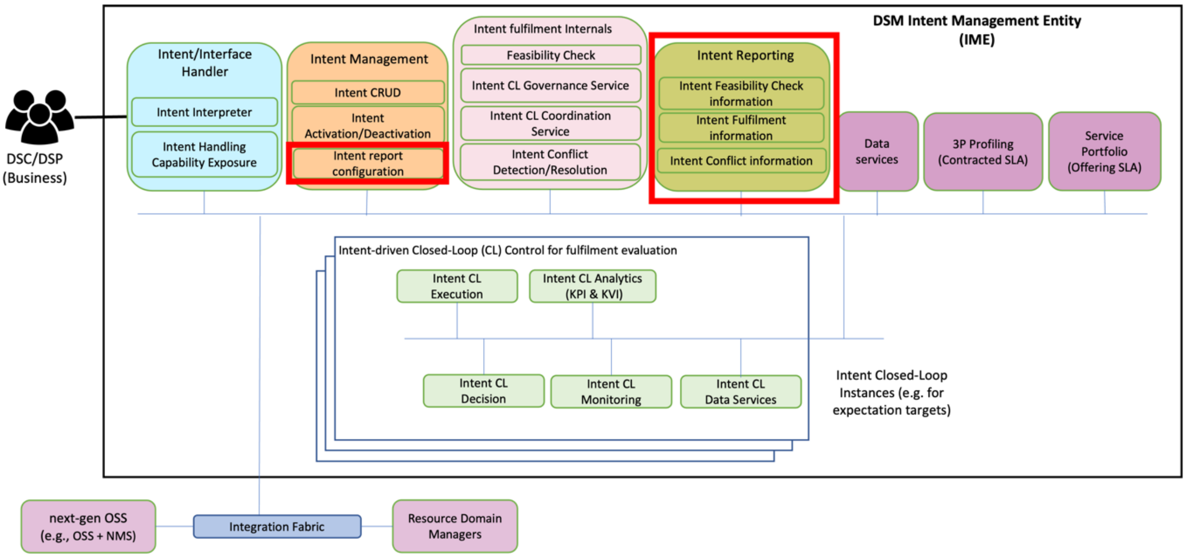

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-17: Intent reporting design mapping to functional blocks of the DSM IME.

4.2.9 3rd party facing services

This enabler aims to specify how the DSP provides a characterization of i) individual tenants accessing the Hexa-X-II system, capturing this information in the form of a 3rd party profile, and ii) service offerings, which will be later linked to tenants according to well-defined SLAs.

4.2.9.1 *Enabling technologies (SotA)*

As to the 3rd party (3P) profile definition, in the 3GPP community, TR 28.804 [28.804] and TR 28.824 [28.824] raised attention on the need to have multi-tenancy support in 3GPP management, especially with the sight set of slicing offering. However, neither of these studies concluded takeaways that justified going for a normative phase; of the main reasons sustaining this decision was the 3GPP did not reach consensus on the scope and meaning of the tenant concept.

With regards to service offering characterization, the reference work as of today is led by the Information Technology Infrastructure Library (ITIL). ITIL is a framework designed to standardize the selection, planning, delivery, maintenance, and overall lifecycle of IT services within a business. At the core of the ITIL framework, it is the service portfolio management work, built upon the following tenets: i) improved efficiency, to reduce operational costs and increase efficiency; ii) increased visibility; iii) optimized service delivery, by focusing on quality and customer experience; and iv) enhanced service lifecycle management.

4.2.9.2 *High-level description (Beyond SotA)*

This enabler aims to go beyond the state-of-the-art reported in the previous section, providing solutions that allows characterizing tenants (i.e., 3Ps) and service offerings under the DSP realm. On one hand, the 3P profile contains the following information:

- The **tenant type** that the 3P represents. Three options available: i) “tenant-1 type”, wherein the tenant is an aggregator (e.g., OTT, hyperscaler marketplace or telco consortium); ii) “tenant-2 type”, wherein the tenant is an enterprise customer accessing the DSP in a wholesale model (e.g., application service providers accessing through aggregators); and iii) “tenant-3 type”, wherein the tenant is an enterprise customer accessing the DSP in a retail model (e.g., vertical customer).
- The **credentials and access control** (authentication & authorization) solutions supported by the 3P back-end/IT systems.

- The **trust level of the 3P**, which is useful in federation scenarios. When delivering services to a 3P, the serving DSP sometimes needs to leverage resources/capabilities from other partner DSP(s). How these partner DSP(s) can trust this 3P, considering they do not have a contractual agreement with it? The trust level aims to precisely fill this gap.
- **Services contracted** by the 3P**, and associated SLAs**.
- Information on **3P subscribers**. “Tenant-3 type” subscribers are enterprise users, while “Tenant-3” subscribers are also DSP users. The DSP needs information on these subscribers (e.g. IP addresses, MSISDN) to appropriately provision services and manage their consents (to comply with privacy and regulation in force).

On the other hand, the service offerings can be individually characterized by the following information:

- **Service name:** unique identifier for the service.
- **Service status**: “defined”, “designed”, “built”, “tested”, “released”, “retired”.
- **Service owne**r: it specifies the name of the owner along with its role in Hexa-X-II system (“DSP”, “Partner DSP”, “3rd party”).
- **Flavors and package variations**: different SLA offerings, different coverage of time zones, different coverage of geographical regions.
- **Costs and pricing**: available pricing schemes for the service provisioning, rules for penalties/charge backs
- **Dependencies with other services**, relevant for those cases in which a (composite) service builds on other (nested) services.

4.2.9.3 Analysis of mapping with functional blocks of the architecture

The impact of this enabler in the DSP architecture is illustrated in the figure below (see Figure 4-18).

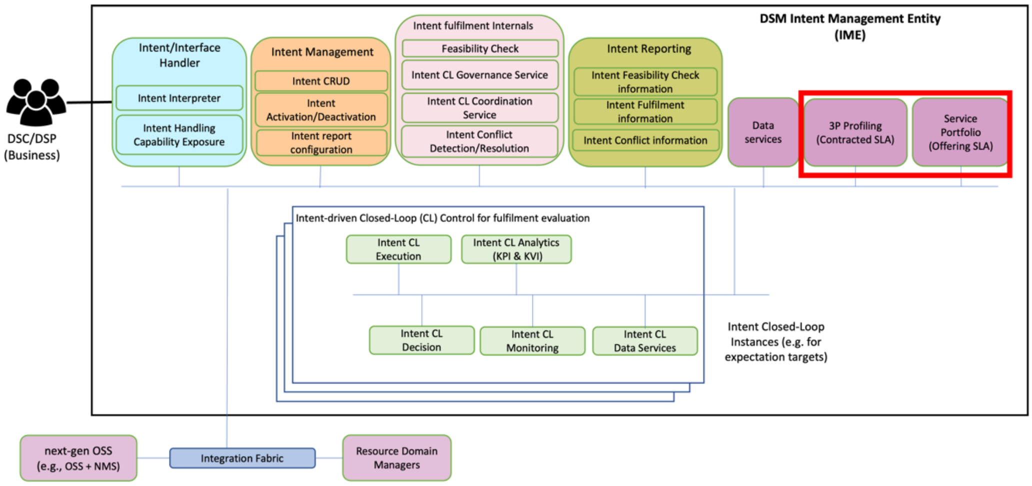

DSM Intent Management Entity (IME)
DSC/DSP (Business)
Intent/Interface Handler
Intent Interpreter
Intent Handling Capability Exposure
Intent Management
Intent CRUD
Intent Activation/Deactivation
Intent report configuration
Intent fulfilment Internals
Feasibility Check
Intent CL Governance Service
Intent CL Coordination Service
Intent Conflict Detection/Resolution
Intent Reporting
Intent Feasibility Check information
Intent Fulfilment information
Intent Conflict information
Data services
3P Profiling (Contracted SLA)
Service Portfolio (Offering SLA)
Intent-driven Closed-Loop (CL) Control for fulfilment evaluation
Intent CL Execution
Intent CL Analytics (KPI & KVI)
Intent CL Decision
Intent CL Monitoring
Intent CL Data Services
Intent Closed-Loop Instances (e.g. for expectation targets)
next-gen OSS (e.g., OSS + NMS)
Integration Fabric
Resource Domain Managers

Figure 4-18: 3rd party facing services design mapping to functional blocks of the DSM IME.

5 Enablers related to security, privacy and system-level resilience

This chapter describes the enablers related to the *trustworthiness KVI*, namely those focused on security, privacy, and resilience. These enablers have been analysed according to the principle of the *6G Delta*, as introduced by the Hexa-X project [HEX23-D13], that is, considering the specific threats posed by the different trends towards 6G, either by the mobile network evolution itself or by the evolution of the security base technologies. This chapter introduces the trends and their potential threats, describing the enablers proposed to address them, and analysing how network security, user privacy and system-wide resilience can benefit from the use of these enablers.

In the particular fields of security, privacy and resilience, enablers are not necessarily intended to provide a new, differential functionality, but to address specific threats, providing mechanisms to detect them and to mitigate their impact in the system performance. An analysis of such enablers needs to be introduced by identifying and describing the potential threats, justifying the requirements on the corresponding enablers. Therefore, this chapter not only focuses on starting from the desired features the enablers are expected to provide, but also it is intended to provide a justification for these features by describing the threats they intend to identify and mitigate.

With this focus in mind, the discussion of the proposed enablers is structured along the *threat families* they intend to address. First of all, the threats associated with foreseen architectural trends, essentially connected with the virtualization, disaggregation and further composition of network infrastructure, are introduced, followed by a discussion on the enablers identified to address them, based on formal methods and confidential computing. The chapter continues with a discussion of the threats implied by the pervasive use of AI, and the enablers to enhance its trustworthiness at all levels. The evolution of the mechanisms to establish trust, and the essential elements for identifying agents in the network, is then discussed, together with the key enablers for these matters: distributed trust fabrics and the evolution of cryptography. To complete the analysis, the evolution of the network physical layer, with new technology paradigms like joint communication and sensing (JCAS) and new relevant trends related to the evolution of physical protection strategies, is considered, together with the corresponding enablers at this layer.

Finally, in the spirit of enhancing the applicability of the different enablers and facilitating their integration with the general 6G system-level design, an analysis of the proposed mechanisms to validate the enablers discussed here is provided. The project plans to report the results of these validations in future deliverables.

5.1 Architectural enablers

We have identified three main architectural trends conforming the security impact of the 6G Delta mentioned above. These are the Network of Networks (NoN) compositional pattern, the use of a cloud continuum as base infrastructural approach, and the application of radical disaggregation mechanisms, especially in the RAN segment. In this section we discuss the implications of these trends, highlighting the main characteristics of the required enablers for guaranteeing trustworthiness, and finally discussing the main enablers associated to these trends, focused on the technologies around *confidential ICT* and the opportunities offered by a wider use of formal security proofs.

5.1.1 Security impact of 6G architectural trends

5.1.1.1 Network of Networks (NoN)

The NoN concept focuses on the dynamic integration of multiple networks or network segments (we will refer to them in general as network domains), dynamically building connectivity services. The NoN concept includes the integration of Body Area Networks (BAN), Car Area Networks (CAN), Non-Terrestrial Networks (NTN), multi-connectivity proposals, etc. Some of them need a dynamic, opportunistic integration (typically, via federation) of different networking components, due to their temporary availability (LEOs, UAV-based nodes) or local range availability (hotspot-like solutions). The NoN concept raises many issues linked with procedures needed for the attachment and separation of the network domains, and procedures related to the proper and optimized operations of the federated network. Such procedures should include federation-oriented mechanisms at the control plane and the management interfaces.

System-level procedures should consider the integration of networking solutions of different operators, with limited information exchange among the integrated networking domains, and eventual prescheduling of the federation operations prepared in advance. Moreover, some federations (for example, D2D-based V2X) can be short-lived. The NoN operations can be done peer-to-peer or hierarchically with dedicated entities external to the federated network domains and selected according to the nature of the applicable use cases.

Making the NoN concepts trustworthy calls for their special handling from the security point of view. The security mechanisms for static multi-domain integration are so far not well defined, and they should consider the following three basic scenarios:

Scenario 1: This scenario lies in the interconnection of fully independent (autonomous) networks, of which each has a complete set of security mechanisms. In this case, modified and generalized non-3GPP network roaming mechanisms can be used. As the integrated networks are secure, there is a need to secure the interconnection between the networks. This can be achieved by dedicated proxies for the control plane and user plane. The modified solutions used for 5G roaming, i.e., Security Edge Protection Proxy (SEPP) for Control Plane information exchange [33.501] and Inter-PLMN UP Security (IPUPS) proxy for User Plane, can be used. The enhanced capability of negotiations (already present in SEPP) should expand both proxy functionalities. Such loosely coupled integration provides no synergy, and tightly coupled integration in which entities optimize their behaviour (this includes redundancy) and for NoN should also be considered. The scenario may cover a temporal integration of Private Mobile Networks with PLMNs, fixed networks, BAN or CAN. There is a need for security policy negotiations and alignment and separate treatment of trusted and non-trusted networks federation. For some use cases, prescheduled integration can be implemented in which the security policies are configured in advance or are known a priori. This is of particular importance for short-lived federations.

Scenario 2: This scenario concerns the dynamic multi-connectivity of users, i.e., a dynamic use of multiple networks technologies. It may include a hotspot-like use case. The scenario is similar to the integration of 3GPP networks with "non-3GPP access networks". In this case, it is assumed that the user is always connected to the 3GPP network and, occasionally, may have access to other (trusted or untrusted) access networks. In the proposed solutions, the main network can be used for control plane messages exchange, which includes all security-related operations. The other access technologies are used as user plane solutions only. Therefore, there is a need to provide confidentiality and integrity protection of the exchanged data.

Scenario 3: This scenario lies on integrating different networking solutions that are not access networks. Such transport networks can be used for providing long-range interconnection of existing networks or as one of several transport networks used in the Core network, which is dedicated to a specific use case or network slice. The integration of such solutions needs mutual authentication between the integrated domains.

In addition, we may consider mixed scenarios in which the NoN is composed of the integration of independent networks, dynamic multi-connectivity, and the addition of diverse transport domains to the Core network. These mixed scenarios deal with multiple access links, multiple transport solutions, and replicated NoN functions, due to the integration of fully independent, disjoint networks, a kind of internal roaming capabilities. In these scenarios, a combination of the considerations above would apply.

The 3GPP has defined some mechanisms for secure 3GPP network integration based on roaming and secure adding of *non-3GPP access networks*, as identified by 3GPP itself [23.501]. The approach is 3GPP-centric, and it does not allow extending the security mechanisms beyond the 3GPP networks, even in the static case.

5.1.1.2 Cloud continuum

The Cloud Continuum concept [HEX23-D13] assumes integrating all infrastructure resources and their uniform application exposure. It includes using virtualized, unreliable, and constrained Far-Edge (i.e., terminal) resources. To that end, the ETSI NFV ISG approach [NFV0062] has security-related limitations.

First, the business interfaces towards infrastructure providers are not defined, only the Infrastructure as a Service (IaaS) model of a private cloud, typically owned by the network operator, is so far defined by ETSI NFV ISG. However, according to [INF001], the NFVI should provide authentication mechanisms to ensure that only authorized entities have access to NFVI resources. This gap is currently being filled by introducing certificate management functionality for management interfaces. Second, it assumes an almost static resource pool. Third, it cannot cope with constrained, unreliable, and dynamic resources. An outline of such a Cloud Continuum Framework has been provided in [KBP23]. The essential component of the framework is the Resource Layer (RL), in which operations are focused solely on the operations on resources. It is worth noting that the RL includes resource orchestrators.

The concept assumes that the RL can integrate resources of different data centres owned by different infrastructure providers via the South Bound Interfaces (SBI) of the RL and expose all the resources to service orchestrators using the North Bound Interfaces (NBI). The SBI should be secure and allow for mutual authentication of dynamically attached data centres and confidential access to data centre resources.

The RL offers an almost infinite resource pool. The concept described in [KBP23] proposes to use multiple service orchestrators atop RL. It proposes, in the first phase of orchestration, a creation of isolated resource partition based on the Network Service (NS) requirements (i.e., placement, delays, cost, etc.) to simplify the orchestration process. Such partition of resources is visible by the orchestrator that will be involved in the orchestration of this NS. It is therefore needed to address the security of the RL SBI and NBI. Moreover, functions related to security features, like observing malicious behaviours related to the RL, should be considered.

5.1.1.3 Disaggregation mechanisms for RAN

In general terms disaggregation implies to create modularity and flexibility by breaking down a complex system into multiple segments. In the context of RAN the tendency of disaggregation evolves over time with clear decomposition of various functional elements and components that conventionally constitute a unified network infrastructure. As such hardware, software and network services are separated into distinct entities and can be managed and operated independently. However, over those benefits such as flexibility and optimization, disaggregated scenarios may introduce new security challenges with RAN and its operations. In a way, it may expand the potential attack surface by making the network more vulnerable since each disaggregate element can be a possible entry point for cyberattacks. With disaggregation, there is also need for robust node authentication, authorization, secure interface management, and verification mechanisms to prevent unauthorized access and data breaches. This may require strong adherence to the zero-trust principles for RAN design [PPR+23]. Moreover, as data traverses through various disaggregated elements, ensuring its confidentiality and integrity becomes paramount. Encryption, integrity checking, and secure data transmission protocols are necessary to protect sensitive information from interception and unauthorized exposure.

Moreover, aside from ensuring integrity and confidentiality, safeguarding the availability of network assets is crucial, particularly against Distributed Denial of Service (DDoS) attacks originating from numerous compromised UE elements. Given that attacks on the network often initiate at the radio interface, it can be prudent and essential to develop a method for detecting DDoS attacks at this level. This enables the implementation of mitigation and countermeasures before malicious packets breach critical points in the network. However, devising a detection algorithm at this stage requires the identification and establishment of new features and methodologies. These innovative approaches are imperative for creating a tailored solution for DDoS detection in this specific context, potentially enhancing the network's ability to proactively thwart DDoS attacks and maintain uninterrupted service for legitimate users.

The considerations made above are equally applicable to other network segments, as disaggregation is applied in IP forwarding, whether at the transport and the core user plane, and a coordination with the proposed mechanisms for the radio protocol stack would aid in devising a new feature set and innovative defence approaches.

External parties have certain control over the network for deploying third party applications in disaggregated RAN architectures. Thus, security architecture should be robust to counter the vulnerabilities introduced by these third-party applications. Identification of such security threats and vulnerabilities in the disaggregated RAN architectures using AI/ML models has received much attention [KKU+22]. In that sense, while using data locally to train models and aggregating model updates to derive a global model, Federated Learning (FL) and its different variants such as hierarchical FL or distributed FL, are highly applicable to RAN disaggregated scenarios. FL models can be trained for anomaly or intrusion detection, and later deployed at the RAN intelligent controllers to process real-time network traffic to detect anomalies, identify root causes, and even identify potential attacks before they impact the RAN infrastructure.

FL-based Deep Reinforcement Learning (DRL) techniques have shown effective results for different RAN optimization operations [ATF+22]. FL techniques can be used in the security domain in a similar manner. FL security agents can be in different locations of the network to collect data and pass them to a trusted data collector before passing them to the inference engine. Security measures should be designed considering the accuracy and latency requirement of the applications to be decisive in real-time. In the next stages of Hexa-X-II, we intend to run experiments on how an FL-based security solution can be applied for RAN-disaggregated scenarios. Some experiments can be conducted from some existing data sets to exemplify the general framework and show its viability for the anomaly detection in disaggregated RAN architectures.

5.1.2 Formal security proofs

The complexity of mobile networks increases due to disaggregation as well as integration of new functionalities. Although the first can offer benefits in terms of security and the latter enables new use cases, the drawback of this increased complexity is an increase in the likelihood for errors and contradictions in the overall specifications which in turn can lead to security or safety threats. Addressing the threats implied by increased complexity calls for the application of formal methods able to support correctness analysis.

One necessary step to mitigate complexity driven threats is to strive for more formal specifications, suitable to be used as the foundation for (semi-automated) formal analysis and verification. Although formal standards and formal verification are highly beneficial, there are quite some challenges which need to be addressed. Besides the effort regarding the formal verification itself (which is not in focus here), writing and reading/ understanding specification documents using formal languages is cumbersome – especially for non-domain experts. Therefore, different “views” tailored for the needs and experiences of the reader/writer are necessary. Given the current advancements in the domain of AI, it seems to be imaginable to achieve this by some form of automatic transformation from one representation to another. This will allow to have the technical essence expressed using formal languages and at the same time the content remains accessible for humans, who are not experts in the domain of formal methods. Although there is currently research ongoing into this direction, the state-of-the-art still does not allow the fully automated translation between natural language-based specifications and its formal counterpart. Therefore, a first step could be a “human-in-the-loop”-based approach.

Methodologies and procedures to support formal verification of standards and specifications are needed and should be introduced in the protocol development process for the next generation mobile networks. This can be ground on scientific findings and related approaches of the past and aligned, e. g. with recent activities of the IETF research group “Usable Formal Methods” [IRTF23].

5.1.3 Confidential network deployment

Network domain composition, disaggregation and the cloud continuum infrastructure will be enabled by cloud-native deployments, typically based on the usage of an orchestrator framework, like Kubernetes [KUB23]. This section considers the base mechanisms for a confidential usage of computing and networking facilities, supporting trust across the different roles in the deployment, and concludes with a set of recommendations for their validation. These roles comprise a cloud service provider and one or several tenants. The cloud service provider is the entity providing and operating the cloud infrastructure, and a tenant is an entity running workloads on top of this infrastructure. The workloads running on the infrastructure are referred to as cloud native functions. The exact demarcation between cloud service provider and the tenant within the cloud stack (e.g., which layer is operated by which entity) might depend on the specific deployment scenario and has significant impact on the resulting security properties.

This section will start with an introduction to confidential computing, which is one of the key components of confidential network deployments, followed by a description of various related practical approaches already applied or emerging. Subsequently the concepts of topology attestation and supply chain security will be introduced, before an outlook about planned practical experiments will conclude the section.

5.1.3.1 Confidential computing

An important aspect in this context is the assumed trust relationship between the cloud service provider and the tenants. In one model the tenant is fully trusting the cloud service provider (for instance, if cloud service provider and tenant are part of the same organization). In this case the main responsibility for security is with the cloud service provider, which needs to protect its own infrastructure from attacks and ensure isolation of workloads. In the other case, the tenant considers the infrastructure offered by the cloud service provider as potentially insecure. This does not necessarily mean that the tenant regards cloud service providers as potentially malicious organizations, but rather that a tenant does not fully rely on cloud service provider’s ability to prevent all attacks using the infrastructure as attack vector. In other words, the tenant wants to run its workloads in a potentially hostile environment in a safe way and wants to have own evidence about the protection. Solutions fulfilling these requirements are often referred to as Confidential Computing [CCC22]. Confidential computing consists of two pillars, remote attestation and trusted execution environments (TEE), as shown in Figure 5-1. The secure application running in a TEE (attester) is sending evidence to a key management server (relying party). The key management server is invoking the attestation server (verifier) to verify the validity of the evidence, i.e., that the evidence indeed originated from an application running in a TEE. Subsequently the key management server sends secrets back to the secure application. The secrets are protected using information included in the evidence, e.g., a public key, whose corresponding private key is known only in the TEE.

Authentication mechanism typically rely on the cryptographic proof of possession of a secret key. Remote attestation is extending this approach through the utilization of verifiable claims about a system (called attester), which include information about the state of the system or how the system is executed. The goal of remote attestation is to verify a set of claims (called evidence) made by a communicating party (i.e., the attester), about itself. Typically, after claims are verified, a secret is provided to the attester, so it can use this secret for instance in the framework of existing security protocols to authenticate itself towards other peers. Although these peers are not directly involved in the remote attestation, they trust the attester, because without previous successful remote attestation the attester would not be in possession of the secret.

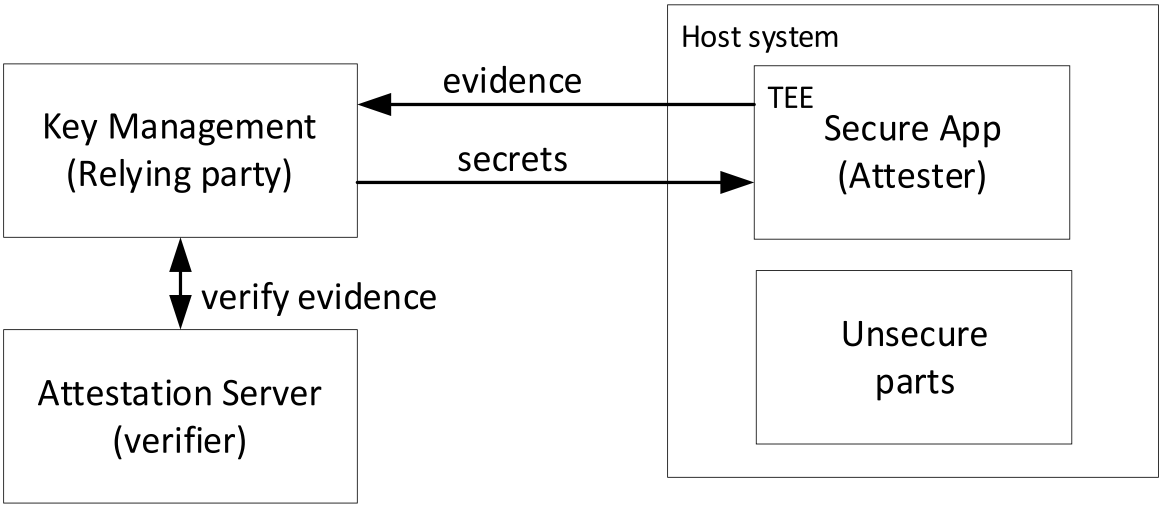


Figure 5-1: The relationship between TEE, remote attestation and provisioning of secrets.

This approach of attested provisioning only makes sense if the attester can protect the secret and if the attester is able to protect its state (i.e., if the state of system during the provisioning of the secret is the same as during usage of the secret). Therefore, remote attestation is most valuable, if it is combined with the usage of trusted execution techniques and if the remote attestation is used to provide a secure channel for provisioning of secrets terminating in the TEE. By definition, a TEE is an isolated environment, which allows execution of code and handling of data protected from other applications, i.e., even an attacker with higher privileges is not able to read or modify data in the TEE ("data confidentially and integrity") or to manipulate the executed code ("code integrity"). Typically, the TEE is realized on the main processor [AMD20, ARM23, INT23]. The relationship between TEE, remote attestation and provisioning of secrets is illustrated in Figure 5-1.

In the context of 5G and 6G networks, one main use case of remote attestation and confidential computing is securing the communication between network functions defined by 3GPP. Thus, the secure provisioning and protection of private keys in a cloud native setting is one main design goal of 6G cloud-native security.

5.1.3.2 Platform integrity using TPM / Measured Boot

The main idea of solutions based on platform integrity is the verification of the integrity of hosts by taking measurements of system components during system boot. The actual measurements are taken by a trust agent running on the host utilizing a Trusted Platform Module (TPM) and sent to a central verification service, which compares the measurement values against known good values [INT23]. The authenticity and correctness of the measurements is ensured by utilizing a TPM and protocols defined by the trusted computing group [TPM2].

Verification of the platform integrity does not include mechanisms for the secure execution of workloads. The solution also does not include a mechanism to use the remote attestation to securely provision a secret into a workload. Thus, it should not be regarded as a confidential computing solution. Rather it is a data centre solution, which can and should be used to ensure basic security of all hosts within a cluster.

In addition to pure verification of platform integrity, workload confidentiality can be achieved by means of image encryption. Image encryption for cloud native functions is a concept introduced by the Open Container Initiative [OCI19]. Before uploading to a registry an image is encrypted with an encryption key, which is preserved in a key management server, operated by the tenant [LXR+21]. During the start-up procedure of the container image the container runtime decrypts the (encrypted) image pulled from the registry. The container runtime can retrieve the decryption key from the key management server if it is running on a host, whose integrity could be successfully verified. In this way the leakage of images can be prevented, for instance to protect intellectual property or to ensure regulatory or policy-based requirements with respect to the placement of functions.

In all solutions based on platform integrity the tenant needs to trust the cloud service provider. That is, the tenant needs to trust the cloud service provider to be not malicious and it needs to trust the cloud service provider to be able to defend itself against other malicious actors, who might want to penetrate the cloud service provider to attack the tenant.

5.1.3.3 Attested identity provisioning

Securing communication between CNFs using Transport Layer Security (TLS) is a general problem. Independent from confidential communication, solutions for the provisioning of TLS (and other) secrets to CNFs exist, like the Secure Production Identity Framework for Everyone (SPIFFE) [FFG+20].

In the SPIFFE approach, a workload retrieves its identity (in the form of a X.509 certificate) and corresponding private key from an agent running on the same node. When the workload requests the identity, the agent verifies the identity of the workload. Instead of relying on another secret, which is pre-provisioned on the workload, the agent uses the (Unix) process identity of the calling workload to interrogate the Unix kernel or the orchestrator to obtain the necessary information about the workload. For instance, the information could be the name or a hash value of the workload executable or the name of the pod or service account used by the workload. Workload and the agent communicate via a Unix socket interface, which is exposed by the agent.

A common design pattern in cloud native deployments is the utilization of a service mesh. One typical aspect of a service mesh is the usage of a communication proxy, which is realized as a sidecar executed next to the actual workload in a pod. The communication proxy can also be used to terminate the TLS connections between workloads, if this concept is combined with the SPIFFE approach, i.e., the communication proxy receives the TLS secrets from the agent.

This approach does not include the concept of a TEE. Thus, security is relying on the CSP. However, the main design of enabling the injection of root secrets without the need to pre-provision other secrets into workloads is leading the way and most likely will be adopted in mature confidential computing solutions.

5.1.3.4 Attested enclaves / secure key caching

Two main approaches for TEEs exist: VM-based (AMD SEV, Intel TDX), and process-based (Intel SGX). The applicability of process-based trusted execution technologies for running complex workloads has gained strong momentum from the introduction of libOS technologies (see for instance [STC+20]). To apply a TEE, applications must be split into a trusted and an untrusted part. The trusted part is protected within the TEE (an enclave, in SGX terminology). The untrusted part of the application communicates with the TEE by means of dedicated library calls, referred to as e-calls. When porting existing applications or designing new applications, a meaningful split between trusted and untrusted parts needs to be made.

In case of Secure Key Caching [BSW+23], this split is done in such a way that the trusted part (i.e., the part running in the enclave) implements the functionality of a Hardware Security Module (HSM). That is, like an HSM, the enclave can store credentials, such as private keys, and execute cryptographic operations, both of which can be used by communication protocols during (TLS) connection establishment. Like in case of HSMs the integration of an application with the enclave can be done via a PKCS#11 interface [OAS23], which facilitates the integration into existing applications supporting PKCS#11 interfaces.

Private keys stored in the enclave never leave the enclave, even when they are used during TLS session setup. After the TLS handshake the entire session is handled within the application without using the enclave. The loading of keys into the enclave is part of remote attestation. The enclave loads the keys from a Key Broker (operated by the tenant) if it can provide necessary evidence about its own integrity.

5.1.3.5 Confidential container

The confidential container approach [COC23] uses encrypted container images, and both the decryption of the workload (executed by the image agent) as its execution happen in trusted execution environments. The image agent receives the decryption key from a key server only after successful remote attestation. After successful decryption the image agent launches the container in another trusted execution environment.

Per se, this design does not include a mechanism for provisioning of secrets into the confidential workload. In theory, secrets could be embedded into the image since the image is encrypted. However, this would lead to large overhead, because a dedicated encrypted image (with different values of the secrets) needs to be created for each instance. Thus, this approach should be combined with a provisioning mechanism for confidential workloads as described in the next section.

5.1.3.6 Confidential workload provisioning

The main idea of this approach is to combine remote attestation for secure provisioning with the concept of a service mesh. In contrast to service meshes, where application containers are executed together with a side car in a pod, the actual application is executed together with an *init* function in an enclave utilizing the libOS concept mentioned above. The *init* function is the entry point into the enclave and it is executed before the application. The *init* function contacts a key management service, executes remote attestation and through the secure channel established as part of remote attestation receives secrets. Remote attestation covers the entire enclave including the *init* function and the application.

The *init* function prepares the received secrets for usage by the application, for instance by setting environment values or creating files with the secrets readable only inside the enclave. Subsequently, the *init* function starts the application, which use the prepared secrets. The distribution of the secrets to the workloads is configured on a key management server through appropriate policies. A policy might define that only workloads running a dedicated image (identified by image hash and image signature) are allowed to receive specific secrets.

One main advantage of this approach is that the design is stateless. The application never needs to persist the secrets. If the application is stopped the secrets automatically disappear and if the application is started again the secrets are reprovisioned. If a scalable service consists of several instances of the application (potentially running on different nodes) the secrets are automatically provisioned to all instances of the application (even if the number of instances is increased during operations). Technical realizations of this approach include [EDG23, OCC23, SOF23].

5.1.3.7 Topology Attestation

In cloud-native deployments of network services, topology is not any longer associated with a set of physical nodes and links, and the parameters associated with them. Function scaling and migration imply great variability of the physical paths network flows have to traverse, and it becomes essential in many cases (think just about guaranteeing the execution of certain security functions at a certain segment, or the assurance of data not leaving a certain domain, to name two examples) to verify that service topology is still valid.

For this purpose, Proof of Transit (PoT) [BBM+21] techniques are to be applied. PoT adds a small piece of metadata to those designated packets in a given flow, or to the packets of specific Operation and Management (OAM) flows. These metadata are updated at each node of the intended service path. Finally, a validator at the end of the path needs to validate the data and verify the packet have crossed the path correctly.

As indicated above, PoT can be applied in two main scenarios. One is the *verification of the path* for the packets in a flow (or alternatively, for a statistically significant number of packets) by means of extending packet headers to incorporate the mechanisms described below. The other scenario considers the *attestation of the topology* of a particular network deployment, by injecting at regular intervals a set of specific OAM packets in order to verify the required paths are established.

When speaking about PoT for a path, there are always two or more nodes and a controller in the topology. The controller will handle the process of distributing the crypto material used to process metadata at each of the nodes, and the nodes will handle the verification process depending on their type. We can consider the *ingress node*, first node in the path that introduces the packets into the PoT path, the *middle nodes*, always placed between two nodes, they forward the PoT packets while updating the metadata values, and finally the *egress node*, last node in the path which handles the verification process and removes the metadata added to each packet.

The most common implementation, as defined originally in [BBM+21], uses the Shamir's Share Secret schema SSS, what does not guarantee proper ordering. A more complete, safer solution, able to guarantee order is defined in [ALP+20], and work is ongoing in exploring other cryptographic methods, like a recent proposal to use vector commitments [CF13]. Experiments evaluating the different choices and their applicability in different scenarios are planned for the next project phase.

5.1.3.8 Supply chain security

Any current cloud-native application is a mix of code from different parties, integrated, and even deployed (when models such as Platform-as-a-Service (PaaS), Software-as-a-Service, (SaaS) or the recent *serverless* trends are followed) following different practices and tools. Considering the supply chain for these applications requires to take into account all the elements related to their lifecycle, from design to maintenance. Cybersecurity organizations [ENISA21] warn about how unnoticed risks induced by lack of visibility of the software supply chain expand the attack surface and can be leveraged by miscreants. The need for mechanisms allowing consumers of software products to know about security practices applied by software suppliers, and incorporate independent evaluation results, has been acknowledged and there is ongoing work to produce an architecture for improving software supply chain security [IETF23].

5.1.3.9 Experimenting with confidential ICT

Platform integrity based on measured boot is and will be essential part for basic data centre security and should be adopted in 6G deployments. Additionally, Confidential Computing is expected to become a requirement for 6G, at least for certain deployment scenarios. For security and practical reasons solutions which ideally allow moving entire CNFs into trusted execution environments seem to be preferred. An automated process for secure provisioning (including remote attestation) into the trusted workloads is needed, as well as the evaluation of the different choices for topology attestation and their applicability in different scenarios.

Furthermore, in the cloud-native networking environment envisaged for 6G, network service providers will play a *prosumer* (both provider and consumer) role in the software supply chain, so an evaluation of this role and the implications the supply chain security mechanisms would have is required and is intended to become the target of a further evaluation.

Further research questions include the feasibility of the existing open-source solutions for telco workloads with respect to performance and impact on the development and operational processes. The most likely impact on standardization will be the definition of combined procedures for attestation, supply chain assurance, and secret provisioning.

5.2 Enablers for trustworthy AI

AI/ML is expected to play a key role in the realization of the data-driven 6G system. Potential use of AI/ML in 6G includes network optimization, network automation and security protection of the 6G system against various attacks. Considering the security by design and privacy by design approaches, trustworthiness of AI/ML [ABB+20] should be considered in the integration of AI/ML in 6G, i.e., prevention mechanisms against the attacks to AI/ML based systems needs to be in place. These attacks can target the security of the system or privacy of information [SPL21]. The aim of security attacks against AI/ML is to make the AI/ML system work unexpectedly or make the AI/ML system behave in the direction the attacker wants, which may result in serious negative effects on the execution of the 6G system. The security attacks can be done during the training or during the inference phase of the AI/ML model lifecycle. The goal of privacy attacks on AI/ML is to learn sensitive data, such as information about the training data, the parameters of the AI/ML model, the AI/ML model itself, inference queries and the responses.

5.2.1 AI security

Recently, it has been shown that AI/ML models may be vulnerable to security attacks [SZS+13], [TCE21]. In reality, even very little, mostly undetected changes in data samples can cause state-of-the-art classifiers to make inaccurate predictions during the inference time. Despite the heterogeneity of the network and the scattered nature of the communication domain, there is still a possibility of an adversarial attack in a telecom environment. This fact poses a possible danger as AI is expected to play a key role in the 6G communication systems. According to prior studies, adversarial attacks with optimized perturbations can impair the functionality of a telecommunication network or service [TKK23]. Therefore, effective defensive strategies are needed to counteract the consequences of such attack threats.

