Partners: OPT, ATO, NXW, CTT, IMEC, UC3, ICC, TNO, EBY, TID, ASA, WIN, EAB, UBW, UMU.

HEXA-X-II D6.5 Deliverable

D6.5 Summary Slides: Final Design of 6G Smart Network Management framework

Hexa-X-II

hexa-x-ii.eu

Table of Contents

Section 1 (Slides 1-9)

Introduction

- Deliverable properties
- Executive Summary
- WP6 objectives
- Evolution of work in WP6

Section 2 (Slides 10-36)

Smart Management Framework Description

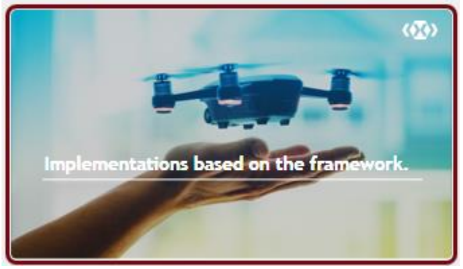
- Overall framework description
- Description of each component in the framework

Section 3 (Slides 37-56)

Framework components implementation.

- Description of the implementation of selected components *of the* framework, along with evaluation results

Section 4 (Slides 57-90)



Implementations based on the framework.

- Description of a set of implementations *based on* the framework, along with evaluation results

Section 5 (Slides 91-118)

Impact

- Publications
- Open-Source Software
- Dissemination and Communication
- KERs and KVIs.
- Quantifiable Targets
- Advisory group recommendations
- Conclusions
- Vision for the future

Acronyms

Acronym	meaning
AI	Artificial Intelligence
AIaaS	AI as a Service
CL	Closed Loop
DLT	Distributed Ledger Technology
E2E	End-to-end
EDA	Event-driven Architecture
ERAB	Evolved Radio Access Bearer
ETSI	European Telecommunications Standards Institute
GAT	Graph Attention
INT	In-band Network Telemetry
IoT	Internet of Things
KER	Key Exploitable Result
KPI	Key Performance Indicator
KVI	Key Value Indicator

Acronym	meaning
LoTAF	Level of Trust Assessment Function
M&O	Management and Orchestration
MAS	Multi-agent System
MCES	Management Capabilities Exposure System
mIoT	Massive IoT
ML	Machine Learning
MLOps	Machine Learning Operations
NDT	Network Digital Twin
NOC	Network Operations Centre
NWDAF	Network Data Analytics Function
OPEX	Operational Expenditure
PLMN	Public land mobile network
PoC	Proof of Concept
QoE	Quality of Experience

Acronym	meaning
QoS	Quality of Service
RBAC	Role-based Access Control
REST	Representational state transfer
RL	Reinforcement Learning
RRC	Radio Resource Control
SDN	Software Defined Network
SLA	Service Level Agreement
TEF	Trust Evaluation Function
TLA	Trust Level Agreement
UE	User Equipment
URLLC	Ultra-Reliable Low Latency Communication
URSP	UE Route Selection Policy
WP	Work Package
XAI	Explainable AI

Executive Summary (1/2)

These slides are a summary of the final public Deliverable D6.5 produced by the Hexa-X-II “Smart Network Management” Work Package (WP6), documenting the evolution in the design and implementation of the envisioned 6G Smart Network Management Framework and a set of evaluation results.

WP6 has defined a framework considering: (i) the contributions to the Hexa-X-II architecture design principles, (ii) the mapping with the envisaged 6G stakeholders, (iii) the definition and the alignment of the WP6 M&O technical enablers with the initial blueprint provided from WP2, and (iv), the envisaged contributions of those M&O enablers towards the future 6G smart networks, as detailed in D6.3 [HEX224-D63].

The primary goal of this framework is to serve as a reference structure for other WPs in the project, and mainly to WP2, to support the design and implementation of the final end-to-end (E2E) system blueprint addressed in such WP. However, the ambition is of course that this WP6 framework can be used as a reference beyond the Hexa-X-II project itself, also becoming a referent for designing and implementing the actual management and orchestration (M&O) systems of the future 6G networks.

The WP6 framework is divided into four main sections:

- The Overall Architectural M&O Solutions, which contains the systems designed to manage and orchestrate resources and services across the entire network continuum, enabling the M&O of resources and services even beyond the technical and administrative boundaries of individual stakeholders.
- The Specific Systems that would be deployed in the scope of specific stakeholders, e.g., network resources programmability systems, provisioning systems, or others.
- The Overall Functionalities that are more specific in scope than the Overall Architectural M&O Solutions. Such functionalities include, for instance, the monitoring functionality or the trust management systems.
- The Algorithms that would be deployed on the stakeholder’s scope. A set of algorithms is provided that were found of interest in WP6, in line with the Hexa-X-II project work programme.

Executive Summary (2/2)

For each section of the framework, details are provided for the designed and developed components. Initially, the design of each component is provided, focusing on the main supported characteristics and its suitability for managing specific functionalities or services of the framework. Following, implementation details are given, considering both the implementation of individual components of the framework, as well as implementation of workflows that engage multiple components of the framework.

Evaluation results are presented in most of the cases based on the status of evaluation activities in the project. For part of the cases, the provided evaluation results are considered final, while for the rest, evaluation activities are going to be continued in the context of the PoC developments in WP2 towards the end of the lifetime of the Hexa-X-II project work programme.

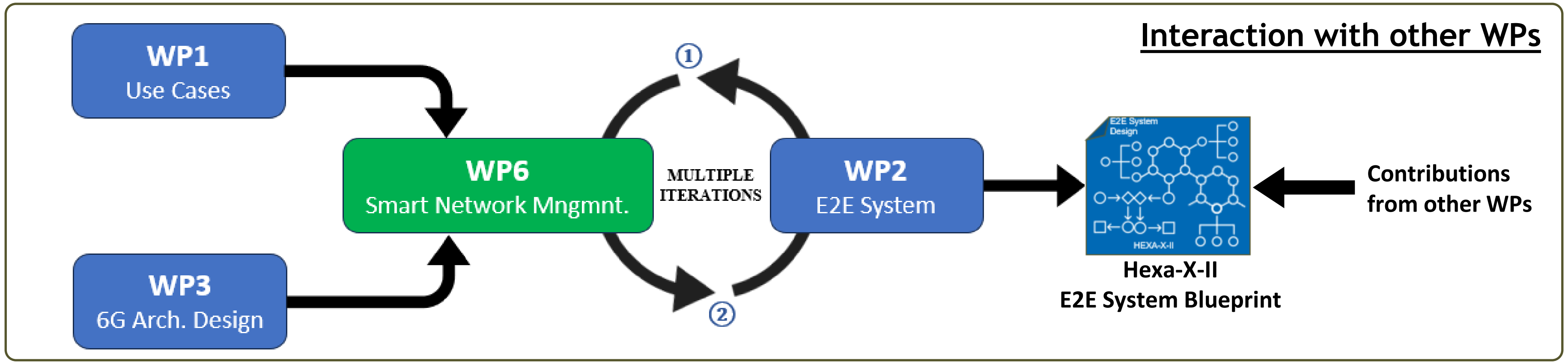
Furthermore, details are presented for the dissemination and standardisation activities realised in the context of WP6, the WP6 contribution to the project Key Exploitable Results, the estimated impact on the Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) defined in the Hexa-X-II project, the contribution from WP6 to the Hexa-X-II project objectives, and the alignment with the recommendations provided by the Advisory Group.

Finally, a set of concluding remarks are detailed for the work done within WP6 in the Hexa-X-II project.

Introduction

Introduction

- The Hexa-X-II project is a flagship initiative of the Smart Networks and Services Joint Undertaking (SNS JU), aimed at pioneering the design of an end-to-end (E2E) system blueprint built on integrated and interoperable technology enablers, and advancing efforts to shape the global communications network of the 2030s.
- Its goal is to provide a framework for delivering innovative services for the next generation (6G) of wireless networks, building on the work of the previous Horizon Europe Hexa-X project, which established the 6G vision and foundational concepts including key candidate technologies.
- The Hexa-X-II project comprises several Work Packages that span over different parts of the envisaged 6G ecosystem. In these slides results from work in WP6 are presented, which targets the network and services management and orchestration aspects. More specifically, this slide set represents the summary of the fifth deliverable from WP6, referred to as D6.5.
- The purpose of this presentation is to summarise the design and implementation of the 6G Smart Network Management Framework, while also providing a set of evaluation results. Furthermore, details are presented for the dissemination and standardisation activities realised in the context of WP6, the WP6 contribution to the project Key Exploitable Results, the estimated impact on the Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) defined in the Hexa-X-II project, the contribution from WP6 to the Hexa-X-II project objectives, and the alignment with the recommendations provided by the Advisory Group.
- The following diagram shows the overall role of this WP6 in the context of the Hexa-X-II project, targeting to contribute to the E2E System Blueprint being addressed in WP2:

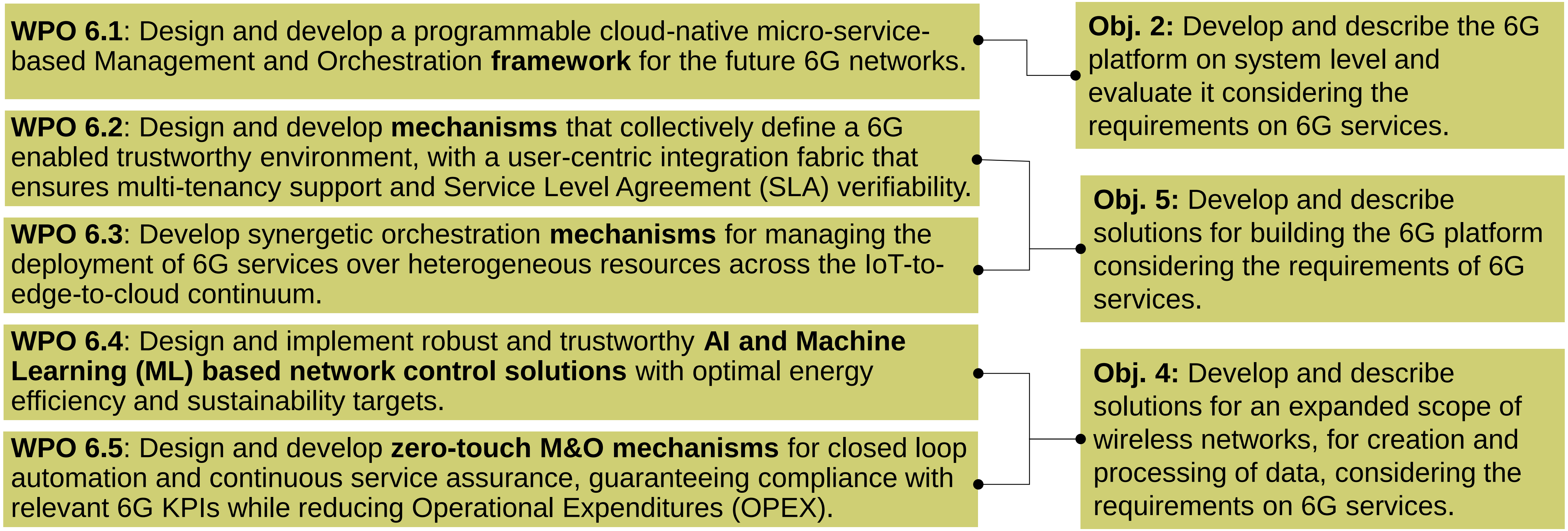


[HEX224-D63]

WP6 Objectives

According to the GA, WP6 has the goal to **design** and **implement** smart network management and orchestration mechanisms, considering both network and cloud resources, to be part of the 6G overall system, while contributing to achieve 6G KPIs and KVIs, such as full automation, programmability, trustworthiness, and sustainability.

The 5 specific objectives defined for WP6 and the 3 overall Hexa-X-II objectives to which they relate are:



WPO 6.1: Design and develop a programmable cloud-native micro-service-based Management and Orchestration **framework** for the future 6G networks.

WPO 6.2: Design and develop **mechanisms** that collectively define a 6G enabled trustworthy environment, with a user-centric integration fabric that ensures multi-tenancy support and Service Level Agreement (SLA) verifiability.

WPO 6.3: Develop synergetic orchestration **mechanisms** for managing the deployment of 6G services over heterogeneous resources across the IoT-to-edge-to-cloud continuum.

WPO 6.4: Design and implement robust and trustworthy **AI and Machine Learning (ML) based network control solutions** with optimal energy efficiency and sustainability targets.

WPO 6.5: Design and develop **zero-touch M&O mechanisms** for closed loop automation and continuous service assurance, guaranteeing compliance with relevant 6G KPIs while reducing Operational Expenditures (OPEX).

Obj. 2: Develop and describe the 6G platform on system level and evaluate it considering the requirements on 6G services.

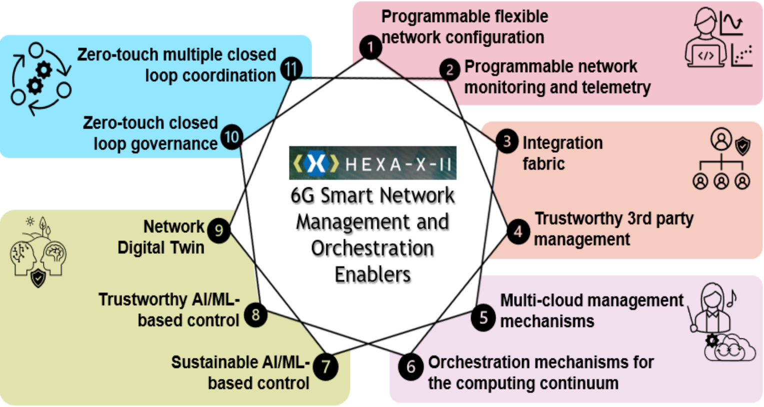
Obj. 5: Develop and describe solutions for building the 6G platform considering the requirements of 6G services.

Obj. 4: Develop and describe solutions for an expanded scope of wireless networks, for creation and processing of data, considering the requirements on 6G services.

Evolution of the work in WP6

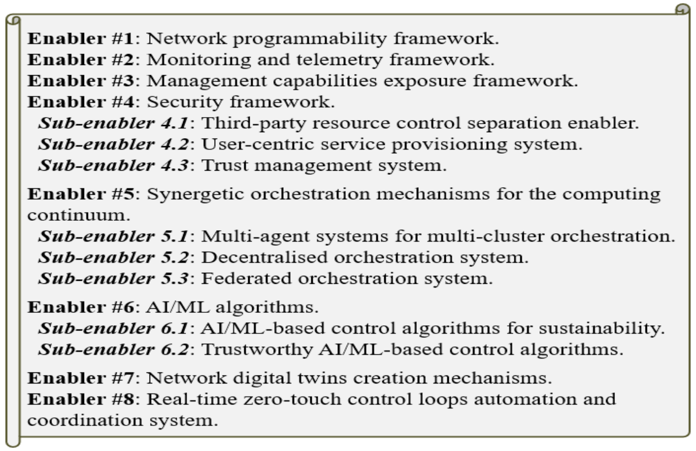
The following diagram summarises the evolution of the work in WP6, from the initial definition of the technical enablers early considered in the WP, to the definition of the final Smart Management Framework, based on these enablers.

Deliverables D6.1 & D6.2


HEXA-X-II
6G Smart Network Management and Orchestration Enablers
1 Programmable 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 Twin
10 Zero-touch closed loop governance
11 Zero-touch multiple closed loop coordination

*Identification of the initial set of the 6G smart network management **technical enablers**. 11 enablers were early identified, grouped into 5 categories.*

Deliverable D6.3



Enabler #1: Network programmability framework.
Enabler #2: Monitoring and telemetry framework.
Enabler #3: Management capabilities exposure framework.
Enabler #4: Security framework.
Sub-enabler 4.1: Third-party resource control separation enabler.
Sub-enabler 4.2: User-centric service provisioning system.
Sub-enabler 4.3: Trust management system.

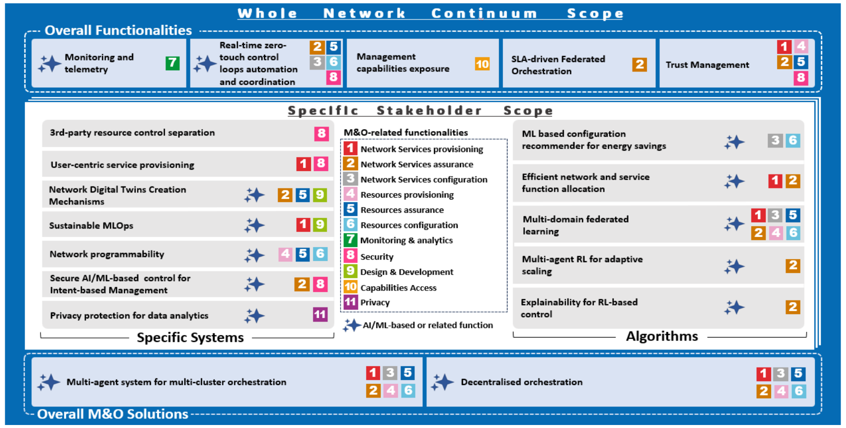
Enabler #5: Synergetic orchestration mechanisms for the computing continuum.
Sub-enabler 5.1: Multi-agent systems for multi-cluster orchestration.
Sub-enabler 5.2: Decentralised orchestration system.
Sub-enabler 5.3: Federated orchestration system.

Enabler #6: AI/ML algorithms.
Sub-enabler 6.1: AI/ML-based control algorithms for sustainability.
Sub-enabler 6.2: Trustworthy AI/ML-based control algorithms.

Enabler #7: Network digital twins creation mechanisms.
Enabler #8: Real-time zero-touch control loops automation and coordination system.

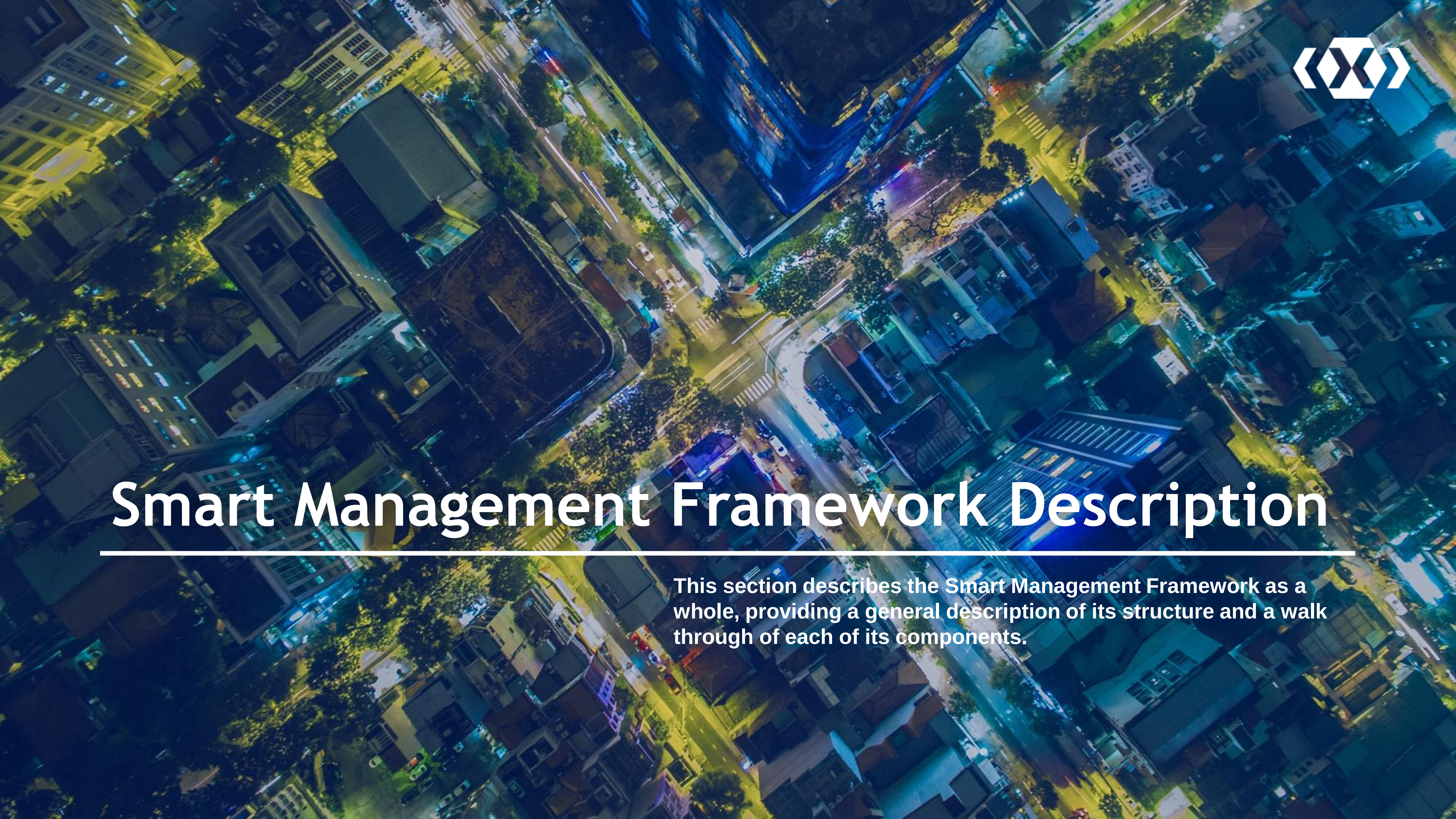
*Enablers streamlined renaming them and introducing the sub-enablers concept. Initial concept of the **Smart Management Framework** based on that set of enablers.*

Deliverables D6.4 & D6.5


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party resource control separation 8
User-centric service provisioning 1 8
Network Digital Twins Creation Mechanisms 2 5 9
Sustainable MLOps 1 9
Network programmability 4 5 6
Secure AI/ML-based control for Intent-based Management 2 8
Privacy protection for data analytics 11
Specific Systems
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based configuration recommender for energy savings 3 6
Efficient network and service function allocation 1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

*Final definition of the WP6 **Smart Management Framework**. Early report in the D6.4 presentation and full description in Deliverable D6.5.*

The following slides present this WP6 Smart Management Framework with more detail.

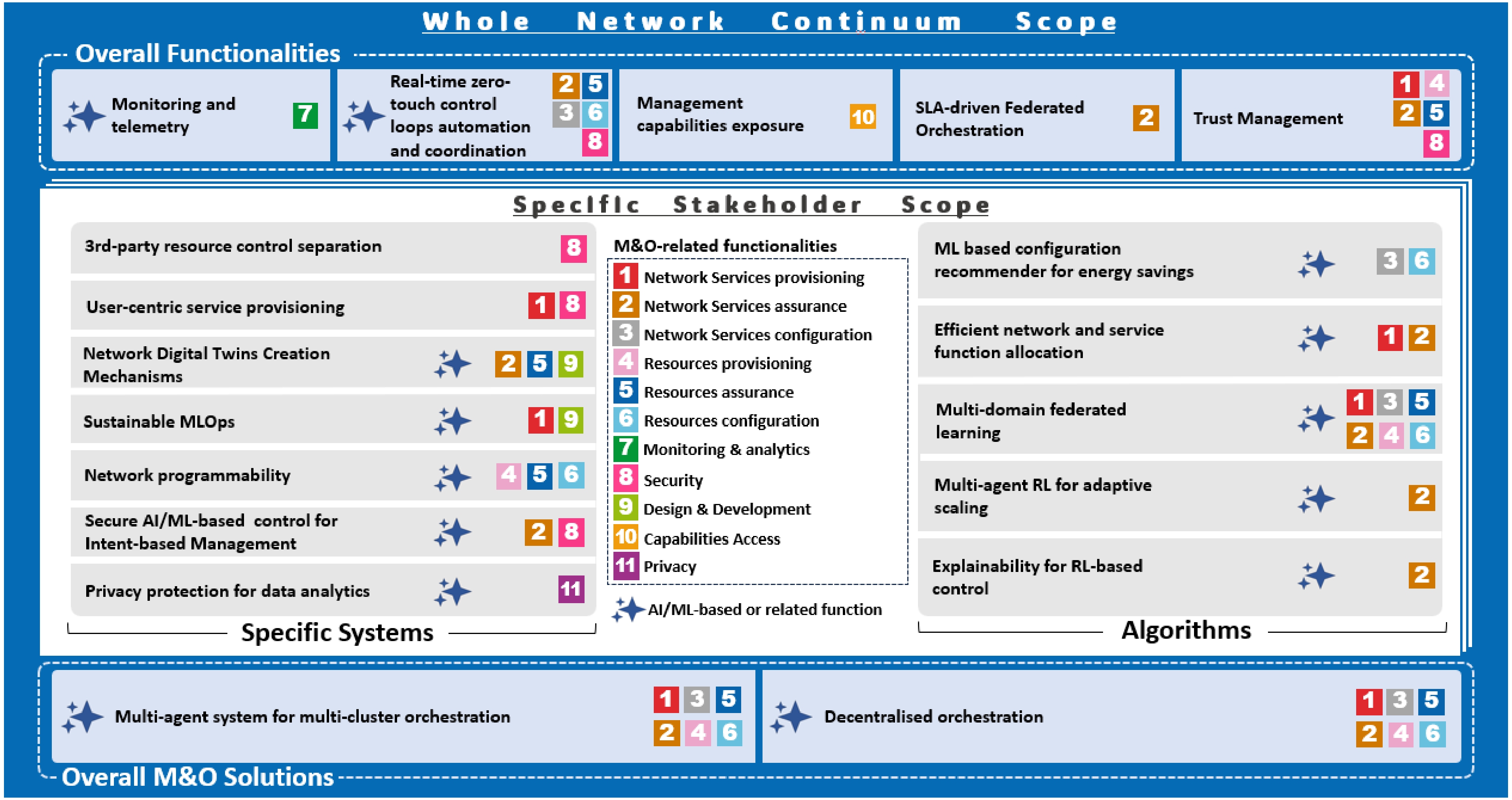


Smart Management Framework Description

This section describes the Smart Management Framework as a whole, providing a general description of its structure and a walk through of each of its components.

Hexa-X-II Smart Management Framework - Overall view (1/2)

The figure below provides an overview of the smart management framework, with its main technical enablers grouped into different categories:



Whole Network Continuum Scope

Overall Functionalities

- ✦ Monitoring and telemetry — 7
- ✦ Real-time zero-touch control loops automation and coordination — 2 5 3 6 8
- Management capabilities exposure — 10
- SLA-driven Federated Orchestration — 2
- Trust Management — 1 4 2 5 8

Specific Stakeholder Scope

Specific Systems

- 3rd-party resource control separation — 8
- User-centric service provisioning — 1 8
- ✦ Network Digital Twins Creation Mechanisms — 2 5 9
- ✦ Sustainable MLOps — 1 9
- ✦ Network programmability — 4 5 6
- ✦ Secure AI/ML-based control for Intent-based Management — 2 8
- ✦ Privacy protection for data analytics — 11

M&O-related functionalities

1	Network Services provisioning
2	Network Services assurance
3	Network Services configuration
4	Resources provisioning
5	Resources assurance
6	Resources configuration
7	Monitoring & analytics
8	Security
9	Design & Development
10	Capabilities Access
11	Privacy
✦	AI/ML-based or related function

Algorithms

- ✦ ML based configuration recommender for energy savings — 3 6
- ✦ Efficient network and service function allocation — 1 2
- ✦ Multi-domain federated learning — 1 3 5 2 4 6
- ✦ Multi-agent RL for adaptive scaling — 2
- ✦ Explainability for RL-based control — 2

Overall M&O Solutions

- ✦ Multi-agent system for multi-cluster orchestration — 1 3 5 2 4 6
- ✦ Decentralised orchestration — 1 3 5 2 4 6

▶ The framework is designed as a supporting structure on which the M&O systems towards 6G can be built, offering a system of rules, ideas, and innovations that can be used to plan and decide on such systems. ◀

Hexa-X-II Smart Management Framework - Overall view (2/2)

Technical enablers are grouped into two overall scopes:

- Those targeting the whole network continuum scope (outer light blue frame), i.e., considering the network domains beyond the individual stakeholder's boundaries.
- Those in the particular scope of specific stakeholders (inner white rectangle).

4

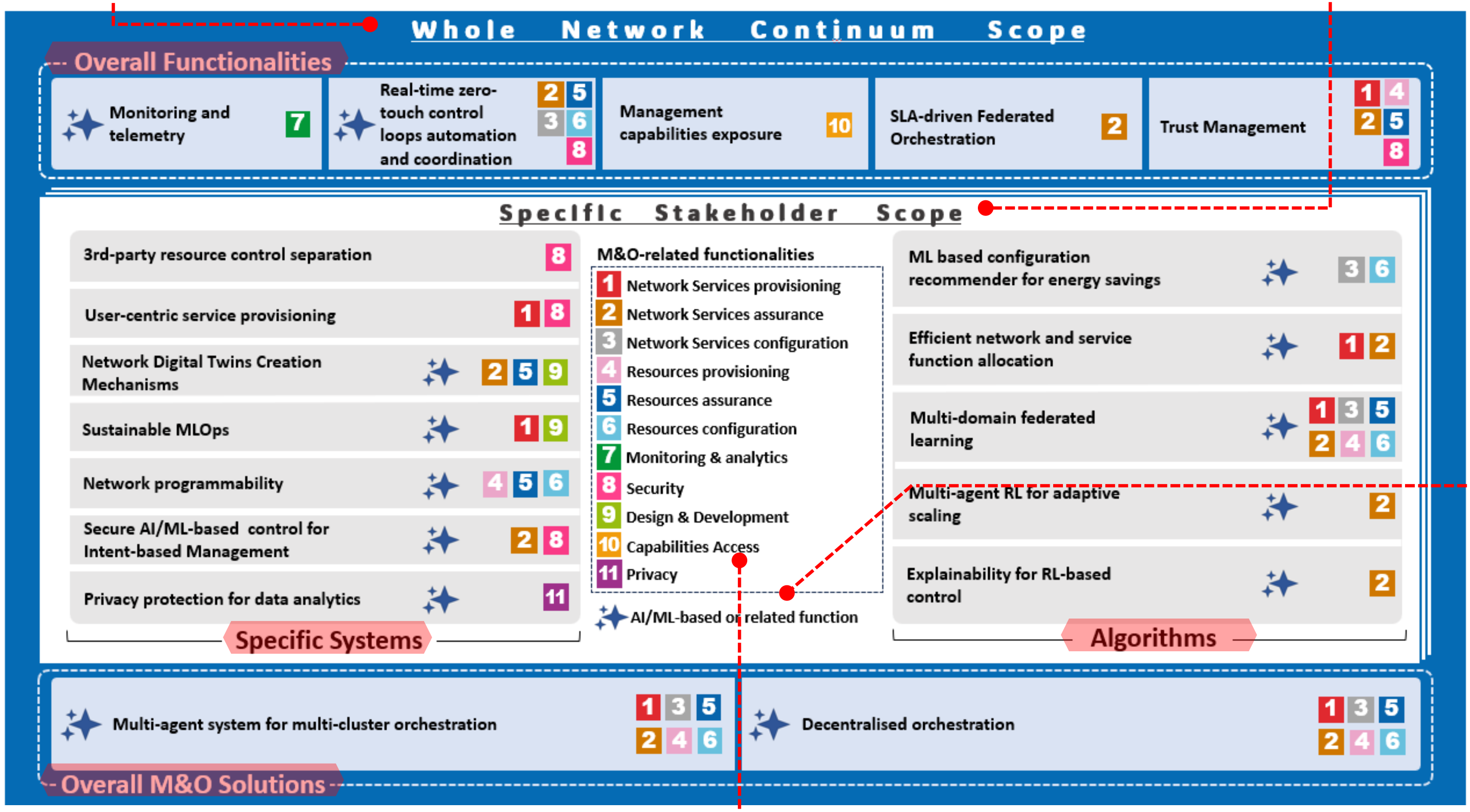
Enablers are also grouped into four main categories highlighted in red in the figure:

a) Overall M&O Solutions (bottom) with enablers integrating an extensive set of M&O mechanisms across the whole network continuum.

b) Overall Functionalities (Top Rectangle): Specific functionalities also targeting the whole network continuum.

c) Specific Systems (Left Grey Blocks): Systems for specific stakeholders.

d) Algorithms (Right Grey Blocks): Selection of algorithms, also to be deployed within the stakeholder's scope.



Whole Network Continuum Scope

Overall Functionalities

✦ Monitoring and telemetry	7
✦ Real-time zero-touch control loops automation and coordination	2 5 3 6 8
Management capabilities exposure	10
SLA-driven Federated Orchestration	2
Trust Management	1 4 2 5 8

Specific Stakeholder Scope

Specific Systems

3rd-party resource control separation	8
User-centric service provisioning	1 8
Network Digital Twins Creation Mechanisms ✦	2 5 9
Sustainable MLOps ✦	1 9
Network programmability ✦	4 5 6
Secure AI/ML-based control for Intent-based Management ✦	2 8
Privacy protection for data analytics ✦	11

M&O-related functionalities

1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy

✦ AI/ML-based or related function

Algorithms

ML based configuration recommender for energy savings ✦	3 6
Efficient network and service function allocation ✦	1 2
Multi-domain federated learning ✦	1 3 5 2 4 6
Multi-agent RL for adaptive scaling ✦	2
Explainability for RL-based control ✦	2

Overall M&O Solutions

✦ Multi-agent system for multi-cluster orchestration	1 3 5 2 4 6
✦ Decentralised orchestration	1 3 5 2 4 6

3

Beyond these common management functions, the Framework also includes cutting-edge technologies, e.g., the extensive usage of AI/ML (represented by the sparkle icon ✦), zero-touch automation through closed control loops, or digital twins creation mechanisms, among others.

2

1

The Framework includes those functionalities commonly found in M&O systems (e.g., service/resource provisioning, assurance, monitoring…), represented by the coloured numbered list in the middle of the figure. These numbers are assigned to specific components based on their functionality.

Hexa-X-II M&O Framework
Overall M&O Solutions

- The Overall M&O solutions are enablers integrating a rich set of M&O mechanisms, targeting to manage and orchestrate network services and resources across the whole network continuum. The “overall” term here refers the extensive set of functionalities integrated in these solutions.
- Two approaches are offered in the framework, namely Hierarchical and Decentralised.
  - The Hierarchical approach is based on a multi-agent system for multi-cluster orchestration and includes a centralised E2E orchestrator and distributed agents to manage resources and services across platforms and domains, including also the extreme-edge domain, and integrating AI/ML techniques for the proactive allocation of service components and to perform predictive M&O actions.
  - The Decentralised approach targets to manage network and services in a fully distributed manner, focusing on providing service continuity, scalability, and optimised resource usage considering the highly heterogeneous and the volatile resources in the extreme-edge domain. It includes zero-touch infrastructure discovery mechanisms and AI/ML for predicting changes in the infrastructure.
- Main challenges addressed:
  - The envisaged highly distributed and multi-stakeholder ecosystem towards 6G.
  - The integration of the extreme-edge domain, with its volatile resources, cloud-native scale, and varied kind of devices.
- Key concept: Utilising the combined computing and storage capacity of the extreme-edge domain to
  - Distribute workloads efficiently,
  - Reduce data communication needs,
  - Lower latency, potentially surpassing current 5G capabilities,
  - Enable new business models for stakeholders.

Multi-agent system for multi-cluster orchestration

1 3 5
2 4 6

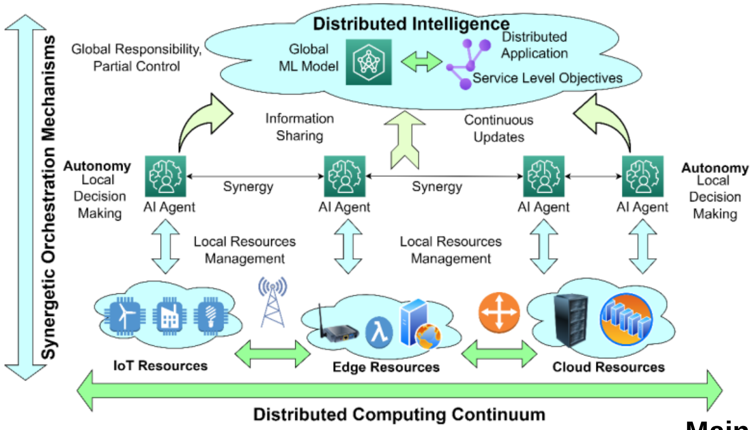
Decentralised orchestration system

1 3 5
2 4 6

Overall M&O Solutions

Hexa-X-II M&O Framework - Overall M&O Solutions
Multi-agent system for multi-cluster orchestration

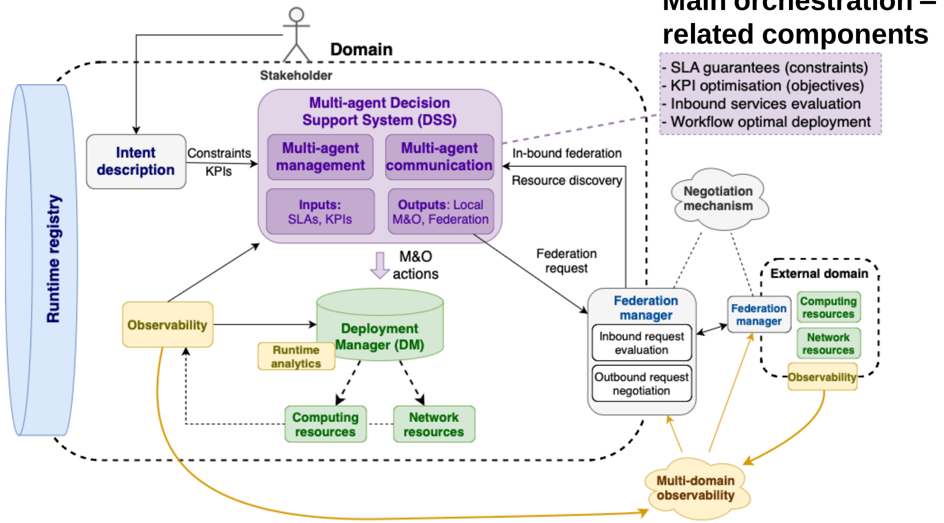
- Provides a hierarchical multi-agent approach for network services and resources provisioning, assurance and configuration that is applicable for management of distributed services over resources in the computing continuum.
- Offers a methodology for combining specific systems and algorithms working together to automate decision making and multi-cluster service orchestration. Increased autonomy and decentralised intelligence in orchestration mechanisms.
- Leverages multi-agent systems applied in a variety of scopes and deployments, e.g. across different services, or different layers of the computing continuum.
- Highly dependent on monitoring and telemetry inputs, facilitates coordination between orchestration algorithms (e.g. Multi-agent Reinforcement Learning –RL– for adaptive scaling), zero-touch automation loops, and multi-cluster management technologies.
- Fine-grained distributed control of resources and services, allowing multi-actor synergetic systems to optimise towards common objectives.
- The enabler targets the following Hexa-X-II design principles defined in WP2 [HEX223-D21]: (1) Support and exposure of 6G services and capabilities, (2) Full automation and optimisation, (3) Flexibility to different network scenarios, (7) Internal interfaces are cloud-optimised, and (10) Minimising environmental footprint and enabling sustainable networks.



Synergetic Orchestration Mechanisms
Distributed Intelligence
Global Responsibility, Partial Control
Global ML Model
Distributed Application
Service Level Objectives
Information Sharing
Continuous Updates
Autonomy Local Decision Making
AI Agent
Synergy
AI Agent
Synergy
AI Agent
AI Agent
Autonomy Local Decision Making
Local Resources Management
Local Resources Management
IoT Resources
Edge Resources
Cloud Resources
Distributed Computing Continuum

Multi-agent orchestration approach with hierarchical decision making [ZFF+24]

Main orchestration – related components



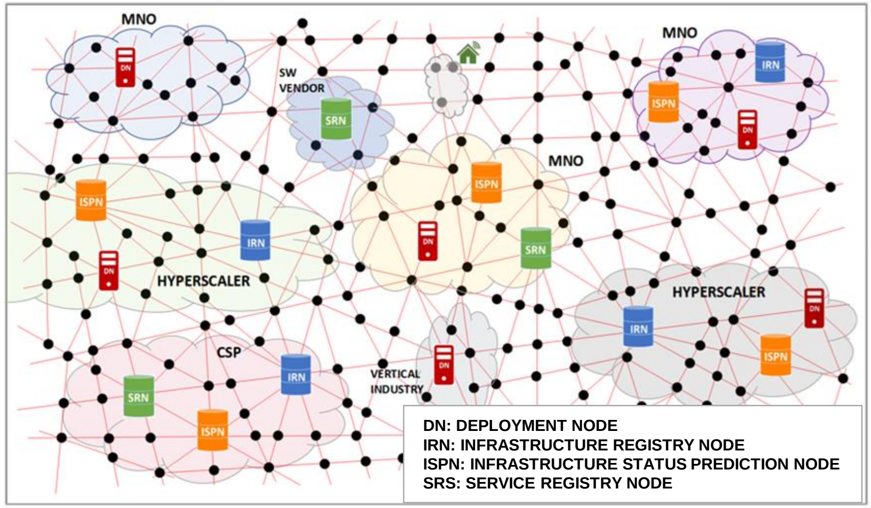
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6

Decentralised orchestration system 1 3 5 2 4 6

Overall M&O Solutions

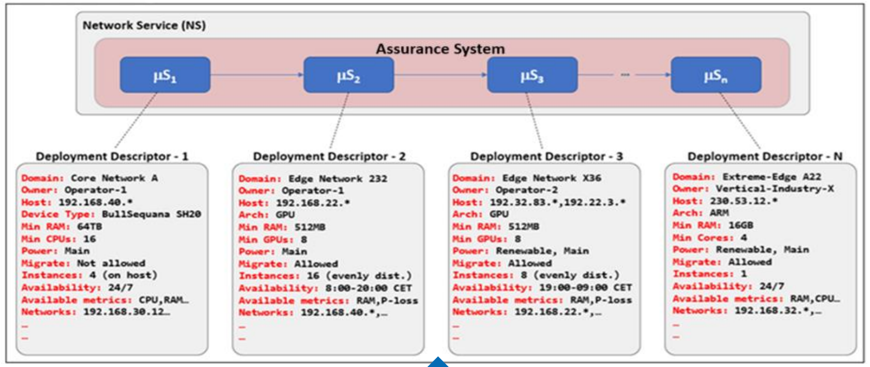
Hexa-X-II M&O Framework - Overall M&O Solutions
Distributed Orchestration System

- Focus on integrating the vast number and diversity of devices in the cloud-native extreme-edge domain, along with the numerous service components that may be deployed on it.
- Inherently multi-domain in a broad sense, enabling service chaining across multiple technological and administrative network domains using cloud-native exposed interfaces and microservices federation, simplifying business and technological agreements between stakeholders to connect their M&O systems.
- The global M&O system is divided into two:
  - The Common Infrastructure Management and Services Provisioning System (CIM & SPS), consisting of four network elements (DN, IRN, ISPM & SRN) spread across the entire network continuum (see figure). It makes possible the onboarding of the network services on the network, considering its large-scale and heterogeneity. It also Includes AI/ML mechanisms to predict the infrastructure changes and optimally allocate the network service components.
  - The service assurance mechanisms, which are tailor-made per service and embedded within network services themselves. Provides service developers the flexibility to design these mechanisms based on the unique needs of the network services they must support.



DN: DEPLOYMENT NODE
IRN: INFRASTRUCTURE REGISTRY NODE
ISPN: INFRASTRUCTURE STATUS PREDICTION NODE
SRS: SERVICE REGISTRY NODE

CIM & SPS network elements deployed through the network continuum [HEX224-D33].



Deployment Descriptor - 1

Domain: Core Network A
Owner: Operator-1
Host: 192.168.40.*
Device Type: BullSequana SH20
Min RAM: 64TB
Min CPUs: 16
Power: Main
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: Operator-1
Host: 192.168.22.*
Arch: GPU
Min RAM: 512MB
Min GPUs: 8
Power: Main
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: Edge Network X36
Owner: Operator-2
Host: 192.32.83.*,192.22.3.*
Arch: GPU
Min RAM: 512MB
Min GPUs: 8
Power: Renewable, Main
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-Industry-X
Host: 230.53.12.*
Arch: ARM
Min RAM: 16GB
Min Cores: 4
Power: Renewable, Main
Migrate: Allowed
Instances: 1
Availability: 24/7
Available metrics: RAM,CPU…
Networks: 192.168.32.*,…
–
–

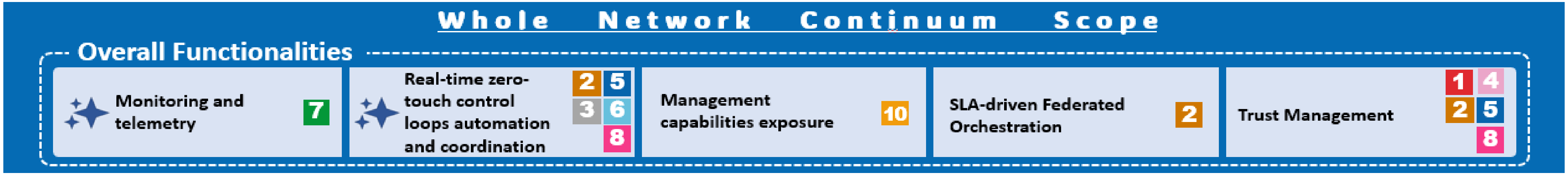
Network Service definition with the embedded tailor-made service assurance system [HEX224-D33].

Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6

Decentralised orchestration system 1 3 5 2 4 6

Overall M&O Solutions

Hexa-X-II M&O Framework Overall Functionalities



Whole Network Continuum Scope

Overall Functionalities

Monitoring and telemetry 7

Real-time zero-touch control loops automation and coordination 2 5 3 6 8

Management capabilities exposure 10

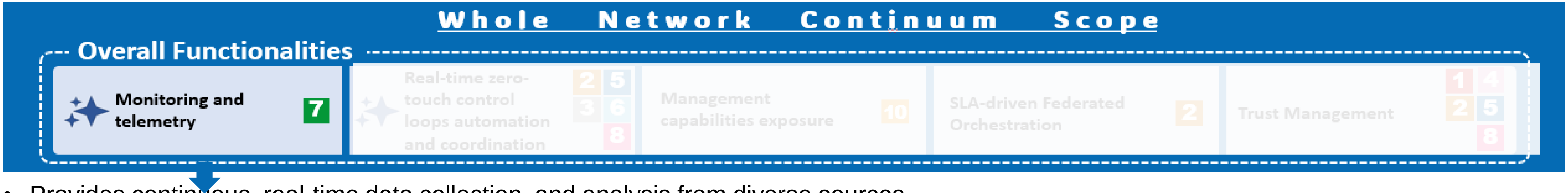
SLA-driven Federated Orchestration 2

Trust Management 1 4 2 5 8

This section outlines the five key functionalities in the framework that, although not so general as the Overall M&O Solutions described before, can complement and enrich those overall solutions, also in the scope of the entire network continuum. They are the following:

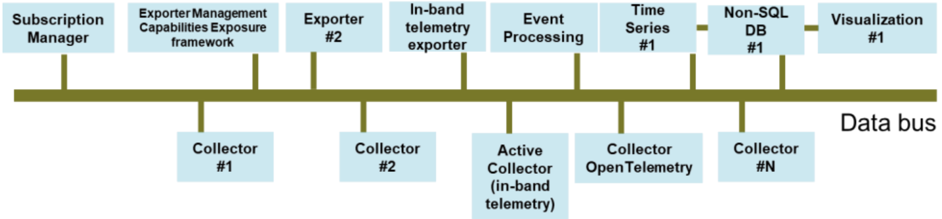
- **Monitoring and telemetry**, enables the integration of diverse monitoring protocols and user-defined metrics across various network domains and support AI/ML models by feeding heterogeneous, multi-domain data for improved decision-making and proactive network management.
- **The real-time zero-touch control loops automation and coordination mechanisms**, intended to automate network management through adaptable and configurable control loops, integrating AI/ML for predictive actions, and allowing autonomous responses to real-time network changes or proactive optimisation actions.
- **The management capabilities exposure** that facilitates cross-domain connectivity, allowing stakeholders to integrate their M&O systems.
- **The SLA-driven federated orchestration mechanism**, supporting dynamic SLA definitions in multi-stakeholder environment using blockchain for autonomous SLA management, reducing 3rd party involvement.
- **The trust management mechanisms**, which can be used to integrate trust evaluation as part of the M&O systems, targeting to ensure secure and efficient resource allocation, particularly in multi-stakeholder environments.

Hexa-X-II M&O Framework - Overall Functionalities
Monitoring and telemetry

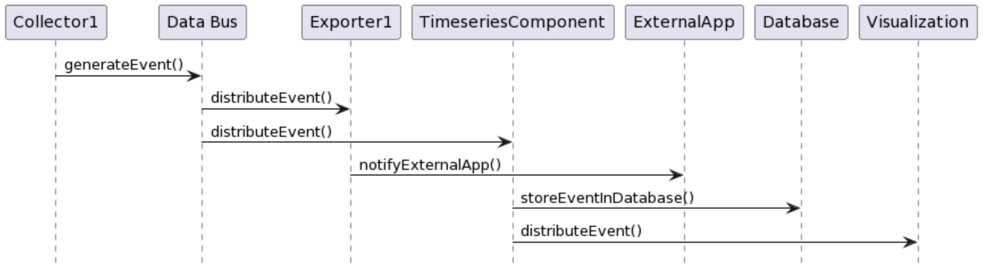

Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
7
Real-time zero-touch control loops automation and coordination
2
5
3
6
8
Management capabilities exposure
10
SLA-driven Federated Orchestration
2
Trust Management
1
4
2
5
8

- Provides continuous, real-time data collection, and analysis from diverse sources, extending the 5G's Network Data Analytics Function (NWDAF) capabilities for 6G networks.
- Cloud-Native Microservices Monitoring and telemetry Architecture:
  - Includes components like Subscription Manager, Exporters (including in-band telemetry), Event Processing, Time Series databases, Non-SQL databases, and Visualisation connected through a common data bus.
  - Incorporates Collectors such as Active Collectors and OpenTelemetry collectors to gather data from virtual and physical components and applications.
- This enabler can also feed multi-domain, heterogeneous data into AI/ML models, enabling advanced decision-making processes.
- Relationship with other enablers in the framework: this enabler could be jointly used with control-loop (CLs) that can use AI/ML for analysis and decision and use this enable to monitor.
- Relationship with WP2 design principles: 2) Full automation and optimisation, 3) Flexibility to different network scenarios, 10) Minimising environmental footprint and enabling sustainable networks.

