Partners: NFI, EAB, AAU, AAP, ASA, EBY, ICC, LTU, TNO, NXW, NGE, OPL, TUD, TIM, UBW, UC3, WIN, TUK, SON

HEXA-X-II D3.3 Deliverable

Initial analysis of architectural enablers and framework

Hexa-X-II

hexa-x-ii.eu

2024-04-30

Table of Content

- Objectives
- 6G Architecture overview and migration
- AI enablers for data-driven architecture
- Network modularisation
- Architectural enablers for new access and flexible topologies
- Network beyond communications
- Virtualization and Cloud transformation (IoT-Edge-Cloud continuum enablers)
- Summary

Chapter 1

Objectives

Deliverable D3.3 objective

- This is the second public deliverable D3.3 from WP3
- The main objective of this deliverable is to make an analysis of the WP3 enablers and at the same time define new requirements that is important for the 6G architecture
- **These slide-sets gives a high-level overview of D3.3**

WP3 objectives

The overarching objective of WP3 is to develop a 6G architecture framework and innovative enablers for beyond communications and data driven architecture to power new services, modular cloud-native network for improved signalling and new access and flexible topologies for improved reliability.

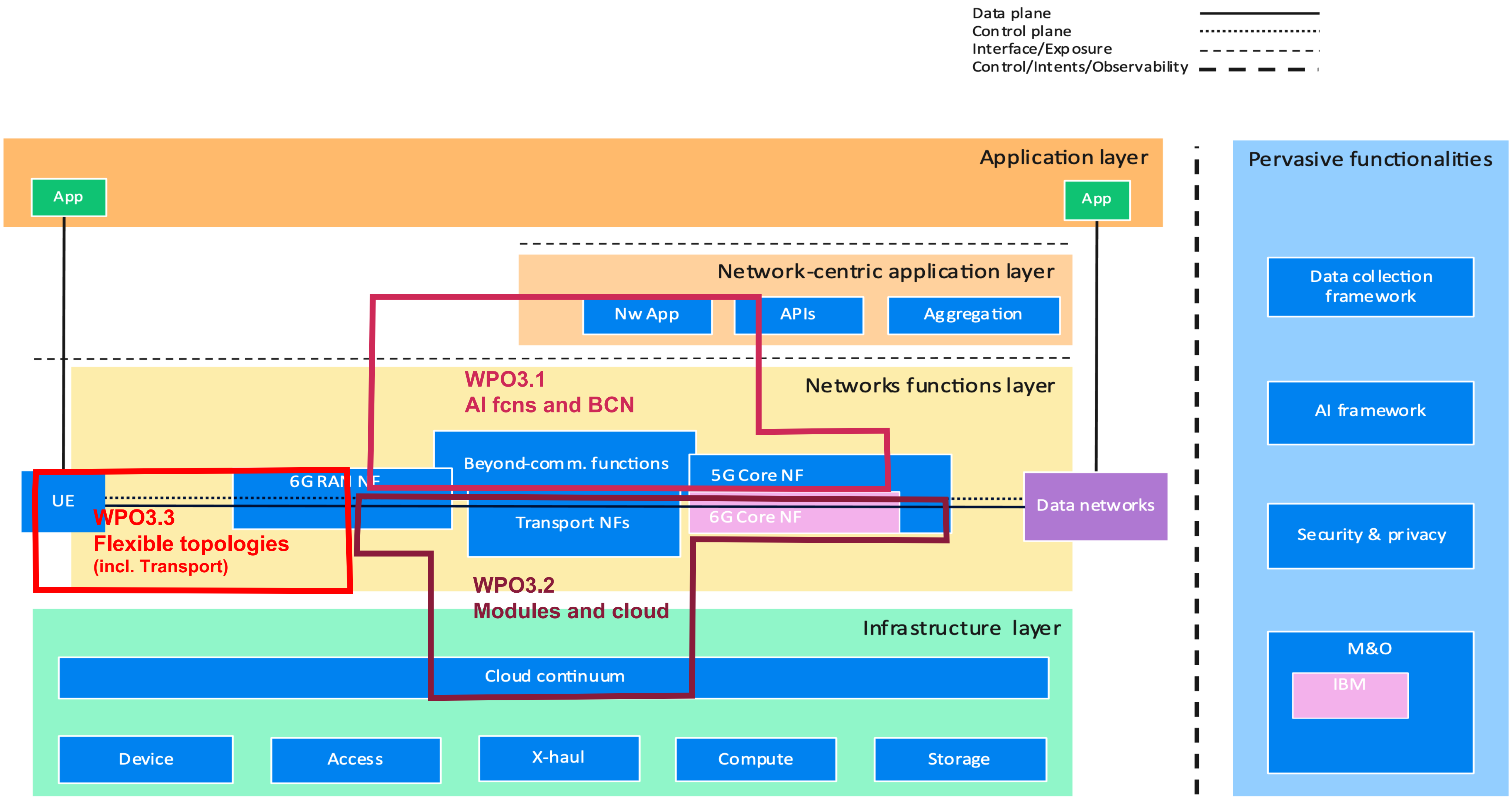
The WP objectives are:

- **WPO3.1**: Develop and analyse 6G architecture framework and new innovative enablers for the beyond communications and data driven architecture, identify the requirements a data-driven architecture will have on protocols, interfaces, data, and network nodes.
- **WPO3.2**: Define and analyse solutions that combine cloud technology flexibility with distributed processing nodes into self-contained modules with minimum dependency that can be used to extend and scale the network deployments in stepwise manner.
- **WPO3.3**: Develop and analyse new access for flexible topologies and local communications, including different types of multi-connectivity, node roles and node coordination, as well as design control and management solutions for programmable and context-aware transport.

Chapter 2

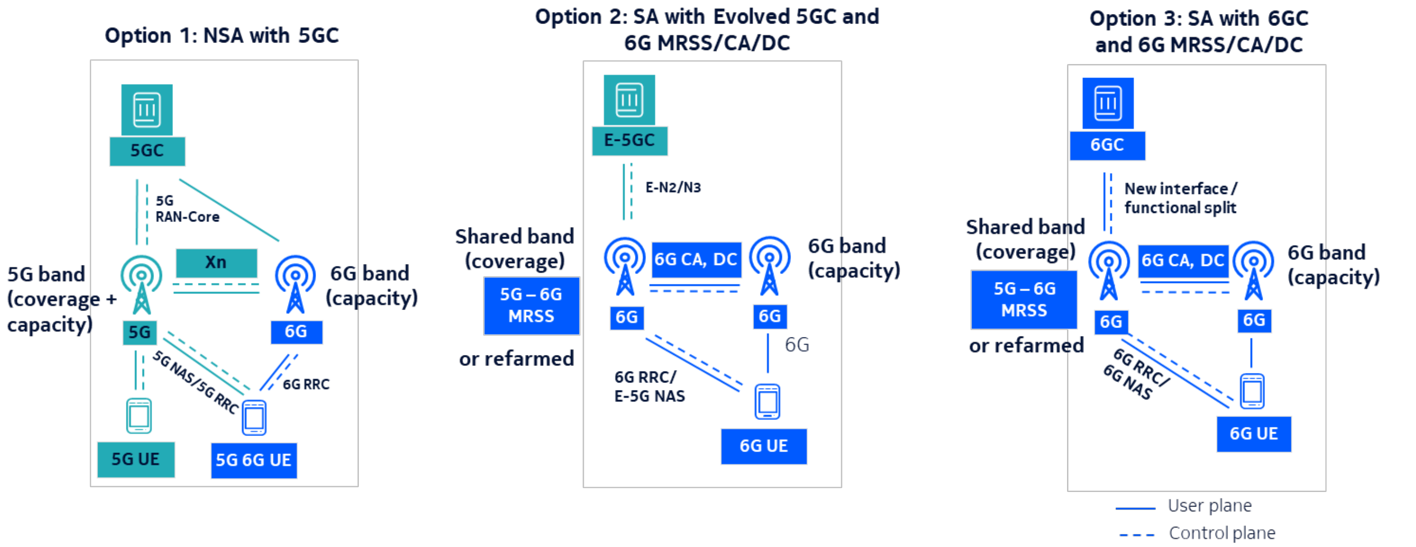
6G Architecture overview and migration

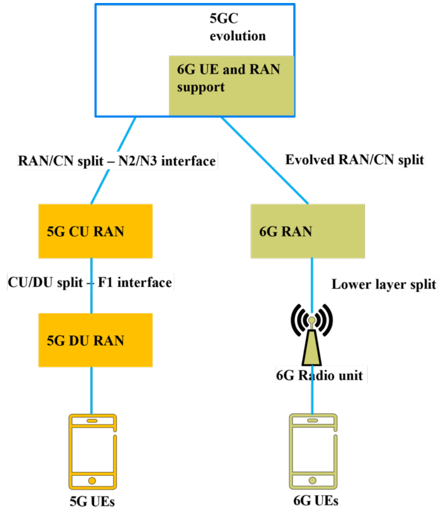
Chapter 2 WP3 objectives mapping on system blueprint



Chapter 2.1 Migration

- The goal is to have fewer architecture options specified the deployment of 5GC
  - Reduce the 5G-6G interworking and deployment complexity
  - Avoid delays introducing key 6G services
- The main architecture options evaluated for 6G migration
  - Option 1: 6G RAN is connected to 5G RAN using DC.
  - Option 2: 6G RAN is deployed standalone and connected to an Evolved 5GC (E-5GC). The E-5GC is built upon 5GC NFs which can be enhanced to handle 6G requirements. The E-5GC serves both legacy 5G RAN and new 6G RAN while supporting the 5G and 6G features at the same time
  - Option 3: 6G RAN is deployed in stand-alone mode with a 6G Core (6GC), which allows for non-backward compatible changes. A clear separation between 5GC and 6GC is presented, where only a few NFs would be shared.
- UEs controlled by a single RAN control function (i.e., LLS)
  - Simple standardization, reduced signalling





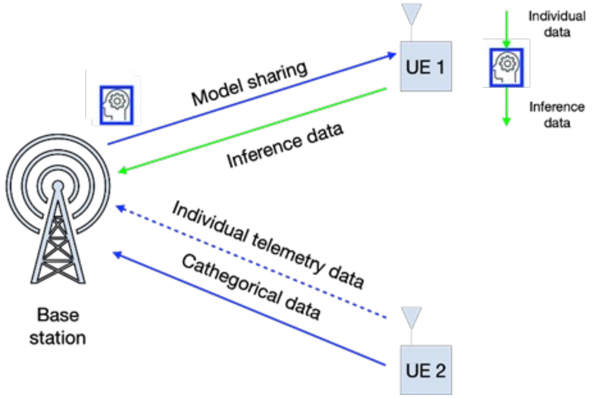
Chapter 3

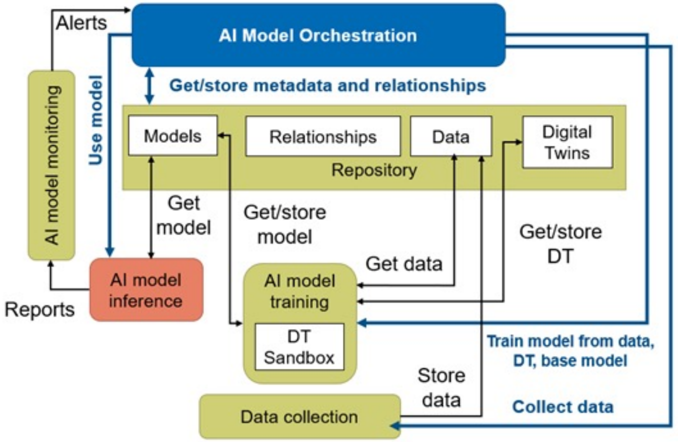
AI enablers for data-driven architecture

Chapter 3.1 Architectural Means and Protocols

- **Background**
  - The structure, design, and principles governing AI component interaction and data handling.

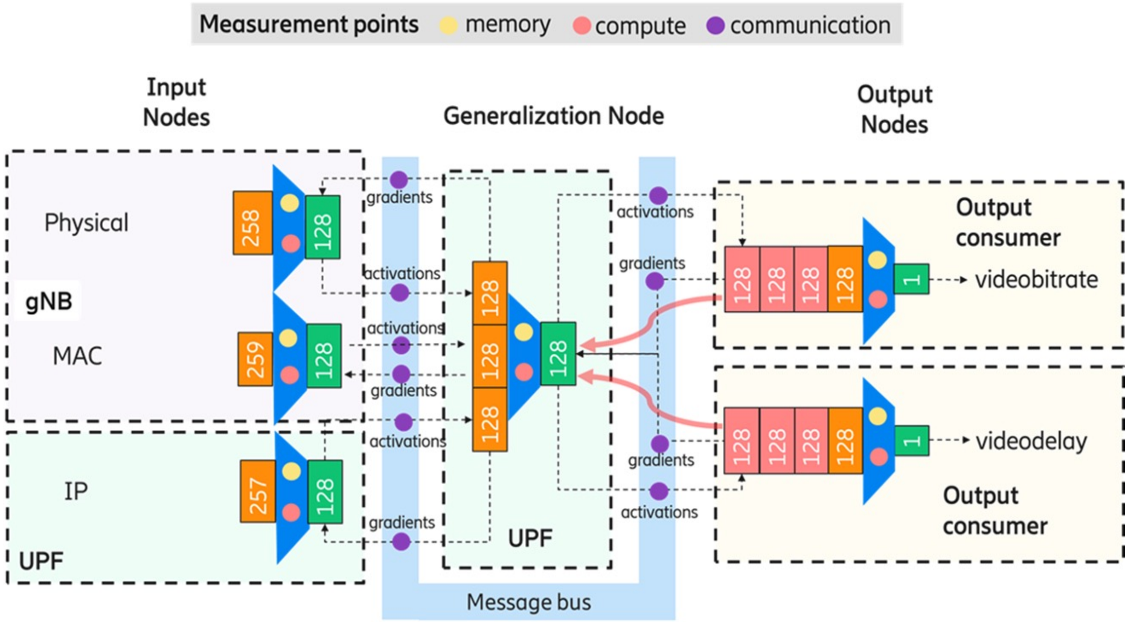
- **Benefits**
  - Scalability
  - API Response Time
  - MLOps, DataOps, and AIaaS depend on architectural means and protocols for seamless integration and efficient operation. Interoperability
  - Principles #1, #2, #4 and #5
-
- **Implication**
  - Standardized protocols for integration,
  - Scalability,
  - Robust security,
  - Metadata management,
  - Interoperability - protocols for cross-functional collaboration, API design, and adaptability to emerging technologies.
  - TR 38.817 for RAN Intelligence