To mitigate these kinds of attacks, there have been studies in the literature that propose the use of adversarial training and explainable AI (XAI) [ZAM22]. Typically, AI/ML models that provide predictions, recommendations, or decisions do not offer clear understanding of how they reach those outcomes, known as black-box AI-algorithms are providing promising solutions to the security of 6G networks [SPL+21]. But these black-box AI algorithms lack transparency in their decisions. Explanations are a significant factor in understanding the attack surface of adversarial machine learning [KKU+22]. XAI also improves the transparency and accountability of black-box AI models and can be used to explain any deterioration in the performance of these black-box models to proactively identify issues in the model training or issues in the data quality. That means, the incorporation of transparency-enhancing XAI techniques can not only strengthen defensive strategies, but also foster trustworthiness by demystifying the decision-making processes of the AI/ML system. This transparency is crucial for ensuring the reliability of 6G communication systems [ITM+21].

5.2.2 AI privacy preservation

Privacy concerns especially increase when there is sensitive data in the AI/ML system. The sensitive data can be the training data, the inference query, the inference result, and the AI/ML model. More precisely, for example, there can be some privacy concerns about sharing the data with the entity that will execute the model training operation using the data. Another example can be that the consumer of the AI/ML inference service may want to keep the inference query and the result secret from the AI/ML service provider. The inference service can also leak information about the AI/ML model, which cannot be acceptable to the service provider. Another privacy-related concern can be the following. Since the AI/ML model is the output of AI/ML model training that uses sensitive training data, the AI/ML model somehow carries this sensitive information. When there are some cases where the AI/ML model needs to be transferred and shared with some other parties or the AI/ML model is provided as an inference service, there can be some potential leakage about the training data used for the AI/ML model training. Membership inference attacks can be given as an example threat.

When we consider the 6G system as a data-driven, multi-vendor and multi-environment system, each use case of AI/ML usage needs to be analysed to identify which kind of threats are valid for the use case and then appropriate protection mechanisms should be considered. Widely used technologies to address privacy aspects are secure multi-party computation, homomorphic encryption, differential privacy, and confidential computing. Also, there are some privacy aware solutions such as federated learning and split learning, but they also need support from privacy enhancing technologies because pure usage of these solutions may not be enough against privacy attacks.

One important point to consider is that focusing only on the privacy aspect can make the AI/ML vulnerable to security attacks, because with privacy solutions the data or local model updates coming from the data owners will be hidden from the AI/ML model trainer. Thus, the AI/ML model trainer will not be able to detect and prevent security attacks towards the training operation. For example, in the federated learning case, when secure aggregation is used for privacy, the server cannot analyse the local model updates to detect poisoning attacks. One approach to mitigate both the security and privacy attacks can be to allow the central AI/ML trainer (the server) to access some pieces of the local model updates in cleartext. With that approach, the server will be able to detect security attacks without learning any information about the training data from the local model updates.

Another consideration about application of privacy and security mechanisms would be taking the possible overhead on performance and speed of learning process into account. In private federated learning for example, it is usually desirable to apply differential privacy only on parts of the model and not all of it. This usually preserves privacy while providing better learning speed and performance.

5.3 Enablers for trust infrastructures

In the orchestration and management of modern mobile networks, Key Performance Indicators (KPIs) and Service-Level Agreements (SLAs) have become foundational concepts. These KPIs, such as throughput and delay, are objective, strictly defined, straightforwardly measurable, and can be assured. This framework can be formulated under the concept of “assurance”, representing a well-established and objective approach to network management [NIST22].

With the advent of 6G, a new dimension that extends beyond traditional assurance metrics emerges. Trustworthiness becomes a critical KVI to guarantee E2E system resilience, reflecting the broader 6G vision of connecting human, digital, and physical worlds [HEX21-D12]. Unlike assurance, trust is subjective and varies among different parties, making it a distinct and equally vital concept in the 6G landscape [NIST22].

Recognizing the need to bridge the gap between assurance and trust, Hexa-X-II introduces the proposal of Trust-Level Agreements (TLAs). TLAs are envisioned to extend beyond traditional SLAs by incorporating the Level of Trust (LoT) of various stakeholders, aligning with the unique demands of 6G. The implementation of TLAs presents several key challenges, especially regarding two main aspects.

On the one hand, LoT monitoring and assessment are required to address trust subjectivity, without a universal standard, in contrast to traditional KPIs, objective and measurable. Creating objective, constrictive terms regarding trust in the agreement requires a universally agreed standard for LoT assessment. This assessment must also consider indirect/social trust or “reputation”, influenced by third parties. This poses challenges to the monitoring mechanism design regarding real-time feasibility, data integrity, security, and privacy. Establishing a standardized framework for trust assessment, including universally agreed standards and metrics (e.g. [ISO/IEC 5723:2022]) and multi-stakeholder consensus, is essential to overcome this challenge, as addressed with the Hexa-X proposal of a LoT Assessment Function (LoTAF) [HEX23-D13].

On the other, new technology pillars are expected to contribute as the solution for assurance and trust in 6G networks. These include but are not limited to Trust as a Service (TaaS) approaches (providing a standardized framework for trustworthiness assessment, addressing the challenge of objective measurement [HEX23-D13]), ML/AI-driven threat detection (allowing dynamic trustworthiness evaluation and addressing the challenges of real-time feasibility and data integrity), Distributed Ledgers, DLT, offering transparency, immutability and non-repudiation, can be utilized to define TLAs and liability of each party, and protect the integrity of AI data [HEX21-D12]), Privacy Preserving Data Publishing (aligning with the shift towards trust-centered frameworks, ensuring confidentiality and privacy), and attestation mechanisms (making identities only bound to the possession of a cryptographic key but to the verification of the overall status of the component).

In conclusion, the transition from classical assurance based on KPIs and SLAs to a trust-centered framework in 6G requires a multifaceted and innovative approach. The proposal of TLAs paves the way for a more secure, resilient, and trustworthy 6G ecosystem. Ongoing research and collaboration among stakeholders will be essential to overcome the challenges and realize the full potential of 6G in delivering a new level of QoE for users [NIST22].

5.3.1 Evolved cryptography

The use of cryptography in today's networks is pervasive, much beyond the common use of TLS for E2E security, and including all aspects related to infrastructure management and network support systems. In this context, two main trends can be identified. First, and the one most widely discussed recently, we have the threat associated to the vulnerability of current crypto algorithms to quantum computing, becoming much more than a theoretical risk [Mos15]. In addition, the generalization of virtualized and cloud-native solutions requires a comprehensive approach to identity infrastructures and key management procedures, very much focused on trust fabrics supported by static procedures and relationships, challenged by the virtual nature of current network functions.

The evolution of mobile network infrastructure requires solutions able to support quantum-safe and cloud-aware crypto mechanisms able to satisfy two main requirements:

- **Agility**, as the possibility to maintain alternate evolution paths as algorithms and technologies evolve.
- **Pliability**, as the capacity of being adapted to network management best practices.

There is a number of relevant trends in the evolution of the cryptography application in mobile networks that are worth exploring:

- The opportunity of taking advantage of the urgent transition to quantum-safe schemas to restructure the trust fabrics and their PKI foundations, aligning them with cloud-native considerations, including identity and key management.
- The proposals to address *crypto agility*, such as the so-called hybrid certificates (including signatures and keys for different crypto algorithms) and the use of non-PKI mechanisms for establishing session keys.
- The extension of operational mechanisms addressing users and their privacy, addressing other network elements, extending identity management to include matters related to attestation and supply chain verification, or to data provenance.
- The choices to bring key and identity management to scale, such as the use of automation procedures based on ACME [BHM+19], simplified revocations based on ACME-STAR [SLG+20], or the use of alternate roots of trust, including self-sovereign identities.

In this context, we plan to execute experiments for analysing how these technologies can be applied to enhance 6G network security, and the results of research and industry initiatives attempting to make the technology evolve to support quantum-safe networking, exemplified by:

- The recently launched European initiative on quantum cryptography, [QSNP], and the recent Horizon Europe projects on PQC, [QUB23], and [PQR23].
- The IETF groups concerned with crypto evolution, especially [PQUIP], [LAMPS], [ACME], and [QIRG].
- The ETSI [ISGQKD] and the TC-CYBER group on Quantum-Safe Cryptography [ETSIQSC].
- The GSMA initiatives on QKD [GSMAQKD] and PQC [GSMAPQC].

The results of these experiments will become feedback to the relevant groups, contributing in this way to the (pre-)standardization work in the area.

5.3.2 Distributed ledgers

Distributed Ledgers (that we will refer as DLT (Distributed Ledger Technology), to avoid confusion with the acronym related to "downlink") are systems that aim to distribute the information storage among the participants and to add a level of security to the stored information by linking the stored pieces. Commonly referred with the term "Blockchain", a DLT is a distributed, secure, and public infrastructure that makes its stored data traceable and immutable. Its main disadvantage is the time and energy required to process and store data in scenarios where latency is important and the possible doubt of certain players to make their information completely public. While on the first aspect, the solution is the evolution of new consensus mechanisms to validate and store data faster, the solution towards the second aspect is the use of Permissioned (i.e., restricted) DLT (PDL) systems such as Hyperledger [Lin22]. Permissioned Blockchain are those networks composed of a limited number of peers with granted access to the stored data, limiting the vision of the information but keeping the rest of the public Blockchain aspects (i.e., consensus mechanism, the use of smart contracts, etc.). By applying a permissioned model, some specific players may share data with other equal peers (e.g., between service providers from different domains) and avoid those other players (e.g., service consumers) may access it.

Beyond the possibility to generate and manage trust in a distributed way that is at the core of DLT, a second layer of trust lies on the definition and evaluation of specific trust parameters associated with each actor involved in the service orchestration and management. To do so, there are two elements that are being proposed in the research literature [AMM+23] and standards [ETSI_PDL_015]. On the one hand, the use of reputation parameters to compute a trust level value and, on the other hand, the definition of trust requirements by means of TLAs. There are multiple works considering how to use and compute trust values, but the most common way is to do it by computing multiple reputation-based parameters and joining them to generate a single value that defines the trustworthiness of an entity. Then, the clients may use the TLA to define the reputation and trust requirements that a provider must fulfil to be selected for the service provisioning.

To assist with the management of the elements previously described, especially the management of SLA, TLA and service resources, the PDL based systems offer a certain level of automation via smart contracts. A smart contract is a small piece of code that may be executed when certain conditions are fulfilled. This way, while the peers are taking care of their internal tasks, they may trigger some cooperative tasks with the other peers. The use of smart contracts brings different benefits but also possible security threats [ETSI_PDL_011]. The main three groups of security threats are: programming errors, internal threats (e.g., transactions ordering, malicious/accidental executions or reporting the wrong parameters) or external threats such as malicious attacks, accidental damages, denial of service and other possible options.

Together with the definition of TLA and their management, DLT solutions could assist to the trust management also on what is called decentralized identity [ETSI_PDL_019]. This allows to avoid the need to have (again) third parties to manage the identities and reduce trust-related threat risks on authentication processes such as data leakage, identity tracking, credential stuffing, etc.

Based on all the previous possibilities, DLT solutions should be a key element towards some possible scenarios such as: a) to give support towards connectivity scenarios "without infrastructure" (e.g., ad-hoc or personal networks) with DLT-based identities, b) a roaming case of applications hosted in cloud-edge environments from multiple operators (e.g., V2X services while travelling), c) to store data from components attestations and results verification in order to verify aspects related to the supply chain, and finally, d) the use of smart contracts to define and verify requirements related to any kind of agreement (e.g. SLA or TLA).

5.4 Physical layer security enablers

To increase the trustworthiness of the mobile network the current architecture incorporates or proposes a range of security measures and controls. Many of them are based on cryptography. It is currently envisioned that the set of existing security measures will be extend by methods from the domain of physical layer security leading to an increasing number of design and deployment options regarding achieving desired security goals. But selecting appropriated security controls is not only driven by the security goals, since most (if not all) security controls induce certain costs (e. g., in terms of increased energy consumption, reduce goodput or increased latency) and rely on certain assumptions (e. g., mathematical assumptions or the non-existence of sufficiently powerful quantum computers). Therefore, the overall selection process can be understood as a trade-off or optimization between the desired LoT and other non-functional as well as functional requirements. The relevant decisions must be made not only during the design and development or deployment phase but also during the runtime. For example, methods for Physical Layer Security (PLS) in particular are dependent on the current conditions of the transmission channel.

5.4.1 AI-enabled attacks

AI can as well be used on the attacker side to support attacks in various ways. Characteristics of AI-enabled attacks are evasiveness, pervasiveness, and adaptiveness [CEPS21]. The assistance by AI has the potential to expand existing threats, to introduce new types of attacks [TMR20] or to change the typical character of threats [BAC+18]. Given the potentially augmented risk and impact of AI-enabled threats, it is important to leverage all available threat detection and mitigation capabilities, in particular the AI-empowered security methods. For the ongoing fight of "good AI" versus "malicious AI", significantly further research is needed.

5.4.2 Context awareness and PLS adaptivity

Context awareness and related (self-)adaptivity methods deal with the trade-off or optimization between the desired level of security and other non-functional as well as functional requirements. The fundamental concept is that the system "understands" the desired (security) goals and the current situation and adapts the security controls, accordingly, selecting and configuring them to achieve the goals in an optimal way, which could include graceful degradation.

Applied in the domain of PLS (and here more specifical PLS-based secret key generation methods) this would imply that the secret key generation rate is adapted according to current and past channel measurements. If e.g., these measurements reveal that the channel is rather static with a strong line of sight component, the key generation rate is usually much lower compared to highly dynamic settings with e.g., lots of interference and reflections. The achievable key generation rate in turn might influence the cryptographic algorithm selected for encryption of the messages. If the key generation rate is above the needed message bandwidth a one-time pad like encryption can be applied whereas if the key generation rate is lower than the required message bandwidth a (less secure) pseudo one time pad construction might be selected.

5.4.3 Security and privacy in joint communication and sensing (JCAS)

Sensing is a process that acquires real-time data about the physical environment [HEX21-D31]. In the context of 5G, sensing has already started to manifest as a significant feature. However, the full potential of sensing is projected to be unlocked with the advent of 6G technology. This will involve an even more precise, ubiquitous, and automated form of data acquisition, empowering a multitude of use-cases spanning from industrial automation to advanced remote health monitoring. 6G innovation in Joint Communication and Sensing (JCAS) needs to be studied to examine early state of JCAS to understand the security and privacy challenges to the JCAS system.

In order to analyze the JCAS system from a security standpoint, we take a broader higher-level view of the system. The Figure 5-2 presents the logical system view, not focused on a particular deployment in mind. In the figure we can see the system separation into four main parts. First, we have the physical environment that we are observing in some ways. Producers are sensors/entities making these observations and producing raw information. The dataflow network then extracts desired level of processed information that can be used by consumer applications. This is an abstract model which will change with given application needs e. g., presence of moving objects versus semantic classification of all physical objects in the scene.

From the JCAS system view, we can extract four main categories of assets: sensing information, systems, value provided to the operator from the sensing service and value provided by sensing applications.

- **Sensing information** contains data about the observable physical environment surrounding mobile radio networks. The information varies widely per sensor capturing it, processing done on it and the

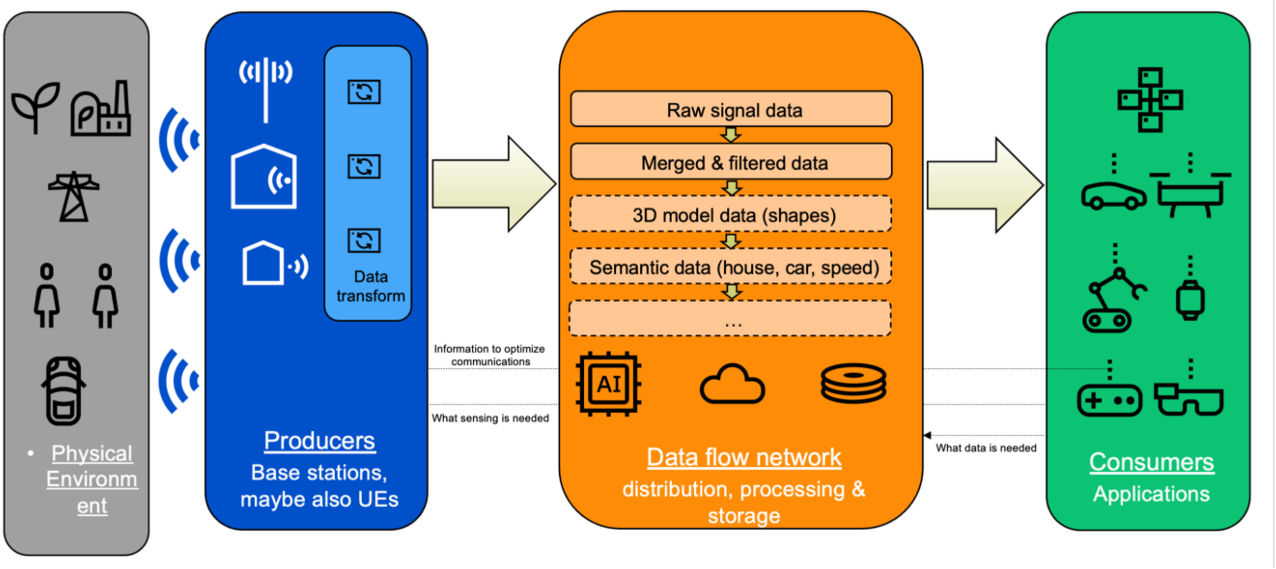


Figure 5-2: JCAS system early overview.

use case. The information includes raw sensing data received directly from the sensor with none or minimal necessary level of processing, processed sensing data that has been extracted from raw sensing data via various processing steps based on the application need, and data about sensing consumers collected to resolve sensing requests for compensation. This list is not exhaustive and other categories could and should be considered in future, for example relevant sensing metadata such as time and location, the processing pipeline description or hierarchy, sensing requests, policies and more.

- **Systems** form a chain of assets covering the whole path from sensing producers to consumers, end users and possibly beyond, such as sensing data generation and processing, sensing control and management, networking, access control and more.
- **Value provided to the operator from the sensing service**: The sensing service needs to generate some value to the operator and as such it is perceived to be an asset. This might be handled in a different way than current subscriber billing.
- **Value provided by sensing applications**: The value provided by sensing applications includes safety, information, efficiency, and similar assets.

Since assets are the valuable entities of the system that needs to be protected from adversaries, identifying assets in the JCAS system helps to focus security efforts on these entities. The assets (processes, sensing data, physical devices, actors such as producers and consumers) and their interactions can be utilized to create a data flow diagram of the JCAS system. Each component of the dataflow network can be assessed thoroughly for any security and privacy threat using threat modelling tools such as [Mic22], [LINDDUN], etc. We reserve the extensive threat and risk modelling work to a later stage when the JCAS system and its architecture will reach a more mature stage. From the high-level system view and JCAS capability we put forth below points related to security and privacy aspects of JCAS.

Two distinct areas are identified in the security domain of JCAS: using sensing for security applications and securing the sensing process itself. In this work, we focus on securing the sensing process.’However, it's important to note that some solutions could benefit from enhancements in the use of sensing by security applications.

5.4.3.1 JCAS security

Sensing and positioning are being explored in terms of their Key Value Indicators (KVI), particularly trustworthiness, underlining the need to secure both the information generated and the processes that generate it. Confidentiality, Integrity, and Availability (CIA) are all required to realize trustworthy JCAS solutions. Confidentiality protects sensitive data that could be an attractive target for malicious actors. It may be hard or not feasible not to sense personal sensitive information on radio level. Availability is important due to the increasing dependency on network sensing, especially in critical and public safety applications. Integrity is also critical in public safety and following the overall increase in dependency on sensing and communication infrastructure. Introduction of sensor fusion expands the possible use cases but also threats. The introduction of third parties for processing further increases the need for integrity. In the CIA triple, confidentiality and integrity can be achieved by a proper use of cryptography (with specific CAS requirements), while availability assurance requires further research.

Despite sensing reusing the communication platform different applications will need adaptation of different security requirements [HEX23-D33] and will impose different levels of privacy risks. The security framework for each application differs as different types of data and functions need to be protected with different levels of requirements. Their interaction further changes the threat dynamics.

5.4.3.2 JCAS privacy

Data collected by JCAS about users and passive subjects raises privacy concerns in matters of strong social sensitivity. So, the privacy issues in JCAS should be well understood in advance. In the framework of JCAS, privacy is about the collection, handling, ownership, and protection of personal data linked to and extracted from sensing measurements and results. This includes personal data as well as meta-information of active participants and of other passive subjects physically present in the sensed area.

Given the sensitive nature of sensing data, ensuring user privacy becomes a paramount consideration in the design and operation of JCAS. It is vital that the evolution of this technology respects privacy rights and operates within the boundaries of established legal frameworks to avoid misuse. In this respect, insightful lessons can be gleaned from existing technologies and systems, such as CCTV networks, public radar systems, automotive radars, and Wi-Fi sensing security research. Also, the methods of privacy protection that can be applied to sensing functionality in 6G need to be studied.

Collecting consent is another big challenge in JCAS. The General Data Protection Regulation (GDPR) defines in its article 6 [GDPRA6] six lawful bases for data processing: consent, contract, legal obligation, protection of vital interests, task in public interest, or legitimate interests. Some of them may be simpler to support technologically in JCAS than others.

The scope of using consent in JCAS scenarios is currently unknown. The rules and requirements obtaining and maintaining consent make it unviable for large-scale collection, especially in public places such as JCAS deployments. In general, obtaining consent from sensing targets is difficult or impossible, due to the data collection indirection. And even if consent can be guaranteed, it would need to be further managed to comply with data subject’s right to access, right to opt out and be forgotten. Additional questions rise when a sensing target does not consent or withdraws at later stage; what happens to the potentially large amount of data that has been already collected and perhaps processed. Furthermore, consent needs to be explicit to given processing thus reuse of sensing measurements for multiple purposes is challenging. And setting consent policies to tackle some of the challenges is against the main principles of GDPR. These challenges may differ in complexities based on local regulations.

Therefore, use cases that fall under legal obligations, protection of vital interests, and legitimate interests could see earlier adoption. Legitimate interest can be applied quite widely and could be leveraged in private network settings, but individual use cases need to be well considered.

5.4.4 Perception of physical anomaly sources

By its nature, wireless communications operate in an open medium, making it vulnerable to interference, no matter whether it is intended or unintended. While the latter typically poses only a small risk in licensed spectrum scenarios, where operators maintain control over interference, intended interference also known as jamming has the potential to break the communications. The threat of jamming is not new for 6G. However, with former generations of mobile communications, which preliminary served non-critical services like phone calls, text messages, and multimedia, the impact and damage of jamming was relatively low compared to the effort. Only a large-scale attack would result in severe consequences from an operator perspective but would require significant effort. In contrast, when mobile communications are foreseen to also serve critical applications, e.g., in autonomous driving, manufacturing, or even health applications, jamming and interrupting small parts of the network could lead to significant economic, physical or even human damage.

As a conclusion, the motivation to jam networks for malicious reasons is much higher in 6G (or critical applications in particular) than before, and critical applications cannot be realized as long as they are prone to relatively simple jamming attacks. Furthermore, opportunities for jamming have evolved as an effect of the maturity of available Software Defined Radio (SDR) technology and open-source software.

As a preliminary step to jamming signal management, the classification of jamming signals is considered to play an important role in the design of a resilient communication system. Moreover, the localization of the jamming sources (e.g., [BNS+23]) is helpful for initiating countermeasures of the jamming sources. With the obtained information on the existing jamming sources, the system can respond more rationally to intentional interference signals.

The jamming signals can impact the quality of communication or even disrupt the communication. The detection of the jamming signals can be performed by analysing certain parameters of the receiver. Making use of the transmission parameters measured over a period of time, the presence or even the type of interference signals can be determined. A multi-stages-based classifier is proposed in [XTZ+05], where the abnormal status of the packet delivery ratio, the consistency of the received signal strength, and the location information are sequentially analysed. Comparing the measured results with the pre-defined thresholds, the existence of the jamming signals within the communication environment and the active period of the jamming signals are detected. Replacing the manual comparator with a deep learning network, the jamming signals can also be identified [KJS20]. The limitation of the aforementioned parameter-based jamming classification method is that it only analyses the time-domain characteristics of the signals and neglects the information in the frequency domain. Further considering the signal behaviour in both frequency and time domains, time-frequency analysis is employed. After processing the received signal using the time-frequency analysis tools, such as short time Fourier transform [MFL19], wavelet [TGE+19], Wigner-Ville distribution [KCH+22], etc., the time-frequency features of the signal are captured. With the usage of deep learning algorithms, the features of the signal are employed to detect the existence of the jamming signal and classify the time-frequency characteristics related jamming type.

With the rapid development of deep learning algorithms, the classification accuracy of the jamming signals can reach up to 99% when employing models such as VGG, ResNet [YJJ+22], and transformer [LG22]. After the identification of the interference signals within the communication environment, the next challenge lies in effectively leveraging this information to enhance the system’s reliability. To further facilitate the integration of interference signal management algorithms into the system, the idea of introducing a lightweight model is proposed in [LG22]. The limitations of the current works are as follows:

- The design of resilient communication systems is only for the presence or absence of interfering signals, but the class of interfering signals is not considered. The narrowband jamming signals within the Wi-Fi environment are classified using a convolutional neural network in [RUO+23], while the other types of jamming signals are not considered.
- The developed countermeasure strategies for interference signals can only be implemented when special conditions are met. However, the practical implementation of the techniques outlined in [ALM+23] necessitates the availability of channel state information at both the transmitter and the receiver, which poses challenges in real-world systems. In an attempt to provide a more comprehensive solution, the authors of [MLN23] propose an alternative approach. They optimize the channel selection algorithm based on monitored interference information, but its applicability is limited to frequency hopping spread spectrum (FHSS) systems.

Therefore, it is a new problem to design a resilient communication system that can cope with multiple types of interference signals.

To eliminate the influence of the jamming signals, the localization of the jammer is helpful for removing the anomaly nodes from the communication environment or further analysing the source of the jamming signals. With this two-stage based jamming management strategy, the information required for initiating jamming countermeasure is considered to be obtained. Identifying the type of jammer has the advantage of providing a first starting point for implementing countermeasures. On the other hand, with supervised learning methods, there is a risk that jammers with characteristics not included in the training data will not be detected [ZMH2010]. This motivates to also investigate unsupervised learning approaches which require only a minimum set of assumptions on possible jammers, e.g., that a jammer needs to transmit an arbitrary signal to interfere with another wireless communications link.

For this, a radio environment monitoring framework for 6G networks capable of detecting anomalies such as intended and unintended interferers is proposed in [KKS+23]. The framework is based on a digital twin (DT) of the radio environment. A DT is a digital representation of a physical system, further explanations can be found in section 5.5.3. The radio environment as considered here consists of transmitters as well as the physical environment (obstacles and their materials, etc.) and propagation characteristics [PJK+14]. The digital representations of the components can also be found in the DT. The framework is based on the following assumptions:

- The framework operates in a licensed band, where the location of each regular transmitter connected to the network is known at a central unit (CU). Of particular interest are non-public networks which require a high level of resilience and operate in a licensed band in a strictly limited area.
- The physical environment is known. In case of an indoor scenario the physical environment means obstacles and walls. The database on the physical environment might originate from JCAS or from any other source.
- Sensing units (SUs) which measure the received signal strength (RSS) are distributed in the area that is monitored. The SUs have a feedback link to the CU.
- The CU has a deterministic propagation model (e.g., ray tracing or ML-based) available to run a DT of the radio environment which is based on the known regular transmitter locations and the database of the physical and environment. In particular, this means that the expected RSS at the locations of the SUs can be estimated in the case of normal operations.

By comparing the expected RSSs at the locations of the SUs with the actually measured RSS values, the system is able to detect sources which contribute additional energy in the monitored band, i.e., intended interferers (jammers) or unintended interferers. Unintended interferers might be for example – considering non-public networks – neighbouring non-public networks which are wrongly configured (e.g., regarding the time synchronization) or violate regulations.

To be able to detect anomalies of the radio environment despite the deviations between the digital and the real-world counterpart – the so-called physical twin (PT) – that occur anyway (see validation in section 5.5.2), approaches need to be researched that are able to distinguish whether a deviation between the DT and the PT originate from model imperfection or whether an anomaly is present.

The proposed approach can identify anomalies in the radio environment such as mentioned above without any assumptions on their characteristics but only by the knowledge that an interferer will lead to an observable difference between the expected and the measured RSS. Furthermore, the proposed approach exploits context awareness, e.g., knowledge on regular transmitters and their location and thereby integrates in the ongoing research on employing DTs for comprehensive monitoring of wireless networks (e.g., [LKD+23]).

5.4.5 Physical layer deception

Conventional PLS solutions primarily aim to prevent potential eavesdroppers from decoding transmitted information, adopting a passive defense approach. A novel paradigm, Physical Layer Deception (PLD), is introduced in [HZS+23]. This approach represents an active defense strategy at the Physical layer, designed to deceive potential eavesdroppers by transmitting compromised information while simultaneously delivering the original data to authorized receivers.

The foundational principle of PLD relies on a meticulously designed symmetric block encryptor. In this scheme, the codeword sets for both plaintext and ciphertext are identical. Specifically:

- For any feasible pair of ciphering key and plaintext, the resulting ciphertext must also be a feasible plaintext.
- For any feasible pair of ciphertext and key, the decrypted plaintext must be feasible.
- For any feasible ciphertext, decrypted plaintexts using different feasible keys must be distinct.

In the absence of encryption, messages are transmitted in plaintext without any multiplexing with a ciphering key. When encryption is enabled, each message undergoes encryption using a randomly selected ciphering key, followed by power-domain multiplexing with the key for transmission. Radio resources are allocated such that:

- The authorized receiver can reliably decode both the ciphertext and the ciphering key, thereby successfully obtaining the original plaintext.
- The potential eavesdropper can reliably decode the ciphertext but experiences a high error rate in decoding the ciphering key, leading to a high likelihood of receiving incorrect plaintext—effectively being deceived.

On the user plane, PLD safeguards data privacy by feeding eavesdroppers false user data, achieving at least the same level of security as classical PLS. In specialized use-cases, such as police or military communications, deceiving adversaries with compromised information may offer advantages over merely blocking their access to data. Furthermore, PLD potential extends to the control plane, where compromised signaling messages could lure eavesdroppers or man-in-the-middle attackers into revealing their presence, thereby enabling their detection.

5.5 Validation mechanisms

The scenarios, threats and specific enablers described above require to be validated, providing evidence of their specific requirements and applicability. This needs to be done in parallel with the development of the proposed 6G architecture and enablers, so we can apply the *security by design* principle. Applying this principle, it will become possible to avoid the common situation of adding security features to existing design, what translates into intricate privacy and security solutions, difficult to be applied by users and service providers, unnecessarily extending the threat surface.

In order to obtain this evidence on security, privacy and resilience properties in the early stages of design, we propose the use of early validation mechanisms to evaluate the preservation of these properties and contribute to architecture definition. The project is committed to apply two kinds of validation mechanisms: the use of simulation for E2E resilience assessment and the evaluation of anomaly detection at the physical layer, and the application of a Network Digital Twin (NDT) environment for evaluation of security and privacy threats and enablers.

5.5.1 Simulation-based resilience assessment

As an essential step for implemented system PoCs validation, network resilience performance assessment can be carried out in a simulation-based environment. The main focusing aspects of this environment are the threats that impact network performance, such as network element failures and traffic changes. Since 6G is becoming complex and even a network of networks, and 6G is connecting more and more new devices and device types, many of these threats become fatal and seriously impact the network, and these impacts can propagate from one part of the network to the rest of the network. To efficiently estimate network resilience, the network can profit from a simulation-based resilience assessment environment.

The assessment is based on discrete-event simulation. Indeed, in the environment, the service packet processing, network failure, network recovery, and management are described by mathematic representation event modules, the *Petri Nets* [Pet77]. They are modelled as events and will interact with each other. On top of these modules, the environment also includes a model of the communication service delivery that connects all the modules, including RAN, TN, and CN modules. The considered RAN and CN are fully virtualized.

Threats are injected either as events or as irregular behaviors of the events. Till now, two types of threats have been considered. One is the internal failure of the network, represented by the network failure module [LDB+22]. The other is the external threats of traffic variation (peak) [LDB+23]. The environment can be extended to adapt to other threats, especially those identified by different WP.

The goal of the environment is to obtain KPIs and KVIs from simulation. The threats and network may behave with randomness/stochastics. A large number of simulations is envisaged to generate more variant risk scenarios and get a more accurate estimation of the performance. The real KPIs and KVIs can then be approached using the average of the simulation results. Until now, different indicators can be estimated from the environment, including availability, reliability, E2E latency, and packet loss rate. The indicator can also be a combined one by considering various SLA metrics. The resilience can be eventually estimated using the *Resilience Triangle* [BCE+03].

For the next step, this environment will be adapted to the system PoC architecture provided in chapter 7. Various scenarios will then be included. KPIs and KVIs from other WPs will be further included in order to prepare the environment for the PoCs validation.

In parallel, another plan is to study the possibility of connecting or implementing the environment to the DT approach so that these two methods can efficiently work together.

5.5.2 Validation mechanisms for perception of physical anomaly sources

To validate the proposed perception concepts in section 5.4.4, the listed technical solutions are planned to be evaluated in a simulation environment. The anomaly sources perception consists of jamming localization and jamming classification/detection. Each of the two stages will be firstly validated separately. After confirming the effectiveness of the proposed technical solution, a more sophisticated simulation environment will be built so that the two stages can be concatenated and evaluated.

The first stage is about the localization of the jamming sources. The purpose of this stage is to find the sources within the assumed communication environment using the available measurements from different access points. After selecting the appropriate measurements according to their sensibility to the jamming sources, the measurements are going to be classified using either conventional approaches or machine learning methods.

The second stage is about the classification of the jamming signal, which is considered to be the following step of the jamming localization. Similar to stage one, the feasibility of using measurements or the raw received signal sequence for jamming classification shall be evaluated. The impact factors for generating the required input of the classifier should include the employed modulation scheme, the channel model, the other potential interference, and the background noise. Afterward, the deep learning classifier will be selected by evaluating its effectiveness in dealing with the classification task. The aforementioned evaluation steps shall be conducted using MATLAB and Python.

For further practical experiments, the possibility of using an access point to capture the required measurements or the I\Q sequences should also be checked.

For the DT-based anomaly detection approach mentioned in section 5.4.4, a simulation framework is planned to evaluate the performance. The framework shall roughly represent a non-public indoor network in a factory environment using a licensed band at a carrier frequency of 3.7GHz. In an initial simulation setup, general results shall be obtained by employing a log-distance path loss model with random shadowing. All transmitters shall use isotropic antenna patterns. The challenge of the DT-based anomaly detection approach is the fact that even though trying to accurately replicate the physical twin, there will always be some deviation between the digital twin and its physical counterpart due to imperfection of available data or the used models. For the first step, inaccuracy in localization shall be considered as well as the random shadowing which mimics inaccuracies in the data on the physical environment. In the simulation-based validation, a high number of samples (each sample comprises the radio environment at one time instance) shall than be created, whereby in each sample a number of regular transmitters and eventually a jammer are present. The DT of the radio environment is then constructed by estimating the regular transmitter positions (this is mimicked by adding random offsets to the transmitter positions) and applying the log-distance path loss model to obtain an estimated radio environment map, i.e., a DT of the radio environment.

As mentioned previously, there might always be a deviation between the PT and the DT due to model imperfections. Thus, approaches which are capable of identifying whether deviations originate from model imperfections or whether there is an anomaly need to be evaluated in the validation phase. For this task, unsupervised learning seems to be a promising approach. In the learning phase, normal data (i.e., no anomaly is present) are fed to the learning algorithm to generalize the characteristic of the normal data. In operation phase, the algorithm shall than decide whether incoming have the same characteristics as the normal training data or if there is a difference, i.e., an anomaly.

In the next step, the simulation framework shall integrate ray tracing to generate synthetical data. Thereby, the scenario shall still orient on a factory hall. The focus of the enhancement step will be to see if the findings from employing the log-distance path loss model for simulations can be extended to a more complex scenario, bringing the framework closer to a real-world deployment. With integrating raytracing, the approach shall then also be truly capable of exploiting information on the physical environment.

5.5.3 Network digital twin (NDT)

A digital twin works with the same (virtual) elements of a real environment (UEs, forwarders, controllers, network functions, etc, in a network), has direct access to (live or recorded) data from the real system, and can incorporate real-time data from sensors and other sources. Digital twins are used to generate data evidence on the effect of different actions on the real environment, making predictions about future performance, and possibly providing control actions to be applied.

The development and creation of a digital twin is associated to three main stages:

- Digital Twin Prototype (DTP): before creating a final physical product, a digital one is made to see what it would look and behave like.
- Digital Twin Instance (DTI): once a product has been produced, the digital twin is used to test different usage scenarios with the virtual rather than the real one.
- Digital Twin Aggregate (DTA): one (or several) twin collects information from the real environment to determine the capabilities of a product, run forecasts and test operational parameters.

By replicating real assets, frameworks, and operations to produce continuous data, a digital twin allows industry to anticipate downtime, react to changing circumstances, test design improvements, evaluate security threats and mitigations, and any other tasks related to impact (including risk) assessment and data collection.

The IETF proposed a definition of an NDT [ZYD+23] as a virtual representation of the real network. Such virtual representation of the network is meant to be used to analyse, diagnose, emulate, and then control the real network based on data, models, and interfaces. To that aim, a real-time and interactive mapping is required between the real network and its virtual twin network. Referring the characteristics of digital twin in other industries and the characteristics of the networking itself, the NDT should involve four key elements: data, mapping, models, and interfaces as shown in Figure 5-3 below.

- **Data**: A NDT should maintain historical data and/or real time data (configuration data, operational state data, topology data, trace data, metric data, process data, etc.) about its real-world twin that are required by the models to represent and understand the states and behaviors of the real-world twin. The data is characterized as the single source of “truth” and populated in the data repository, which provides timely and accurate data service support for building various models.
- **Models**: Techniques that involve collecting data from one or more sources in the real-world twin and developing a comprehensive representation of the data (e.g., system, entity, process) using specific models. These models are used as emulation and diagnosis basis to reflect the dynamics of the live real network operation, and to generate reasoning data utilized for *decision-making*. Various models such as service models, data models, dataset models, or knowledge graph can be used to represent the real network element and then, instantiated to serve various network applications.
- **Interfaces**: Standardized interfaces can ensure the interoperability of NDT. There are two major types of interfaces:
  - The interface between the NDT platform and the real network infrastructure.
  - The interface between NDT platform and applications.