Monitoring and telemetry functionality architecture



Monitoring and telemetry functionality sequence diagram



Hexa-X-II M&O Framework - Overall Functionalities
Real-time zero touch control loops automation and coordination

Whole Network Continuum Scope

Overall Functionalities

Monitoring and telemetry 7

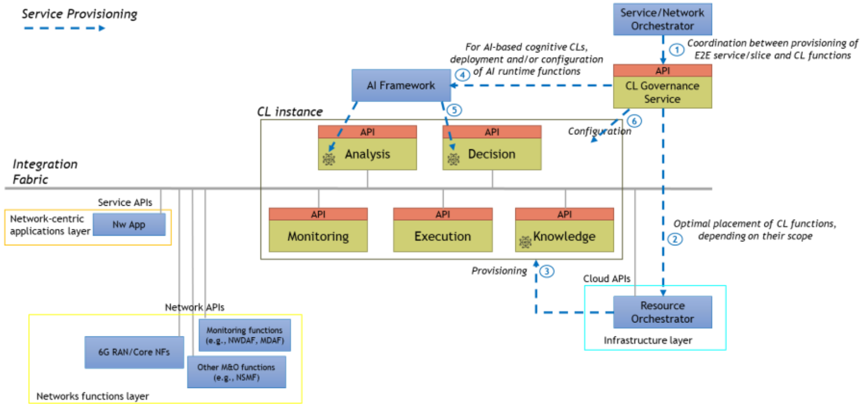
Real-time zero-touch control loops automation and coordination 2 5 3 6 8

Management capabilities exposure 10

SLA-driven Federated Orchestration 2

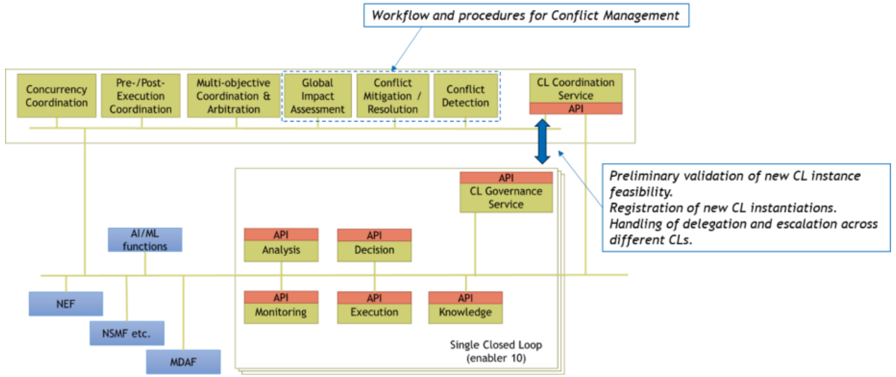
Trust Management 1 4 2 5 8

- Real-time zero-touch Closed Loops provide the foundation for mobile network automation, implementing self-configuration, self-adaptation, and self-optimisation logic.
- Provided M&O functionalities:
  - Network Services assurance and configuration, for CLs working at service or network layer, since they automate the network re-configuration and guarantee the continuous Quality of Service (QoS) level and functionalities to be delivered by Network Services;
  - Resource assurance and configuration, for CLs working at infrastructure layer.
- Relationship with other enablers:
  - CLs can use AI/ML for analysis and decision, use monitoring to operate over real-time data and exploit orchestration and network programmability for dynamic re-configurations. Orchestration can support CL Governance for CL functions' provisioning, while Digital Twins can offer protected sandboxes for CL decisions' validation in CL coordination. Management Capabilities Exposure can mediate the interaction among CL functions.
- Beyond SotA: CL Governance allows to deploy CLs in customizable and automatic manner over the edge/cloud continuum, adjusting their placement, scaling, and configuration to the dynamicity of the network. CL Coordination allows efficient cooperation among CLs.
- Relationship with WP2 design principles: Full automation and optimisation (2), and Network scalability (4).



Service Provisioning
Service/Network Orchestrator
1 Coordination between provisioning of E2E service/slice and CL functions
For AI-based cognitive CLs, deployment and/or configuration of AI runtime functions 4
AI Framework
API
CL Governance Service
5
6
CL instance
Configuration
API
Analysis
API
Decision
Integration Fabric
Service APIs
Network-centric applications layer
Nw App
API
Monitoring
API
Execution
API
Knowledge
Optimal placement of CL functions, depending on their scope 2
Provisioning 3
Cloud APIs
Network APIs
Monitoring functions (e.g., NWDAF, MDAF)
6G RAN/Core NFs
Other M&O functions (e.g., NSMF)
Networks functions layer
Resource Orchestrator
Infrastructure layer

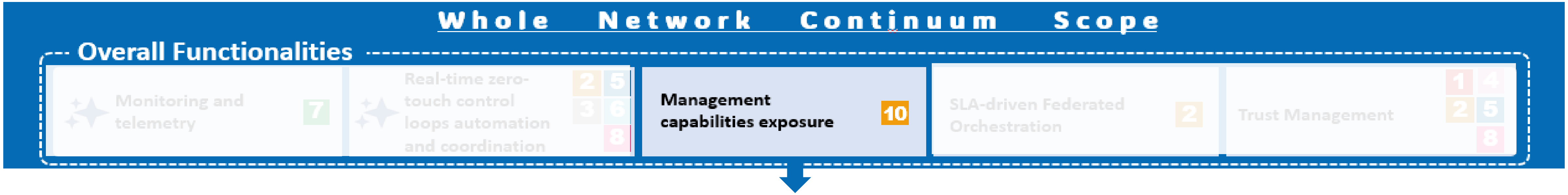
CLs Provisioning [HEX224-D63]



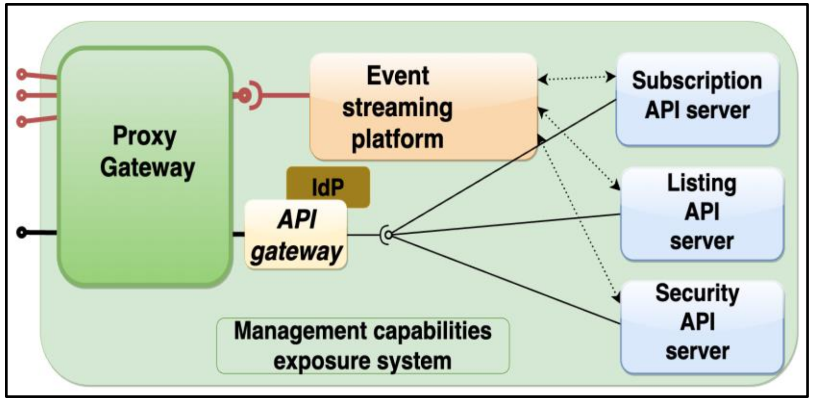
Workflow and procedures for Conflict Management
Concurrency Coordination
Pre-/Post-Execution Coordination
Multi-objective Coordination & Arbitration
Global Impact Assessment
Conflict Mitigation / Resolution
Conflict Detection
CL Coordination Service
API
Preliminary validation of new CL instance feasibility.
Registration of new CL instantiations.
Handling of delegation and escalation across different CLs.
API
CL Governance Service
AI/ML functions
API
Analysis
API
Decision
NEF
NSMF etc.
MDAF
API
Monitoring
API
Execution
API
Knowledge
Single Closed Loop (enabler 10)

CL Coordination [HEX224-D63]

Hexa-X-II M&O Framework - Overall Functionalities
Management Capabilities Exposure


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
7
Real-time zero-touch control loops automation and coordination
2
5
3
6
8
Management capabilities exposure
10
SLA-driven Federated Orchestration
2
Trust Management
1
4
2
5
8

- It serves as a centralised API connector, enabling secure and scalable interactions across M&O systems with seamless, event-driven communication, in line with the European Telecommunications Standards Institute (ETSI) ZSM Integration Fabric specification [ZSM002].
- Enables cross-domain communication and access to M&O services via Representational State Transfer (REST) asynchronous APIs, facilitating multi-stakeholder integration across the network continuum.
- Can work with the following other enablers in the framework:
  - The **Monitoring and Telemetry** functionality, to expose real-time data for optimisation.
  - With the **Multi-agent system for multi-cluster orchestration** and **Decentralised orchestration system** for dynamic orchestration across the network.
  - The **Trust Management** for TLA
  - With the **Closed Loop Automation** and **Network programmability** functionalities, to enable autonomous network adjustments through real-time exposure of network data.
- Enables plug-and-play integration and event-driven coordination at scale, enhancing cross-domain resource management and agility in the envisaged complex 6G ecosystems.



MCE internal architecture

- It aligns with **WP2's principles 1, 3 and 7**, acting as a communication connector within the M&O framework, and external API producers and consumers at service level.

Hexa-X-II M&O Framework - Overall Functionalities
SLA-driven Federated orchestration system

Whole Network Continuum Scope

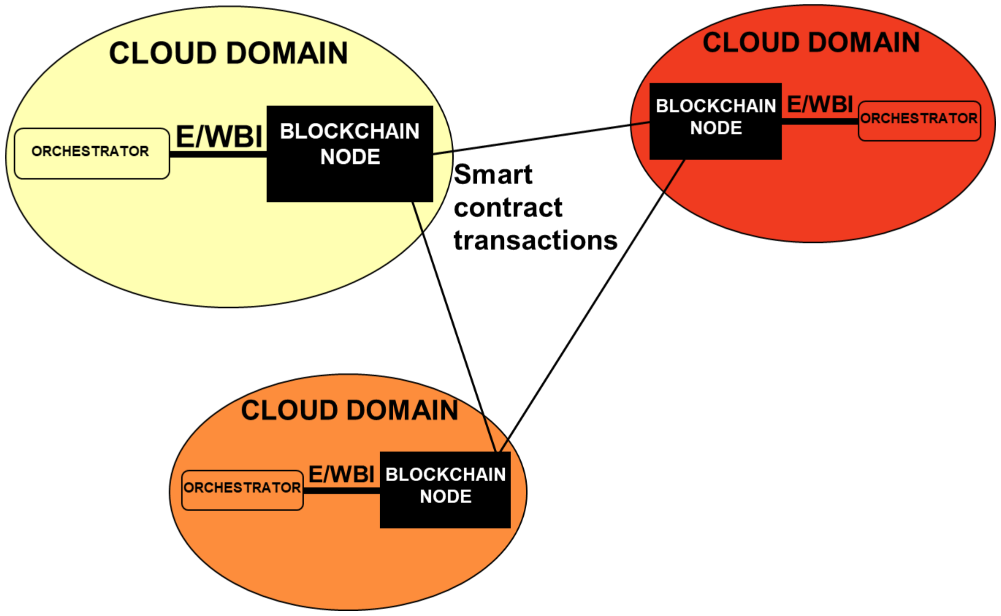
Overall Functionalities

- Monitoring and telemetry 7
- Real-time zero-touch control loops automation and coordination 2 5 3 6 8
- Management capabilities exposure 10
- **SLA-driven Federated Orchestration 2**
- Trust Management 1 4 2 5 8

Provides dynamic SLA creation and policing mechanism using Distributed Ledger Technologies (DLT) i.e. blockchain-based smart contracts to facilitate service continuity beyond a service provider network.

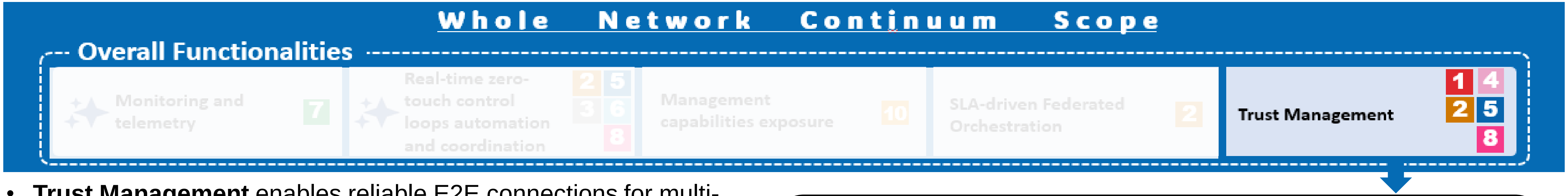
Main features:

- Facilitates Network service assurance – ensuring that a service keeps running in dynamic roaming scenarios.
- It is considered a specific system given that supports service provisioning.
- This enabler would be jointly used with the Monitoring and Telemetry enabler to provide triggers for the smart contracts.
- The broad range of 6G use cases mean that providers beyond Public Land Mobile Network (PLMN) operators might need to be contracted on a short-term basis leveraging the dynamic SLA mechanism provided.
- Supports all the WP1 use cases.
- The enabler is related to the Hexa-X-II architecture design principles: support and exposure of 6G services and capabilities, full automation and optimisation, and provide flexibility to different network scenarios.

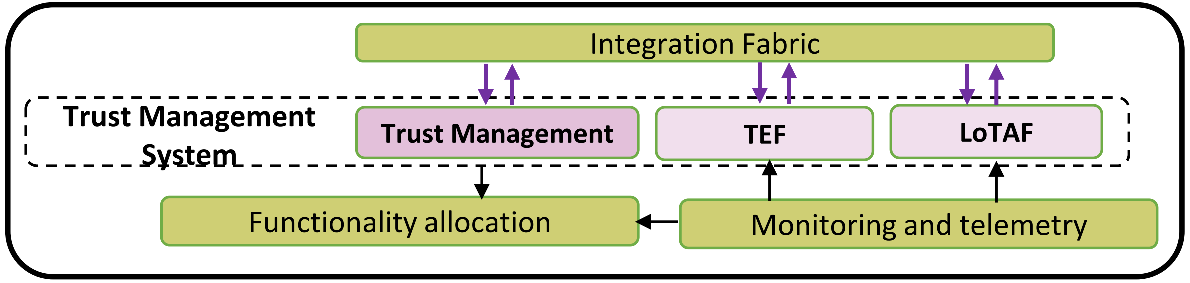


Schema of federated orchestration [HEX224-D63]

Hexa-X-II M&O Framework - Overall Functionalities
Trust Management


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
7
Real-time zero-touch control loops automation and coordination
2
5
3
6
8
Management capabilities exposure
10
SLA-driven Federated Orchestration
2
Trust Management
1
4
2
5
8

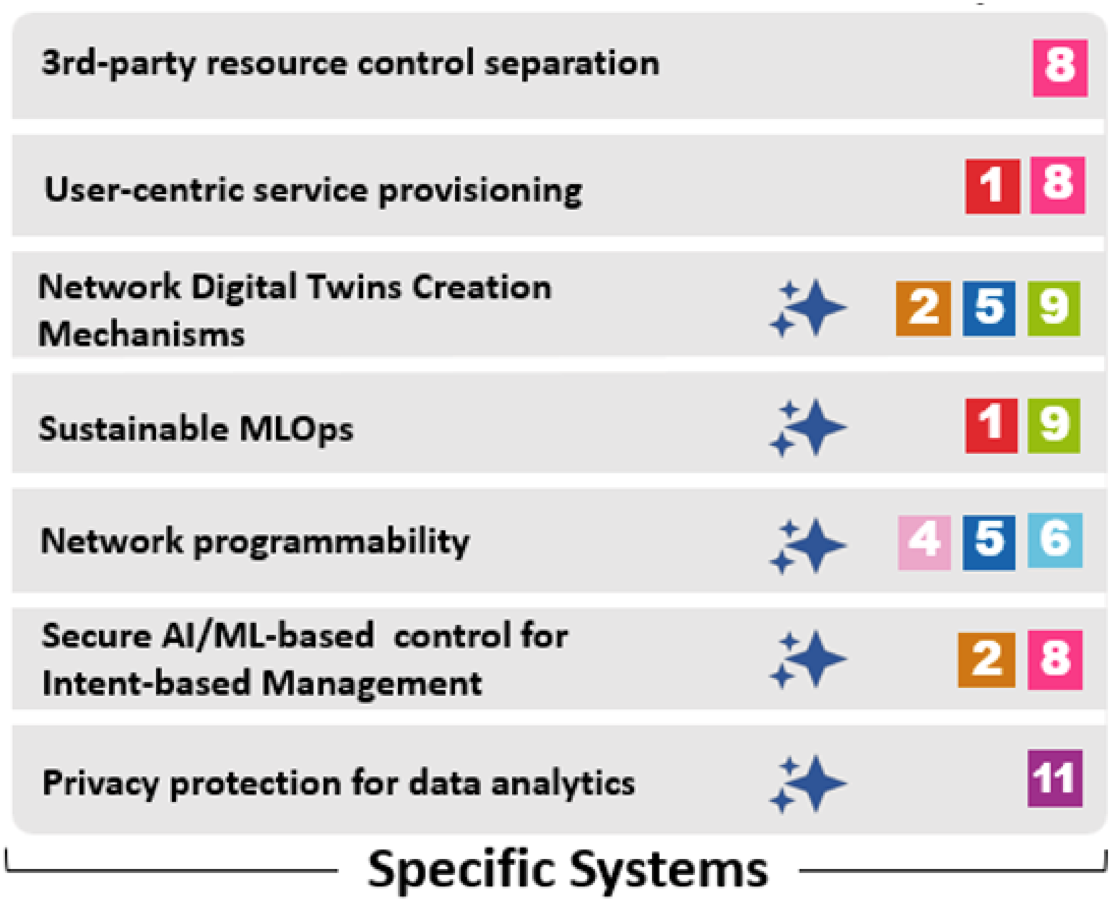
- **Trust Management** enables reliable E2E connections for multi-stakeholder and multi-domain scenarios on the **whole network continuum**. It consists of two main functions:
  - **Trust Evaluation Function** (**TEF**), which enhances **trust assessment** results in near real-time through advanced monitoring mechanisms and trust quantification, providing the trust indexes of **compute nodes** used by cloud orchestration engines for trustworthy **workload** placement.
  - **Level of Trust Assessment Function** (**LoTAF**), which presents an E2E intent-based trust management solution to **assess** and **ensure** the **trustworthiness** of network service or resource **provisioning**.



Trust management system architecture and interactions with other framework's components

- The enabler would be jointly used with the Monitoring and Telemetry enabler (to gather real-time performance parameters), the User-centric Service Provisioning enabler (to validate Trust Level Agreements - TLAs), the Management Capabilities Exposure enabler (to share trust indexes among sub-enablers), and the Functionality Allocation enabler (to provide trustworthiness and optimised workload placement).
- As a **main novelty** in literature [SCITT-09], **LoTAF** promotes the foundations of **transparent evidence service** concerning network services by verifying TLAs and transparently recording trust events from multiple issuers. **TEF** provides adaptive, **network-centric trust** from devices to the cloud, quantifying and verifying trust in real-time of hardware (HW), software (SW), and applications, considering the initial establishment of trust even at the low-level HW/SW elements.
- Trust Management contributes to the Hexa-X-II **design principle 6** (persistent security and privacy) by ensuring trustworthiness and reliable, secure connections across multi-stakeholders and multi-domain scenarios.

Hexa-X-II M&O Framework Specific Systems



3rd-party resource control separation 8

User-centric service provisioning 1 8

Network Digital Twins Creation Mechanisms ✦ 2 5 9

Sustainable MLOps ✦ 1 9

Network programmability ✦ 4 5 6

Secure AI/ML-based control for Intent-based Management ✦ 2 8

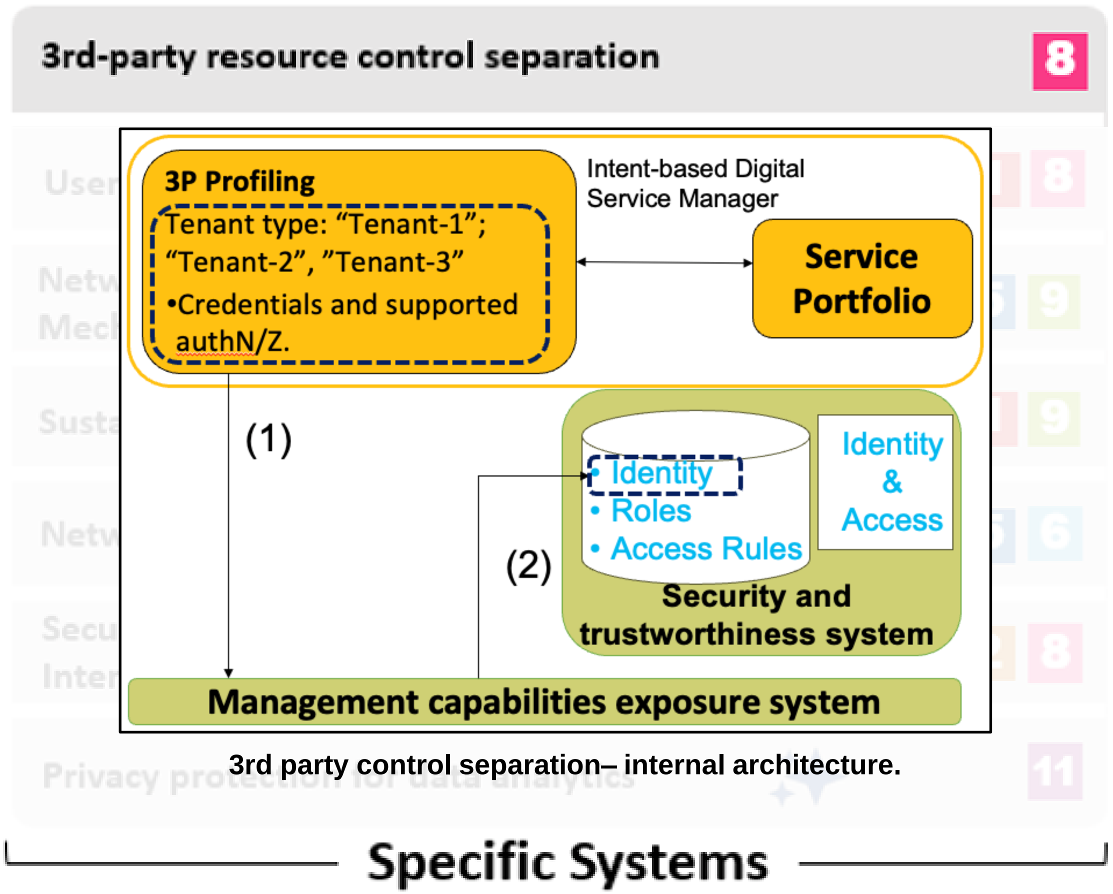
Privacy protection for data analytics ✦ 11

Specific Systems

This block groups the key systems operating at the stakeholder level in the management framework. The following technical enablers are included here:

- **The 3rd party resource control separation system**, which makes possible to establish separate M&O spaces per tenant in multi-tenant environments for secure and precise control over services, applications, and resources.
- **The User-centric service provisioning system**, which enables more dynamic and flexible SLA definitions.
- **The Network Digital Twins creation mechanisms**, to create virtual models of network environments for safe pre-production testing and real-time insights through connection with live systems.
- **A sustainable ML Operations (MLOps) asset**, to create automated AI/ML-based service workflows, enabling also the monitoring of the energy consumption in the different stages of those workflows.
- **Network programmability enabler**, which integrates the Software Defined Networks (SDN) technology for the future 6G networks with a cloud-native model and offering new interfaces for emerging devices.
- **Secure AI/ML-based Control** to support intent-based management systems (addressed in the context of WP2) to enhance their security.
- **A privacy-preserving data analytics enabler**, to ensures sensitive data protection during AI/ML processes, and enabling privacy-preserving analytics.

Hexa-X-II M&O Framework - Specific Systems
3rd party resource control separation

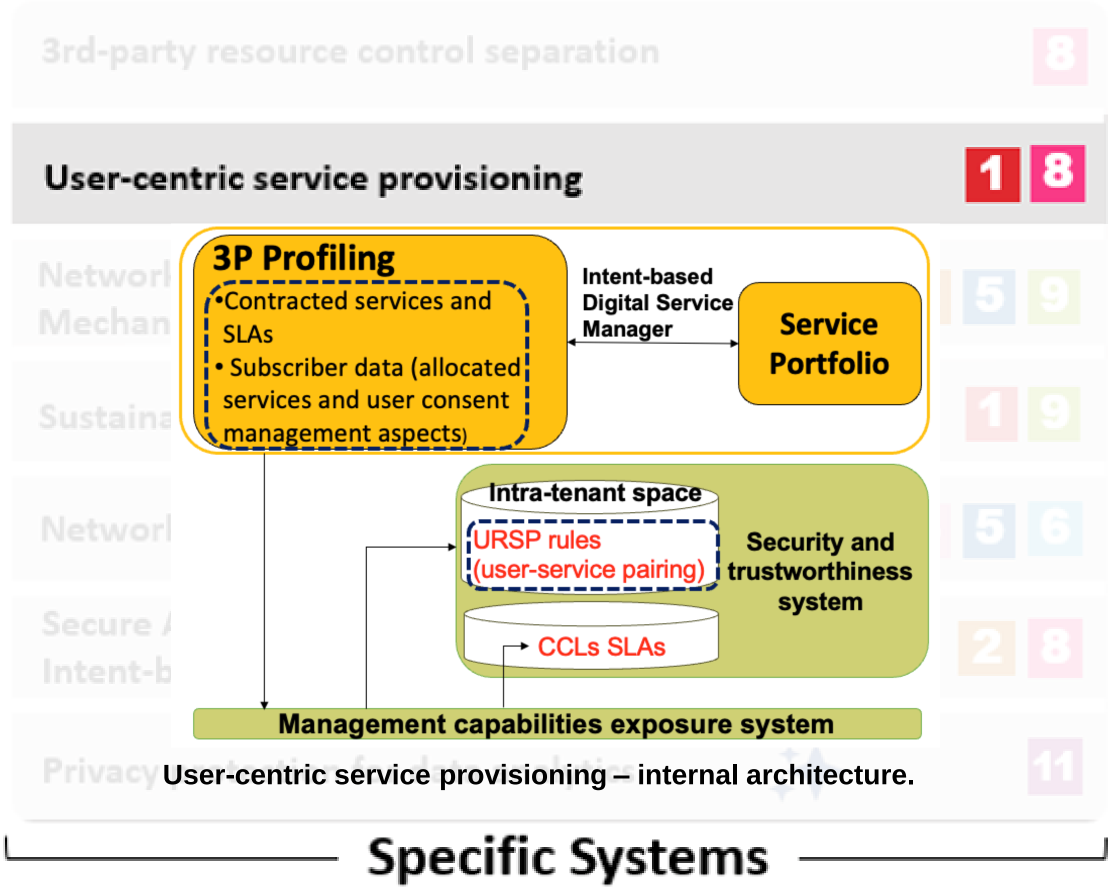


3rd-party resource control separation
8
User
8
Netw
Mech
9
Susta
9
Netw
6
Secu
Inter
8
Privacy protection for data analytics
11
3P Profiling
Tenant type: "Tenant-1"; "Tenant-2", "Tenant-3"
•Credentials and supported authN/Z.
Intent-based Digital Service Manager
Service Portfolio
(1)
(2)
• Identity
• Roles
• Access Rules
Identity & Access
Security and trustworthiness system
Management capabilities exposure system
Specific Systems

3rd party control separation– internal architecture.

- Provides **segregated management spaces** for tenants in multi-tenant 6G environments, ensuring **secure resource control** and **privacy** through dynamic, model-driven access control.
- Enables **granular resource control** and **conflict management** for tenant-specific resource sharing, positioned within the **Specific Systems Block** for handling secure resource allocation and third-party interactions.
- Targets **specific stakeholders** across the network continuum, ensuring secure, conflict-free resource control.
- Can work with the following other enablers in the framework:
  - The Management Capabilities Exposure for APIs exposure,
  - The User-Centric Provisioning system, and
  - The Trust Management functionality for trust monitoring.
- Beyond the state-of-the-art regarding what is offered by other common approaches, such as Role-based Access Control (**RBAC**), with **dynamic, model-driven permissions** tailored to tenants, crucial for the 6G multi-tenant setup.
- Supports **principles 5 (Resilience and availability), 6 (Persistent security and privacy) and 8 (Separation of concerns of network functions)** with fine-grained access control, ensuring safe multi-tenant operations.

Hexa-X-II M&O Framework- Specific Systems
User-centric service provisioning system


3rd-party resource control separation
8
User-centric service provisioning
1
8
Networ
Mechan
5
9
Sustaina
1
9
Networ
5
6
Secure /
Intent-b
2
8
Privacy
11
3P Profiling
•Contracted services and SLAs
• Subscriber data (allocated services and user consent management aspects)
Intent-based Digital Service Manager
Service Portfolio
Intra-tenant space
URSP rules (user-service pairing)
CCLs SLAs
Security and trustworthiness system
Management capabilities exposure system
Specific Systems

User-centric service provisioning – internal architecture.

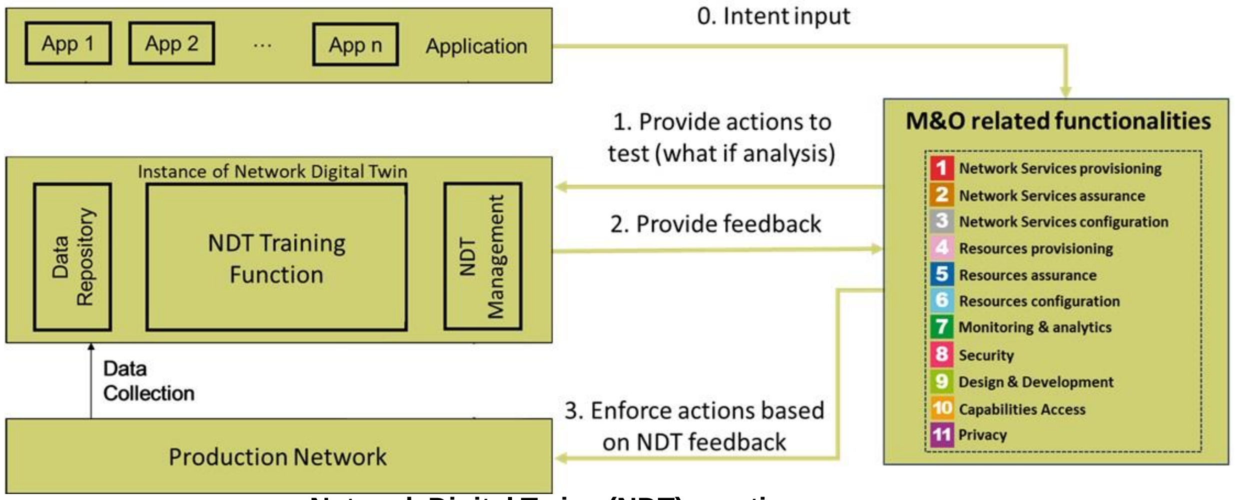
- Enables **personalised service provisioning** for tenants based on specific **SLAs**, ensuring optimal Quality of Experience (QoE) and secure service access.
- Provides **customised service policies** and **closed-loop automation** to monitor SLAs for **tenant-specific service management** and SLA enforcement.
- Applies to **specific stakeholders**, tailoring service provisioning to meet the dynamic needs of **individual tenant users** across the network.
- It can integrate with other enablers in the management framework, e.g.:
  - Can work with the resource controllability for 3rd parties enabler to ensure tenant-specific access control,
  - The Trust Management functionality, to handle trust validation during service provisioning.
  - The Management Capabilities Exposure (MCE) for secure API exposure
  - The Federated Orchestration system for SLA enforcement in federated multi-domain environments.
- Innovates beyond state-of-the-art models by using dynamic URSPs (User Equipment Route Selection Policies) for personalised service activation, with closed-loop automation ensuring SLA compliance.
- Supports WP2 design principles 3 (Flexibility to different network scenarios) and 6 (Persistent security and privacy) by customising services for diverse users, and ensuring secure access and SLA assurance.

Hexa-X-II M&O Framework - Specific Systems

Network Digital Twins Creation Mechanisms

- 3rd-party resource control separation 8
- User-centric service provisioning 1 8
- **Network Digital Twins Creation Mechanisms** 2 5 9
- Sustainable MLOps 1 9
- Network programmability 4 5 6
- Secure AI/ML-based control for Intent-based Management 2 8
- Privacy protection for data analytics 11

Specific Systems



App 1 · App 2 · ··· · App n · Application

0. Intent input

1. Provide actions to test (what if analysis)

2. Provide feedback

3. Enforce actions based on NDT feedback

Instance of Network Digital Twin: Data Repository · NDT Training Function · NDT Management

Data Collection

Production Network

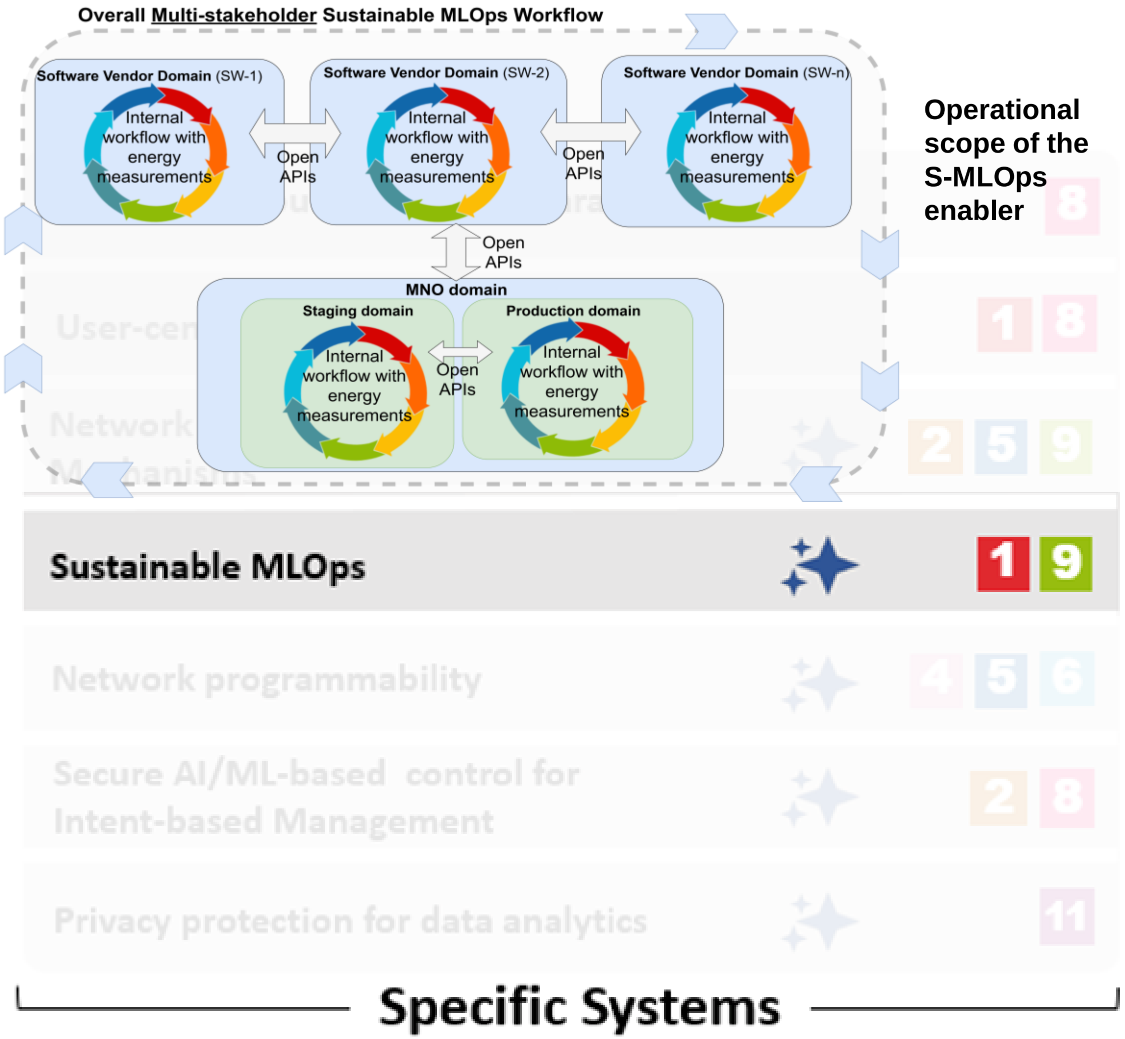
M&O related functionalities

1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy

Network Digital Twins (NDT) creation process

- Network Digital Twins (DTs) are a tool that enables safe orchestration and control by generating means for “what-if” scenarios testing or training AI/ML management algorithms as illustrated in the figure above.
- Graph Attention (GATs) networks, trained by the NDT Training Function, is one such mechanism to create the Network Digital Twins (NDTs) to make them apt representatives of the complex cloud-native network envisioned towards 6G.
- GATs allow the NDT to efficiently model the network.
- This enabler would be jointly used with the Monitoring and Telemetry enabler to collect data from the production network.
- It can potentially support all WP1 use cases.
- Related to the targeted WP2 design principles - full automation and optimisation, and providing support for the automation of control loops.

Hexa-X-II M&O Framework - Specific Systems
Sustainable MLOps


Overall Multi-stakeholder Sustainable MLOps Workflow
Software Vendor Domain (SW-1)
Software Vendor Domain (SW-2)
Software Vendor Domain (SW-n)
Internal workflow with energy measurements
Open APIs
MNO domain
Staging domain
Production domain
User-cen
Network
Mechanisms
Sustainable MLOps
1
9
Network programmability
4
5
6
Secure AI/ML-based control for Intent-based Management
2
8
Privacy protection for data analytics
11
Specific Systems

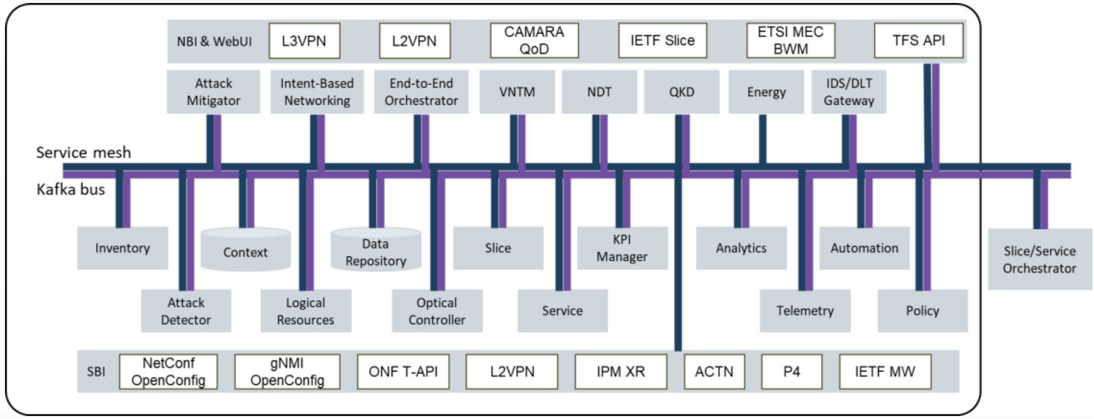
Operational scope of the S-MLOps enabler

The Sustainable MLOps (S-MLOps) enabler automates, monitors, and optimises the workflows involved in the development, deployment, and operation of AI/ML-based network services, also considering the measurement of energy consumption.

Main Features:

- Supports the design, development, and provisioning workflows of AI/ML models or AI/ML-based network services, considering also the energy consumption associated to the different stages in these workflows.
- Could be jointly used with enablers that provide enhanced functionality in an MLOps environment, such as the Network Digital Twins Creations Mechanisms or the Privacy Protection for Data Analytics enablers.
- The main contribution beyond the state-of-the-art regarding the regular MLOps approaches is the possibility to work in line with the multi-stakeholder service development ecosystem in the telecommunications industry, and also, including energy consumption related measurements in the different stages of the MLOps workflows to contribute to the use of AI/ML techniques in a more sustainable way.
- The enabler can support all the WP1 use cases, as all of them could rely on AI/ML techniques.
- Targets the following WP2 design principles: flexibility to different network scenarios (3), and minimising environmental footprint and enabling sustainable networks (10).

Hexa-X-II M&O Framework - Specific Systems
Network Programmability System



Network programmability framework architecture

Sustainable MLOps 1 9

Network programmability 4 5 6

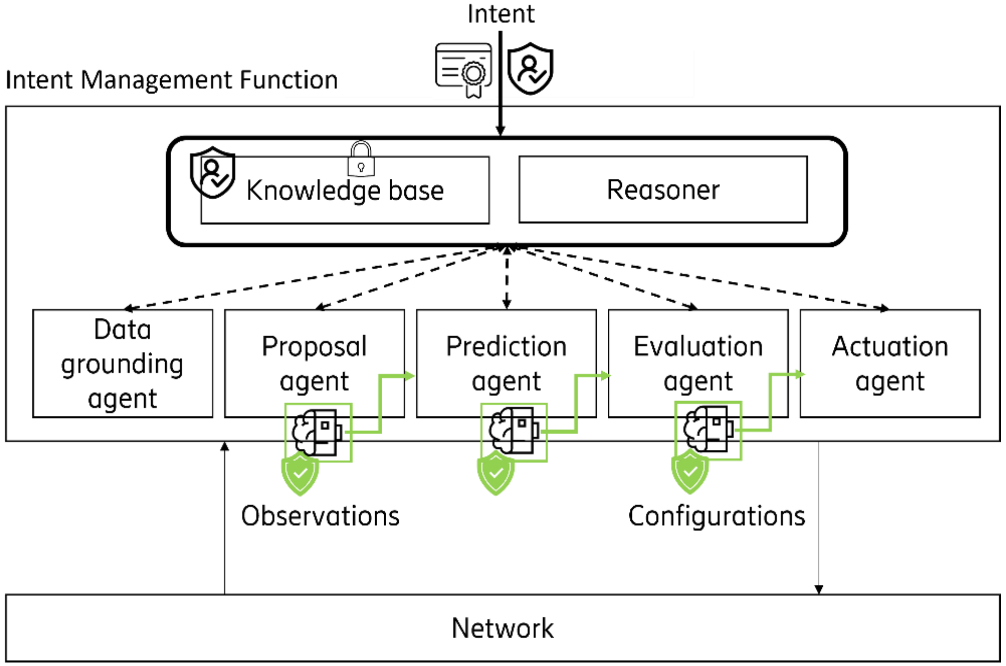
Secure AI/ML-based control for Intent-based Management 2 8

Privacy protection for data analytics 11

Specific Systems

- Enables a scalable and flexible network control and management solution based on SDN, APIs, and cloud platforms. It features an E2E SDN Orchestrator as a so-called *parent* controller, and technological domain SDN controllers (IP, Optical, TSN/DetNet) as child controllers within a single administrative domain.
- Provides the necessary elements to provision connectivity services, ensure a certain degree of QoS for the requested services and configure the underlying network elements to provide those services. The Network Programmability Enabler integrates the SDN technology into the framework. Beyond the traditional benefits of SDN, it aligns with the cloud-native model and cloud continuum concept, while also providing new interfaces for emerging devices. This system is primarily based on the ETSI-hosted TeraFlowSDN project [TFS].
- Beyond the SotA features: SmartNIC Transceiver Support; Time Sensitive Networking (TSN) and Deterministic Networking (DetNet); Automated Transport Network Re-Configuration; MEC BandWidth Management (BWM) Service for Gaming; Integration with Standards and Frameworks.
- Relation to the WP2 design principles: Contributes to full automation and optimisation (2), flexibility to different network scenarios (3), network scalability (4), resilience and availability (5), and the optimisation of internal interfaces (7) in a cloud native way.
- Relationship with other enablers in the WP6 Framework: it collaborates with Monitoring and Telemetry to provide the necessary tools to monitor current networks state and take AI-based decisions on the offered services that might re-configure network elements.

Hexa-X-II M&O Framework - Specific Systems
Secure AI/ML-based control for Intent-based Management



Secure AI/ML-based control overview

Secure AI/ML-based control for Intent-based Management 2 8

Privacy protection for data analytics 11

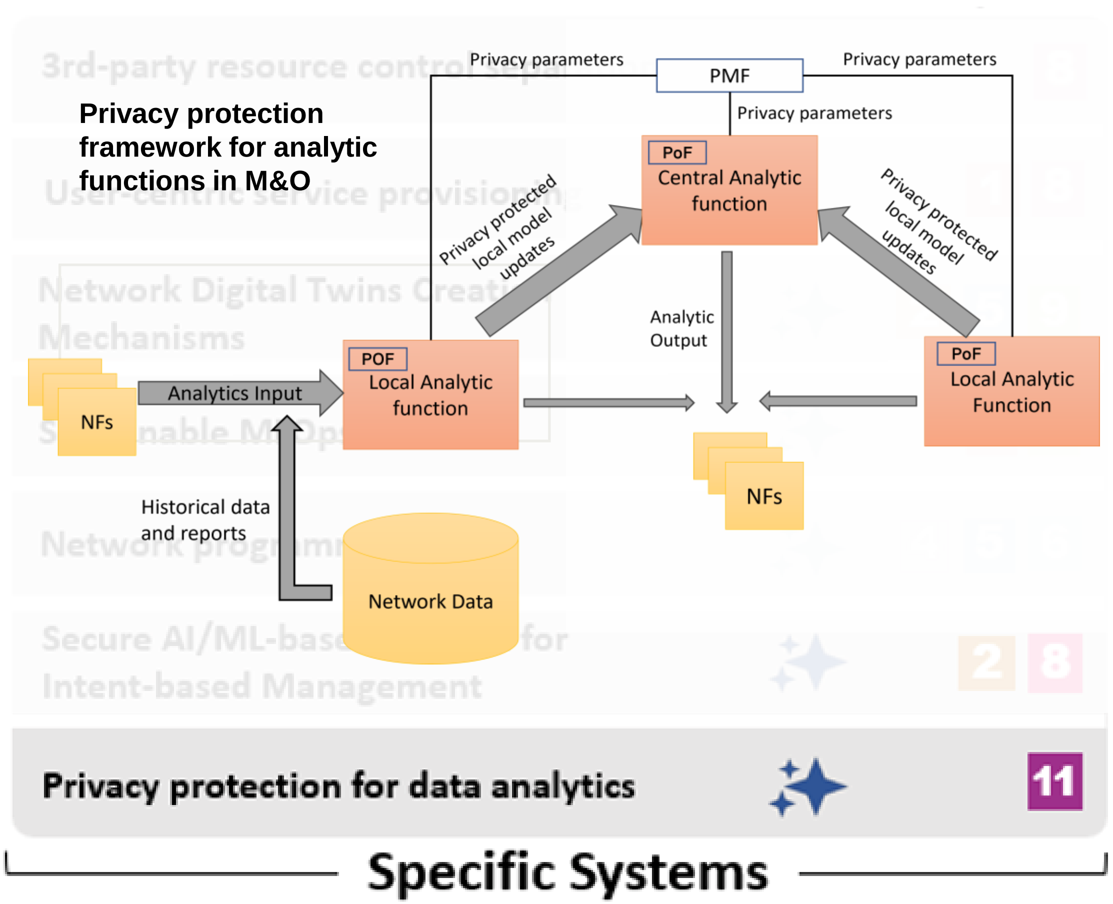
Specific Systems

The AI/ML-based control for the intent-based management algorithm ensures that decisions inferred by AI/ML models deployed on agents are resistant to adversarial attacks and are not manipulated or compromised.

Main features:

- The enabler provides network service assurance and security functionalities to the WP6 smart management framework. It contributes to the Hexa-X-II design principle 6 (persistent security and privacy).
- The secure AI/ML-based control is an algorithm that incorporates adversarial examples during the training process to improve model robustness against attacks. These adversarial examples can be generated through other different algorithms.
- This enabler could be used with the ML-based Configuration Recommender for Energy Saving enabler to provide more robust configuration recommendations for power consumption reductions. Also, it would be jointly used with the Monitoring and Telemetry enabler to collect data.
- This enabler is considered to contribute to the effectiveness and resilience of next-generation networks by facilitating real-time and automated network management with enhanced security.
- The enabler can support WP1 use cases where AI/ML has a main role such as E-health for all, immersive smart cities, autonomous supply chains and cooperating mobile robots.