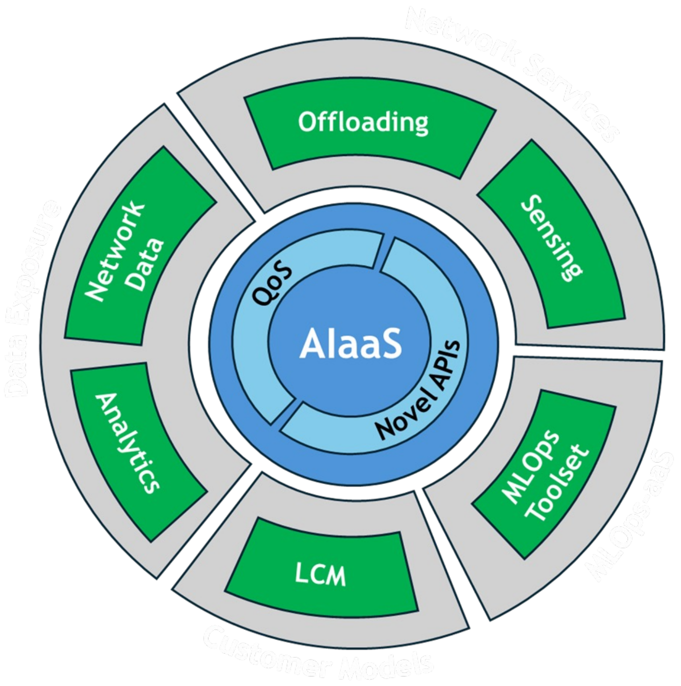
Chapter 3.2 Machine Learning Operations (MLOps)

- **Background**
  - Streamline of the entire life cycle of ML models during development, automating deployment through CI/CD pipelines.
- **Benefits**
  - Ensure model accuracy and effectiveness through continuous monitoring and improvement,
  - Maintain the relevance of models by automating (re)training and deployment with new data,
  - Implement robust monitoring systems for early issue detection,
  - Smooth deployment and monitoring of ML models enabling seamless deployment and monitoring throughout the data lifecycle.
  - Architectural means and protocols support collaboration and data integration, crucial for MLOps.
  - AIaaS relies on MLOps for seamless integration and operation, aligning communication protocols and data exchange formats for scalability and cost-effectiveness.
  - Principles # 2, #3, #6, #10
- **Implication**
  - high-quality data, scalable storage, communication between computation nodes, data processing, model training and inference, data and model alignment, and data versioning.
  - Automation via CI/CD pipelines and robust monitoring are crucial, as are model explainability and interpretability.
  - Security, compliance, feedback loops for continuous improvement, and resource optimization are also essential.
  - TR 38.817 for RAN Intelligence



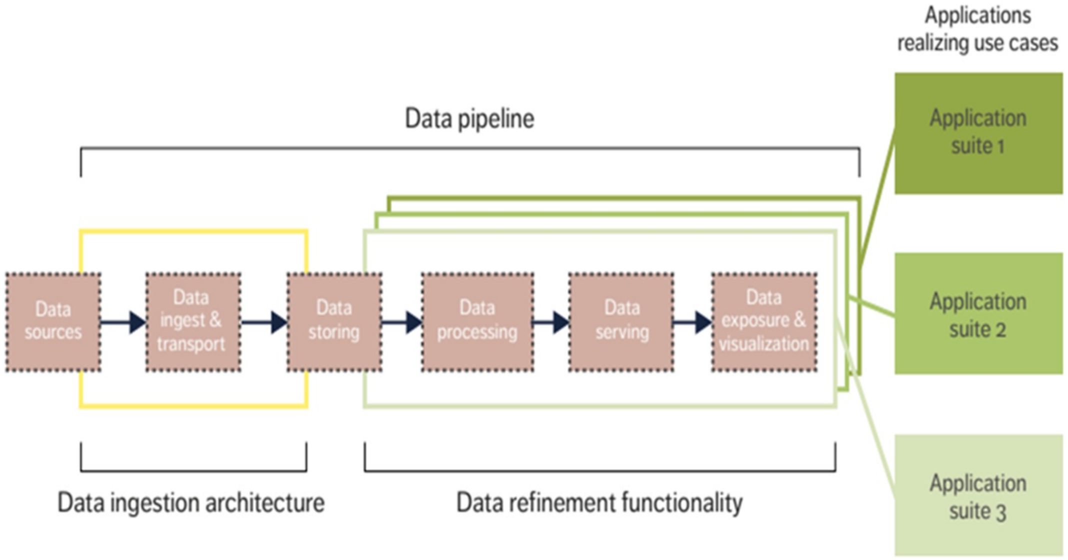
Chapter 3.3 Artificial Intelligence as a Service (AIaaS)

- **Background**
  - Paradigm shift where the mobile network becomes a catalyst for innovative use cases by providing accessible AI capabilities through a variety of APIs, thereby eliminating the need for application developers to build and manage their AI infrastructure.
- **Benefits**
  - model efficacy, scalability, and API availability/reliability.
  - DataOps to streamline data workflows by incorporating AI capabilities for enhanced data processing and analysis.
  - Architectural means and protocols can be designed to seamlessly integrate AIaaS, aligning communication protocols and data exchange formats to accommodate AI-driven processes efficiently.
  - MLOps can utilize AIaaS for model deployment, monitoring, and optimization, ensuring the operational efficiency and effectiveness of ML models within the architecture.
  - Principles #1, #2, #3, #4 and #8
- **Implication**
  - high-quality data, scalable storage, and processing, and data versioning.
  - Automation through CI/CD pipelines and robust monitoring is crucial, along with a focus on model explainability and interpretability.
  - Security measures and compliance with regulations, feedback loops for continuous improvement and resource optimization are essential.
  - ETSI GS ZSM 012, 3GPP TR 23.700-82
  - EU AI Act



Chapter 3.4 Data Operations (DataOps)

- **Background**
  - DataOps ensures data quality by cleansing and validating data, crucial for precise AI model training. It fosters collaboration and efficiency accelerating development through automation, integration, and monitoring
- **Benefits**
  - accuracy rate, automation rate, and data drift detection rate
  - Architectural means and protocols can integrate DataOps seamlessly, ensuring compatibility and interoperability.
  - MLOps relies on DataOps for efficient data preparation, enabling smooth deployment and monitoring of ML models.
  - AIaaS benefits from DataOps by leveraging its streamlined data workflows and efficient processes
  - Principles #1, #2, #3, #4, #5, #6, #7 and #8
- **Implication**
  - Implementing robust data quality management,
  - establishing automated end-to-end data pipelines,
  - implementing version control for data artifacts



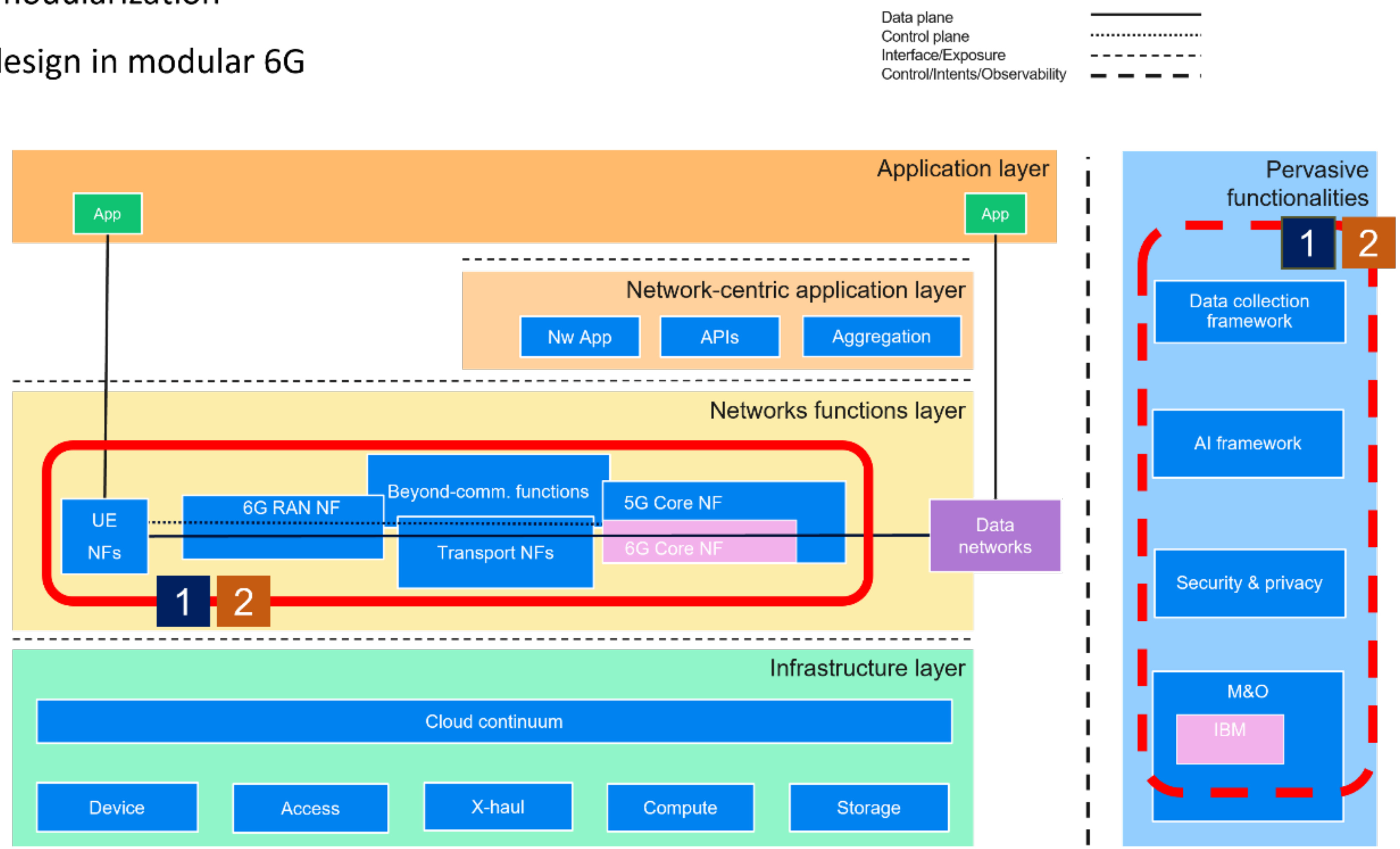
Chapter 4

Network modularisation

Mapping of network modularization enablers on system blueprint

1 6G network modularization

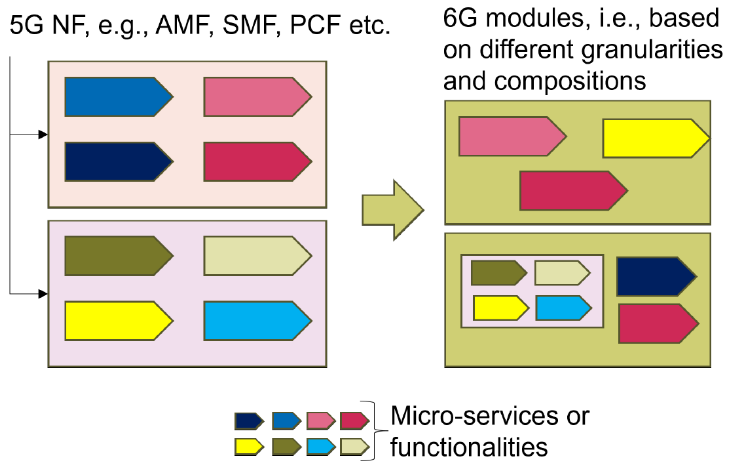
2 E2E service design in modular 6G



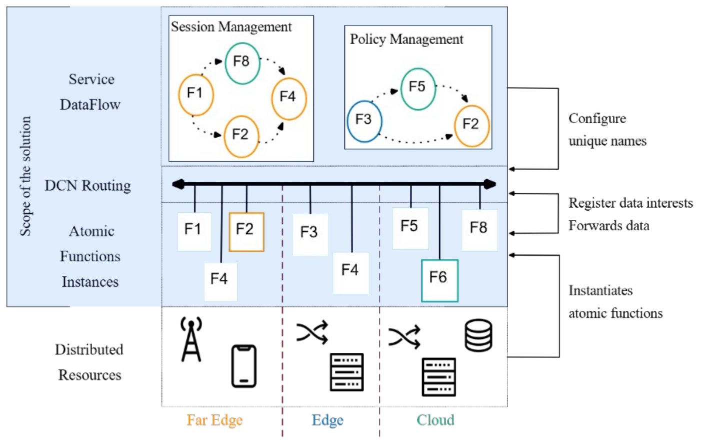
6G Network modularisation

- Description
  - Different granularities of a module, how to design a module and the evolution of the modules based on different deployment locations and use cases
  - Investigates the streamlining the interactions between different modules/domains
- Benefits
  - KPI improvements
    - Decrease latency (E2E), procedure completion time, CP signalling, and it will reduce complexity.
    - Increase efficiency, enhancing network scalability, flexibility (i.e., both deployment and execution), and reliability. It will provide uniform and reliable service for all users
  - It can serve as a basis for all other enablers, including but not limited to E2E service design in modular 6G, JCAS protocols, and AIaaS.
  - Principles: #3, #4, #5, #6, #7, #8,#9
- Implication
  - Requires the management functions between RAN and CN to be revisited and changing the core NF design and implementation.
  - As the new NF compositions evolve, new interfaces might be needed for inter-node network-level coordination.
  - Redesign of the architecture considering new approaches is needed which might need a “clean-slate” approach.
  - The network composition changes as well as the relevant interfaces and interactions would impact all the 3GPP standards including but not limited to [23.501], [23.502], [38.413].
  - All the solution is mostly software based

Decomposition of 5G NFs with changing granularity

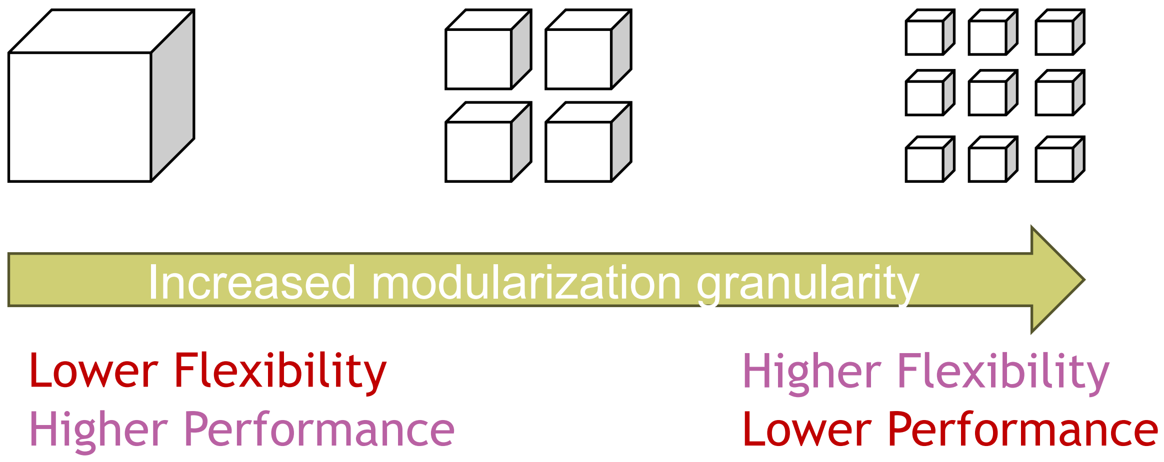


Data-centric Service-Based Architecture for Edge-Native 6G Network



Advantages and Disadvantages of increased modularization granularity

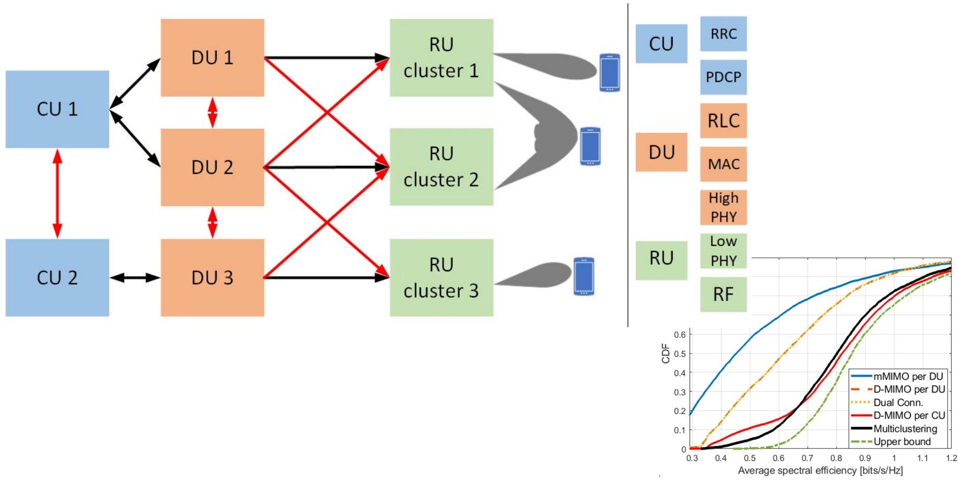
Advantages	Disadvantages
Efficient and targeted scaling of modules	Lower performance in terms of delay
Easier to identify faulty modules	Additional interfaces definition between modules
Faster development cycles and flexible selection of development tools	Management overhead of a bigger number of modules
Flexibility for distributed deployments	More signaling and data exchange is needed
Efficiency by reusing fine-grained modules	More memory transactions are needed to read and store stateful data