  The first one provides real-time data collection and control on the real network. The second one helps in delivering application requests to the NDT platform and exposing the various platform capabilities to applications.

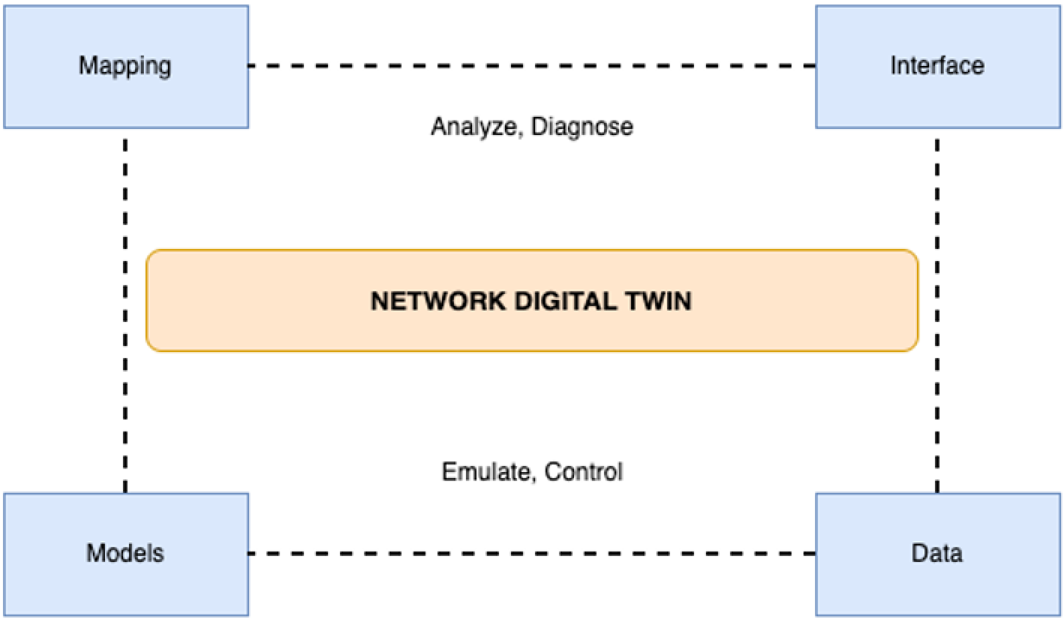


Figure 5-3: Elements in a network digital twin (NDT).

- **Mapping**: Used to identify the digital twin and the underlying entities and establish a real-time interactive relation between the real network and the twin network or between two twin networks. The mapping can be:
  - One to one (pairing, vertical): Synchronize between a real network and its virtual twin network with continuous flows.
  - One to many (coupling, horizontal): Synchronize among virtual twin networks with occasional data exchange.

Two phases of the methodology for the development of the Digital Twin can be identified, as described below:

In a first phase:

- **Data Acquisition:** Data collection of the scenario to be emulated. The main objective of this phase is to obtain topological information, information about hardware resources presents in the network, route tables, cost metrics between links, bandwidth measurements, and specific parameters of the elements that make up the network.
- **Data Modelling:** Once the previous step is complete, the information collected by the agents using various technologies and tools (ALTO, SNMP, ICMP, Iperf3, OpenFlow, etc.) is stored on the central computer. In this process, correlation and data inference are applied to obtain a semantic representation of the network that includes as much detail as possible based on the data collected. The aim of this phase is to obtain a simplified model of the real network and to avoid a complete replication of the real network as a digital twin, which would not be effective from the point of view of cost and resource optimization.
- **Data Adaptation:** Once the network data modelling stage is complete, a subset of information is extracted from the data model about the physical parameters required to deploy the virtual resource in the NFV architecture that will host the Network Digital Twin (hardware resources, software resources, network interfaces, network segments, etc.). The final outcomes of this phase are the descriptors (scripts) of network functions, the configuration, and the topology of all the elements of which the virtualization infrastructure is made up.

- **NDT Descriptors:** Set of scripts obtained from the three previous stages. It consists of a series of descriptors of both, the real network to be emulated and the parameters to be monitored during the “experiment”, as well as other relevant information.

In the second phase:

- **Network Digital Twin Deployment:** The Network Digital Twin descriptors mentioned above are the virtual infrastructure manager inputs. A set of software agents (VDUs) are deployed and managed within a virtualized infrastructure by an orchestrator.
- **Integration:** At this point, the NDT is set up and deployed, but it still does not meet one of the key requirements of the digital twin technology, which is the connection between the real twin and the digital twin. If this requirement is not met, we would only be dealing with a conventional network emulator that does not have real-time processing capabilities. This phase presents two major challenges: the security of communication between the twins and the management of communication between the dual elements within each twin. At this point the NDT is able to consume data from the real network in order to update its topology and/or configuration.
- **Closed loop:** The last phase consists in activating a feedback loop between the two twins. The twins start monitoring tasks to exchange information to perform higher-level tasks by applying artificial intelligence techniques, such as topology optimization, dataset acquisition, preventive actions in stressful situations in the network, etc. This last stage would end with the highest level of digital twin sophistication.

Given the status of the 6G technology development and the goals of the project, the NDT to be used in security, privacy and resilience experiments will be limited to the validation of threats and mitigations, and the collection of evidence data on them.

6 Overall 6G E2E system design

This chapter first presents an analysis of the preliminary overall integration of Hexa-X-II enablers in 6G E2E system (see section 6.1). The criteria under consideration for the integration of the various enablers that are under development in the Hexa-X-II project are defined. Some criteria are elaborated for indicating the impact of each enabler on the E2E system. Other criteria are established to evaluate the conformance to the design principles and the migration path to 6G as defined in [HEX223-D21]. As stated in [HEX223-D21], ten system design principles defined in Hexa-X-II are listed as follows: 1) Support and exposure for 6G services and capabilities; 2) Full automation and optimization; 3) Flexibility to different network scenarios; 4) Network scalability; 5) Resilience and availability; 6) Persistent security and privacy; 7) Internal interfaces are cloud optimized; 8) Separation of concerns of network functions; 9) Network simplification in comparison to previous generations; 10) Minimizing environmental footprint and enabling sustainable networks. Criteria about the dependence with other enablers, as well as the expected contribution on targeted key performance and key values are then completing the list.

Using the defined criteria, a preliminary analysis is conducted for a first set of Hexa-X-II enablers that relates to E2E management and orchestration (M&O). The objective is to provide in section 6.1.36.1.3, some recommendations for integration in the E2E system and for validation in the first system PoC (system PoC A)which is further detailed in chapter 7), by integrating some of them to demonstrate concrete use cases. Then, in section 6.2 6.2, the E2E system blueprint defined in [HEX223-D21] is updated to consider the analysis of the selected enablers, refining the functionalities in the different layers as well as for the pervasive functionalities. Besides, one focuses more specifically onto the M&O block in the blueprint as most of the current enablers analyzed so far are part of this pervasive functionality, capturing the dependence with the different layers (resource, function and application) of the blueprint and with the other pervasive functionalities. As M&O of 6G services requires the involvement of multiple stakeholders (different owners of the set of capabilities involved in 6G services) to deliver and assure the 6G services, an E2E intent-based service management architecture framework is introduced to detail the aspects of the system architecture in the multi-stakeholder scope. Being of special importance the description presented in section 6.2.2.2 6.2.2.2 regarding the need to evolve from well-known roles such as Network Operator and CSP towards a new set of roles more adapted to the proposed E2E intent-based service management architecture framework, with special emphasis on the Digital Service Provider (DSP) and Capabilities Operator (CO) that were also referenced by some enablers in section 4.2.

6.1 Enabler integration in 6G E2E system design

6.1.1 Key criteria for enabler integration in 6G E2E system

The iterative system design process elaborated in [HEX223-D21] includes a bottom-up approach in which the enablers (and their components) designed in an independent manner are analyzed in order to achieve the 6G E2E system architecture objectives. The analysis results will serve to continually update the 6G system blueprint as well as the component design, as enablers and components become mature within the project. As part of the process, the enabler analysis considers pros and cons of each potential enabler and component developed or considered for achieving the 6G E2E system requirements. For this purpose, some criteria are proposed of what needs to be considered in technical components/enablers design to align with the E2E performance and operation targets. Such criteria can be used as a checklist for on-going development of enablers. The following definition of criteria for the integration of the enablers in the E2E system is established in Table 6-1.

Table 6-1: Criteria for enabler integration in 6G E2E system

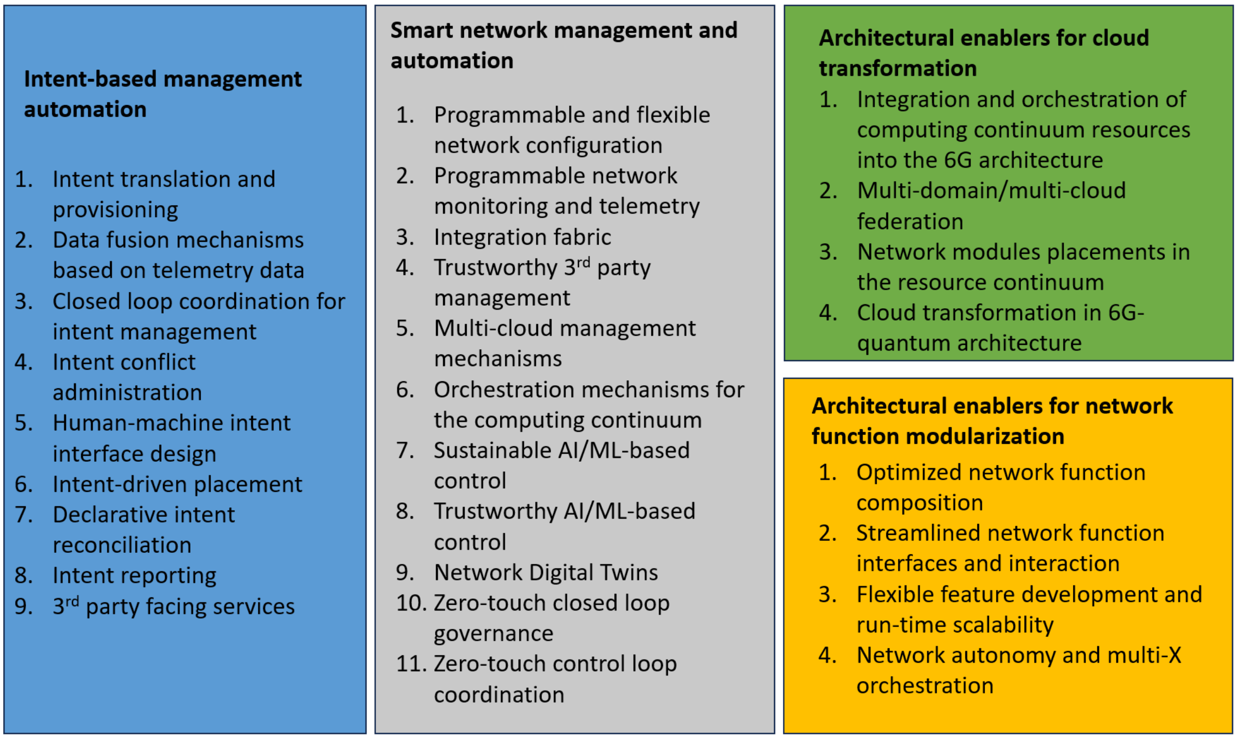
Criterion	Description
Relevance and significance of enabler towards E2E system design	It indicates the perimeter of the enabler (e.g., RAN domain only) and the importance of the enabler in efficient and effective in the control, coordination, and optimization of the 6G system.
Impact of the enabler on the E2E system design	It indicates which functionality is provided and if it has other requirements toward the rest of the system.
How the enabler fits with the system design principles	identifies to which system design the enabler contribute and how it fulfils the design principles.
Feasibility (estimation) of enabler vs migration options	It specifies whether the enabler is aligned with the recommendation from 5G to 6G migration options.
Dependency with other enablers	It indicates the relationship with other enablers.
Any proposed updates to E2E system design and architecture design principles	It identifies some update to the blueprint in terms of new functionalities, interfaces, etc. It also indicates requirements for setting a new design principle.
Network performance, security/privacy, flexibility, resilience/robustness, and sustainability/energy efficiency	Those criteria report the expectation from the enabler in terms of network performance, security/privacy, flexibility, resilience/robustness, and sustainability/energy efficiency.

These criteria provide a framework used to determine the worth of the Hexa-X-II enablers in the E2E system. The important insights are reported in the next section for a first set of enablers and are used to provide recommendations to the tasks working on the further iteration on the enabler design.

6.1.2 First iteration of enabler analysis

The analysis is carried out on a preliminary set of Hexa-X-II enablers that have been identified as important technology innovations for the use case of cobot cooperating in the context of an industrial environment that is under study in the first system PoC of Hexa-X-II and described in [HEX223-D21].

Three categories of enablers have been identified (see Figure 6-1) that are directly derived from the use case requirements as described in chapter 2: i.e., ***1) Enablers for intent-based service management automation*** (section 4.2); ***2) Enablers for smart network management*** [HEX223-D62]; ***3) Enablers for the virtualization and cloud continuum transformation***. Indirectly connected to the previous ones, a fourth category of enablers has also been identified as important to consider in this first iteration of the system design. It covers the fourth category ***4) Enablers for the modularization of the network function***. Network modules will be new managed objects that have to be manipulated by the M&O of the E2E 6G system. As such, the enablers for the network function modularization will facilitate a flexible and customized composition and placement of modules to fulfill the service intent expectation. All the considered four types of enablers are listed in Figure 6-1 and described in the following subsections. Moreover, summary of considered system design principles in these enablers are presented in Annex 9.



Intent-based management automation

1. Intent translation and provisioning
2. Data fusion mechanisms based on telemetry data
3. Closed loop coordination for intent management
4. Intent conflict administration
5. Human-machine intent interface design
6. Intent-driven placement
7. Declarative intent reconciliation
8. Intent reporting
9. 3rd party facing services

Smart network management and automation

1. Programmable and flexible network configuration
2. Programmable network monitoring and telemetry
3. Integration fabric
4. Trustworthy 3rd party management
5. Multi-cloud management mechanisms
6. Orchestration mechanisms for the computing continuum
7. Sustainable AI/ML-based control
8. Trustworthy AI/ML-based control
9. Network Digital Twins
10. Zero-touch closed loop governance
11. Zero-touch control loop coordination

Architectural enablers for cloud transformation

1. Integration and orchestration of computing continuum resources into the 6G architecture
2. Multi-domain/multi-cloud federation
3. Network modules placements in the resource continuum
4. Cloud transformation in 6G-quantum architecture

Architectural enablers for network function modularization

1. Optimized network function composition
2. Streamlined network function interfaces and interaction
3. Flexible feature development and run-time scalability
4. Network autonomy and multi-X orchestration

Figure 6-1: List of enablers considered for the first iteration enabler analysis.

6.1.2.1 *Enablers for intent-based management automation*

The key elements from the analysis criteria for the integration in the E2E system are discussed for the following enablers that are detailed in chapter 4 of this deliverable related to intent-based management automation.

6.1.2.1.1 Intent translation and provisioning

This enabler (detailed in section 4.2.1) has the main function to manage the process to reach a service and requirements agreements with the user and translate that (intent-based) agreement into the right number of system-internal requests to manage network services. In essence, this enabler should be a possible access point for the user to interact with the system itself. The impact of this enabler on the E2E system design is on the fact that this enabler should allow the interaction between a user (i.e., intent owner) and the system (i.e., intent handler), so the user may define what it needs in an easy procedure and avoiding the need for the user to have a very specific and technical knowledge to request a service.

Regarding the system design principles, this enabler suits to three of them: a) “support and exposure of 6G services and capabilities” (principle 1) due to the implementation of capability called “intent handling capability exposure” that is focused on presenting to the user the intent-based services available, b) “full automation and optimization” (principle 2) due to the fact of using intent and removing technical responsibilities from the final users and placing them on the intent-based system itself and finally, c) “flexibility to different network scenarios” (principle 3) due to the objective of making this enabler capable to manage E2E intent-based services across different and multiple resource (i.e., networks and/or computing) domains. This enabler does not bring significant constraints towards feasibility considerations. To achieve a complete management and control of the intents, Figure 6-2 shows the primarily identified interaction with other enablers.

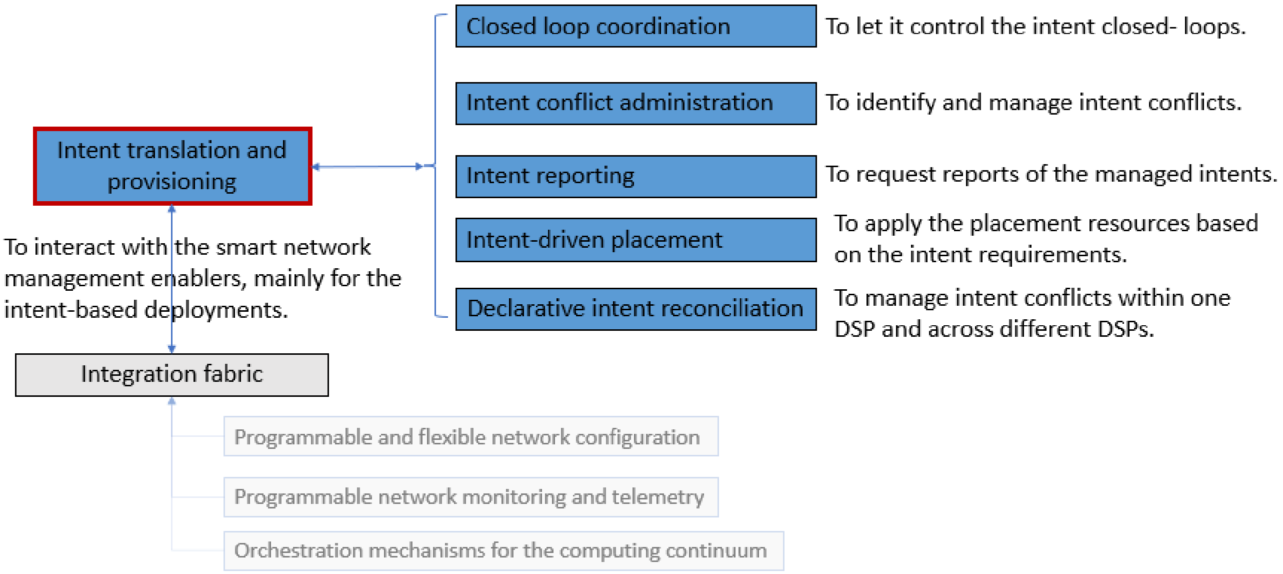

Closed loop coordination
To let it control the intent closed- loops.
Intent conflict administration
To identify and manage intent conflicts.
Intent translation and provisioning
Intent reporting
To request reports of the managed intents.
To interact with the smart network management enablers, mainly for the intent-based deployments.
Intent-driven placement
To apply the placement resources based on the intent requirements.
Declarative intent reconciliation
To manage intent conflicts within one DSP and across different DSPs.
Integration fabric
Programmable and flexible network configuration
Programmable network monitoring and telemetry
Orchestration mechanisms for the computing continuum

Figure 6-2 - Correlation between “the enabler "intent translation and provisioning" and other enablers.

6.1.2.1.2 Data fusion mechanisms based on telemetry data

This enabler (detailed in section 4.2.2) focuses on the collection and fusion of information coming from various signals (metrics, traces, logs). It can be used by network management platforms as well as orchestration platforms for edge/cloud computing applications with strict QoS requirements. The enabler can assist the development of intelligent orchestration mechanisms within an E2E 6G system by providing access to advanced insights regarding failures, abnormalities, or misperformances that exist in the operation of the infrastructure, as well as a repository of homogenized time series data that can be used for the development of analysis processes. An API is envisaged to be made available to enable usage and integration of the enabler outcomes by other components/platforms of the 6G architecture. The enabler is mainly related to the design principle 1 (support and exposure of 6G services and capabilities) based on the provision of interfaces for access to the fused data, and the design principle 5 (Resilience and availability) since it can enable the monitoring and assurance of conformance to SLAs.

The enabler will provide an interface to all services that require real-time telemetry data for their decision-making mechanisms, such as *Orchestration mechanisms for the computing continuum* and *Intent-driven placement*. Network telemetry data collection from the *Programmable network monitoring and telemetry* enabler will contribute to the integration of QoS metrics to the data fusion schema. Finally, the enabler will utilize the *Integration Fabric* to broadcast real-time data to other services that need to access the interfaces through the corresponding data bus. The correlation between this enabler with other enablers is primarily identified as in Figure 6-3.

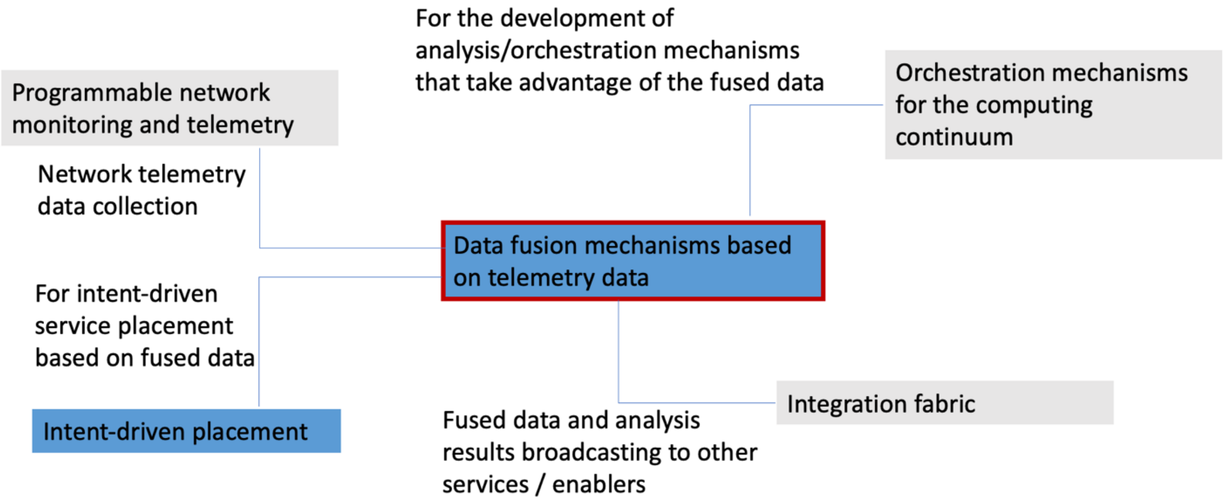

For the development of analysis/orchestration mechanisms that take advantage of the fused data
Programmable network monitoring and telemetry
Orchestration mechanisms for the computing continuum
Network telemetry data collection
Data fusion mechanisms based on telemetry data
For intent-driven service placement based on fused data
Integration fabric
Intent-driven placement
Fused data and analysis results broadcasting to other services / enablers

Figure 6-3: Correlation between the enabler "data fusion mechanisms based on telemetry data" with other enablers.

6.1.2.1.3 Closed loop coordination for intent management

The closed loop coordination (CLC) for intent management is an enabler (introduced in section 4.2.3) that is under widely investigation in the context of smart network management in section 6.1.2.2.11 and [HEX223-D62]. Here it expresses the same concepts and fits into the same architectural principles of its counterpart zero-touch multiple closed loop coordination enabler, although applied at the level of intent. In this regard, a comprehensive discussion related to the relevance and impact of this enabler to the E2E system architecture, its feasibility, and dependencies, is provided in section 6.1.2.2.11.

6.1.2.1.4 Intent conflict administration

The Intent conflict administration enabler (introduced in section 4.2.4) can be related to some enablers in the Hexa-X-II for smart network management. For example, the enablers “programmable flexible network configuration” and “programmable network monitoring and telemetry” are important to the enabler “intent conflict management” as they deal with the network programmability and monitoring respectively. To be more precise, the first one deals with solutions to enable a network to be flexible and to be configured by the human or by the network itself. After using the solutions of the second enabler to get information about the status of the network, the intent fulfilment evaluation is performed to check if the intent is being fulfilled). With this information about the network, the *intent manager* can check if the intent is being fulfilled. If this is not the case, the *intent manager* can take actions to improve the intent expectations, and this can be done by using the solutions of the enabler “programmable flexible network configuration” to reconfigure the network. However, this reconfiguration can lead to intent conflict. For example, consider two intents A and B. The intent A has the target of throughput > 50 Mbps and the intent B has the target of low resource consumption. Using the network monitoring tools, they can be checked that the intent A is not being fulfilled. One possible solution is to increase the throughput by modifying some aspects of the network using solutions of the enabler “programmable flexible network configuration”. However, this is consuming more network resources which can result in a conflict with the intent B and some solution needs to be developed to solve this intent conflict which is a task of the “intent conflict administration” enabler. It is important to highlight that, after the deployment of one Intent in the network, a closed-loop is created and instantiated to execute and monitor the Intent expectations. Therefore, for intent conflict, it is important to develop solutions for the management of the closed-loops and the management of simultaneous closed-loop. Specially the simultaneous closed-loop management is related to the “intent conflict administration” as in this scenario multiple closed loops will be instantiated. Therefore, the Intent conflict administration enabler are also related to the “Zero-touch closed loop coordination” enabler as to solve an intent conflict it is necessary to coordinate multiple closed loops.

Also, the “Intent reporting” enabler can provide useful information to detect and handle intent conflicts. Figure 6-4 describes the correlation between the “intent conflict administration” enabler and the other enablers.

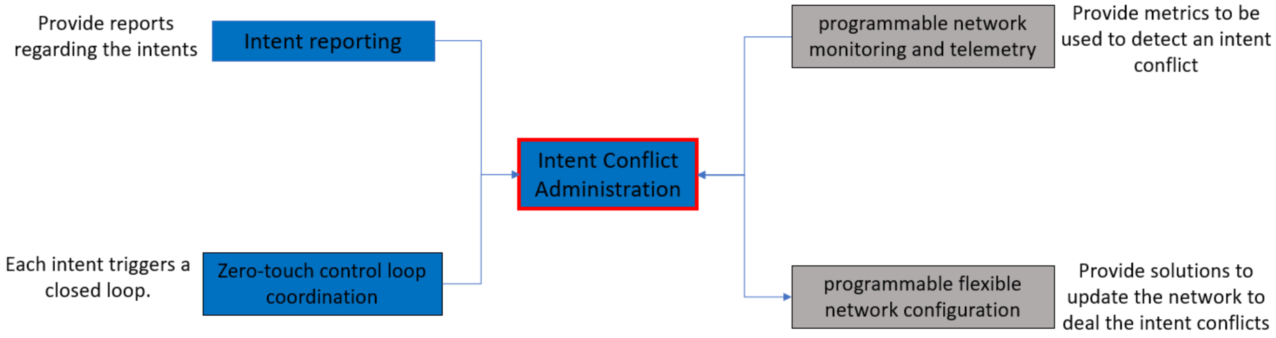

Provide reports regarding the intents
Intent reporting
Each intent triggers a closed loop.
Zero-touch control loop coordination
Intent Conflict Administration
programmable network monitoring and telemetry
Provide metrics to be used to detect an intent conflict
programmable flexible network configuration
Provide solutions to update the network to deal the intent conflicts

Figure 6-4: Correlation between the “intent conflict administration” enabler and other enablers.

6.1.2.1.5 Human-machine intent interface design

This enabler (introduced in section 4.2.5) aims for designing a human-machine intent interface that bridges the gap between humans, machines, and networks. It aims to provide a way for humans and applications to express their requirements in their respective domain languages (without in-depth telco knowledge), while maintaining a meaningful and actionable feedback from the network. Accordingly, the E2E system should be able to handle H2M/M2H interfaces as well as M2M interfaces. It should also be able to efficiently translate application intents, map it to the right network objectives, and provide actionable feedback to the applications.

As a result, this enabler fits system design principle 2 (full automation and optimization), and principle 3 flexibility to different network scenarios where the network should be fully automated without human interaction and should support increased application awareness and adaptive QoS/QoE. According to a preliminary analysis, this enabler is related to other enablers as presented in Figure 6-5.

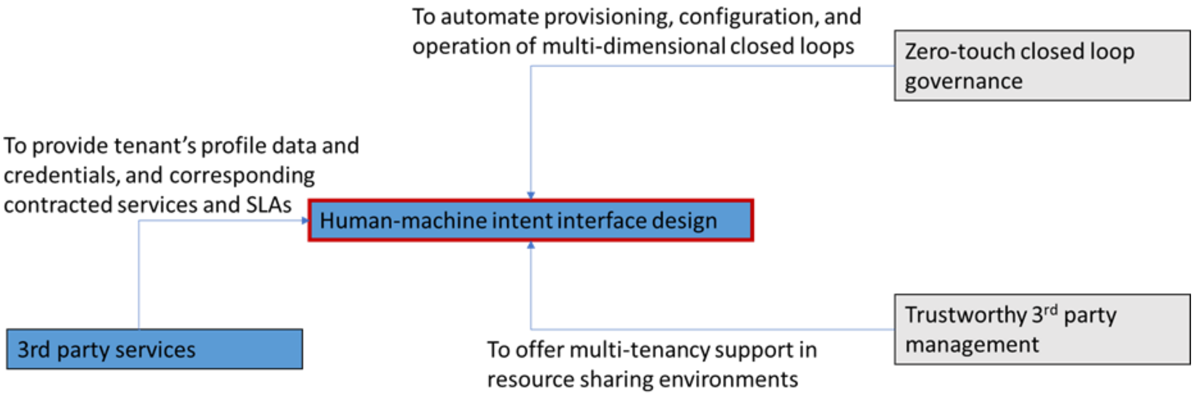

To automate provisioning, configuration, and operation of multi-dimensional closed loops
Zero-touch closed loop governance
To provide tenant’s profile data and credentials, and corresponding contracted services and SLAs
Human-machine intent interface design
3rd party services
To offer multi-tenancy support in resource sharing environments
Trustworthy 3rd party management

Figure 6-5: Correlation between the enabler “human-machine intent interface design” and other enablers.

6.1.2.1.6 Intent-driven placement

This enabler (introduced in section 4.2.6) supports coarse grained intent-driven compute placement along the compute continuum. As such it fits system design principle 1 (Support and exposure of 6G services and capabilities) and principle 2 (full automation and optimization). It depends on the enabler “intent translation and provisioning” as a basis for intent management and the enabler “data fusion mechanisms based on telemetry data” to support decision within placement-oriented control loops where the service orchestration request should be placed. Finally, it largely relies on the enabler “orchestration mechanisms for the computing continuum” which will expose APIs for access on distributed agents for managing and scaling microservices across the continuum and for resource allocation, migration and scheduling actions on heterogeneous computing nodes across the continuum Figure 6-6. represents these correlations with other enablers.

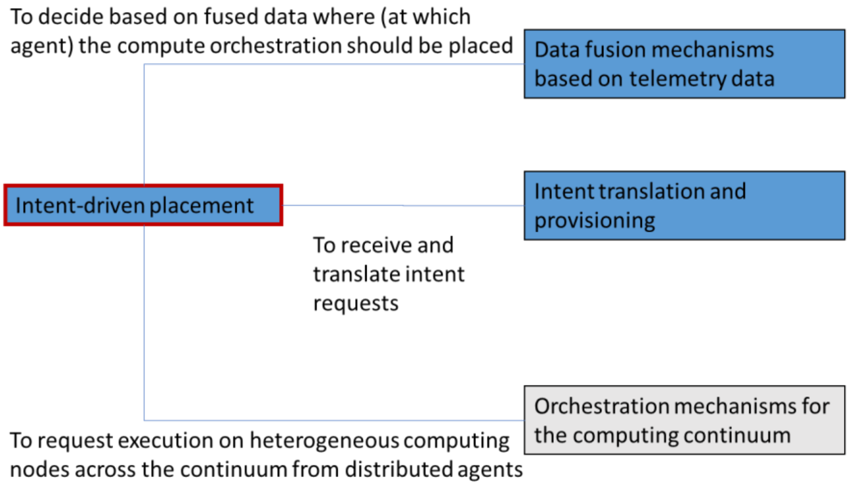

To decide based on fused data where (at which agent) the compute orchestration should be placed
Data fusion mechanisms based on telemetry data
Intent-driven placement
Intent translation and provisioning
To receive and translate intent requests
Orchestration mechanisms for the computing continuum
To request execution on heterogeneous computing nodes across the continuum from distributed agents

Figure 6-6: Correlation between the enabler “intent-driven placement” and other enablers.

6.1.2.1.7 Declarative intent reconciliation

This enabler (introduced in section 4.2.7) aims to manage intent specifications and automate its life cycle by adopting a declarative approach across management domains controlled by various stakeholders. Particularly, intent parameters and management pipelines can be described, configured and version-controlled in a single source of truth to ensure the consistency. Upon an intent update, the corresponding pipeline is executed, and its progress is orchestrated to fulfil the new requirement.

The enabler declarative intent reconciliation fits system design principle 1 (Support and exposure of 6G services and capabilities) and principle 2 (full automation and optimization). According to the preliminary analysis, this enabler is related to the other enablers as presented in Figure 6-7.

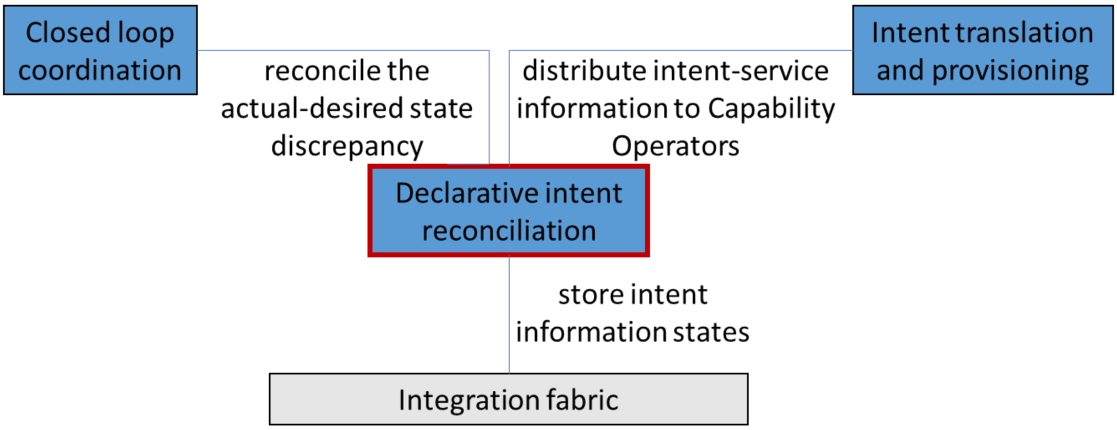

Closed loop coordination
reconcile the actual-desired state discrepancy
distribute intent-service information to Capability Operators
Intent translation and provisioning
Declarative intent reconciliation
store intent information states
Integration fabric

Figure 6-7: Correlation between the enabler “declarative intent reconciliation” and other enablers.

6.1.2.1.8 Intent reporting

This enabler (introduced in section 4.2.8) aims to specify an information model for intent report, including details on intent fulfilment (intent fulfilment status and current performance values for corresponding expectation targets), intent conflict (conflict type and possible solution recommendations) and intent feasibility. This intent report model should be designed in such a way that allows a tenant, as intent owner, to configure how it wants to be reported; in other words, the consumer shall be able to receive intent reports with different content and intervals according to its specified requirements.

The impact of this enabler on the E2E system design is notably significant, as it provides an internal yet self-contained capability within the DSP’s intent-based digital service manager. The tangible architectural outcome resulting from this capability is the definition of a new sub-system, called "intent reporting", which aligns with one of the core system design principles 2 full automation and optimization.

Finally, this enabler is closely related with other three enablers, as depicted in Figure 6-8.

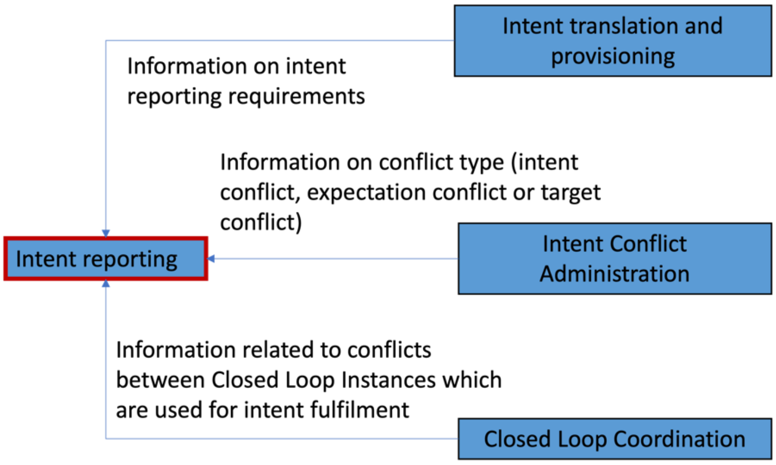

Intent translation and provisioning
Information on intent reporting requirements
Information on conflict type (intent conflict, expectation conflict or target conflict)
Intent reporting
Intent Conflict Administration
Information related to conflicts between Closed Loop Instances which are used for intent fulfilment
Closed Loop Coordination

Figure 6-8: Correlation between the enabler "intent reporting" and other enablers.

6.1.2.1.9 3rd party facing services

This enabler (introduced in section 4.2.9) aims to specify how the DSP provides a characterization of i) individual tenants accessing 6G system, capturing this information in the form of a 3rd party profile, and ii) service offerings, which will be later linked to tenants according to well-defined SLAs through smart contracts.

In terms of E2E system design, this enabler translates into two new sub-systems within the DSP's intent-based digital service manager: 3P profiling and service portfolio. Their collaborative functional scope provides the DSP with zero-touch means when provisioning customer-facing capabilities, contributing to the realization of the principles 1 (support and exposure of 6G services and capabilities) and principle 2 (full automation and optimization). According to the preliminary analysis, this enabler is related to the other enablers as presented in Figure 6-9.