Hexa-X-II M&O Framework - Specific Systems
Privacy protection for data analytics


Privacy protection framework for analytic functions in M&O
3rd-party resource control
User-centric service provisioning
Network Digital Twins Creation Mechanisms
Network programmability
Secure AI/ML-based for Intent-based Management
2
8
PMF
Privacy parameters
Privacy parameters
Privacy parameters
PoF
Central Analytic function
Privacy protected local model updates
Privacy protected local model updates
Analytic Output
POF
Local Analytic function
PoF
Local Analytic Function
NFs
Analytics Input
NFs
Historical data and reports
Network Data
Privacy protection for data analytics
11
Specific Systems

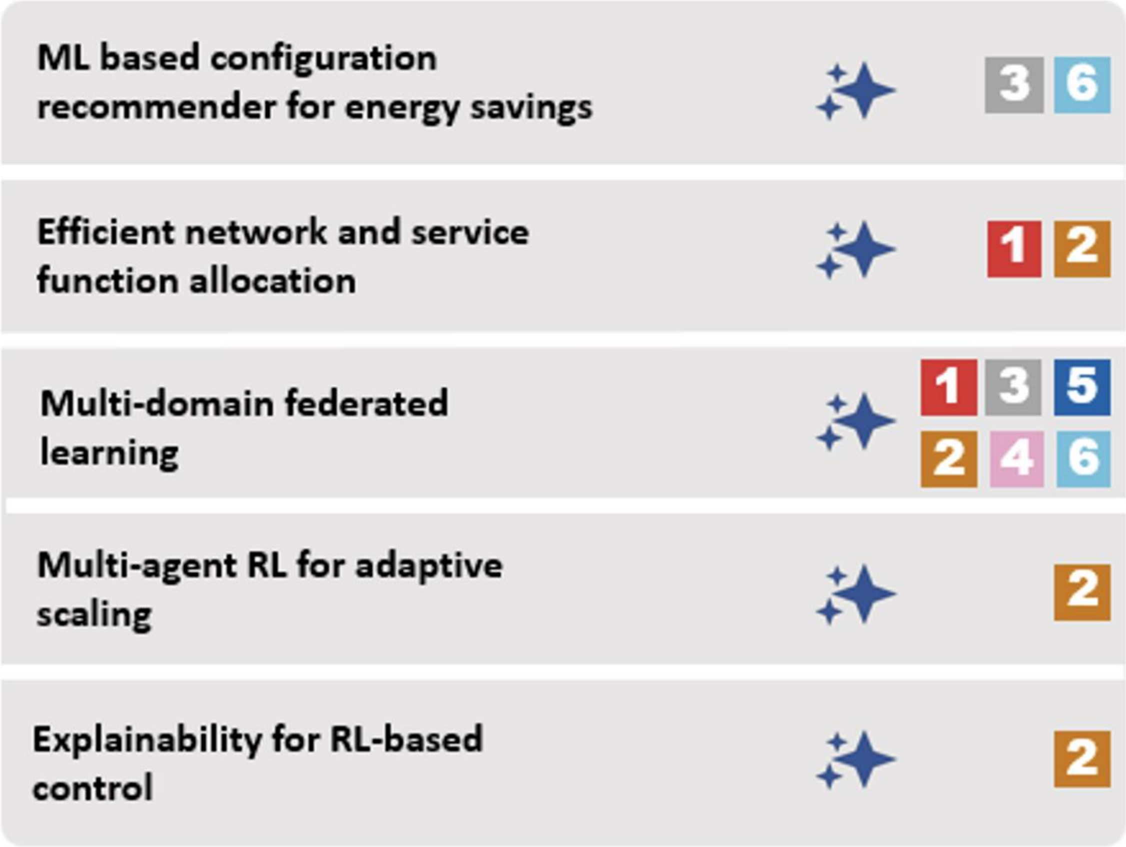
The Privacy Protection System for data analytics in M&O ensures that sensitive data is protected while still enabling effective analysis and decision-making. A privacy management function (PMF) and a privacy operation function (POF) are proposed. The POF is responsible for making data privacy-preserving using policies created by the PMF, which is responsible for selecting the privacy operation to be used by the POF and the generation and distribution of key pairs for privacy operations, which take place at local and central analytic functions in a federated learning setting. The main features of the system are.

- Addresses the WP2 design principle 6, regarding persistent security and privacy.
- By proposing a new privacy-enhancing framework for data analytics in M&O, ensures that data is managed in a private way for the cases that privacy aware data analytics is required. In this way it provides the privacy functionalities to the Smart Management Framework.
- This enabler is considered as a specific system which provides privacy protection against data leakage in M&O using privacy-enhancing technologies, specifically in the multi-vendor environments of 6G.
- This enabler could be jointly used with the Monitoring and Telemetry enabler to collect data from the network to be used for analytics.
- The enabler is independent from the use cases. It is related to privacy requirements in data analytics in the M&O, and can support WP1 use cases that has a close relation with data analytics in the M&O.

Hexa-X-II M&O Framework Algorithms

This section of the Framework includes a set of selected algorithms for future 6G systems. It emphasises the importance of incorporating certain algorithms into the management framework with a primary focus on AI/ML algorithms. The selected algorithms are the following:

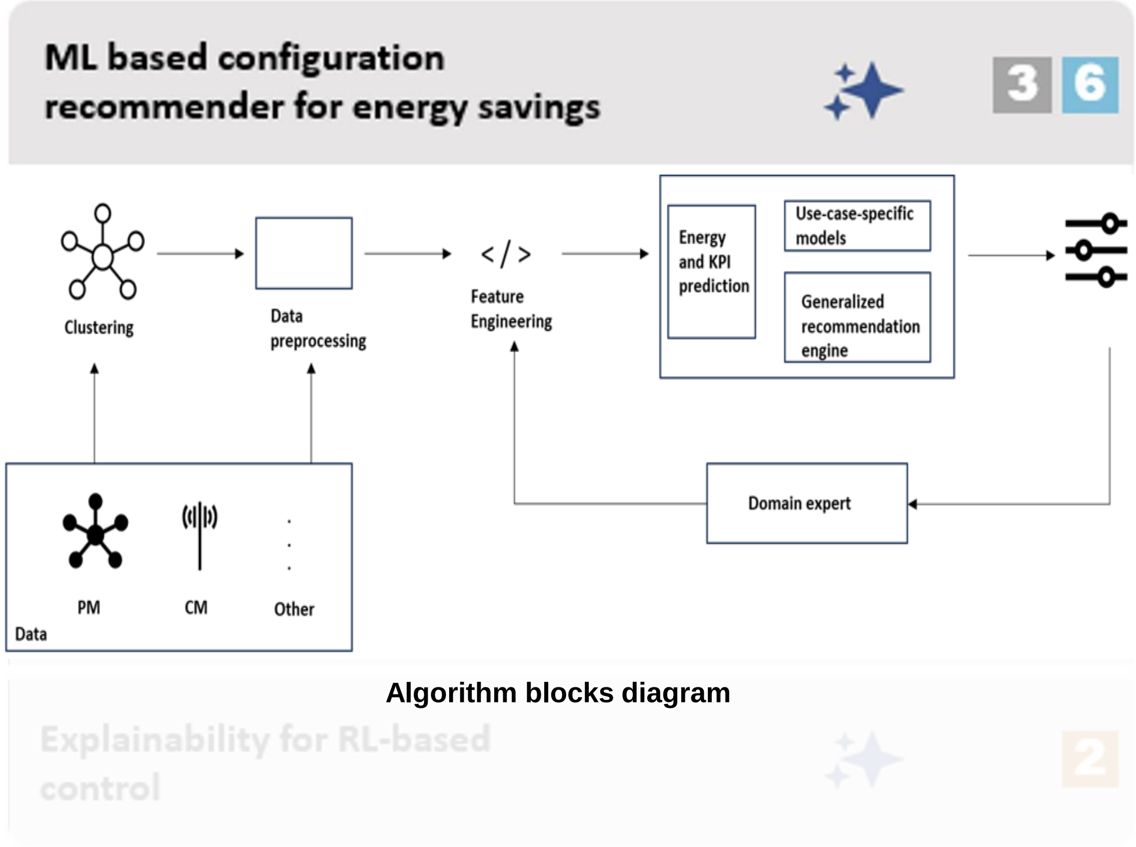
- **ML-based Configuration Recommender**, to optimises energy efficiency in 6G base stations by reducing power consumption.
- **Algorithms for Efficient Network and Service Function Allocation**, focusing on energy-efficient network function allocation. Suitable for resource-limited edge devices.
- **Multi-domain federated Learning Algorithm**, which optimises compute resource allocation for federated learning with minimal impact on energy consumption and other network services.
- **Multi-Agent RL Algorithm**, which enables adaptive resource scaling to meet service-level objectives, such as latency and energy efficiency, in dynamic network conditions.
- **Explainability Algorithms for RL Control**: Provides human-understandable explanations for decisions made by RL-based control algorithms, enhancing transparency and trust.



ML based configuration recommender for energy savings	3 6
Efficient network and service function allocation	1 2
Multi-domain federated learning	1 3 5 2 4 6
Multi-agent RL for adaptive scaling	2
Explainability for RL-based control	2

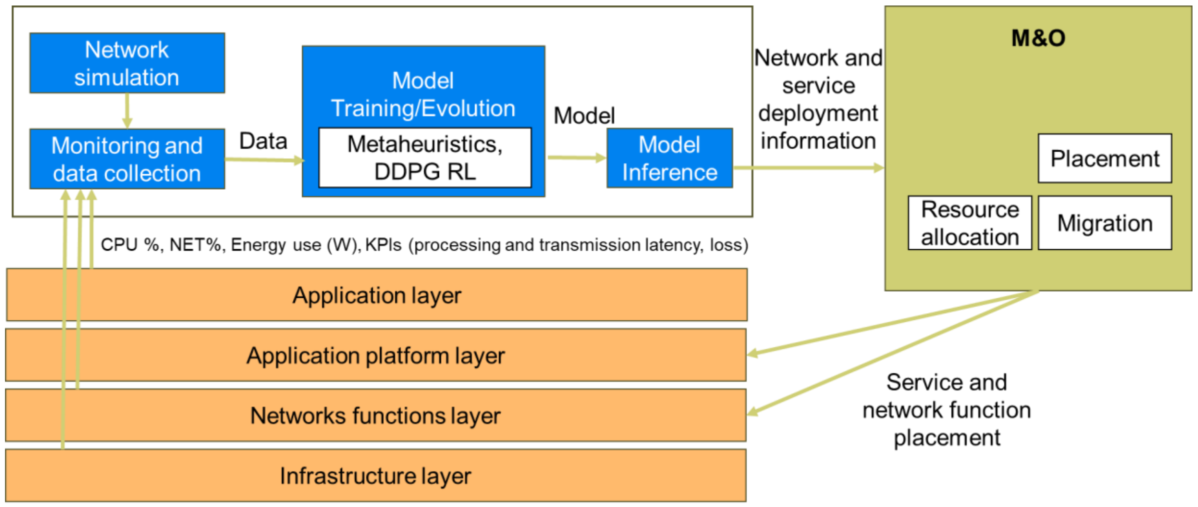
Hexa-X-II M&O Framework - Algorithms
ML-based configuration recommender for energy savings

- This enabler aims at providing a solution for the energy reduction of the future 6G networks with the aid of AI/ML and without degrading the network performance.
- The enabler provides a generic structure where different types of algorithms (e.g., AI-based or rule-based algorithms) targeting this objective can be deployed and implemented.
- It contributes to energy savings by targeting both network services and network resources configuration.
- This enabler can be in interaction with the monitoring entity and also can utilise the closed-loop operations further automation. It can also work in interaction with the Monitoring and Telemetry enabler.
- Instead of static and manual operation, which can result in a poor performance in dynamic network conditions, with the aid of AI/ML more energy can be saved, and more robustness can be achieved under harsh network conditions. In this use-case, base-stations are put into sleep mode to save energy depending on the predicted traffic.
- The enabler aligns with the principle of minimising the environmental footprint and enabling sustainable networks (Principle 10 under WP2). The most relevant related value is sustainability, since the algorithms are designed to save energy in different parts of the network.


ML based configuration recommender for energy savings
3
6
Clustering
Data preprocessing
Feature Engineering
Energy and KPI prediction
Use-case-specific models
Generalized recommendation engine
Domain expert
PM
CM
Other
Data
Explainability for RL-based control
2

Algorithm blocks diagram

Hexa-X-II M&O Framework - Algorithms
Efficient network and service function allocation



Network simulation
Monitoring and data collection
Data
Model Training/Evolution
Metaheuristics, DDPG RL
Model
Model Inference
Network and service deployment information
M&O
Placement
Resource allocation
Migration
CPU %, NET%, Energy use (W), KPIs (processing and transmission latency, loss)
Application layer
Application platform layer
Networks functions layer
Infrastructure layer
Service and network function placement

AI/ML-based energy efficient M&O

- Allows an intelligent, optimised, and energy-efficient M&O for 6G networks and services. AI/ML models and algorithms are used to tackle the complexity of service requirements and network dynamics.
- Provide orchestration actions (i.e., placement, resource allocation and migration) aiming to minimise energy or resource use, and matching the capabilities of the infrastructure nodes and devices.
- In this enabler, the network state is used as input to the decision algorithm, and to measure the effect of applying those actions. This enabler would be jointly used with the Monitoring and Telemetry enabler to gather the data.
- If harvested energy is available, the information is included in the optimisation process to encourage using greener energy.

ML based configuration recommender for energy savings 3 6

Efficient network and service function allocation 1 2

Multi-domain federated learning 1 3 5 2 4 6

Multi-agent RL for adaptive scaling 2

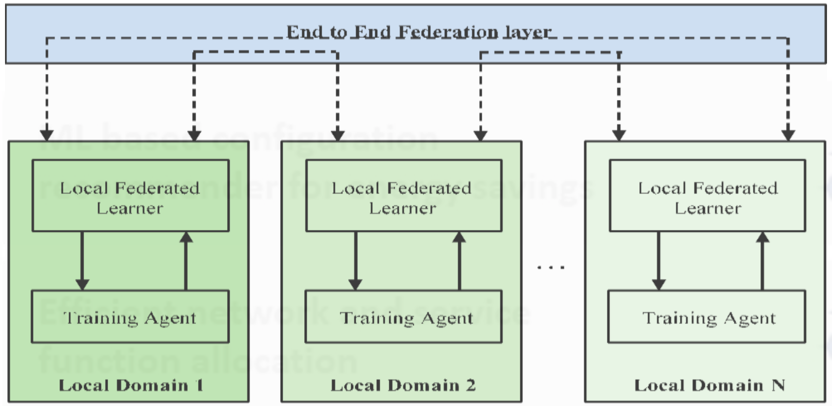
Explainability for RL-based control 2

Hexa-X-II M&O Framework - Algorithms
Multi-domain federated learning

This enabler provides a way for operators to leverage rich data sources to provision resources for secure decentralised AI training network services for verticals.

Main Features:

- Facilitates network service provisioning and compute resource assurance and configuration by directing the orchestration of resources to deliver on AI model training services.
- It provides a structured process to enable the network deliver novel AI model training services using network edge compute and storage resources.
- The enabler leverages the Monitoring and Telemetry enabler to manage the life cycle of the AI model training services.
- The enabler supports Compute as a Service use cases, projected to be a growth market in 6G given the large number of verticals envisaged.
- It is applicable to all WP1 Use Cases leveraging networking for AI.
- The enabler supports the following design principles defined in WP2:
  - Principle 1: Support and exposure of 6G services and capabilities, enabling Compute as a Service.
  - Principle 2: Full automation and optimisation, facilitating the automatic placement of service workloads.
  - Principle 3: Flexibility to different network scenarios, since the enabler is versatile in its approach to the verticals supported.



Schematic of multi-domain federated learning [HEX224-D63]

Multi-domain federated learning

1 3 5
2 4 6

Multi-agent RL for adaptive scaling

2

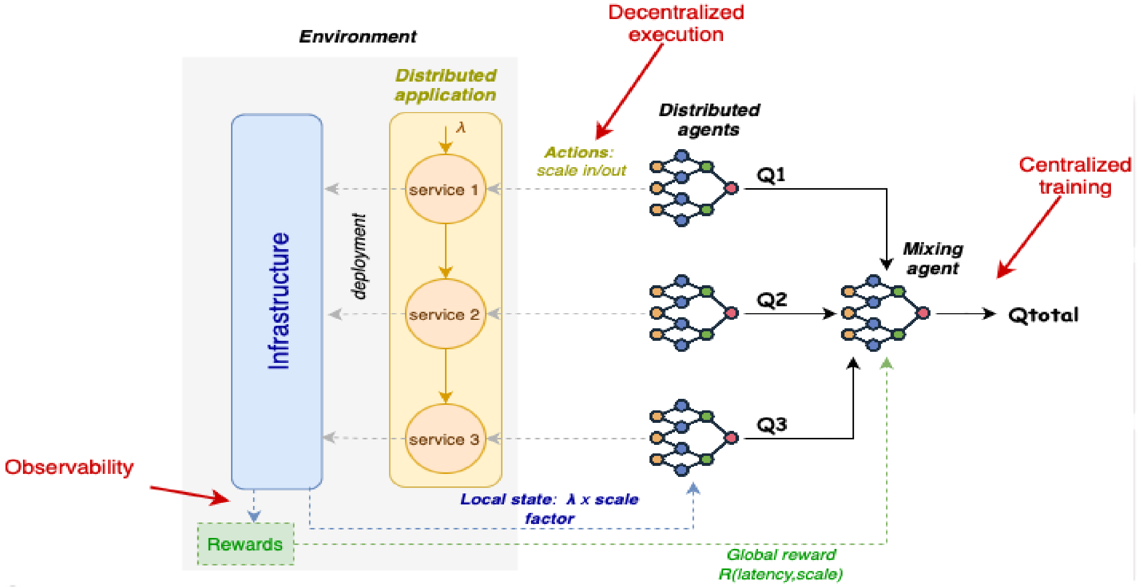
Explainability for RL-based control

2

Hexa-X-II M&O Framework - Algorithms
Multi-agent RL for adaptive scaling

- The enabler provides a multi-agent system (MAS) approach for service autoscaling and migration across the computing continuum. It improves automation in scaling actions while respecting SLAs.
- This specific algorithm follows the approach of the Multi-agent system for multi-cluster orchestration Overall M&O Solution by assigning agents to interconnected services
- Decisions are taken based on analysing service performance in terms of latency and resource consumption values supporting variable workloads.
- Focuses on service assurance for seamless operation.
- The input to this enabler would be provided by Monitoring and Telemetry enabler. Based on such, this enabler would generate actions such as service horizontal autoscaling and migration across infrastructure clusters.
- Enables collaboration among services with separate local environments but common global objectives.
- The enabler targets the following Hexa-X-II design principles defined in WP2:
    - Full automation and optimisation (principle 2).
    - Internal interfaces are cloud-optimised (principle 7).
    - Minimising environmental footprint and enabling sustainable networks (principle 10).

Multi-agent setting for autoscaling



Multi-agent RL for adaptive scaling ✦ 2

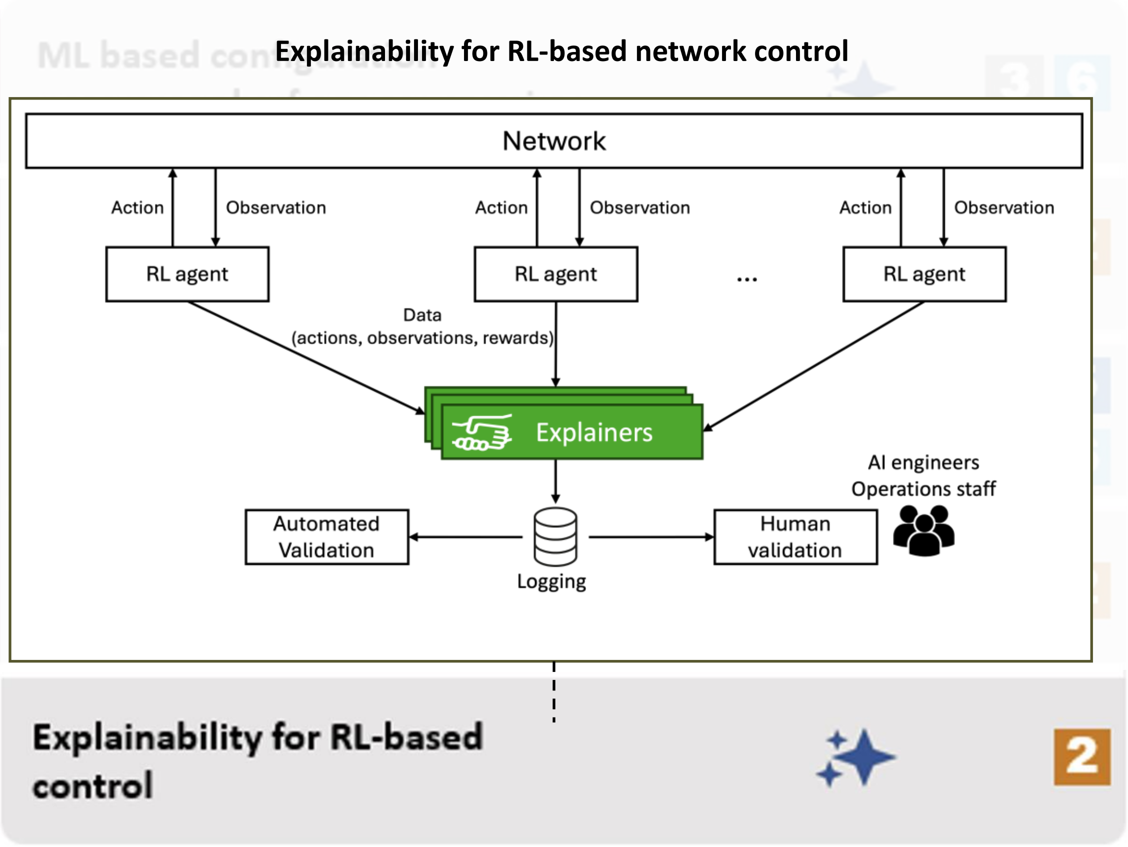
Explainability for RL-based control ✦ 2

Hexa-X-II M&O Framework - Algorithms
Explainability for RL-based control

This enabler provides human-interpretable explanations for decisions made by Reinforcement Learning (RL) AI models to enhance their trustworthiness.

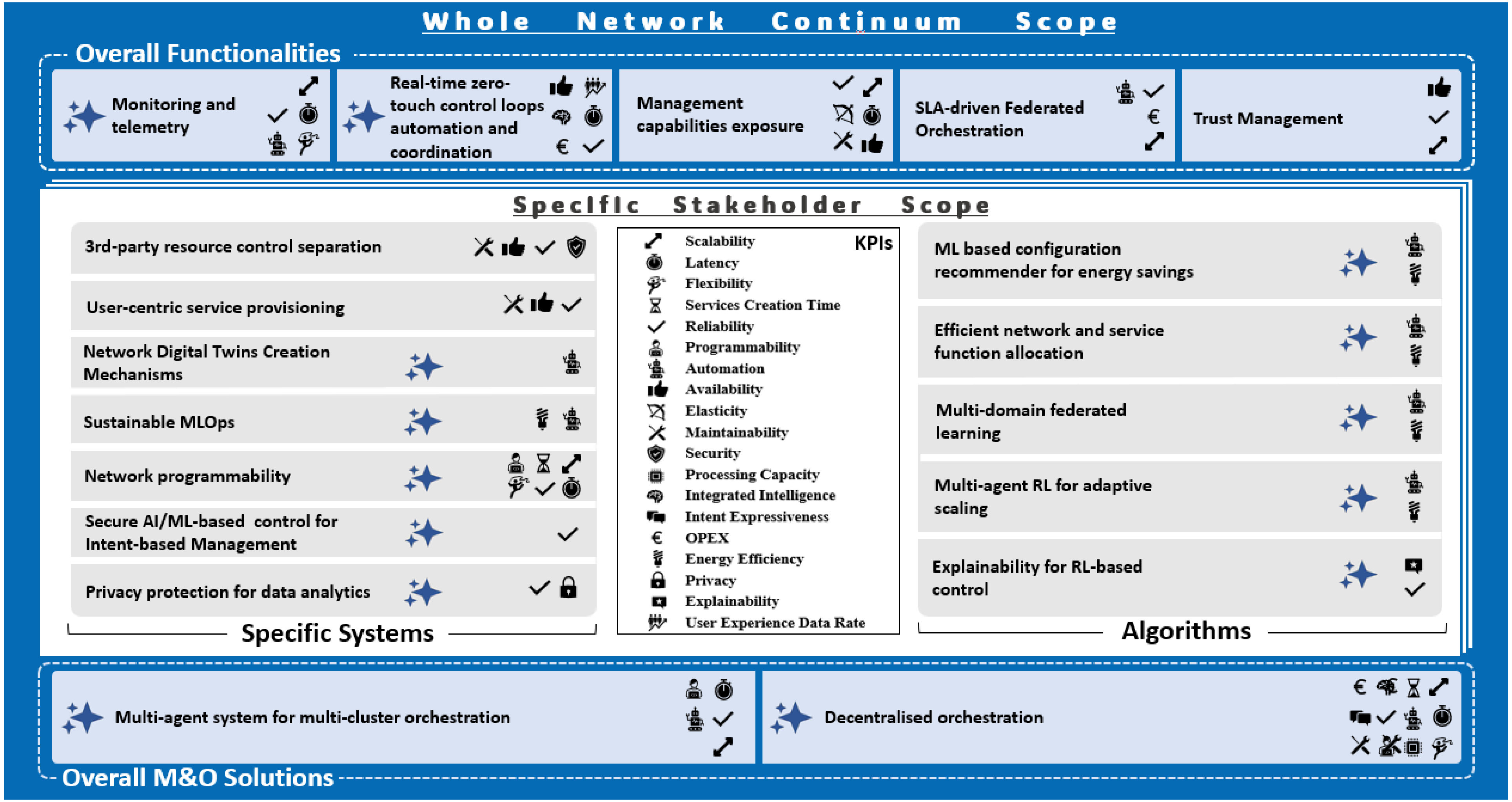
Main features:

- The explanations can be validated through automated procedures and/or human inspection – e.g., by AI engineers or the operations staff – to help detect anomalies, such as adversarial attacks, and contribute to network services assurance.
- The solution relies on eXplainable AI (XAI) algorithms to analyse the RL models and produce explanations.
- This enabler could be used jointly with other RL-based enablers (multi-agent RL for adaptive scaling and energy efficient service function allocation) and the Secure AI/ML-based Control for Intent-based Management enabler.
- This enabler will enhance the transparency, accountability, and security of AI-based operations.
- It is considered to support those WP1 use cases involving service provisioning through an AI-based network optimisation – e.g., cooperating mobile robots or network-assisted mobility, among others.



Related KPIs

The following figure shows the mapping between the impacted KPIs identified in the previous Deliverable D6.3 (listed in the middle of the figure), and the different assets of the management framework:



Whole Network Continuum Scope

Overall Functionalities

- Monitoring and telemetry
- Real-time zero-touch control loops automation and coordination
- Management capabilities exposure
- SLA-driven Federated Orchestration
- Trust Management

Specific Stakeholder Scope

Specific Systems

- 3rd-party resource control separation
- User-centric service provisioning
- Network Digital Twins Creation Mechanisms
- Sustainable MLOps
- Network programmability
- Secure AI/ML-based control for Intent-based Management
- Privacy protection for data analytics

KPIs

- Scalability
- Latency
- Flexibility
- Services Creation Time
- ✓ Reliability
- Programmability
- Automation
- Availability
- Elasticity
- Maintainability
- Security
- Processing Capacity
- Integrated Intelligence
- Intent Expressiveness
- € OPEX
- Energy Efficiency
- Privacy
- Explainability
- User Experience Data Rate

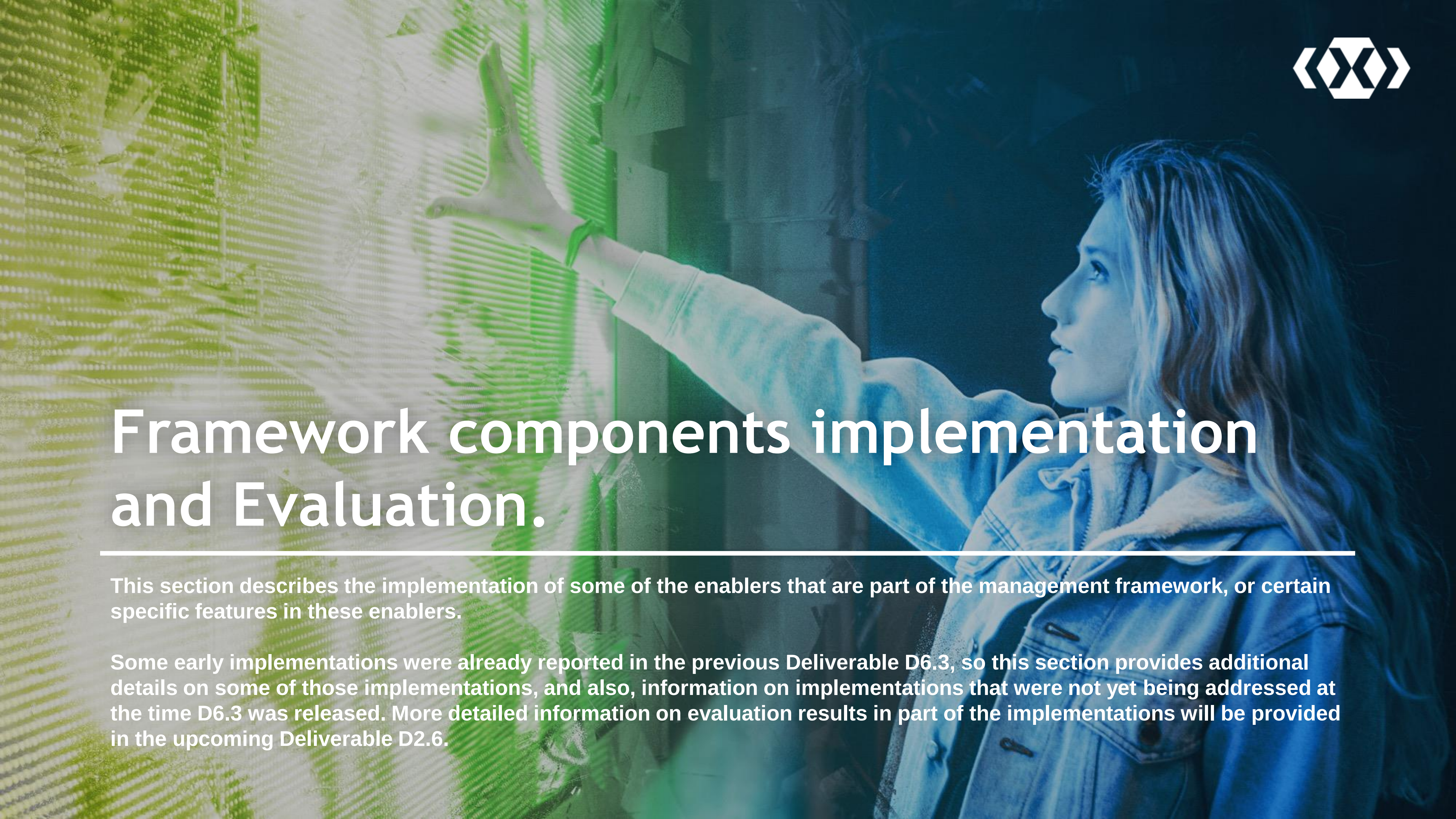
Algorithms

- ML based configuration recommender for energy savings
- Efficient network and service function allocation
- Multi-domain federated learning
- Multi-agent RL for adaptive scaling
- Explainability for RL-based control

Overall M&O Solutions

- Multi-agent system for multi-cluster orchestration
- Decentralised orchestration

Values of these KPIs can be found in the following slides regarding the practical implementations carried out during the reporting period of Deliverable D6.5, or will be included in the next D2.6 regarding the results in the final PoCs of the project being addressed in the context of WP2.



Framework components implementation and Evaluation.

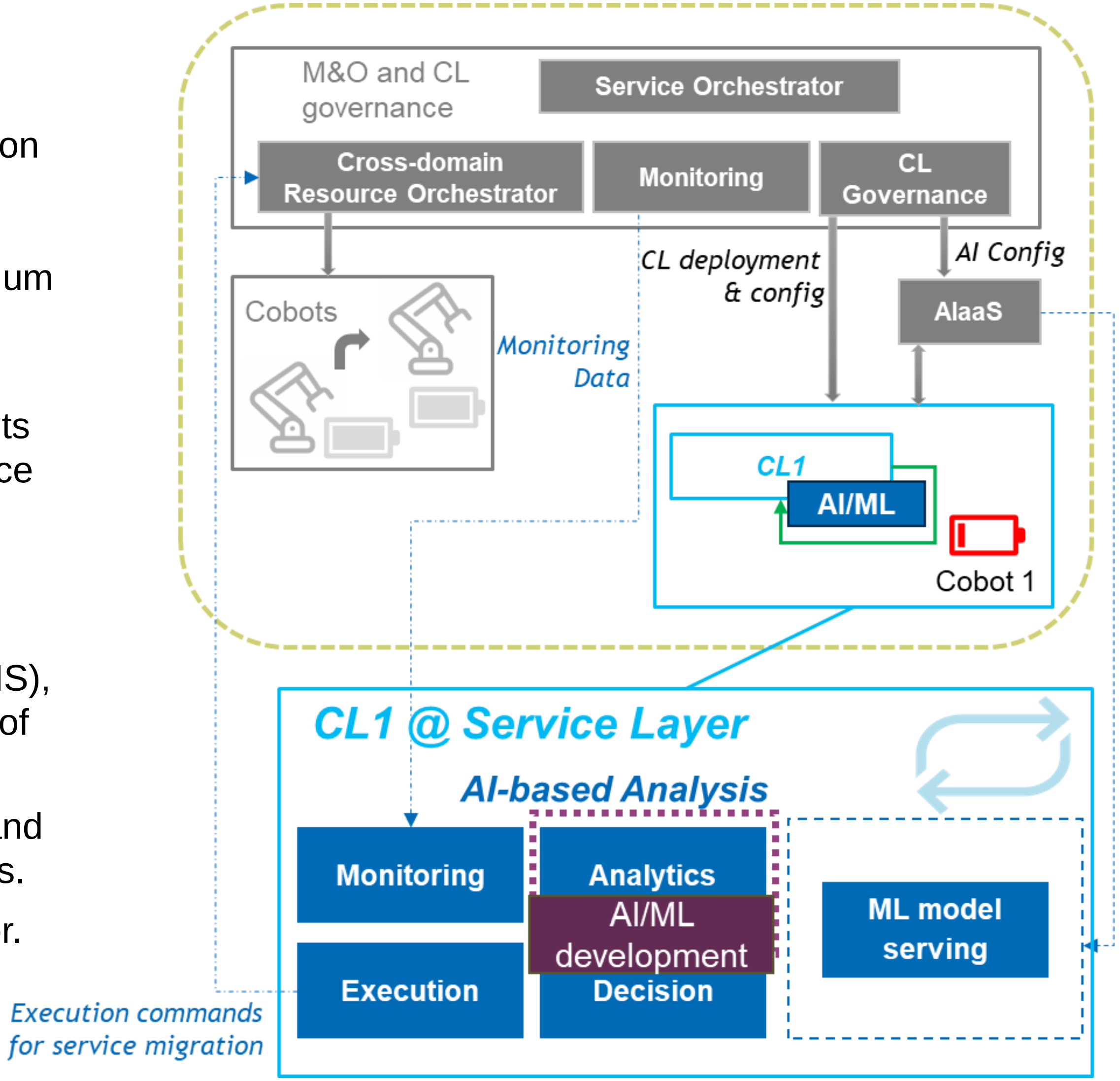
This section describes the implementation of some of the enablers that are part of the management framework, or certain specific features in these enablers.

Some early implementations were already reported in the previous Deliverable D6.3, so this section provides additional details on some of those implementations, and also, information on implementations that were not yet being addressed at the time D6.3 was released. More detailed information on evaluation results in part of the implementations will be provided in the upcoming Deliverable D2.6.

AI-enabled RT zero-touch control loop analysis function (1/2)

Implementation of an example of Real-time Zero-touch CLs automation, with focus on AI/ML-based Analysis function.

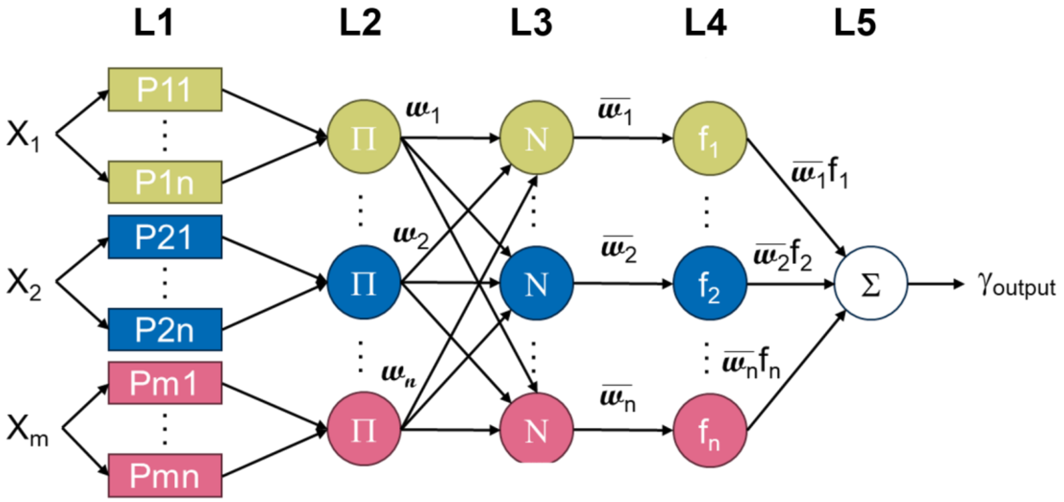
- Service-layer CL automating application services migration across edge continuum relying on an AI-enabled CL analysis that leverages AI-as-a-Service (AIaaS) to consume a model serving instance for continuous inference calls.
- Dynamic and optimal allocation of edge resources, migrating service components according to ever-changing resource status, with the objective of reducing service downtime.
- Trigger based on forecasted states of extreme edge nodes (cobots) and service requirements.
- Versatile ML model based on an Adaptive Neuro Fuzzy Inference System (ANFIS), trained using historical data from cobots. Experiments performed in the context of PoC C, on a video surveillance cobots use case.
- CL Governance used to deploy and configure CL functions in the edge nodes and coordinating the interfacing with AIaaS for ML model configuration in CL analysis.
- Service migration decisions executed interacting with the Resource Orchestrator.



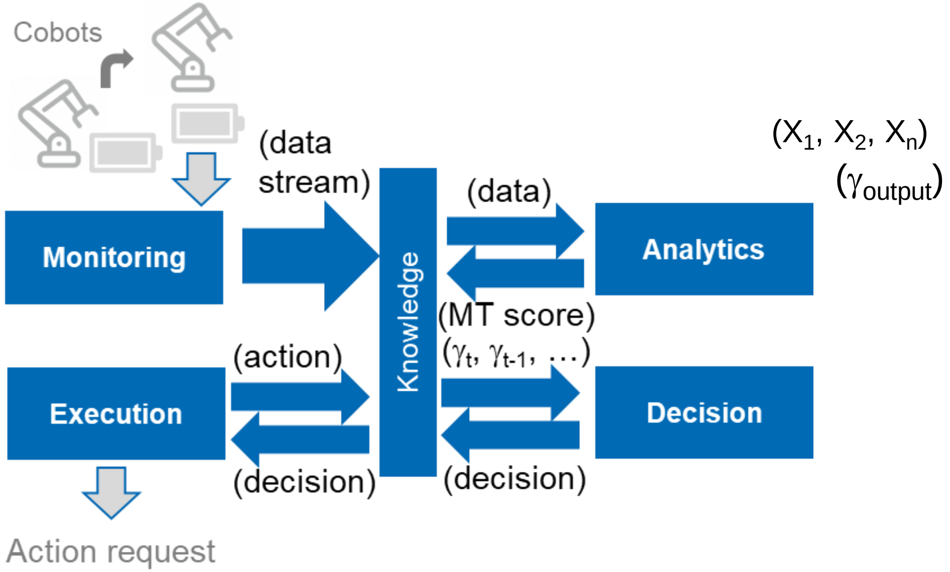
Real-time zero-touch CLs automation implementation for PoC B.1 using the AI-enabled analytics function

AI-enabled RT zero-touch control loop analysis function (2/2)

- CL Analysis based on ANFIS model to obtain insight (goutput) on the managed entity based on input monitoring data (X1, X2, Xn).
- Monitoring CL function collecting position and battery level from cobots, which are used as input to the ANFIS model.
- Computation of a Migration Trigger (MT) Score, evaluated with historical scores by a Decision CL function, to decide when to migrate the service to reduce service downtime.
- Containerised model serving for the trained ANFIS model deployed and validated on the experimental PoC C cobots testbed.



ANFIS Model



Interactions among CL functions

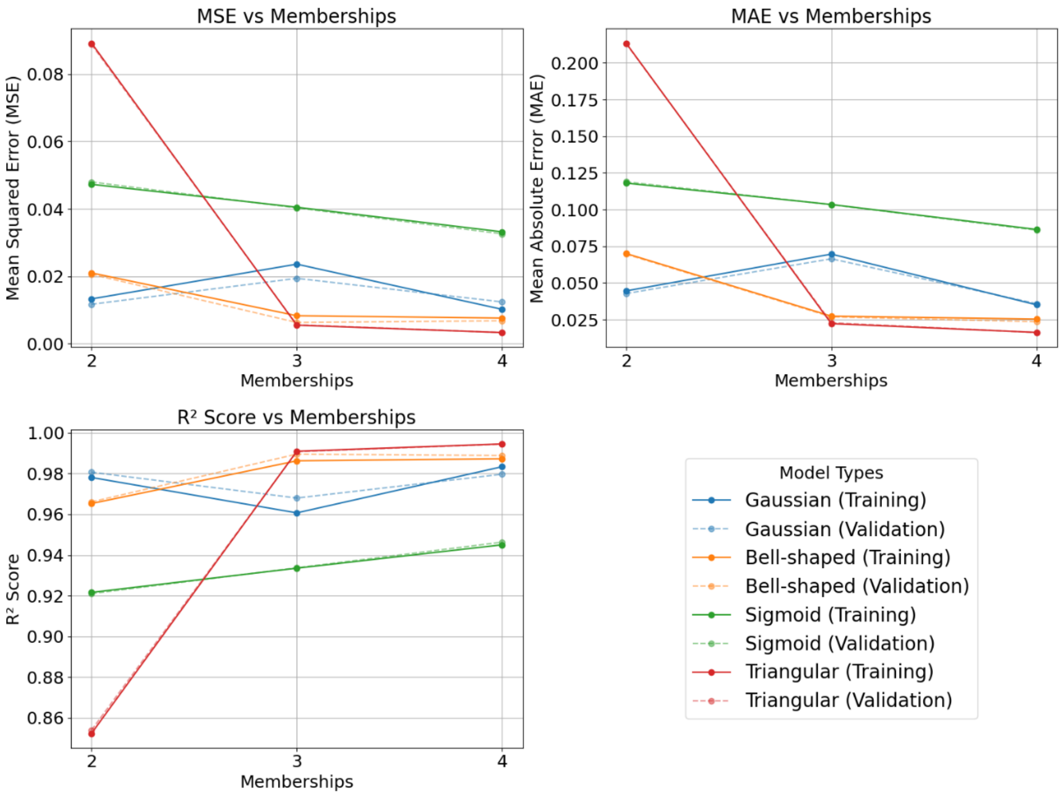
AI-enabled RT zero-touch control loop analysis function - Evaluation Results

Model trained using labelled historical data from cobot PoC testbed

- Continual stream of cobots’ positions and battery levels as they perform mobile video surveillance monitoring around the testbed area
- Total of 6.5 hours of data at 1 second sampling rate.

ANFIS model goal to determine the scale in [-1;1] range of the estimated efficiency gain from service migration based on cobot battery level and relative position.

- Mean Squared Error (MSE), Mean Absolute Error (MAE), and R2 obtained for training and validation datasets for 4 different membership function types (Gaussian, bell-shaped, sigmoidal, and triangular)
- For each membership type, the number of membership functions, n, per metric was evaluated for n= 2,3,4. For each evaluation, 50 epochs with batch size of 64.
- The triangular membership type slightly outperforms when the number of memberships is greater than 2.

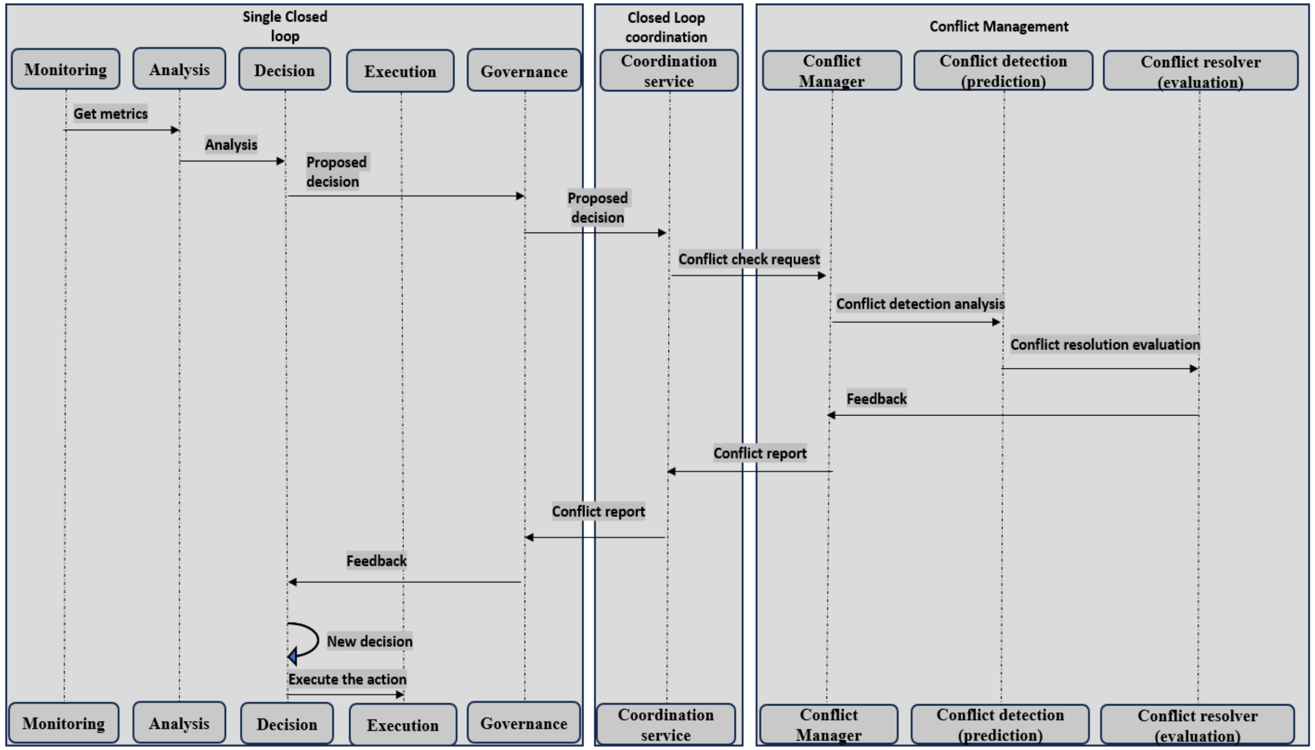


MSE, MAE and R2 results

Penalty-based management of concurrent service CLs

This slide describes a specific implementation associated to the real-time zero-touch CLs automation and coordination functionality. The implementation has been fully done in the context of the Hexa-X-II project, and with the overall purpose to enable a conflict detection and resolution mechanism among CLs. It basically relies on the following:

- Each CL has its own decision stage. In this stage, also referred as proposal stage, each CL can propose actions independently, but their actions are first predicted and evaluated in the Conflict Management block. This step is important because a decision that can be made by each single CL could affect the performance of other CLs.
- The practical development of this implementation has been done using a realistic network emulator which can support conversational video, Ultra-Reliable Low Latency Communication (URLLC), and massive Internet of Things (mIoT) service traffic.
- These services have different expectations, such as the Quality of Experience for video, Latency for URLLC, and packet loss for mIoT.
- In the development, the CLs coordination system interacts with a single CL and with the Conflict Management module. Besides, analysis and decision stages are combined as a proposal agent. This agent is responsible for the creation of one or more actions proposals with the goals to meet the expectations.
- The Monitoring and Telemetry functionality of the framework was also used to collect and analysis the network data.



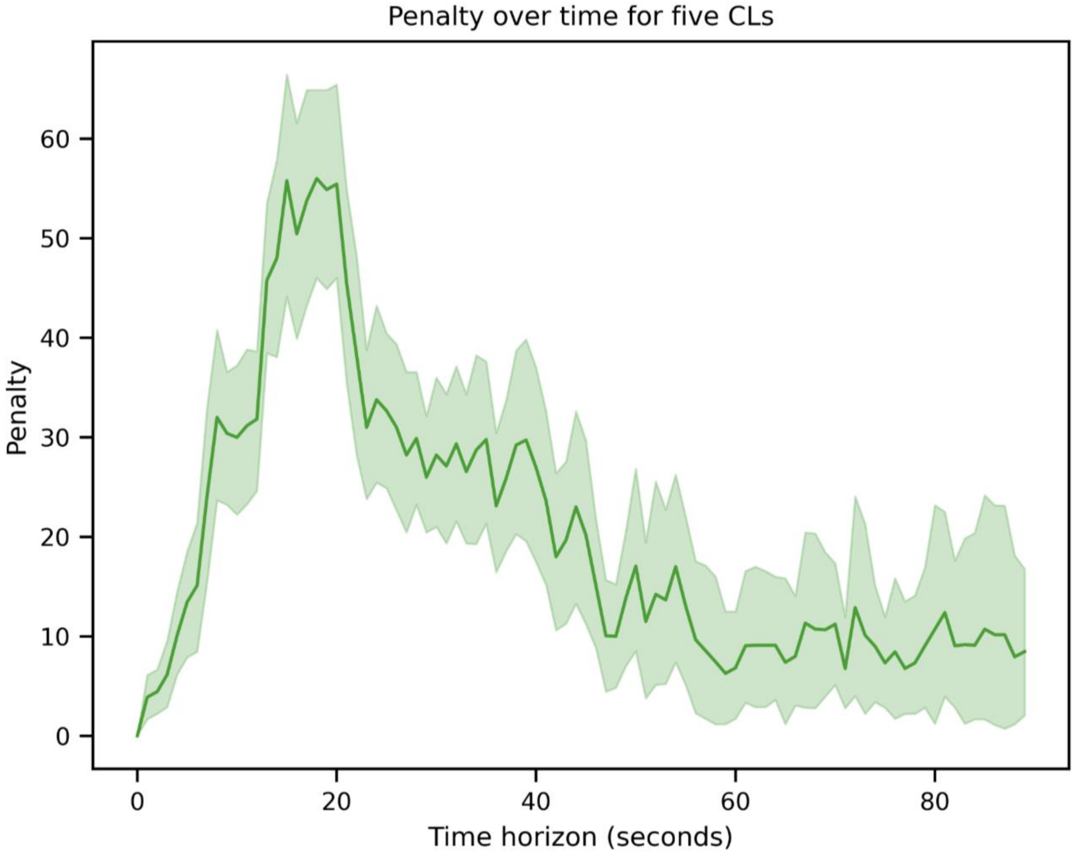
Proposed workflow for the concurrent closed loop conflict management

Penalty-based management of concurrent service CLs - Evaluation Results

- For evaluation results, experiments were conducted in a testbed to evaluate the management performance of five different closed loops. Each CL has a specific service with KPIs that needs to be met. If a CL does not meet its KPI, a penalty value is added following the equation:

$$P = \sum_{p=0}^{n-1} p_i$$

- Where n refers to the number of CLs and the index i refer to the current penalty of the Closed-Loop CL_i. If the Closed-loop CL_i is meeting the KPIs, then the penalty p_i is equal to zero.
- The goal is to take actions to minimize the penalty value by proposing different actions (e.g., increasing the computational power)
- The Figure shows the performance in the management of five concurrent CLs. The goal is to minimise the penalty value. In the implementation, the experiment was executed 22 times, and the average penalty was calculated over time. The Figure on the right shows the results obtained, with the x-axis representing the experiment execution time and the y-axis represents the penalty value. The overall penalty is obtained by the sum of the penalties.
- It is possible to see in the Figure that in the beginning of the experiment, the value of the penalty increases until it reaches the value closer to 55. This pattern reflects the fact that the closed loops are being deploying at different time stamps and the system is starting to take the decisions to minimize the overall penalty. After that, the penalty tends to decrease until reaches a stable value around 10.
- As future work, it is necessary to evaluate the performance for a higher numbers of CLs. The proposed solutions works for different numbers of CLs but a deeper study to evaluate the scalability is necessary



Numerical results for five CLs. The goal is to minimise the penalty value.