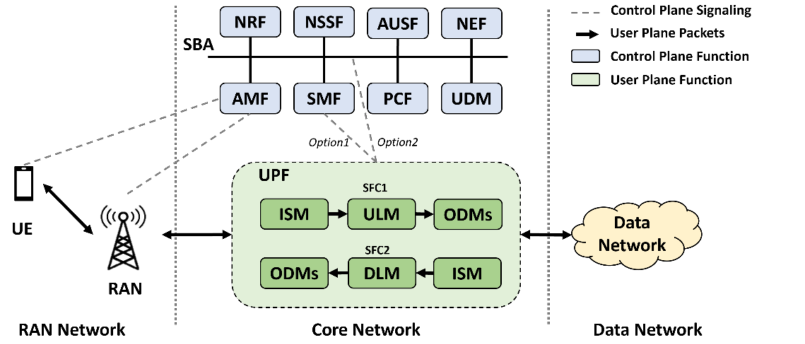
E2E service design in modular 6G

- Description
  - The E2E modular 6GS design for services, that includes, the modular UP/CP design, network autonomy and adaptiveness via modularization and flexible orchestration.
- Benefits
  - KPI improvements:
  - Decreased latency, CP signalling and the number of hops. Increased user spectral efficiency, same front-haul user-plane traffic.
    - Increased efficiency by lower number of interfaces, scalability, flexibility in terms of deployment and execution, and resiliency
  - Depends on the findings on 6G network modularization. As this enabler is a fundamental enabler that has implications to all parts of the 6GS, it has impact on all the other enablers and functionalities within RAN, CN and M&O.
  - #1, #2, #3, #4, #5, #7, #8, #9
- Implication
  - The legacy implementations need to be analysed and fixed where needed.
  - All the proposed changes need to be considered within the gains and loses with respect to the legacy implementations and all non-backward compatible changes need to be justified.
  - The proposed changes need to support new use cases and services and an analysis of different solution options needs to be considered.
  - Due to the its impact on the 6G NF/module design and the respective interactions between RAN and CN, the major standards that are envisioned to be impacted by this enabler includes but not limited to [23.501], [23.502], [38.410], [29.510], [24.501], RAN modules defined in [38.401], O-RAN considers an open-interfaced disaggregated RAN.

Logical RAN architecture enabling distributed cell-free operation.



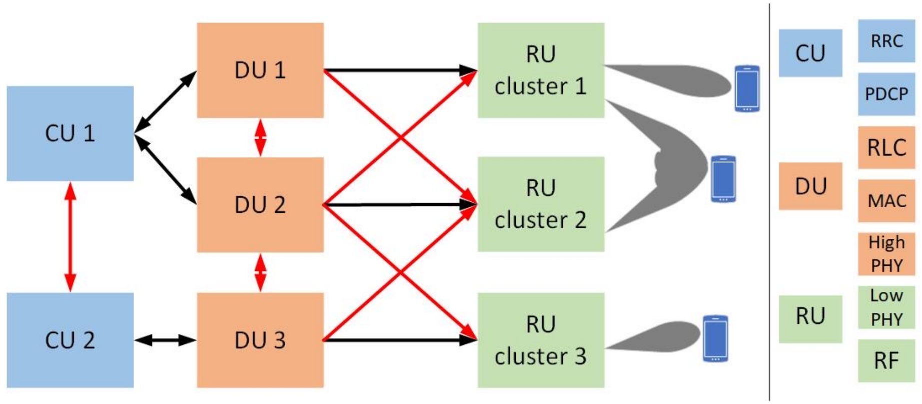
Modular UPF design integrated in the E2E mobile network system



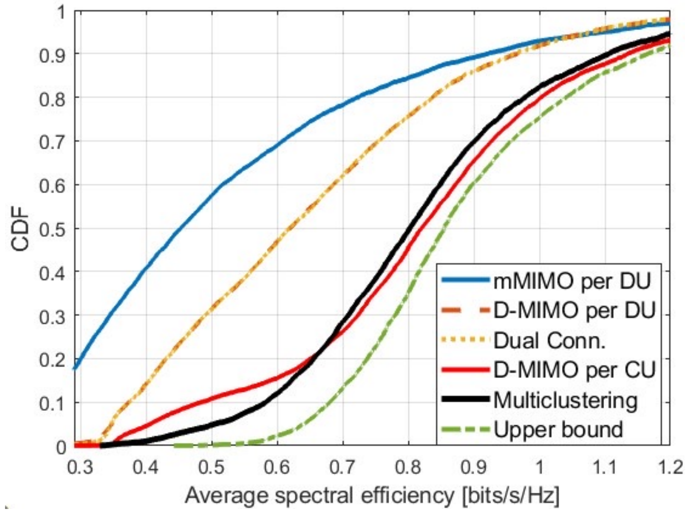
RAN Modularity

- Disaggregation of RAN functionalities in different entities is contemplated as an architectural option for 6G RANs. Functional splits are defined for that purpose.
- Disaggregation and virtualization of RAN components allow for increased reconfigurability, interoperability and resiliency of the networks.
- User-centric cell-free massive MIMO networks are a strong contender for 6G.
- To enable cell-free operation, focus should be put on suitable functional splits.
- Comparing the results of the proposed approach (multi-clustering) to other benchmarks, such as single clustering D-MIMO and collocated Massive MIMO, it can be seen that it provides better service for users, especially for those who suffer from the cell-edge problem.
  - The gain is higher in users which suffer from high co-channel interference in single clustering D-MIMO, as can be seen by comparing the spectral efficiency of the 5th percentile.
  - Additionally, since the functional split defines the front-haul as carrying frequency domain samples, and the partition is also done in the frequency domain, the proposed solution does not increase the fronthaul traffic, compared to a single-clustering solution (D-MIMO).

Logical RAN architecture enabling distributed cell-free operation.



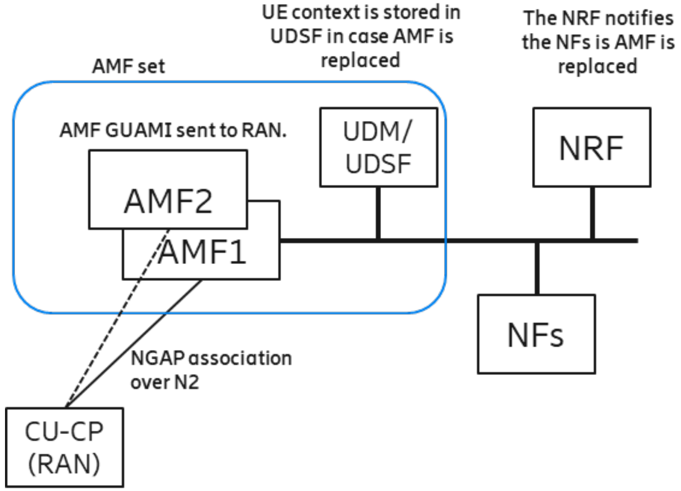
Empirical CDF of the user average spectral efficiency, comparing RRH multi-clustering with different benchmarks.



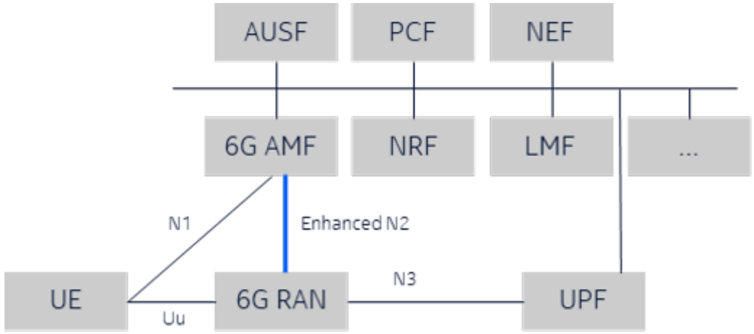
Investigation of possible RAN-CN SBI

P2P connection between RAN and CN

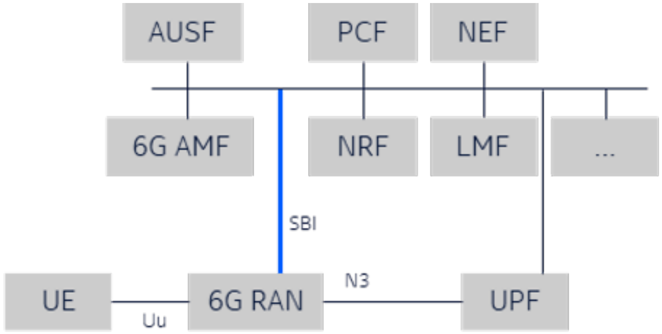
- Interface between the RAN and AMF is called N2; relies on NGAP
- Expected RAN and CN CP changes to support new 6G radio features and new services