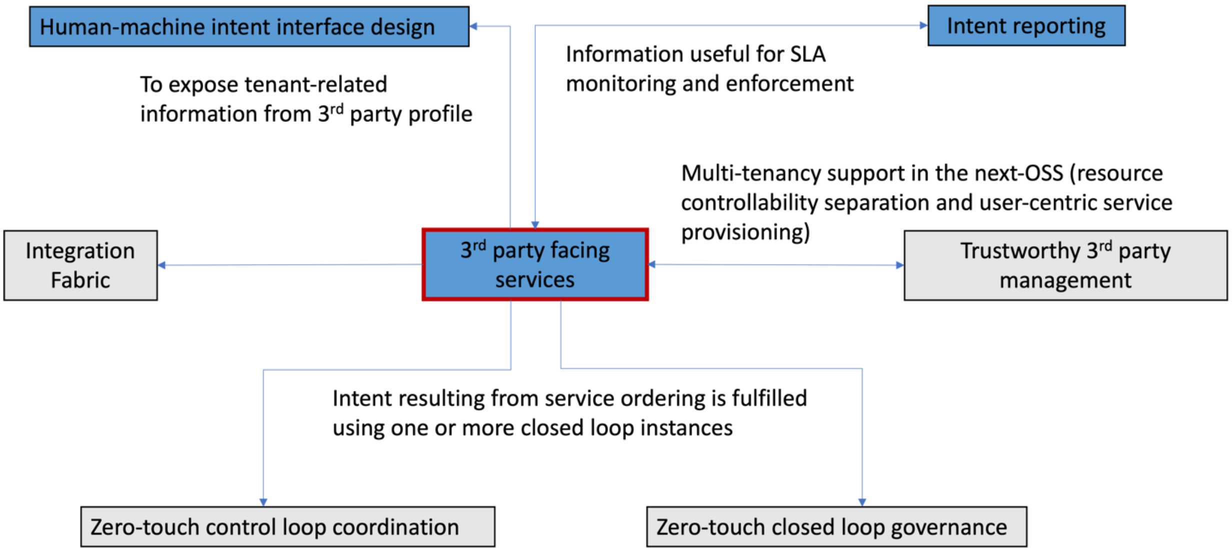

Human-machine intent interface design
Intent reporting
Information useful for SLA monitoring and enforcement
To expose tenant-related information from 3rd party profile
Multi-tenancy support in the next-OSS (resource controllability separation and user-centric service provisioning)
Integration Fabric
3rd party facing services
Trustworthy 3rd party management
Intent resulting from service ordering is fulfilled using one or more closed loop instances
Zero-touch control loop coordination
Zero-touch closed loop governance

Figure 6-9: Correlation between the enabler "3rd party facing services" and other enablers.

6.1.2.2 *Enablers for smart network management*

The key elements from the analysis criteria for the integration in the E2E system are discussed for the following enablers proposed in [HEX223-D62].

6.1.2.2.1 Programmable and flexible network configuration

The enabler addresses the need for advanced control and management capabilities in a rapidly evolving technological landscape, making it a crucial component of E2E system design. The impact of this enabler on the E2E system design is notably significant as it provides a T-NSSMF (Transport Network Slice Specific Management Function) interface. This interface serves as a bridge between the hierarchical SDN controller and the transport network, facilitating seamless communication and control. This impact ensures that the E2E system design can incorporate the necessary management functions efficiently, contributing to enhanced system performance and agility.

The integration of the Cloud-Native Hierarchical SDN Controller aligns seamlessly with the core system design principles. This enabler is inherently cloud-native, emphasizing the use of cloud-based technologies and principles in its operation. Moreover, it targets full automation and optimization, aligning with principle (2) of efficient system operation. Additionally, it plays a role in enabling sustainable networks, in accordance with principle (10), by promoting efficient resource utilization and network sustainability through advanced control mechanisms. Feasibility considerations for the adoption of this enabler indicate no significant constraints. Its integration into the E2E system design does not impose undue challenges or limitations. This enabler can be implemented and utilized effectively, making it a feasible choice in the design and deployment of E2E systems. Figure 6-10 shows the primarily identified interaction with other enablers.

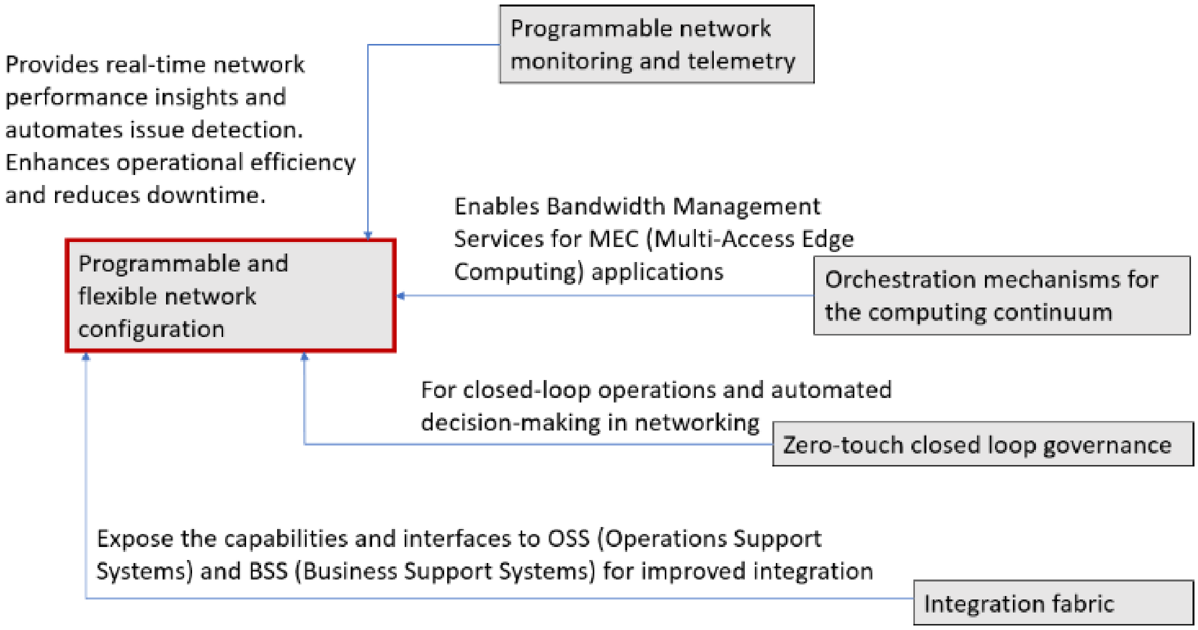

Programmable network monitoring and telemetry
Provides real-time network performance insights and automates issue detection. Enhances operational efficiency and reduces downtime.
Programmable and flexible network configuration
Enables Bandwidth Management Services for MEC (Multi-Access Edge Computing) applications
Orchestration mechanisms for the computing continuum
For closed-loop operations and automated decision-making in networking
Zero-touch closed loop governance
Expose the capabilities and interfaces to OSS (Operations Support Systems) and BSS (Business Support Systems) for improved integration
Integration fabric

Figure 6-10: Correlation between the "programmable and flexible network configuration" and other enablers.

6.1.2.2.2 Programmable network monitoring and telemetry

The enabler plays a crucial role in the automated collection of cross-domain data, including data from the extreme edge of the network. This data collection encompasses various aspects such as network performance, energy monitoring, and time-sensitive networking (TSN) monitoring, ultimately impacting the user experience.

The impact of this enabler on the E2E system design is extensive, as it influences numerous pervasive functionalities. It contributes to M&O capabilities by enabling automated data collection and fusion, which is essential for E2E service assurance, optimization, and anomaly detection. Additionally, the enabler facilitates the establishment of a robust data and AI framework of the 6G blueprint that proves invaluable for achieving these objectives. Its impact extends across various facets of the system design, enhancing its overall capabilities.

The enabler aligns seamlessly with several key system design principles. Firstly, it fits into the principle targeting full automation and optimization (principle 2), as it directly contributes to automating data collection and aids in optimizing network performance and user experience. Furthermore, it supports the principles related to pervasive security and privacy (principle 6) by ensuring that the collected data is managed securely and with privacy considerations in mind. Lastly, this enabler aligns with the principle of minimizing the environmental footprint and enabling sustainable networks (principle 10), as it assists in energy monitoring and sustainability efforts through data-driven insights.

Feasibility considerations for the adoption of this enabler indicate that there are no significant constraints that would impede its implementation. Its integration into the E2E system design can be achieved without undue complications.

In terms of dependency on other enablers, this enabler interacts with various M&O enablers that contribute to zero-touch management. Notably, it interfaces with enablers such as AI/ML-based control, computing continuum orchestration, network digital twin, and DataOps. For example, it can feed the data ingestion pipeline of the DataOps enabler, providing valuable data inputs for further processing and analysis.

Figure 6-11 shows the primarily identified interaction with other enablers.

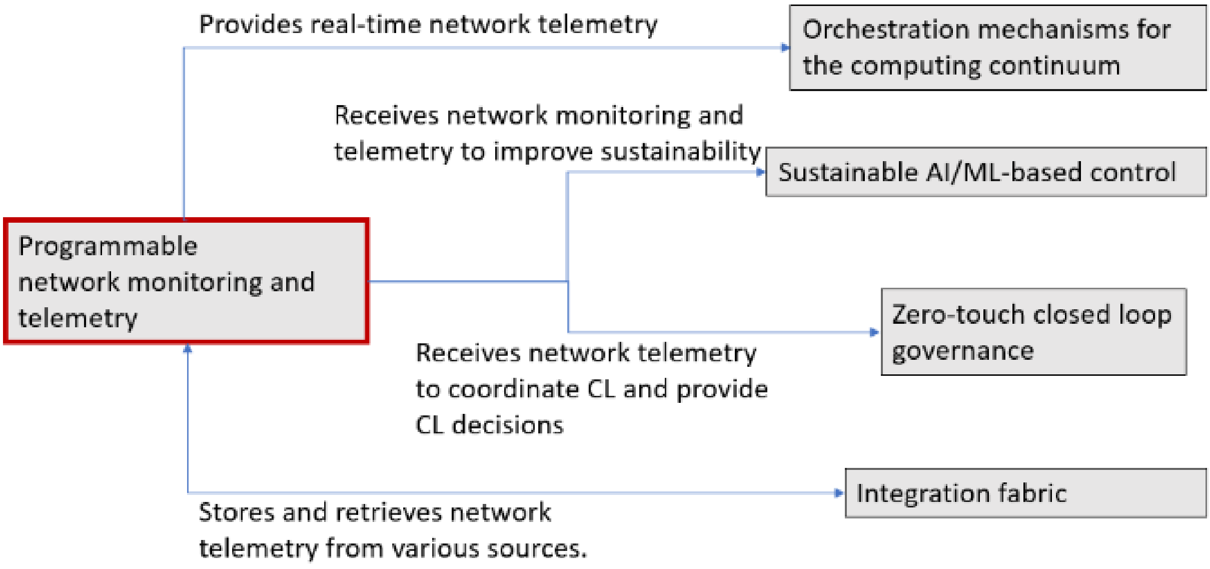

Provides real-time network telemetry
Orchestration mechanisms for the computing continuum
Receives network monitoring and telemetry to improve sustainability
Sustainable AI/ML-based control
Programmable network monitoring and telemetry
Zero-touch closed loop governance
Receives network telemetry to coordinate CL and provide CL decisions
Integration fabric
Stores and retrieves network telemetry from various sources.

Figure 6-11: Correlation between the "programmable network monitoring and telemetry" and other enablers.

6.1.2.2.3 Integration fabric

This enabler empowers the definition of a SBMA (Service Based Management Architecture) for Hexa-X-II management layer, by offering a liquid interoperation medium between API producers and consumers, regardless their administrative domains nor the actual capability in scope. Integration fabric essentially is Hexa-X-II capability connector, because it provides a service bus to connect microservices, which can be scaled and containerized independently, building up Hexa-X-II resources/functions. It implements connectivity by offering a set of communication channel, service discovery and registration. It allows services access control and observability making it a single point for intent-based digital service manager to access system internal capabilities (management, network, cloud, and AI) (as introduced in chapter 4).

The enabler fit with two of the system design principles because it unlocks the support and the exposure of 6G capabilities (principle 1), and it is completely based on cloud-native solutions, so all internal interfaces are cloud optimized (principle 7). Feasibility considerations for the adoption of this enabler indicate no significant constraints.

Given the central position in the general system architecture of this enabler, it is related to all enablers for intent-based management automation, as the integration fabric acts as a single point of contact with intent-based digital service manager. Instead regarding smart network management, it has point of contact with multiple enablers such as programmable flexible network configuration, programmable network monitoring and teleme^{tr}y, trustworthy 3rd party management, multi-cloud management mechanisms, zero-touch closed loop governance and zero-touch multiple closed loop coordination.

6.1.2.2.4 Trustworthy 3rd party management

This enabler allows offering multi-tenancy support in resource sharing environments. From the standpoint of Hexa-X-II resource providers, which are in charge of provisioning infrastructure solutions, this requires applying the enablers for confidential network computing and trust infrastructures reported in sections 5.1.3 and 5.3, respectively. From the standpoint of a Hexa-X-II Capability Operator (COP) (defined in section 6.2.2.2), this requires a controllable and auditable exposure of capabilities to individual tenants (3rd parties), according to their profiles. Controllable exposure means that the Hexa-X-II system can regulate the specific set of resources each tenant is allowed to access and under which conditions. Auditable exposure means that every interaction between Hexa-X-II system and tenant systems needs to be logged with accurate timestamps (for traceability) and support non-repudiation (for SLA verification).

The impact of this enabler on the E2E system design is notably significant, as it allows defining a sub-system called "3rd party provisioning", which serves as a bridge between the DSP's intent-based digital service manager and the rest of architecture components under the realm of COP, managing the workflows across them at provisioning time. This impact ensures that the E2E system design have means to capture multi-tenancy requirements coming from DSP, and enforce them timely in the following terms: i) resource controllability separation, providing tenants with segregated yet customized management spaces; ii) user-centric network management, defining policies for tenant subscribers in relation to service delivery and consumption, touching on user experience, privacy and regulation aspects; and iii) SLA enforcement, including KPI/KVI and TLA translation, assurance and verifiability.

The capabilities offered by the "3rd party provisioning" sub-system aligns with two of the core system design principles: (principle 1) support and exposure of 6G services and capabilities, and (principle 6) persistent security and privacy. Feasibility considerations for the adoption of this enabler indicate no significant constraints; actually, it can be easily registered and plugged into the integration fabric.

In terms of dependency with other enablers, this enabler has relationship with the other enablers such as 3rd party services, zero-touch closed loop governance and zero-touch multiple closed loop coordination. 3rd party services enabler allows the DSP to provide a full characterization of Hexa-X-II service offerings, and of the individual tenants accessing to them; this information is issued in the form of intent to the "3rd party provisioning" sub-system, which makes necessary actions to get this intent fulfilled. Apart from policy definition with regards to resource controllability separation and user-centric network management, the intent fulfilment will require the instantiation/configuration of one or more closed loop instances. For it to happen with automation means, this enabler leverages on enablers on zero-touch closed loop governance and coordination.

Finally, as to the potential extensions of E2E system design, further investigation on interfaces between the "3rd party provisioning" sub-system and the security artifacts captured in sections 5.1.3 and 5.3 is required.

6.1.2.2.5 Multi-cloud management mechanisms

This enabler addresses challenges related to the management of compute and network resources over of network services and applications over multi-cloud infrastructure. The enabler is highly relevant with the design of the E2E system, since in a 6G ecosystem there is evident the need to manage the deployment of software components across infrastructure that may span in various clusters across the computing continuum or belong to different providers. In the case of a single provider that gets access to multi-cluster infrastructure, the enabler provides solutions for multi-cluster management of both compute and network resources. Open-source multi-cluster management tools for compute resources are going to be adopted and extended accordingly to be interoperable with the resources management mechanisms that are going to be developed (e.g., for supporting autoscaling, compute offloading functionalities). In the case of multiple providers, federated management schemes are going to be considered enabling the interaction among them. In parallel, exploitation of APIs – especially northbound APIs provided by network management platforms – is going to take place to develop solutions that can tackle the end-to-end provision and orchestration of network services and applications over heterogeneous infrastructure. The enabler is strongly related with the next enabler (i.e., orchestration mechanisms for the computing continuum) and the enablers related with the management of multi-cloud infrastructure.

6.1.2.2.6 Orchestration mechanisms for the computing continuum

This enabler provides orchestration of network services and applications over resources that span across the computing continuum from the extreme edge to the edge to the central cloud part of it. Such challenges emerge with the increasing adoption of IoT based technologies and the need to provide services and applications with very strict requirements in terms of latency, especially in the extreme edge part of the infrastructure. The work in the enabler includes the adoption of intent-driven orchestration approaches, the development of synergetic orchestration mechanisms taking advantage of multi-agent systems, and the adoption – where applicable – of ML techniques to support automation in the various orchestration actions. All the aforementioned mechanisms should be considered in the design of a 6G E2E system to enable distributed intelligence and automation in the management of services across the computing continuum. The primarily identified correlations between this enabler and the other enablers are presented in Figure 6-12. With regards to the system design principles, there is high relevance with the principle for the "support and exposure of 6G services and capabilities" in terms of orchestration functions, the principle of "full automation and optimization" in terms of automated orchestration mechanisms (e.g., autoscaling mechanisms), and the principle for the development of "internal interfaces that are cloud optimized" given the adoption of a microservices-based development approach.

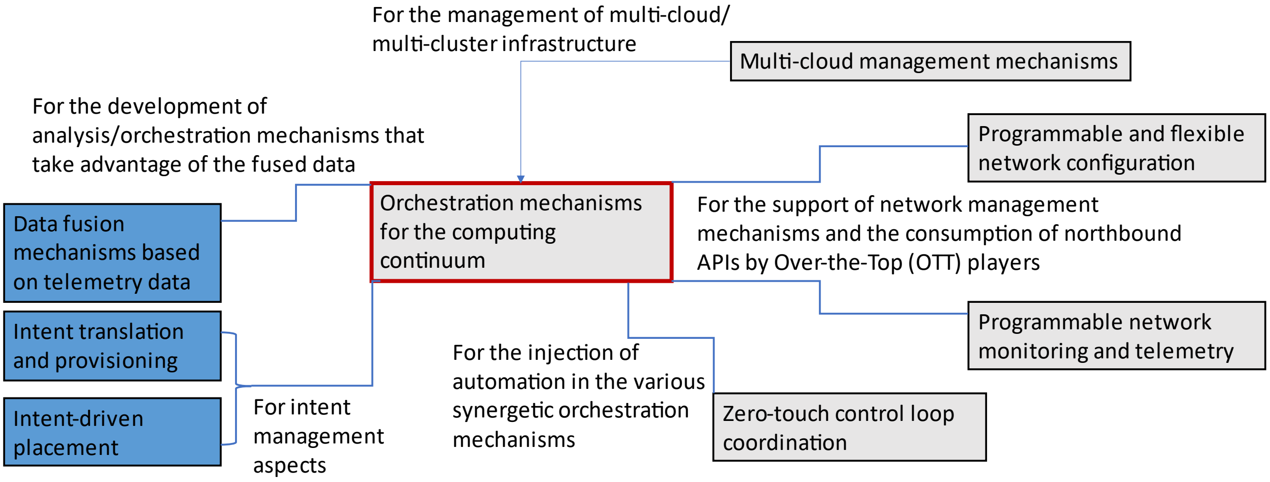

For the management of multi-cloud/ multi-cluster infrastructure
Multi-cloud management mechanisms
For the development of analysis/orchestration mechanisms that take advantage of the fused data
Programmable and flexible network configuration
Orchestration mechanisms for the computing continuum
For the support of network management mechanisms and the consumption of northbound APIs by Over-the-Top (OTT) players
Data fusion mechanisms based on telemetry data
Intent translation and provisioning
Programmable network monitoring and telemetry
For the injection of automation in the various synergetic orchestration mechanisms
Intent-driven placement
For intent management aspects
Zero-touch control loop coordination

Figure 6-12: Correlation between the enabler "orchestration mechanisms for the computing continuum" and the other enablers.

6.1.2.2.7 Sustainable AI/ML based control

This enabler's focus is the development of AI/ML algorithms for network management with the objective of achieving energy sustainability and reducing carbon footprint while satisfying the network and service performance objectives. Energy efficiency could then be achieved by performing optimal resource allocation and usage, by providing only the necessary resources to reach the KPI values that satisfy the user's SLAs. On the other hand, using AI/ML mechanisms can result in increased energy consumption. Thus, the enabler will also work on designing and implementing mechanisms to keep energy consumption to a minimum, moreover reducing CO2 emission pursuing environmental sustainability target.

This enabler is relevant for the system's autonomous adaptivity and overall increase of efficiency as well as decrease of carbon footprint by providing tools to enable fully autonomous and sustainable network systems. Indeed, its output will increase network's performance in terms of optimal resource and energy use. As such, this enabler aligns with key system design principles 2 (full automation and optimization) and 10 (minimizing environmental footprint and enabling sustainable networks).

In terms of dependencies, Figure 6-13 shows the correlation between this enabler and other enablers.

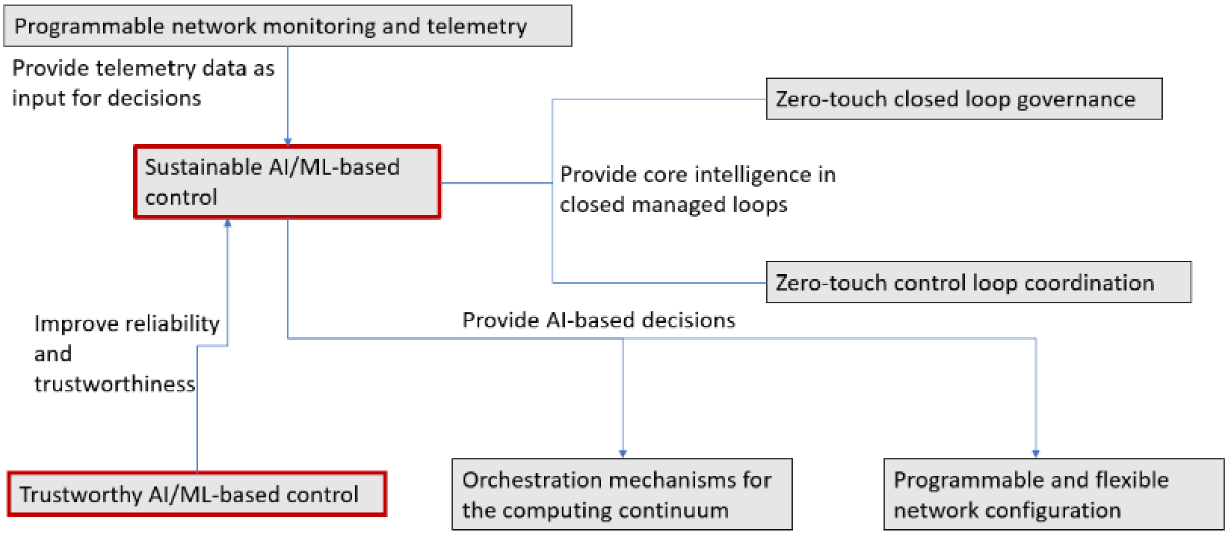

Programmable network monitoring and telemetry
Provide telemetry data as input for decisions
Zero-touch closed loop governance
Sustainable AI/ML-based control
Provide core intelligence in closed managed loops
Zero-touch control loop coordination
Improve reliability and trustworthiness
Provide AI-based decisions
Trustworthy AI/ML-based control
Orchestration mechanisms for the computing continuum
Programmable and flexible network configuration

Figure 6-13: Correlation between two enablers "sustainable AI/ML-based control" and "trustworthy AI/ML-based control" with the other enablers.

6.1.2.2.8 Trustworthy AI/ML-based control

While AI/ML methods are expected to improve system efficiency and performance, they are also vulnerable to privacy attacks that might compromise sensitive data, as well as adversarial attacks where the decisions of models are influenced by introducing deceitful input data. Lastly, AI/ML models are seen as complex "black boxes" for which the logic behind the provided decisions cannot be explained, which raises concerns in terms of accountability and hinders the adoption of AI/ML in complex systems.

This enabler is relevant to the end-to-end system design by improving security and privacy aspects of ML elements in the system architecture and management, and thus fits with the resilience and availability system design principle. The mechanisms developed within this enabler will result in increased security and privacy for AI/ML based procedures and an improved network resilience to privacy and adversarial attacks. Additionally, the enabler will develop explainability mechanisms to provide human-readable explanations to the decisions that are provided by the AI/ML algorithms, which allows to understand, interpret and trust those decisions. However, to ensure trustworthiness and privacy, this enabler may impose limits on information exposure within the system.

This enabler will focus on improving the reliability and trustworthiness of the sustainable AI/ML based solutions developed in the enabler for sustainable AI/ML-based control (see Figure 6-13).

6.1.2.2.9 Network Digital Twins

Efficient training of network management models is challenging due to the user data privacy concerns, as well as the need to apply and evaluate different actions in order to obtain an optimal policy. Indeed, applying sub-optimal decisions in a network system would impact system performance. In this context, the Network Digital Twin developed within this enabler will provide a near real-time representation of the network, which will allow a safe and efficient training of AI/ML models. Indeed, the Network Digital Twin has similar characteristics and behavioral patterns as the real network infrastructure and provides the feedback necessary for model training without carrying out the actions on the real network.

This enabler is relevant for closed loop control and management procedures, as it can be integrated into ML model control loops for a fully automated and safe model training and update. Thus, this enabler fits within the design principle 2 for the full automation and optimization, and principle 1 for support and exposure of 6G services and capabilities. However, it may increase the requirements for system monitoring elements in case of continuous model update loops. The correlation between the given enabler with the other enablers are presented in Figure 6-14.

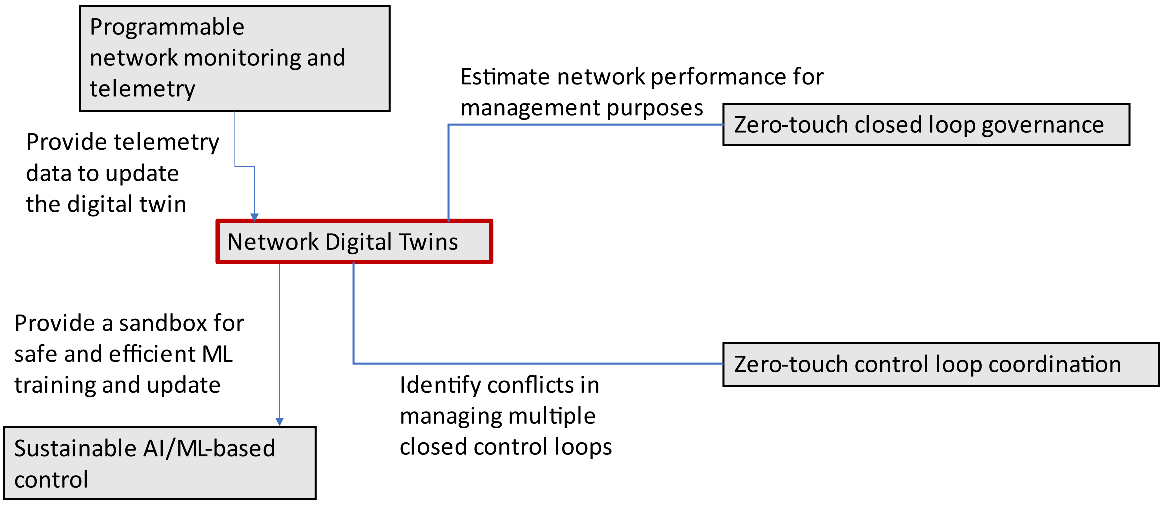

Programmable network monitoring and telemetry
Provide telemetry data to update the digital twin
Estimate network performance for management purposes
Zero-touch closed loop governance
Network Digital Twins
Provide a sandbox for safe and efficient ML training and update
Zero-touch control loop coordination
Identify conflicts in managing multiple closed control loops
Sustainable AI/ML-based control

Figure 6-14: Correlation between the enabler "Network Digital Twins" and the other enablers.

6.1.2.2.10 Zero-touch closed loop governance

The enabler allows to instantiate and manage the lifecycle and the execution of closed loops' functions (monitoring, analysis, decision, actuation), exposing interfaces to interact with the CL functions and interacting with a CL coordination when multiple closed loops need to be jointly coordinated. At the E2E system level, suitable interfaces must allow to collect real-time monitoring data related to different layer, domains, technologies, in order to feed the analysis stage of the CL. The re-configuration actions defined at the CL decision stage need to be actuated through other interfaces enabling the dynamic programmability of the network. These interactions should be mediated through a common integration fabric.

In terms of architecture design principles, as defined in [HEX223-D21], this enabler mainly contributes to principle 2 (Full automation and optimization). Since the governance can be applied to various types of closed loops specialized for particular scenarios, it is also aligned with principle 3 (Flexibility to different network scenarios). When the governance is applied to a closed loop related to service recovery and restoration, it can be considered associated to principle 5 on resiliency and availability. No additional design principle or E2E system design updates are required to support this enabler.

The control loop governance enabler can be implemented through enhancements to the current management systems, mostly relying on virtualization and data analytics functionalities, without any additional constraints. Figure 6-15 presents the primarily identifies correlations between the given enabler and the other enablers.

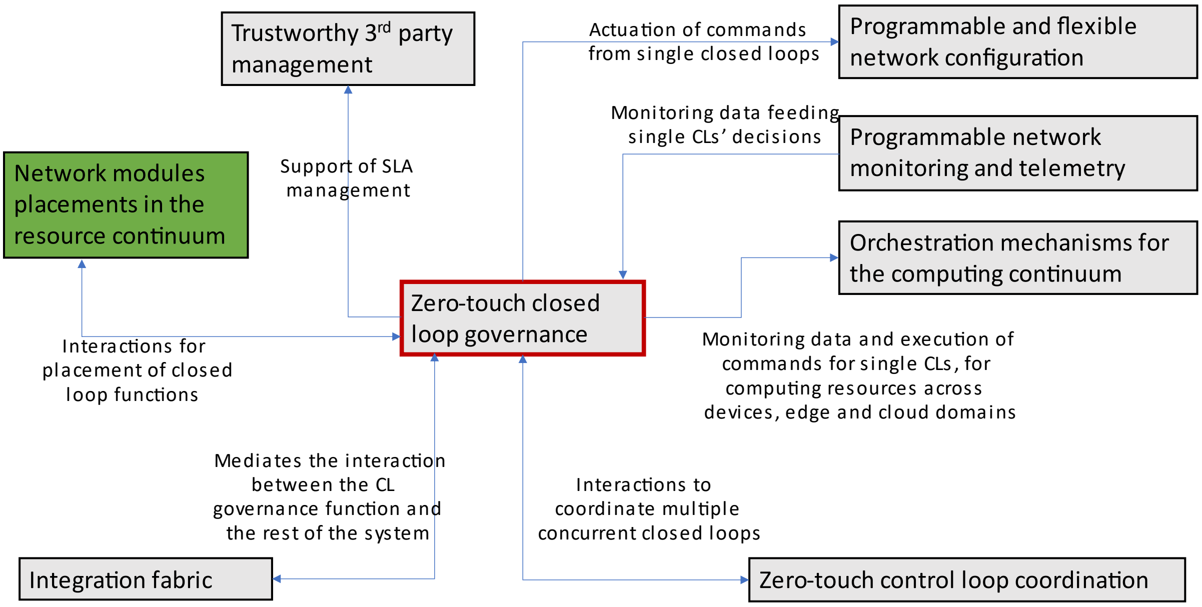

Trustworthy 3rd party management
Actuation of commands from single closed loops
Programmable and flexible network configuration
Monitoring data feeding single CLs' decisions
Programmable network monitoring and telemetry
Network modules placements in the resource continuum
Support of SLA management
Orchestration mechanisms for the computing continuum
Zero-touch closed loop governance
Interactions for placement of closed loop functions
Monitoring data and execution of commands for single CLs, for computing resources across devices, edge and cloud domains
Mediates the interaction between the CL governance function and the rest of the system
Interactions to coordinate multiple concurrent closed loops
Integration fabric
Zero-touch control loop coordination

Figure 6-15: Correlation between the enabler "zero-touch closed loop governance" and the other enablers.

6.1.2.2.11 Zero-touch multiple closed loop coordination

The enabler It allows to coordinate various instances of closed loops that can operate at different layers of the E2E system (e.g., at resources layer, networks functions layer, and application layer) or, within a given layer, in different technological domains (e.g., operating over access resources, over compute resources, over transport NFs, etc.), in different segments (e.g., at edge or cloud domains) or within the scope of specific applications. Closed loops can refer to different categories of objectives, for example they can be used for SLA management, for intent management, etc., and the coordination mechanisms should be able to identify potential conflicts and enforce actions for conflict mitigation and resolution. This impacts the E2E system in terms of interfaces and data/capabilities exposures from the other elements of the E2E system, mostly for monitoring and actuation of re-configuration decisions. The adoption of a common integration fabric is foreseen to mediate the various interactions with the rest of the E2E system.

In terms of architecture design principles, as defined in [HEX223-D21], the enabler clearly constitutes one of the core building blocks for principle 2 (Full automation and optimization). The possibility to apply the closed loop concept to various domains, layers, services and applications is aligned with principle 3 (Flexibility to different network scenarios), while the combination of multiple closed loops through coordination and collaboration mechanisms contributes to reach higher level of network scalability, matching principle 4. Finally, since closed loops can be specialized to guarantee the mobile connection and service continuity, they indirectly contribute to principle 5 on resiliency and availability. No additional design principle or E2E system design updates are required to support this enabler.

The control loop coordination enabler can be implemented through enhancements to the current management systems, without any particular constraints preventing the migration to solutions based on closed loops. The adoption of suitable coordination mechanisms will need to be put in place to reduce the possibility of conflicting decisions and automation actions to avoid any potential instability. Figure 6-16 presents the primarily identified correlations between the given enabler and the other enablers. In addition to that, there is a clear relationship with the enabler on intent conflict administration, especially when intents' fulfilment is provided by the usage of closed control loops to be coordinated together (see Section 4.2.3).

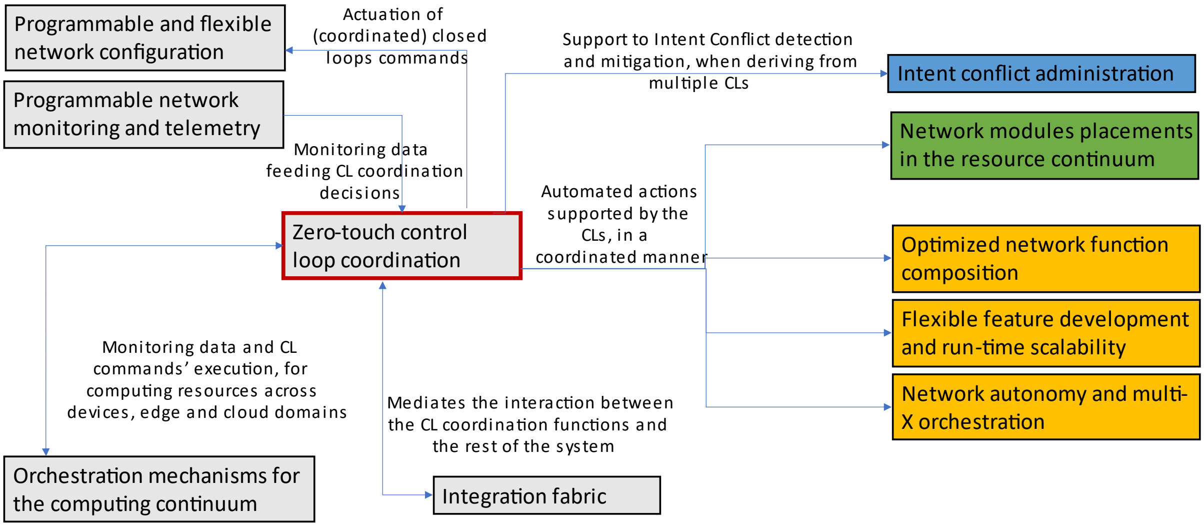

Programmable and flexible network configuration
Actuation of (coordinated) closed loops commands
Support to Intent Conflict detection and mitigation, when deriving from multiple CLs
Intent conflict administration
Programmable network monitoring and telemetry
Network modules placements in the resource continuum
Monitoring data feeding CL coordination decisions
Automated actions supported by the CLs, in a coordinated manner
Zero-touch control loop coordination
Optimized network function composition
Flexible feature development and run-time scalability
Monitoring data and CL commands' execution, for computing resources across devices, edge and cloud domains
Network autonomy and multi-X orchestration
Mediates the interaction between the CL coordination functions and the rest of the system
Orchestration mechanisms for the computing continuum
Integration fabric

Figure 6-16: Correlation between the enabler "zero-touch control loop coordination" and other enablers.

6.1.2.3 *Enablers for virtualization and cloud continuum transformation*

The key elements from the analysis criteria for the integration in the E2E system are discussed for the following enablers proposed in [HX223-D32].

6.1.2.3.1 Integration and orchestration of computing continuum resources into the 6G architecture

This enabler addresses the design principles of the resource layer in the E2E 6G system for integration and orchestration of extreme edge devices into the compute continuum. The impact of this enabler on the E2E 6G system design is significant as it provides the needed architectural modifications, interfaces, and protocols to include, orchestrate and manage extreme edge devices as part of the compute continuum. This enabler servers as an extension of exiting compute continuum beyond the radio access part of the network.

This enabler enhances the full automation and optimization (principle 2), including the end devices as part of the optimization process. Moreover, it contributes towards the flexibility to adapt to different topologies (principle 3) by enabling a more dynamic and elastic compute continuum. The design principles and interfaces taken in consideration in this enabler are cloud native (principle 7).

In terms of dependency with other enablers from the E2E 6G system, this enabler has relationships with several other enablers, see Figure 6-17. The programmable network monitoring and telemetry, orchestration mechanisms for the computing continuum and programmable and flexible network configuration enablers can be directly affected of the interfaces and design of this enabler when considering the end devices as part of the resource layer. Finally, the inclusion of this enabler as part of the E2E 6G system design aligns well with the exiting principles, no architectural updates are needed and recently has been taken in consideration by different standardization bodies (e.g., 3GPP, ETSI).