Conflict detection for the reactive activities requested by closed loops (1/4)

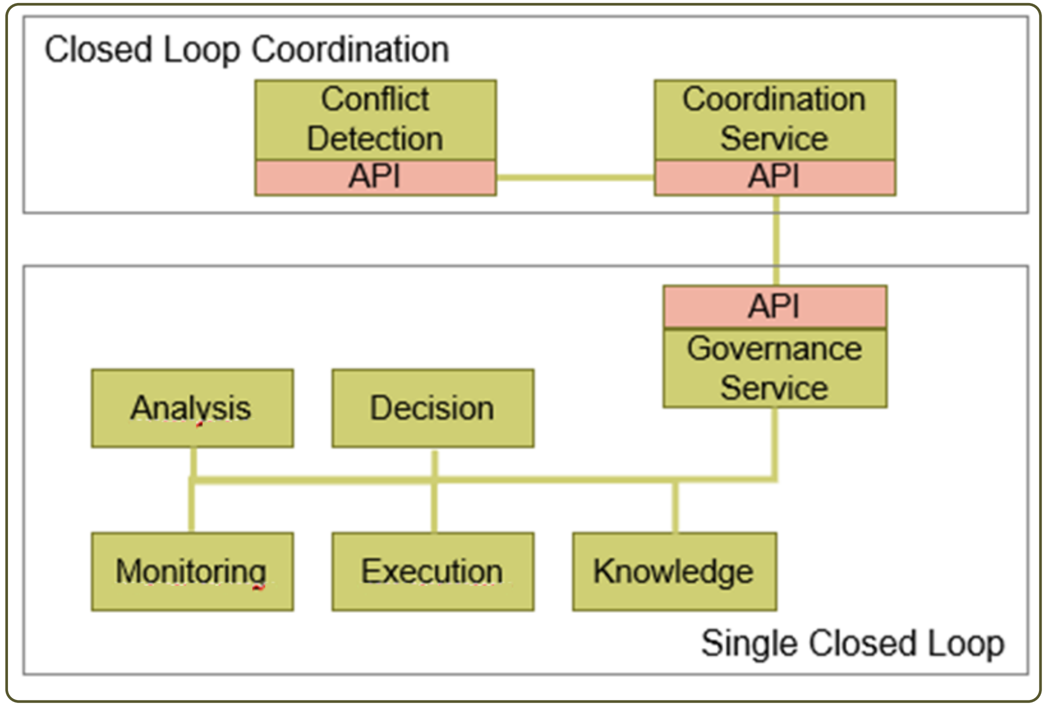
- This implementation is related with the Real-time Zero-touch CLs Automation and Coordination functionality, specifically in what regards the CLs Conflict Detection solution, which operates within the CL Coordination Service in this enabler.

- The implementation targets to automatically handle action plans proposed by interacting CLs, each aiming to maintain or improve specific system KPIs and goals. The Conflict Detection solution ensures that:
  - Non-conflicting action plans are executed directly.
  - Duplicated actions are removed to avoid redundancy.
  - Conflicting actions are resolved based on policies (priority, cost, penalties) to determine which action plan proceeds, either partially or fully.

- The development, specifically developed for this Hexa-X-II project, is a lab experiment intended to simulate and test the conflict management component's ability to:
  - Identify potential conflicts between concurrent action plans proposed by running closed loops.
  - Ensure that the best combination of action plans is selected based on system policies.

- For simplicity reasons, the tests were performed with the use of static priorities, but these priorities could be dynamic and reconfigurable (e.g. increasing over time) to avoid the case of an activity that is never executed due to its low priority.

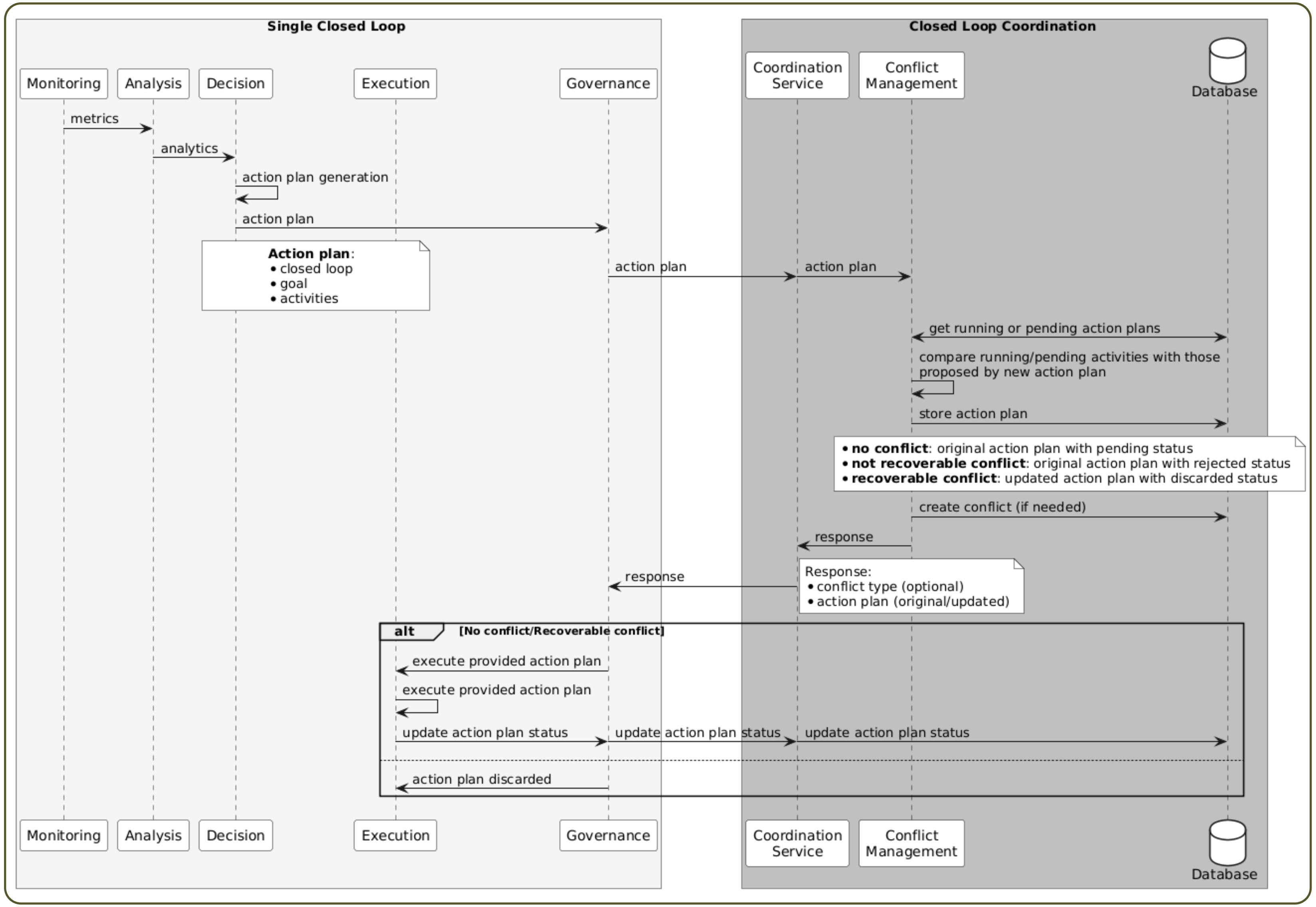
- This implementation considers three levels of priority to select or disable the conflicting activities in the action plans: high, medium, and low. These priorities could be defined based on the user profiles (silver/gold/platinum) and the committed SLAs, the final use of the managed entities (commercial use, security and rescue forces, etc) or the costs (economic, energy consumption, number of resources used, etc.).



Implementation blocks diagram

Conflict detection for the reactive activities requested by closed loops (2/4)

Conflict detection sequence diagram



Single Closed Loop

Closed Loop Coordination

Monitoring

Analysis

Decision

Execution

Governance

Coordination Service

Conflict Management

Database

metrics

analytics

action plan generation

action plan

Action plan:
- closed loop
- goal
- activities

action plan

action plan

get running or pending action plans

compare running/pending activities with those proposed by new action plan

store action plan

- **no conflict**: original action plan with pending status
- **not recoverable conflict**: original action plan with rejected status
- **recoverable conflict**: updated action plan with discarded status

create conflict (if needed)

response

response

Response:
- conflict type (optional)
- action plan (original/updated)

alt [No conflict/Recoverable conflict]

execute provided action plan

execute provided action plan

update action plan status

update action plan status

update action plan status

action plan discarded

The figure shows the scenario where a **Single Closed Loop** generates an action plan containing the closed loop identifier, desired goal, and the necessary activities based on the monitoring, analysis, and decision phases.

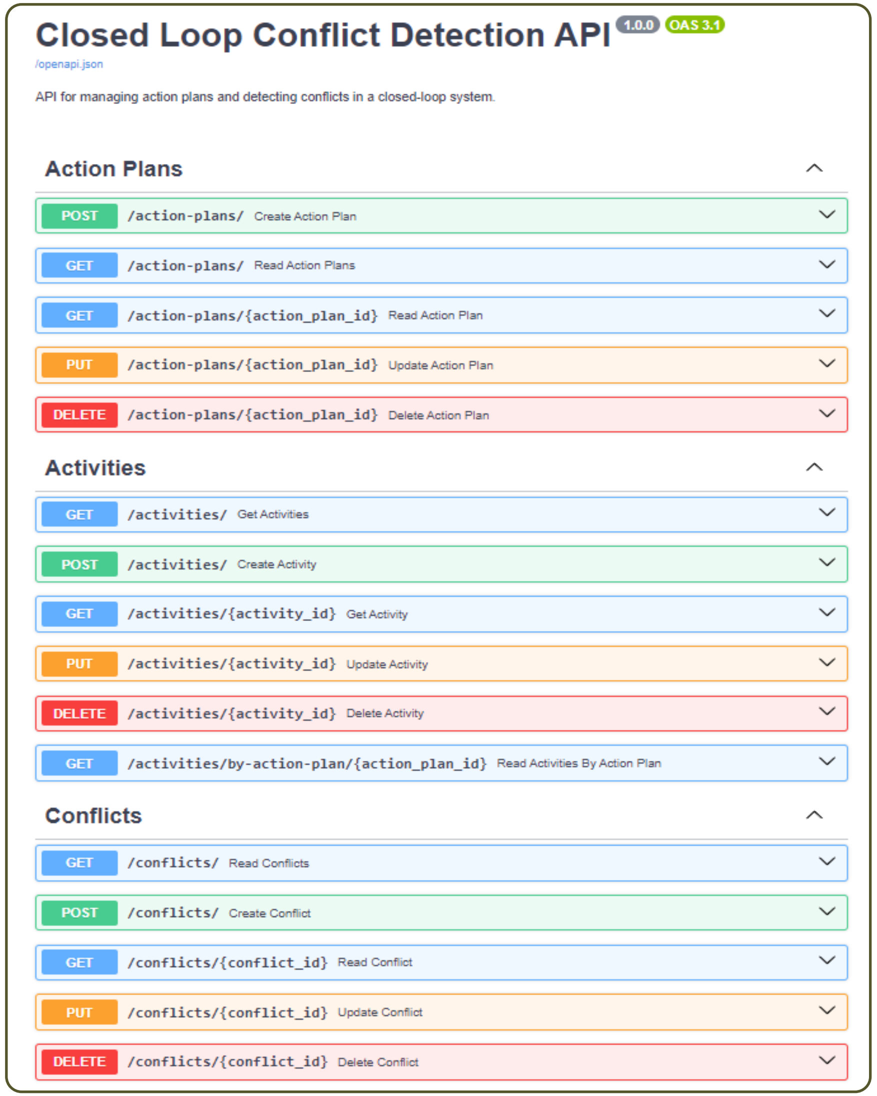
The action plan is then sent to the **Coordination Service**, which asks the **Conflict Management** component to check for conflicts with any running or pending action plans.

The **Conflict Management** module compares the new plan with existing ones and provides a response: either no conflict (allowing the original plan), an unrecoverable conflict (discarding the plan), or a recoverable conflict (suggesting an updated plan).

Based on this, the **Single Closed Loop** either executes the original/updated plan or discards it if execution is not feasible.

Conflict detection for the reactive activities requested by closed loops (3/4)

Open API:



Closed Loop Conflict Detection API 1.0.0 OAS 3.1

/openapi.json

API for managing action plans and detecting conflicts in a closed-loop system.

Action Plans

- POST `/action-plans/` Create Action Plan
- GET `/action-plans/` Read Action Plans
- GET `/action-plans/{action_plan_id}` Read Action Plan
- PUT `/action-plans/{action_plan_id}` Update Action Plan
- DELETE `/action-plans/{action_plan_id}` Delete Action Plan

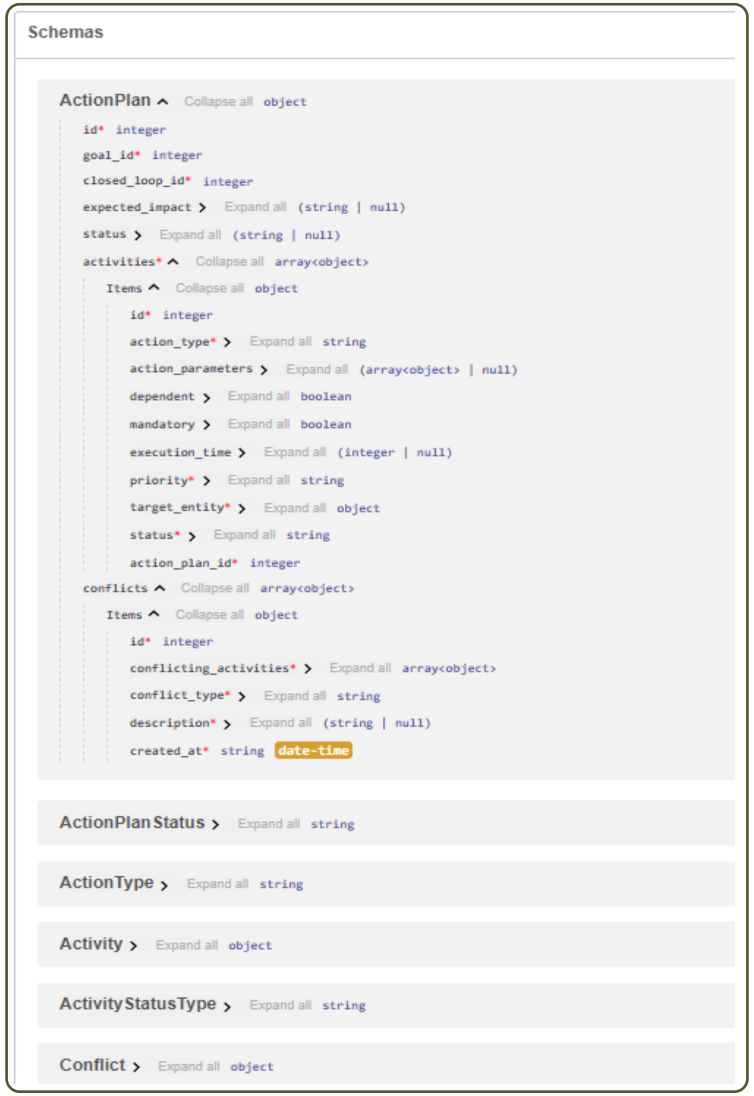
Activities

- GET `/activities/` Get Activities
- POST `/activities/` Create Activity
- GET `/activities/{activity_id}` Get Activity
- PUT `/activities/{activity_id}` Update Activity
- DELETE `/activities/{activity_id}` Delete Activity
- GET `/activities/by-action-plan/{action_plan_id}` Read Activities By Action Plan

Conflicts

- GET `/conflicts/` Read Conflicts
- POST `/conflicts/` Create Conflict
- GET `/conflicts/{conflict_id}` Read Conflict
- PUT `/conflicts/{conflict_id}` Update Conflict
- DELETE `/conflicts/{conflict_id}` Delete Conflict

Schemas definition:



Schemas

```
ActionPlan ^  Collapse all  object
  id*  integer
  goal_id*  integer
  closed_loop_id*  integer
  expected_impact >  Expand all  (string | null)
  status >  Expand all  (string | null)
  activities* ^  Collapse all  array<object>
    Items ^  Collapse all  object
      id*  integer
      action_type* >  Expand all  string
      action_parameters >  Expand all  (array<object> | null)
      dependent >  Expand all  boolean
      mandatory >  Expand all  boolean
      execution_time >  Expand all  (integer | null)
      priority* >  Expand all  string
      target_entity* >  Expand all  object
      status* >  Expand all  string
      action_plan_id*  integer
  conflicts ^  Collapse all  array<object>
    Items ^  Collapse all  object
      id*  integer
      conflicting_activities* >  Expand all  array<object>
      conflict_type* >  Expand all  string
      description* >  Expand all  (string | null)
      created_at*  string  date-time

ActionPlanStatus >  Expand all  string

ActionType >  Expand all  string

Activity >  Expand all  object

ActivityStatusType >  Expand all  string

Conflict >  Expand all  object
```

The solution is exposed via an OpenAPI server developed with Python and FastAPI [FAPI24].

The API allows the CL Coordination Service to submit action plans, retrieve conflicts, and fetch the status of running or pending action plans.

Key endpoints include submitting new action plans, detecting conflicts, and resolving those conflicts in real-time.

SQLAlchemy [SQLA24] is used for database interaction to store and retrieve action plans, resources, and conflict data.

Left figure shows the crud operations defined for action plans, activities, and conflicts plans. Activities and conflicts have their own schema definition (middle figure), but the relationships were established to allow an action plan to contain the set of activities to be executed (including its parameters and the target entities on which it acts), as well as the possible conflicts that may arise, showing which are the conflicting activities and the type of conflict.

Conflict detection for the reactive activities requested by closed loops (4/4)

Lessons learnt:

- It has been seen that, the most important and, at the same time, the most difficult part for this implementation was the definition of the **ActionPlan data model** so that it can reflect all the requested actions in the CLs allowing to establish relationships with activities, entities and conflicts. This complex data model was approached by breaking it down into the key components, using reusable schemas, clearly modelling relationships, and avoiding excessive nesting.
- The experiments were performed with the assumption that activities in **"running" action plan** cannot be disabled, but future versions should include the consideration that not all the activities of an action plan are launched at the same time and some of them could still be pending (and can be disable).

Addressed KPIs:

- Although conflict detection has impact on **availability** of services and connectivity, as well as in the reduction of **operational costs** in terms of energy consumption (avoiding the execution of conflicting actions that should be fixed) and maintenance costs, the most important indicator addressed is **Reliability,** as the ability of the system to perform without failure, and expressed according to this formula:

$$Reliability(\%) = \left(\frac{Number\ of\ correct\ action\ plans}{Number\ of\ total\ action\ plans\ generated}\right) \times 100$$

where the number of correct action plans is the sum of not conflicting action plans executed and the conflicting action plans that were updated and executed (the reliability increases thanks to the detection and update of conflicting action plans)

Next steps:

- This conflict detection implementation follows a comparison approach where new proposed action plans are compared with the existing ones. This could be extended with **AI/ML-based** techniques that could help with the task of identifying conflicts, providing a more optimal selection of activities for their execution.

Conflict detection for the reactive activities requested by closed loops – Evaluation Results

- In the testbed context where the implementation was executed it was not possible to obtain specific and relevant KPI values since, as the implementation was not integrated into a real system, the Action Plans needed to be generated artificially which does not allow for the extraction of precise information from the KPIs for conflicts and their possible interdependencies. Therefore, the tests were focused on validating the detection of conflicts itself, and not on performance.

- As a valid reference in the state of the art, it has been seen that in the context of the Open Radio Networks [DMIC23], at least a 2% increase in mean bitrate was achieved by relying on conflict detection techniques, which is considered to validate the approach taken in this implementation.

Human-assisted training of cognitive closed loops functions for network automation (1/3)

- This section describes a specific implementation also associated to the Real-time Zero-touch CLs Automation and Coordination functionality, specifically in what regards the Analysis and Decision stages of a CL in this enabler. The implementation has been specifically developed for the Hexa-X-II project.

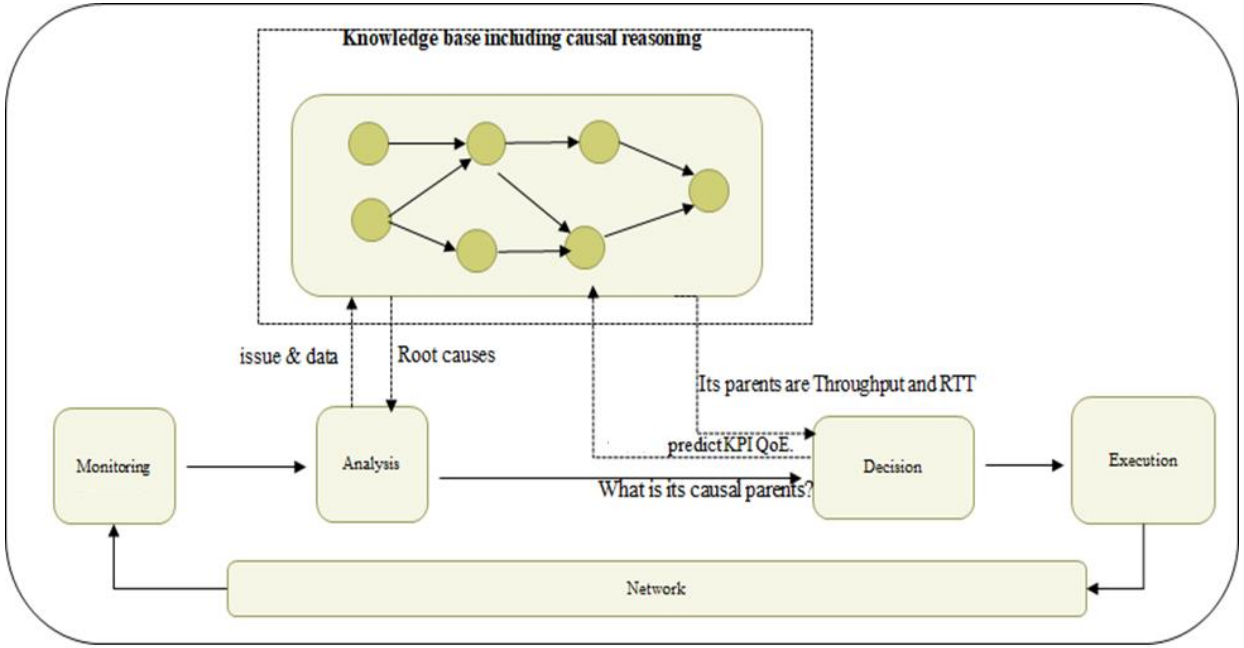
- The implementation overall purpose is to enable cognitive CLs that can make reasoning on their decision and predictions, targeting the following:
  - That non-conflicting action plans are executed directly.
  - That duplicated actions are removed to avoid redundancy.
  - Resolve conflicting actions based on policies to determine which action plan proceeds, either partially or fully.

- For this, the objective is to find causal relations between the network KPIs and network configurations or actions, which are discovered through a data-driven approach. This is done relying on the concept of casual graphs [KAR24]. Once the graphs are built, the Analysis and Decision stages of CLs can utilise that information to make causal reasoning that can bring not only a more effective system solution, but also, explainable understanding behind the choice of different network configurations.

- This implementation has been deployed on a realistic network emulator which can support real video and URLLC traffic. Using this emulator:
  - Different network configurations can be taken.
  - Different network KPIs can be measured, such as Throughput, Packet Loss, Round-trip Time, and QoE.

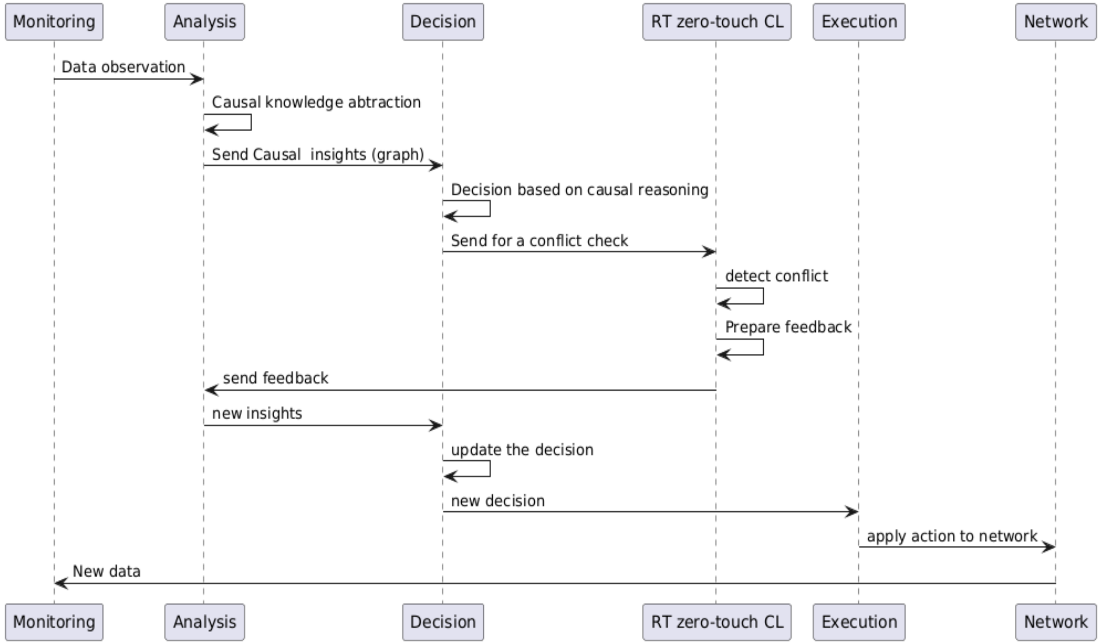
- Besides the real-time zero-touch CLs automation and coordination functionality, the Monitoring and Telemetry enabler of the management framework has been also implemented for the experiments, in order to perform the collection and analysis of the network data.

Human-assisted training of cognitive closed loops functions for network automation (2/3)



Knowledge base including causal reasoning
issue & data
Root causes
Its parents are Throughput and RTT
predictKPI QoE.
What is its causal parents?
Monitoring
Analysis
Decision
Execution
Network

Components of a Cognitive CL



Workflow associated to a Cognitive CL

- Analysis and Decision stages of a CL can interact with the causal graph that can be located in a knowledge base.
- Root-causes of the problem can be identified at the Analysis stage.
- Depending on the root causes, the Decision Stage can make more suitable actions.

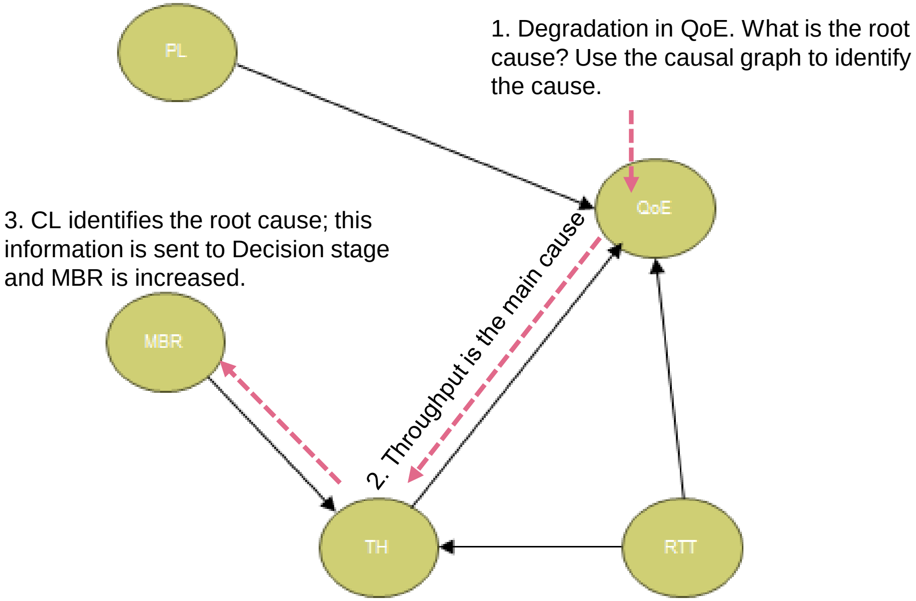
- This workflow shows how causal reasoning can be integrated in the end-to-end decision making and conflict resolution process of CLs.
- Causal reasoning can play a role in conflict resolution as well.

Human-assisted training of cognitive closed loops functions for network automation (3/3)

- Achieving a decent causal graphs is an important step towards to cognitive CL automation. Experiments have been performed to verify the feasibility of finding causal relations between the network KPIs and configurations, which are discovered through a data-driven approach.
- Based on the state-of art [PC91] [SCR02] a new causal discovery algorithm has been developed and tested using an in-house network emulator.
- The implementation was performed using data coming from such in-house network emulator, which can provide realistic video, URLCC, and mIoT traffic.
- Accuracy of the achieved causal graph is critical. It should be checked with domain knowledge, if possible, to have a better assessment.

Human-assisted training of cognitive closed loops functions for network automation - Evaluation Results

- Two network configurations were used for testing: Priority levels of services, and Maximum Bit Rate. For them, the following service KPIs were measured:
  - Throughput, with values up to ~3 Mbps.
  - QoE, varying between 1 to 4.
  - Round-trip Time, up to ~200 ms.
  - Packet Loss: in the range from 0 to 1%.
- The casual graph shown in the figure below was obtained, which is considered aligned with the consortium domain knowledge with an accuracy of more than %90.



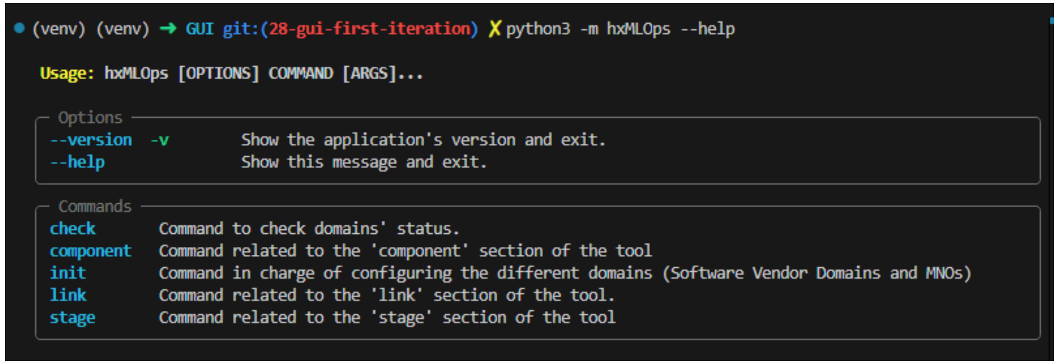
1. Degradation in QoE. What is the root cause? Use the causal graph to identify the cause.

3. CL identifies the root cause; this information is sent to Decision stage and MBR is increased.

Achieved Final Causal Graph in the Experiment

Sustainable MLOps

- This implementation showcases how the Sustainable MLOps enabler in the framework could be carried out in practice. It represents a minimalist version of the enabler that could be deployed in the scopes of the different stakeholders participating in the MLOps workflows (network operators, software vendors…), and consists of a tool based on a command line interface (to define the MLOps workflows) and a GUI (to visualise the energy consumption metrics at the different stages of the workflows). The objective of this tool is to serve as a basic implementation example, which could obviously be extended with additional functionalities.
- The command line interface (CLI) provides a set of commands to define the components participating in the MLOps workflows (e.g., energy measurement probes, AI/ML models repositories, AI/ML serving platforms…), as well as their different stages. This tool has been specifically developed in the context of the Hexa-X-II project and has been released to the Open-Source community (ref. in Slide 83).
- The energy consumption and carbon production translation, in the scope of the software providers and MNOs (based on location and source distribution), is collected from the execution environments, represented in dashboards with different level of granularity, and grouped by MLOps workflow stage, to support sustainability-related decision making. Besides, this information can also be included as metainformation in the AI/ML models themselves.

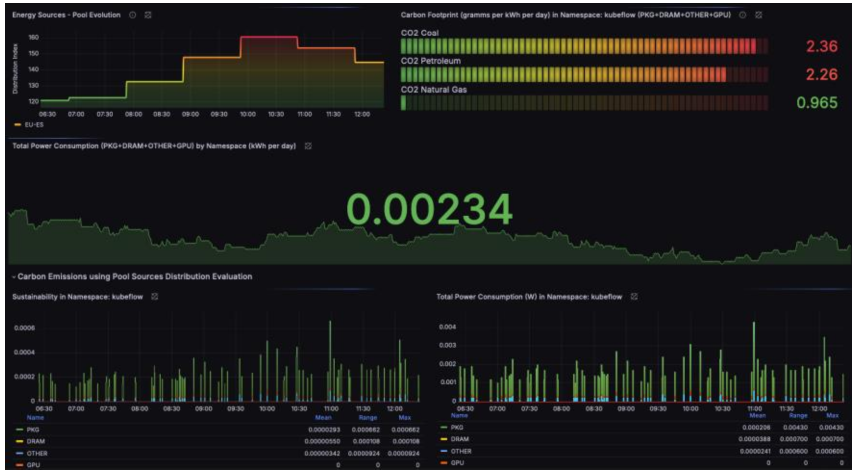


```
● (venv) (venv) ➜ GUI git:(28-gui-first-iteration) ✗ python3 -m hxMLOps --help

 Usage: hxMLOps [OPTIONS] COMMAND [ARGS]...

╭─ Options ──────────────────────────────────────────────────────────────────╮
│ --version  -v        Show the application's version and exit.              │
│ --help               Show this message and exit.                           │
╰────────────────────────────────────────────────────────────────────────────╯
╭─ Commands ─────────────────────────────────────────────────────────────────╮
│ check        Command to check domains' status.                             │
│ component    Command related to the 'component' section of the tool        │
│ init         Command in charge of configuring the different domains (Software Vendor Domains and MNOs) │
│ link         Command related to the 'link' section of the tool.            │
│ stage        Command related to the 'stage' section of the tool            │
╰────────────────────────────────────────────────────────────────────────────╯
```

MLOps Main command line interface

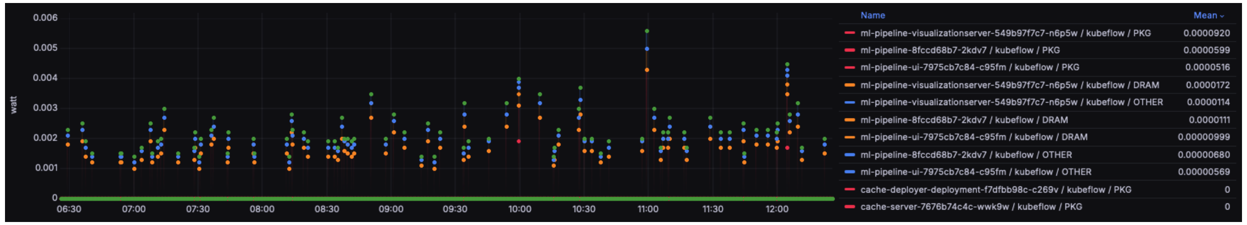


Energy Sources - Pool Evolution
Distribution Index
160
150
140
130
120
06:30
07:00
07:30
08:00
08:30
09:00
09:30
10:00
10:30
11:00
11:30
12:00
EU-ES
Carbon Footprint (gramms per kWh per day) in Namespace: kubeflow (PKG+DRAM+OTHER+GPU)
CO2 Coal
2.36
CO2 Petroleum
2.26
CO2 Natural Gas
0.965
Total Power Consumption (PKG+DRAM+OTHER+GPU) by Namespace (kWh per day)
0.00234
Carbon Emissions using Pool Sources Distribution Evaluation
Sustainability in Namespace: kubeflow
0.0006
0.0004
0.0002
0
Total Power Consumption (W) in Namespace: kubeflow
0.004
0.003
0.002
0.001
0

Name	Mean	Range	Max
PKG	0.0000293	0.000662	0.000662
DRAM	0.00000550	0.000108	0.000108
OTHER	0.00000342	0.0000924	0.0000924
GPU	0	0	0

Name	Mean	Range	Max
PKG	0.000206	0.00430	0.00430
DRAM	0.0000388	0.000700	0.000700
OTHER	0.0000241	0.000600	0.000600
GPU	0	0	0

Energy Consumption dashboard



Name	Mean
ml-pipeline-visualizationserver-549b97f7c7-n6p5w / kubeflow / PKG	0.0000920
ml-pipeline-8fccd68b7-2kdv7 / kubeflow / PKG	0.0000599
ml-pipeline-ui-7975cb7c84-c95fm / kubeflow / PKG	0.0000516
ml-pipeline-visualizationserver-549b97f7c7-n6p5w / kubeflow / DRAM	0.0000172
ml-pipeline-visualizationserver-549b97f7c7-n6p5w / kubeflow / OTHER	0.0000114
ml-pipeline-8fccd68b7-2kdv7 / kubeflow / DRAM	0.0000111
ml-pipeline-ui-7975cb7c84-c95fm / kubeflow / DRAM	0.00000999
ml-pipeline-8fccd68b7-2kdv7 / kubeflow / OTHER	0.00000680
ml-pipeline-ui-7975cb7c84-c95fm / kubeflow / OTHER	0.00000569
cache-deployer-deployment-f7dfbb98c-c269v / kubeflow / PKG	0
cache-server-7676b74c4c-wwk9w / kubeflow / PKG	0

Carbon production results in the different workflow stages

Sustainable MLOps

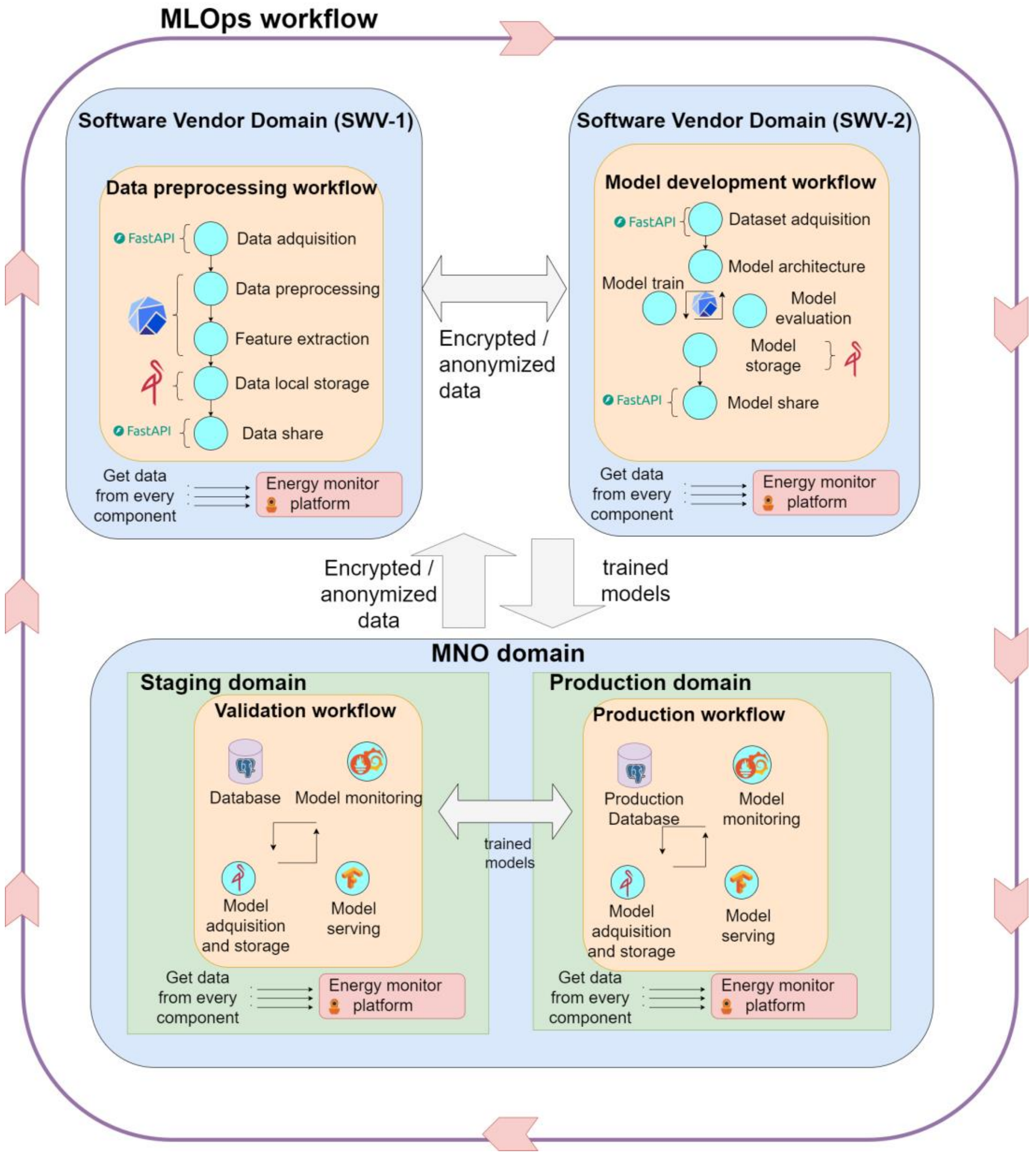
The figure on the right represents an example workflow created with the tool involving three stakeholders:

- A Software Vendor (SWV-1), implementing a specific workflow simulating to perform data preprocessing related activities.
- Another Software Vendor (SWV-2), with another specific workflow to manage the AI models development.
- A Mobile Network Operator (MNO), that provides the first SWV with training data, and in turn, receives the trained models from the second SWV-2. This MNO also implements its own inner workflows, to test the received models (within its validation domain) and to deploy them on the production environment.

Energy consumption data for assessing the sustainability of workflows is executed at both: the software vendors and the MNO environments:

- In the software vendors scope, the data collection is performed for their whole workflows and its individual stages and are served for monitoring and as metainformation in the AI models to support the improvement and eligibility criteria of the models in what regards their generated carbon footprint.
- In the MNO scope, the workflows the target was to validate the AI Model performance in terms of energy consumption indexes.

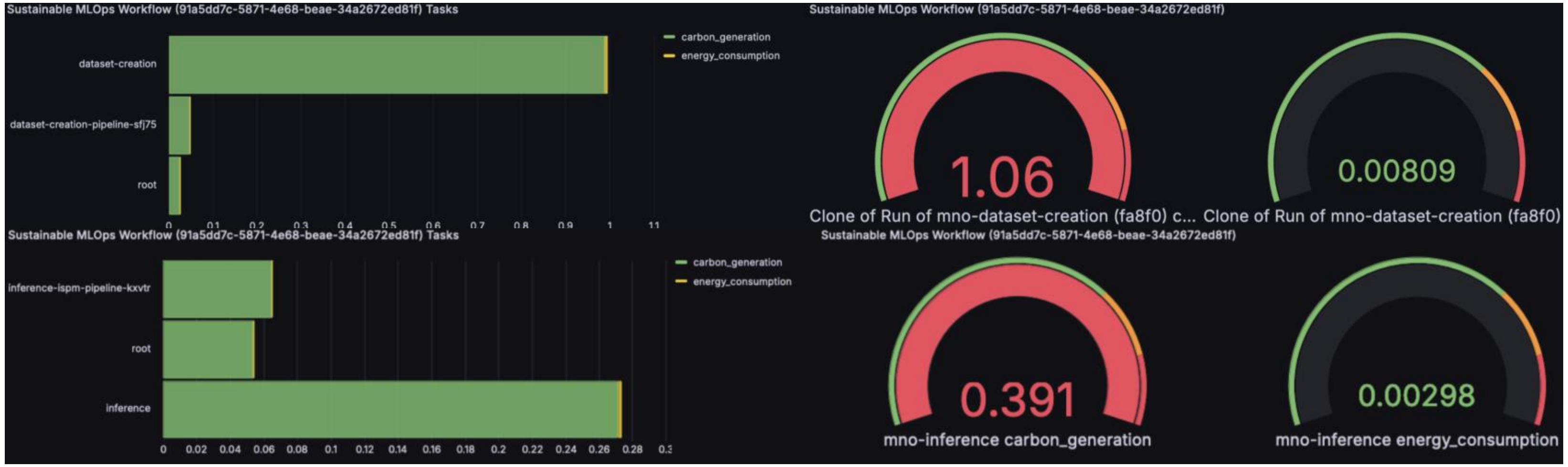
The screenshots in the previous slide show the plots that can be obtained from the tool. A more detailed information regarding these results will be provided in the upcoming Deliverable D6.5.



Example workflow set up using the S-MLOps asset

Sustainable MLOps – Evaluation Results

- The evaluation was made using a test bed composed by different Kubernetes and Kubeflow environments representing the infrastructure of SWV-1, SWV-2 and MNO.
- S-MLOps framework, available in all environments, oversees execution of MLOps workflows as described in D6.5, obtain and monitor measurements and compose Sustainable metrics and publish models with extended metainformation regarding S-MLOps scenario.
- Over each test bed scenario, different developed MLOps workflows, based on SWV and MNO roles, were deployed and measured, obtaining metrics related with Sustainability for every workflow stage, published and shared as model metainformation in model sharing API.



Sustainable MLOps Workflow (91a5dd7c-5871-4e68-beae-34a2672ed81f) Tasks
dataset-creation
dataset-creation-pipeline-sfj75
root
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
carbon_generation
energy_consumption
Sustainable MLOps Workflow (91a5dd7c-5871-4e68-beae-34a2672ed81f)
1.06
Clone of Run of mno-dataset-creation (fa8f0) c...
0.00809
Clone of Run of mno-dataset-creation (fa8f0)
Sustainable MLOps Workflow (91a5dd7c-5871-4e68-beae-34a2672ed81f) Tasks
inference-ispm-pipeline-kxvtr
root
inference
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.3
carbon_generation
energy_consumption
Sustainable MLOps Workflow (91a5dd7c-5871-4e68-beae-34a2672ed81f)
0.391
mno-inference carbon_generation
0.00298
mno-inference energy_consumption

Energy Consumption and Carbon Generation Metric by MLOps stage dashboard examples (inference and Dataset creation pipelines)

- As a result, with test executions over the test bed, we have been capable to demonstrate the capability of obtaining, monitoring, sharing and using Sustainable metrics in MLOps scenarios to improve stages and overall workflow efficiency and as a key factor for model eligibility in a multi stakeholder scenario.

ETSI TeraFlowSDN

Network Programmability Framework Implementation

- The Smart Network Management Network Programmability Framework relies on TeraFlowSDN (TFS).
- Introduces Data-Plane in-a-Box, enabling rapid PoC integration and transport SDN demonstrations.
- Integrated into System PoC#B.1 and planned for PoC#C.

SmartNIC Transceiver Support (OpenConfig Extensions)

- Provides SDN-enabled anomaly detection by extending TFS with SmartNIC support.
- Planned for TFS Release 5.Presented at [VVG+24]. Implemented in the ADRENALINE Testbed.

Time-Sensitive & Deterministic Networking (TSN/DetNet)

- Implements an East-Westbound (EW) Control Architecture for modular DetNet networks.
- Uses multiple Centralised Network Controllers (CNCs) to manage L2 & L3 segments.
- Enables bounded latency & zero packet loss via TSN (TAS) and strict priority queueing.

Automated Transport Network Reconfiguration

- Included in TFS Release 4 and extended in Release 5.
- Will be demonstrated in the ADRENALINE Testbed.

MEC BandWidth Management (BWM) & TFS Synergy

- Part of ETSI MEC PoC 14 on network resource allocation for gaming.
- Introduced in TFS Release 3.
- Ensures QoS-aware bandwidth allocation with MEC-BWM and TFS traffic orchestration.

Standardisation & API Convergence

- Integrated TM Forum APIs (TMF640, TMF664) with ETSI TFS NBI for holistic network management.
- Unifies business-oriented APIs (TM Forum) with intent-driven operations (ETSI).
- Facilitates YANG-based network management for enhanced automation and flexibility.
- Detailed in [HEXA223-D63].

Monitoring and Telemetry

TeraFlowSDN Event-Driven Monitoring

Implemented in TeraFlowSDN Release 4, aligned with event & data-driven architectural principles.

Integrated into System PoC#B.1.

Introduces workflow diagrams for real-time data extraction, processing, and visualisation.

Energy Monitoring & Carbon Footprint Analysis

Designed for monitoring energy consumption across the compute continuum (cloud, edge, extreme edge).

Validated in laboratory environments [MNC+17].

Aims to optimise 6G sustainability and energy efficiency.

Monitoring Platform for Closed-Loop Automation

Integrated into PoC#B, particularly in cobot-driven automation scenarios [PBM+24].

Supports Zero-Touch Closed Loops (ZT-CLs) with hierarchical control models.

Enables delegation & escalation strategies for autonomous decision-making.