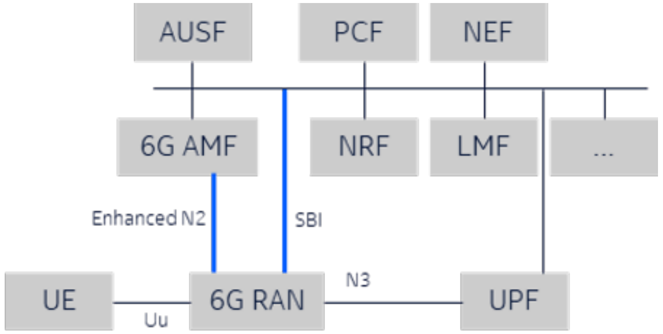
Investigated options



Option A: Enhancing NGAP

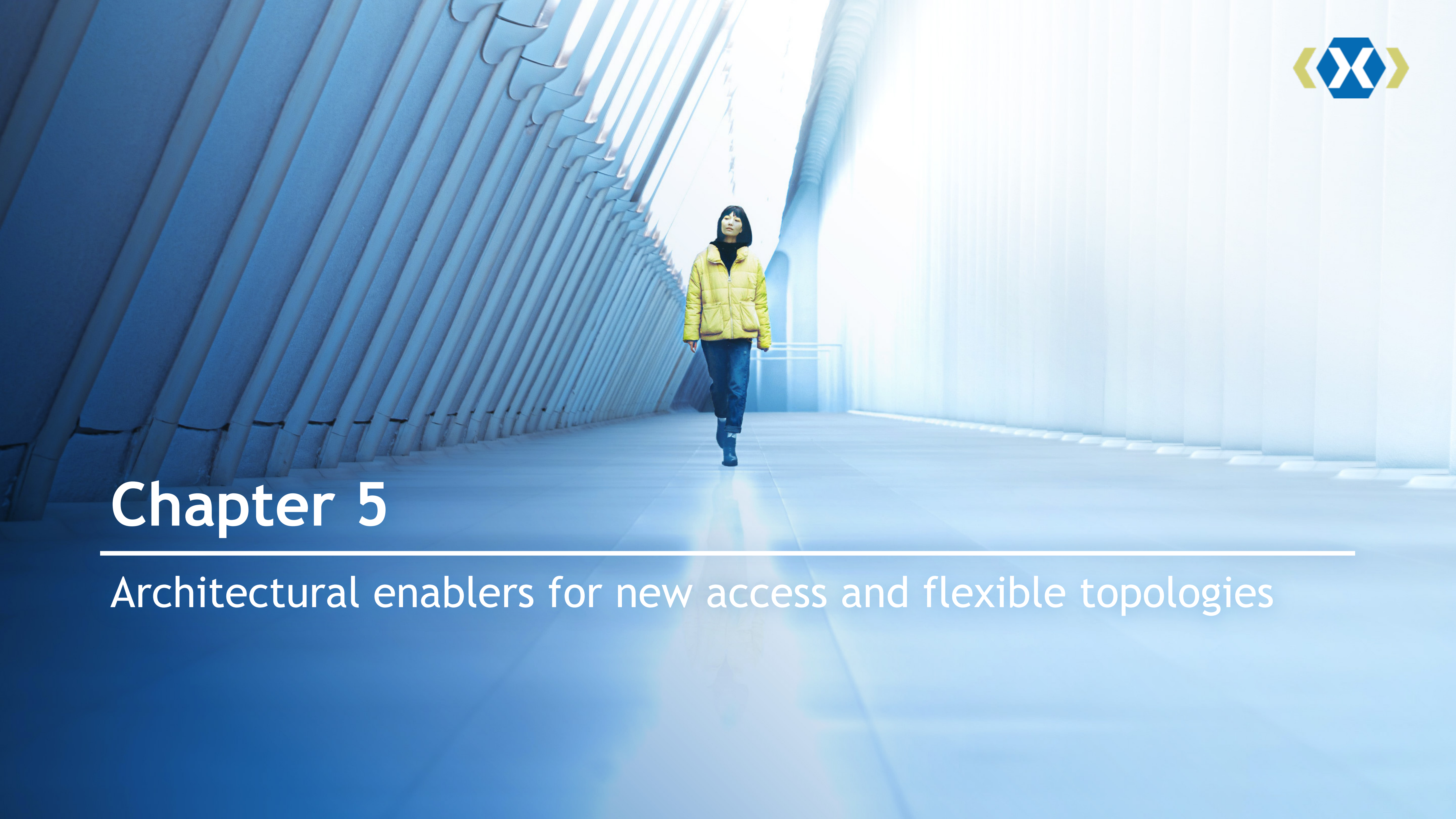


Option B: RAN-CN Control Plane SBI



Option C: RAN-CN Control Plane SBI+NGAP

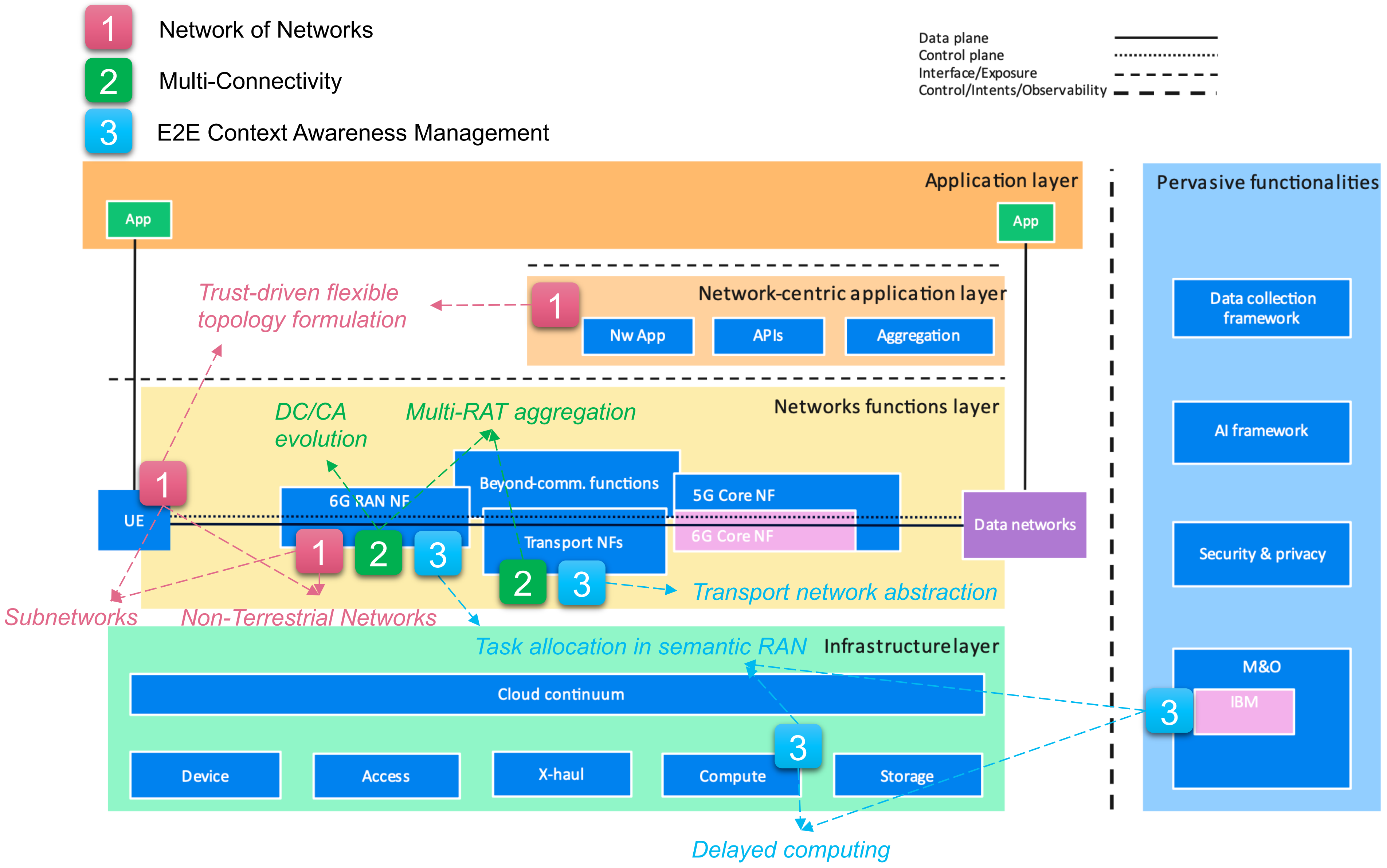
SBA can replace the NGAP functionality however, there are no clear benefits.



Chapter 5

Architectural enablers for new access and flexible topologies

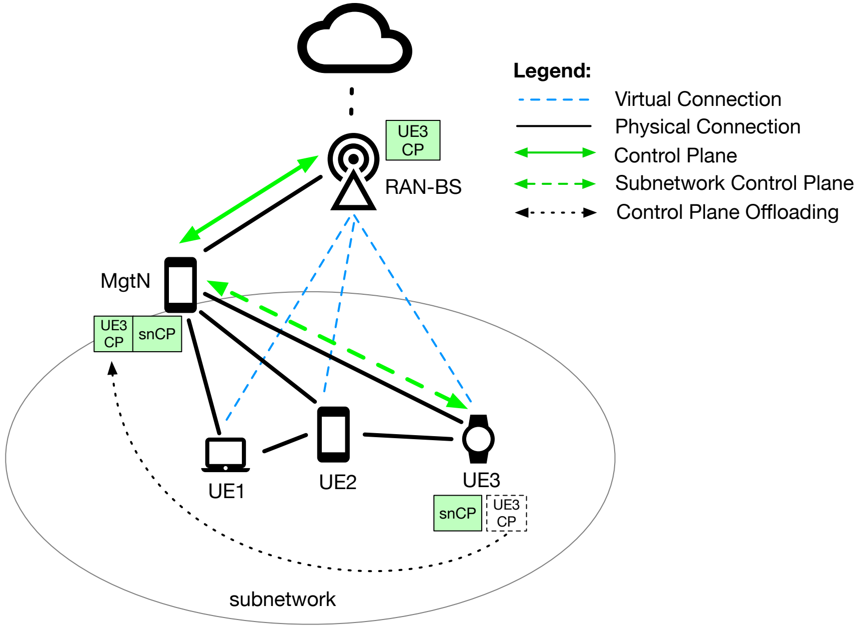
Mapping of architectural enablers for new access and flexible topologies on system blueprint



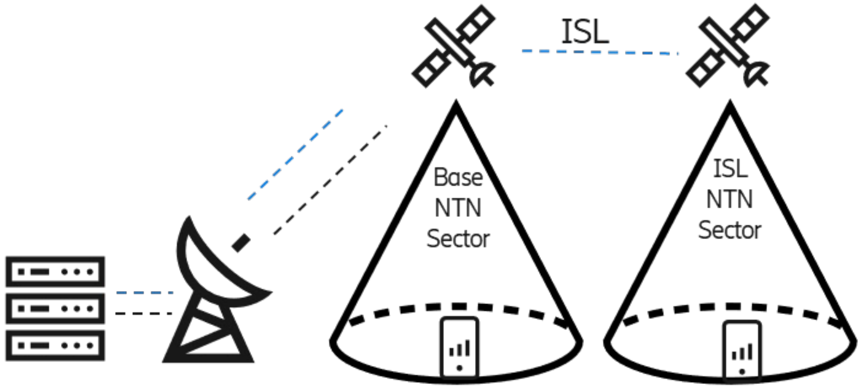
Network of networks

- Design of terrestrial subnetworks and NTN to create a seamless and ubiquitous communication system
- **KPI improvements:** Coverage extension, reduced interruption time and time in outage, increased availability and reliability, reduced complexity, enhanced mobility, service continuity
- **Relation to other enablers:** Mobility and flexible radio protocol design should take this enabler into consideration. Both low-cost and high-capability devices could use the subnetwork framework
- **Requirements**
  - The Inter-Satellite Link (ISL) solution for 6G NTN depends on which NTN RAN architecture is selected (e.g., BS or RU onboard)
  - TN-NTN dual connectivity procedure to switch faster to NTN when there is no good TN coverage
  - Two architectural options for subnetworks: transparent to the NW and non-transparent to the NW
  - New lightweight subnetwork CP (snCP) between the Management Node (MgtN) and the UEs in a subnetwork
  - Device-centric or network-centric trust of diverse nodes
- **Standardisations & regulations**
  - Regulators will have to develop new measures to regulate national coverage based on NTN
  - Architecture of subnetworks, including the new UE roles and their responsibilities: 3GPP RAN2
  - TN-NTN dual connectivity procedures.: 3GPP RAN2

Subnetworks architecture: Subnetwork control plane with the aid of the MgtN



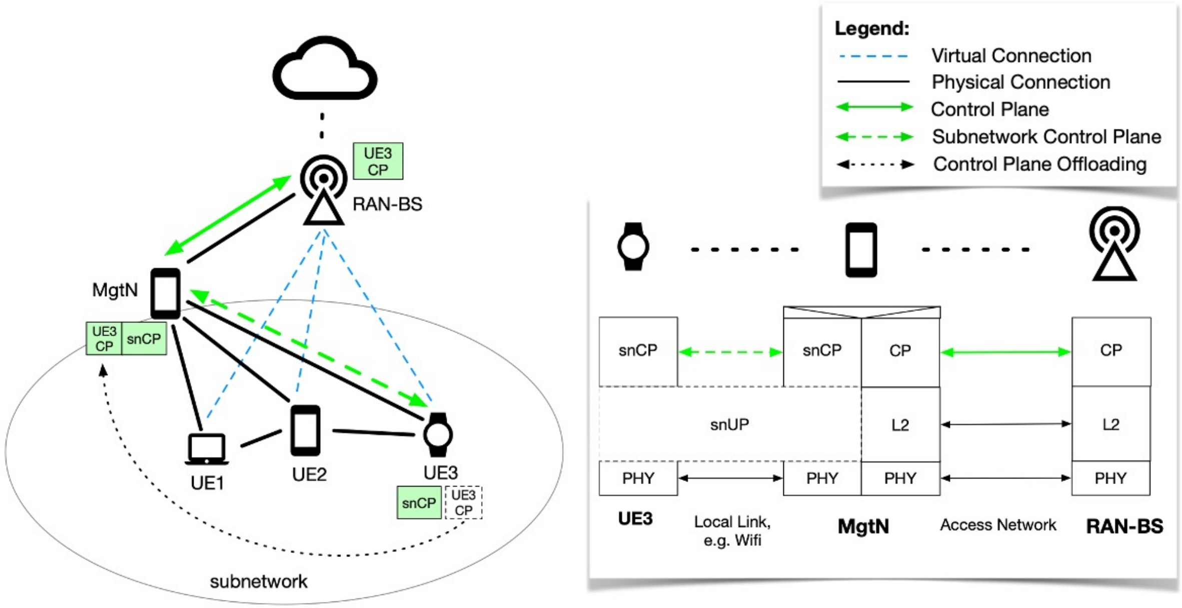
NTN architecture: Multi-hop ISL with the BS being on the ground



Subnetworks architecture

- Architectural options:
  - Transparent:
    - The subnetwork is unknown to the global NW.
    - The NW is unaware of the Management Node (MgtN) acting as man-in-the-middle.
    - The BS communicates virtually with the individual UEs, but the physical connection is always with the MgtN.
    - The MgtN needs to be capable of instantiating multiple UP/CP entities for all devices, as well as of sending and receiving on behalf of every device in parallel
    - The MgtN's device capabilities shall not be exceeded
    - Licensed spectrum cannot be used in the subnetwork
  - Non-transparent:
    - The subnetwork is known to the global NW
    - The NW is aware of the MgtN and its role (e.g., manage local devices and aid other UEs in the subnetwork)
    - UP for all UEs should be handled through the MgtN-BS link
    - BS-MgtN coordination is enabled: For example, licensed spectrum can be used in the subnetwork, data of different Ues can be multiplexed, mobility procedures can be simplified
- CP procedures of local devices can be flexibly deployed on the MgtN
- Lightweight SubNW CP (snCP): transparent to the NW
- Higher UP layers may terminate at the local device. Lower UP layers may terminate at the MgtN.

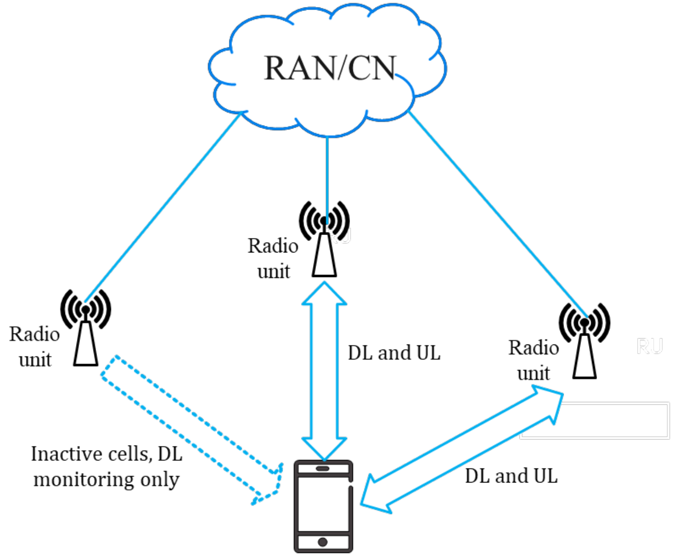
Subnetworks architecture: UE CP deployed at the MgtN, use of snCP within the subnetwork



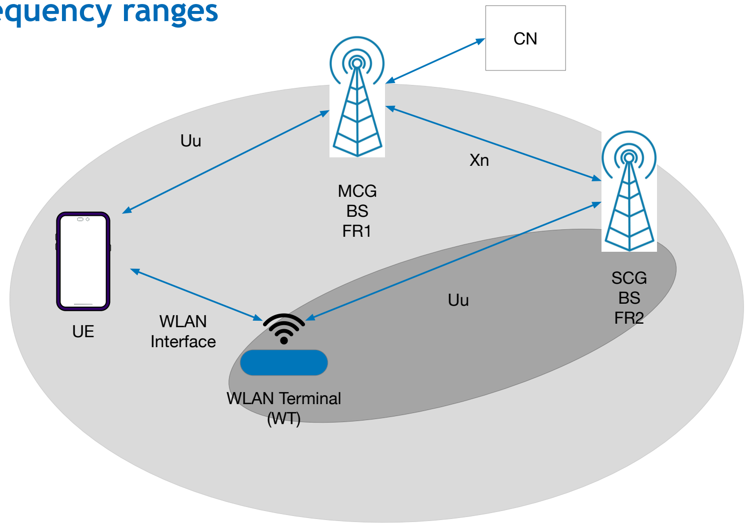
Multi-connectivity

- Multi-connectivity enables simultaneous connection to different carriers, which may belong to physically separated nodes, different radio access technologies or access networks
- **KPI improvements:** Increased user and system data rate, increased robustness and reliability, more efficient management of network resources
- **Relation to other enablers:** Mobility procedures should take the multi-connectivity proposals into consideration
- **Requirements**
  - Depending on the CA/DC solution, new interfaces and protocols between nodes may be needed, which may lead to an increased complexity in coordinating different NW nodes
  - Faster addition of cells compared to 5G is required
  - Need for a more flexible use of the UL so that the SCell may take over the role of control signalling in the UL
  - New mechanisms and procedures for aggregation of different radio technologies, such as 6G cellular and WLAN, should be defined. This would impact the layers where the aggregation takes place, e.g., 6G RAN, transport
- **Standardisations & regulations**
  - RRC protocol modifications and procedures for CA/DC evolution could be defined in 3GPP RAN2
  - RAN protocols for aggregation with other RANs could also be defined in 3GPP RAN2.