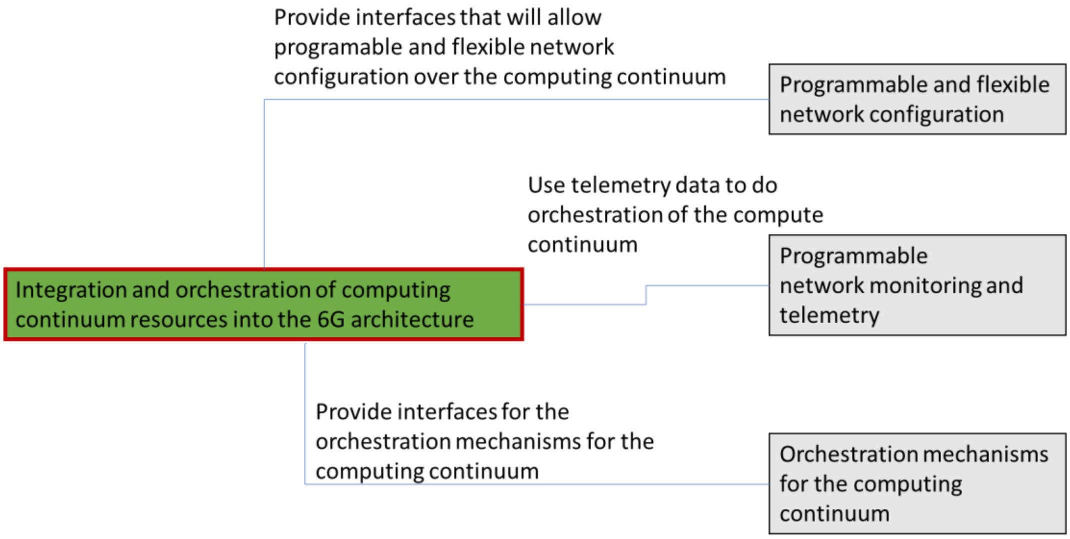

Provide interfaces that will allow programable and flexible network configuration over the computing continuum
Programmable and flexible network configuration
Use telemetry data to do orchestration of the compute continuum
Integration and orchestration of computing continuum resources into the 6G architecture
Programmable network monitoring and telemetry
Provide interfaces for the orchestration mechanisms for the computing continuum
Orchestration mechanisms for the computing continuum

Figure 6-17: Correlation between the enabler “integration and orchestration of computing continuum resources into the 6G architecture” with the other enablers.

6.1.2.3.2 Multi-domain/multi-cloud federation

This enabler plays a crucial role in enabling telco and service providers to efficiently extend the resource layer by using resources or/and services from external sources. In other words, if a telco provider has a set of available resources in the resource layer, it can lease them to an external telco provider. The impact of this enabler on the E2E 6G system design is significant as it redesigns the M&O component by proposing design solutions and secure interfaces between different M&O for multi domain federation.

This enabler aligns with several key system design principles. First, Resilience and availability (principle 5) will be improved mainly in situation when service and telco providers are facing infrastructure failure (e.g., accidents in data centers or sudden increase of number of users). The designed internal interfaces are naturally cloud optimized (principle 7) in the multi-domain/multi-cloud federation since they are using infrastructures of cloud and telco operators that are cloud-native.

In terms of dependency, the Network autonomy and multi-X (where X indicates domain, plane or stakeholder etc.) orchestration and multi-cloud management enablers can be directly affected by the design principles and interfaces of this enabler for their successful execution. In a conclusion, the inclusion of this enabler as part of the E2E 6G system design aligns well with the exiting principles and no architectural updates are needed.

6.1.2.3.3 Network modules placement in the resource continuum

This enabler addresses the need for placement of the data and control plane functionalities of the network modules in the complete compute layer in order to deliver the expected QoS to the end users. The impact of this enabler on the E2E system design is important as it is defining the APIs needed to expose the underling capabilities of the resource continuum. This covers different compute levels, ranging from central clouds to extreme edge clouds which are located near the end user. In addition, this enabler will impact the design techniques to perform the network module placement in the resource continuum environment so that it meets the module requirements.

The enabler complies with fundamental 6G design principles. Resilience and availability (principle 5) for the network modules is improved by exposing the extreme edge devices capabilities and considering them as additional host for network module placement. This enabler defines cloud optimized internal interfaces (principle 7) naturally since this trend was started with 5G where the architectural design and interfaces were cloud native.

For what considers the dependence and correlation with other enablers in the E2E 6G system, the orchestration mechanisms for the computing continuum enabler interacts with this enabler as the orchestration mechanisms will be highly depended on the interfaces and APIs that will expose the underling capabilities of the resource continuum, especially the extreme edge devices. The primarily identified correlations between this enabler and other enablers are presented in Figure 6-18.

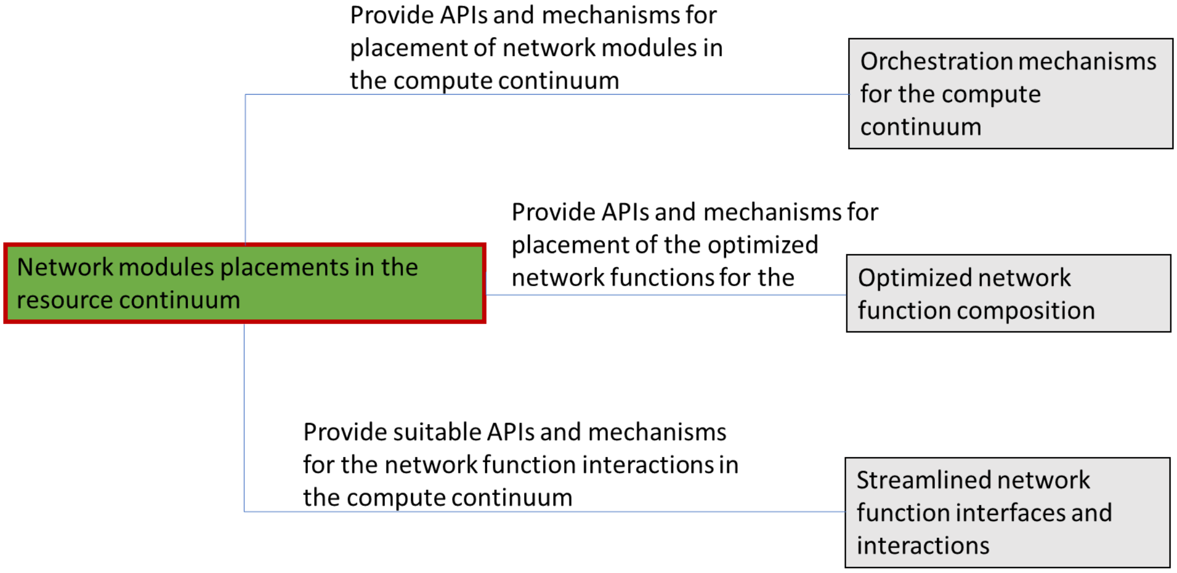

Network modules placements in the resource continuum
Provide APIs and mechanisms for placement of network modules in the compute continuum
Orchestration mechanisms for the compute continuum
Provide APIs and mechanisms for placement of the optimized network functions for the
Optimized network function composition
Provide suitable APIs and mechanisms for the network function interactions in the compute continuum
Streamlined network function interfaces and interactions

Figure 6-18 Correlation between the enabler "network modules placement in the resource continuum" with the other enablers.

The current E2E system design and architecture design principle are completely complaint with this enabler and no updates are required. This enabler will focus on the M&O component of the E2E 6G system where the interfaces, APIs and techniques will be defined to perform the network module placement in the resource continuum.

6.1.2.3.4 Cloud transformation in 6G-quantum architecture

The mentioned enabler introduces the inclusion of quantum technologies in the current network stack thereby creating a hybrid quantum-enhanced classical network. Inclusion of quantum technology enables an efficient encoding and processing of classical data and reduces the computational overhead logarithmically. Encoding of classical data is a crucial step which can drastically increase as the development towards 6G progresses, thereby creating an overhead for the hypervisor in terms of the processing of an abundance of data. In such a scenario quantum encoding of classical data helps in reducing the traffic by harnessing the property of quantum superposition.

This enabler fits suitably with several crucial system design principles such as minimizing environmental footprint (principle 10) and network scalability (principle 4). Firstly, n-dimensional bit strings can be encoded into log(n) qubits which makes it extremely storage-efficient which directly impacts energy efficiency. Secondly, it is scalable to the continuous inclusion of various layers to provide new functionalities. To include several parameters in 6G nodes, quantum algorithms are scalable with the increasing size and complexity of the network.

Cloud transformation in 6G by the inclusion of quantum technology depends on several other enablers such as "multi-domain/multi-cloud federation" and "network modules placements in the resource continuum". As it is mentioned in enabler "multi-domain/multi-cloud federation", the resource layer is extendable by exploiting external sources. The quantum hardware, on top of which virtualized layers can be added, could possibly be such a resource that can be leased. The quantum notion as a module is another crucial aspect which depends directly on the enabler "network modules placements in the resource continuum".

6.1.2.4 *Enablers for network function modularization.*

The key elements from the analysis criteria for the integration in the E2E system are discussed for the following enablers proposed in [HX223-D32]

6.1.2.4.1 Optimized network function composition

This enabler focusses on the optimization of the network function composition for specific KPIs and deployment options. It starts by analyzing the performance of 5G network function composition and the 5G procedures in terms of including but not limited to the delay and signaling. The enabler defines the dependencies between different modules and NFs. This enabler will increase flexibility, optimized signalling, and efficient resource usage.

In terms of the dependencies, it does not depend on any of the other enablers but may impact several other enablers, such as network exposure, intents etc. Being a fundamental change on the core functions, it can align with all the design principles. More precisely, the module can be customized to meet certain design principles, such as resilience and availability (principle 3), network scalability (principle 4) or flexibility to different network scenarios (principle 5). It will have impact on the CN NF design in 6G since the design need to be different from 5G.

6.1.2.4.2 Streamlined network function interfaces and interaction

This enabler focuses on how the modular design should change based on specific use cases as well as placement locations. The network modules and their interfaces need to support the coexistence of these use cases as well as the related services. Therefore, this enabler will focus on how the interfaces and interactions should evolve to meet the requirements.

Regarding the envisioned impact of this enabler, it is expected to extend the support for new and existing use cases as they could be optimized based on the NF (or network module) placement choices (e.g., centralized and distributed cloud deployments). Being a fundamental enabler that designs the interactions and the interfaces of different modules, it can respond to all design principles here. It depends on the enabler “optimized network function composition”, to draw the methodology as well as outlining the different modularization strategies and their implications. It impacts the NF design and 5G procedures and points the need for new interfaces and interaction.

6.1.2.4.3 Flexible feature development and run-time scalability

This enabler explores the possible enhancements to the E2E modularization (e.g., network slicing in 5G) to optimize the network functionality. It analyses how the modularization affects the network slices and the implications, e.g., being able to provide slice as a meta-module that incorporates different modules. It provides enhanced network slicing and performance, flexibility via modularization, customization of E2E functionality. It will have E2E impacts as the design and placement of network modules through the cloud continuum (e.g., cloud, edge, access, extreme edge etc.) would be revisited. As presented in Figure 6-19, it is built upon enablers “optimized network function composition” and “streamlined network function interfaces and interaction”, and this is a fundamental enabler. Therefore, it can respond to all the design principles.

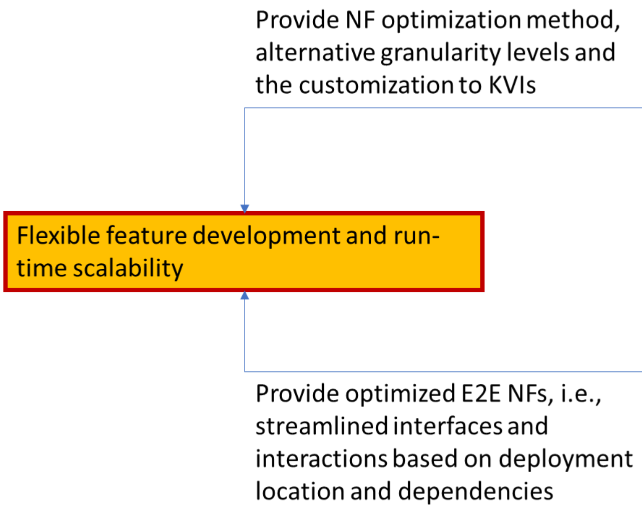

Provide NF optimization method, alternative granularity levels and the customization to KVIs
Optimized network function composition
Flexible feature development and run-time scalability
Streamlined network function interfaces and interaction
Provide optimized E2E NFs, i.e., streamlined interfaces and interactions based on deployment location and dependencies

Figure 6-19: Correlation between the enabler “flexible feature development and run-time scalability” with the other enablers.

6.1.2.4.4 Network autonomy and multi-X orchestration

This enabler focuses on the closed loop orchestration and multi-domain orchestration. It is motivated from the slice management in 5G. In 5G, network slicing was a key enabler to facilitate the co-existence of various use cases with demanding and often conflicting requirements. The M&O are built upon open loop slice configurations and semi-static parameters from SLAs which often result in low resource utilization. It enables improved data-based slice management. With more autonomic and closed-loop based slice orchestration mechanisms it will be possible to address the orchestration of the network services including the extreme-edge domain, which is highly dynamic, heterogeneous, and volatile.

This enabler is expected to have E2E impacts as the NF placement decisions through cloud continuum would be optimized with a higher time granularity based on the network dynamics. It requires closed-loop control and more flexible orchestration mechanisms as well as an enhanced exposure process. It will require also to define a comprehensive information model capturing the peculiarities of those devices in extreme-edge domain. It has dependencies to other three enablers described in section 6.1.2.4 (see Figure 6-20) and can support all the design principles.

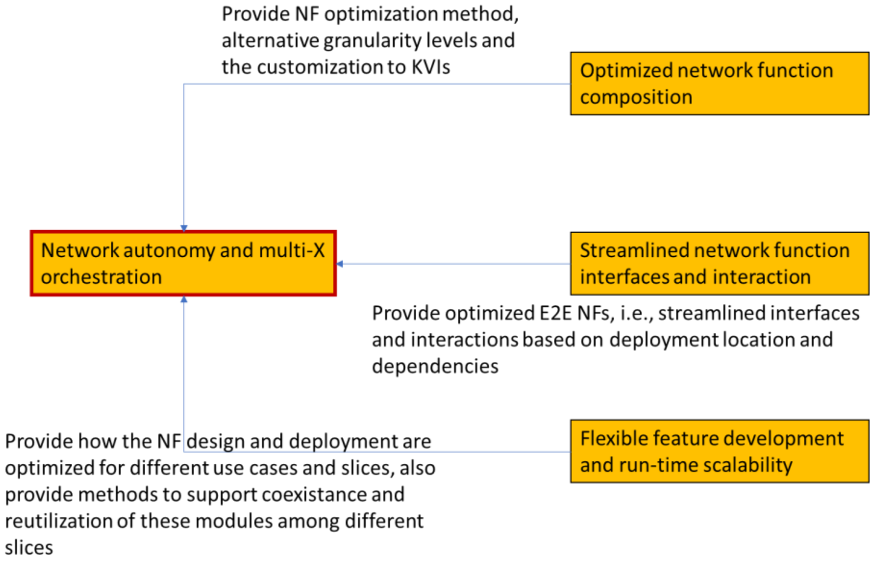

Provide NF optimization method, alternative granularity levels and the customization to KVIs
Optimized network function composition
Network autonomy and multi-X orchestration
Streamlined network function interfaces and interaction
Provide optimized E2E NFs, i.e., streamlined interfaces and interactions based on deployment location and dependencies
Provide how the NF design and deployment are optimized for different use cases and slices, also provide methods to support coexistance and reutilization of these modules among different slices
Flexible feature development and run-time scalability

Figure 6-20: Correlation of the enabler “Network autonomy and multi-X orchestration” with other enablers.

6.1.3 Recommendations for enabler integration

6.1.3.1 Recommendations related to enabler integration in 6G E2E system

The analysis of the enablers under development in the various tasks of the project revealed some potential synergies between different enablers addressing similar aspects and functionalities. Further work will be done in the following interaction clarifying the scope and complementarities of enablers. In this perspective, the concept of an enabler in Hexa-X-II has been further formalized. It is a technical 6G enabler defined as any technical asset that makes it possible to realize or enhance a 6G capability. It is recursive, e.g., 6G system enables new use cases, 6G radio is an enabler of 6G system to achieve system requirements. A 6G technical enabler can be further classified into different types that are extensible, e.g., architecture, system component, process, algorithms, etc. Future work will refine the architecture enablers with the new functionalities and interfaces for all the different options envisaged, particularly if they are contradictory, and relate the mechanisms of the different architecture options to the implementation solutions of components, processes or algorithms enablers.

For the integration in the 6G system blueprint, it is recommended that only enablers that fit with the migration path from 5G to 6G are selected. This does not mean that disruptive enablers are not important, but that their development should be encouraged to prepare for longer-term changes in the network. Cloud transformation in 6G-quantum architecture is one example of exploratory work that is too early for selection consideration.

The rest of the analysed enablers related to intent-based service management automation, to the smart network management and to the virtualization and cloud continuum transformation have been identified for the integration in the system blueprint of the enablers, as detailed in the next section 6.2. The network function modularization concept should be integrated in the system blueprint. However, how and if all related enablers should be part of the system blueprint is reserved for a further iteration as more work will become available to examine their integration.

Beyond the criteria established in this first analysis, some additional information must be taken into consideration for processing the E2E alignment defined in the system design process elaborated in [HEX223-D21]. To be able to properly select the enablers, it is important that enablers are accompanied by the following information:

- In which use cases enablers could be used: Are enablers designs can participate positively to satisfy use cases KPIs and KVIs.
- Dependency between multi-layers enablers (interface, protocol, etc.):
    - How enablers interact with each other between different layers: protocols, interfaces, latency, throughput, etc.)
    - Enablers recommendation for different views: control view, resource (infrastructure) view, network view, etc.
    - In the next iteration, some enablers and components developed from other SNS Stream B projects should be also considered in addition with enablers developed in Hexa-X-II project. Moreover, in the first iteration of enabler analysis, some inconsistencies were identified with the correlation between different enablers. Therefore, in the next iteration, it is recommended to follow up the progress of the design process of those Hexa-X-II enablers while aligning with other SNS-JU projects relevant enablers.

Furthermore, enabler proponents are encouraged to incorporate in their analysis any implications on the security, privacy or general resilience impact the enabler may have, any correction measure that can be applied to address these impacts, and any validation or evaluation experiment suitable to collect additional evidence on impact depth and/or effect of the measures. The different enabler classes discussed in chapter 5, each one including the related threat types they are intended to address, can be used as guidelines for these considerations.

6.1.3.2 Recommendations related to enabler integration in E2E system-PoC

All of the enablers studied within the context of Smart Network Management, have a great value to add in the E2E system evaluation and validation, alas, not all enablers are or will be able to reach the technological maturity that would permit their integration in the system-PoC. Nevertheless, their possible integration or not should not impact whether the recommendations suggested here should be taken into consideration during the enablers' development or not. All these enablers should ensure that the new functionalities and interfaces introduced are well defined and provide a mapping of each enabler to the system-PoC's implementation architecture. Additionally, in the case where there exist multiple development options while evolving the enablers, those should be narrowed down and consider the ones that best serve the E2E functionalities from an architectural perspective.

With regards to the security aspects in system-PoC integration as far as AI/ML-based control is concerned, it is recommended to harden the AI model to be trustworthy, as well as robust against adversarial attacks where the aim of such attacks is degrading the performance of the model. It is important to create more robust and reliable models to guarantee the decisions made by AI-based systems. To create a robust model, one method is adversarial training where some noises are added to data samples to be used in the training phase. In this way, the model will not be affected by the noises which are added to the input of the model by the attacker at inference time. In addition, it is important to ensure the privacy of user in AI-based systems to enhance user trust and adoption to such systems. Another important aspect related to the integration procedures of the E2E system-PoC, concerns the integration fabric component and more specifically the importance of accommodating the interaction between the data producers and consumers. To this end, the provisioning of adapting a common data schema for all the involved systems, infrastructure, nodes, application components, etc. It is necessary that all the messages that will be exchanged, including application layer data, network metrics, logs, etc., are aligned, towards ensuring that the contents and context of all messages are intelligible by all users.

6.2 Foundation of Hexa-X-II 6G E2E system blueprint

This subsection covers the updates to the Hexa-X-II 6G E2E system blueprint and the preliminary perspective of M&O with respect to the system blueprint. It also describes the E2E intent-based service management automation framework with the evolution of telco ecosystem.

6.2.1 Updates to the 6G E2E system blueprint

Figure 6-21 presents an iteration of the E2E system blueprint introduced in [HEX223-D21].

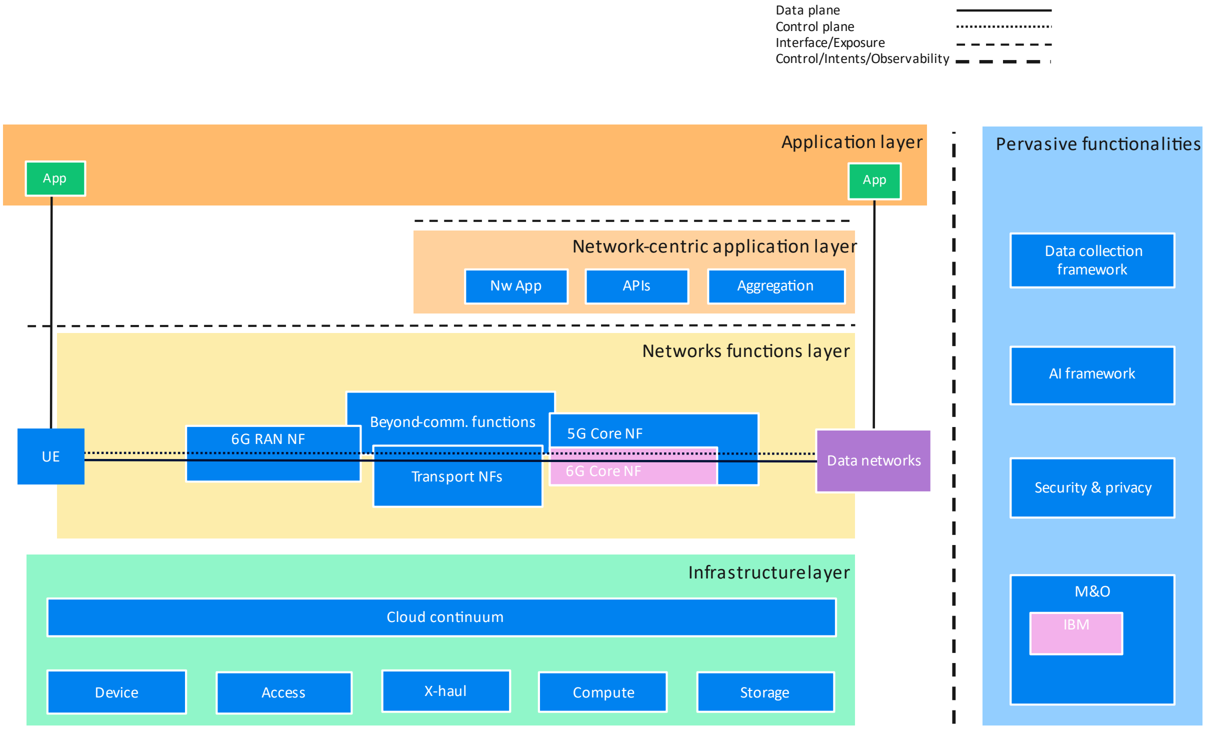


Figure 6-21: Updated E2E system blueprint.

It includes the following updates performed in the latest version of the 6G E2E system blueprint:

- Firstly, RAN NF has been re-named as the 6G RAN NF.
- Next, as part of the migration strategy discussed in the previous blueprint [HEX223-D21], it indicates that the 6G system will use existing 5G Core NF and certain new core NFs specific to 6G represented as 6G Core NFs.
- For architectural homogeneity purpose, the UE abstraction has been set to the same level as the one of the RAN and the Core.
- Furthermore, to better capture the nature of the device, access, x-haul, compute, and storage blocks, which are also represented as part of the device-edge-cloud continuum shorten as cloud continuum, we have reformulated the Infrastructure and Compute layer from the blueprint in [HEX223-D21] to the Infrastructure layer. This is because these blocks provide network and compute resources over which the various functionalities of the 6G system are executed and /or orchestrated.
- Based on the various enablers discussed in section 6.1.2, it is apparent that intent-based management related enablers will touch various aspects of the 6G E2E system. Additionally, they will function under the broad description of what is expected from the M&O block towards 6G system. Henceforth, to reflect such functionality of intent-based management aspects, in Figure 6-21 we introduce an IBM block within the M&O block. A dedicated E2E M&O view covering all of these aspects has been discussed in section 6.2.2.
- An additional aspect that has to be mentioned here, and whose visual representation has been altered in Figure 6-21, as compared to initial blueprint in [HEX223-D21] is the sub-networks feature. While sub-networks will provide capabilities related to device-to-device communication and extend capabilities to just network controlled devices, it is important to state that they represent a topological aspect of the broader 6G system, then only present for certain deployment scenarios. Hence, considering the general aspects of a 6G E2E system blueprint, it is considered that sub-networks are sufficiently covered via the UE and devices in the infrastructure layer and enablers related to sub-networks will still be considered for any future updates to the E2E system blueprint, if mandated by them.

Lastly, it’s worth emphasizing that the different pervasive functionalities interact with each other throughout the lifetime of the network operations that will happen between each of the pervasive network functionalities.

The E2E system blueprint can be further specified with different views that give a focused perspective on some parts and on different aspects of the overall system such as deployment. In this first iteration of the E2E system design, a preliminary iteration of the E2E management and orchestration view provides the focused perspective on the M&O block in the pervasive functionalities as essential for the efficient control and coordination of the complex 6G system. The key ingredients of the E2E management and orchestration view are further detailed in section 6.2.2.

6.2.2 Early E2E management and orchestration view of the system blueprint

First, those enablers described in section 6.1 which encompass the M&O mechanisms are mapped onto the 6G system blueprint in section 6.2.2.1. It highlights how those enablers are tightly integrated with the other pervasive functionalities and interacting with the different layers of the system. Then, the plan is to elaborate the E2E M&O system architecture starting from the baseline M&O architecture from Hexa-X [HX22-D62] evolved with novel Hexa-X-II enablers toward systemization. One element toward this goal relates to the E2E intent-based service management automation framework presented in section 6.2.2.2. The framework is built-upon the analysis of the business relationship between the different 6G stakeholders. This gives the early foundation for further defining the overall M&O functional system architecture in the next deliverable D2.3.

6.2.2.1 Mapping M&O enablers in 6G E2E system blueprint

Figure 6-22 represents the set of enablers analyzed in this document that are part of the “M&O” block in the pervasive functionalities of the E2E system blueprint. It’s worth emphasizing that the given enablers encompass the M&O mechanisms at the various layers of the system, i.e., at the infrastructure, network function, and application levels. At the application management layer, the enablers provide the tool, technologies, and processes for managing a set of digital services beyond only communications to users and applications. Those enablers interfaces on the northbound with the 6G platform customers and automate the processing of service requests expressed as intent, involving the lifecycle management of multiple intent-driven closed loops for E2E digital service orchestration. Such enablers are part of the intent based digital service manager entity functionality and are described in Chapter 4 of this document. They also support aggregating or federating services from other digital service managers owned by distinct service providers to realize the E2E 6G digital services in the 6G multi-stakeholder environment, further detailed in the next section. In their southbound, they interface with the enablers at the network management level. Those latter comprise both the technology enablers related to smart network management from [HEX223-D62] which provide the tools, technologies and processes and the architectural enablers from [HEX223-D32] which concentrate on the evolution of the functional entities and interfaces for the M&O related to the cloud transformation. Some of the M&O enablers are specifically highlighted in the infrastructure layer as they tightly related and acting onto the resource of this layer. Some enablers are also part of other pervasive functionalities of the system blueprint. This implies that each enabler can be comprised of one or several components. There are some components from the programmable network monitoring and telemetry enabler that are part of the data collection framework. Some components related to the sustainable and trustworthy AI/ML based control enablers are identified as part of the AI framework. Some components related to the trustworthy 3rd party management belong to the security and privacy functionality.

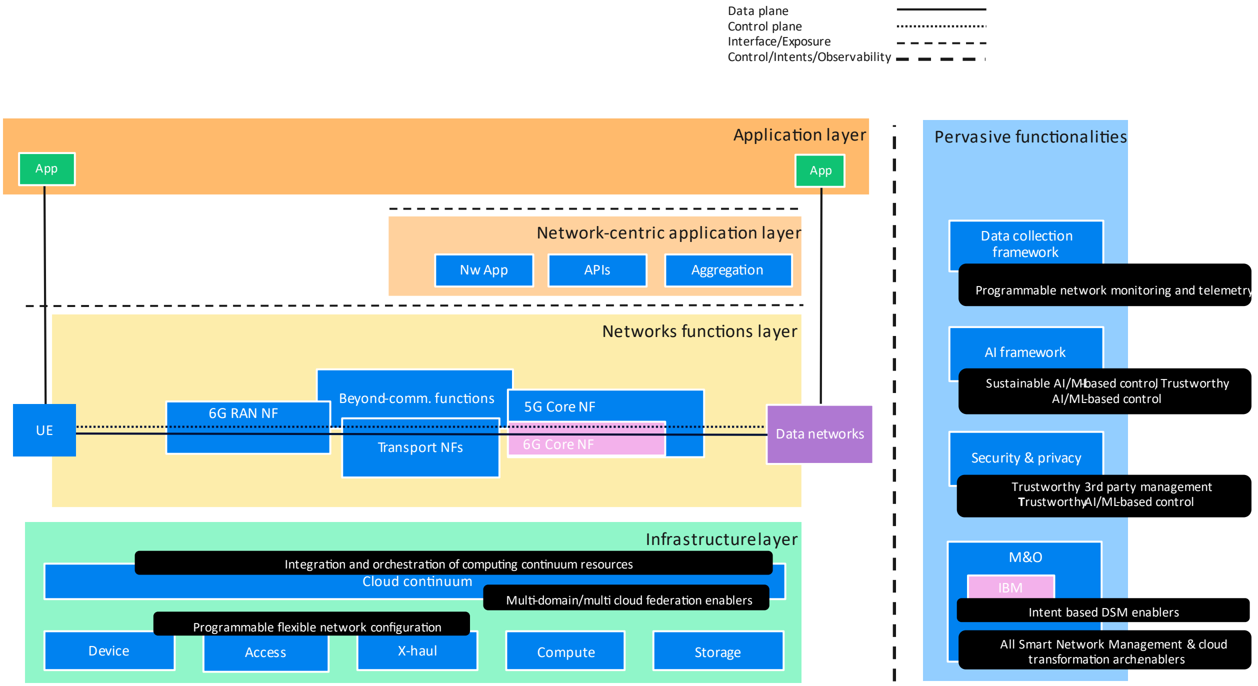

Data plane
Control plane
Interface/Exposure
Control/Intents/Observability
Application layer
App
App
Network-centric application layer
Nw App
APIs
Aggregation
Networks functions layer
UE
6G RAN NF
Beyond-comm. functions
Transport NFs
5G Core NF
6G Core NF
Data networks
Infrastructurelayer
Integration and orchestration of computing continuum resources
Cloud continuum
Multi-domain/multi cloud federation enablers
Programmable flexible network configuration
Device
Access
X-haul
Compute
Storage
Pervasive functionalities
Data collection framework
Programmable network monitoring and telemetry
AI framework
Sustainable AI/ML-based control, Trustworthy AI/ML-based control
Security & privacy
Trustworthy 3rd party management
Trustworthy AI/ML-based control
M&O
IBM
Intent based DSM enablers
All Smart Network Management & cloud transformation arch.enablers

Figure 6-22: Mapping of M&O enablers in 6G E2E system blueprint.

6.2.2.2 E2E intent-based service management automation framework

Within the different work actions done in the Hexa-X-II project, one of them is the definition of the context for the intent-based and automation service management. This subsection focuses on the baseline telco framework with regards to the management of E2E services, and how Hexa-X-II proposes to evolve it towards 'TechCo' framework, with a wider scope in terms of service offerings and flatter roles. This TechCo framework embraces new services beyond traditional connectivity, with focus on digital and application-centric services resulting from an innovation ecosystem leveraging frictionless interactions between network and 3rd party application providers. In pursue of this objective, the sub-section is structured as follows:

- SotA review on system architecture (section 6.3.2.2.1) and federation approaches (section 6.3.2.2.2) in telco environments. These constitute baseline solutions.
- From Telco to TechCo ecosystems, and the need to re-define today's operational roles (section 6.3.2.2.3).
- E2E Intent-based service management automation framework (section 6.3.2.2.4), resulting from evolving baseline solutions into something useful for TechCo ecosystems.

6.2.2.2.1 Baseline telco OAM system architecture

Telco systems are complex, as it involves the administration, operation and maintenance (OAM) of a myriad of hardware and software resources that are i) deployed and installed in a distributed infrastructure; ii) combined into cloud and network functions, which collectively support the execution of end-to-end communication services to multiple customers, from different market segments. The creation of simplicity out of this complex ecosystem requires applying the principles of *abstraction* and *separation of concerns* when designing these systems. Taking these recommendations into account, Tier-1 telcos (i.e., large mobile network operators worldwide) typically structure their OAM systems into different layers, each with a confined scope that can evolve independently from the rest of layers.

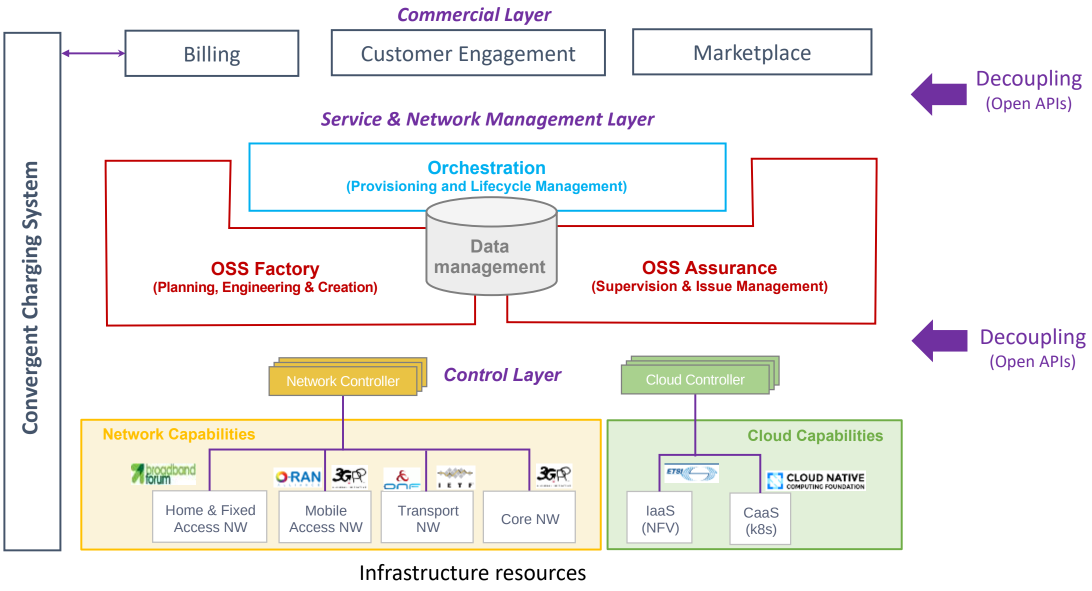


Figure 6-23: Baseline telco OAM system architecture.

Figure 6-23Figure 6-23 pictures a simplified representation of an archetypal telco system architecture for OAM activities, aligned with the principles noted above. As seen, this architecture consists of three layers.

- **Control Layer**: it integrates multiple controllers, each handling the semantics of one or more functions. As for the cloud functions, typical controllers include SOL005-based NFVO (for IaaS solutions) and k8s (for CaaS solutions). As for the network functions, we may find Software-Defined Transport Controller [TIP21] (for the transport network domain) or 3GPP Element Management Systems (for the mobile access and core network domains).
- **Service and Network Management Layer**: it conveys bespoke Operation Support System (OSS) functions, policies, and workflow handlers that telcos may use to manage their assets across their entire lifetime. It aggregates the capabilities made available by the control layer, and use them to have a holistic, E2E view.
- **Commercial Layer:** This is typically referred to as Business System Support (BSS). It is formed of multiple components with their own cycles of technology maturity and adoption curve [Abr22].

6.2.2.2.2 Telco federation approaches (SotA)

There are situations where the network domains building out a beyond-5G service construction span two or more administrative domains. To keep service behavior consistent throughout the E2E path connecting the service endpoints, the management systems from the involved administrative domains need to communicate with each other. This approach, referred to as *federation*, ensures coordination across the domains on decisions related to service lifecycle management, like resource segregation and allocation, connectivity configuration, or scaling.

Federation has traditionally built upon a connectivity model based on wholesale agreements involving Tier-1 telcos and operators of interconnection networks. This model from a technical and legal perspective requires establishing long-term relationships in advance to account for aspects, among others, such as operational reliability, financial compensation, regulatory constraints, and capacity management. Though workable in 4G and 5G NSA, this approach is rather static and do not fit well with the dynamism inherent to 5GSA onwards, in terms of traffic load, resource allocation and service lifetime. In the new ecosystems, novel solutions need to be explored as assessed, by putting the focus on the following open issues:

- How the interfaces, protocols and APIs for the federation should look like? Which body should be responsible for their standardization?
- Should federation be based on peer interactions between administrative domains, or should a third-party act as a broker to mediate these interactions?
- What discovery, routing and access control solutions apply in federation scenarios?
- How to charge when federating administrative domains that are owned by different stakeholders? What are the charging levers?
- How cloud-native deployments can easily, transparently, and dynamically roam onto federated environments, along with all the security, resiliency requirements entailed?

The literature review on telco federation comprises multiple sources, including standardization bodies, H2020 projects, and research papers. The following paragraphs elaborate on them.