Time-Sensitive & Deterministic Networking (TSN/DetNet) Monitoring

Ensures end-to-end latency & packet loss compliance.

Implements multiple monitoring strategies:

- Passive Monitoring: Collects network statistics (bytes sent, lost packets, etc.) without interference.
- Active Probing: Injects artificial traffic to measure end-to-end delay.
- In-Band Telemetry: Monitors network performance with low overhead, capturing real-time statistics.

Data Fusion for Signal Correlation & Remediation

Combines active & passive telemetry for comprehensive failure mitigation.

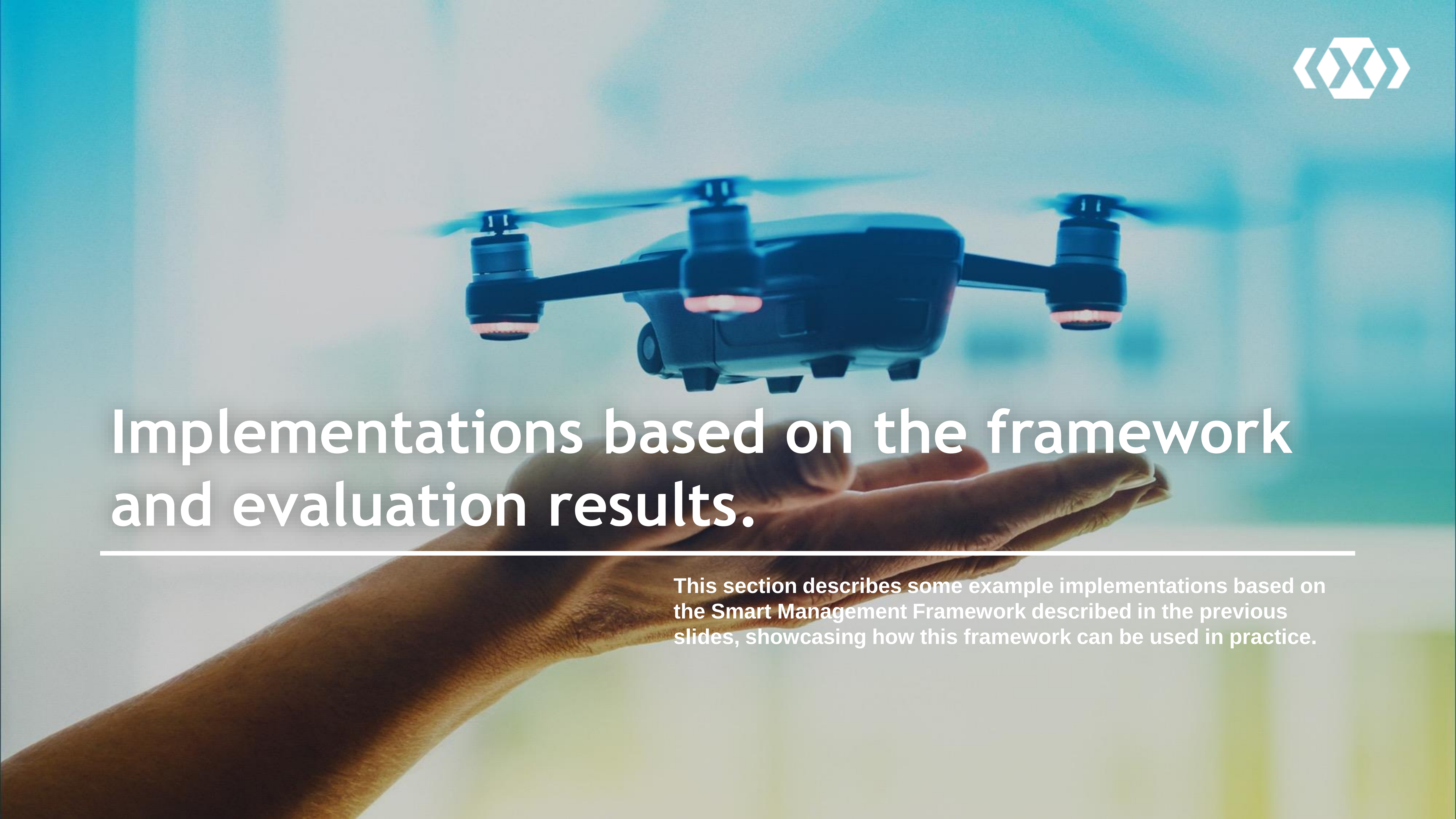
Uses real-time analysis to predict and prevent system & network failures [TAZ+22].

OpenTelemetry-Based Observability

Implements OpenTelemetry [OTL24] for real-time signal access & system orchestration.

Supports distributed data collection with two architectural models:

- OTLP Gateway: Centralised data collection from directly accessible sources.
- OTLP Agent: Distributed data exporters forwarding signals to centralised locations.

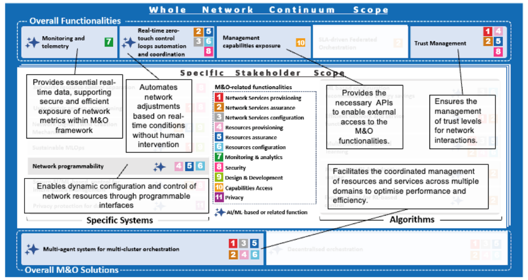


Implementations based on the framework and evaluation results.

This section describes some example implementations based on the Smart Management Framework described in the previous slides, showcasing how this framework can be used in practice.

Overview

The following implementations have been performed, combining different enablers of the management framework (highlighted in red in the figures):


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
Real-time zero-touch control loops automation and coordination
Management capabilities exposure
SLA-driven Federated Orchestration
Trust Management
Specific Stakeholder Scope
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
Network programmability
Specific Systems
Algorithms
Multi-agent system for multi-cluster orchestration
Decentralised orchestration
Overall M&O Solutions

Provides essential real-time data, supporting secure and efficient exposure of network metrics within M&O framework

Automates network adjustments based on real-time conditions without human intervention

Provides the necessary APIs to enable external access to the M&O functionalities.

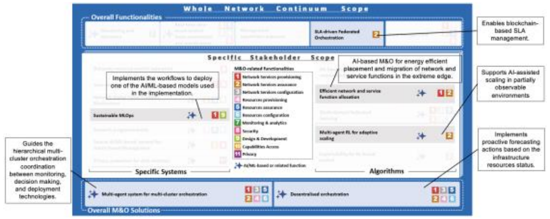
Ensures the management of trust levels for network interactions.

Enables dynamic configuration and control of network resources through programmable interfaces

Facilitates the coordinated management of resources and services across multiple domains to optimise performance and efficiency.

1. Usage of the management capabilities exposure system.

Showcases the usage of the Management Capabilities Exposure enabler considering different use cases: the integration of a vertical industry, and failure detection and recovery scenarios.


Whole Network Continuum Scope
Overall Functionalities
SLA-driven Federated Orchestration
Specific Stakeholder Scope
M&O-related functionalities
Sustainable MLOps
Efficient network and service function allocation
Multi-agent RL for adaptive scaling
Specific Systems
Algorithms
Multi-agent system for multi-cluster orchestration
Decentralised orchestration
Overall M&O Solutions

Enables blockchain-based SLA management.

AI-based M&O for energy efficient placement and migration of network and service functions in the extreme edge.

Implements the workflows to deploy one of the AI/ML-based models used in the implementation.

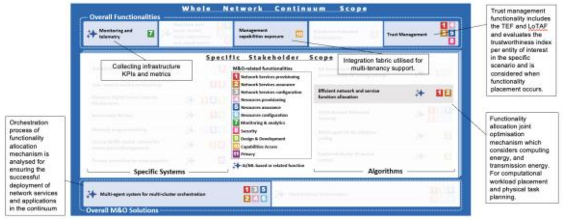
Supports AI-assisted scaling in partially observable environments

Implements proactive forecasting actions based on the infrastructure resources status.

Guides the hierarchical multi-cluster orchestration coordination between monitoring, decision making, and deployment technologies.

2. Orchestration on the network continuum.

Showcases different M&O scenarios on the network continuum, including proactive migration of service components, MLOps, services federation, and trust management.


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
Management capabilities exposure
Trust Management
Specific Stakeholder Scope
M&O-related functionalities
Efficient network and service function allocation
Specific Systems
Algorithms
Multi-agent system for multi-cluster orchestration
Overall M&O Solutions

Trust management functionality includes the TEF and LoTAF and evaluates the trustworthiness index per entity of interest in the specific scenario and is considered when functionality placement occurs.

Collecting infrastructure KPIs and metrics

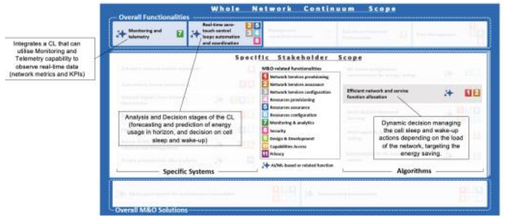
Integration fabric utilised for multi-tenancy support.

Functionality allocation joint optimisation mechanism which considers computing energy, and transmission energy. For computational workload placement and physical task planning.

Orchestration process of functionality allocation mechanism is analysed for ensuring the successful deployment of network services and applications in the continuum

3. Functionality allocation in a cobot-powered warehouse inventory system.

Orchestration of a network service providing an automated inventory management solution for accurate and efficient warehousing operations using collaborative robots.


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
Real-time zero-touch control loops automation and coordination
Specific Stakeholder Scope
M&O-related functionalities
Efficient network and service function allocation
Specific Systems
Algorithms
Overall M&O Solutions

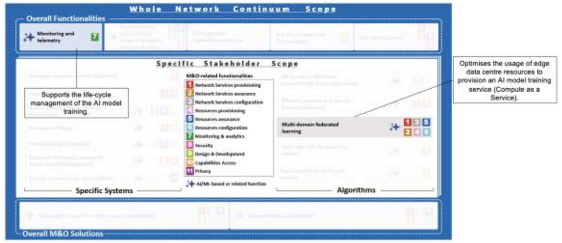
Integrates a CL that can utilise Monitoring and Telemetry capability to observe real-time data (network metrics and KPIs)

Analysis and Decision stages of the CL (forecasting and prediction of energy usage in horizon, and decision on cell sleep and wake-up)

Dynamic decision managing the cell sleep and wake-up actions depending on the load of the network, targeting the energy saving.

4. ML based configuration recommendation for energy saving.

Fully automated CL-based solution for correcting flow misconfigurations in a scalable way in deterministic networks. The system takes autonomous decision for cells sleep and wake-up.

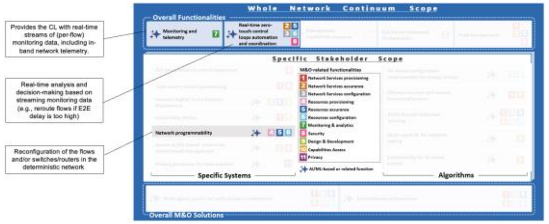

Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
Specific Stakeholder Scope
M&O-related functionalities
Multi-domain federated learning
Specific Systems
Algorithms
Overall M&O Solutions

Supports the life-cycle management of the AI model training.

Optimises the usage of edge data centre resources to provision an AI model training service (Compute as a Service).

5. Resource assignment for federated learning.

Showcases how telco operators could leverage on rich data sources of connected devices and on the deployment of edge compute resources to provision AI model training as a service, in line with the CaaS paradigm.


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry
Real-time zero-touch control loops automation and coordination
Specific Stakeholder Scope
M&O-related functionalities
Network programmability
Specific Systems
Algorithms
Overall M&O Solutions

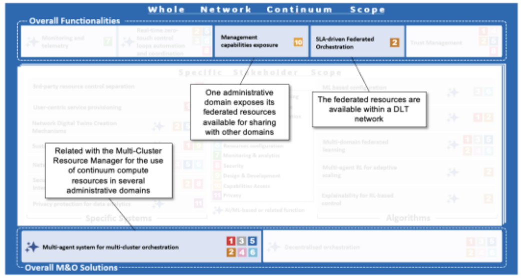
Provides the CL with real-time streams of (per-flow) monitoring data, including in-band network telemetry.

Real-time analysis and decision-making based on streaming monitoring data (e.g., reroute flows if E2E delay is too high)

Reconfiguration of the flows and/or switches/routers in the deterministic network

6. Flow Reconfiguration based on Dynamic Monitoring and Closed Loops in Deterministic Networks.

Provides an automated CL-based solution for correcting flow misconfigurations in a scalable way in deterministic networks by exploring trade-offs between different forms of telemetry.


Whole Network Continuum Scope
Overall Functionalities
Management capabilities exposure
SLA-driven Federated Orchestration
Multi-agent system for multi-cluster orchestration
Overall M&O Solutions

One administrative domain exposes its federated resources available for sharing with other domains

The federated resources are available within a DLT network

Related with the Multi-Cluster Resource Manager for the use of continuum compute resources in several administrative domains

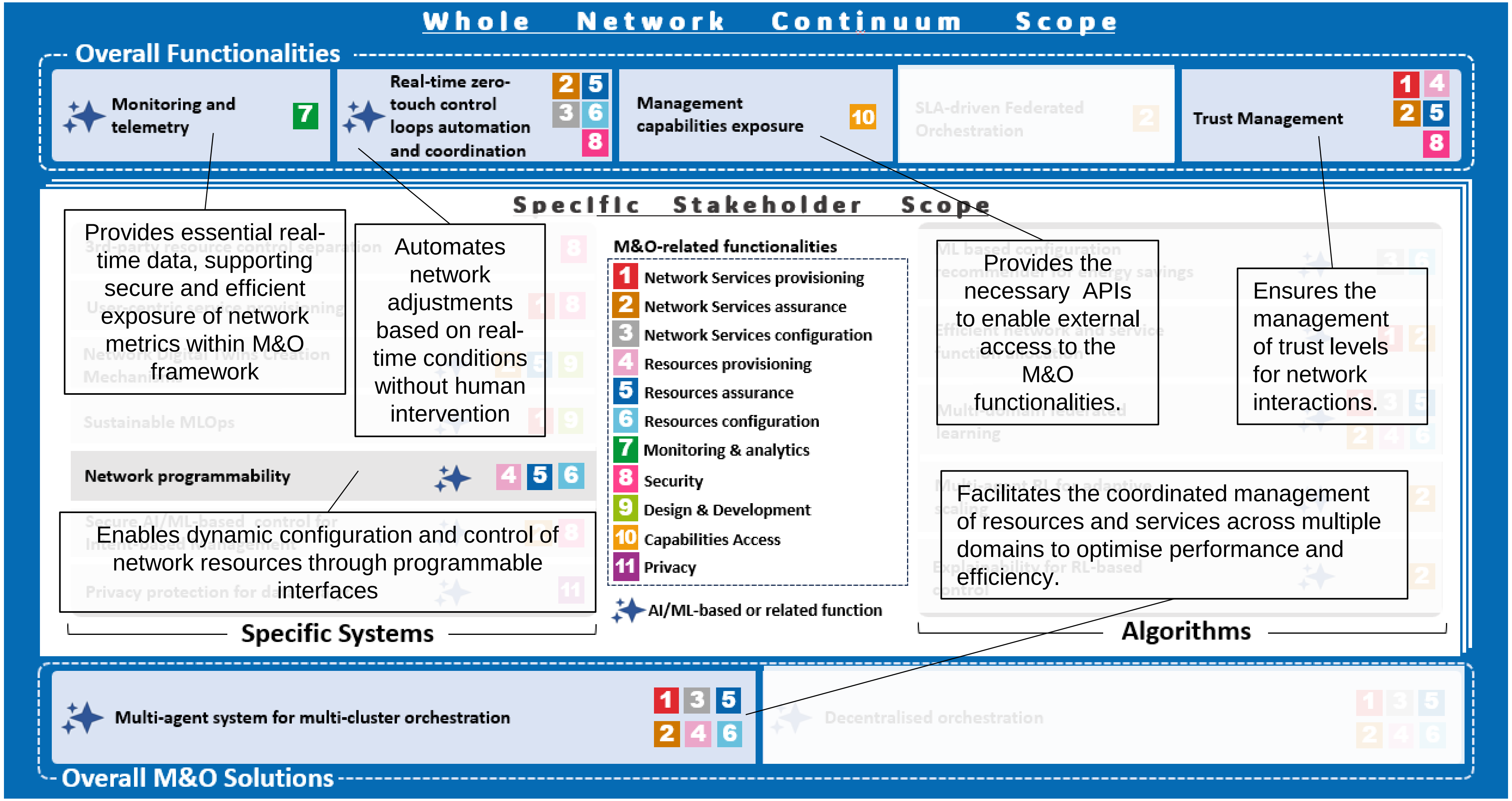
7. Edge convergence over federated resources for the computing continuum.

Explores the capabilities of the CAMARA EdgeCloud APIs in the management of the compute resources in the network continuum, and the possibility to extend them to be used with federated resources of external administrative domains.

The following slides provide additional details on each of these implementations

Implementation 1: Usage of the Management Capabilities Exposure System - Introduction (1)

The purpose of this implementation is to showcase the usage of the Management Capabilities Exposure System, together with other enablers in the management framework. The following figure shows the different enablers involved, along with the functionality they perform in the implementation:


Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party resource control separation 8
User-centric service provisioning 1 8
Network Digital Twins Creation Mechanisms 2 5 9
Sustainable MLOps 1 9
Network programmability 4 5 6
Secure AI/ML-based control for Intent-based management 2 8
Privacy protection for data 11
Specific Systems
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based configuration recommender for energy savings 3 6
Efficient network and service function allocation 1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

Provides essential real-time data, supporting secure and efficient exposure of network metrics within M&O framework

Automates network adjustments based on real-time conditions without human intervention

Provides the necessary APIs to enable external access to the M&O functionalities.

Ensures the management of trust levels for network interactions.

Enables dynamic configuration and control of network resources through programmable interfaces

Facilitates the coordinated management of resources and services across multiple domains to optimise performance and efficiency.

Implementation 1: Usage of the Management Capabilities Exposure System - Introduction (2)

This implementation targets the following topics:

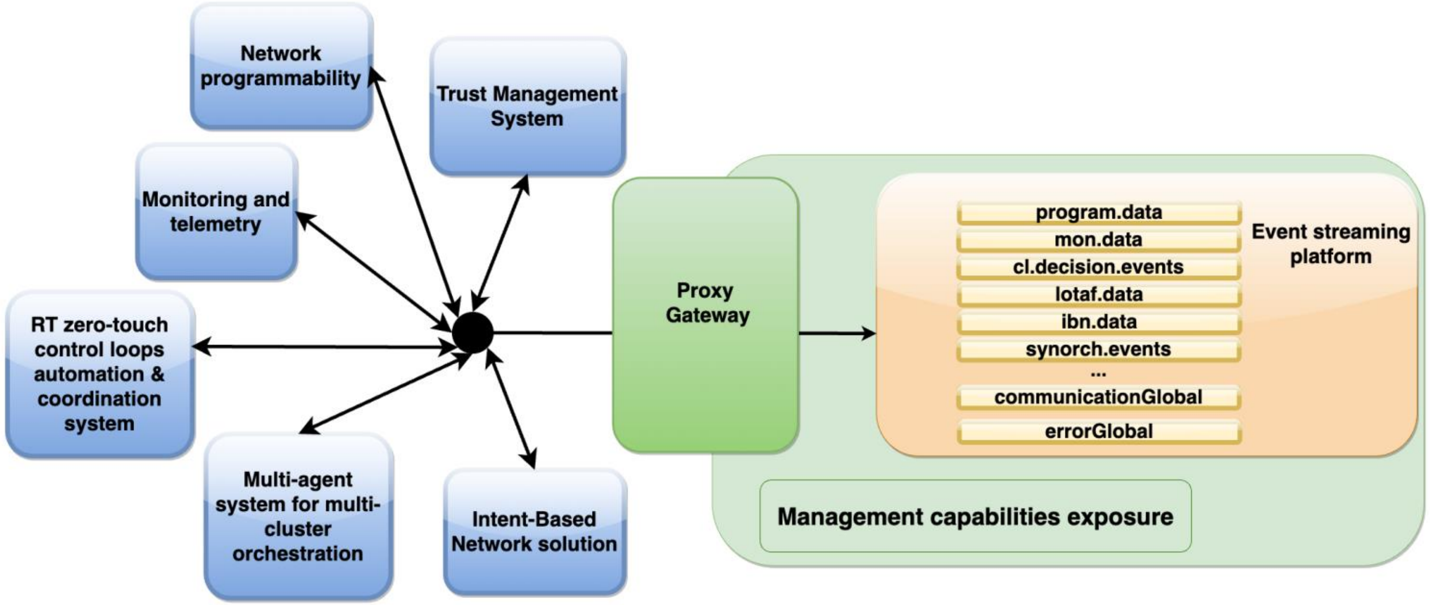
a) Highlights how the MCE facilitates the access at service-level of M&O mechanisms by providing real-time monitoring and telemetry across multiple network domains. This enables dynamic performance insights, supporting automated service tuning and proactive fault management using an event-driven architecture (EDA).

b) Enhances automation and orchestration processes within the management framework by enabling secure and efficient data exchange. This empowers components, such as the Real-time Zero-touch CLs functionality, the multi-agent system for multi-cluster orchestration, and the intent-based network solution targeted in the context of WP2, among others, to make informed decisions and dynamically adjust configurations for optimal performance and reliability.

c) Employs RBAC and secure APIs, ensuring that only authorised entities can perform critical operations such as service reconfiguration, fault recovery, and trust levels assessment, so ensuring reliability, high operational efficiency, and security.

Beyond the lab experiments, the implementation has been made publicly accessible within the PoC B.1 and will be integrated and evaluated as part of the project PoC C, contributing to demonstrate the following :

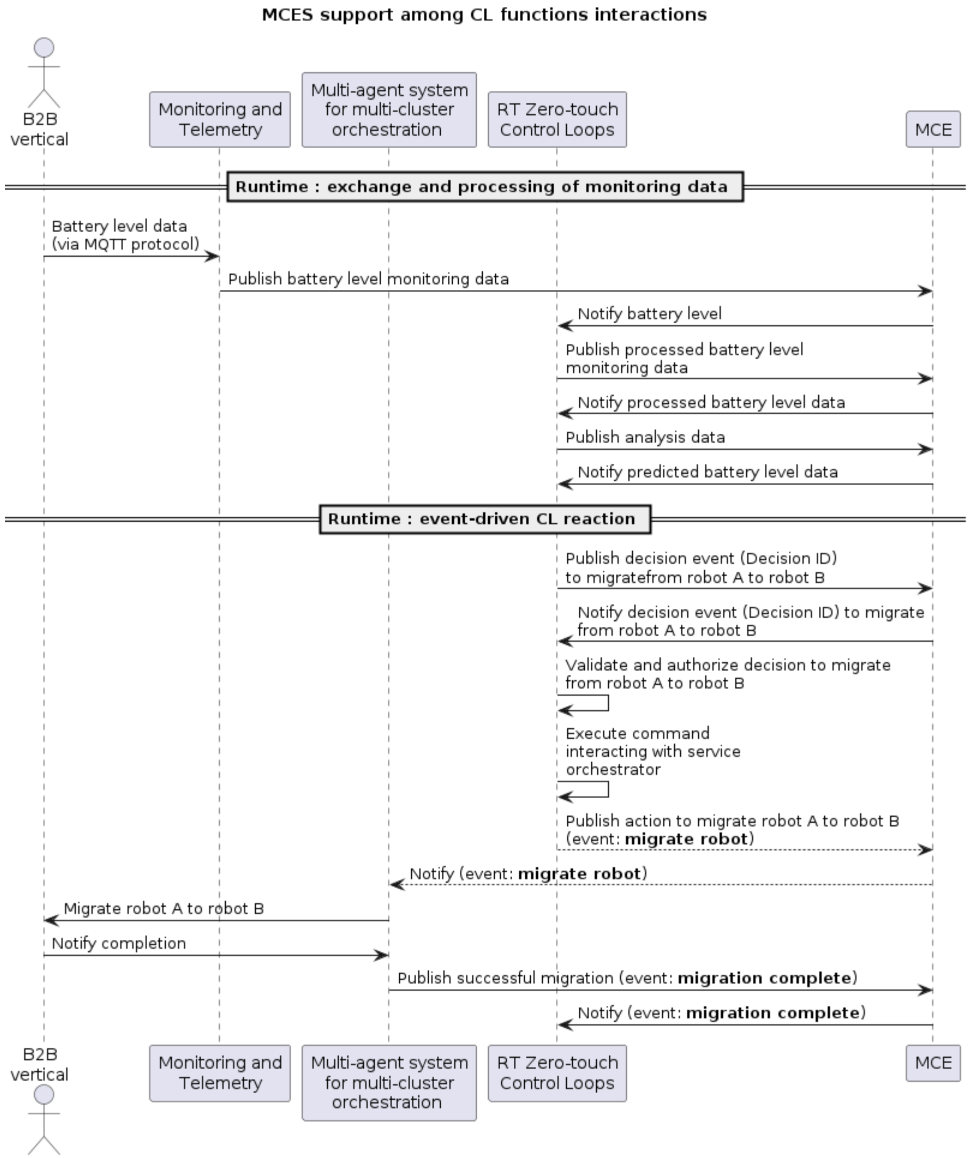
1. CL Functions Interaction with the MCE: Demonstrates how closed-loop functions can use the MCE for secure and real-time data exchange, optimising automated network operations, linked to industrial robots' operations.
2. Dynamic Fault Recovery in SDN using the MCE: Highlights how the MCE enables SDN environments to dynamically detect and recover from faults, enhancing network reliability.
3. Autonomous Service Recovery in Edge Networks Using the MCE: Explains how the MCE helps in facilitating autonomous service recovery in such edge networks, ensuring service continuity through intelligent and automated responses.



M&O framework components event driven interactions overview

Implementation 1: Usage of the Management Capabilities Exposure System - Use case 1

- This implementation is based on the management of certain service components deployed on a set of industrial robots.
- More specifically, the migration of certain service components is triggered based on the measurement of the robot’s battery level. For this, a CL monitors, analyse, and predict battery levels, and based on that, decides and executes the network service components migration when the battery level is envisaged to reach a certain threshold to ensure the continuous operation of the robots.
- The MCE plays a critical role by providing the continuous real-time monitoring data, collected by the Monitoring and Telemetry system, which is used for the analysis and the decision making to improve the efficiency of the robot’s operations. Particularly, the MCE provides the data bus that collects the data sent by the robots, such as the battery levels. This is used by the CL to predict when battery depletion might affect operations. That information is also broadcasted through the MCE, to be consumed by other components involved, which in turn trigger the network service migration requests for these robotic systems, ensuring that tasks are seamlessly transferred between robots based on their power availability. By automating this migration process, the MCE minimises downtimes and improves business continuity, demonstrating its value in managing dynamic, data-driven decisions in such environments.



MCES support among CL functions interactions

B2B vertical

Monitoring and Telemetry

Multi-agent system for multi-cluster orchestration

RT Zero-touch Control Loops

MCE

Runtime : exchange and processing of monitoring data

Battery level data (via MQTT protocol)

Publish battery level monitoring data

Notify battery level

Publish processed battery level monitoring data

Notify processed battery level data

Publish analysis data

Notify predicted battery level data

Runtime : event-driven CL reaction

Publish decision event (Decision ID) to migratefrom robot A to robot B

Notify decision event (Decision ID) to migrate from robot A to robot B

Validate and authorize decision to migrate from robot A to robot B

Execute command interacting with service orchestrator

Publish action to migrate robot A to robot B (event: **migrate robot**)

Notify (event: **migrate robot**)

Migrate robot A to robot B

Notify completion

Publish successful migration (event: **migration complete**)

Notify (event: **migration complete**)

B2B vertical

Monitoring and Telemetry

Multi-agent system for multi-cluster orchestration

RT Zero-touch Control Loops

MCE

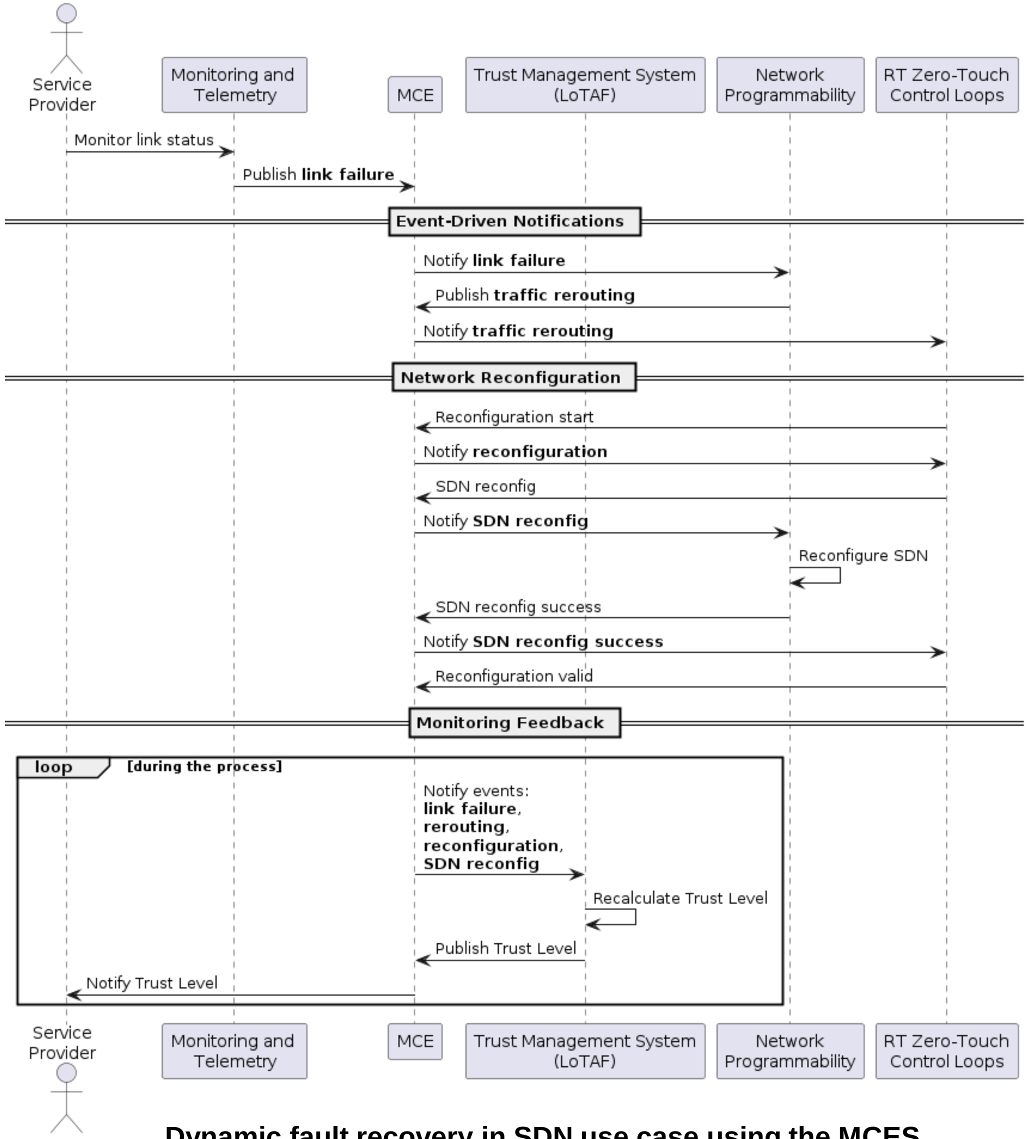
CL functions interactions with the MCE

Implementation 1: Usage of the Management Capabilities Exposure System - Use case 2

- This second implementation is about dynamic fault recovery in SDN, targeting to ensure the real-time detection and the automated recovery of network link failures.

- More specifically, the implemented key processes involve detecting the failure, adjusting trust levels (i.e., recalculate the Level of Trust based on TLAs and collected data), reconfiguring the network, and notifying the involved stakeholders of the successful recovery, all while maintaining service continuity through MCE interactions.

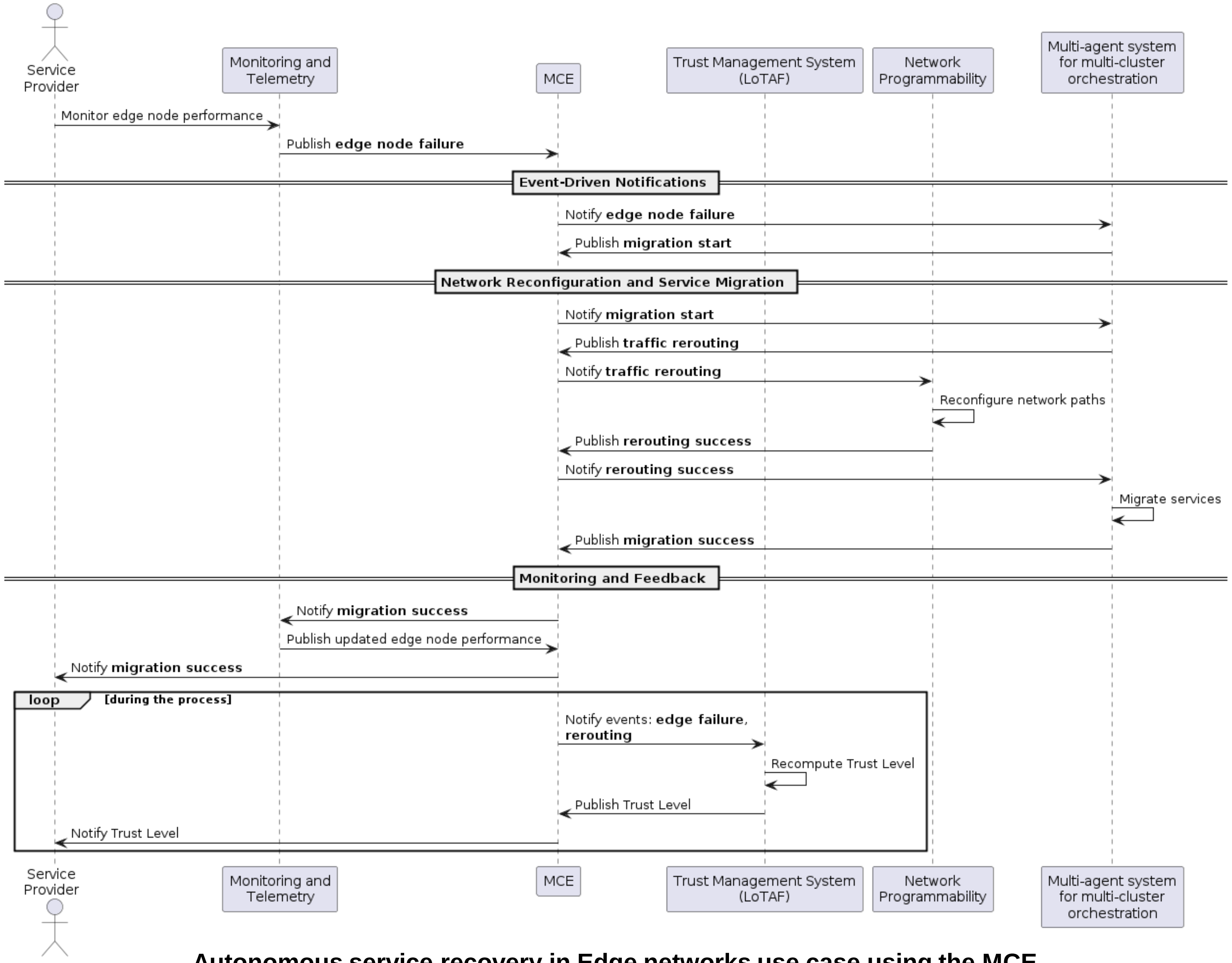
- For this implementation, the MCE is essential for ensuring network resilience. Here, MCE enables rapid identification and automated handling of link failures by continuously exposing monitoring network status in real time and broadcasting the alerts to trigger rerouting and reconfiguration processes. Working alongside other enablers, such as the Trust Management System and the Network Programmability system, it efficiently contributes to the reconfiguration of the SDN when faults occur, minimising disruptions. Through its integration with the Zero-touch CLs enabler, the MCE also supports a seamless and autonomous fault-recovery mechanism, improving the network's ability to self-heal and maintaining the service trust level.



Dynamic fault recovery in SDN use case using the MCES

Implementation 1: Usage of the Management Capabilities Exposure System - Use case 3

- This third implementation targets an edge node failure recovery, considering the real-time failure detection, the migration of network service components deployed on the failing node, and the node recovery.
- The MCE play an important role here for maintaining network stability and reliability. By continuously exposing edge node performance data, the MCE can help detecting failures by providing real-time information to the Network Programmability system, which triggers the automated recovery processes. When an edge node failure is detected, the MCE works with other components to trigger event-driven notifications and manages the service migration to healthy nodes. This ensures that traffic is dynamically re-routed, and that network resources are reallocated to prevent service disruptions. In addition, by interacting with the multi-agent system orchestration solution and the Trust management system, the MCE helps to successfully acknowledge these recovery actions and adjust trust levels accordingly, helping to maintain operational continuity and increasing the trustworthiness in autonomous edge environments.



Service Provider

Monitoring and Telemetry

MCE

Trust Management System (LoTAF)

Network Programmability

Multi-agent system for multi-cluster orchestration

Monitor edge node performance

Publish **edge node failure**

Event-Driven Notifications

Notify **edge node failure**

Publish **migration start**

Network Reconfiguration and Service Migration

Notify **migration start**

Publish **traffic rerouting**

Notify **traffic rerouting**

Reconfigure network paths

Publish **rerouting success**

Notify **rerouting success**

Migrate services

Publish **migration success**

Monitoring and Feedback

Notify **migration success**

Publish updated edge node performance

Notify **migration success**

loop **[during the process]**

Notify events: **edge failure**, **rerouting**

Recompute Trust Level

Publish Trust Level

Notify Trust Level

Autonomous service recovery in Edge networks use case using the MCE

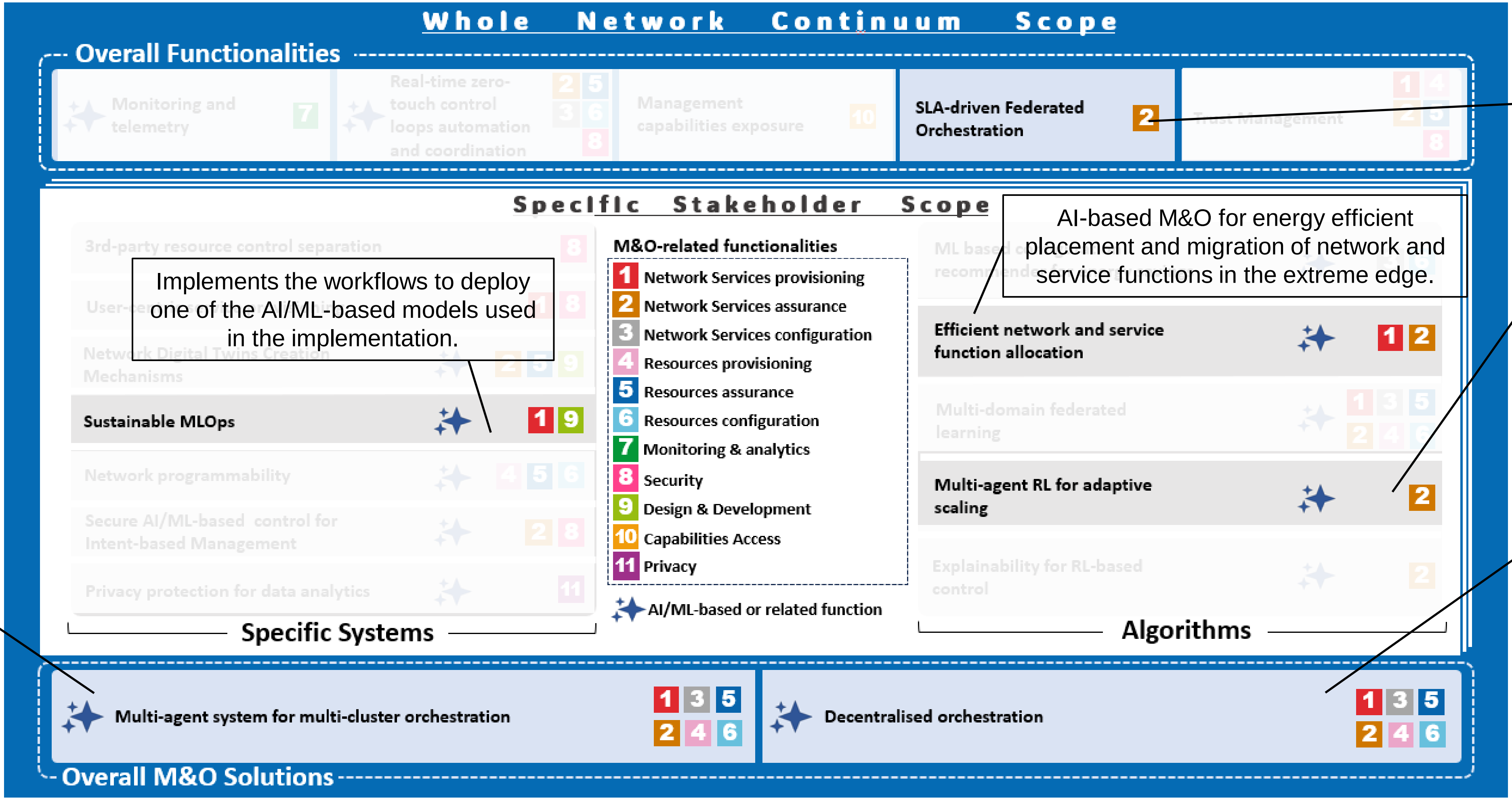
Implementation 1: Analysis of KPIs

In the table below, an overview of the KPIs improvements that the usage of MCE brings in the previous presented implementations. Further analysis and more detailed information about the metrics used to measure these KPIs will be reported in the final project PoC. These numerical values will be showcased in the context of last WP2 deliverable (D2.6), associated to the final project PoC.

KPI	Industrial Robots Service Migration	Dynamic Fault Recovery in SDN	Autonomous Service Recovery in Edge Networks	Metrics
Scalability	MCE scales with additional robots by managing battery data for each.	-	-	Kafka metrics (throughput, partition distribution, consumer lag)
Latency	Low-latency data transmission ensures prompt migration based on battery levels.	Real-time data exposure supports fast fault detection and reconfiguration.	MCE triggers real-time recovery actions, minimising downtime.	E2E latency communication from message production to consumption. Latency related to the onboarding of a new component
Reliability	MCES reliability affects predictive data broadcast. Timely migration reduce downtime from battery depletion.	MCES reliability affect automated detection and recovery enhance network reliability.	MCE reliability affects reliable service continuity. Because data broadcasted helps in migrating services during edge node failures	Consumer lag during failures (track consumer lag if brokers fail),replication count, producer and consumer error rate, metrics of data persistence stored.
Maintainability	-	MCE automates fault alerts and reconfigurations, reducing manual intervention for SDN faults and easing maintenance tasks. The EDA implicitly improves maintenance and upgrades by decoupling the services.	MCE provides real-time failure notifications, automating service migration and minimising manual troubleshooting in edge environments. EDA implicitly improve maintenance and upgrade by decoupling the services	Mean Time to Resolution (MTTR), Manual Intervention Frequency, Automated Recovery Success Rate.
Availability	Continuous monitoring enables proactive service migration, enhancing uptime. EDA with Kafka ensures redundancy and resilience by distributing events across multiple nodes or consumers. This helps maintain service continuity even if individual components fail.	Automated rerouting and fault recovery support high availability, thanks to the information broadcasted by the MCE. EDA with Kafka ensures redundancy and resilience by distributing events across multiple nodes or consumers. This helps maintain service continuity even if individual components fail.	Automated service migration and resource reallocation prevent service disruptions thanks to the information broadcasted by the MCE. EDA with Kafka ensures redundancy and resilience by distributing events across multiple nodes or consumers. This helps maintain service continuity even if individual components fail.	Uptime Percentage, Mean Time to Recovery (MTTR), Incident Response Time.

Implementation 2: Orchestration on the network continuum - Introduction (1)

The purpose of this implementation is to showcase the usage of different components of the framework targeting the orchestration on the network continuum. The following figure shows the different enablers involved, along with the functionality they perform in this specific implementation:



Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party resource control separation 8
User-ce 1 8
Network Digital Twins Creation Mechanisms 2 5 9
Sustainable MLOps 1 9
Network programmability 4 5 6
Secure AI/ML-based control for Intent-based Management 2 8
Privacy protection for data analytics 11
Specific Systems
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based c recommende
Efficient network and service function allocation 1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

Enables blockchain-based SLA management.

Implements the workflows to deploy one of the AI/ML-based models used in the implementation.

AI-based M&O for energy efficient placement and migration of network and service functions in the extreme edge.

Supports AI-assisted scaling in partially observable environments

Guides the hierarchical multi-cluster orchestration coordination between monitoring, decision making, and deployment technologies.

Implements proactive forecasting actions based on the infrastructure resources status.

Implementation 2: Orchestration on the network continuum - Introduction (2)

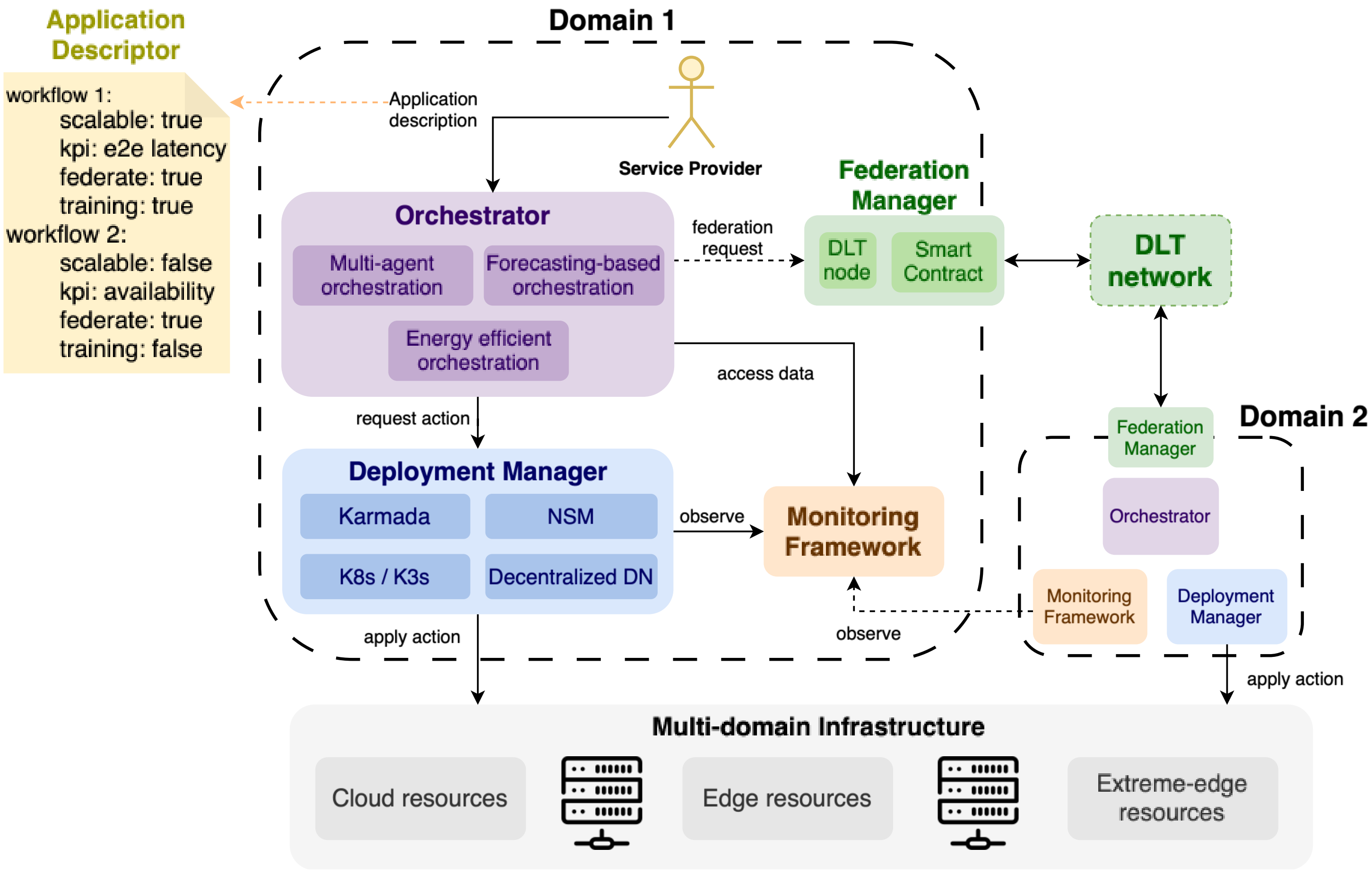
The implementation targets the deployment and continuous operation of service components through:

i. Multi-agent Reinforcement Learning algorithms (based service autoscaling and migration).
ii. Service components proactive migration based on the infrastructure availability status, targeting the extreme-edge domain integration.
iii. Deployment of required AI/ML-model using the Sustainable MLOps enabler.
iv. Service federation to support deployment in multiple domains.
v. Service trust management to support trusted execution in multiple clusters.

This implementation is part of the project Component PoC B.1 [HEX224-D24].

Implementation main steps:

1. Service description and registration.
2. Service mapping to orchestration mechanisms.
3. Service deployment over multi-cluster resources in the continuum.
4. Continuous service lifecycle management (monitoring and scaling actions).
5. Federated mechanisms for multi-domain management.