6G multi-connectivity solution overview

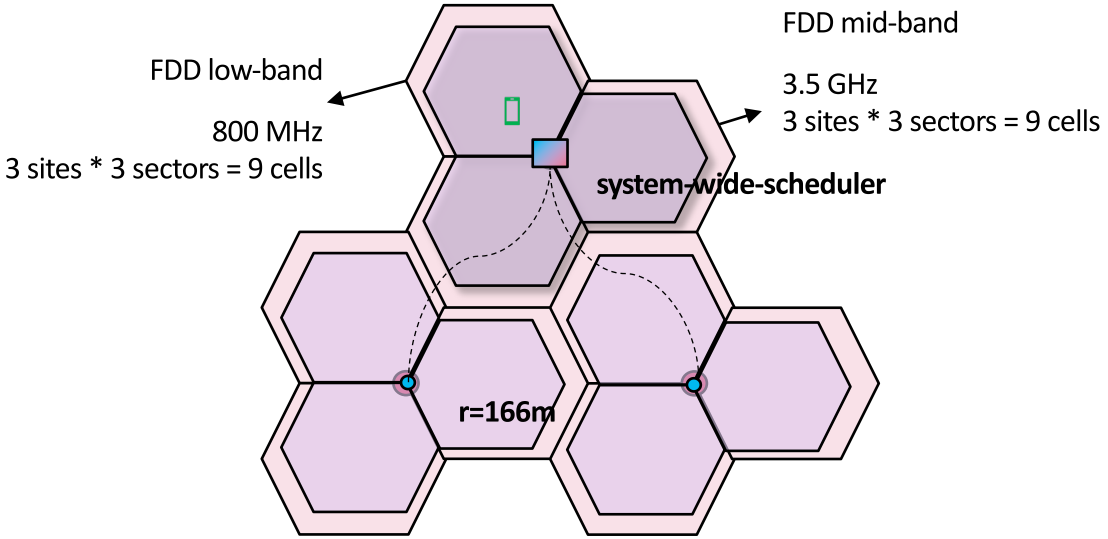


WLAN-Cellular Aggregation (WCA): The UE and the WT are connected to different BSs with different frequency ranges



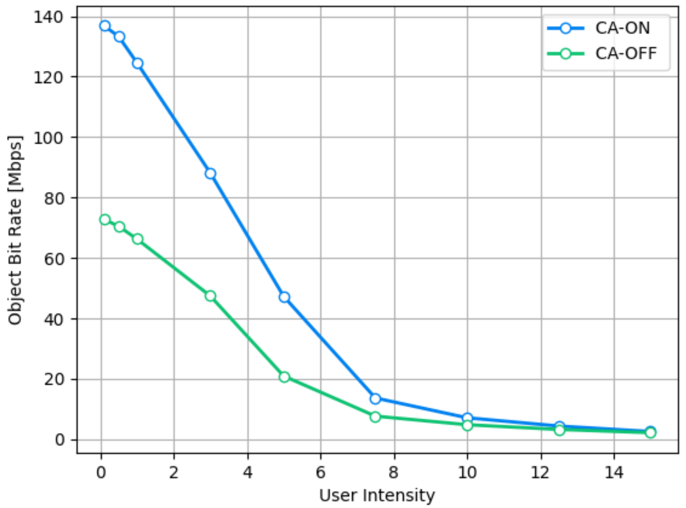
Carrier Aggregation Evaluation Scenario

	Reference scenario		Realistic scenario	
Band	Low-band	Mid-band	Low-band	Mid-band
Carrier frequency	800 MHz	3500 MHz	800 MHz	3500 MHz
Bandwidth	10 MHz	10 MHz	10 MHz	100 MHz
Tx/Rx antennas at gNB	2	16	2	16
FDD or TDD	FDD	TDD	FDD	TDD

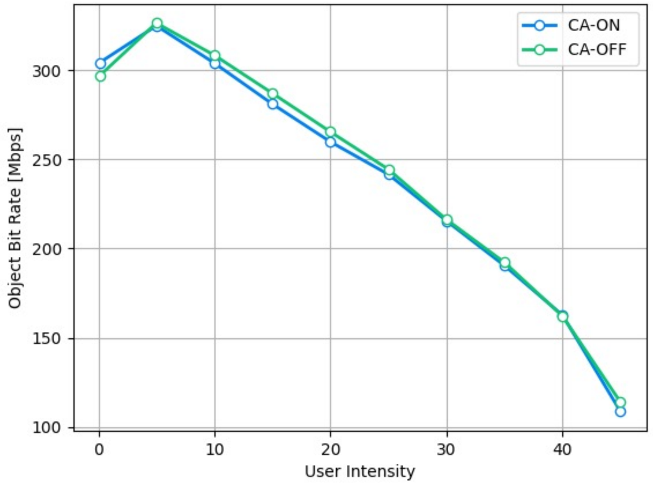


- PCC: If mid-band carrier's RSRP > -110dBm, then PCC is the mid-band carrier; else PCC is the low-band carrier
- CA-OFF: Only the PCC is activated
- CA-ON: Both the PCC and the SCC are activated
- Cell radio: 166m
- Traffic: 10 MBytes FTP download
- Min RSRP to select or add a carrier: -110 dBm
- Results: User bitrate vs user intensity (i.e., mean number of users entering the system per second)
- No big gain for the realistic scenario, because the SCC adds a small percentage of extra BW
- Balanced BWs are required among CCs

Reference scenario



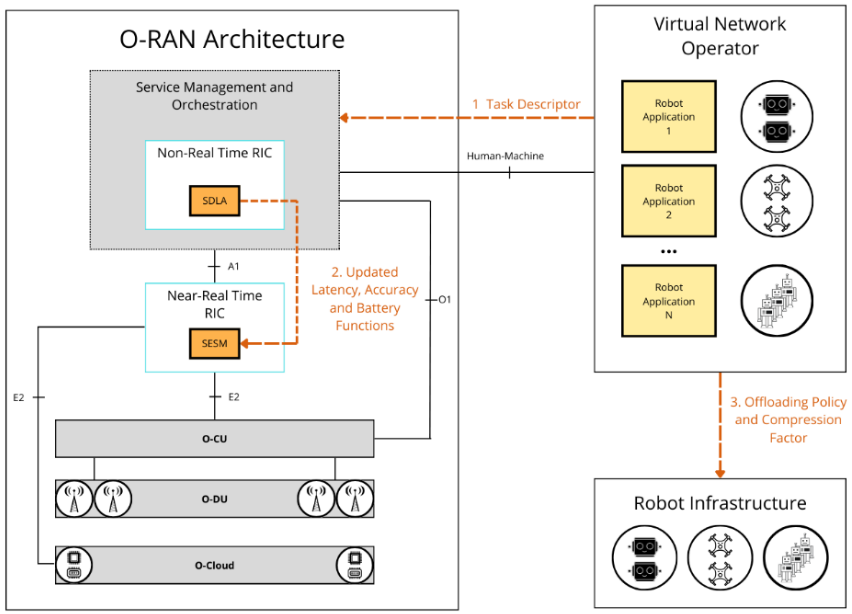
Realistic scenario



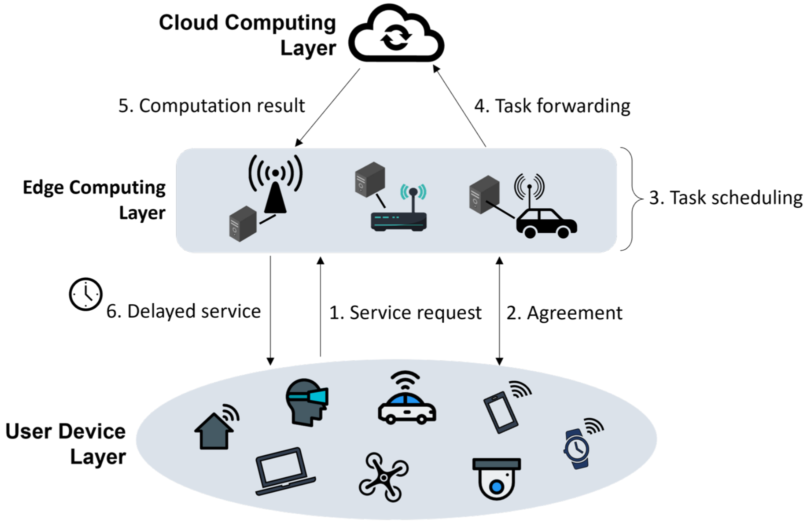
E2E context awareness management

- Mechanisms to allow network components to dynamically adapt to the context to ensure the expected E2E QoS.
- **KPI improvements:** Reduced network overhead, optimal allocation of edge resources
- **Relation to other enablers:** Privacy preserving techniques should be used to perform cross-network function distributed ML model training to fuse the context model with other network models
- **Requirements**
  - Different network components (e.g., RAN, CN, transport) should become aware of the context and need to interact, implying the need for signalling and synchronization
  - A resource orchestrator would have to create an abstracted view of transport resources and to employ a software-defined transport controller for resource management, ensuring that the required QoS is met
  - Effective resource allocation and orchestration mechanisms that operate even when incomplete or partial context awareness is available should be designed
  - Mobile robotic context: a correct resource exposure of the robot status should be ensured. Automatic translation of requirements from one component to another would be required.
- **Standardisations & regulations**
  - Task allocation in semantic RAN may have an impact on the ETSI MEC architecture, on the harmonization with 3GPP and on the O-RAN architecture

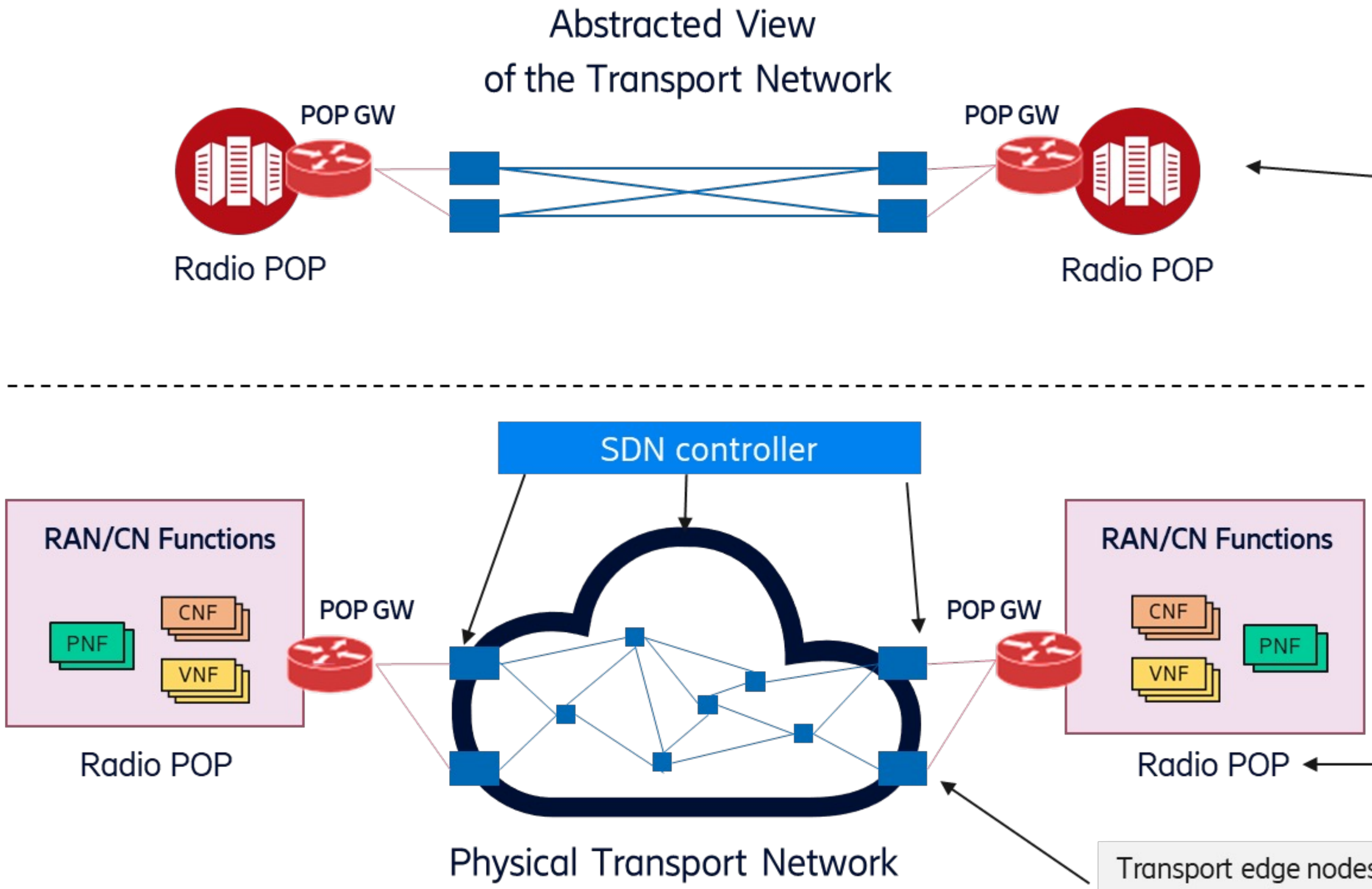
O-RAN architecture: Integration of the semantic context of a robotic task



Delayed computing paradigm



Transport network abstraction



The abstracted view is a **logical network** constituted by the **edge nodes** of the physical transport network connected by **virtual links** which capability, in terms of performances (bandwidth, latency, resiliency…), summarizes the "**cumulated capabilities**" of the real links between the respective end nodes.

Internal site connectivity can be done by one or more switches. The internal switches(s) could also include GW functions

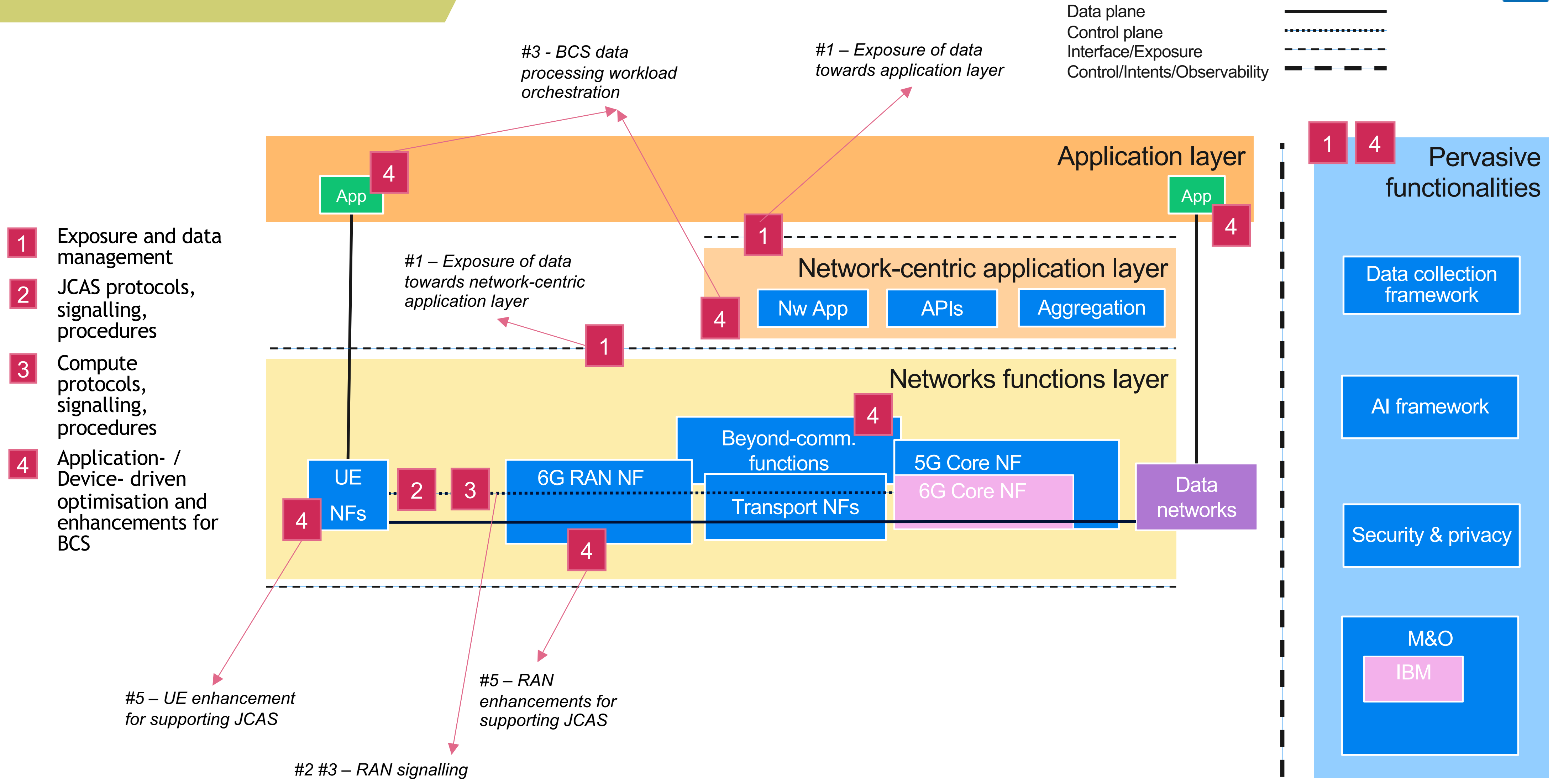
Radio **POP** concept represents an internal site. [ETSI reference]

Transport edge nodes exposed at the abstraction level.