From standards viewpoint, **ITU-T** has provided guidelines and specifications [ITU18] for network federation in inter-cloud computing, to help ensure interoperability between different networks and facilitate the efficient use of resources. **ETSI** has also elaborated on this topic across two ISGs: ZSM, which defines a reference system architecture [ZSM002] leveraging a cross-domain integration fabric to federate resources from different management domains; and NFV, which reports on potential architecture options to support the offering virtualized network services across multiple administrative domains. These options focus on virtualization execution environments, and include hierarchical (i.e., reuse of carrier-grade interfaces like Or-Vi and Os-ma-nfvo) and peer-to-peer approaches (i.e., definition of novel interfaces like Or-Or [IFA030] [SOL011]). **MEF** pretends to go a stop beyond, scoping also non-virtualized environments and considering federation at multiple layers (e.g., business, services, resources). These features are illustrated in the Lifecycle Service Orchestration (LSO) reference architecture [MEF19], which considers three types of domains for the definition of federation interfaces: service provider domain (e.g. network operator), partner domain (e.g., partner operator or any 3rd party acting as broker) and customer domain (e.g., enterprise customer in B2B market segment, and end-user for B2C). These interfaces include LSO:CANTATA and LSO:Allegro. A comparative analysis of MEF and ETSI NFV federation interfaces are reported in [OTR23]. Finally, **GSMA** is working out details on East-Westbound Interface (E/WBI) [OTR20] to federate capabilities across different telcos for more than three years, since the kick-off of the Operator Platform Group (OPG) [OPG].

Different H2020 projects have also explored the CSP federation problem, defining their own solutions at both services and resource levels. Some projects have proposed solutions based on peer-to-peer interactions between orchestrations, as it occurs in occurs in 5G-Exchange [5GEx17-D22] and 5G-TRANSFORMER project [5GTr19-D24]. However, there are others like the 5G!Pagoda project [5GPa19-D43], where a hierarchical approach is taken, with one parent orchestrator interacting with different domains, each operated with a child orchestrator. 5G PPP Phase 3 projects fled from the 'peer-to-peer vs hierarchy' binomial and moved the federation discussion towards the concept of controllable capability exposure, built on the idea that one single federation interface is enough, as long as the primitives and data model contained in this interface can be tailored to the needs of the different partner operators. Examples of these solutions can be found in projects such as 5G-VINNI (ICT-17) [5GV19-D31], 5Growth (ICT-19) [5Gr19-D21] and 5G-CLARITY (ICT-20) [5Gr19-D21]. Hexa-X enriches these solutions and unifies them into one single layer, referred as to the API exposure Gateway [HEX22-D62].

Finally, as for field of research papers, we find a variety of works, including surveys [VMA+18],[TAS+19],[LGL+21],[SKR+20] and transaction-style articles, with focus on the design, specification and validation of solutions to address different 5G and beyond scenarios. Within the literature review, there are three clusters of solutions to implement federation that worth noting: interface/API solutions ([OTB21], [PMC22]), AI/ML driven solutions ([AMG+20], [BMB21],[ARR+22]) and blockchain ([BLS+22], [AB20],[AB22]).

6.2.2.2.3 From 'Telco' to 'TechCo' ecosystem

In the definition of actor-role model for 5G [28.530], 3GPP notes two main roles in telco ecosystem:

- Network Operator (NOP), focused on managing network and cloud resources, and their composition into upper-layer network constructions, including domain-specific (e.g., sub-networks, network slice subnets) and E2E constructions (e.g., network slices).
- Communication Service Provider (CSP), focusing on building communication services and delivering them to targeted customers, including end-users (i.e., Business to Consumer -B2C-), enterprise customers (i.e., Business to Business -B2B-) or other CSPs (i.e., B2B to X -B2B2X-).

Tier-1 telcos typically embrace both roles, mapping them to different organization's units. In the telco system architecture pictured earlier, one can easily note that i) control layer is within the scope of NOP; ii) commercial layer is within the scope of CSP; iii) the network & service management layer is targeted by NOP and CSP, with the first focusing on management of domain specific and E2E network constructions, while the latter on design, provisioning and assurance of communication services.

In the 6G ecosystem, it is expected for Tier-1 telcos to become TechCos [STL23], with a wider scope on service offerings, and further flexibility on the composition and operation of managed resources. This means:

- TechCos service offering will not be limited to communication services (collection of PDU sessions exchanged between devices and data networks), but to other digital services (e.g., Web3, big data, security services). All these services will be also offered to new customers (e.g., aggregators) through APIs (instead of traditional channels). All this new casuistry motivates the transition from CSP to a DSP (Digital Service Provider).
- TechCos will articulate their systems with much more granularity, breaking functions into its fundamental components (microservices), each representing stand-alone capabilities that can be individually programmed and chained on a per service/use case needs. This, alongside the ever-increasing offloading models in telco industry (e.g., with new emerging actors such as TowerCos, FiberCos, neutral hosts), also motivates the decomposition of the NOP role into capability operator (COP) and resource provider roles.

6.2.2.2.4 High-level functional description of E2E intent-based service management automation framework

The framework involves different actors and entities and their interactions with the Intent-based DSM IME (section 4.1.2). It also comprises the key element on the management and control of intent requests. The architecture aims to be valid for two different options: DSP aggregation (Figure 6-24) federation (Figure 6-25).

With the objective to enable the transition to the “TechCo” ecosystem, the framework is designed with four layers: the Digital Service Customers (i.e., tenants) on top, the Digital Service Provider (DSP) which its internal architecture was presented in subsection 4.1.2, the Capability (e.g., Network, Cloud, etc.) Operators layer and, finally, Resource Providers layer.

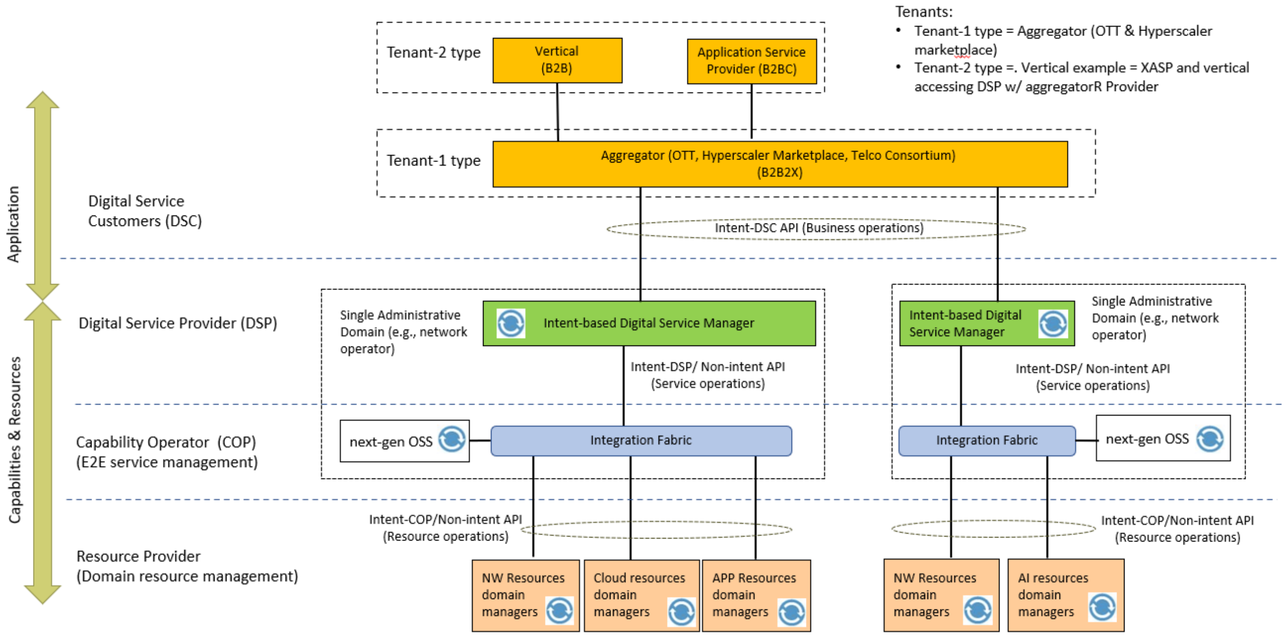

Tenants:
• Tenant-1 type = Aggregator (OTT & Hyperscaler marketplace)
• Tenant-2 type =. Vertical example = XASP and vertical accessing DSP w/ aggregatorR Provider
Tenant-2 type
Vertical (B2B)
Application Service Provider (B2BC)
Tenant-1 type
Aggregator (OTT, Hyperscaler Marketplace, Telco Consortium) (B2B2X)
Intent-DSC API (Business operations)
Application
Capabilities & Resources
Digital Service Customers (DSC)
Digital Service Provider (DSP)
Single Administrative Domain (e.g., network operator)
Intent-based Digital Service Manager
Intent-DSP/ Non-intent API (Service operations)
Capability Operator (COP) (E2E service management)
next-gen OSS
Integration Fabric
Intent-COP/Non-intent API (Resource operations)
Resource Provider (Domain resource management)
NW Resources domain managers
Cloud resources domain managers
APP Resources domain managers
AI resources domain managers

Figure 6-24: E2E Intent-based service management automation framework (Option A: DSP aggregation).

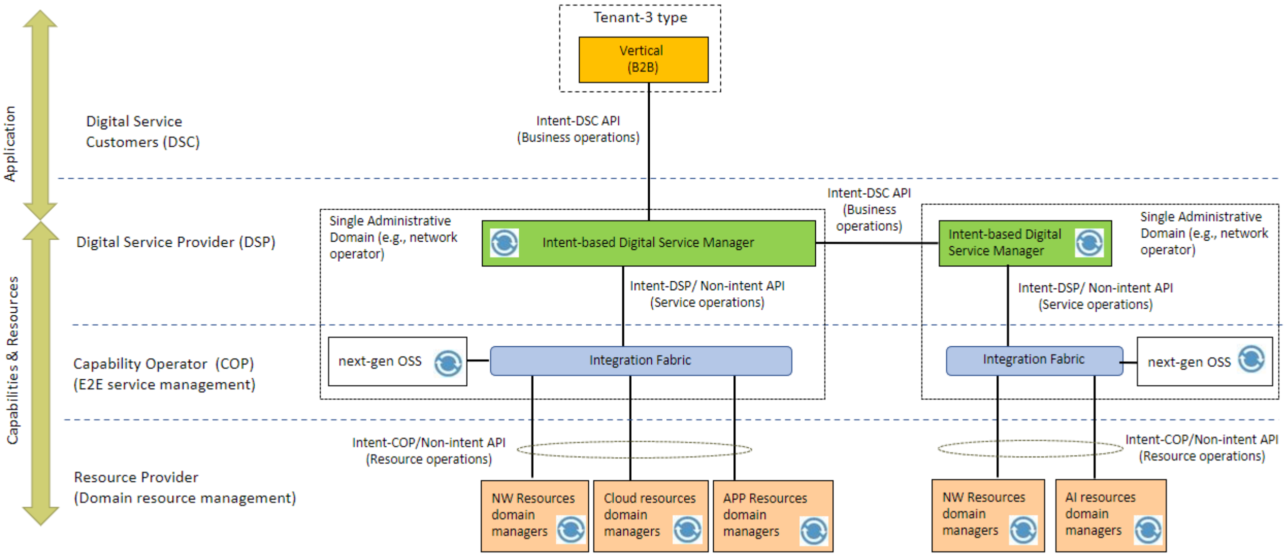

Tenant-3 type
Vertical (B2B)
Intent-DSC API (Business operations)
Application
Capabilities & Resources
Digital Service Customers (DSC)
Digital Service Provider (DSP)
Single Administrative Domain (e.g., network operator)
Intent-based Digital Service Manager
Intent-DSC API (Business operations)
Intent-DSP/ Non-intent API (Service operations)
Capability Operator (COP) (E2E service management)
next-gen OSS
Integration Fabric
Intent-COP/Non-intent API (Resource operations)
Resource Provider (Domain resource management)
NW Resources domain managers
Cloud resources domain managers
APP Resources domain managers
AI resources domain managers

Figure 6-25: E2E Intent-based service management automation framework (Option B: DSP federation).

As given in figures and regarding the tenants, three different tenant types interacting with the DSP are defined:

1) An aggregator tenant which includes actors such as Hyperscaler Marketplaces, Telco Consortium, etc. that follow the B2B2X model and offer the intent-based services to a second type of tenants.
2) This type of tenant groups either Verticals following a B2B model such as an XR provider or Application Service Providers following a Business to Business to Consumer (B2B2C) model.

These two possible tenants, obtain the intent-based services through the interaction of an aggregation (tenant-type 1).

3) Verticals may be a third type of tenants by having direct access to the offered intent-based services, for example banking companies.

These three types of tenants are able to interact (directly or not) with the Intent-based DSM at the DSP layer to generate and agree on the intent request with the expected service and targets associated to deliver the service to the final user.

Below the tenants, the DSP is the actor in charge to receives the incoming intent-based requests from the tenants and manage them to achieve the right translation into a set of specific requests for a selected set of the available Capability Operators (COP) below. At the DSP layer, an Intent-based DSM is placed at each single administrative domain leading the management (i.e., provisioning, monitoring, etc.) of intents using the resources offered in the layers below by the capability operators and providers through the “Integration Fabric” enabler introduced in subsection 6.1.2.2.3. While one single DSP has a lot to of services to offer, it is expected that will not always be able to deliver all services and for this reason, a federation approach (Figure 6-25) was included to allow the interaction of different DSP through their Intent-based DSM using what is called as the Intent-DSC API. Regarding the internal architecture of functionalities and capabilities offered by the Intent-based DSM, this topic was properly addressed previously in subsection 4.1.2.

The third identified actor within the framework are the COPs, which are operating those entities/elements in charge of managing the domain resources available. A COP is handling the control and management elements such as a Network Slice Management Function (NSMF) or other domains specific managers, all based on enablers such as those identified in subsection 6.1.2.2. Among enablers, the most important in terms of allowing the right communication between the DSP and the COP layers is the Integration Fabric.

Finally, the last layer involves the Resource Providers across different domains. The Resource Providers offer a set of different domain resources such as applications, AI, network (including RAN, transport, and core) and finally, cloud resources (extreme-edge, edge, and cloud). The type of resources domain manages that each single administrative domain has available may vary depending on the needs, the two previous figures illustrated one of the many possibilities that may exist.

7 Preliminary E2E system-level evaluation results

This chapter describes the process of the Hexa-X-II system assessment with respect to the proposed KPIs and KVIs, both described in section 7.3. The validation process of the E2E system will be driven by the design and implementation of Proof-of-Concepts for applications that leverage 6G capabilities, such as robotics and extended reality. Within the scope of the E2E validation, various 6G enablers will be assessed, including aspects related to management and orchestration over the 6G continuum, network transformation, network control programmability and telemetry, 6G devices, radio protocols and sensing. The basic framework for the development of system PoCs was first described in D2.1 [HEX223-D21]. In this deliverable a detailed description of System-PoC A is now given, accompanied by some early results. In section 7.1, a brief summary of all system-PoCs is provided along with a short introduction to the different scenarios that will be employed to showcase system-PoCs' progress, in section 7.2 the System-PoCs' architecture is described, followed by the components that constitute System-PoC A. In section 7.3 the related KPIs and KVIs to the System-PoCs' evaluation results are analyzed, and in section 7.4 System-PoC A's preliminary results are presented. Lastly, in section 7.5 is discussed a simulation / digital twin-based approach, used to enable further evaluation of selected KPIs of the E2E communication system.

7.1 Overview of the E2E system evaluation and validation activities

Hexa-X-II project's goal of transitioning from the early development phase of 6G technology to a coherent 6G systemization driven by the need for sustainability, trustworthiness, and inclusion, requires a holistic view encompassing devices, infrastructure, novel radio, and network capabilities, E2E management and orchestration, along with security, and system-level resilience considerations. To this end, three system-PoCs are designed to be developed in a gradual manner.

- The first system-PoC (A), which is currently being developed and its design and preliminary results are presented in this chapter, focuses on smart network management aspects for demonstrating management mechanisms, leveraging Hexa-X [HEXA] achievements and ensuring the progression of 6G journey from incorporating strong inputs from Hexa-X and other 6G EU, national and international projects, to the first public PoC.
- The second system-PoC (B) will mainly focus on network architecture elements and refinements of the management mechanisms, introduced in System-PoC A.
- The third system-PoC (C) will focus on radio and devices aspects while enablers introduced in the two previous system-PoC versions, i.e., 6G architecture design, and smart network management will be further evolved.

Even though system-PoCs B and C are designed in such a way so that new enablers (e.g., flexible topologies, new radio devices, etc.) are to be incorporated in each one in an incremental manner, each system-PoC is also intended to be enhanced throughout the course of this project. This ongoing process is planned to be achieved by iteratively giving feedback to the corresponding integrated enablers of the system-PoCs for improvement and optimization, where in turn these updated enablers are to be incorporated in the respective system-PoCs. Furthermore, three different scenarios are planned to be executed along the project's duration for encompassing the development or evolution of the involved enablers, along with adding new features to enhance the PoCs' functionalities. A description of the proposed scenarios is given in the following section.

7.2 E2E system-PoC architecture and components

7.2.1 E2E system-PoC architecture

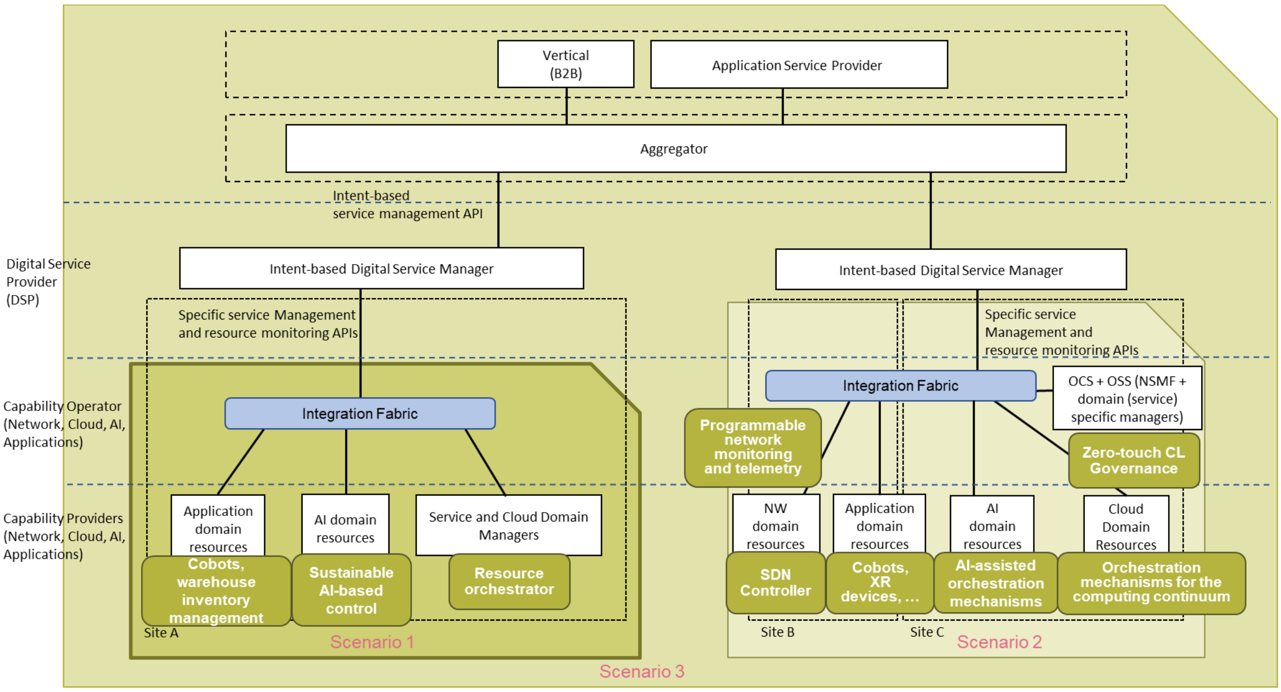


Figure 7-1: An overview of the system-PoCs from the DSP architecture perspective.

The E2E system-PoC architecture is being presented in Figure 7-1. This architecture is based on the proposed system architecture as shown in Figure 6-24, and it consists of 3 main levels, the Digital Service Provider (DSP) level, the Capability Operator level, and the Capability Providers level. The M&O enablers that are exhibited in system-PoCs are located mainly in the two lower levels of Capability Operator and Capability Providers, with the exception of the Intent-based Digital Service Provider that lies within the DSP level. In the presented architecture, three different sites (site A, B and C) are also being considered for showcasing multi-site / multi-domain synergetic orchestration and monitoring.

Three scenarios are selected to showcase the evolution of system-PoCs. These scenarios are designed in an incremental way, in order to complement system-PoCs' evolvement. The first scenario, displayed on site A in Figure 7-1, corresponds to an autonomous, single-domain scenario; the second depicts an E2E, multi-domain scenario; and last, in the third scenario the use of intent-based enablers is introduced for managing different domains from an E2E perspective. More precisely,

- For the first scenario, a simple network configuration is assumed, where all network components lay in a single domain, since this scenario deploys one site A. Here, an integration fabric, at the capability operator's level, is used to facilitate the integration of the various application domain resources, AI domain resources and service and cloud domain managers that lay in the capability provider's level.
- In the second scenario exhibited, the previous configuration is expanded by assuming that not all of the network and cloud domain resources, application components and devices are contained in the same domain and/or geographical location (depicted as sites B and C in Figure 7-1). Therefore, it is required to be able to orchestrate all these components in a centralized way. A network slice management function will be introduced along with domain specific managers.
- In the third, and last, scenario of the system-PoCs' progression, the objective is to reach the correct performance of the whole system composed of different domains being managed from an E2E point of view through the use of intent-based enablers.

As illustrated in Figure 7-1, each of the two domains has a layer on top with an Intent-based DSM component. The Intent-based DSM brings the capability to receive intent-based requests originated from the final users (e.g., a vertical or an application service provider) and manage them to deliver the desired service.

The Intent-based DSM should be able to deal with the whole process to receive what the user expects, and find the way (i.e., how) to achieve it using the M&O elements offered at the Capability Operator layer. Due to the nature of intents (they define the what, not the how), the Intent-based DSM has the responsibility to interact in both ways; first with the user to reach an agreement between them without doubts, and secondly, with the network management elements below to find the best configuration possible (i.e., how) to reach what is expected by the user. Moreover, a third interaction is required between Intent-based DSMs in case they need resources that are not available in their own domain, but in other Intent-based DSM domains. To do so, a federation model is expected to be designed and implemented.

7.2.2 System-PoC A components

A detailed description of the components of the E2E architecture follows, along with the use cases that are to be executed in system-PoC A.

7.2.2.1 Application, service, and cloud-domain resource orchestration

For System-PoC A, two different configurations are being assumed for exhibiting the system’s capabilities.

In the first configuration, the resource domain provides a testbed to develop and evaluate a cutting-edge automated inventory management (audit) solution that utilizes drones/unmanned aerial vehicles (UAVs) and autonomous mobile robots (AMRs) for precise and efficient warehousing operations. The key components include state-of-the-art fusion of computer vision and sensor data, to ensure that objects are accurately identified, counted, and localized in real-time, and a dynamic translation system between symbolic warehouse locations and 3D geometric coordinates. This facilitates seamless drone and AMR navigation along with accurate inventory pinpointing.

In this cobot-powered warehouse inventory management scenario several user stories will be studied and developed utilizing the proposed infrastructure. *Resource allocation* is one such user story, for which preliminary results are presented later in this chapter. In the resource allocation scenario, the objective is to optimize the placement of a) the inventory management services (e.g., allocation of item scanning cobot role, among UAVs/AMRs), and b) workloads requiring considerable computational resources (e.g., computer vision tasks, such as cobot video camera feed real time processing for obstacle detection across an AMR’s path), based on current workload, energy availability (for mobile, battery-operating devices), hardware capabilities (e.g., ground/aerial node), as well as physical environment parameters, such as real-time proximity to the inventory locations. Some of the services/roles that will be offered/needed in this scenario include *path planning*, *object detection*, *quality inspection*, *warehouse digital twin*, and *inventory management*.

In Figure 7-2 the system architecture is provided, comprising the orchestration and monitoring components, the AI domain resources (e.g., Energy efficient Functionality Allocation algorithm), the inventory management-specific services (e.g., Object detection, Path planning, etc.), the user interfaces, as well as the network domain resources. The ultimate goal of this configuration is to showcase resilient and trustworthy operation scenarios in Warehouse and Manufacturing environments, leveraging AI-assisted, Trust- and Energy-driven optimization, considering network reliability, energy availability, and compute continuum nodes performance.

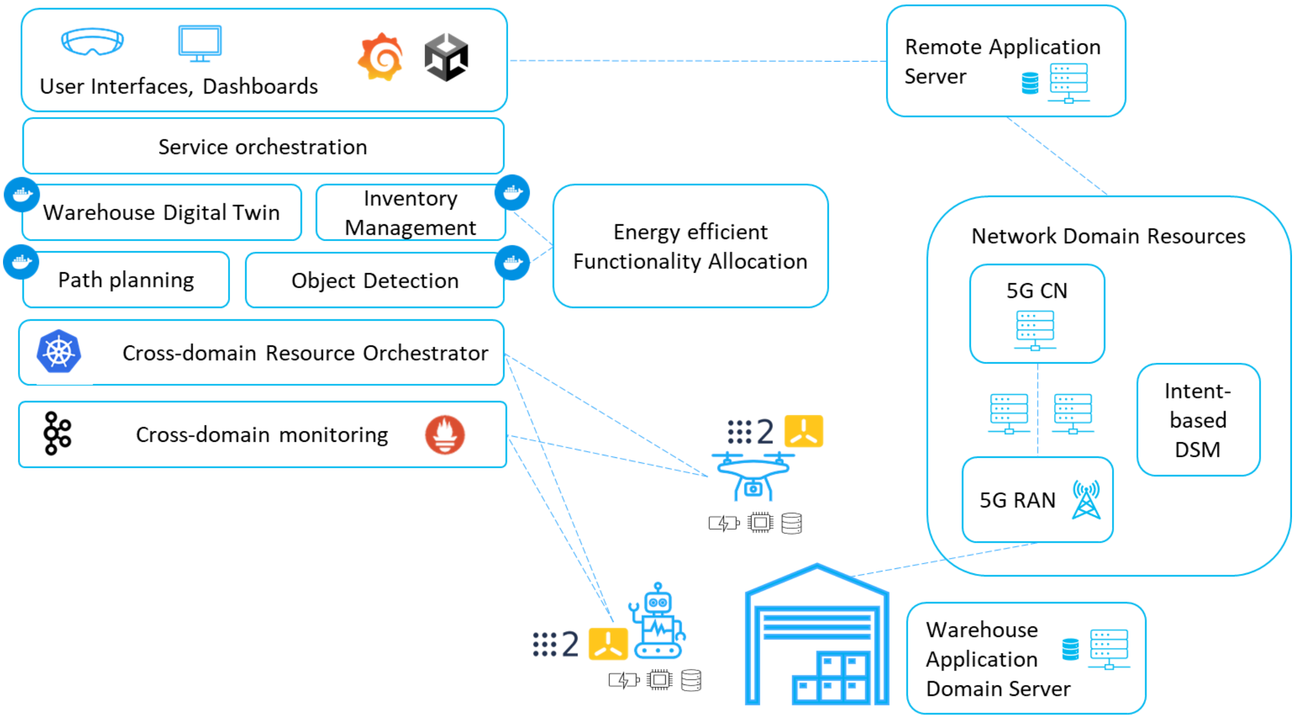


Figure 7-2: An overview of the platforms and tools configuration for the warehouse inventory management.

In the second configuration, the resource domain is assumed to provide a testbed to implement and evaluate two cobot use cases as examples: 1) Task Continuity and 2) Remote repairing through XR. The Task Continuity scenario consists of a robot on a surveillance mission offloading mission-critical computation to the edge cloud. To ensure zero downtime, when the robot is about to run out of battery, a new robot will go to the latest surveillance location, and the old robot will go to the charging station. This scenario will use intent-based enablers to maximise the reliability there by improve trustworthiness of the network for the end devices. In the Remote Repairing Through XR scenario, a device malfunction alert is triggered, and technicians are sent to the site. With XR devices, technicians can gain a better understanding of the scene beforehand by identifying faulty equipment, monitoring the temperature on overlays, visualizing poor network coverage, or displaying repair instructions. The development of this scenario will take part in a future stage.

As shown in Figure 7-3 below, the PoC architecture consists of two interconnected domains: the application domain and the network layers. The application domain encompasses multiple services: The robot services, such as remote control and surveillance, and the collection of application metrics, XR services used for remote repair, and other application server services, such as object decomposition algorithms (e.g., YOLO) and messaging applications (e.g., an MQTT broker). In practice, each application service is containerised and deployed through Kubernetes to facilitate high-level application service networking and container management for ease of deployment. The network domain provides connectivity and manages the network resources of all system devices. In the given resource layer, a 5G Stand-Alone network is used to connect all devices. Additionally, an SDN controller manages network slices, which guarantees QoS to the services running on the system.

In PoC-A, a strong interconnection exists between the application and network domains as described in the above paragraph. Changes at the application layer will affect the network configuration and vice versa. For instance, in the task continuity scenario, to ensure network resilience, the surveillance robot is allocated a network slice. When replacing the robot for task concurrency, the network will reallocate the network slice from the old robot to the new robot. In Figure 7-4, the PoC architecture is detailed, displaying logical and physical connections among components.

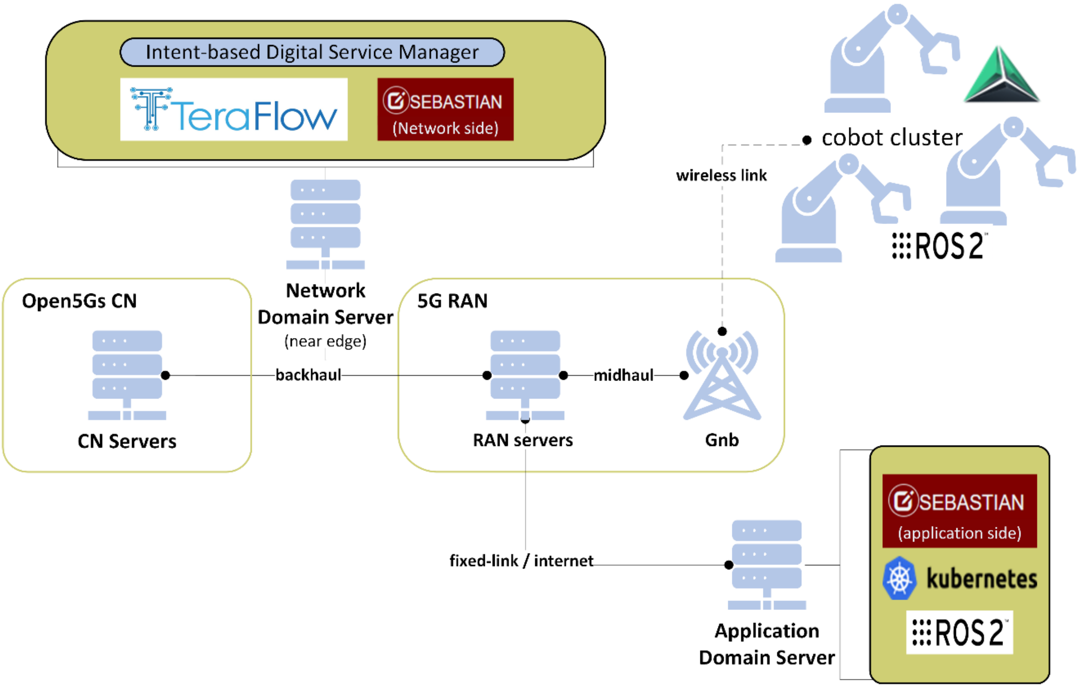


Figure 7-3: PoC-A system architecture for task continuity and remote repairing scenarios.

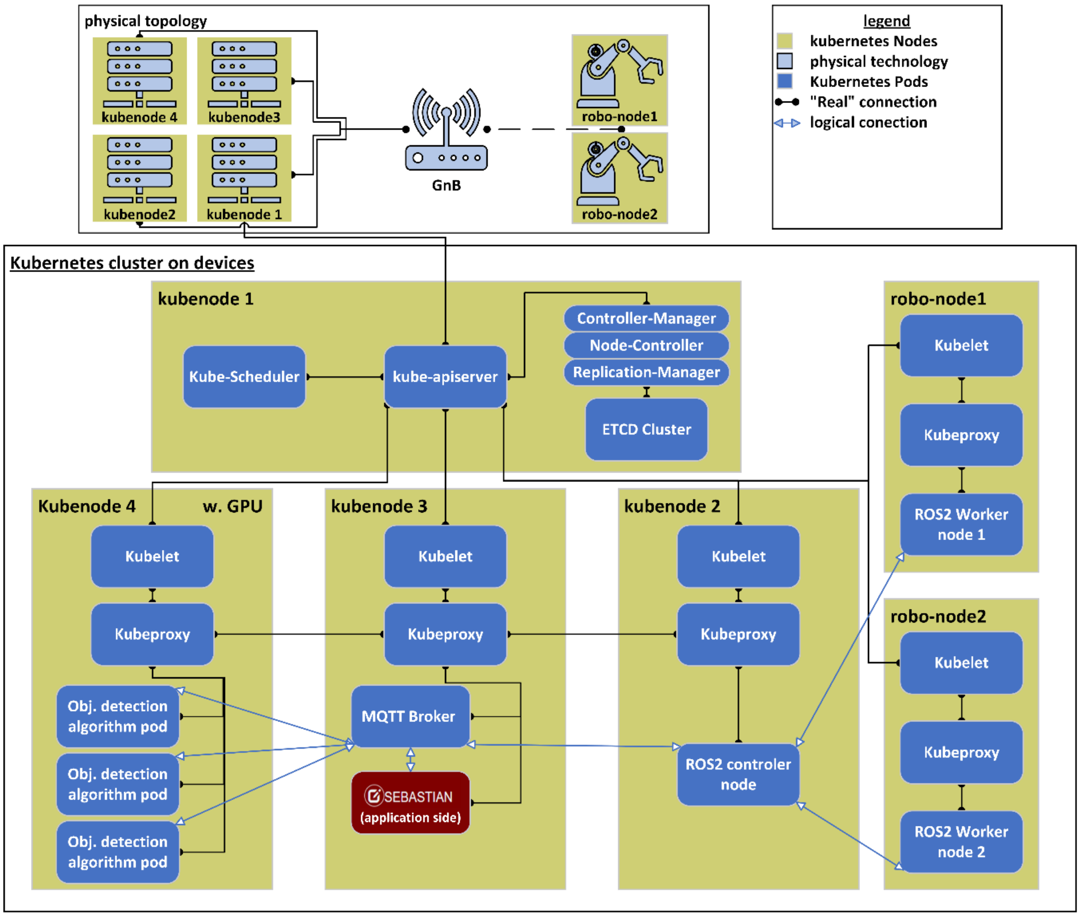


Figure 7-4: PoC-A application software diagram for task continuity scenario.

7.2.2.2 Closed-loop automation

System-PoC A aims at demonstrating Closed-loop (CL) automation via a progressive integration of technological enablers coming from smart network management. The main idea is to implement 3 workflows characterized by a growing complexity able to guarantee the automatic management for the service based on cobot applications described in Section 7.2.2.3. It is important to note that such workflows require a maturity level on the involved components that is expected to be achieved in the long period, beyond the PoC A timeframe. For this reason, the work to implement the CL started in PoC A will also continue in PoC B.

The foreseen workflows are the following:

> **Workflow 1 -** *Threshold based closed-loop and governance*. In this workflow, the CL decision is **reactive**, i.e., based on real-time measurement of cobots' battery level.
>
> **Workflow 2 -***AI/ML based closed loop and governance*. The CL decision is **predictive** i.e., on cobots' battery level prediction (AI/ML-driven).

Workflow 3 -*Multiple closed loops with governance & coordination*. This scenario involves two different CLs that act at Service and Network level, respectively that required to be coordinated to guarantee the service continuity. Workflow 1 is described in Figure 7-5. The scenario considers a set of cobots connected to the mobile network. Specific applications for them run on a couple of edge servers, linked to the UPF through a somehow complex transport network. This environment is the same per each workflow.

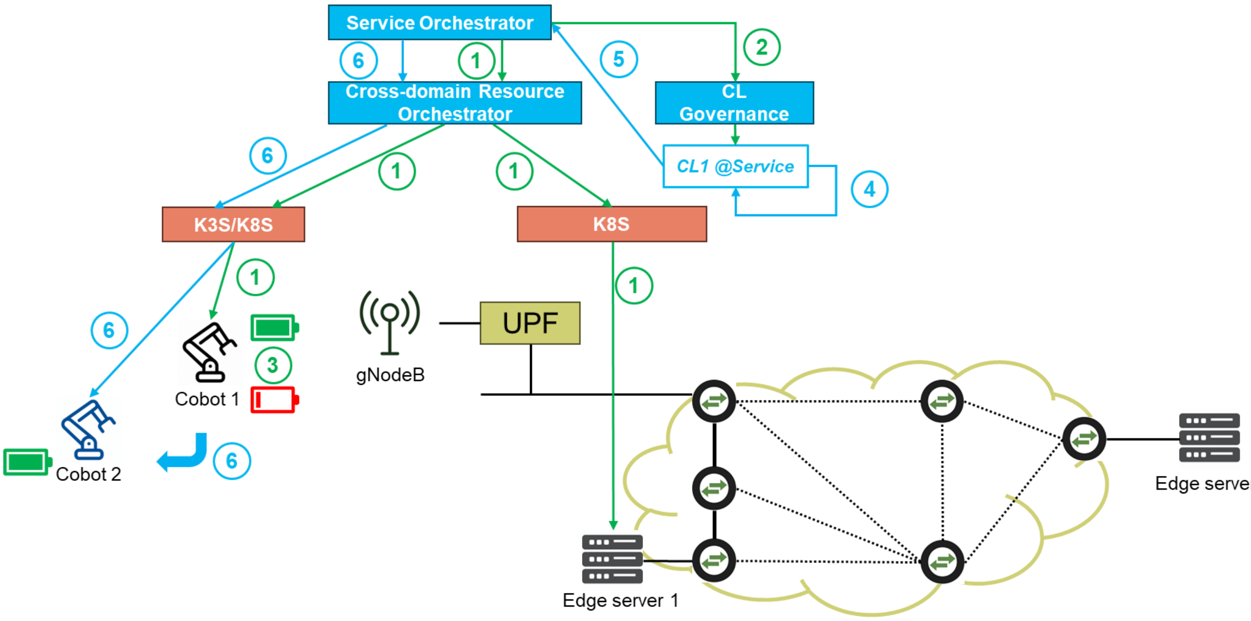


Figure 7-5: Reactive CL based on real-time measurement of cobots' battery level

Step-1. The E2E service is provisioned. Cobot 1 and Edge Server 1 are configured accordingly while the transport network i.e., the path between the UPF and the Edge Server 1, is considered pre-allocated.