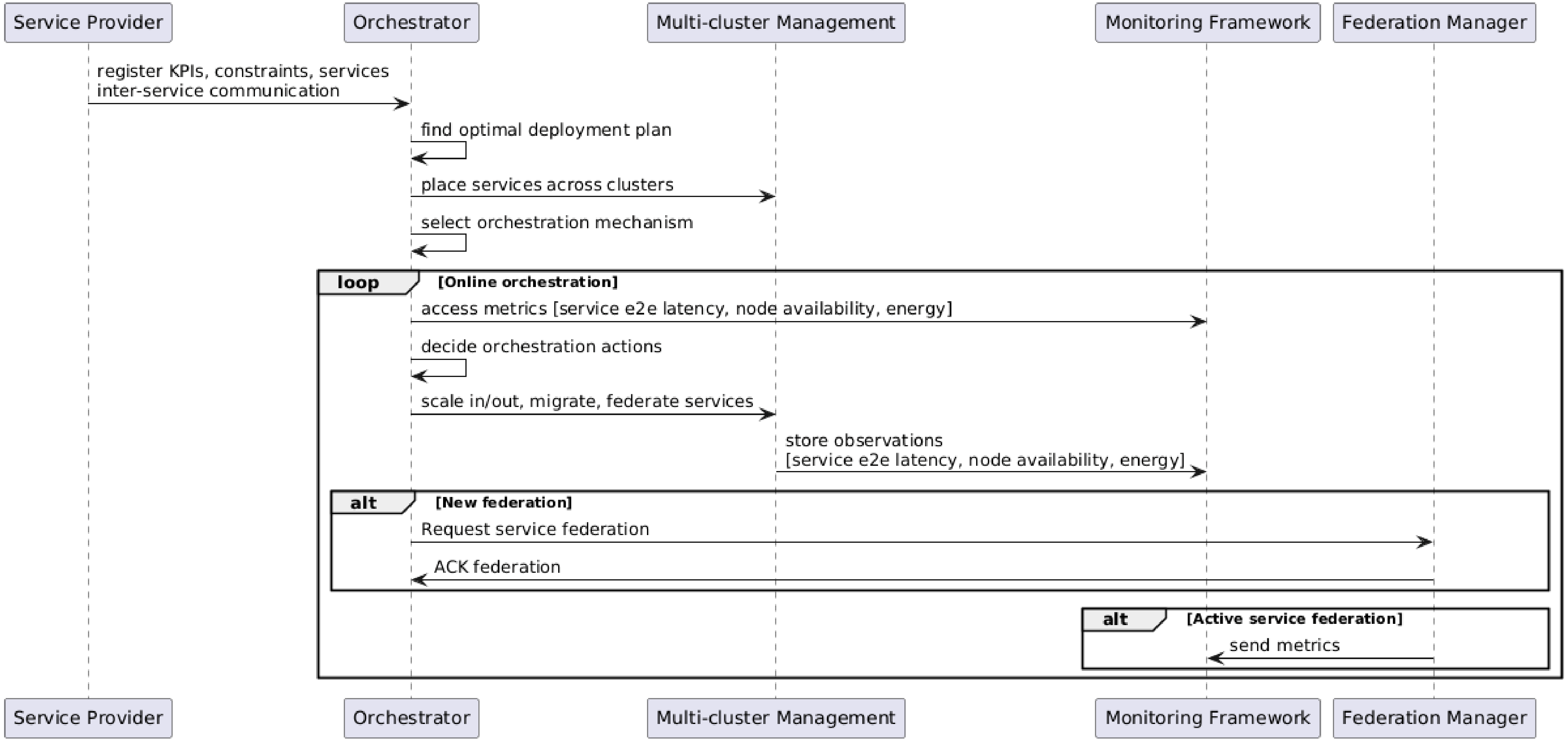


Application Descriptor

```
workflow 1:
        scalable: true
        kpi: e2e latency
        federate: true
        training: true
workflow 2:
        scalable: false
        kpi: availability
        federate: true
        training: false
```

Orchestration mechanisms over resources in the continuum

Implementation 2: Orchestration on the network continuum - Workflow



Service Provider

Orchestrator

Multi-cluster Management

Monitoring Framework

Federation Manager

register KPIs, constraints, services
inter-service communication

find optimal deployment plan

place services across clusters

select orchestration mechanism

loop [Online orchestration]

access metrics [service e2e latency, node availability, energy]

decide orchestration actions

scale in/out, migrate, federate services

store observations
[service e2e latency, node availability, energy]

alt [New federation]

Request service federation

ACK federation

alt [Active service federation]

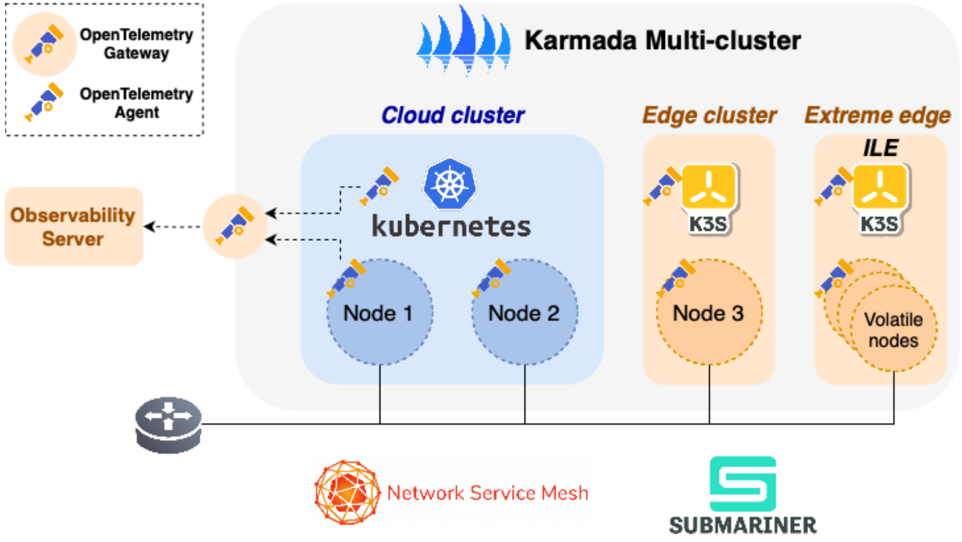
send metrics

Network service lifecycle management across multiple domains

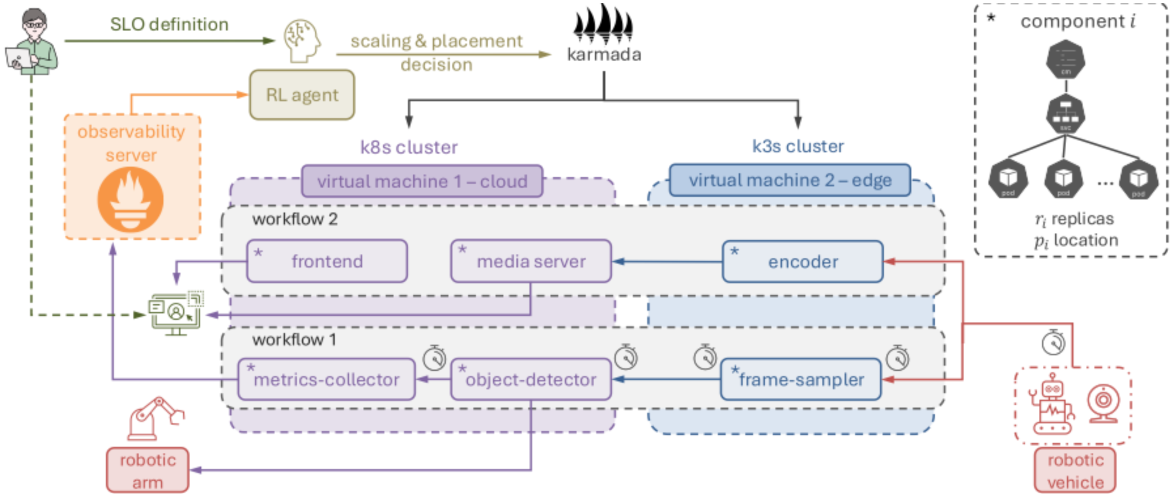
- The service provider registers the service's requirements and constraints.
- The orchestrator decides the initial deployment plan and instructs the multi-cluster manager to place the microservices across the clusters.
- The orchestrator maps microservices to appropriate orchestration mechanisms.
- Online orchestration loop: decide orchestration actions based on observed metrics.
- If necessary (i.e., deployment across multiple domains), request service federation.

Implementation 2: Orchestration on the network continuum - Experiment and results

- A real edge-cloud multi-cluster experimentation setup was used in coordination with an Infrastructure Layer Emulator (ILE), emulating the network extreme-edge domain. This ILE has been specifically developed for the Hexa-X-II project and released to the Open-Source community*.
- Regarding the deployed network service, a latency-sensitive service was used for evaluating the developed mechanisms. This service has also been specifically developed for the project and released as Open-Source**.
- Main relevant KPIs: Scalability (stable performance for up to 20 agents), Latency (< 22ms E2E latency), Service Creation Time (< 1 minute), Automation (> 10% efficiency compared to HPA), Elasticity (> 10% efficiency compared to HPA).
- Results showcase:
  - How joint decision making for service scaling and migration by multiple agents can make latency-focused guarantees.
  - How availability forecasting-based migration can guarantee streaming continuous operation.
  - How AI/ML-models can be deployed using the S-MLOps asset.
  - How service federation can exploit multi-domain resources.
  - How the usage of open-source APIs can implement resource federation in the computing continuum.



Multi-cluster Testbed



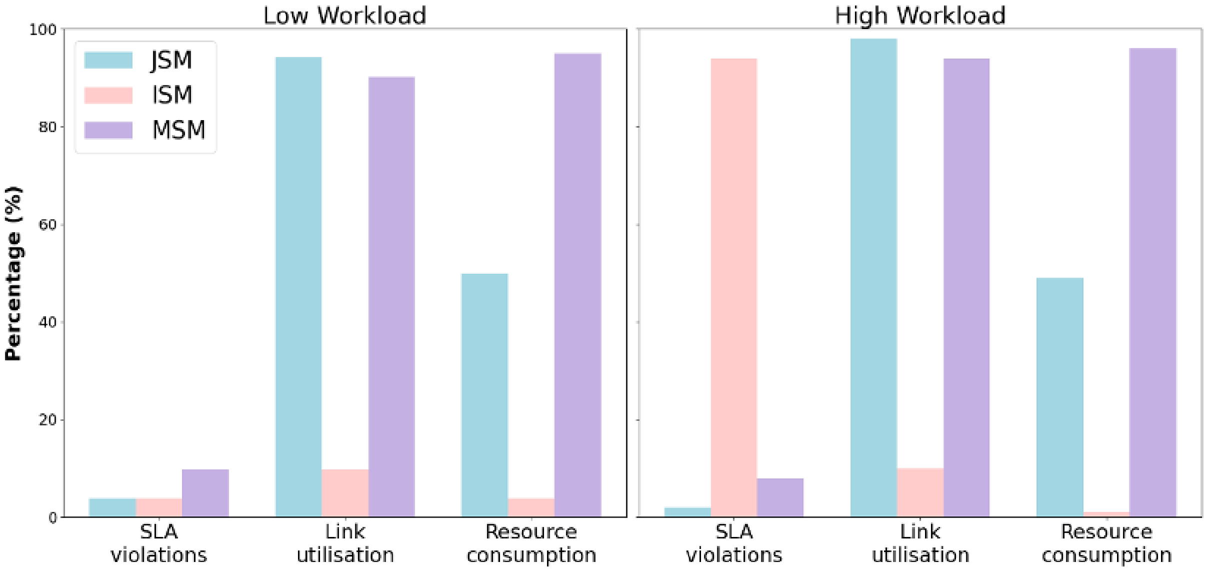
6G latency-sensitive service

* https://gitlab.com/decentralized-continuum-orchestration/infrastructure-layer-emulator

** https://gitlab.com/netmode/6g-latency-sensitive-service

Implementation 2 - Evaluation Results

- **Hierarchical Mechanism Implementation:** A MARL-based approach is used where Agent 1 handles service migration, and Agent 2 manages scaling.
- **Component PoC Evaluation:** Different agent networks are assessed for their ability to handle tasks collaboratively or independently. Three different agent setups are used:
    - Joint Scaling - Migration (JSM)
    - Independent Scaling - Migration (ISM)
    - Mixed Scaling - Migration (MSM)

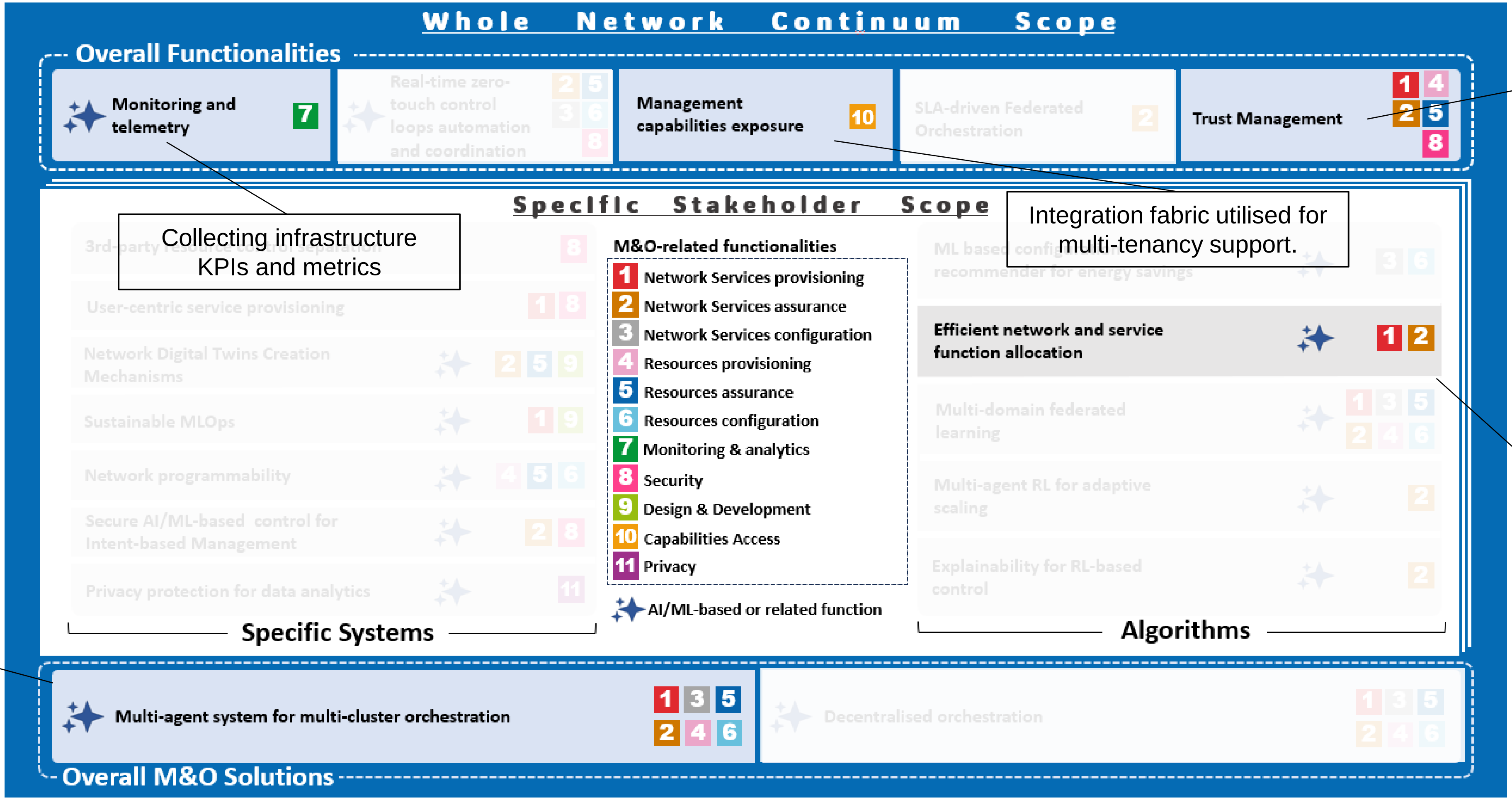


Multi-agent settings comparison

- **JSM** meets the SLA requirements for both low (1 frame/sec) and high (10 frames/sec) workloads, at the cost of high link utilisation and medium to high resource consumption.
- **ISM** demonstrates minimal link utilisation and resource consumption but fails to satisfy SLA requirements for high workloads, as its agents operate independently.
- **MSM** introduces collaboration between agents, reducing SLA violations but increasing link utilisation and resource consumption, even for low workloads (1 frame/sec).

Implementation 3: Functionality allocation in a cobot-powered warehouse inventory management (1/3)

The purpose of this implementation is to showcase the usage of certain components of the management framework in a cobot-powered warehouse inventory use case. The following figure shows the different enablers involved, along with the functionality they perform in the implementation:



Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party ... 8
User-centric service provisioning 1 8
Network Digital Twins Creation Mechanisms 2 5 9
Sustainable MLOps 1 9
Network programmability 4 5 6
Secure AI/ML-based control for Intent-based Management 2 8
Privacy protection for data analytics 11
Specific Systems
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based configuration recommender for energy savings 3 6
Efficient network and service function allocation 1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

Collecting infrastructure KPIs and metrics

Integration fabric utilised for multi-tenancy support.

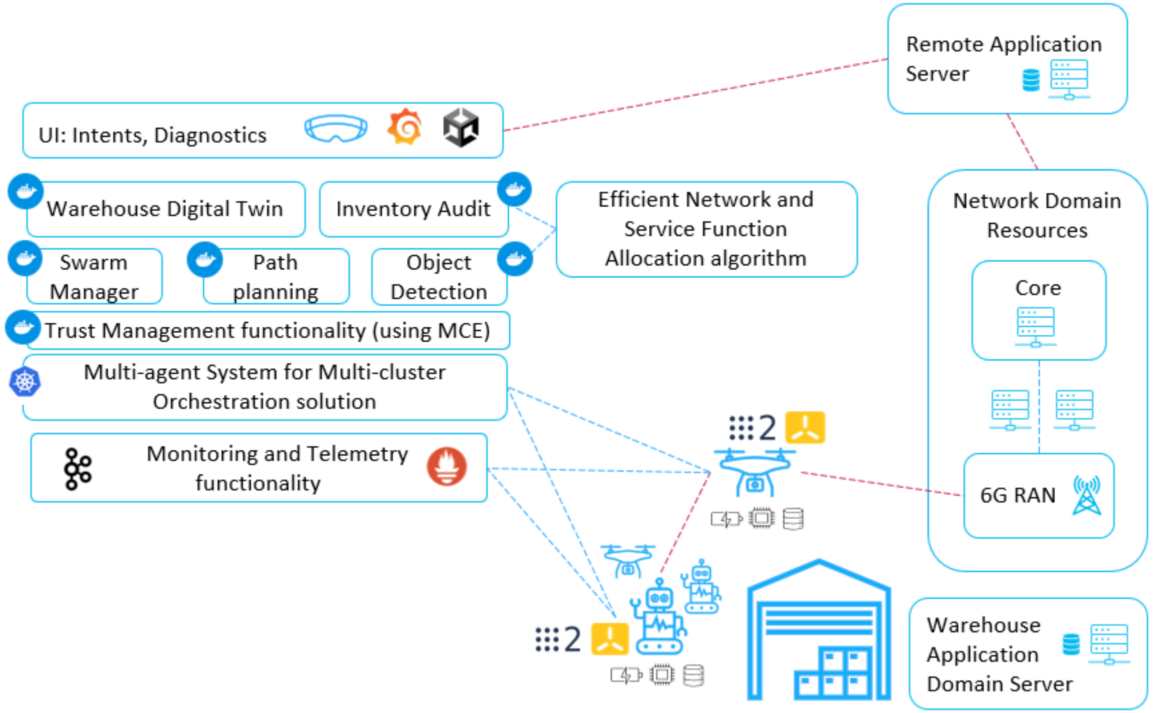
Trust management functionality includes the TEF and LoTAF and evaluates the trustworthiness index per entity of interest in the specific scenario and is considered when functionality placement occurs.

Functionality allocation joint optimisation mechanism which considers computing energy, and transmission energy. For computational workload placement and physical task planning.

Orchestration process of functionality allocation mechanism is analysed for ensuring the successful deployment of network services and applications in the continuum

Implementation 3: Functionality allocation in a cobot-powered warehouse inventory management (2/3)

- The implementation targets to provide an automated inventory management solution for accurate and efficient warehousing operations using collaborative robots. Overall, the deployed service is used to check a warehouse inventory, ensuring that stored items match the records in the system, so that records are always accurate and reliable. Specifically, the implementation showcases the management mechanisms associated to this service.
- The implementation integrates a functionality allocation M&O component, responsible for the optimal placement and planning of computational and physical workloads and robotic tasks towards energy efficiency and trustworthiness. Key services, workloads, and tasks considered are object detection, path planning, inventory management, quality inspection, and the creation of the warehouse digital twin.
- The implementation has been integrated as part of the System-PoCs A & B in a testbed including the cobots, consisting in unmanned aerial vehicles (UAVs) and autonomous mobile robots (AMRs).


Remote Application Server
UI: Intents, Diagnostics
Warehouse Digital Twin
Inventory Audit
Efficient Network and Service Function Allocation algorithm
Network Domain Resources
Swarm Manager
Path planning
Object Detection
Core
Trust Management functionality (using MCE)
Multi-agent System for Multi-cluster Orchestration solution
2
Monitoring and Telemetry functionality
6G RAN
2
Warehouse Application Domain Server

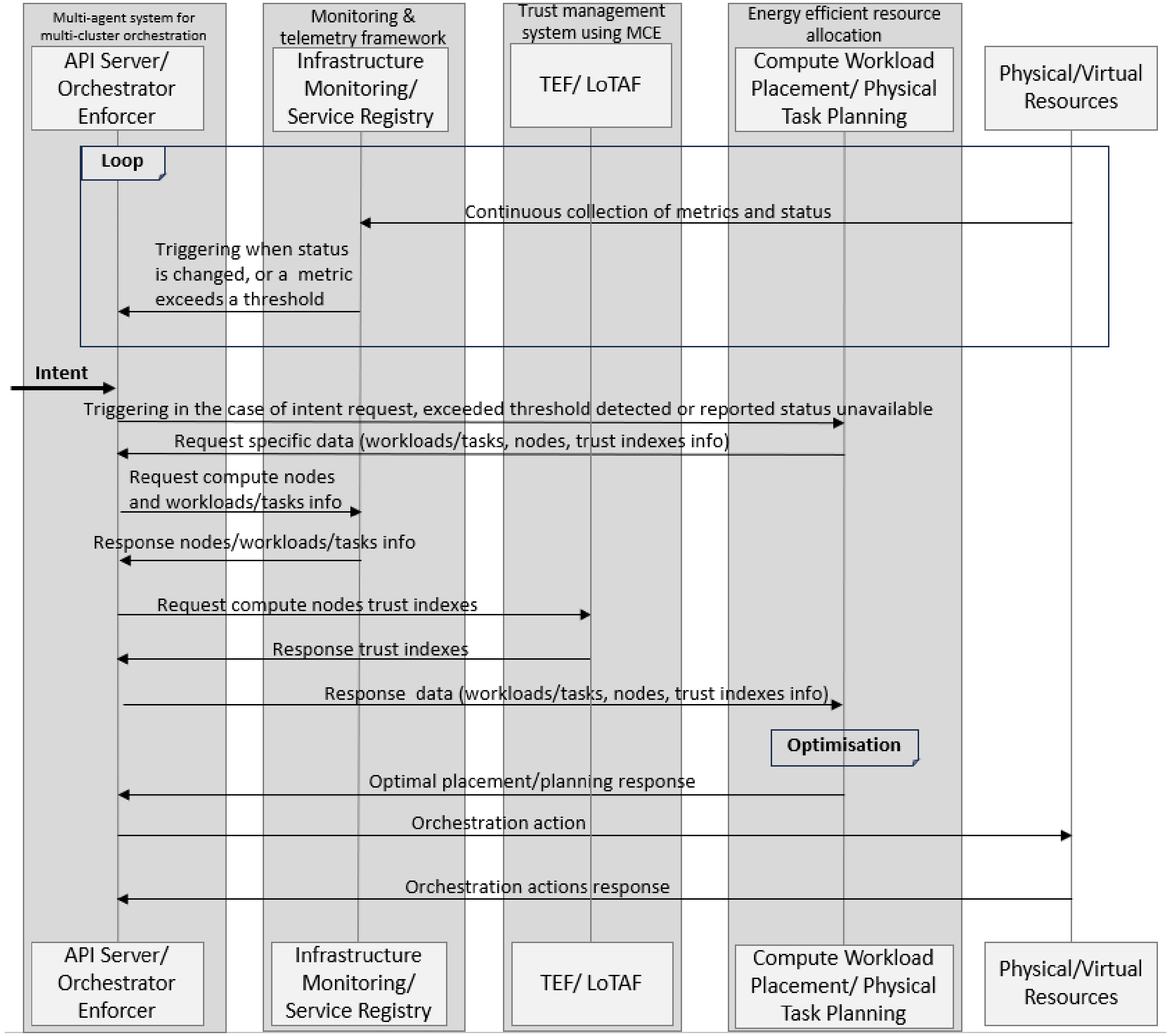
Representation of the implementation showing the inventory management-specific services (e.g., object detection, path planning), the user interfaces, the functionality allocation mechanism, and the network domain resources.

Main evaluation results and outcomes obtained were compared to the round-robin load balancing approach, a simulation scenario of 50 fixed compute nodes and increasing number of computational workloads. The reference and target values can be found in [HEX224-D63], with the following main results:

- Up to 50.9% energy consumption gains, related to the energy efficiency KPI.
- Up to 43% increase of trustworthiness (computed as the sum of the trust indexes of the compute nodes utilised for the placement, which are expressed as the weighted sum of relevant KPI values, such as availability, reliability, and security, among others).

Implementation 3: Functionality allocation in a cobot-powered warehouse inventory management (3/3)

Main Workflow



Multi-agent system for multi-cluster orchestration

API Server/ Orchestrator Enforcer

Monitoring & telemetry framework

Infrastructure Monitoring/ Service Registry

Trust management system using MCE

TEF/ LoTAF

Energy efficient resource allocation

Compute Workload Placement/ Physical Task Planning

Physical/Virtual Resources

Loop

Continuous collection of metrics and status

Triggering when status is changed, or a metric exceeds a threshold

Intent

Triggering in the case of intent request, exceeded threshold detected or reported status unavailable

Request specific data (workloads/tasks, nodes, trust indexes info)

Request compute nodes and workloads/tasks info

Response nodes/workloads/tasks info

Request compute nodes trust indexes

Response trust indexes

Response data (workloads/tasks, nodes, trust indexes info)

Optimisation

Optimal placement/planning response

Orchestration action

Orchestration actions response

API Server/ Orchestrator Enforcer

Infrastructure Monitoring/ Service Registry

TEF/ LoTAF

Compute Workload Placement/ Physical Task Planning

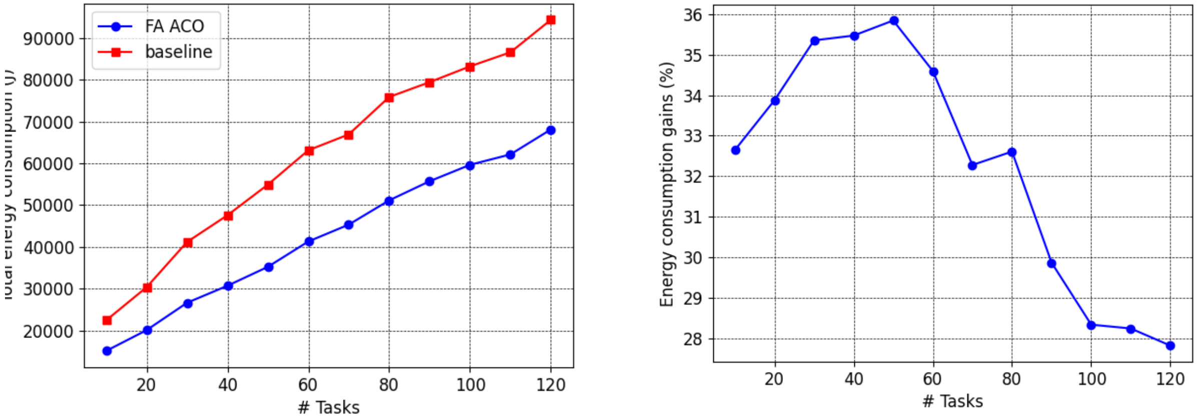
Physical/Virtual Resources

Workflow of the cobot-powered warehouse inventory management implementation.

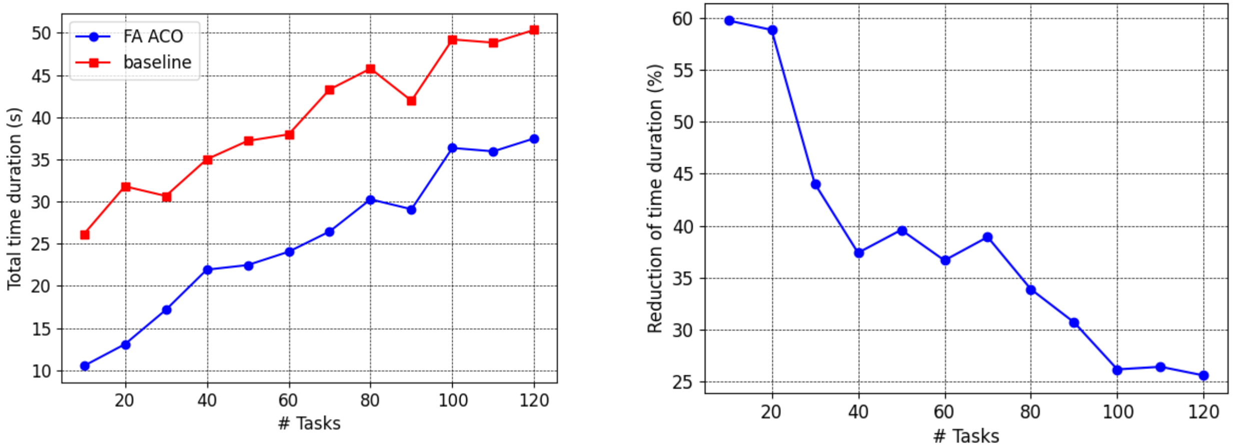
The interacting components are:

- The API server (related to the multi-agent system for multi-cluster orchestration enabler). This is the interface orchestration manager component responsible for checking the feasibility of the various requests and responds coming from the various components, receiving intents/requests, among others.
- Infrastructure monitoring and Service Registry (related to the monitoring and telemetry functionality). The former collects KPIs and information related to the status of the physical and virtual resources, checking also that some thresholds are respected. The latter stores the information and requirements of the various computational workloads and tasks.
- The Trust Management functionality. It evaluates the trustworthiness index per entity of interest in the specific scenario been considered when functionality placement occurs.
- Functionality allocation (related to the efficient network and service function allocation algorithms). It computes the workload placement and the physical task planning towards energy efficiency and trustworthiness.

Implementation 3 – Evaluation Results



Total (left) energy consumption and the gains (right) using the physical task planning algorithm based on ACO algorithm compared to the nearest neighbour heuristic (baseline).



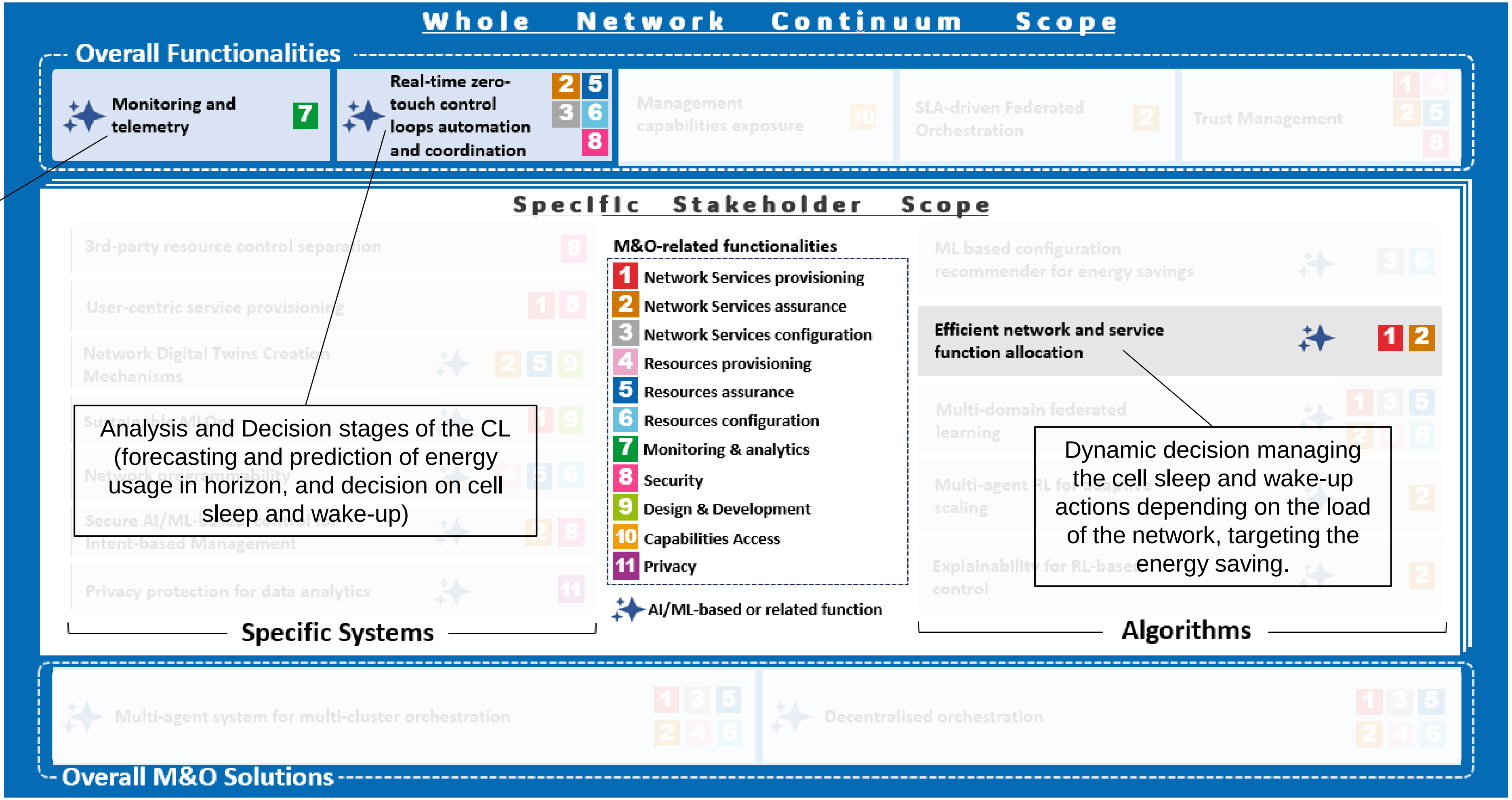
Total (left) duration time and the reduction (right) using the physical task planning algorithm based on ACO algorithm compared to the nearest neighbour heuristic (baseline).

- Gains of energy consumption and total duration time of completing all tasks.
- The baseline algorithm for these measurements was the **nearest neighbour heuristic** and is compared with the developed physical task planning-functionality allocation algorithm **based on Ant Colony Optimisation** (ACO) algorithm.
  - The validation scenario comprised of a fixed number of virtualised AMRs and UAVs (= 30) having various capabilities and increasing number of physical tasks (computer vision tasks) in varied warehouse locations.
- **Observations**:
  - **Up to 35.9% energy consumption gains and up to 60% reduction in duration time .**
  - The gains gradually decrease as the number of tasks increases, due to the **scarcity of robots/resources** which resulted in **lower gain potential**.

Implementation 4: ML based configuration recommendation for energy savings (1/3)

The main purpose of this implementation is to showcase the usage of the ML-based configuration recommendation algorithm for energy savings, which is part of the management framework. To execute this algorithm, the implementation also relies on the Monitoring and Telemetry and the Real-time Zero Touch CLs functionalities in the framework, as represented in the figure below:

Integrates a CL that can utilise Monitoring and Telemetry capability to observe real-time data (network metrics and KPIs)



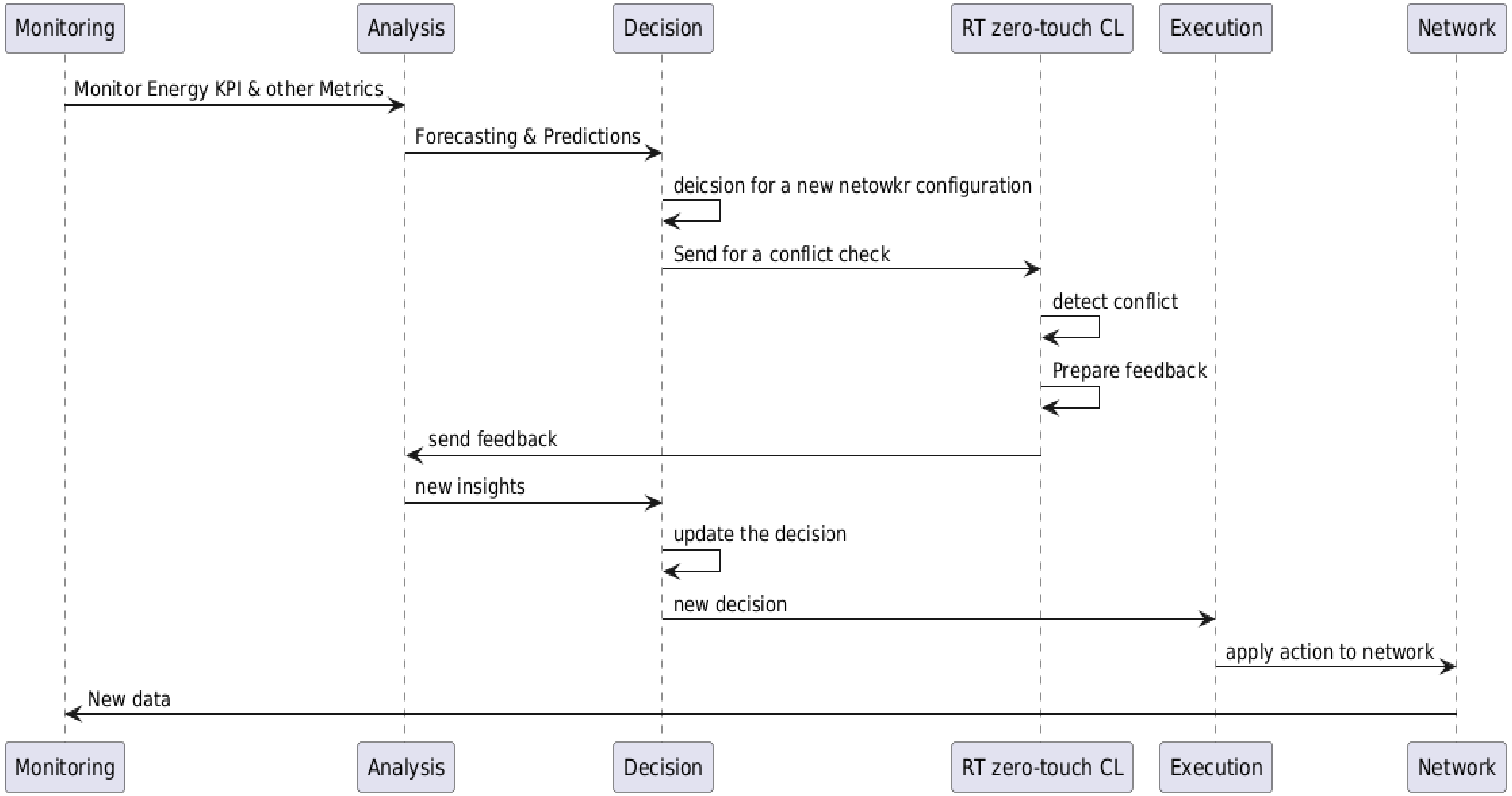
Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party resource control separation 8
User-centric service provisioning 1 8
Network Digital Twins Creation Mechanisms 2 5 9
Sustainable MLOps 1 9
Network programmability 4 5 6
Secure AI/ML-based control for Intent-based Management 2 8
Privacy protection for data analytics 11
Specific Systems
M&O-related functionalities
1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based configuration recommender for energy savings 3 6
Efficient network and service function allocation 1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

Analysis and Decision stages of the CL (forecasting and prediction of energy usage in horizon, and decision on cell sleep and wake-up)

Dynamic decision managing the cell sleep and wake-up actions depending on the load of the network, targeting the energy saving.

Implementation 4: ML based configuration recommendation for energy savings (2/3)

Main Workflow



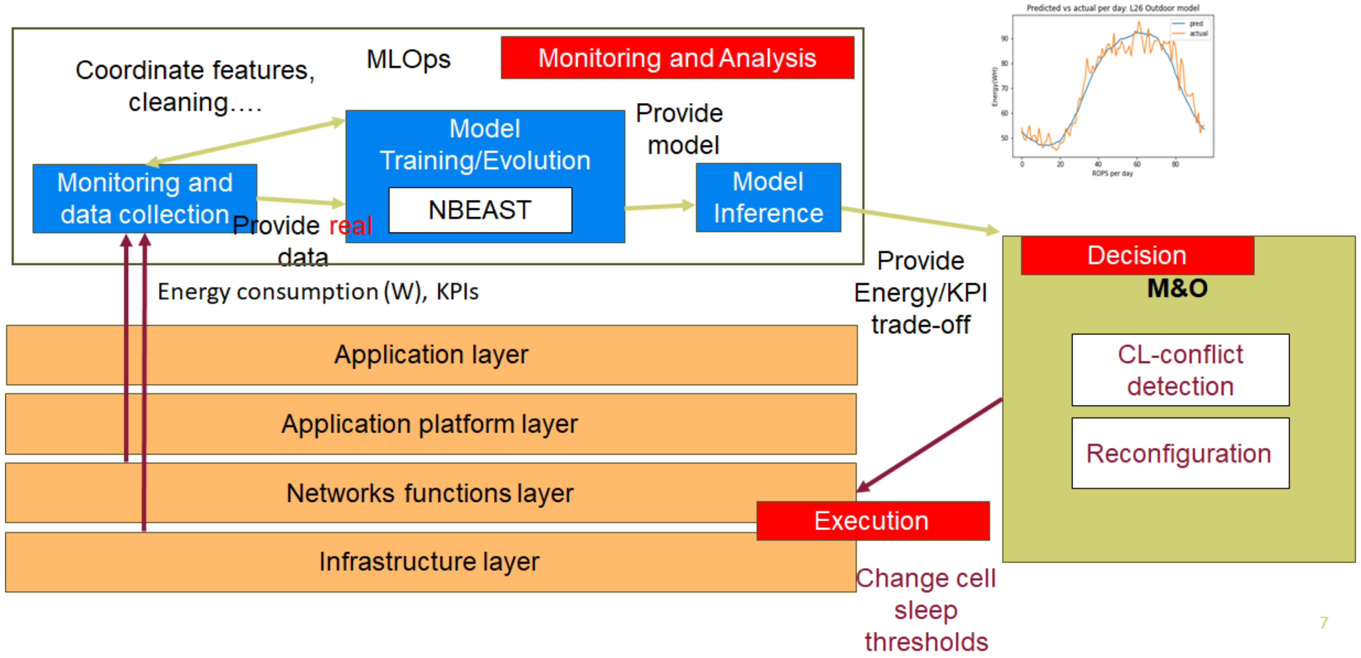
As it can be appreciated, this implementation deploys a fully automated CL for correcting flow misconfigurations in a scalable way in deterministic networks. This is done by exploring trade-offs between different forms of telemetry – including in-band telemetry (INT) – and applying sketching methods for lightweight analysis and detection. The algorithm takes autonomous decisions for cell sleep and wake-up.

The Monitoring and Real-time zero touch CL automation functionalities in the framework are utilised here to:

- Monitor the relevant KPIs: energy, throughput, latency, etc.
- Analyse the situation and make short-term forecasts.
- Decision based on the forecasts are done, making the cells to sleep or wake-up.
- In case the sleep decision could cause a significant conflict in the network, it can be postponed, so that the cell remains in wake-up mode.

Implementation 4: ML based configuration recommendation for energy savings (3/3)

- For the initial testing and implementation, real data from a Tier-1 operator were used.
- The implementation has been integrated as part of the project Component PoC B.1 using the mentioned real network and monitoring relevant data (energy usage, throughput, latency, and traffic).
- From the infrastructure layer, fetch of the relevant KPIs and make an analysis with forecasting the possible energy saving in the future. The decision on the cell status is done at M&O level, and the final action is executed on the network.



High level view of ML based configuration recommendation for energy saving.

Implementation 4 – Evaluation Results

- The dynamic threshold at cell level at which a capacity cell is awakened or put to sleep based on Radio Resource Control (RRC) Connections and PRB Utilisation. The following results are achieved:

  - Achieved a **10-12%** reduction in energy consumption across pilot sites compared to a static method relying on fixed thresholds.

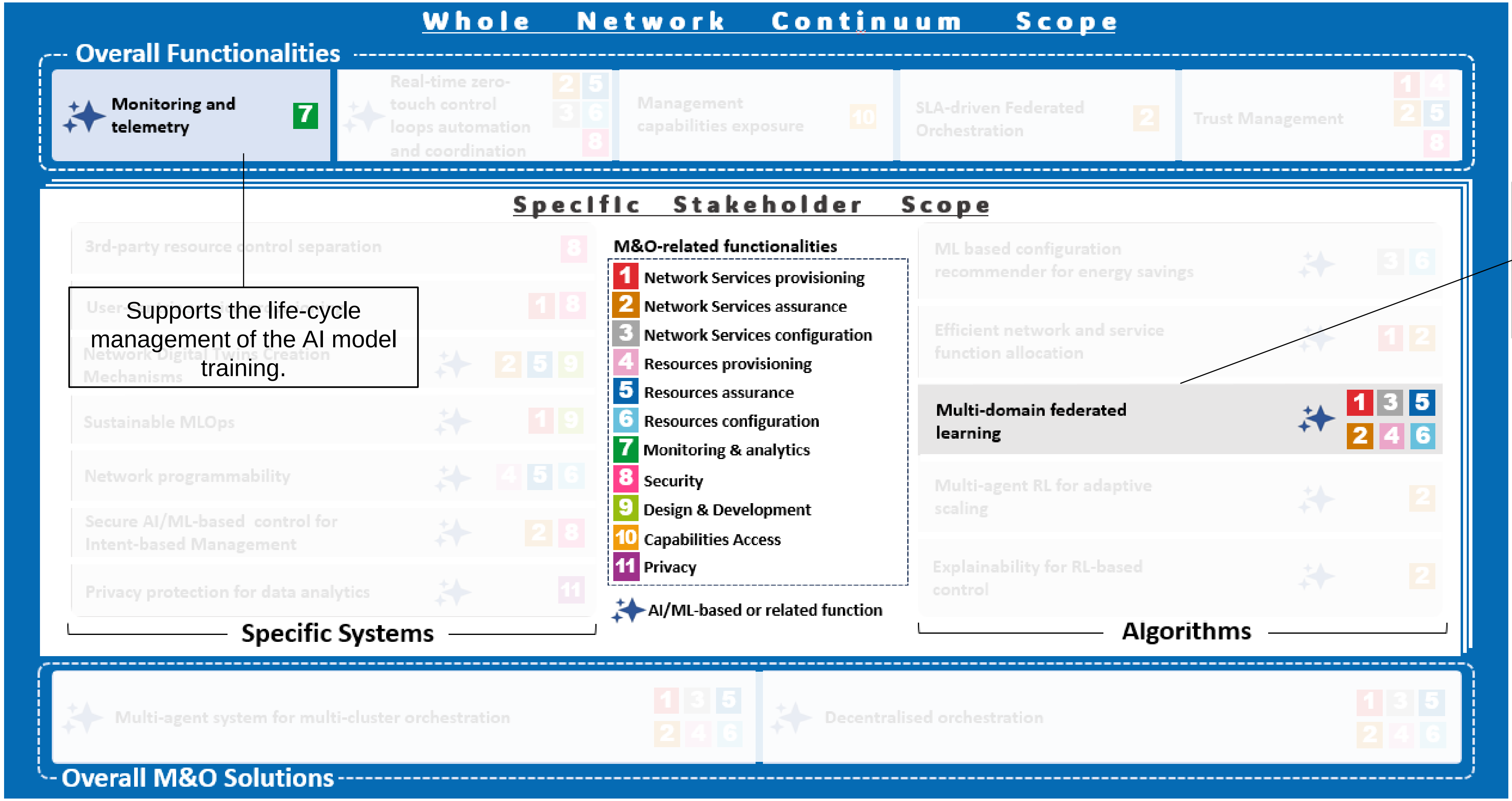
  - Uncompromised Performance: No degradation observed RRC, Evolved Radio Access Bearer (ERAB) success rates, or call drop rates across all frequency bands.

  - Consistent Metrics: Maintained stable traffic volume, mobility success rates, and latency.

  - Steady Throughput: Downlink and uplink throughput, along with other primary KPIs, remained consistent with historical trends.

Implementation 5: Resource assignment for federated learning (1/3)

This implementation focuses on showcasing the functionality of the Multi-domain Federated Learning algorithm highlighted in the figure below. As it can be appreciated, in also makes use of the Monitoring and Telemetry functionality in the management framework.