Chapter 6

Networks beyond communication

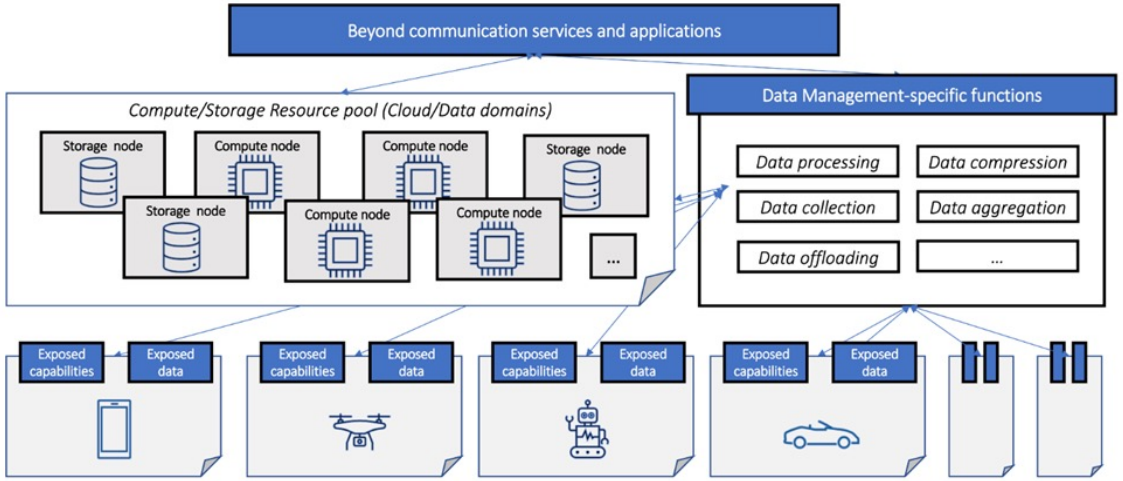
Mapping of networks beyond communication on system blueprint



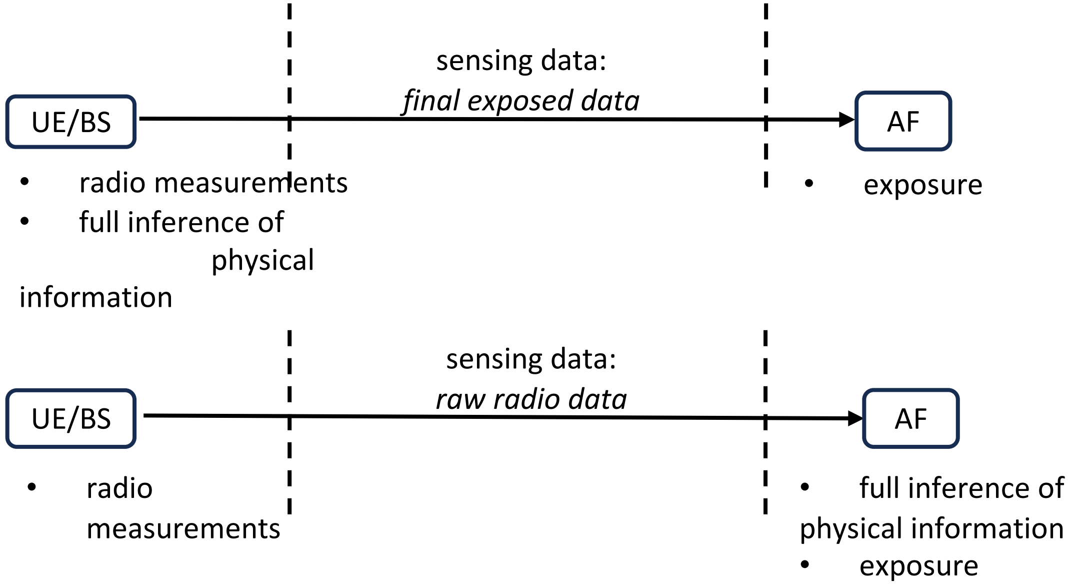
Exposure and Data Management

- **Description**
  - Encompasses mechanisms for the exposure, storage, processing, and trust management of data generated by various producers towards network-centric and application layers
- **Benefits**
  - **KPI Improvements**
    - Efficient data management leading to controlled data traffic/overhead, and higher data availability and low latency.
    - Increased number of vertical services consuming BCS data
    - Novel services exploiting BCS data contribute to safety, trust, and sustainability.
    - Enhanced scalability.
  - **Dependencies**
    - Compute offloading protocols, signalling and procedures, DataOps, MLOps, E2E context awareness management, Integration and orchestration of extreme edge resources in the computing continuum
  - **Design principles**
    - Support and exposure of 6G services and capabilities.
    - Flexibility to different network scenarios.
    - Network scalability.
    - Internal interfaces are cloud-optimized
    - Separation of concerns of network functions
- **Implications**
  - **Requirements**
    - Requirements for exposure of data collection, aggregation, and labeling between consuming functions in core and 3rd party applications.
    - Device capabilities exposure for beyond communications service resource control.
  - **Standardisations**
    - 3GPP TSG SA WG 1, TS 22.137, 26.531, 29.503, 23.288.
  - **Required resources**
    - Storage and computing resources for BCS data management.
    - Bandwidth for BCS data flow and E2E delay for URLLC use cases.

Exposure and data management enabler concept



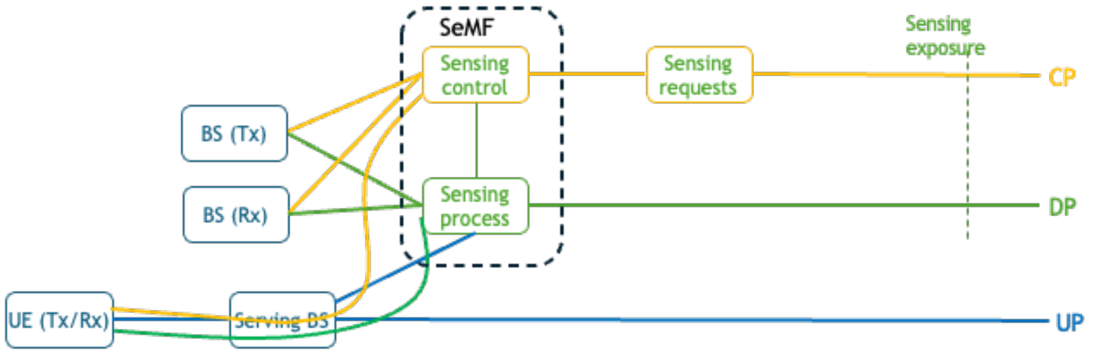
Canonical forms of sensing



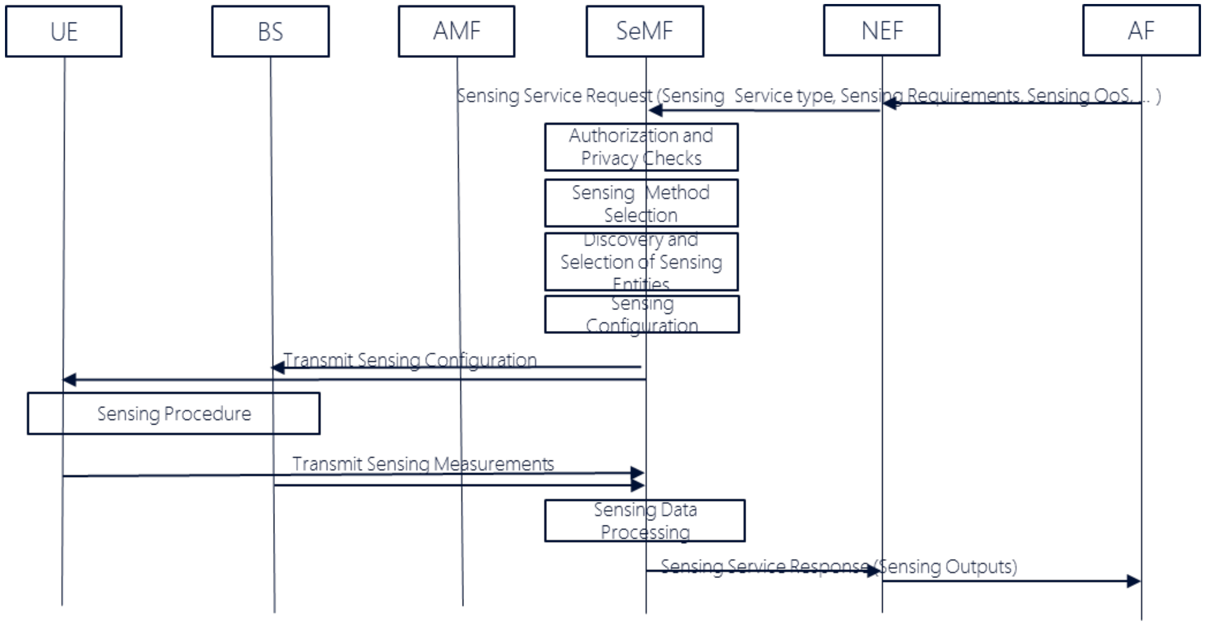
JCAS Protocols, Signalling, and Procedures

- **Description**
  - Enhances architecture and protocols to integrate sensing services in communication systems.
- **Benefits**
  - **KPI Improvements**
    - Enables high-quality sensing as a service.
    - Improves safety, environmental monitoring, and human motion monitoring.
  - **Dependencies**
    - Exposure and data management.
  - **Design principles**
    - Support and exposure for 6G services and capabilities.
    - Sensing flexibility and persistent security and privacy.
- **Implications**
  - **Requirements**
    - Integration of sensing capability to the RAN nodes.
    - Identification and management of sensing resources and sensing report.
    - Authorization, configuration of UEs for sensing, and facilitation of data transfer among RAN entities, UEs, and processing units for sensing data aggregation.
  - **Standardisations**
    - 3GPP TSG SA WG 1, TS 23.501, 23.502, 22.137 .
  - **Required resources**
    - Sensing, bandwidth, computation, and storage resources for sensing data management.

Functional architecture, network-based sensing with UE involvement for a bi-static sensing setup



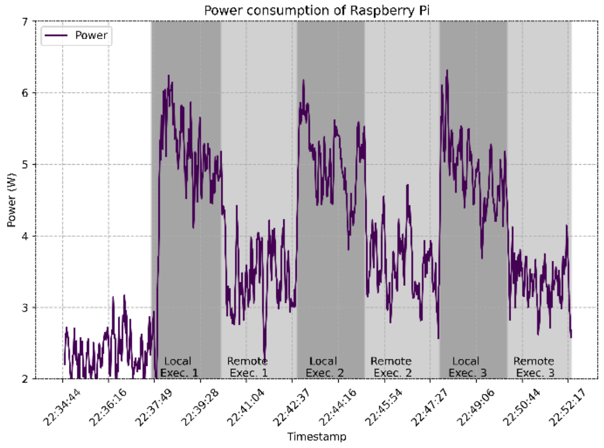
Example Deployment of Sensing Services in a Communication System



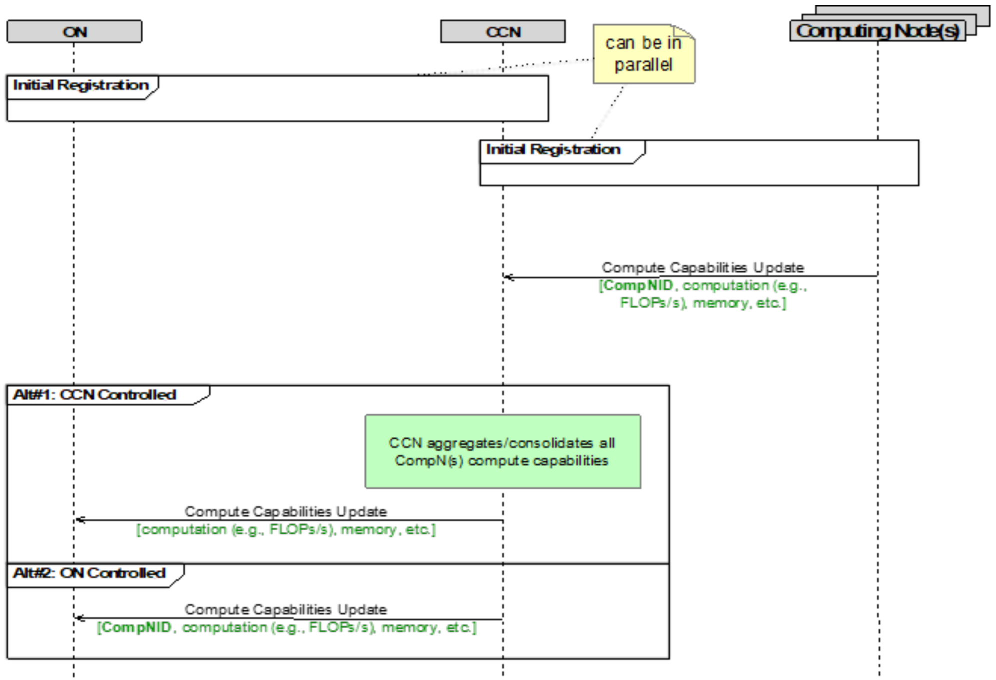
Compute Offloading Protocols, Signalling, and Procedures

- **Description**
  - Supports compute offloading from mobile devices to network nodes to satisfy latency, trustworthiness, power consumption, and resilience without increasing communication protocol complexity.
- **Benefits**
  - **KPI Improvements**
    - Energy savings
    - Increased application scalability and reduced latency
    - Enhanced privacy/security
    - Computation resiliency.
    - Quality of computation/ data accuracy
    - Device complexity
  - **Dependencies**
    - Exposure and data management.
    - Architectural means and protocols
    - Integration and orchestration of extreme edge resources in the computing continuum
    - Multi-domain/Multi-cloud federation
  - **Design principles**
    - Resilience and availability.
    - Support and exposure of 6G services.
- **Implications**
  - **Requirements**
    - Need for seamless connection between device offload functions and network offload functions.
    - Define RAN signalling and procedures
    - Adjust core signalling.
  - **Standardisations**
    - 3GPP TS 23.501, RRC 38.331.
  - **Required resources**
    - Compute resources at various network locations.