Step-2. The Service Orchestrator invokes the CL Governance to provision the CL at Service Level.

Step-3. The battery level of Cobot 1 falls below a given threshold.

Step-4. The CL detects the battery is draining and decides to provision Cobot 2 to guarantee the service continuity.

Step-5-6. CL request the Service Orchestrator to enforce the decision on the system.

Workflow 2 follows the same steps. The advancement with respect to the previous workflow is given by the analysis and decision steps of the CL that this time are AI/ML based and able to take a decision on a predictive manner. In this way it is possible to optimize the usage of the cobots and the time for re-charging them.

The last workflow aims at demonstrating the coexistence of multiple CLs and their coordination: one loop is the same as previous workflow (service level), while a second is deployed to automatize the management of the transport network. For that reason, the scenario enhanced with a network slice manager (NSMF) that manages the transport network though an SDN controller, and with a CL Coordination service coordinate both the loops.

To ease the explanation, this workflow is described by using two different figures (Figure 7-6 and Figure 7-7). Furthermore, for the sake of simplicity, the cobots' service is considered already provisioned.

Step-1. The service is considered pre-provisioned. In this case, the CL Governance provisions an additional CL (CL2) for the network management.

Step-2. An outage happens on one of the network nodes belonging to the pre-provisioned path between the UPF and the Edge Server 1 (where cobots' edge application is running).

Step-3-4 The CL2 detects the network outage and decides to contact the NSMF to address the issue and guarantee the service continuity.

Step-5-6. The NSMF requests a new path to the SDN controller that accordingly configures the network nodes.

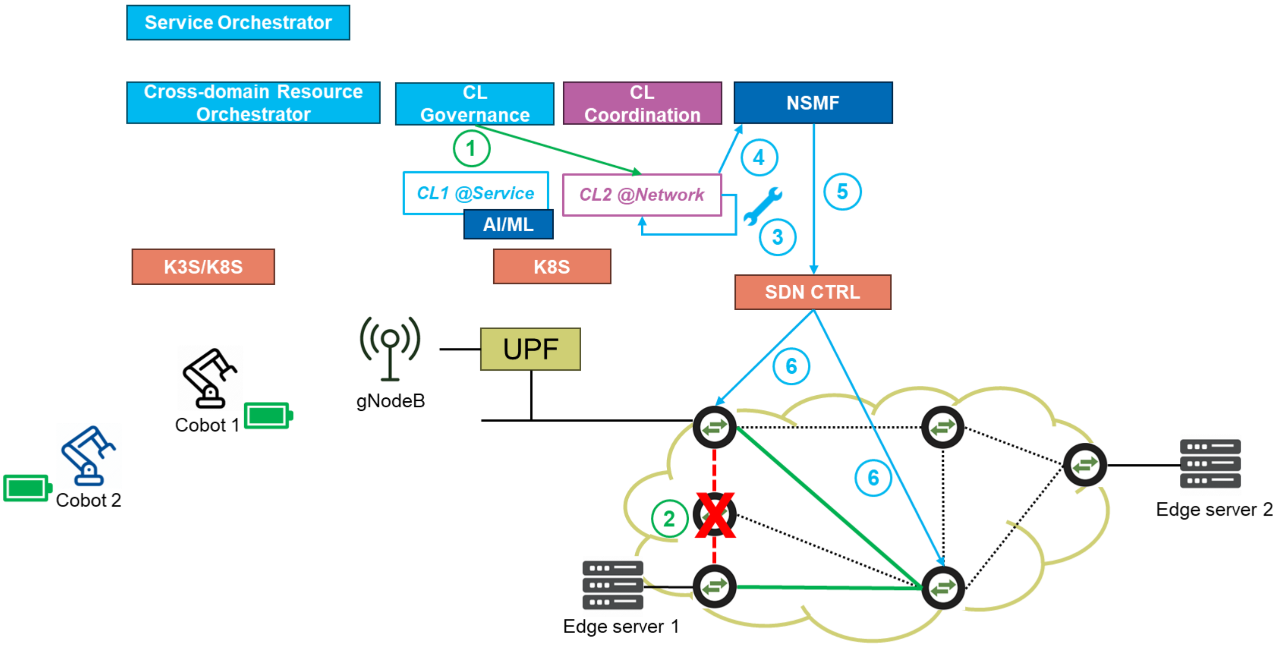


Figure 7-6: Workflow 3 part 1 – Zero-touch transport network management

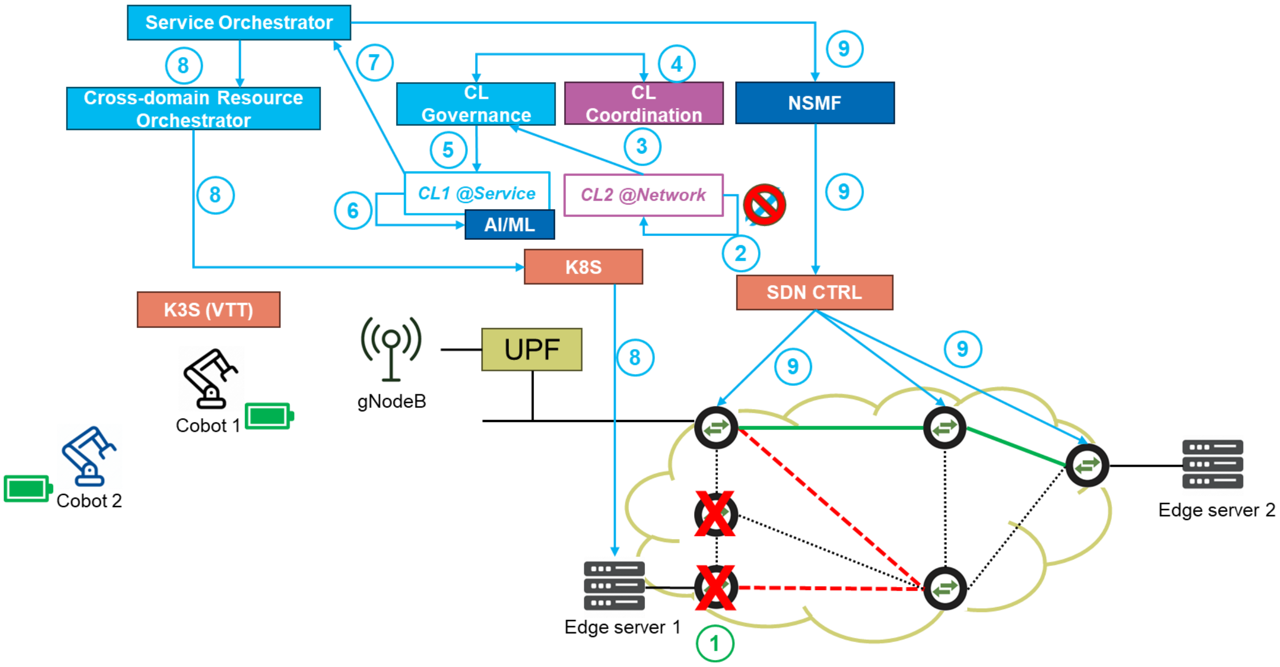


Figure 7-7: Workflow 3 part 2 – Closed-loop coordination for service continuity

Step-1. A second outage happens on one of the network nodes in a way that it is no longer possible to reach Edge Server.

Step-2-3. The CL2 detects the network outage that cannot be fixed: there is no path available towards the Edge Server 1. CL2 contacts the CL Governance to notify the impossibility to address the issue.

Step-4-6 CL Governance interact to CL Coordinator that escalates to CL1 to solve the problem.

Step-7-9. CL1 takes its decision and requests the Service Orchestrator to update the service. It requests the migration of the cobots' application to Edge Server 2 while requests the provisioning of a specific path towards the new server.

7.2.2.3 Intent-based services management solution

Among the objectives planned in system-PoC A, there is the use of an intent-based solution to manage the service (and network) resources across all the domains involved. The idea is to make use of intent-based requests to allow a non-technical user to manage the elements located in the different physical domains and let the system to manage any possible issue without the need for the service resources owner to participate.

To properly achieve the objective, there are two main actions to work on:

- The integration of the different intent-based management automation enablers previously described in subsection 4.2: The first set of intent-based enablers has been proposed in this document and an initial description of each one of them presented. In future actions, these enablers will have to keep evolving to define more details such as their possible architectures with the components and interfaces, the workflows illustrating the interactions of the architectural components and other aspects. Moreover, initial validation and experimental outcomes may be obtained from those that may be developed.
- The intent-based solution implementation and integration with the other system-PoC A elements: As earlier presented, the system-PoCs are organized in three main scenarios (i.e., versions) in which at each scenario adds and integrates new elements to complete the whole designed PoC. Among these elements, there are solutions based on enablers from different Hexa-X-II project WPs such as smart network management enablers and intent-based management automation enablers. The integration of the intent-based solution aims to follow a two-step procedure. First with a single intent-based solution managing a single-domain scenario, and once this option is properly validated and works as expected,

the next step is going to have an intent-based multi-domain scenario where two intent-based solutions interact between them to deploy and assure the expected intent requests from the users.

7.2.2.4 Integration fabric

System-PoC A aims to show the core role of the integration fabric in the E2E sight of the system. As shown in section 6.1.2.2.3, the integration fabric represents a communication bus between all the microservices of the whole system with the aim of unlocking an SBMA. To implement that communication layer two approaches were considered, i.e., service mesh and message broker approach. The first offer a more polyhedric solution, just by design. However, the versatility and the power of this approach is balanced by a quite complex deployment and a high maintenance and management cost. The message broker-based architecture is not originally oriented to be a fully interconnected layer to facilitate the communication within services, but rather a more general asynchronous communication bus. With a correct design it can be exploited to satisfy all the requirements that is needed by the integration fabric. The solution at the beginning may need a development effort that is bigger than what is previously mentioned, but in the end, it gains in versatility in deployment and management. In addition, it offers by design a much lower computational cost to be maintained. Looking to the needs of both smart network management enablers and intent-based management automation enablers, the decision has fallen on the choice of a message broker architecture. A general overview of the component is shown in Figure . The entire architecture will be built on top of open-source solutions. The chosen message broker framework (RabbitMQ, NATS, Pulsar or Kafka) will be exploited to create a more complex solution, enriched by other boundary framework, to achieve the wanted architectural design objectives.

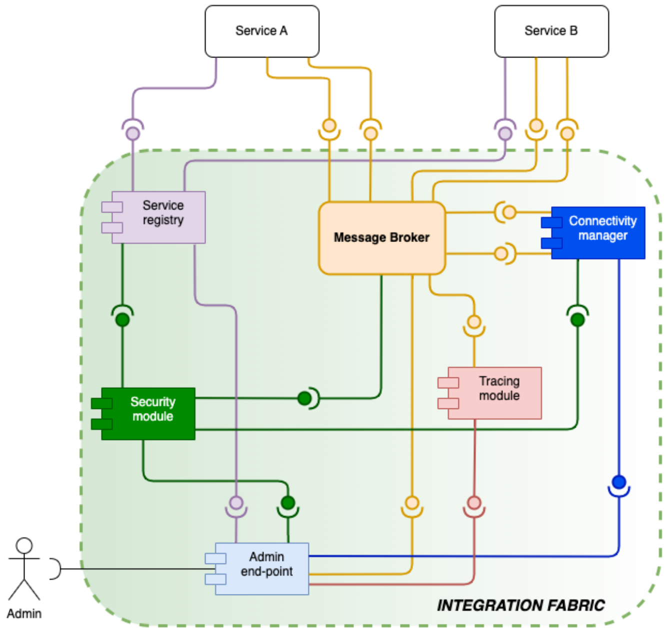


Figure 7-8: Integration fabric message broker architecture.

7.2.2.5 Network domain resources: Programmable and flexible network configuration

Within the System-PoC A, one of the resource domains is focused on transport network resources, to this end, the use of TeraFlow SDN is presented to become the solution towards the implementation of the enabler to achieve a programmable and flexible network configuration. TeraFlow SDN allows a set of multiple actions such as the configuration of connectivity services to compose Transport Network Slices. Based on this, further steps and actions towards the use of the TeraFlow SDN within the PoC, such as the discussion of the interfaces and other aspects, are under study.

7.3 System-PoC A evaluation results for environmentally and socially sustainable orchestration in 6G systems

7.3.1 KPIs & KVIs related to system-PoC A

Below are the targeted KVIs in system-PoC A and the KPIs that will be measured to ensure the reliability and performance of the system (see detailed list of KPIs below). Given that the system-PoC A deals with management and orchestration aspects, it is important to measure specific attributes that can show the impact of intents and decisions for the problem resolution. Such attributes can expand beyond common, usual ones (e.g., throughput, capacity) towards reliability, intent deployment time, recovery time etc. The set of KPIs and KVIs include generic indicators for assessing aspects related to environmental and social sustainability, as well as specific indicators that are associated with the assessment of the performance of orchestration mechanisms.

7.3.2 Sustainability aspects

The system-PoC A in Hexa-X-II mainly focuses on the feasibility to achieve the target KVIs of environment and social sustainability. Environmental sustainability aspects regard both the operation of the application per se, as well as the operation of the networking and computing infrastructure in a 6G ecosystem. The operations of the cobots with an energy efficient perspective will maximize the lifetime of the operations which will contribute to the environmental sustainability of the system. The improved human-machine interaction in the system-PoC A with the intelligent cooperation among cobots via management and orchestration of the 6G continuum, and resource-usage efficiency will increase the overall environmental sustainability. This will impact the overall operations in the manufacturing process by maximizing the lifetime of the operations in a resource limited environment. In parallel, enforcement of energy-efficient orchestration policies in the deployment and lifecycle management of the application can reduce the overall energy consumption across the infrastructure and help to achieve environmental sustainability targets. Moreover, the trustworthiness (as part of the social sustainability key values) refers to the reliability, security and overall integrity of the network and its provided services. Handover of the operations in a general failure, as well as the injection of distributed intelligence and automation characteristics will assure the resilience of the system and thereby enhance the trustworthiness.

7.3.3 Technical KPIs

A set of technical KPIs are related to orchestration aspects with regards to the PoC A. These KPIs aim to assess the performance of the developed orchestration mechanisms and platforms. An extended list of the KPIs detailed in the D2.1 [HEX223-D21] is provided, as follows.

- **Reliability**: Network reliability refers to the percentage of time that the network is available and functioning correctly. High network reliability is required to minimize disruptions to service and prevent lost revenue. Application/Service reliability refers to the percentage of time that the application or the service is fully functional and responsive to the posed requests/workload in accordance with the defined SLA.
- **Latency**: Network/Link latency measures the time it takes for data to travel between two points in the network. It is typically expressed in milliseconds (ms) or microseconds (μs). In the context of 6G, low latency is crucial for applications that require real-time interactions, such as virtual reality, telemedicine, and autonomous vehicles. Software latency (distributed traces) measures the time required for the interaction between software components. Having information for both type of latencies is helpful to identify whether a misperformance or a bottleneck is due to network or application performance aspects.
- **Provisioning Time**: this KPI measures the time taken to enforce a provisioning request of a managed entity (CNF, Network Service, Network Slice, Application etc.) to the underlying infrastructure, measured from when the request reaches the provisioning interface up to the time that the managed entity provisioning is fulfilled.

- **Termination Time**: Time to terminate a managed entity (CNF, Network Service, Network Slice, Application etc.), from the termination request up to the release of its assigned resources. This provides a measure of the promptness in re-availability of the resources after released by a service.
- **Recovery Time**: this KPI measures the time to recovery of a managed entity (CNF, Network Service, Network Slice, Application etc.) after an outage, providing a measure of the reactiveness of the network in minimizing service downtime.
- **Intent Deployment Latency**: Time to have the complete E2E intent-based service request properly deployed and available to be used by for the final user. This latency will start when the service user intent-based requests reach the IBM system solution until the confirmation of the intent-based service is available for the user.
- **Intent Conflict Resolution Latency**: Time to achieve the complete resolution (i.e., service completely working in normal status) since an intent-based conflict is detected, up until it is solved.
- **Scaling Time**: Time to apply horizontal or vertical scaling actions. It is measured from the time that a scaling request is triggered till the time the new or updated instances of a service or application component are operational.
- **APIs Performance**: a set of indicators can be considered for measuring the performance of the provided APIs (e.g., the Northbound APIs provided by the DSP). These indicators include the average and maximum latency for serving a request, requests served per minute, errors per minute, number of concurrent tenants.

Complementary to the above, a preliminary list of KPIs/KVIs specific to above-presented applications follow:

- **Power Consumption per AMR/UAV:** This KPI measures the power consumption of the robot for a pre-defined set of configured roles/actions, per unit of time.
- **Overall System Power Consumption:** The power consumption measured for all involved system components, for the E2E service execution, per unit of time.
- **Overall System Trust:** A set of indicators for measuring the trust of all the entities/nodes, allocated in the E2E service, including end-devices (e.g., cobots), as well edge/cloud compute nodes.
- **Path Planning Efficiency:** The time required to calculate optimal paths per robot.
- **Object Detection Accuracy/Performance:** Performance metrics related to the object detection service, such as accuracy, precision, recall.
- **Warehouse Digital Twin positioning accuracy:** The metric which assesses the error introduced in the digital representation of the physical objects, in real-time comparing to the actual one.
- **High-Resolution, Real-Time Cobot Camera Feed Latency:** The amount of time that it takes for a single frame of video to transfer from the robot’s camera to the Digital Twin’s display.
- **Cobot Tele-Operation Command Latency:** The time required for teleoperation command packets, from the tele-operation user interface (edge/cloud server) to reach the cobot tele-operation service.
- Business KPIs:
  - Warehouse task operational time.
  - Number of accomplished tasks per time window.
  - Percentage of workload covered with the use of the AMRs/UAVs.

7.4 Preliminary technical results

This subsection presents some preliminary results on the current architecture of System-PoC A, in relation to the Power Consumption, Provisioning Time and Recovery Time KPIs, listed above.

As it is illustrated in Figure 7-9, the component called *Functionality Allocation* (FA) has been developed. This component calculates the close to optimal placement of functionality, compute workloads/tasks, services to the various available compute nodes, robotic units, edge, and cloud servers towards energy efficiency (part of the sustainable AI/ML based-control enabler [HEX223-D62]). The input to FA mechanism is the computation and functional requirements of the compute workloads, the capabilities of the available compute nodes, and data producers (edge devices). The output of this mechanism is handled by the orchestrator which enforces the proposed (re)allocations to the system. The mechanism is triggered by the monitoring system when a need of reallocation emerges (e.g., detected increased latency, power consumption, malfunction in a robot’s part).

The development of FA mechanism is in progress. The current version leverages a metaheuristic algorithm, which can demonstrate high scalability for large scale experimentations. Figure 7-9 shows the gains in power consumption with the use of the proposed FA mechanisms compared to two baseline algorithms, namely a Round-Robin Placement (RRP) and a Feasible Random Placement (FRP) algorithm. In this graph the two lines represent the reduction of power consumption obtained by the FA mechanism compared to the FRP algorithm (red line) and compared to the RRP algorithm (green line). The validation scenario comprised 7 compute nodes (3 robotic units, 2 edge servers and 2 cloud servers) and an increasing number of compute workloads/tasks (4-28 workloads). The proposed algorithm provides higher power consumption gains when there is higher number of workloads, varying from 8.8-28.6%.

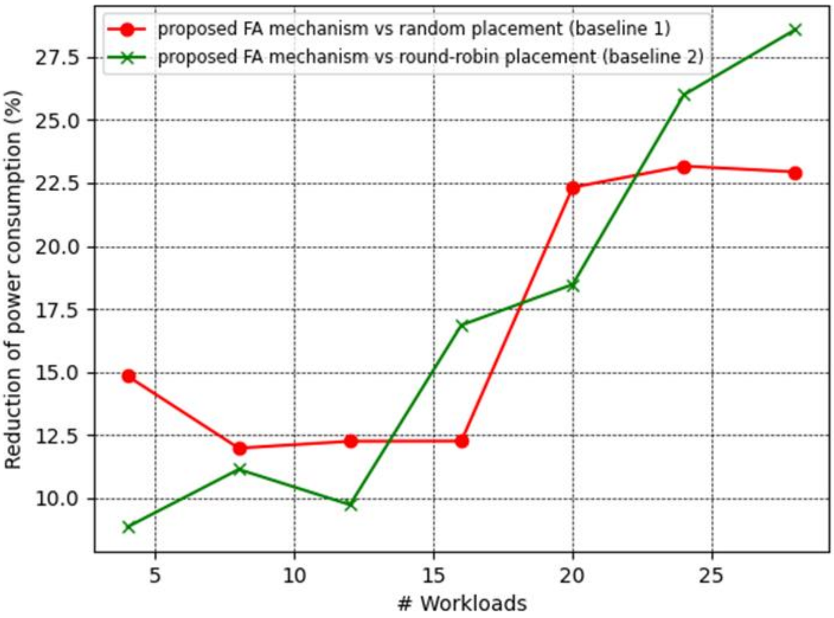


Figure 7-9: Reduction of power consumption with increasing number of compute workloads of the proposed FA mechanism compared with two baseline algorithms.

Additionally, Figure 7-10 and Figure 7-11, present the enhancements that functionality allocation optimization mechanism and performance diagnosis bring to management and orchestration (M&O) operations. We compare the typical M&O workflow (notification, action) as baseline, with the optimized M&O workflow which uses performance diagnosis and the functionality allocation mechanism. Both workflows are used to handle four types of events/operation descriptions that can happen in the industrial context of the PoC. For each of these types, ten instances of events are triggered manually, following the typical patterns of the industrial automation service and the average of them is presented in Figure 7-10 and Figure 7-11. The four events are:

a. The redeployment of functionalities to existing resources caused by robot malfunction.

b. The scaling of functionalities to new resources caused by increasing load or low battery.

c. The deployment of functionalities to new resources caused by robot malfunction.

d. The redeployment of functionalities to the maximum number of resources caused by significant load increase.

For each of these events we measured the following time periods:

- The notification time which is the time the monitoring system needs to check for the status of the node/component.
- The detection time which is the time the monitoring system needs to detect that there is an issue including possible timeouts, retries, etc.

- The reaction time which is the time the corrective actions take to be triggered on the respective component.
- The operations time (provisioning time) which is the time it takes for the corrective actions to be completed, e.g., functionality reallocation, scaling by commissioning resources etc.
- The application time which is the time the service needs to be restored (mainly due to service initialization or management operations in case it became unavailable).
- The recovery time which is the time of the appearance of the event, till the service is available again.

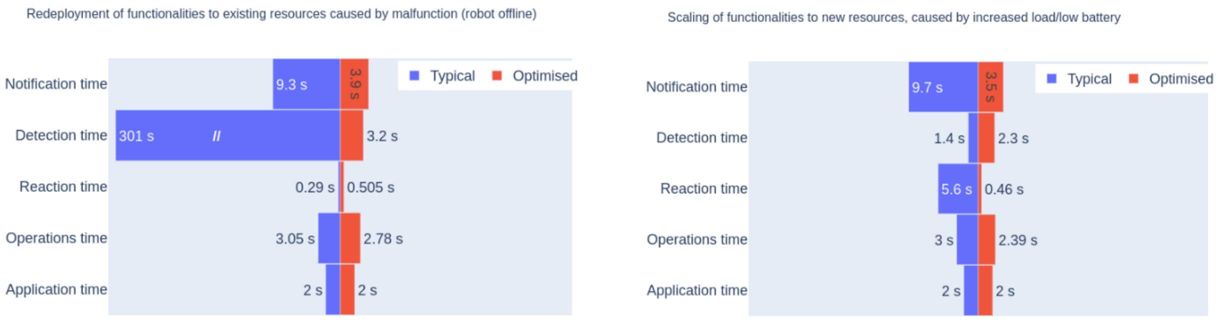


Figure 7-10 Collected time measurements during unexpected event a: redeployment of functionalities to existing resources caused by robot malfunction -displayed on the left-hand side graph-, and during unexpected event b: scaling of functionalities to new resources caused by increasing load or low battery -displayed on the right-hand side graph

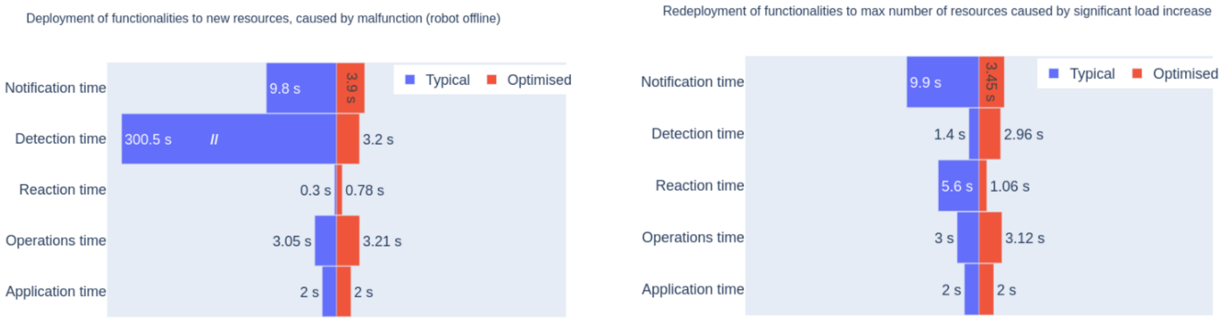


Figure 7-11 Collected time measurements during unexpected event c: deployment of functionalities to new resources caused by robot malfunction -displayed on the left-hand side graph-, and during unexpected event d: redeployment of functionalities to the maximum number of resources caused by significant load increase -displayed on the right-hand side graph

When the robot goes offline due to malfunction, as in events (a) and (c), the Kubernetes workflow starts to detect, wait for a response, and finally move the unavailable pods from that node to another one. This workflow uses a default timeout of 300 seconds to avoid moving pods unnecessarily due to short network failures. Also, for events (a) and (c), the first M&O workflow assumes that there are enough nodes and compute resources in the cluster since it does not have automated commissioning. The optimised M&O workflow, on the other hand, can dynamically commission resources from the available ones using the intelligent orchestration functionalities to optimize energy efficiency and load distribution.

Results in Figure 7-10 and Figure 7-11 show that the optimized M&O workflow, having performance diagnosis and the functionality allocation mechanism, has the best recovery time. If we compare the recovery time of the typical M&O, which is 21.7-315.65 s, with the optimized M&O recovery time (10.65-13.09 s), we can say that we have important gains. Finally, the operations time, which is related to the provisioning time KPI, ranges from 2.39 to 3.21s in both workflows. The measurement of the optimized M&O workflow is slightly lower in the (a) and (b) events and it is slightly higher in (c) and (d) events.

More evaluations will follow with these components when the trust manager component is integrated as well, which will ensure the maximum trustworthiness of the system and will be reported to the upcoming deliverable.

7.5 Further developments on E2E system evaluation from other approaches

The use of virtual simulation / emulation-based modeling, and digital twinning is envisaged to increase remarkably in the 6G solutions design and performance evaluation. That will be enabled by the enhanced computation power of computers' central and graphical processing units, as well as sophisticated software development. Virtual solutions have a huge potential to improve the design processes' cost-efficiency in comparison to traditional approach that require measurements to be performed in the field or in lab environments. Virtual solutions enable also testing of the new solutions before manufacturing the physical prototypes, leading to decreased cost and improved resource-efficiency, which supports sustainable development process. Therefore, a simulation / digital twin-based approach is used also in this project to enable evaluation of selected KPIs of E2E communication system, as well as to get insights about the virtual modelling requirements of the 6G system.

Deliverable D2.1 [HEX223-D21] shortly introduced the simulation and digital twin-based performance evaluation approaches which are planned to be performed during the project. Here an updated plan will be given, which gives more insights to the virtual E2E performance evaluation to be done in the later stage of the project, once the technical enablers are more matured.

7.5.1 E2E simulation framework for connectivity

As was introduced in D2.1 [HEX223-D21], an E2E simulation framework is planned to be developed and used for selected 6G connectivity enablers performance evaluation. As a starting point a 5G simulation platform which is based on disaggregated radio access network (RAN) composed of central unit (CU), distributed unit (DU) and radio unit (RU) will be used. The high-level architecture of the E2E simulation platform its mapping to the 6G system blueprint, is shown in Figure 7-12. As the Figure 7-12 illustrates, the E2E simulation framework includes partial functionalities of each layer of the system blueprint. Application layer includes service application, which creates traffic to be communicated between the core network (CN) and user equipment (UE). Network-centric application layer includes the RAN intelligent control (RIC) application, which is here assumed to be based on the O-RAN approach, i.e., RIC, xApps and E2 interface. However, similar mapping could be done in general for the CPU based application which is controlling the RAN intelligence. Network functions layer includes CN, CU, DU, RU and UE. Resource layer includes here additional computing resources, which can be used, e.g., for computing tasks needed for AI/ML -based RIC applications learning data preparation, or for accurate 3D modelling of the environment and corresponding channel models.

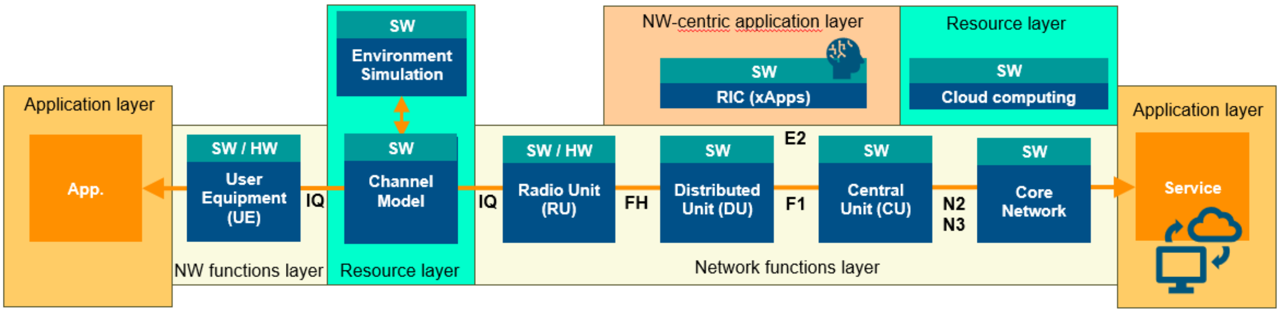


Figure 7-12: High-level architecture of E2E simulation framework and its' mapping to the system blueprint.

Introduced E2E simulation framework components and interfaces will be developed further to enable performance evaluation of the selected 6G technical enablers as a part of the E2E network. As the technical enablers are just being developed and individually evaluated, a detailed selection of the enablers to be evaluated in the virtual E2E system, has not yet been done. However, at this point we have identified certain technical enablers, e.g., D-MIMO, beamforming, RIS and JCAS, which may be useful and feasible to be evaluated by using E2E simulation framework.

To be able to simulate selected technical enablers as part of the introduced virtual E2E system, implementation of the technical enabler features / algorithms to the RU / UE, as well as enabling the control of the intelligent radio(s) by using RIC application(s), is required. In addition, for example in case of the D-MIMO solutions, there will be a need to enable simulation of multiple RUs under single Du, and / or multiple DUs under single CU. In the next phases of the project, the technical enablers, and their performance evaluation needs as a part of E2E system, will be analysed in detail. Then the performance evaluation simulations as a part of the E2E system will be performed, to gain a more comprehensive insight on their performance as well as on the integration requirements and challenges.

8 Conclusions and next steps

WP2 is the main technical hub of Hexa-X-II project and drives its work towards designing a system blueprint aiming at a sustainable, inclusive, and trustworthy 6G platform, and to provide the E2E system validation. Accordingly, WP2 provides the overall design of the 6G E2E system blueprint and harmonizes the E2E design principles considering 6G use case requirements as well as the KPIs and KVIs. Moreover, WP2 aims to provide overall design of the radio interface and protocols of the 6G platform, design an intent based E2E service management automation framework and develop a validation framework focused on security, privacy, and resiliency issues. Finally, WP2 is developing E2E level system PoCs to evaluate, at the E2E level to validate if the system can reach the 6G targeted KPIs/KVIs.

This deliverable has summarized the work carried out in WP2 since the completion of the first D2.1. In pursuit of the objectives outlined in Hexa-X-II project as well as those established in WP2, the efforts have been undertaken through the collective contribution of five tasks within WP2, in collaboration with other technical work packages. The report has provided a preliminary set of system requirements identified for 6G E2E system with respect to use cases as well as the operational aspects. In the next deliverable, considering the inputs coming from other technical WPs and considering the use case and operational requirements, it is expected refine those requirements of the 6G E2E system.

In this deliverable, enablers related to radio interface and protocols are discussed, with ambition for 6G radio interface and protocols considering lessons learned in 5G and support of new expanded scope and capabilities. Some selected topics within the RAN user plane and the RAN control plane have been analysed in depth, including data recovery, cyphering and integrity, DL control and mobility procedures. An initial analysis of the impact on the radio interface and protocols from a subset of enablers developed in other WPs has been performed based on an initial analysis of their anticipated functionalities. Those enablers include energy efficient radio design, sensing and positioning, compute offloading, AI, subnetworks, distributed MIMO, RIS, and energy neutral device.

This deliverable has defined the E2E intent-based service management automation architecture, defining the relationship between the customers and the DSP exposing services from its own and from the third parties based on an intent-based API. Several different actors have been identified, defining the management of E2E services in a multi-DSP scenario A preliminary functional architecture for the intent-based digital service manager of the DSP has also been proposed. Additionally, the first set of enablers that will contribute to the E2E system M&O, and an initial analysis of the proposed architecture and enablers from other work packages have also been reported.

Enablers for enhancing E2E security, privacy and resilience have been identified and classified, according to the characteristics of the *6G delta* (the direct implications of 6G technology evolution) and the new threats and enhancements it implies. To collect additional evidence during the development of the project, specific methods have been identified and described, focused on the use of synthetic environments, via simulation, emulation or a combination of both. The project team is currently analysing and prioritizing the execution of validation experiments, according to their interest and feasibility.

As a continuation of the work reported in D2.1 [HEX223-D21], this report (D2.2) provides the updates to the 6G E2E system blueprint considering the use case requirements and the enabler integration from other technical WPs. Key criteria under consideration for the analysis of the enablers for their integration in the E2E system have been defined by providing a framework to be used by the technical WPs of Hexa-X-II for the further development of their enablers. It constitutes a first checklist of what can be considered in technical components/enablers for the alignment with the E2E performance and operation targets and to provide recommendations for on-going development of enablers. The E2E alignment process serves to continually update the 6G system blueprint as well as the component design, as enablers and components become mature within the project. As the blueprint provide the holistic view of the system, it is complemented by different views whose purpose is to provide more specific details for the pervasive functionalities and the different layers of the blueprint. In this document, one started to dive into the view on the M&O functionality and the interaction with the enablers as well as with the various 6G stakeholders with more refinements to come in the coming months.

Lastly, the deliverable presents the first system-PoC (i.e., system-PoC A), which is currently being developed focusing on smart network management aspects for demonstrating management mechanisms. A detailed and stable design for system-PoC A has been presented along with the preliminary results and a plan for future developments for system-PoC B and system-PoC C. During the coming months more evaluation results of system-PoC A are expected and the integration of flexible topologies in order to progress with system-PoC B.

The work presented in this deliverable set up the framework for the further iterations of the system design in a top-down approach and for the refinement of the enablers under development in the project in a bottom-up approach. In the next deliverable D2.3 [HEX224-D23], technical enablers and components developed from other SNS Stream B projects, which are relevant to E2E system, will be also considered in addition to those developed in Hexa-X-II. The refined description of enablers related to radio interface and protocols, intent-based service management automation, and for enhancing E2E security, privacy and resilience that will be considered with other selected components from other Hexa-X-II work packages for the second iteration of the System-PoC (B) will be reported in deliverables D2.3 and D2.4 and will be published in Hexa-X-II webpage [HEXA2].

References

[23.501] 3GPP TS23.501 “System architecture for the 5G System (5GS)”, version 16.6.0 Release 16, October 2020. [28.312] 3GPP, TS 28.312 “Intent driven management services for mobile networks” v18.1.1, Sept. 2023.

[28.530] 3GPP TS 28.530, “Management and orchestration; Concepts, use cases and requirements”, V17.4.0, March 2023.

[28.804] 3GPP TR 28.804, “Telecommunication management; Study on tenancy concept in 5G networks and network slicing management”, v16.0.4, October 2019.

[28.824] 3GPP TR 28.824, “Study on network slice management capability exposure”, v18.0.0, June 2023.

[33.501] 3GPP TS 33.501, “Security architecture and procedures for 5G System”, v17.11.0, September 2023.

[36.300] 3GPP, TS 36.300, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2,” v17.5.0, July 2023

[38.323] 3GPP TS 38 323, “NR; Packet Data Convergence Protocol (PDCP) specification”, v17.5.0, June 2023

[38.305] 3GPP TS 38 305, “5G; NG Radio Access Network (NG-RAN); Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN,” v17.6.0, Sept. 2021.

[38.331] 3GPP, TS 38.331, “NR; Radio Resource Control (RRC); Protocol specification,” v17.6.0, September 2023.

[5GEx17-D22] 5GEx project deliverable D2.2: “5GEx Final System Requirements and Architecture”, Dec. 2017.

[5GPa19-D43] 5G!Pagoda project deliverable D4.3: “End-to-end Network Slice”, Jan. 2019.

[5Gr19-D21] 5Growth project deliverable D2.1: “Initial design of 5G End-to-End Service Platform”, 2019.

[5GTr19-D24] 5G-Transformer project deliverable D2.4: “Final design and implementation report on the MTP”, Jan. 2019

[5GV19-D31] 5G-VINNI project deliverable D3.1: “Specification of services delivered by each of the 5G-VINNI facilities”, June 2019.

[AB20] K. Antevski and C. J. Bernardos, “Federation of 5G services using distributed ledger technologies”, Internet Technology Letters, 3(6), e193, 2020.