Whole Network Continuum Scope

Overall Functionalities

Monitoring and telemetry 7

Real-time zero-touch control loops automation and coordination 2 5 3 6 8

Management capabilities exposure 10

SLA-driven Federated Orchestration 2

Trust Management 1 4 2 5 8

Specific Stakeholder Scope

3rd-party resource control separation 8

User-… 1 8

Network Digital Twins Creation Mechanisms 2 5 9

Sustainable MLOps 1 9

Network programmability 4 5 6

Secure AI/ML-based control for Intent-based Management 2 8

Privacy protection for data analytics 11

Specific Systems

M&O-related functionalities

1	Network Services provisioning
2	Network Services assurance
3	Network Services configuration
4	Resources provisioning
5	Resources assurance
6	Resources configuration
7	Monitoring & analytics
8	Security
9	Design & Development
10	Capabilities Access
11	Privacy

✦ AI/ML-based or related function

ML based configuration recommender for energy savings 3 6

Efficient network and service function allocation 1 2

Multi-domain federated learning 1 3 5 2 4 6

Multi-agent RL for adaptive scaling 2

Explainability for RL-based control 2

Algorithms

Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6

Decentralised orchestration 1 3 5 2 4 6

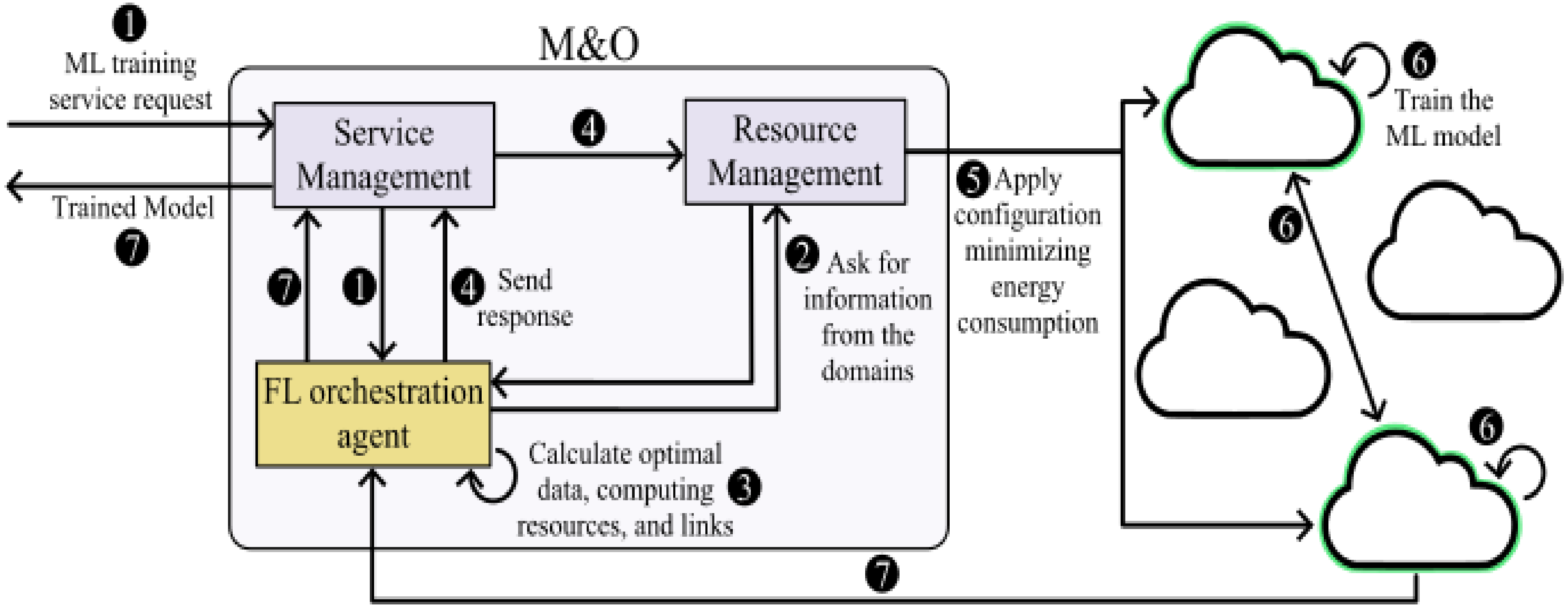
Overall M&O Solutions

Supports the life-cycle management of the AI model training.

Optimises the usage of edge data centre resources to provision an AI model training service (Compute as a Service).

Implementation 5: Resource assignment for federated learning (2/3)

- The implementation showcases how telco operators could leverage connected devices as rich data sources, as well as the deployment of edge compute resources to provision AI model training as a service, in line with the Compute as a Service paradigm.
- As a whole, it demonstrates the optimisation of the compute and the network resources usage via simulations.
- Simulations were carried out on a testbed environment consisting of containerised networked components and Pytorch for training AI Models.
- It also relies on the Monitoring and Telemetry enabler for the life-cycle management of the model training service.

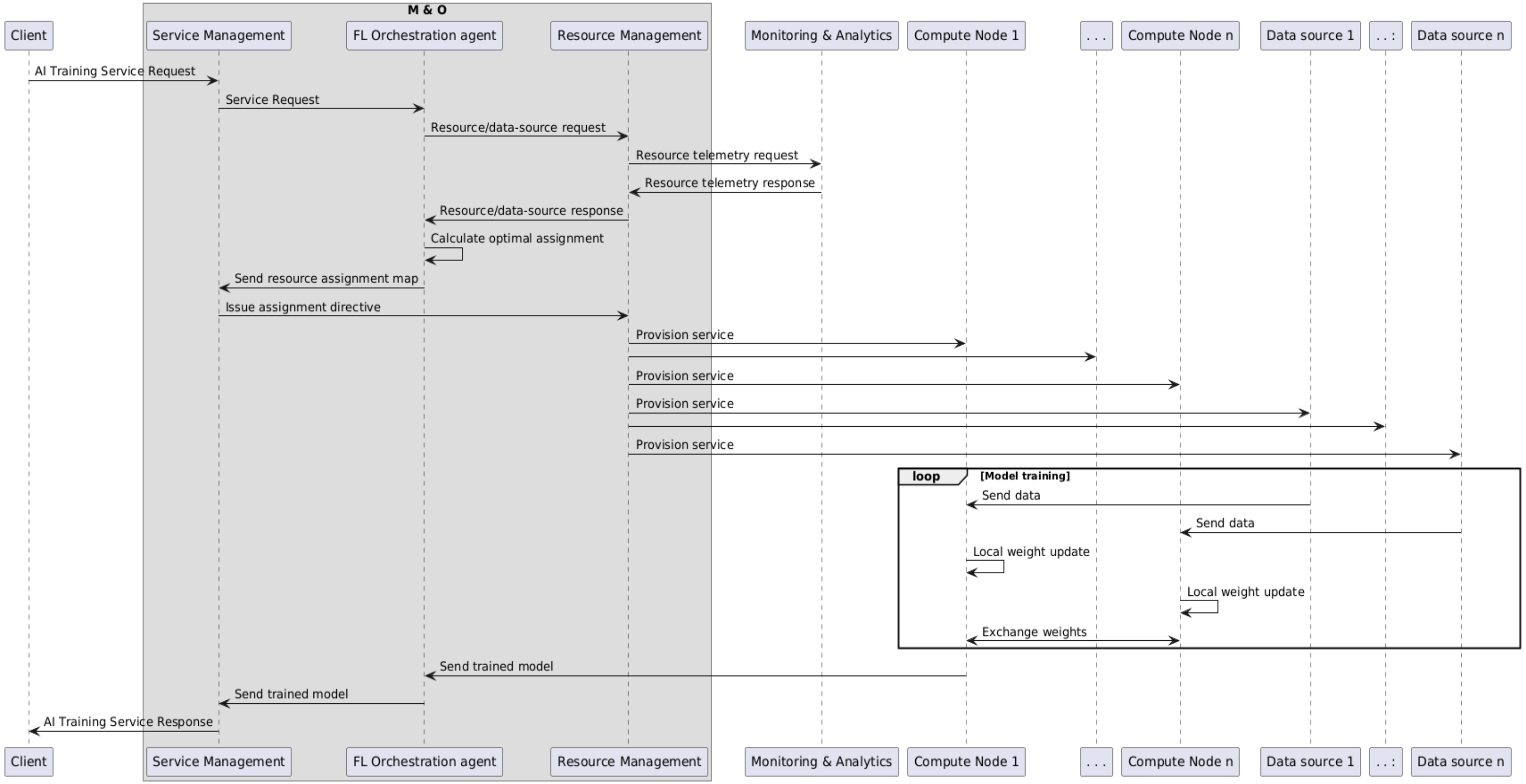


❶ ML training service request
M&O
Service Management
❹
Resource Management
Trained Model ❼
❼
❶
❹ Send response
❷ Ask for information from the domains
❺ Apply configuration minimizing energy consumption
FL orchestration agent
Calculate optimal data, computing resources, and links ❸
❻ Train the ML model
❻
❻
❼

High-level flow diagram of the implementation

Implementation 5: Resource assignment for federated learning (3/3)

Workflow



M & O

Client | Service Management | FL Orchestration agent | Resource Management | Monitoring & Analytics | Compute Node 1 | . . . | Compute Node n | Data source 1 | . . : | Data source n

AI Training Service Request

Service Request

Resource/data-source request

Resource telemetry request

Resource telemetry response

Resource/data-source response

Calculate optimal assignment

Send resource assignment map

Issue assignment directive

Provision service

Provision service

Provision service

Provision service

loop [Model training]

Send data

Send data

Local weight update

Local weight update

Exchange weights

Send trained model

Send trained model

AI Training Service Response

Sequence diagram of interactions among components in this implementation

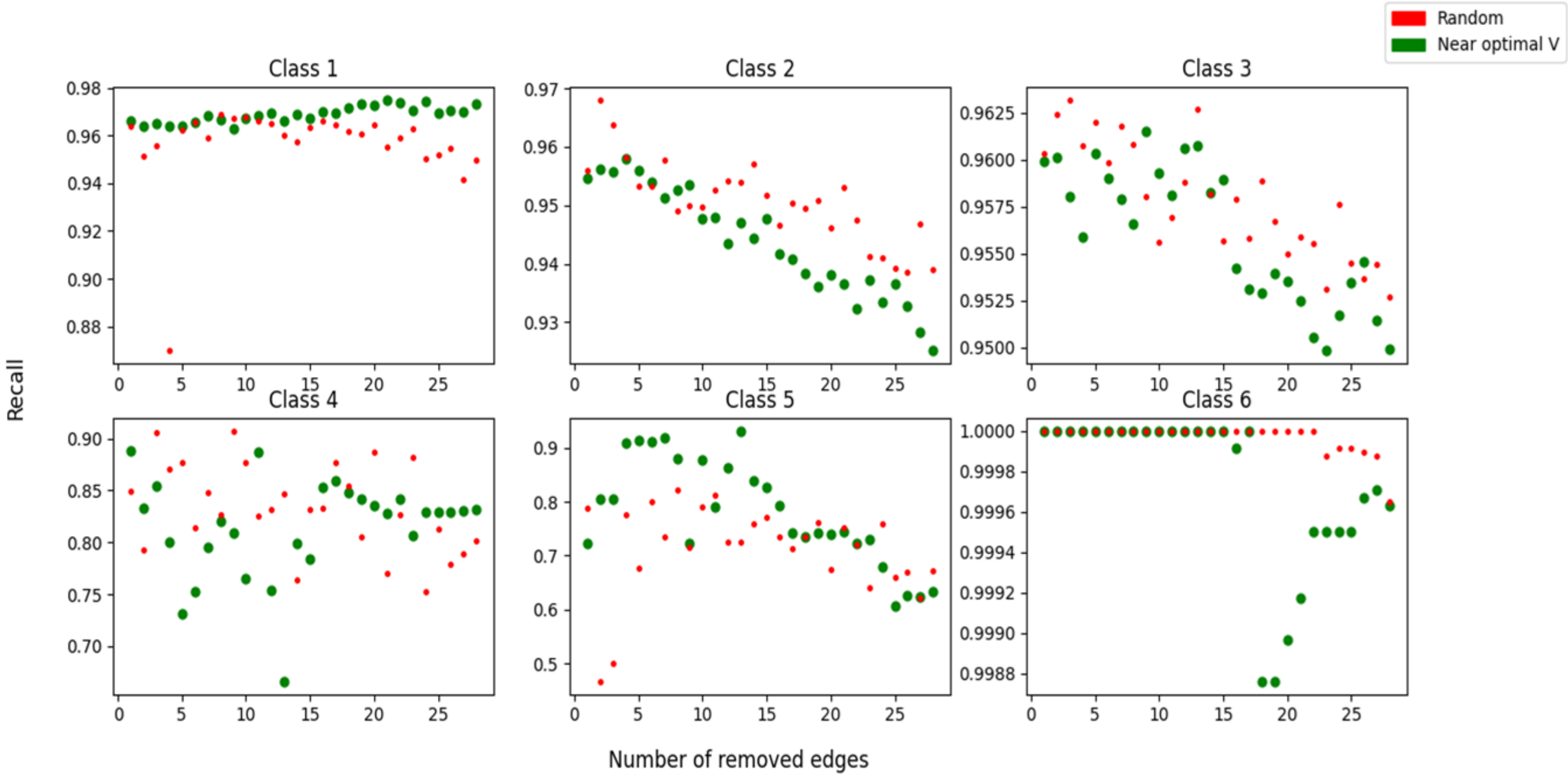
The scenario describes the provisioning process involving a federated learning agent. As it can be appreciated:

- A vertical client (which could be, e.g., a connected industry), requests for the training of a model.
- The federated learning agent algorithm determines which data sources and corresponding compute resources should be used.
- The resource manager provisions the training service, and the service management handles its life-cycle.
- Finally, the trained model is sent to the requesting vertical client.

Implementation 5: Evaluation Results

The plots here represent the recall (true positive) rate of different classification models obtained by training on selected data resources (compute, storage and network). As shown:

- The algorithm optimises the use of data and compute sources in training, thereby improving the energy efficiency KPI.
- Most models are robust to some pruning (removal) of data resources.
- The model leverages the Frobenius norm (a.k.a "V distance") to measure the differences between data sources.
- Removing edges (pruning) that are far from the near optimal V distance shows slightly better and less noisy performance (green dots) in comparison with random removal (red dots).



Recall rate vs pruned data resources

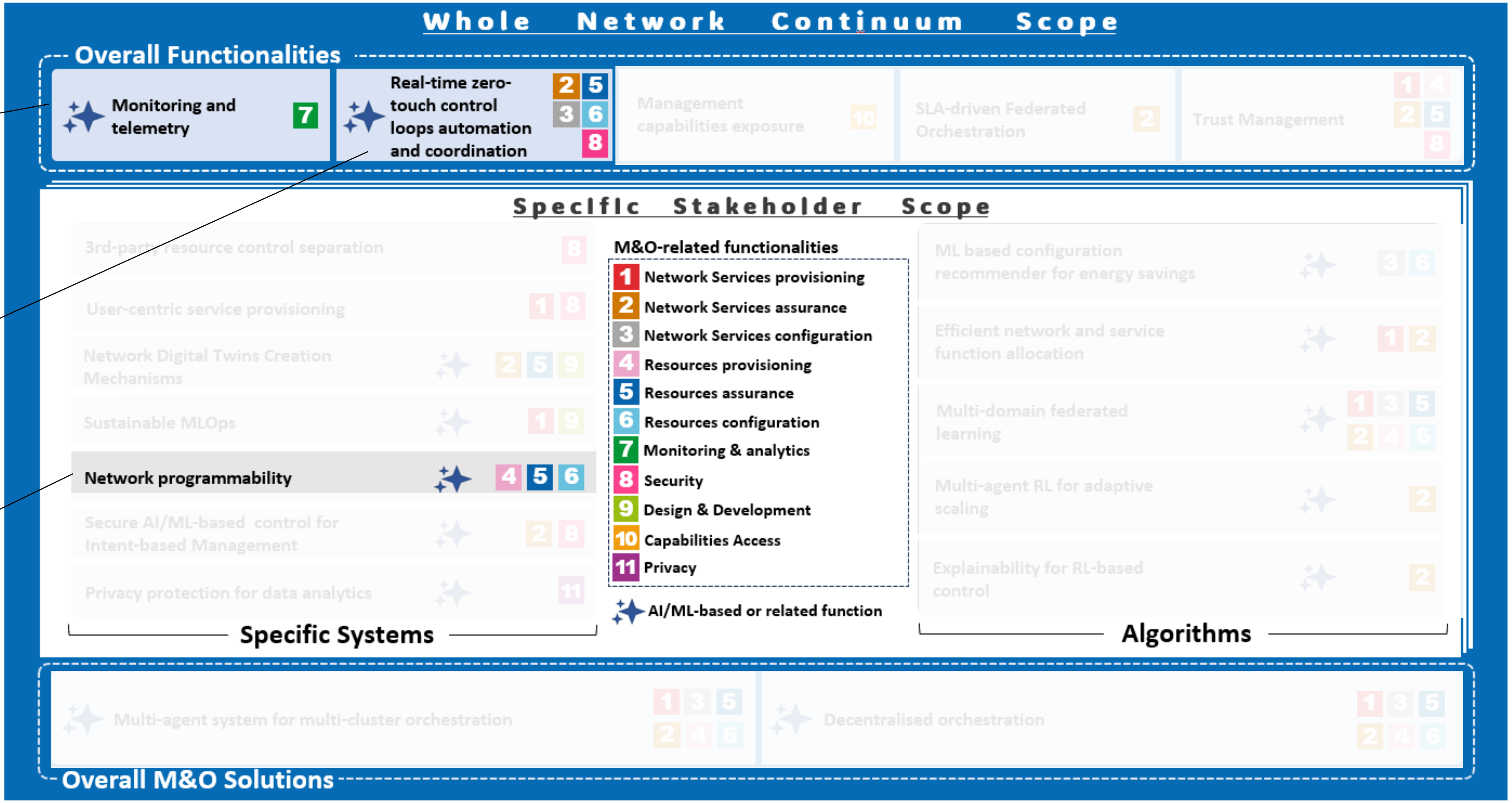
Implementation 6: Flow Reconfiguration based on Dynamic Monitoring and Closed CLs in Deterministic Networks (1/3)

The purpose of this implementation is to showcase mechanisms for the flow reconfiguration based on dynamic monitoring and closed control loops in deterministic networks. The following figure shows the involved enablers in this implementation with a brief explanation of their role:

Provides the CL with real-time streams of (per-flow) monitoring data, including in-band network telemetry.

Real-time analysis and decision-making based on streaming monitoring data (e.g., reroute flows if E2E delay is too high)

Reconfiguration of the flows and/or switches/routers in the deterministic network



Whole Network Continuum Scope

Overall Functionalities

Monitoring and telemetry 7

Real-time zero-touch control loops automation and coordination 2 5 3 6 8

Management capabilities exposure 10

SLA-driven Federated Orchestration 2

Trust Management 1 4 2 5 8

Specific Stakeholder Scope

3rd-party resource control separation 8

User-centric service provisioning 1 8

Network Digital Twins Creation Mechanisms 2 5 9

Sustainable MLOps 1 9

Network programmability 4 5 6

Secure AI/ML-based control for Intent-based Management 2 8

Privacy protection for data analytics 11

Specific Systems

M&O-related functionalities

1 Network Services provisioning
2 Network Services assurance
3 Network Services configuration
4 Resources provisioning
5 Resources assurance
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy

AI/ML-based or related function

ML based configuration recommender for energy savings 3 6

Efficient network and service function allocation 1 2

Multi-domain federated learning 1 3 5 2 4 6

Multi-agent RL for adaptive scaling 2

Explainability for RL-based control 2

Algorithms

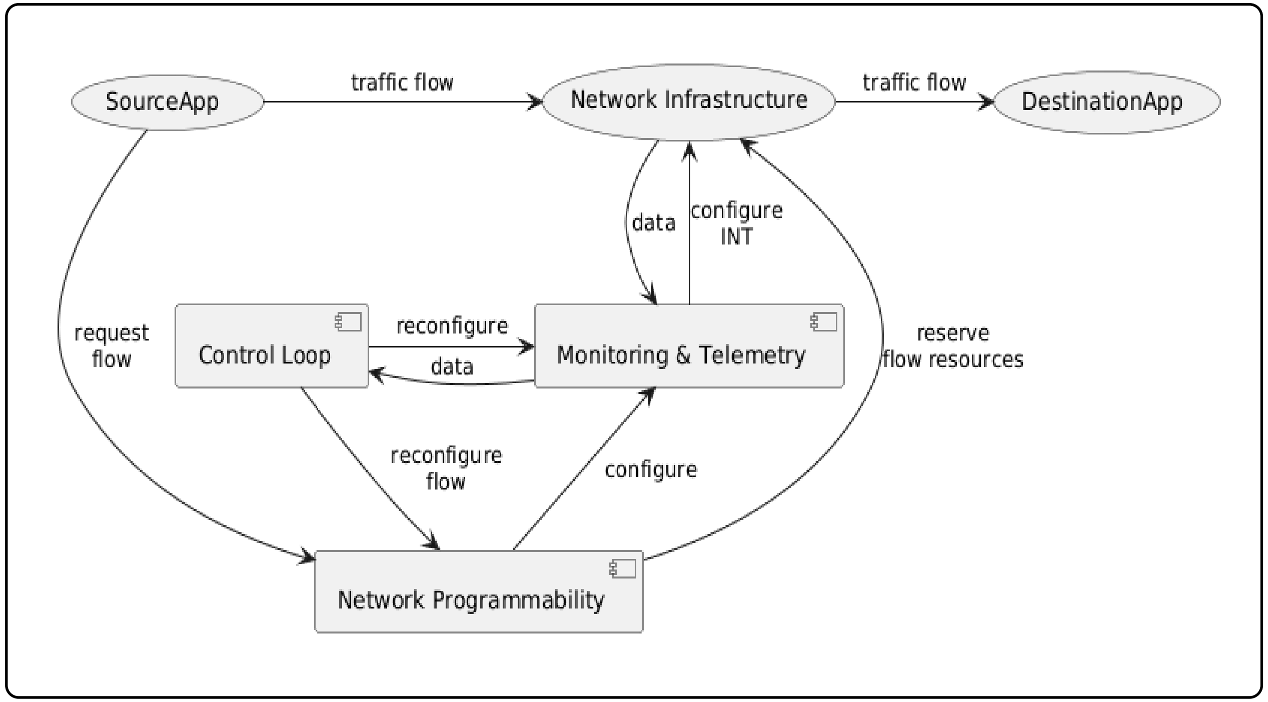
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6

Decentralised orchestration 1 3 5 2 4 6

Overall M&O Solutions

Implementation 6: Flow Reconfiguration based on Dynamic Monitoring and Closed CLs in Deterministic Networks (2/3)

- The implementation provides a fully automated CL for correcting flow misconfigurations in a scalable way in deterministic networks by exploring trade-offs between different forms of telemetry – including in-band telemetry – and applying sketching methods for light-weight analysis and detection.
- Beyond the lab experiments, the implementation has also been included as part of the project PoC#B.1, demonstrating continuous monitoring and reconfiguration of latency-sensitive flows (e.g., among cluster environments).
- Interactions:
  - SourceApp requests a flow through the Network Programmability system, which reserves resources in the network infrastructure and configures the Monitoring and Telemetry functionality. SourceApp then sends traffic to DestinationApp over the network infrastructure.
  - The Monitoring and Telemetry functionality configures the in-band telemetry on the network infrastructure and continuously receives data from the infrastructure.
  - The Monitoring and Telemetry functionality feeds data to the CL, which detects issues and sends reconfiguration commands back.
  - The CL requests flow reconfiguration from the Network Programmability system, which then reserves resources in the network infrastructure as needed for the flow.

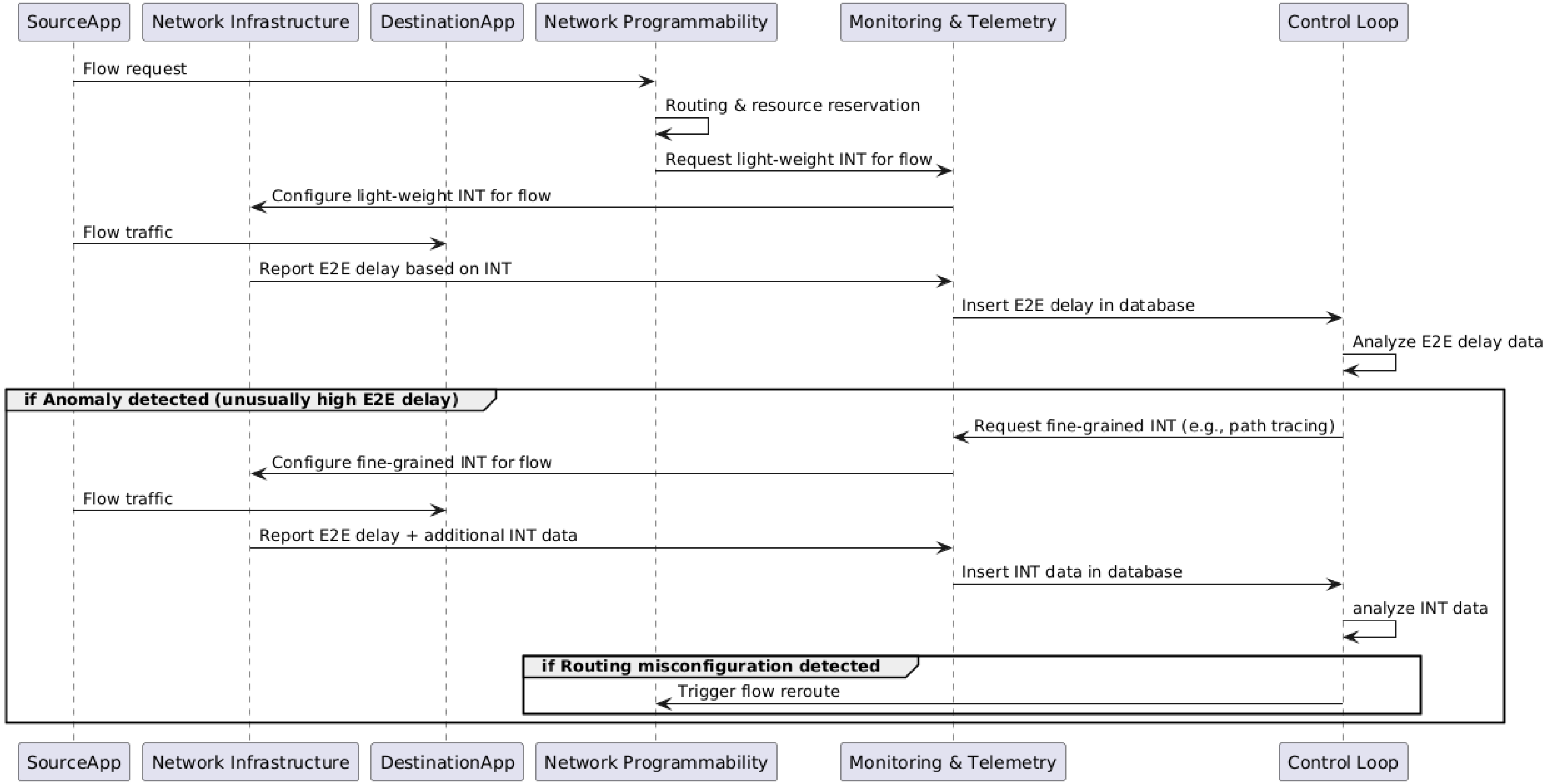


Block diagram showing the interactions between the involved components in this implementation (framework components are shown as rectangular blocks)

Implementation 6: Flow Reconfiguration based on Dynamic Monitoring and Closed CLs in Deterministic Networks (3/3)

This scenario describes the process of adjusting the monitoring and routing configuration by the CL during the lifetime of a critical traffic flow. As it can be seen:

- The source app requests a flow to the destination app.
- Initially, the flow uses light-weight in-band telemetry to measure only E2E delays.
- If an anomaly (high E2E delay) is detected, the CL adjusts the monitoring configuration of the flow to include more telemetry data.
- Then, the CL can analyse the additional INT data, trying to pin-point the issue.
- Finally, a flow reroute can be triggered if a routing misconfiguration is detected to solve the high E2E delays.


SourceApp
Network Infrastructure
DestinationApp
Network Programmability
Monitoring & Telemetry
Control Loop
Flow request
Routing & resource reservation
Request light-weight INT for flow
Configure light-weight INT for flow
Flow traffic
Report E2E delay based on INT
Insert E2E delay in database
Analyze E2E delay data
if Anomaly detected (unusually high E2E delay)
Request fine-grained INT (e.g., path tracing)
Configure fine-grained INT for flow
Flow traffic
Report E2E delay + additional INT data
Insert INT data in database
analyze INT data
if Routing misconfiguration detected
Trigger flow reroute
SourceApp
Network Infrastructure
DestinationApp
Network Programmability
Monitoring & Telemetry
Control Loop

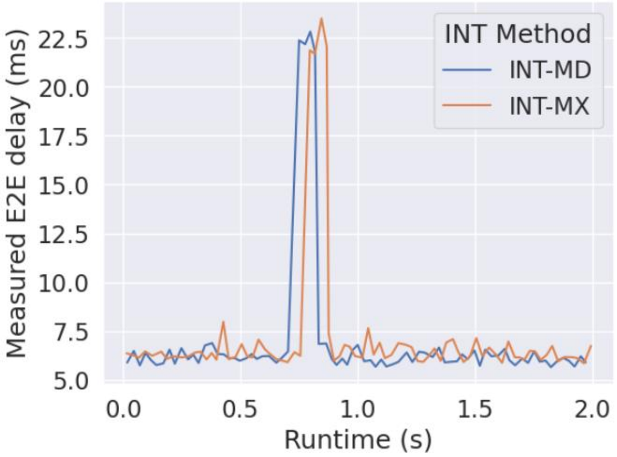
Sequence diagram of interactions between the involved components in this implementation in the scenario of an anomaly within the network

Implementation 6: Evaluation results

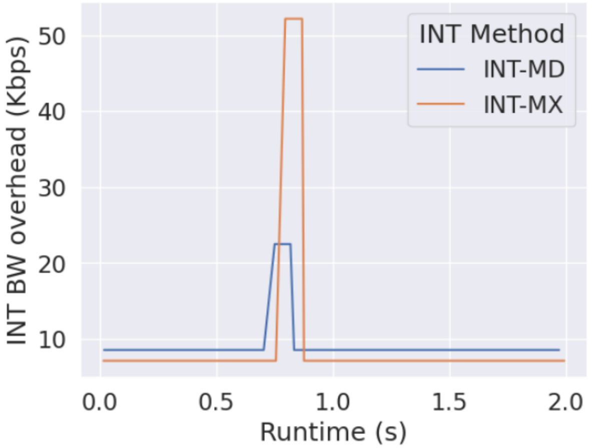
The experiment was run twice using in-band telemetry MD and MX modes, with the anomaly introduced at 0.75s.

Impacted KPIs:

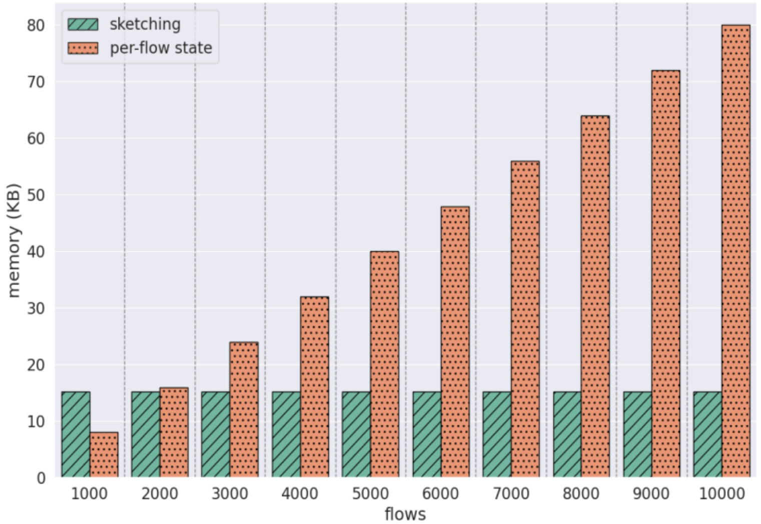
- **Availability**: Mitigates unacceptable E2E delays, ensuring application QoS with <2% of packets impacted in experiments.
- **Automation**: Fully autonomous Closed Loop (CL) with communication overhead capped at milliseconds for fine-grained telemetry collection and SDN analysis.
- **Operational Efficiency (OPEX)**: Reduces monitoring and analysis resource requirements; no human intervention needed.
- **Latency**: High E2E delays detected and corrected at runtime, reducing flow latency to acceptable levels.
- **Reliability**: Recovery time of ~120 ms (~4 packets) after detecting high E2E delay.
- **Scalability**: Dynamic monitoring keeps INT bandwidth overhead under 10 kbit/s on average. CL uses constant O(1) memory, avoiding linear scaling of per-flow state.



1. CL detects and corrects high E2E delays during flow runtime as seen by the measured E2E delays



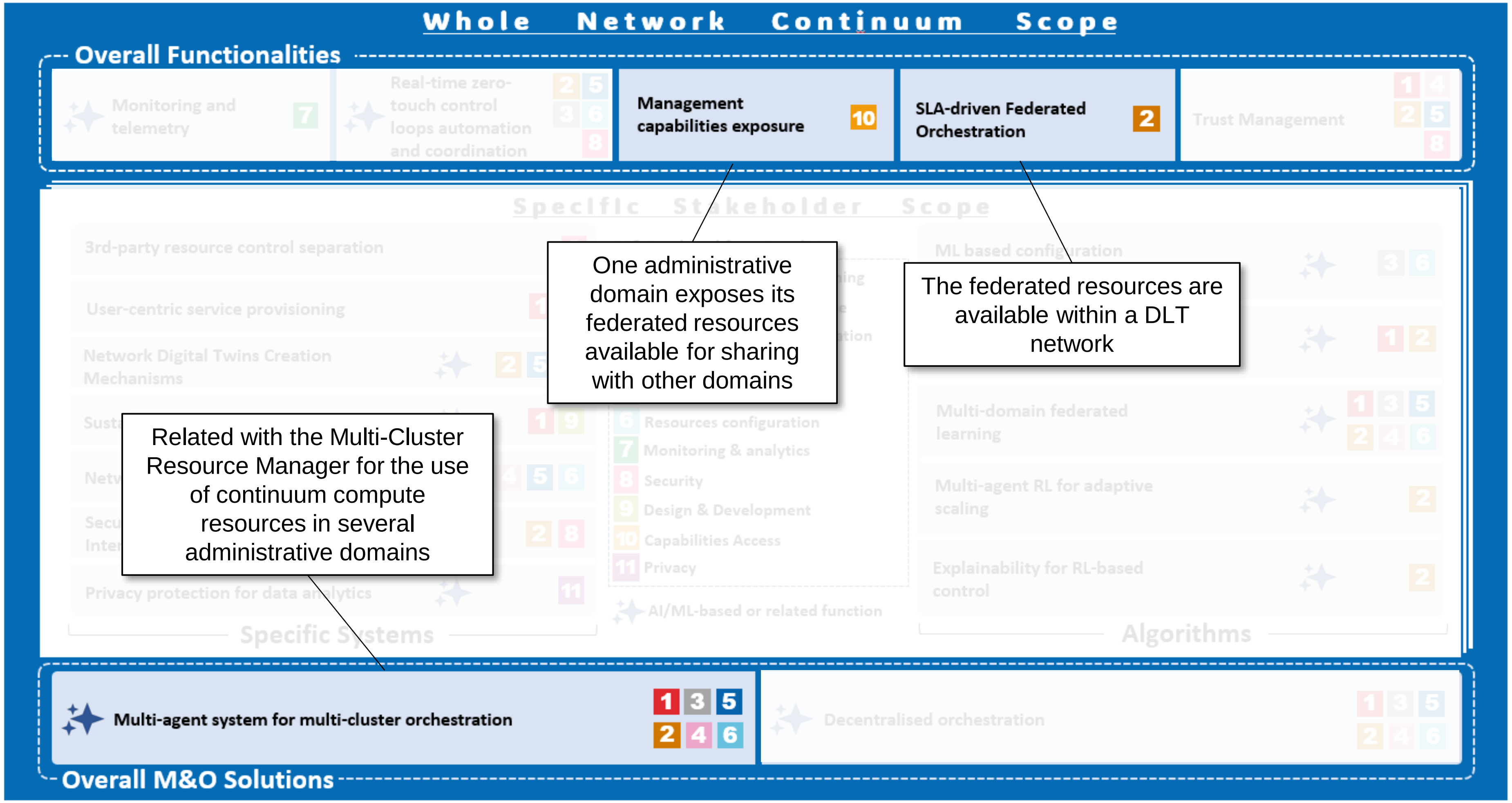
2. INT bandwidth overhead during flow runtime (increase during high E2E delays)



3. Memory analysis of sketching-based methods for detecting high E2E delays vs per-flow state

Implementation 7: Edge convergence over federated resources for the computing continuum (1/3)

This implementation targets to explore the capabilities of the CAMARA EdgeCloud APIs [CEC24] for the management of the compute resources in the network continuum, as well as the possibility to extend them to be used with federated resources of external administrative domains. The figure shows the different enablers involved in this implementation.



Whole Network Continuum Scope
Overall Functionalities
Monitoring and telemetry 7
Real-time zero-touch control loops automation and coordination 2 5 3 6 8
Management capabilities exposure 10
SLA-driven Federated Orchestration 2
Trust Management 1 4 2 5 8
Specific Stakeholder Scope
3rd-party resource control separation
User-centric service provisioning 1
Network Digital Twins Creation Mechanisms 2 5
Susta… 1 9
Netw… 4 5 6
Secu… Inter… 2 8
Privacy protection for data analytics 11
Specific Systems
6 Resources configuration
7 Monitoring & analytics
8 Security
9 Design & Development
10 Capabilities Access
11 Privacy
AI/ML-based or related function
ML based configuration 3 6
1 2
Multi-domain federated learning 1 3 5 2 4 6
Multi-agent RL for adaptive scaling 2
Explainability for RL-based control 2
Algorithms
Multi-agent system for multi-cluster orchestration 1 3 5 2 4 6
Decentralised orchestration 1 3 5 2 4 6
Overall M&O Solutions

One administrative domain exposes its federated resources available for sharing with other domains

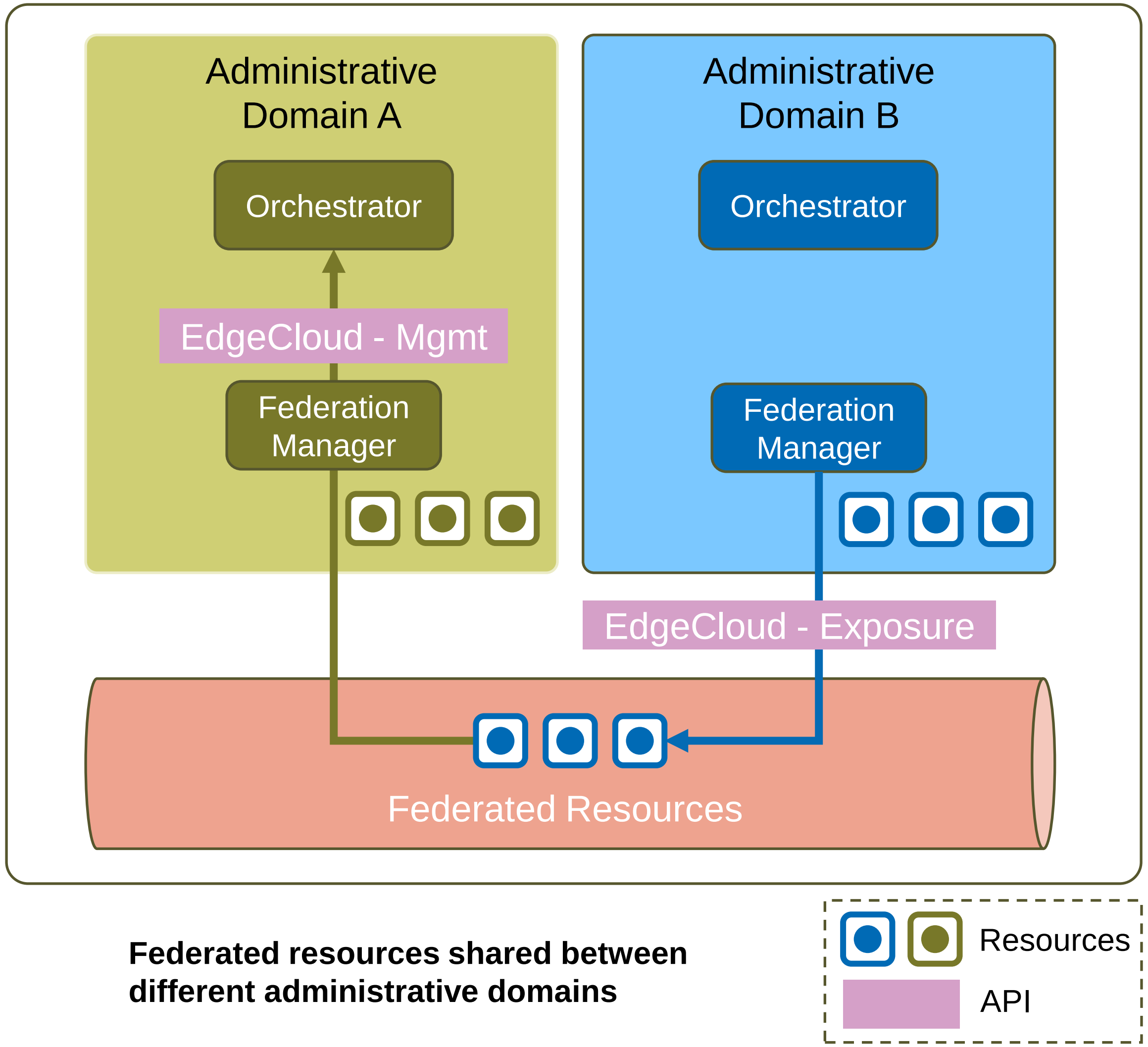
The federated resources are available within a DLT network

Related with the Multi-Cluster Resource Manager for the use of continuum compute resources in several administrative domains

Implementation 7: Edge convergence over federated resources for the computing continuum (2/3)

- The federated resources are available within a DLT network. Based on this, the implementation extends the EdgeCloud APIs [CEC24] for accessing federated resources in a seamless way with minimal changes in the APIs.

- This implementation has been deployed on a testbed with two administrative domains, and using certain smart contracts associated to an agreement stablished among parties after a defined negotiation, in the following way:
  - One administrative domain exposes its federated resources available for sharing with other domains over the DLT network.
  - The other administrative domain needs resources to fulfil a service. These resources could be their own resources or external ones, i.e., applying the network continuum by including the exposed resources in the DLT by the other administrative domain. If the external resources were required, the service provisioning area would request for an agreement (needed to be created if it doesn’t exist).
  - Using the updated EdgeCloud APIs, the continuum among partners is extended, increasing the resources capability in an agile way.



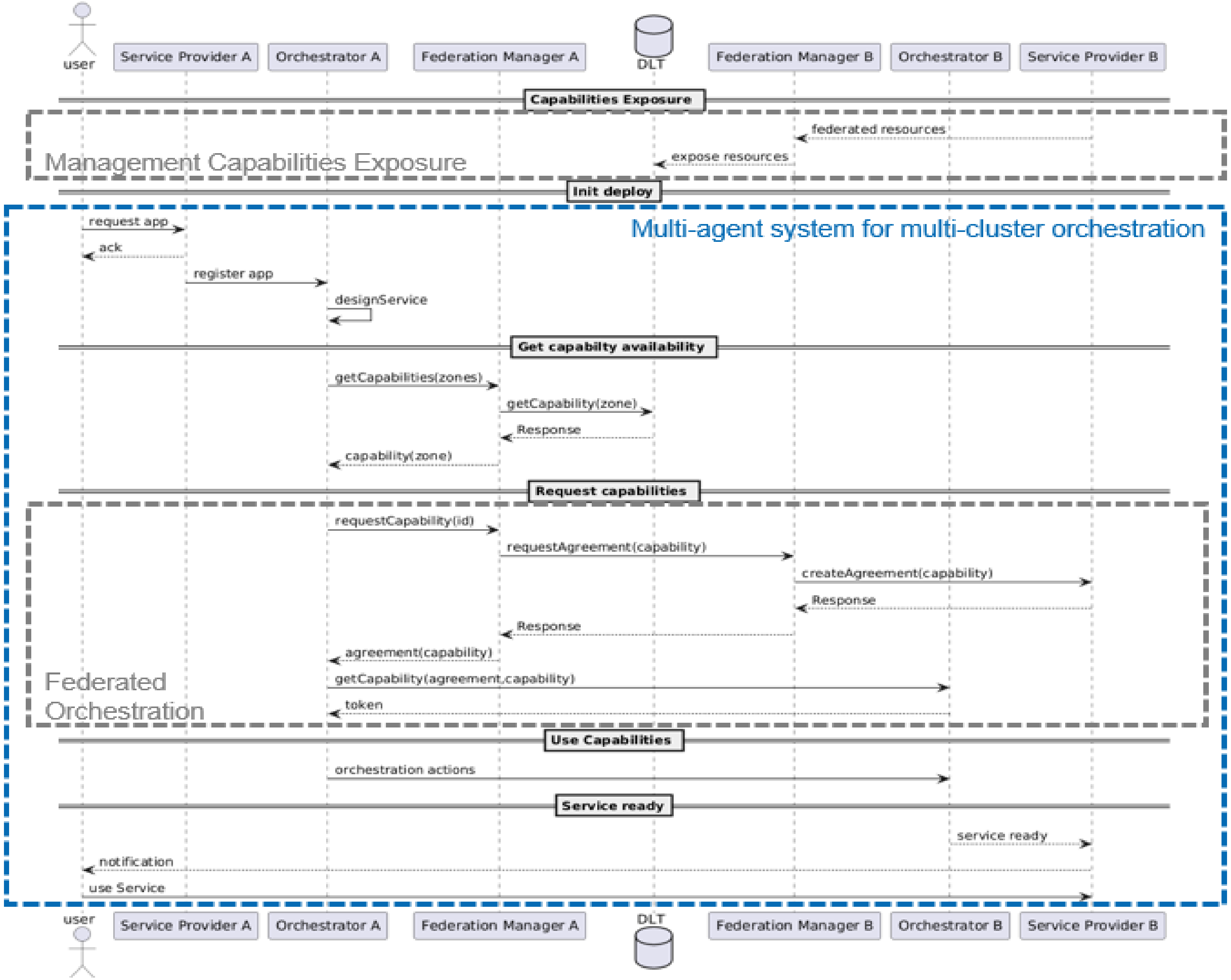
Federated resources shared between different administrative domains

Implementation 7: Edge convergence over federated resources for the computing continuum (3/3)

Functional sequence diagram

The scenario describes a complex multi-domain environment where two providers (A and B) and their respective orchestrators interact to deliver services:

- Service provider B exposes its available resources to the “federation” through the Management Capabilities Exposure enabler.
- The multi-agent system for multi-cluster orchestration enabler is responsible for the service orchestration. To do that, the service provider A requests a service instance to the orchestrator on its domain.
- Orchestrator A needs resources from third parties to perform the design of the service instantiation, so:
  - It gets the available resources from the federated ones for a specific zone.
  - Using the Federated Orchestration enabler it orchestrates the service (agreement negotiation, token request, etc.).
- Finally, the service is ready to be used from the domain of the A provider.



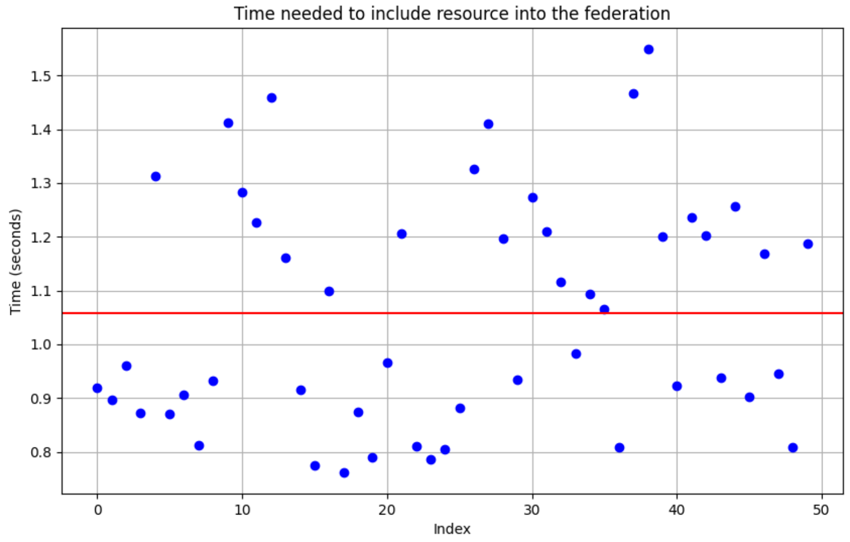
Implementation 7: Evaluation results (1/2)

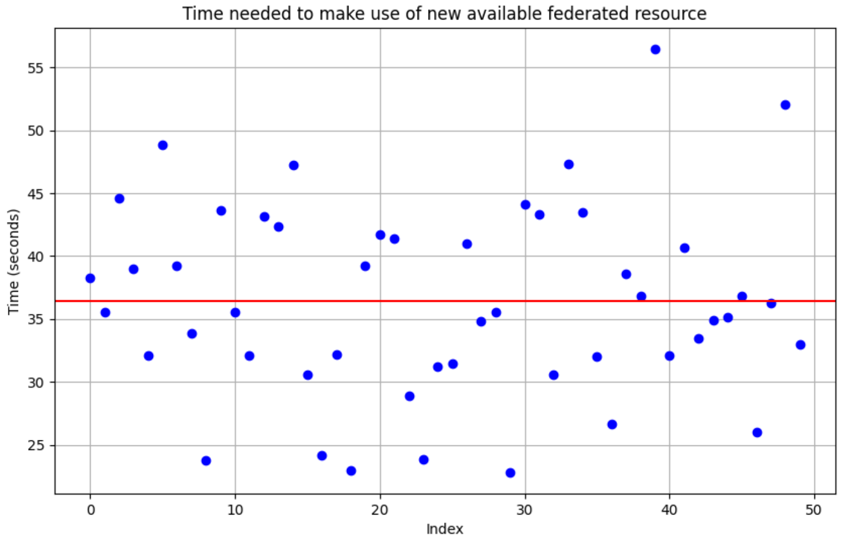
- The CAMARA API EdgeCloud definitions were validated as effective for federated exposure and management of the compute capabilities. These definitions successfully considered service availability zones, ensuring harmonisation of capabilities across different environments.
- Smart contract agreements were effectively integrated using TMForum Agreement definitions within orchestrator actions. This integration, managed by business or E2E layers, employed DLT and the CAMARA Identity and Consent Management API.
- Modifications were identified as necessary in the CAMARA APIs to support the discovery, management, and deployment of services in a federated environments. These modifications will be described in the upcoming deliverable D6.5
- A common DLT technology was proposed to enable secure and seamless integration across multiple stakeholders. This would simplify operations in complex environments and promote easier usage across the ecosystem.
- Simplifyed operations in multi-cluster orchestration scenarios. This was achieved through coordinated and synergetic actions, ensuring smoother operation across multiple stakeholders and clusters.
- The testbed evaluated the impact of utilising external resources, particularly focusing on the time required to link these resources and the associated entities. It also considered the necessary information for this process.
- The results from this implementation will be shared with the CAMARA project for further evaluation, offering valuable insights for future improvements.