Dynamic offloading of a CPU demanding operation from the device to the network



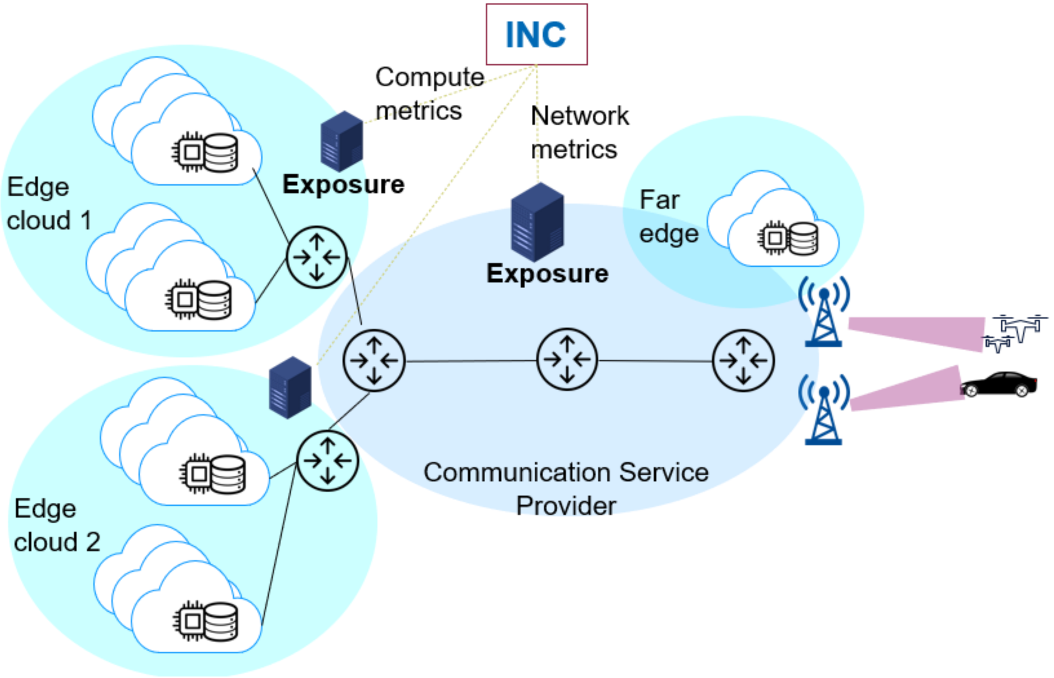
Node Discovery



Application-/Device-specific BCS Data Consuming Functions

- **Description**
  - Optimizes resource allocation and placement for BCS data consumers, improving data privacy, security, and efficiency.
- **Benefits**
  - **KPI Improvements**
    - Enhances data privacy, security, and network efficiency.
    - E2E computation and communication delay, network load, BW/energy efficiency.
    - Supports a higher number of services with optimized JCAS capabilities.
  - **Dependencies**
    - Exposure and data management.
    - Protocols, signalling and procedures
    - Distributed CaaS,
    - AIaaS, Intent based Management (Zero Touch),
    - E2E context awareness management, MLOps, DataOps
  - **Design principles**
    - Scalability
    - Efficient data management and exposure for various AFs
    - Emphasis on security and energy efficiency.
- **Implications**
  - **Requirements**
    - Advanced algorithms and robust data exposure mechanisms required.
    - Efficient resource allocation strategies.
    - New UE roles in JCAS, coordination with infrastructure for sharing sensing data and utilizing radios for enhanced sensing
  - **Standardisations**
    - 3GPP TS 23.501, 23.502, 23.503, 23.288, IEEE 802.11 az.
  - **Required resources**
    - Computational resources for data processing and AI/ML tasks.

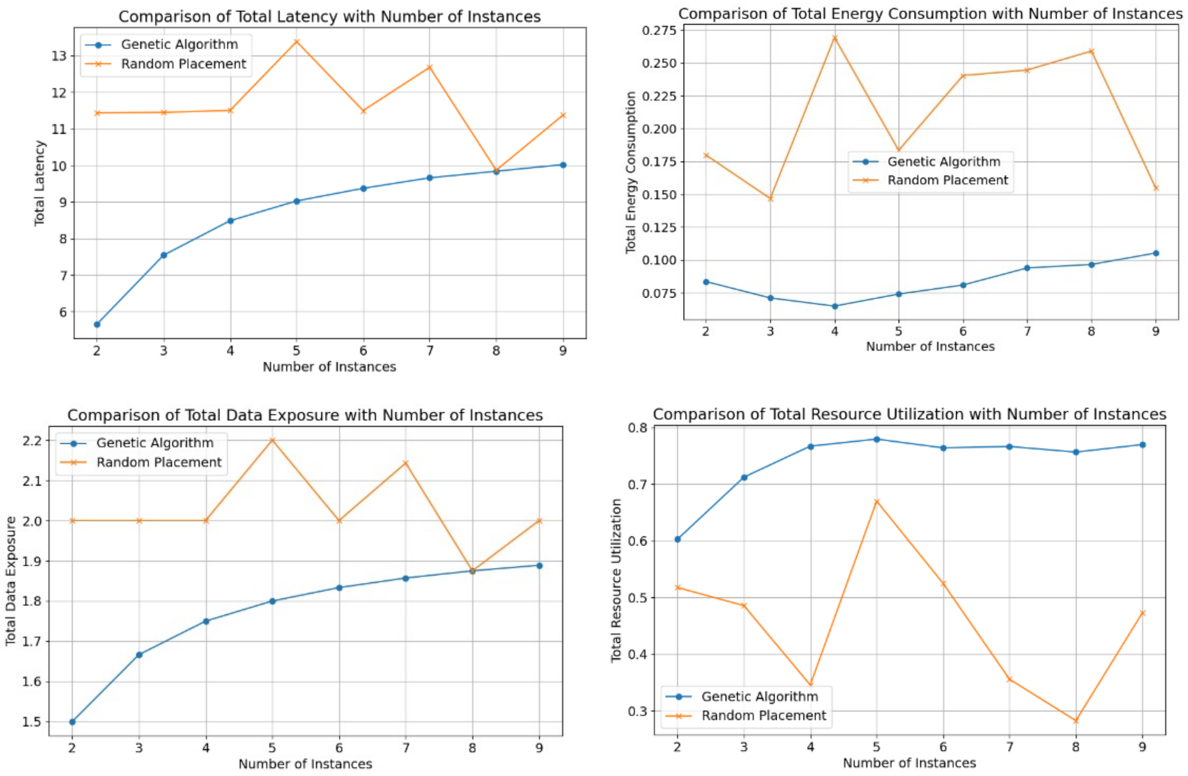
INC (Integration of Network and Compute) server collects network and compute metric to decide optimized placement of application



BCS data consumer application-driven optimization

- Background of this study
  - Evaluation of the efficacy of application placement strategies within a simulated network environment; focus on optimizing BCS data consuming application function placement to meet various QoS requirements.
- Description
  - Utilizing a Python-based simulation, this study explores genetic algorithm (GA) versus random placement strategies across multi-layer network nodes with varying resources.
  - The optimization of application instance placements based on latency, energy, data exposure, and resource utilization is assessed.
- Key take aways
  - Optimization Efficiency: GA improves network performance significantly by optimizing application placements, beneficial for managing 6G networks.
  - Resource Utilization: GA optimizes network resources, avoiding congestion and underutilization, crucial for 6G's high-demand environments.
  - Latency and Energy: GA reduces latency and energy use, important for IoT and edge computing's efficiency and responsiveness.
  - Data exposure: GA enhances security by keeping applications close to data sources, reducing exposure and increasing privacy in 6G networks.
  - Sustainability: GA's efficient placements lower energy use, aiding telecom's shift toward greener operations in the 6G era.
  - Strategic Planning: The study supports using advanced algorithms like GA for strategic network design, ensuring diverse QoS needs are met efficiently.

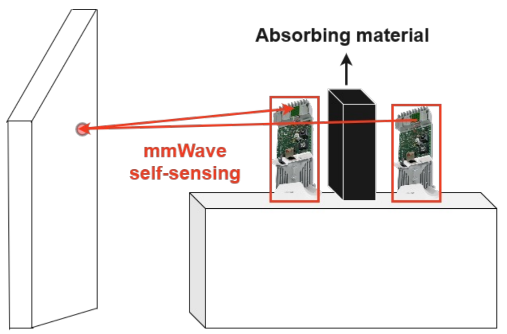
Application component placement optimisation evaluation: E2E Latency (a), Energy Consumption (b), Data Exposure (c), Resource Utilisation (d)



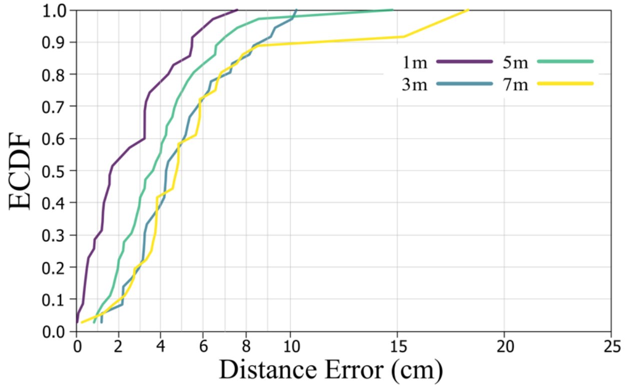
Enhancing JCAS capabilities – Indoor mapping

- Background of this study
  - Evaluation of the feasibility of the self-sensing concept via the development of a simple prototype: Two mmWave devices with the same orientation are placed on a table, separated by absorbing material that eliminates undesired paths, prevents interference, and increases the directivity of the antennas
- Description
  - In order to grasp the capabilities and constraints of the self-sensing concept, an analysis is performed in order to delve into the accuracy of the estimations relative to the distance between the devices and the walls, and the angle assessments.
  - An assessment takes place of the estimated distances by positioning the self-sensing system facing a wall and progressively increasing the distance between mmWave devices and the wall, from 1 meter to 7 meters, with a 2-meter step.
  - For angle estimations, Angle of Arrival (AoA) measurements are conducted, at various angles (-40 to 40) at distinct positions from the wall i.e.1m, 3m, and 5 m. For every distance and angle estimation, 50 measurements are conducted.
- Key take aways
  - With the self-sensing concept centimetre-level precision can be attained across all distances, maintaining distance error under 10 cm in 80% of instances
  - The FTM distance error rises with increasing distance
  - The farther from an obstacle, the greater the likelihood of encountering second-order or higher reflections within the system, leading to larger errors, as with yellow line (7 m) indicates
  - The angle estimation error also escalates with the enlargement of the rotation angle of the mmWave devices
  - Angle estimations indicate that in self-sensing the mmWave devices should consistently face the object/material surface at angular orientations ranging from ±40, and the estimated distances indicate that we can obtain up to 7 meter-range, which is comparable to commercially available Lidars.

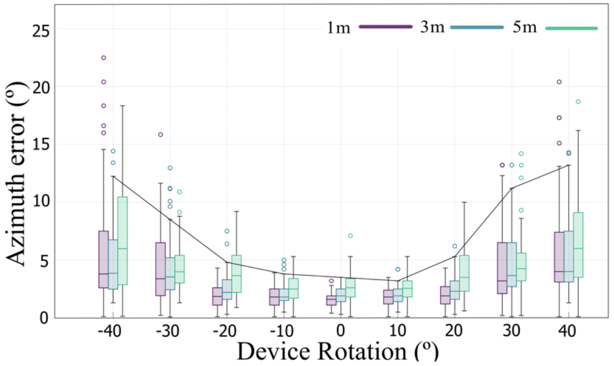
Self-sensing prototype



ECDF for the FTM distance errors



Azimuth errors for different rotation



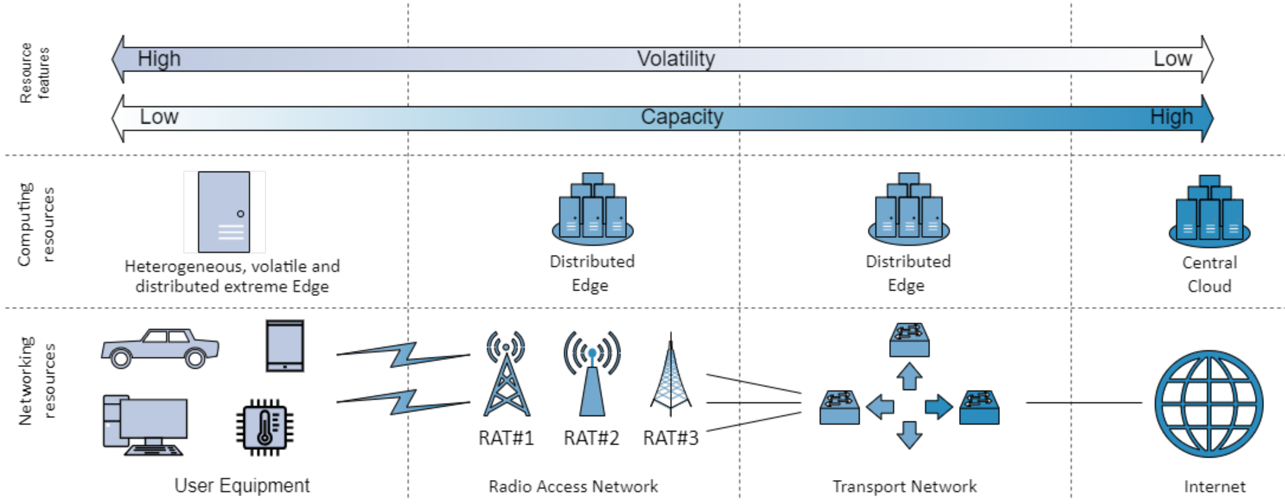
Chapter 7

Virtualization and Cloud transformation (IoT-Edge-Cloud continuum enablers)

Integration and Orchestration of extreme-edge resources in the compute continuum

- Description
  - The integration and orchestration of extreme-edge resources in the Compute Continuum require the definition of dedicated architectural interfaces and components
    - seamless orchestration and management of the compute resources that are placed beyond the radio access segment of the network
    - address the key challenges that the introduction of this extreme-edge resources bring (mobility, volatility, compute and location constraints, etc.)
  - The extended Cloud Continuum could be used to run vertical applications, but also network functions and modules.
- Benefits
  - KPI improvement
    - Latency, Energy consumption, Service availability, Increase data privacy.
  - Design principles improvement
    - Full automation and optimization
    - Scenarios flexibility
    - Cloud-native approach: internal interfaces are cloud optimized
  - Dependencies
    - Programable network monitoring and telemetry
    - Orchestration mechanisms for the computing continuum
    - Programmable and flexible network configuration
- Implication
  - Requirements
    - UE roles and responsibilities, coordination and interaction of the UE with the functions in the continuum, UE resource orchestration and management.
  - Standardisation
    - 3GPP TS 23.558 [23.558]
    - ETSI GR MEC 036 [MEC035]
  - Resources
    - Cloud, edge and extreme-edge computing resources control and usage will be jointly automated and optimized