[AB22] K. Antevski and C. J. Bernardos, "Federation in Dynamic Environments: Can Blockchain Be the Solution?,"IEEE Communications Magazine, vol. 60, no. 2, pp. 32-38, February 2022, doi: 10.1109/MCOM.001.2100585.

[ABB+20] P. Ala-Pietilä, Y. Bonnet, U. Bergmann, M. Bielikova, C. Bonefeld-Dahl, W. Bauer, and A. Van Wynsberghe, “The assessment list for trustworthy artificial intelligence (ALTAI),” European Commission 2020.

[Abr22] J. Abraham, “The BSS building blocks of the future,” Analysis Mason, 2022 [Online]. Available: https://d1a5bopfc3yb9e.cloudfront.net/reports/Analysys-Mason-The-BSS-Building-Blocks.pdf?Expires=1686764408&Signature=l6RbiNxfNo7FQIJv-zGQU9AXFt6aPWjZMT~zcebo15~cYAkRFD1aeYy7~62PodMDH0v0SVyxnbSfoj0S3ryxI8qZAw2AKOQtokJg6YFLQDwyQGmyCSDDiPeI9kOv0S8znl5~9OAT

[ACME] IETF, “Automated Certificate Management Environment”. https://datatracker.ietf.org/wg/acme/about/

[ALM+23] V. C. Andrei, X. Li, U. J. Mönich and H. Boche, "Sensing-Assisted Receivers for Resilient-By-Design 6G MU-MIMO Uplink," 2023 IEEE 3rd International Symposium on Joint Communications & Sensing (JC&S), Seefeld, Austria, pp. 1-6. 2023. doi: 10.1109/JCS57290.2023.10107512.

[ALP+20] A. Aguado, D. R. Lopez, A. Pastor, V. Lopez, J. P. Brito, M. Peev, A. Poppe, and V. Martin, “Quantum cryptography networks in support of path verification in service function chains,” J. Opt. Commun. Netw., vol. 12, no. 4, pp. B9–B19, Apr. 2020. DOI: 10.1364/JOCN.379799 [aLTEr] D. Rupprecht, K. Kohls, T. Holz and C. Pöpper, "Breaking LTE on Layer Two," 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2019, pp. 1121-1136, doi: 10.1109/SP.2019.00006.

[AMD20] AMD, "AMD SEV-SNP: Strengthening VM Isolation with Integrity Protection and More", White Paper, January 2020, https://www.amd.com/system/files/TechDocs/SEV-SNP-strengthening-vm-isolation-with-integrity-protection-and-more.pdf.

[AMG+20] K. Antevski, J. Martín-Pérez, A. Garcia-Saavedra, C. J. Bernardos, X. Li, J. Baranda, J. Mangues-Bafalluy, R. Martnez, and L. Vettori, "A Q-learning strategy for federation of 5G services," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020, pp. 1-6.

[AMM+23] P. Alemany, R. Muñoz, J. Martrat, A. Pastor, R. Díaz, D. Lopez, R. Casellas, R. Martínez, and R. Vilalta, "A Blockchain-based Trust Management Collaborative System for Transport Multi-stakeholder Scenarios", submitted in the Journal of Optical Communications and Networking, June 2023.

[ARM23] ARM, "Introducing ARM Confidential Compute Architecture", Version 2.0, Release Information, 2023, https://developer.arm.com/documentation/den0125/0200.

[ARR+22] J. Ahmed, M. A. Razzaque, M. M. Rahman, S. A. Alqahtani and M. M. Hassan, "A Stackelberg Game-Based Dynamic Resource Allocation in Edge Federated 5G Network," in IEEE Access, vol. 10, pp. 10460-10471, 2022.

[ATF+22] A. Abouaomar, A. Taik, A. Filali, and S. Cherkaoui, "Federated deep reinforcement learning for open ran slicing in 6g networks," IEEE Comm. Mag. Vol. 61, no. 2, pp. 126-132, 2022.

[BAC+18] M. Brundage, S. Avin, J. Clark, et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation", Oxford, 2018.

[BBM+21] F. Brockners, S. Bhandari, T. Mizrahi, S. Dara, and S. Youell, "Proof of Transit," Internet Engineering Task Force, Internet-Draft https://datatracker. ietf.org/doc/draft-ietf-sfc-proof-of-transit/ 2021.

[BCE+03] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallance, and D. von Winterfeldt, et al. "A framework to quantitatively assess and enhance the seismic resilience of communities," Earthquake spectra, vol. 19, no. 4, pp. 733-752, 2003.

[BHM+19] R. Barnes, J. Hoffman-Andrews, D. McCarney, J. Kasten, "Automatic Certificate Management Environment (ACME)". March 2019. https://www.rfc-editor.org/rfc/pdfrfc/rfc8555.txt.pdf

[BJZ+22] A. C. Baktir, A. D. N. Junior, A. Zahemszky, A. Likhyani, D. A. Roeland, D. Biyar, E. D. Ustok, R. F. Orlic, and M. D. Angelo, "Intent-based cognitive closed-loop management with built-in conflict handling," 2022 IEEE 8th International Conference on Network Softwarization (NetSoft), Milan, Italy, pp. 73-78, 2022. doi: 10.1109/NetSoft54395.2022.9844074.

[BLS+22] B. Coll-Perales, M. C. Lucas-Estañ, T. Shimizu, J. Gozalvez, T. Higuchi, S. Avedisov, O. Altintas, and M. Spulcre, "Improving the Latency of 5G V2N2V Communications in Multi-MNO Scenarios using MEC Federation," 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 2022, pp. 1-5.

[BMB21] B. Bakhshi, J. Mangues-Bafalluy and J. Baranda, "R-Learning-Based Admission Control for Service Federation in Multi-domain 5G Networks," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6.

[BMC21] A. Banerjee, S. S. Mwanje and G. Carle, "Optimal configuration determination in Cognitive Autonomous Networks," 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), Bordeaux, France, pp. 494-500, 2021.

[BMC21a] A. Banerjee, S. S. Mwanje and G. Carle, "An Intent-Driven Orchestration of Cognitive Autonomous Networks for RAN management," 2021 17th International Conference on Network and Service Management (CNSM), Izmir, Turkey, pp. 380-384, 2021. doi: 10.23919/CNSM52442.2021.9615505.

[BNS+23] I. Bizon, A. Nimr, P. Schulz, M. Chafii and G. Fettweis, "Blind Transmitter Localization Using Deep Learning: A Scalability Study," in Proceedings of 2023 Wireless Communications and Networking Conference (WCNC 2023), Glasgow, United Kingdom (Great Britain), Mar 2023.

[BSB+23] B. Briscoe, K. De Schepper, M. Bagnulo, G. White, "Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture", January 2023. https://www.rfc-editor.org/rfc/rfc9330.html

[BSW+23] M. Bartock, M. Souppaya, J. Wheeler, T. Knoll, M. Ramalingam, S. Righi, "Hardware Enabled Security: Hardware-Based Confidential Computing", NIST Interagency Report 8320D ipd, 2023, https://doi.org/10.6028/NIST.IR.8320D.ipd.

[CCC22] The Confidential Computing Consortium, "A Technical Analysis of Confidential Computing", v1.3, 2022, https://confidentialcomputing.io/wp-content/uploads/sites/10/2023/03/CCC-A-Technical-Analysis-of-Confidential-Computing-v1.3_unlocked.pdf

[CCG+21] A. Clemm, L. Ciavaglia, L. Granville, and J. Tantsura, "Intent-based networking-concepts and definitions," 2021. [Online]. Available: https://tools.ietf.org/html/draft-irtf-nmrgibn-concepts-definitions-02,

[CCG+22] A. Clemm, L. Ciavaglia, L.Z. Granville, and J. Tantsura, "Intent-Based networking – Concepts and Definitions", RFC 9315, doi 10.17487/RFC9315, Oct 2022.

[CEPS21] Centre for European Policy Studies (CEPS), "Artificial Intelligence and Cybersecurity - Technology, Governance and Policy Challenges", Task Force Report, 2021. https://www.ceps.eu/ceps-publications/artificial-intelligence-and-cybersecurity-2/

[CF13] D. Catalano, and D. Fiore, "Vector Commitments and Their Applications". Lecture Notes in Computer Science, vol 7778. Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-36362-7_5

[COC23] Confidential Containers, "Architecture", 2023, https://github.com/confidential-containers/confidential-containers/blob/main/architecture.md.

[EDG23] Edgeless Systems, "Welcome to Marblerun", https://docs.edgeless.systems/marblerun/, 2023.

[ENISA21] ENISA, "Threat Landscape for Supply Chain Attacks", https://www.enisa.europa.eu/publications/threat-landscape-for-supply-chain-attacks

[Eri21] Ericsson, "Planning in-building coverage for 5G: from rules of thumb to statistics and AI," Ericsson Mobility Report, June 2021

[Eri23] Ericsson Blog, "What to expect from 6G: Here are nine important takeaways from early global research," 2023, [Online]. Available: https://www.ericsson.com/en/blog/2023/2/6g-early-research-global-takeaways

[ETSI_PDL_011] ETSI, "ETSI GS PDL 011; Permissioned Distributed Ledger (PDL); Specification of Requirements for Smart Contracts' Architecture and Security", V1.1.1, December 2021

[ETSI_PDL_015] ETSI, "ETSI GS PDL 015; Permissioned Distributed Ledger (PDL); Reputation management", V 1.1.1, January 2023

[ETSI_PDL_019] ETSI, "ETSI GR PDL 019; PDL Services for Decentralized Identity and Trust Management", V1.1.1, May 2023

[ETSIQSC] ETSI, "Quantum-Safe Cryptography". https://www.etsi.org/technologies/quantum-safe-cryptography

[FAN22] S. Fan,W. Ni, H. Tian, Z. Huang, R. Zeng, "Carrier Phase-Based Synchronization and High-Accuracy Positioning in 5G New Radio Cellular Networks," IEEE Transactions on Communications, vol. 70, Issue 1, Jan. 2022

[FFG+20] D. Feldman, E. Fox, E. Gilman, I. Haken, F. Kautz, U. Khan, M. Lambrecht, B. Lum, A. M. Fayó, E. Nesterov, A. Vega, M. Wardrop, "Solving the Bottom Turtle – a SPIFFE Way to Establish Trust in Your Infrastructure via Universal Identity", ISBN: 978-0-578-77737-5, 2020.

[GBH+21] Gomes, P.H., Buhrgard, M., Harmatos, J., Mohalik, S.K., Roeland, D. and Niemöller, J. 2021. Intent-driven Closed Loops for Autonomous Networks. Journal of ICT Standardization. 9, 2 (Jun. 2021), 257–290. DOI:https://doi.org/10.13052/jicts2245-800X.929.

[GDPRA6] General Data Protection Regulation (GDPR) Article 6 "Lawfulness of processing". https://gdpr-info.eu/art-6-gdpr/

[GLC19] Z. Guan, J. Lin, and P. Chen, “On anomaly detection and root cause analysis of microservice systems,” In Proceedings of the Service-Oriented Computing (LNCS, Vol. 11434). Springer, Cham, pp. 465–469, 2019. DOI: https://doi.org/10.1007/978- 3- 030- 17642- 6_45

[GPW+20] X. Guo, X. Peng, H. Wang, W. Li, H. Jiang, D. Ding, T. Xie, and L. Su, “Graph-based trace analysis for microservice architecture understanding and problem diagnosis,” In Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, New York, pp. 1387–1397, 2020. DOI: https://doi.org/10.1145/3368089.3417066

[GSMAPQC] GSMA, “Post Quantum Telco Network Impact Assessment Whitepaper”, February 2023. https://www.gsma.com/newsroom/wp-content/uploads/PQ.1-Post-Quantum-Telco-Network-Impact-Assessment-Whitepaper-Version1.0.pdf

[GSMAQKD] GSMA, “Quantum Networking and Service”, De. 2021. https://www.gsma.com/newsroom/wp-content/uploads/IG-12-Quantum-Networking-and-Service.pdf

[GZH+19] Y. Gan, Y. Zhang, K. Hu, D. Cheng, Y. He, M. Pancholi, and C. Delimitrou, “Seer: Leveraging big data to navigate the complexity of performance debugging in cloud microservices,” In Proceedings of the 24th International Conference on Architectural Support for Programming Languages and Operating Systems. ACM, New York, pp. 19–33, 2019. DOI: https://doi.org/10.1145/3297858.3304004

[HEX21-D31] Hexa-X Deliverable D3.1 "Localisation and sensing use cases and gap analysis,” December 2021, available at https://hexa-x.eu/wp-content/uploads/2022/02/Hexa-X_D3.1_v1.4.pdf

[HEX223-D12] Hexa-X-II, “Deliverable D1.2: 6G use cases and requirements,” June 2023, available at https://hexa-x-ii.eu/

[HEX223-D21] Hexa-X-II, “Deliverable D2.1: Draft foundation for 6G system design,” June 2023, available at https://hexa-x-ii.eu/

[HEX223-D32] Hexa-X-II, “Deliverable D3.2: Initial Architectural Enablers,” October 2023, available at https://hexa-x-ii.eu/

[HEX223-D42] Hexa-X-II Deliverable D4.2, “Radio Design and Spectrum Access requirements and key enablers for 6G Evolution”, November 2023. [Online]. Available: https://hexa-x-ii.eu/wp-content/uploads/2023/11/Hexa-X-II_D4_2_final.pdf

[HEX223-D52] Hexa-X-II Deliverable D5.2, “Characteristics and classification of 6G device classes”, October 2023. To be uploaded at “https://hexa-x-ii.eu/”.

[HEX223-D62] Hexa-X-II Deliverable D6.2, “Foundations on 6G Smart Network Management Enablers”, October 2023. To be uploaded at “https://hexa-x-ii.eu/”.

[HEX224-D23] Hexa-X-II Deliverable D2.3, “Interim overall 6G system design”, To be published June 2024.

[HEX22-D62] Hexa-X deliverable D6.2: “Design of service management and orchestration functionalities”, April 2022.

[HEX23-D13] Hexa-X Deliverable D1.3 “Targets and requirements for 6G - initial E2E architecture,” Hexa-X project, 2022, [Online]. Available: https://hexa-x.eu/wpcontent/uploads/2022/03/Hexa-X D1.3.pdf.

[HEX23-D33] Hexa-X Deliverable D3.3 "Final models and measurements for localisation and sensing,” April 2023. [Online]. Available: https://hexa-x.eu/wp-content/uploads/2023/05/Hexa-X_D3.3_v1.4.pdf

[HEXA] Hexa-X Project [Online] https://hexa-x.eu/.

[HEXA2] Hexa-X-II Project [Online] https://hexa-x-ii.eu/

[HSW05] T. Hongal, J. Saperia, and S. Waldbusser, “Policy Based management MIB”, IETF RFC 4011, 2005. [Online]. Available: https://datatracker.ietf.org/doc/rfc4011/

[HZS+23] B. Han, Y. Zhu, A. Schmeink, and H. D. Schotten, “Non-orthogonal multiplexing in the FBL regime enhances physical layer security with deception,” to appear in the 24th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Shanghai, China, September 2023. [Online]. Available: https://arxiv.org/abs/2304.06402

[IETF23] IETF, “Charter for the Working Group on Supply Chain Integrity, Transparency, and Trust (scitt)”, https://datatracker.ietf.org/wg/scitt/about/

[IFA030] ETSI GS NFV-IFA 030: “Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Multiple Administrative Domain Aspect Interfaces Specification”, 2019.

[Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/030/03.03.01_60/gs_NFV-IFA030v030301p.pdf

[INF001] ETSI GS NFV-INF 001, Network Functions Virtualisation (NFV); Infrastructure Overview, V1.1.1, ETSI, Sophia Antipolis, Jan. 2015.

[INT23] Intel Security Libraries for Datacenter, “Platform Integrity Attestation”, 2023, https://intel-secl.github.io/docs/5.1/product-guides/Foundational%20%26%20Workload%20Security/60%20Platform%20Integrity%20Attestation/.

[IRTF23] IRTF, “Usable Formal Methods Proposed Research Group”, 2023. [Online]. Available: https://datatracker.ietf.org/rg/ufmrg/about/

[ISGQKD] ETSI, “Quantum Key Disitribution”. https://www.etsi.org/technologies/quantum-key-distribution

[ITM+21] R. Inam, A. Terra, A. Mujumdar, E. Fersman, and A. Vulgarakis, “Explainable AI – How humans can trust AI,” Ericsson White Paper, Apr. 2021, [Online]. Available: https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai

[ITU18] ITU Y.3101, “Requirements of the IMT-2020 network”, 2018. [Online]. Available: https://www.itu.int/rec/T-REC-Y.3101/en

[ITUR23] ITU-R, “WP5D Temporary Document 905, Draft new Recommendation ITU-R M.[IMT.FRAMEWORK FOR 2030 AND BEYOND] - Framework and overall objectives of the future development of IMT for 2030 and beyond, “ Jun. 2023, https://www.itu.int/md/R19-WP5D-230612-TD-0905/en

[JCY+17] T. Jia, P. Chen, L. Yang, Y. Li, F. Meng, and J. Xu, “An approach for anomaly diagnosis based on hybrid graph model with logs for distributed services,” In 2017 IEEE International Conference on Web Services. IEEE, New York, pp. 25–32, 2017. DOI: https://doi.org/10.1109/ICWS.2017.12

[JPR21] A.S. Jacobs, R.J. Pfitscher, and R.H. Ribeiro, “Hey, Lumi! Using Natural Language for Intent-Based Network Management”, 2021 USENIX Annual Technical Conference (USENIX ATC 21), pp. 625-639, ISBN 978-1-939133-23-6, Jul 2021.

[JYC+17] T. Jia, L. Yang, P. Chen, Y. Li, F. Meng, and J. Xu, “Logsed: Anomaly diagnosis through mining time-weighted control flow graph in logs,” In Proceedings of the IEEE 10th International Conference on Cloud Computing. IEEE, New York, 447–455, 2017. DOI: https://doi.org/10.1109/CLOUD.2017.64

[KBP23] S. Kukliński, J. M. Batalla and J. Pieczerak, "Dynamic and Multiprovider-based Resource Infrastructure in the NFV MANO Framework," NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium, Miami, FL, USA, pp. 1-4, 2023. doi: 10.1109/NOMS56928.2023.10154398.

[KCH+22] T. Kuang, H. Chen, L. Han, R. He, W. Wang, and G. Ding, ”Abnormal signal recognition with time-frequency spectrogram: a deep learning approach,” arXiv preprint arXiv:2205.15001, 2022.

[KJS20] G. S. Kasturi, A. Jain, and J. Singh, “Machine learning-based RF jamming classification techniques in wireless Ad Hoc networks,” In International Conference on Wireless Intelligent and Distributed Environment for Communication, pp. 99–111. Springer, 2020.

[KKS+23] A Krause, MD Khursheed, P Schulz, F Burmeister, G Fettweis: “Digital Twin of the Radio Environment: A Novel Approach for Anomaly Detection in Wireless Networks”, accepted for publication in IEEE Global Communications Conference Workshops (IEEE Globecom Workshops), Kuala Lumpur, 2023

[KKU+22] F. Klement, S. Katzenbeisser, V. Ulitzsch, J. Krämer, S. Stanczak, Z. Utkovski, I. Bjelakovic, and G. Wunder, “Open or not open: Are conventional radio access networks more secure and trustworthy than open-ran?” 2022. [Online]. Available: https://arxiv.org/abs/2204.12227

[KLM+22] H. Kokkonen, L. Lovén, N. H. Motlagh, J. Partala, A. Gonz'alez-Gil, E. Sola, I. Angulo, M. Liyanage, T. Leppanen, T. Nguyen, P. Kostakos, M. Bennis, S. Tarkoma, S. Dustdar, S. Pirttikangas, and J. Riekki, “Autonomy and Intelligence in the Computing Continuum: Challenges, Enablers, and Future Directions for Orchestration,” ArXiv, 2022. abs/2205.01423

[KSS13] M. Kim, R. Sumbaly, and S. Shah. 2013. Root cause detection in a service-oriented architecture. ACM SIGMETRICS Performance Evaluation Review 41, 1 (2013), 93–104. DOI: https://doi.org/10.1145/2494232.2465753

[KUB23] The Kubernetes Authors, Kubernetes Documentation, 2023, https://kubernetes.io/docs/home/

[LAMPS] IETF, "Limited Additional Mechanisms for PKIX and SMIME". https://datatracker.ietf.org/wg/lamps/about/

[LCZ18] J. Lin, P. Chen, and Z. Zheng, "Microscope: Pinpoint performance issues with causal graphs in micro-service environments," In Proceedings of the Service-Oriented Computing (LNCS, Vol. 11236). Springer, Cham, pp. 3–20, 2018. DOI: https://doi.org/10.1007/978- 3- 030-03596- 9_1

[LDB+22] R. Li, B. Decocq, A. Barros, Y. Fang, and Z. Zeng, "Petri net-based model for 5g and beyond networks resilience evaluation," In 2022 25th Conference on Innovation in Clouds, Internet and Networks (ICIN), pp. 131-135, 2022.

[LDB+23] R. Li, B. Decocq, A. Barros, Y. P. Fang, and A. Zeng, "Estimating 5G network service resilience against short timescale traffic variation," IEEE Transactions on Network and Service Management. Early Access, 2023.

[LG22] B. Lang, and J. Gong, "JR-TFViT: A Lightweight Efficient Radar Jamming Recognition Network Based on Global Representation of the Time–Frequency Domain," *Electronics* vol. *11*, pp. 2794, 2022. https://doi.org/10.3390/electronics11172794

[LGL+21] X.Li, C. Guimarães, G. Landi, J. Brenes, J. Mangues-Bafalluy, J. Baranda, D. Corujo, V. Cunha, J. Fonseca, J. Alegria, A. Z. Orive, J. Ordonez-Lucena, P. Iovanna, C. J. Bernardos, A. Mourad,and X. Costa-Pérez "Multi-Domain Solutions for the Deployment of Private 5G Networks," IEEE Access, vol. 9, pp. 106865-106884, 2021.

[Lin22] The Linux Foundation, "https://www.hyperledger.org/." [Online]. Accessed in February 3, 2022.

[LINDDUN] "A framework for privacy threat modelling". https://linddun.org

[LKD+23] X. Lin, L. Kundu, C. Dick, E. Obiodu, T. Mostak, and M. Flaxman, "6G Digital Twin Networks: From Theory to Practice," in IEEE Communications Magazine, doi: 10.1109/MCOM.001.2200830.

[LXR+21] B. Lum, H. Xia, P. R, S. S. Bangalore, "Enabling advanced key usage and management in encrypted container images", IBM, 2021.

[MEC21] ETSI Multi-Access Edge Computing, July 2021 [Online] Available: https://www.etsi.org/technologies/multi-access-edge-computing

[MEF19] Metro Ethernet Forum, "MEF 55; Lifecycle Service Orchestration (LSO): Reference Architecture and Framework", 2019.

[MES+01] B. Moore, E. Ellesson, J. Strassner, and A.Westerinen, "Policy Core Information Model", IETF RFC 3036, 2001. [Online]. Available: https://www.ietf.org/rfc/rfc3060.txt

[MFL19] R. Morales Ferre, A. de la Fuente, and E. S. Lohan, ”Jammer classification in GNSS bands via machine learning algorithms," Sensors, vol. 22, pp. 4841, 2019. https://doi.org/10.3390/s19224841

[Mic22] Microsoft, "Microsoft Threat Modeling Tool. STRIDE", 2022. [Online]. Available: https://learn.microsoft.com/en-us/azure/security/develop/threat-modeling-tool-threats#stride-model

[MJS+21] L. Meng, F. Ji, Y. Sun, and T. Wang," Detecting anomalies in microservices with execution trace comparison," Future Generation Computer Systems 116, pp. 291–301, 2021. DOI: https://doi.org/10.1016/j.future.2020.10.040

[MLN23] C. Marche, V. Loscri and M. Nitti, "A Channel Selection Model based on Trust Metrics for Wireless Communications," in IEEE Transactions on Network and Service Management, 2023. doi: 10.1109/TNSM.2023.3277578.

[MMP+18] L. Mariani, C. Monni, M. Pezzé, O. Riganelli, and R. Xin, ”Localizing faults in cloud systems," In Proceedings of the 2018 IEEE 11th International Conference on Software Testing, Verification and Validation. IEEE, New York, pp. 262–273, 2018. DOI: https://doi.org/10.1109/ICST.2018.00034

[Mos15] M. Mosca, "Cybersecurity in a quantum world: will we be ready?" Invited talk at NIST workshop on Cyber Security in a Post-Quantum World. (Gaithersburg, 2015).

https://csrc.nist.gov/csrc/media/events/workshop-on-cybersecurity-in-a-post-quantum-world/documents/presentations/session8-mosca-michele.pdf

[MWZ+13] H. Mi, H. Wang, Y. Zhou, M. Rung-Tsong Lyu, and H. Cai, "Toward fine-grained, unsupervised, scalable performance diagnosis for production cloud computing systems," IEEE Transactions on Parallel and Distributed Systems vol. 24, no. 6, pp. 1245–1255, 2013. DOI: https://doi.org/10.1109/TPDS.2013.21

[NCK19] S. Nedelkoski, J. Cardoso, and O. Kao, "Anomaly detection and classification using distributed tracing and deep learning," In Proceedings of the 2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE, New York, pp. 241–250, 2019. DOI: https://doi.org/10.1109/CCGRID.2019.00038

[NFV006] ETSI GS NFV 006, Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Architectural Framework Specification, v4.4.1, ETSI, Sophia Antipolis, 12/2022.

[NGA22] NextG Alliance "6G Technologies," Report, June 2022

[NIST22] NIST, "Engineering Trustworthy Secure Systems", November 2022. https://csrc.nist.gov/pubs/sp/800/160/v1/r1/final

[NMA+16] A. Nandi, A. Mandal, S. Atreja, G. B. Dasgupta, and S. Bhattacharya, "Anomaly detection using program control flow graph mining from execution logs," In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, 215–224, 2016. DOI: https://doi.org/10.1145/2939672.2939712

[OAS23] Oasis Open, "PKCS #11 Specification Version 3.1", https://docs.oasis-open.org/pkcs11/pkcs11-spec/v3.1/os/pkcs11-spec-v3.1-os.pdf, 2023.

[OCC23] Occlum team, "Occlum RA Flow in a real demo", https://github.com/occlum/occlum/tree/master/demos/remote_attestation/init_ra_flow, 2023.

[OCI19] Open Container Initiative, "Proposal to add Encrypted Layer Mediatype", 2019, https://github.com/opencontainers/image-spec/pull/775

[ONF_TR523] ONF TR-523 – Intent NBI – Definition and Principles, Oct. 2016

[OPG] GSMA OPG.04, "East-Westbound Interface APIs", v1.0, October 2022 [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2022/10/Operator-Platform-Group-4.0-v1.0-October22.pdf

[OTB21] J. Ordonez-Lucena, C. Tranoris and B. Nogales, "Automated Network Slice Scaling in Multi-site Environments: The ZSM PoC#2 report", 2021.

[OTL23] OpenTelemetry, 2023. https://opentelemetry.io

[OTR20] J. Ordonez-Lucena, C. Tranoris and J. Rodrigues, "Modeling Network Slice as a Service in a Multi-Vendor 5G Experimentation Ecosystem," in 2020 IEEE International Conference on Communications Workshops (ICC Workshops), 2020, pp. 1-6.

[Pet77] J. L. Peterson, "Petri nets," ACM Computing Surveys (CSUR), vol. 9, no. 3, pp. 223-252, 1977.

[PJK+14] M. Pesko, T. Javornik, A. Košir, M. Štular, and M. Mohorčič, "Radio environment maps: The survey of construction methods," KSII Transactions on Internet & Information Systems, vol. 8, no.11, 2014.

[PMC22] M. Park, Y. Ma and T. Choi, "OneVisionNFP - Design and Implementation of Agile Private/Public 5G/B5G Network Federation Platform," 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, Republic of, 2022, pp. 947-949,

[PMD+22] S. K. Perepu, J. P. Martins, R. S. S and K. Dey, "Intent-based multi-agent reinforcement learning for service assurance in cellular networks," GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022, pp. 2879-2884, doi: 10.1109/GLOBECOM48099.2022.10001426.

[PPR+23] P. Porambage, J. Pinola, Y. Rumesh, C. Tao, and J. Huusko, "Xcaret: Xai based green security architecture for resilient open radio access networks in 6g," in 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). IEEE, pp. 699–704, 2023.

[PQR23] PQREACT (Post Quantum Cryptography Framework for Energy Aware Contexts) project, 2023. https://cordis.europa.eu/project/id/101119547

[PQUIP] IETF, “Post-Quantum Use in Protocols”. https://datatracker.ietf.org/wg/pquip/about/

[QIRG] IRTF, “Quantum Internet Research Group”, https://datatracker.ietf.org/rg/qirg/about/

[QSNP] Quantum Secure Networks Partnership. https://qsnp.eu

[QUB23] QUBIP (Quantum-oriented Update to Browsers and Infrastructures for the PQ Transition) project, 2023. https://cordis.europa.eu/project/id/101119746

[RUO+23] C. P. Robinson, D. Uvaydov, S. D'Oro, and T. Melodia, ”Narrowband Interference Detection via Deep Learning,” arXiv preprint arXiv:2301.09607, 2023.

[SB22] J. Soldani and A. Brogi, “Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey,” ACM Comput. Surv., vol.55, no. 3, Article 59 pp. 1-39, 2022. https://doi.org/10.1145/3501297

[SKR+20] N. Slamnik-Kriještorac, H. Kremo, M. Ruffini and J. M. Marquez-Barja, "Sharing Distributed and Heterogeneous Resources toward End-to-End 5G Networks: A Comprehensive Survey and a Taxonomy," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1592-1628, Q3 2020.

[SLG+20] Y. Sheffer, D. Lopez, O. Gonzalez de Dios, A. Pastor Perales, T. Fossati, “Support for Short-Term, Automatically Renewed (STAR) Certificates in the Automated Certificate Management Environment (ACME)”, March 2020. https://www.rfc-editor.org/rfc/rfc8739.pdf

[SMP+14] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, ”Fundamental tradeoffs among reliability, latency and throughput in cellular networks,” In 2014 IEEE Globecom Workshops (GC Wkshps), pp. 1391-1396, 2014.

[SOF23] SOFAEnclave, “Attestation based Enclave Configuration Service”, https://github.com/SOFAEnclave/enclave-configuration-service, 2023.

[SOL011] ETSI GS NFV-SOL 011: “Network Functions Virtualisation (NFV) Release 3; Protocols and Data Models; RESTful protocols specification for the Or-Or Reference Point”, 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-SOL/001_099/011/03.03.01_60/gs_nfv-sol011v030301p.pdf

[SPL+21] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila "AI and 6G security: Opportunities and challenges." 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). IEEE, 2021

[STL23] STL Partners, “From Telco to Techno: Six tenets for success”, August 2023 [Online]. Available: https:// https://stlpartners.com/research/telco-to-techco-six-tenets-for-success/

[Szi21] P. Szilágyi, “I2BN: Intelligent Intent Based Networks,” Journal of ICT Standardization, vol. 9, no. 2, pp. 159–200, 2021. https://doi.org/10.13052/jicts2245-800X.926

[SZS+13] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erham, I Goodfellow, and R. Fergus, “Intriguing properties of neural networks,” CoRR arXiv:abs/1312.6199, 2013.

[TAS+19] T. Taleb, I. Afolabi, K. Samdanis and F.Z. Yousaf, “On multi-domain network slicing orchestration architecture and federated resource control”, IEEE Network, vol. 33, no. 5, pp. 242-252, 2019.

[TAZ+23] I. Tzanettis, C. M. Androna, A. Zafeiropoulos, E. Fotopoulou, S. Papavassiliou, Data Fusion of Observability Signals for Assisting Orchestration of Distributed Applications. Sensors 2022, 22, 2061. https://doi.org/10.3390/s22052061

[TCE21] Ö. F. Tuna, F. O. Catak, and M. T. Eskil, “Exploiting epistemic uncertainty of the deep learning models to generate adversarial samples,” Multimedia Tools and Applications, vol. 81, pp. 11479 – 11500, 2021.

[TGE+19] O. A Topal, S. Gecgel, E. M. Eksioglu, and G. K. Kurt, "Identification of smart jammers: Learning-based approaches using wavelet preprocessing," Physical Communication, vol. 39, pp. 101029, 2020.

[TIP21] Telecom Infra Project, “Open Transport SDN Architecture Whitepaper”, 2021 [Online]. Link: https://cdn.brandfolder.io/D8DI15S7/at/jh6nnbb6bjvn7w7t5jbgm5n/OpenTransportArchitecture-Whitepaper_TIP_Final.pdf

[TKK23] Ö. F. Tuna, F. E. Kadan, and L. Karaçay, “Practical Adversarial Attacks Against AI-Driven Power Allocation in a Distributed MIMO Network,” arXiv preprint, 2023. arXiv:2301.09305

[TMF_IG1253] TMForum - IG1253C Intent Life Cycle Management and Interface v1.1.0, Nov. 2021.

[TMR20] TrendMicro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol's European Cybercrime Centre (EC3), "Malicious Uses and Abuses of Artificial Intelligence", 2020

[TPM2] Trusted Platform Group, TPM 2.0 Library, https://trustedcomputinggroup.org/resource/tpm-library-specification/

[VMA+18] L. Valcarenghi, B. Martini, K. Antevski, C. J. Bernardos, G. Landi, M. Capitani, J. Mangues-Bafalluy, R. Martinez, J. Baranda, I. Pascual, A. Ksentini, C. F. Chiasserini, F. Malandrino, X. Li, D. Andrushko, and K. Tomakh, "A Framework for Orchestration and Federation of 5G Services in a Multi-Domain Scenario", Workshop on Experimentation and Measurements in 5G (EM-5G'18). ACM, pp. 19- 24, 2018.

[W3C14] W3C (World Wide Web Consortium), "RDF 1.1 concepts and abstract syntax," 2014.

[WTE+20] L. Wu, J. Tordsson, E. Elmroth, and O. Kao, "MicroRCA: Root cause localization of performance issues in microservices," In NOMS 2020 IEEE/IFIP Network Operations and Management Symposium. IEEE, New York, 1–9, 2020. DOI: https://doi.org/10.1109/NOMS47738.2020.9110353

[WZC+20] L. Wang, N. Zhao, J. Chen, P. Li, W. Zhang, and K. Sui, "Root-cause metric location for microservice systems via log anomaly detection," In Proceedings of the 2020 IEEE International Conference on Web Services. IEEE, New York, pp. 142–150, 2020. DOI: https://doi.org/10.1109/ICWS49710.2020.00026

[XGH+20] M. Xie, P. H. Gomes, J. Harmatos and J. Ordonez-Lucena, "Collaborated Closed Loops for Autonomous End-to-End Service Management in 5G," 2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Leganes, Spain, 2020, pp. 64-70, doi: 10.1109/NFV-SDN50289.2020.9289902.

[XTZ+05] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," In Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, pages 46–57, 2005.

[STC+20] Y. Shen, H. Tian, Y. Chen, K. Chen, R. Wang, Y. Xu, Y. Xia, S. Yan, "Occlum: Secure and Efficient Multitasking Inside a Single Enclave of Intel SGX", ASPLOS '20: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems, March, 2020, pages 955–970, https://doi.org/10.1145/3373376.3378469.

[ZAM22] W. Zhao, S. Alwidian, and Q. H. Mahmoud, "Adversarial Training Methods for Deep Learning: A Systematic Review." Algorithms vol. 15, no. 8. pp. 283, 2022.

[ZLF22] X. Zheng, A. Leivadeas, M. Falkner, "Intent Based Networking management with conflict detection and policy resolution in an enterprise network", Computer Networks, Vol. 219, pp. 109457, 2022, https://doi.org/10.1016/j.comnet.2022.109457.

[ZSM002] ETSI ZSM, "Zero-touch network and Service management (ZSM); Architecture Reference", v1.1.1, Group Standard, ETSI, Aug. 2019.

[ZSM011] ETSI ZSM. "Zero-touch Network and Service Management; Intent-driven autonomous networks; Generic Aspects," v1.1.1, Group report, ETSI, February 2023.

[ZYD+23] C. Zhou, H. Yang, X. Duan, D. Lopez, A. Pastor, Q. Wu, M. Boucadair, C. Jacquenet, "Digital Twin Network: Concepts and Reference Architecture", April 2023. https://datatracker.ietf.org/doc/draft-irtf-nmrg-network-digital-twin-arch/

9 Annex

First iteration of enabler analysis: Summary of enablers and their adherence to Hexa-X-II system design principles.

Enablers	Design principles									
	Support and exposure of 6G services and capabilities	Full automation and optimization	Flexibility to different topologies	Network scalability	Resilience and availability	Persistent security and privacy	Internal interfaces are cloud optimized	Separation of concerns of network functions	Network simplification in comparison to previous generation	Minimize environmental footprint and enabling sustainable networks
Enablers for intent-based management automation										
Intent translation and provisioning	X	X	X							
Data fusion mechanisms based on telemetry data	X				X					
Closed loop coordination		X	X	X	X					
Intent conflict administration										
Human-machine intent interface design		X	X							
Intent-driven placement	X	X								
Declarative intent reconciliation	X	X								
Intent reporting		X								
3rd party facing services	X	X								
Enablers for smart network management										
Programmable and flexible network configuration		X								X
Programmable network monitoring and telemetry		X				X				X
Integration fabric	X						X			
Trustworthy management	X					X				
Multi-cloud management mechanisms		X			X					
Orchestration mechanisms for the computing continuum		X			X					
Sustainable AI/ML-based control		X			X					X
Trustworthy AI/ML-based control					X					

Network Digital Twins	X	X								
Zero-touch closed loop governance		X			X					
Zero-touch control loop coordination		X	X	X	X					
Architectural enablers for cloud transformation										
Integration and orchestration of computing continuum resources into the 6G architecture		X	X				X			
Multi-domain/multi-cloud federation					X		X			
Network modules placements in the resource continuum					X		X			
Cloud transformation in 6G-quantum architecture				X						X
Architectural enablers for network function modularization										
Optimized network function composition			X	X	X					
Streamlined network function interfaces and interaction	X	X	X	X	X	X	X	X	X	X
Flexible feature development and run-time scalability	X	X	X	X	X	X	X	X	X	X
Network autonomy and multi-X orchestration	X	X	X	X	X	X	X		X	X