Implementation 7 – Evaluation Results (2/2)

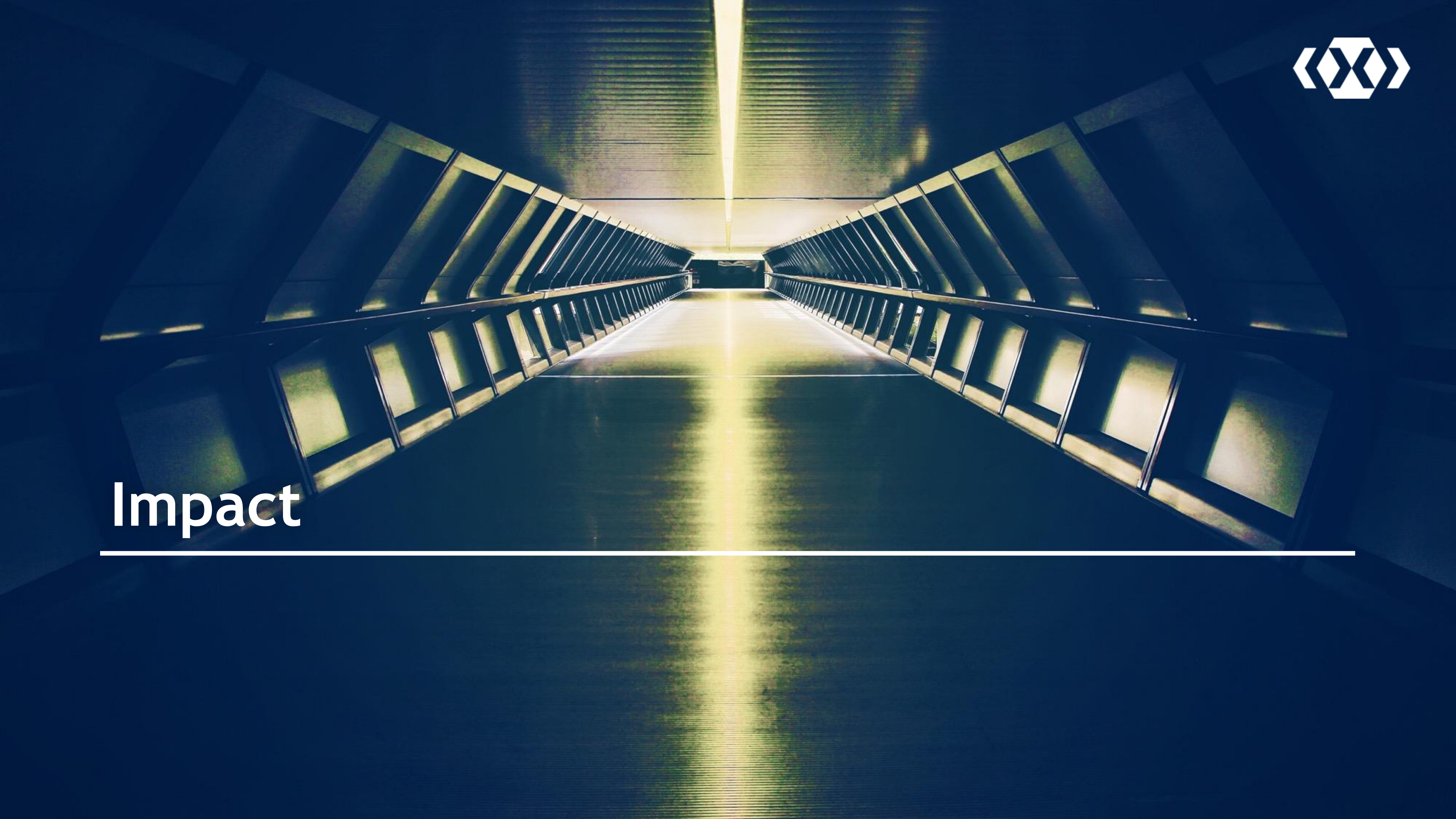
Regarding the KPIs address, the following can be highlighted:

- Scalability: regarding the usage of resources from a larger catalogue (combination of own resources and federated ones), using service components from several providers and locations.
- Latency: with the use of resources with better capabilities and closer to the user. The use of federated resources in the nearby areas (device to access network) resulted in 12ms (measurements from in-house testbed), compared to the 40ms measured for the core domain resources, and the 60ms for hyperscalers.
- Reliability: easing service migration between domains, facilitating new paths when the defined service is not working.
- Availability: being able to create services with better features to comply with the personalised SLAs of client services. Also, the time between adding a federated resource in the network and the time for connecting it to the federated resource management was less than 1 minute.
- Automation: with the use of external capabilities in an agnostic way (a service provider could use external resources in a similar way as their own resources).ed, the following can be highlighted:


Time needed to include resource into the federation


Time needed to make use of new available federated resource

Impact



Publications (1/2)

Below the WP6 publications during the reporting period:

Title	Type of Publication	Status
Applying Digital Twins to Optical Networks with Cloud-native SDN Controllers	Article in Journal	Published
IntentLLM: An AI Chatbot to Create and Explain Slice Intents in TeraFlowSDN	Publication in conference proceeding/workshop	Published
Utilizing Causal Learning for Cognitive Management of 6G Networks	Publication in conference proceeding/workshop	Published
Exploiting Queue Information for Scalable Delay-Constrained Routing in Deterministic Networks	Article in Journal	Published
An East-Westbound Control Architecture for Multi-Segment Deterministic Networking	Publication in conference proceeding/workshop	Published
Secure AI/ML-based control in Intent-based Management System	Publication in conference proceeding/workshop	Published
Enabling Traffic Forecasting with Cloud-native SDN Controller in Transport Networks	Article in Journal	Published
Network Resource Allocation for Gaming Using MEC API and TeraFlowSDN	Publication in conference proceeding/workshop	Published

Publications (2/2)

Below the WP6 publications during the reporting period:

Title	Type of Publication	Status
A Cloud-Native Approach for Orchestrating 6G-Enabled Services at the Computing Continuum	Publication in conference proceeding/workshop	Published
Towards an AI/ML-driven SMO Framework in O-RAN: Scenarios, Solutions, and Challenges	Publication in conference proceeding/workshop	Published
Attention to Virtualisation: Making Network Digital Twins aware of Network Slicing	Article in Journal	Submitted
Trust-based Intent Management for 6G: A Level of Trust Assessment Function	Article in Journal	Submitted
AI-Driven Orchestration of 6G-Enabled Services Across the Computing Continuum	Publication in conference proceeding/workshop	Submitted
Transport DataPlane-in-a-box: Using the TeraFlowSDN Controller to Manage Packet-Optical Transport Networks	Publication in conference proceeding/workshop	Published
Providing Anomalous Behaviour Profiling by extending SmartNIC Transceiver support in Packet-Optical Networks	Publication in conference proceeding/workshop	Published

Open-Source Software

Below the Open-Source components released or updated during the reporting period:

Title	SW Origin	Licence	Available in	PoC / Demo	Related Enabler / Use Case
ETSI TeraFlowSDN	Existing SW - Enhanced in project	Apache 2.0	TeraFlowSDN Website	PoC B.1	Network Programmability
Digital Ledger Technology (DLT) for service federation	New SW - Created in project	Apache 2.0	GitLab	PoC B.1	Federated orchestration system
Infrastructure Layer Emulator (ILE)	New SW - Created in project	Apache 2.0	GitHub	PoC B.1	Decentralised Orchestration System
HX MLOps	New SW - Created in project	Apache 2.0	GitHub	Testbed	Sustainable MLOps
Routing protocols for deterministic networks	Existing SW - Enhanced in project	Apache 2.0	GitHub	Testbed	Network Programmability
6G latency sensitive service for Smart Manufacturing	New SW - Created in project	Apache 2.0	GitLab	PoC B.1	Overall M&O System enablers.

Open APIs

Below the Open APIs released or updated during the reporting period:

API	SW Origin	Licence	Available in	Related Component / UseCase	PoC / Demo
DLT Service Federation Open API	New - Created in project	CC Attr. 4.0 International	Zenodo	Federated Orchestration	PoC#B.1
Sustainable MLOps models sharing API	New - Created in project	Apache 2.0	Zenodo	Sustainable MLOps	Testbed
TMF640 Service Activation Management API	Existing - Enhanced in project	Apache 2.0	GitHub Zenodo	Network programmability	Testbed
TMF644 Resource Function Activation Management API	Existing - Enhanced in project	Apache 2.0	GitHub Zenodo	Network programmability	Testbed
Conflict Detection for CLs Automation and Management API	New - Created in project	CC Attr. 4.0 International	Zenodo	Real-time zero-touch control loops automation and coordination functionality	Testbed
MEC Bandwidth Management (MEC 014)	Existing	BSD-3-Clause	Github	Network programmability	Testbed
CAMARA QoD	Existing	Apache 2.0	Github	Network programmability	Testbed
Monitoring jobs configuration	Existing – Enhanced in project	Apache 2.0	Zenodo	Monitoring and Telemetry	PoC#B.1
Query historical monitoring data	Existing	MIT	InfluxDB site	Monitoring and Telemetry	PoC#B.1
Multi-cluster extreme-edge resource orchestration API	New - Created in project	Apache 2.0	Zenodo	Multi-agent system for multi-cluster orchestration	PoC#B.1
Closed Loop Governance - Catalogue API	New - Created in project	Apache 2.0	Zenodo	Real-time zero-touch control loops automation and coordination functionality	PoC#B.1
Closed Loop Governance – Lifecycle Management API	New - Created in project	Apache 2.0	Zenodo	Real-time zero-touch control loops automation and coordination functionality	PoC#B.1
MEC exposure and experience management API	Existing-Modified in project	CC Attr. 4.0 International	Zenodo	Edge convergence over federated resources for the computing continuum	Testbed
Sustainable MLOps Workflow Info Collector API	New – Created in project	Apache 2.0	Zenodo	Sustainable MLOps	Testbed
Security API	New - Created in project	CC Attr. 4.0 International	Zenodo	Management Capabilities Exposure	Testbed
Subscription API	New - Created in project	CC Attr. 4.0 International	Zenodo	Management Capabilities Exposure	Testbed
Listing API	New - Created in project	CC Attr. 4.0 International	Zenodo	Management Capabilities Exposure	Testbed
Trust Management API	New - Created in project	GPL-3.0	Zenodo	Trust Management System	Testbed
Trust Evaluation Function API	New - Created in project	GPL-3.0	Zenodo	Trust Management System	PoC#A.1
LoTAF API	New - Created in project	GPL-3.0	GitHub Zenodo	Trust Management System	Testbed
DLT Service Federation Open API	New - Created in project	CC Attr. 4.0 International	Zenodo	Federated Orchestration	Testbed

Dissemination activities

WP6 participated in the following dissemination activities during the reporting period:

Title:	**Participation in the working group “Trustworthiness of the 6G-Platform Program”.**
Type of activity:	Meeting.
Target audience reached:	Research community, Industry, EU institutions, Business partners
Objective:	Common understanding and comparison of Trustworthiness for 6G with other projects around the world like the 6G Platform Program Germany.
Date:	20/09/2024.
Partners involved:	TID and UMU.

Title:	**Presentation at EuCNC Special Session: Jazz Networks: A proposal for deploying network services in the 6G cloud continuum.**
Type of activity:	Conference.
Target audience reached:	Research community, Industry, business partners, Innovators, EU Institutions
Objective:	Sharing ideas and proposal regarding the continuum orchestration concept towards 6G.
Date:	04/06/2024.
Partners involved:	ASA

Communication activities

Below the communication activities related to WP6 during the reporting period:

Description	Audience	Channel	Date	Partner(s)
News item on the Hexa-X-II website: Hexa-X-II Deliverable D6.3 focuses on the initial design of 6G smart network management framework	Research community	Website	01/07/2024	OUL, ASA
D6.3 – Initial Design of 6G Smart Network Management Framework added to the Hexa-X-II website	Research community	Website	01/07/2024	OUL, ASA
LInkedIn post on D6.3	Public	Social media	01/07/2024	NFI
FNS & Hexa-X-II joint workshop. Presentation on the Hexa-X-II Architecture Aspects: Foundations of 6G smart management and orchestration design.	Research community	Event	23/09/2024	ASA

Media

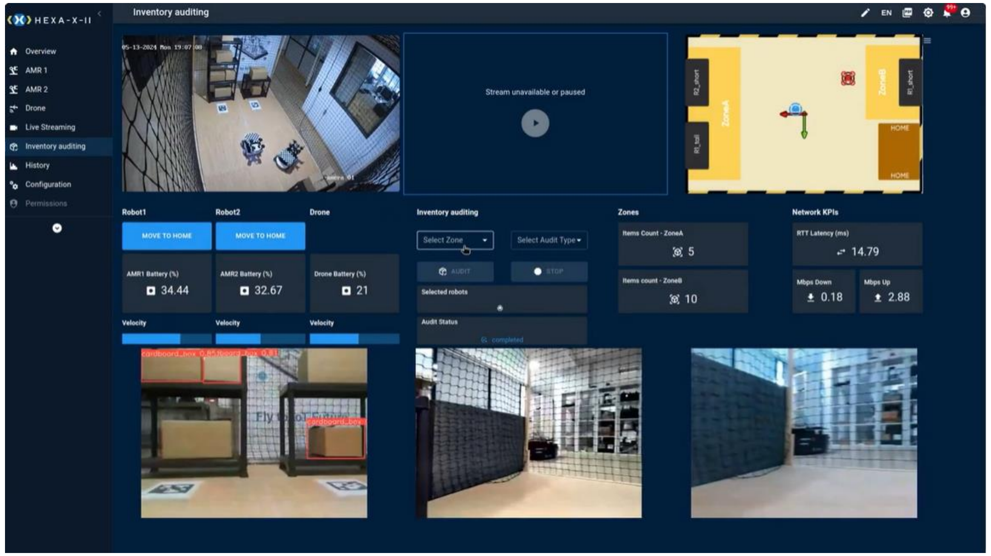
Demo: Advanced M&O, Flexible topologies and Network beyond communications enablers in a Cobots-powered warehouse.

Click to play video

HEXA-X-II

System-PoC A/B & Component-PoC #B.3

Demo: Advanced M&O, Flexible topologies and Network beyond communications enablers in Cobot-powered Warehouse Inventory Management


Inventory auditing

Robot1 — MOVE TO HOME — AMR1 Battery (%) 34.44 — Velocity

Robot2 — MOVE TO HOME — AMR2 Battery (%) 32.67 — Velocity

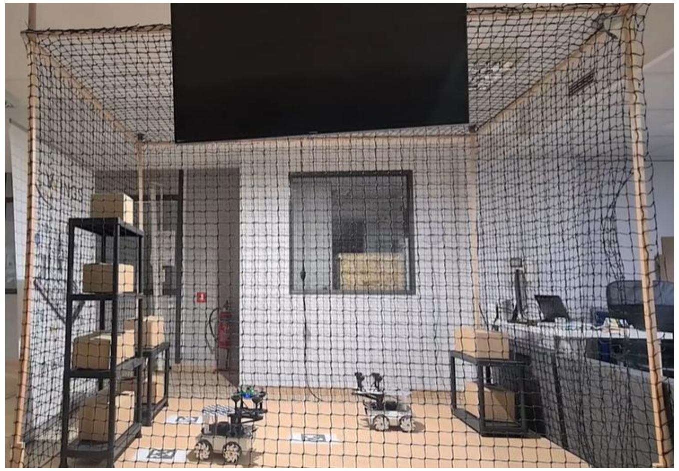
Drone — Drone Battery (%) 21 — Velocity

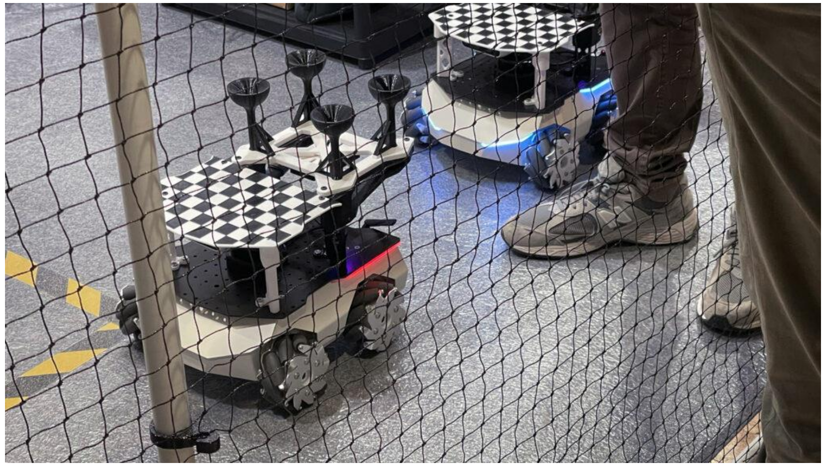
Inventory auditing — Select Zone — Select Audit Type — AUDIT — STOP — Selected robots — Audit Status: completed

Zones — Items Count - ZoneA: 5 — Items count - ZoneB: 10

Network KPIs — RTT Latency (ms): 14.79 — Mbps Down: 0.18 — Mbps Up: 2.88

Stream unavailable or paused





This demo showcases resilient and trustworthy operation scenarios in warehouse and manufacturing environments, leveraging beyond-5G / 6G, AI/ML-, trust- and energy-driven optimisation enablers, flexible topologies, network reliability, energy availability and compute continuum nodes performance. An early version of this demo was presented at EuCNC & 6G Summit 2024.

The final version of this demo is planned to be submitted for demonstration at EuCNC & 6G Summit 2025.

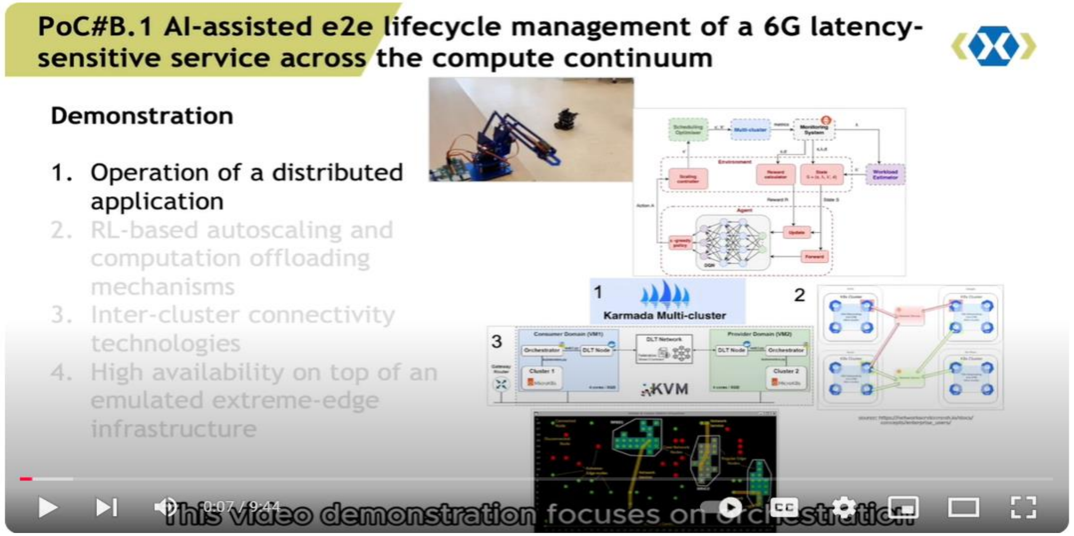
Media

Other videos on YouTube:

INTENT-BASED SERVICE PROVISIONING WITH INTEGRATED CLOSED-LOOP FOR COBOT SERVICE MIGRATION BASED ON BATTERY THRESHOLD

HEXA-X-II EUCNC - ANTWERP -2024

[Click to play video]



PoC#B.1 AI-assisted e2e lifecycle management of a 6G latency-sensitive service across the compute continuum

Demonstration

1. Operation of a distributed application
2. RL-based autoscaling and computation offloading mechanisms
3. Inter-cluster connectivity technologies
4. High availability on top of an emulated extreme-edge infrastructure

This video demonstration focuses on orchestration

[Click to play video]

This video presents a demo intent-based service provisioning with an integrated closed-loop for service components migration in cobots triggered by their battery level measurements. This demo was presented at the EuCNC & 6G Summit 2024.

This video presents another demo showcasing the E2E and AI-assisted orchestration process of a latency sensitive service across the compute continuum. The demo was also presented at the EuCNC & 6G Summit 2024.

Contribution to the Key Exploitable Results (1/2)

The work performed during the reporting period in this WP6 is considered to impact on the following Key Exploitable Results defined in WP7:

KER	Rationale
KER 2.1 – E2E system blueprint of the sustainable, inclusive, and trustworthy 6G platform.	The work performed in this WP6 is actually intended to support the E2E system blueprint design being addressed in WP2. Besides, the topics of sustainability and trustworthiness are specifically addressed by different enables in the Smart Management Framework.
KER 2.3 – Intent-based digital service management.	Though Intent Based Management (IbM) is a work topic specifically assigned to WP2 according to the Hexa-X-II work programme, the WP6 Smart Management Framework also considers this topic in certain technical enablers, specifically the following: - In the “Secure AI/ML-based control algorithms for Intent-based Management” enabler. - In the “Trust Management System”, which considers an E2E intent-based trust management solution to assess and ensure the trustworthiness of network services or resource provisioning. - In the “Decentralised Orchestration System”, which also considers IbM as a way to define the deployment descriptors for the service components to be deployed.
KER 2.4 - Integrated Network Digital Twins for E2E security, privacy, and resilience assessment.	The Network Digital Twins technology is specifically addressed from the “Network Digital Twins Creation Mechanisms” enabler in the M&O framework.
KER 3.1 – Integration of AI in the 6G data-driven architecture.	AI/ML techniques are extensively applied in the Smart Management Framework. All those enabler in the framework highlighted with the sparkle icon (✨), which are most of them, are AI/ML based or related with AI/ML functions.
KER 3.2 – Easily deployable modular and scalable architecture	The Overall M&O Solutions in the management framework are considered well aligned with this KER.
KER 3.3 – Network of networks	All the enablers targeting the whole network continuum scope in the management framework are specifically aligned with this topic.
KER 3.5 – Cloud transformation	The M&O technical enablers in the management framework are intended to adapt the cloud for the 6G requirements, as requested for this KER.

Contribution to the Key Exploitable Results (2/2)

KER	Rationale
KER 6.1 – Programmable flexible network configuration of transport networks	All these KERs have associated the following specific enablers in the Smart Management Framework: • The Network Programmability system, targeting KER 6.1. • The Monitoring and Telemetry functionality, targeting KER 6.2. • The Management Capabilities Exposure system, targeting KER 6.3. • The Energy Efficient Service Function Allocation algorithms, targeting KER 6.7. • The Trust Management functionality, the User-centric Service Provisioning system, and the 3rd Party Resource Control Separation system, all of them contributing to KER 6.9. • The Network Digital Twins Creation Mechanisms system, targeting KER 6.10. • The Rear-time zero-touch control loops automation & coordination functionality, targeting both KER 6.11 and KER 6.12.
KER 6.2 – Programmable network monitoring and telemetry.	
KER 6.3 – Integration fabric.	
KER 6.7 – Sustainable resource allocation for network management	
KER 6.9 – Methods for trustworthiness of AI/ML in network management	
KER 6.10 – Network Digital Twins for autonomous network management	
KER 6.11 – Closed loop governance	
KER 6.12 – Multiple Closed loops coordination	
KER 6.4 – Resource controllability separation for multi-tenancy support	The 3rd Party Resource Control Separation system in the framework specifically targets this KER.
KER 6.5 – Multi-cluster resource management mechanisms	The Overall M&O Solutions in the management framework contribute to these two KERs.
KER 6.6 – Intelligent orchestration mechanisms for the computing continuum	
KER 6.8 – Methods for network management ML pipeline sustainability	The Sustainable MLOps system from the management framework contributes to this KER.

Impacted KVIs

- **Key Value Indicators (KVIs)** are an important concept in the Hexa-X-II project. It aims to address **societal, environmental,** and **economic challenges,** while ensuring technological innovation.
- KVIs are based on the broader **"Key Value" (KV) concept**, which assesses the impact of use cases and technology on high-level **human and planetary goals,** aligned with the **United Nations SDGs**: https://sdgs.un.org/goals.
- A KVI represents a **high-level qualitative or quantitative metric** designed to measure the performance and alignment of technology development with broader societal values and project goals. Unlike well-known KPIs (focused on technical parameters like latency, throughput, and energy efficiency), KVIs emphasise **sustainability, inclusiveness and trustworthiness**, helping to ensure that **6G innovations** are not only technically advanced, but also aligned with **EU goals** for sustainability, digital inclusion and global competitiveness.
- The main work on **KVIs** has been driven from **WP1** (Value, Requirements, and Ecosystem). In alignment with WP1**,** in **WP6** have been identified the relevant Human and Planetary Goals for each technical enabler in the Smart Management Framework (outlined in D6.5) and the specific values and associated KVIs linked to these goals.
- Multiple values and the indicators for them were identified, covering aspects such as **trust management, resiliency, hardware reutilisation and sharing, increased digital inclusion, environmental sustainability,** and others.
- Based on this, D6.5 provides a detailed list of the specific values and KVIs associated with each technical enabler in the management framework. However, the work on KVIs is scheduled to be continued in WP1 after the release of Deliverable D6.5.

Quantifiable targets towards the project objectives

In line with the Hexa-X-II workplan, the following Quantifiable Targets (QTs) have been addressed in the context of WP6:

Regarding Objective 4(*):

- QT 4.2: (>20%) improvement in performance in at least one of energy efficiency, latency, bit rate or area capacity through use of sensing, localisation, traffic data, or mobility patterns for AI-based optimisation in selected use cases.
- QT 4.3: Trustworthy communication and compute network services for distributed AI applications in large scales (applications with >1000 collaborating AI components).

Regarding Objective 5:

- QT 5.2: (>25%) reduction in OPEX by using zero-touch automation.

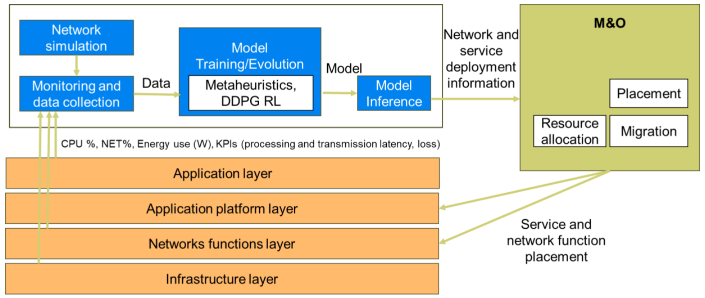
The following slides summarise how these QTs have been addressed.

(*) These refer the project objectives in Slide 8.

QT 4.2 - Improvement in performance

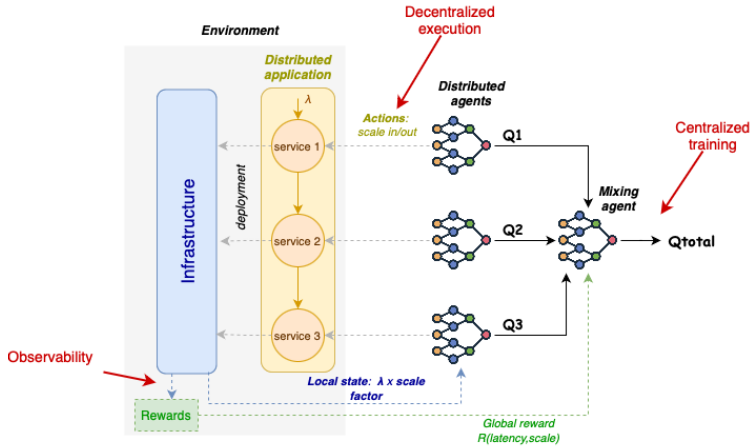
QT4.2 (>20%) improvement in performance in at least one of energy efficiency, latency, bit rate or area capacity through use of sensing, localisation, traffic data, or mobility patterns for AI-based optimisation in selected use cases.

- AI/ML derived actions: transmit power change, function allocation, resource allocation, function scaling
- Metrics: energy consumption, latency
- Improvements: 10-60% => QT is fulfilled
- Technical enablers in the framework contributing to this:

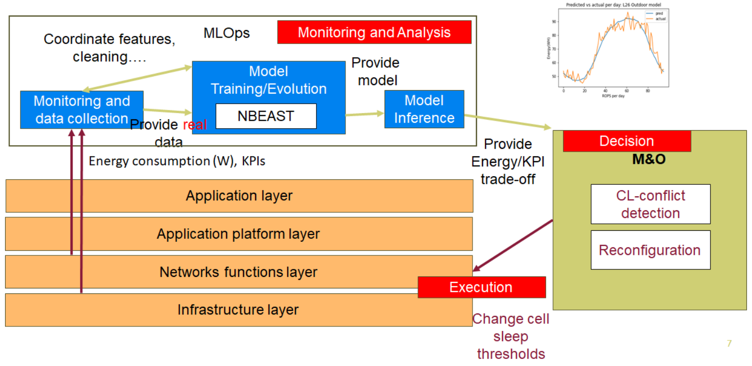


Network simulation
Monitoring and data collection
Data
Model Training/Evolution
Metaheuristics, DDPG RL
Model
Model Inference
Network and service deployment information
M&O
Placement
Resource allocation
Migration
CPU %, NET%, Energy use (W), KPIs (processing and transmission latency, loss)
Application layer
Application platform layer
Networks functions layer
Infrastructure layer
Service and network function placement

Efficient network and service function allocation



Multi-agent RL for adaptive scaling

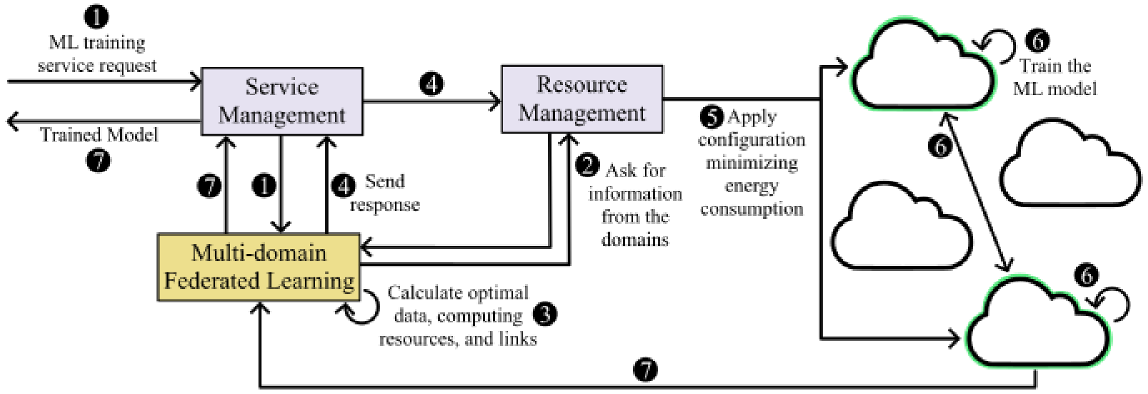


ML-based configuration recommender for energy savings

QT 4.3 - Trustworthy communication and compute network services

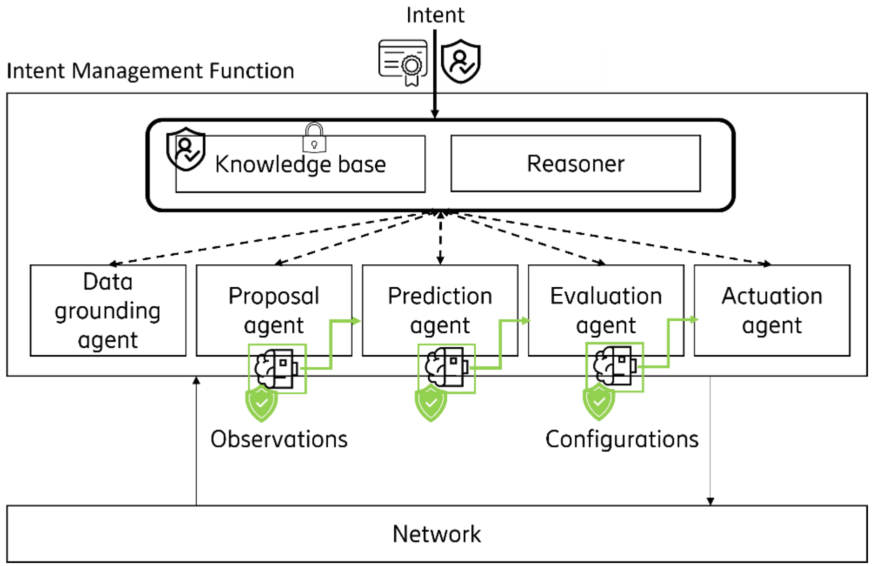
QT4.3 Trustworthy communication and compute network services for distributed AI applications in large scales (applications with >1000 collaborating AI components).

- Actions: adversarial training of ML model, resource allocation.
- Metric: accuracy under attack (MSE), training efficiency.
- Improvements: 30% in model accuracy under attack, 37% less communication overhead, 30% fewer hops per epoch between learning peers and 10% fewer epochs overall => partially fulfilled
- Technical enablers in the framework contributing to this:



❶ ML training service request
Service Management
❹
Resource Management
Trained Model ❼
❼
❶
❹ Send response
❷ Ask for information from the domains
❺ Apply configuration minimizing energy consumption
Multi-domain Federated Learning
Calculate optimal data, computing resources, and links ❸
❻ Train the ML model
❻
❻
❼

Resource assignment for federated learning



Secure AI/ML-based control for Intent-based Management

QT 5.2 - Reduction in OPEX by using zero-touch automation (1/2)

QT 5.2 aims to achieve >25% reduction in OPEX by using zero-touch automation techniques

Assessed considering three scopes:

1. Multi-agent orchestration to improve automation and decision-making in deployment and management of network services over computing continuum.
   - Using DQN agents in scaling and migration scenarios, OPEX reduction higher than 10%, due to the reduction in the operational and energy costs related to the computational infrastructure.
   - Performance of multi-agent RL techniques for automated scaling, it is shown that by applying the obtained scaling policy resulted in a 10% end-to-end latency and request rate improvement, while using 20% less resources.
2. CLs and CLs coordination for resource allocation, service provisioning, automated recovery, reconfiguration of cells' operation mode from active to sleeping mode, targeting energy consumption reduction.
   - 10-12% gain measured for energy consumption in the RAN segment.
   - Applicability of automation mechanisms for decreasing labour costs, maintenance time and costs, reducing human errors in routine operational actions, increasing resource allocation efficiency and accelerating service operation actions.
   - Extensive literature analysis showing increasing efficiency up to 20% achieved by European Telcos through zero-touch network automation and benefits from autonomous networks around 20% improvement in operational efficiency and 18% network OPEX savings, on average.

QT 5.2 - Reduction in OPEX by using zero-touch automation (2/2)

3. The Management and Coordination Engine (MCE) enables seamless automation and coordination, inspired by ETSI ZSM integration fabric, leveraging an event-driven architecture for efficient service management.
   - Adoption of an event driven architecture for microservices coordination reduces OPEX through asynchronous, decoupled communication, minimizing system-wide failures, and improving scalability. Using an event streaming platform automates real-time analytics cutting labor costs by ~30%.
   - ETSI ZSM integration fabric standardises zero-touch automation across multi-vendor networks, reducing manual configuration efforts by 70% and energy costs by 10–12% in RAN segments.
   - Enables end-to-end automation, lowering OPEX through resilient resource allocation (5% resource usage vs. 50% in legacy systems) and proactive SLA adherence (~2% violation rates).

It can be concluded that the network automation mechanisms proposed in Hexa-X-II can provide a strong contribution to OPEX reduction, with **up to 20% of reduction fully demonstrated**, even though not achieving the entire 25% reduction when not applied in combination with other enablers.

Alignment with the Advisory Group recommendations

The Hexa-X-II Advisory Group (AG), consisting of experts and representatives from key public and private stakeholders and organisations, aims to ensure that the project remains open to integrating relevant results and approaches from key stakeholders beyond the project partners. During the lifetime of the project, the GA has made several recommendations, some of them with a clear scope within the WP6 work topics. These recommendations and the answers provided are summarised below*:

Rec. 1: Attention should be paid on sustainability values, and inside that, also on social sustainability like inclusion and trust.

Answer: *Based on the work performed regarding the KVIs (see previous slides) it is considered that WP6 is in line with this recommendation. Sustainability values have been considered for all the technical enablers in the smart management framework, and in some cases specifically addressing the increase of the digital inclusion and to ensure ICT trustworthiness.*

Rec. 2: Show contrast and improvement in 6G compared to 5G.

Answer: *WP6 introduces several key improvements in management and orchestration (M&O) towards 6G, including enhanced continuum orchestration, allowing operators to manage services across multiple domains, even beyon their own network infrastructure. Also, the adoption of DevOps methodologies has expanded, with innovations like Sustainable-MLOps, optimizing ML pipeline energy efficiency. AI/ML is also increasingly integrated for complexity management, enabling advanced functionalities such as intent-based networking and resource-efficient deployments. Besides, the dynamic integration of new infrastructure resources enhances network continuum management, supported by real-time zero-touch automation. Additionally, the incorporation of Network Digital Twins (NDTs) improves M&O by enabling testing and verification, driving the shift from automation to full autonomy.*

(*) Please, consider that the content in these slides is just a summary of the answers, which are further explained in Deliverable D6.5.

Alignment with the Advisory Group recommendations (cont.)

Rec. 3: Explain how security aspects are considered in the context of WP6.

Answer: *The WP6 smart management framework emphasises security by incorporating a dedicated M&O-related functionality (labeled with the number 8 in the framework figure), assigned to various technical enablers. These include the real-time zero-touch control loops for automated security responses enabler, the 3rd-party resource control system ensuring secure multi-stakeholder management, the AI/ML-based intent management for enhanced security, the trust management for secure resource allocation functionality, and the user-centric service provisioning for flexible and secure SLAs. Additionally, security is reinforced through the MCE functionality, which rely on RBAC and secure APIs to restrict critical operations to authorised entities, ensuring secure onboarding, reconfiguration, fault recovery, and trust assessment.*

Rec. 4: Beyond the traditional way to automate network management and operations, it is considered quite important to figure out how to improve the capability to monetise the network, and how the network capabilities could be expanded beyond communication, since traditional communication is only a lowest level expectation from customers.

Answer: *While business aspects are beyond the scope of WP6 (which is a technical WP), the integration of the extreme-edge domain into M&O processes in WP6 is seen as a means to enable new and profitable business models. It is considered that this approach not only enhances revenue potential but can also helps reduce operational costs. Interoperability through microservices and APIs brings benefits such as market expansion, cross-sector collaboration, access to or elimination of technological silos, faster service deployment in new regions, and reduced time-to-market. Additionally, orchestration of compute and network resources can enhance automation and decentralised intelligence, further driving OPEX reduction.*

Alignment with the Advisory Group recommendations (cont.)

Rec. 5: In line with some considerations within the GSMA, is recommended to consider on how to provide Open APIs to expose functionalities to external players.

Answer: *WP6 contributes to interoperability and service management in several ways. The "edge convergence over federated resources" implementation explores the use of GSMA CAMARA EdgeCloud APIs for managing compute resources in the network continuum and extending them to federated external domains. Additionally, the Management Capabilities Exposure (MCE) functionality, based on ETSI ZSM Integration Fabric, enhances interoperability within telecom operators by providing event-driven service-level interfaces. While similar to GSMA's approach, it offers a different perspective on system interactions. Furthermore, WP6 includes a diverse set of OpenAPIs, further supporting seamless integration and management across domains.*

Rec. 6: It was recommended that Hexa-X-II could help to provide a sort of platform to the customers, so that, e.g., applications could be installed on the network (which is considered also as a way to monetise the networks).

Answer: *It is considered that the Decentralised Orchestration system part of the management framework presented in D6.5 is well in line with this recommendation.*

Rec. 7: It was recommended to consider GenAI-related techniques.

Answer: *The WP6 smart management framework is AI/ML-agnostic, allowing the integration of various techniques without enforcing specific algorithms. However, some implementations are based on Gen-AI models, such as Conditional Variational Autoencoders (CVAE) for scenario optimisation. Also, other state-of-the-art Generative AI approaches could also be integrated, including GenAI-based NLP for intent-based management, GANs for generating synthetic network data, and Transformer-based networks could be used in intelligent network traffic management by analysing large volumes of network traffic data in real time, with application on real-time traffic analysis, demand prediction, or anomaly detection, among others. This flexibility ensures adaptability to evolving AI advancements in network management.*

Conclusions

Conclusions

Deliverable D6.5 provides a fundamental contribution towards the design of the Hexa-X-II E2E 6G System Blueprint, presenting the design of the 6G Smart Network Management Framework. The detailed framework is the outcome of the work in the lifetime of WP6 in the HEXA-X-II work programme. It has been produced, following the specification of a set of enablers to support smart network management functionalities, as detailed in D6.3 [HEX224-D63]. Based on the detailed enabling technologies, specific components have been derived and constitute part of the framework. The set of components are classified as overall M&O solutions, overall functionalities, specific systems and algorithms. Each one of the components has a specific role, while their synergy is required to support end-to-end network service management.

Upon the specification of the components of the 6G Smart Network Management Framework, their design and development are detailed. Implementation details are provided for the supported mechanisms and technologies per component. Following, a set of evaluation results are presented for individual components as well as workflows that consider the collaboration among multiple components. Part of these results is produced based on the work in progress in the PoCs. A set of KPIs are defined and considered in the evaluation.

In the provided 6G Smart Network Management Framework, a set of novel mechanisms and approaches are specified that will play dominant role in the evolution towards the 6G networks. Decentralised intelligence, automation, privacy and security characteristics are strongly considered and supported in the various management and orchestration mechanisms that are detailed. The exploitation of AI/ML technologies is included in most of the developments within WP6, while challenges related to the need for training and evaluation of the ML models to achieve the desired accuracy are identified. A variety of solutions for supporting orchestration actions is introduced based on the emergence of multi-agent approaches, cognitive control loops, decentralised schemes, and federation techniques. Their adoption and/or combination for tackling orchestration challenges for resources spanning across the network continuum is suggested.

Insights and recommendations for the future (1/3)

These and the following slides provide an overview of the next steps considered for the development of the concepts and technical enablers in future iterations.

- Trends and future work for the overall M&O solutions
  - Multi-agent M&O techniques:
    - Consider the combination of multi-agent techniques with reliable and explainable AI frameworks to improve adoption of mechanisms.
    - Consider the emergence of agentic AI frameworks where multiple agents can collaborate and undertake different tasks (e.g., recommenders, RL-driven actions).
    - Study the opportunity to combine these approaches with modern observability stacks (e.g., based on open telemetry specifications) to assist decision making by agents.
    - Consider to provide support regarding the integration of networks with far edge/edge/cloud technologies for time critical operations (e.g., real time data processing of high data volumes).
    - Examine trust, security and privacy aspects by the agents.
  - Decentralised M&O techniques:
    - Further development of components to tackle scalability aspects.
    - Study and integration of alternative AI/ML models for the ISPM component.
    - Examination of synergies with other M&O approaches.

Insights and recommendations for the future (2/3)

- Trends and future work regarding the overall functionalities in the framework
  - Monitoring and telemetry:
    - It is considered that the innovations in TeraFlowSDN, MEC integration, and automation frameworks can pave the way for a more flexible, intelligent, and future-proof networks, aligning with emerging 6G requirements.
    - Plan for the refinement of the supported AI/ML mechanisms for automation, as well as the promotion of the implementations toward standardisation.
  - Zero-touch RT network automation:
    - New descriptors would be provided for unified CL modelling, mechanisms and workflows for governance and coordination.
    - New mechanisms to support real-time CL functions and governance, combined with monitoring and telemetry functionalities, and multi-cluster resource orchestration mechanisms.
  - Management Capabilities Exposure:
    - In the context of Hexa-X-II an Integration Fabric has been developed as inspired by ETSI ZSM, acting as a new integration/exposition based on event-driven approach. This approach could be generalised in the future, elevating it as an active component to decentralise the way to coordinate network functions and components.

Insights and recommendations for the future (3/3)

(cont. from the previous “overall functionalities“ bullet)

- SLA-driven federated orchestration
  - improvements are planned on prediction-based triggers of Smart Contracts for tighter latency bounds
- Trust management:
  - a trust assessment function and an associated ontology have been designed for the network continuum, along with the definition of internal interfaces and data models between TEF and LoTAF
  - it is envisaged that the common trust management model will be able to share information across multiple domains, trust will be delivered as a transparent notary service (TNS) to promote external audits, trust level agreement will be declared to formalise end-user’ requirements, while intent-based trust management will be supported

- Trends and future work for specific systems and algorithms:
  - The Network Digital Twins (NDT) concept has been introduced in Hexa-X-II to improve automation and support optimal decision making. This work could be considered as a starting point for future implementations.
  - The Sustainable MLOps approaches could be further extended to support different AI/ML paradigms and ML workloads.
  - Network programmability solutions are expected to expand significantly, supporting advanced orchestration across the network continuum in the future. These solutions are considered essential, particularly when integrated with Open APIs that enable convergence with edge/cloud orchestration (e.g., CAMARA APIs).
  - Third-party resource control separation and user-centric service provisioning mechanisms explored in Hexa-X-II are envisaged to be promoted within the 3GPP SA5 activities. These solutions extend Role-based Access Control (RBAC) mechanisms with dynamic, model-driven permissions tailored to tenants, by using dynamic URSPs (User Equipment Route Selection Policies) for personalised service activation, as well as the integration with closed-loop automation ensuring SLA compliance.

References (1/2)

[HEX223-D21]	Hexa-X-II, “Deliverable D2.1 Draft foundation for 6G system design”, June 2023.
[HEX224-D24]	Hexa-X-II, “Deliverable D2.4 End-to-end system evaluation results from the interim overall 6G system. Draft foundation for 6G system design”, Sept. 2024.
[HEX224-D33]	Hexa-X-II, “Deliverable D3.3: Initial analysis of architectural enablers and framework”, April 2024.
[HEX224-D63]	Hexa-X-II, “Deliverable D6.3: Initial Design of 6G Smart Network Management Framework”, June 2024.
[JAN93]	J.S.R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665–685, May/Jun. 1993, doi: 10.1109/21.256541.
[PC91]	P. Spirtes and C. Glymour, “An algorithm for fast recovery of sparse causal graphs,” Social Science Computer Review, vol. 9, no. 1, pp. 62–72, 1991.
[SCR02]	D. M. Chickering, “Optimal structure identification with greedy search,” Journal of machine learning research, vol. 3, no. Nov, pp. 507–554, 2002.
[KAR24]	M. Karaca, J. Sadasivan, A. C. Baktir, A. Palaios and A. Zahemszky, "Utilizing Causal Learning for Cognitive Management of 6G Networks," 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), Stockholm, Sweden, 2024, pp. 234-239, doi: 10.1109/ICMLCN59089.2024.10624867.
[ZSM002]	Zero-touch network and Service Management (ZSM); Reference Architecture (https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf)
[SQLA24]	SQLAlchemy, https://github.com/sqlalchemy/sqlalchemy, rel 2.0.35, Sept. 2024
[FAPI24]	FastAPI, https://github.com/fastapi/fastapi, rel 0.115.2, Oct. 2024
[DMIC23]	Adamczyk, C., & Kliks, A. (2023). Detection and mitigation of indirect conflicts between xApps in Open Radio Access Networks. arXiv preprint arXiv:2305.13464.
[MNC+17]	Muñoz, R., Nadal, L., Casellas, R., Moreolo, M.S., Vilalta, R., Fàbrega, J.M., Martínez, R., Mayoral, A. and Vílchez, F.J., The ADRENALINE testbed: An SDN/NFV packet/optical transport network and edge/core cloud platform for end-to-end 5G and IoT services. EUCNC 2017.
[PBM+24]	R. Pires, H. Blue, J. Malinen, M. De Angelis, P.G. Giardina, G. Landi, M. Laukkanen, K. Aloha, and P. Porambage, "Closed-Loop Automation in 6G for Minimum Downtime Task Continuity in Surveillance Cobots", EuCNC and 6G Summit, June 2024.

References (2/2)

[TAZ+22]	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(5):2061. https://doi.org/10.3390/s22052061
[OTL24]	OpenTelemetry, available at: https://opentelemetry.io/
[CEC24]	CAMARA Project. (n.d.). MEC Exposure and Experience Management API Specification. Retrieved from https://github.com/camaraproject/EdgeCloud/blob/main/documentation/SupportingDocuments/API%20proposals/Discovery/MEC%20exposure%20and%20experience%20management.yaml, 2024
[VVG+24]	R. Vilalta, F. J. Vílchez, Ll. Gifre, C. Manso, J.L. Carcel-Cervera, R. Leira, J. Aracil-Rico, J.P. Fernández-Palacios, R. Martínez, R. Casellas, R. Muñoz, Providing Anomalous Behaviour Profiling by extending SmartNIC Transceiver support in Packet-Optical Networks, OFC, 2024.
[SCITT-09]	Draft-IETF-SCITT-Architecture-09: “An Architecture for Trustworthy and Transparent Digital Supply Chains”, October 2024.
[ZFF+24]	A. Zafeiropoulos, N. Filinis, E. Fotopoulou and S. Papavassiliou, "AI-Assisted Synergetic Orchestration Mechanisms for Autoscaling in Computing Continuum Systems," in IEEE Communications Magazine, doi: 10.1109/MCOM.001.2200583.

HEXA-X-II

HEXA-X-II.EU //

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.