Volatility and Capacity of the compute and network resources in the compute continuum

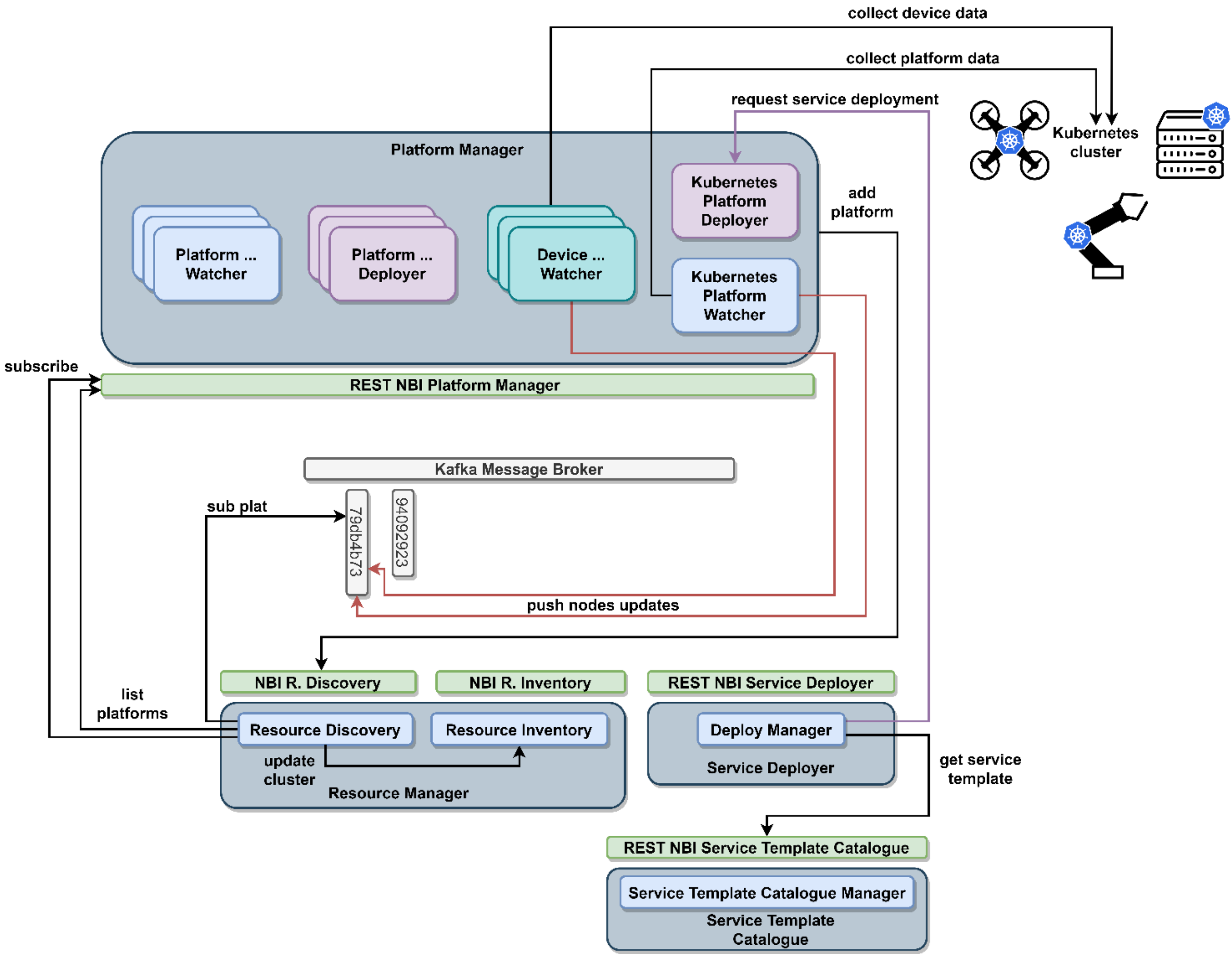


Cloud Continuum made of centralized cloud resources, distributed edge resources, and heterogeneous, mobile, and volatile extreme-edge resources.

Management of continuum resources for E2E service orchestration

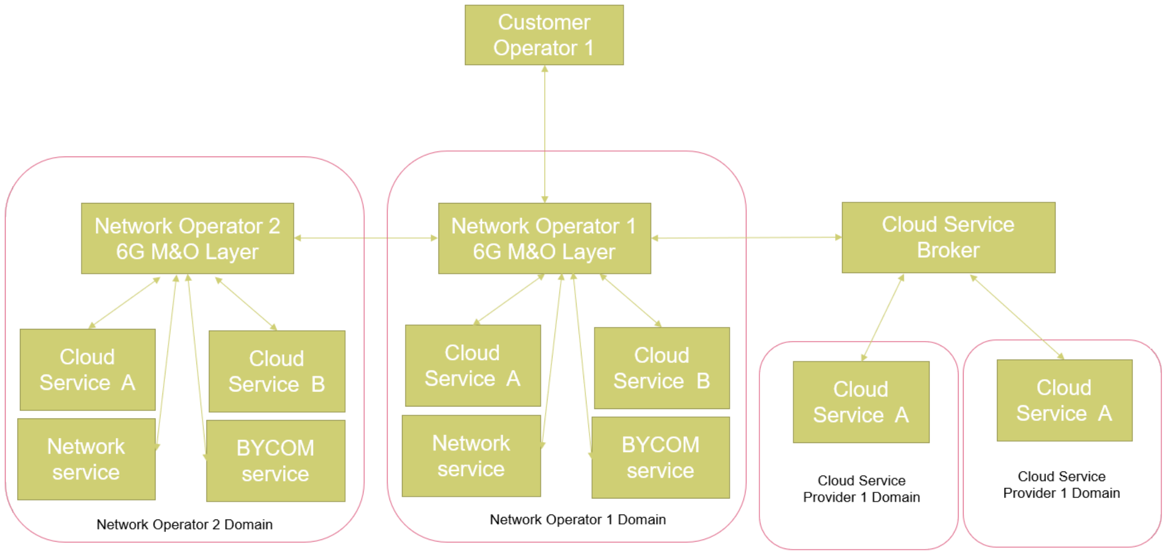
- Background of this study
  - The resources across the extreme-edge, edge and cloud continuum and their programmability become key with 6G and they can be leveraged to efficiently deploy and distribute applications that have proximity constraints/requirements (including AI).
- Description
  - A Continuum Multi-Technology Management and Orchestration Platform could be designed to address the challenges that the coverage of the extreme-edge resources introduces on the Compute Continuum
    - Taking into account heterogeneous virtualization platforms (e.g., Kubernetes, K3s, Microk8s, OpenStack, etc.) …
    - … and extreme edge devices (e.g., IoT devices, Sensors and Actuators, Robots and Cobots, etc.)
    - improving the placement mechanisms to efficiently deploy, migrate and distribute network functions that have proximity constraints and/or requirements exposing a unified interface
    - automating the discovery mechanisms of virtualization platforms and extreme edge devices information
- Key take aways
  - The design of the Continuum Multi-Technology Management and Orchestration Platform has been completed
  - The implementation is being carried out in the context of task 6.3
  - The orchestrator is being currently used in the PoC-B for the monitoring of the dedicated Kubernetes cluster and the cobots within it, and for the deployment of a network monitoring application and its associated closed-loop functions

High level software architecture of the compute continuum M&O



Multi-domain/multi-cloud federation

- The main aim of the federation model is to foster interconnectivity among diverse infrastructures. High level concept design and placement algorithms are studied to allow leveraging a distributed network of computing power, storage, and services.
- Benefits
  - A two-stage algorithm has been proposed to evaluate the number of needed Cloud Continuum locations to provide virtualized ultra-low latency service available 'anywhere'
  - A more conceptual part of the enabler explores the requirements of the federation model and how it could reduce complexity for cross-domain deployment
- Dependencies
  - Cloud continuum, 6G Core network and beyond communication functions module should provide intent-based interfaces
- Implication
  - Requirements
    - Intent-based interfaces for every component and M&O layer that is able to harmonize all the different parts
  - This enabler will impact the orchestration of network functions. Impacts includes but are not limited to 3GPP TS [23.501], [23.502].



Chapter 8

Proof of Concepts

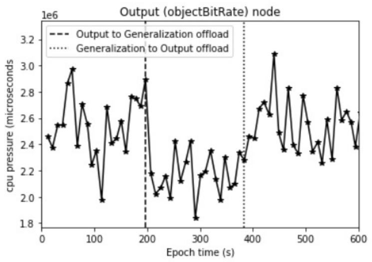
Component-PoC#B.2: Distributed Model Training and Inference

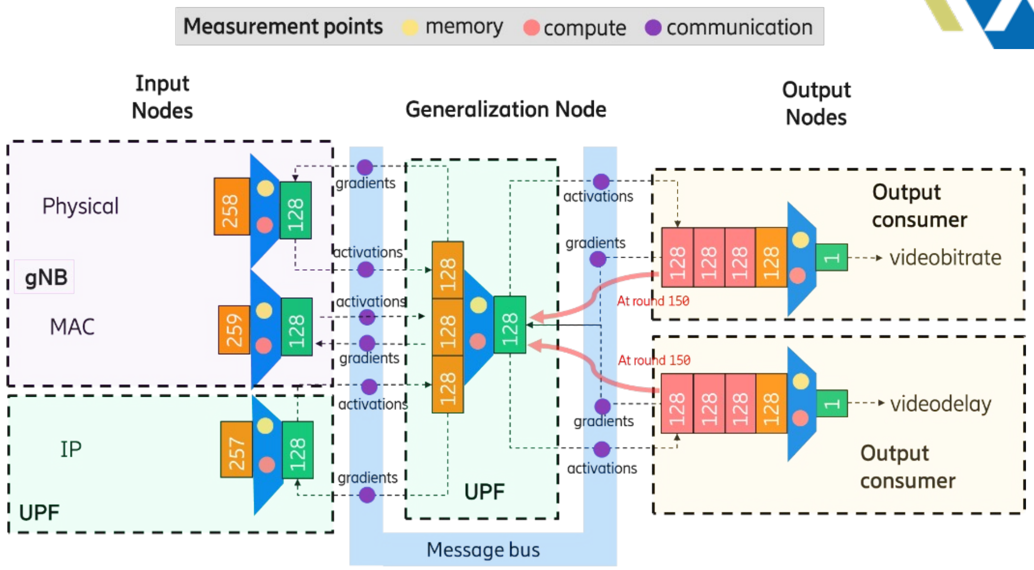
Background: Need for addressing research questions related to scenarios when datasets cannot be moved in between different network entities, and finding new and efficient techniques to enable collaborative distributed model training.

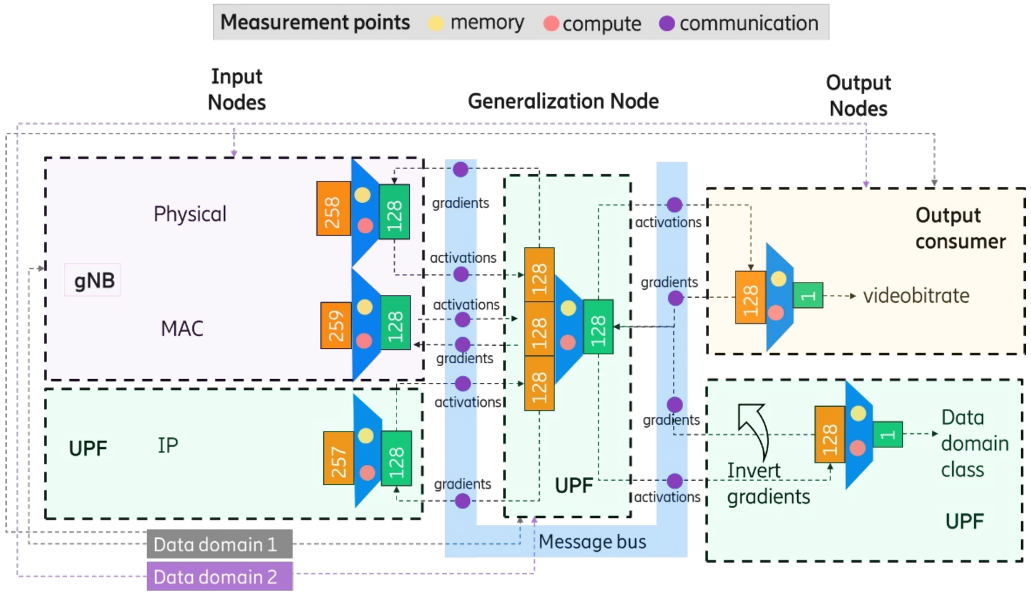
Description: This proof of concept (PoC) presents an AI-enabled technology called *split learning* and showcases benefits and trade-offs on different scenarios. The implementation is performed as a multi-headed, multi-tailed parallel split learning in Kubernetes. Neural networks that belong to different entities are placed in distributed manner to multiple Kubernetes pods.

Key take aways:

- Cross network function distributed ML model training
    - privacy preserving collaborative model training
    - enables data protection
    - improved accuracy (0.44→0.68) as compared to training in isolation, i.e., without collaboration
    - sustaining the accuracy of centralized training (if data were to be moved)
    - data compression by encoding input attributes to lower dimensionality via neural network before transmission
    - reduced communication overhead
- Model generalization
    - Reduction in cost and memory is proportional to the number of use cases (for 2 use cases → 50% reduction)
    - Sustained model accuracy with joint optimization via split learning
    - Improved model accuracy by 11% as compared to direct model transfer in the case of missing labels at the target domain adaptation
- Neural Network model layer offloading (assuming less compute availabilities at the client)
    - Reduction in CPU pressure (16%) at the client where the model layer is offloaded from
    - Reduced memory footprint
    - Reduced training time 45% to 75% for two use cases
    - Sustained accuracy (but this cannot be generalized as it highly depends on the choice of use cases)



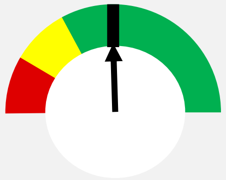




Summary and Conclusions

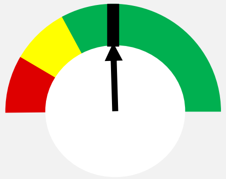
D3.3 progress summary

WPO3.1: Develop and analyse a 6G architecture framework and new innovative enablers for the beyond communications and data driven architecture, identify requirements a data-driven architecture will have on protocols, interfaces, data, and network nodes.



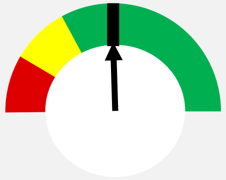
On track - Enablers for AIaaS, JCAS and Compute

WPO3.2: Define and analyse solutions that combine cloud technology flexibility with distributed processing nodes into self-contained modules with minimum dependency that can be used to extend and scale the network deployments in stepwise manner.



On track - New modular design of the network functions, transform the cloud to handle telco grade

WPO3.3: Develop and analyse new access for flexible topologies and local communications, including different types of multi-connectivity, node roles and node coordination, as well as design control and management solutions for programmable and context-aware transport



On track - Improve the network's flexibility via network of networks and context aware transport

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.